A Characterization of Online Search Behavior

Ravi Kumar Andrew Tomkins
Yahoo! Research
701 First Ave.
Sunnyvale, CA 94089, USA.
{ravikumar,atomkins}@yahoo-inc.com

Abstract

In this paper we undertake a large-scale study of online user search behavior based on search and toolbar logs. We identify three types of search: web, multimedia, and item. Together, we show that these different flavors represent almost 10% of all online pageviews, and indirectly result in over 21% of all pageviews.

We study search queries themselves, and show that more than half of them contain direct references to some type of structured object; we characterize the types of objects that occur in these queries. We then revisit the relationship search and navigation specifically in the context of e-commerce, and consider how search aids users in online shopping tasks.

1 Introduction

The online environment has shifted dramatically in the last fifteen years, with orders of magnitude growth in both users and content, as well as significant expansion of the capabilities users expect. Every day, new websites emerge seeking to transform online paradigms and woo users with new types of offerings. Social networking sites are skyrocketing in popularity and levels of user engagement, while traditional online communications paradigms like email continue to see significant usage. Search over pages, listings, and multimedia is increasing in usage and well as sophistication of result sets. And increasingly complex user tasks, from purchasing replacement laptop batteries to booking family vacations in foreign lands, are migrating from offline to online venues. We do not have a complete understanding of these dynamics, partly because the rate of change is high, and partly because there no publicly available data sources that offer a large-scale picture of cross-web user behavior.

In this paper we undertake such a study based on data collected through the Yahoo! toolbar. We analyze a large sample of over fifty million user pageviews collected over an eight-day period in March of 2009 from users who have installed the Yahoo! toolbar and agreed to collection of their data for purposes including this type of analysis. In addition, we augment this dataset in certain areas with data taken from Yahoo! search logs and editorially annotated in various ways.

Employing toolbar data as our primary form of data collection introduces some selection bias on the users with installed toolbars, and captures only web pageviews. Mobile usage is not included in our study, nor is use of AJAX for asynchronous fetching. Nonetheless, we provide a characterization of pageviews, which remains the dominant mechanism by which users interact with the web.

Our goal in this study is to characterize search behavior, which we interpret broadly. We do not restrict our attention to web search, but we consider also multimedia search as well as search of databases of items, as provided by the search boxes of eBay or Amazon. We also consider user behavior after leaving the search results page. Main web search represents 6.2% of all pageviews, multimedia search 1.4%, and item search another 1.4%. These 9% of search pageviews are responsible for an additional 8.9% of pageviews that are direct referrals, and yet another 3.5% of indirect referrals, resulting in a total of 21.4% of overall pageviews that are either on search, or reached through search.

Bulletin of the IEEE Computer Society Technical Committee on Data Engineering

Thus, search is "responsible" for roughly one in five pageviews online, so it is of great interest to understand exactly what these search queries cover. There have been a wide range of studies of search query types starting with the seminal work of Broder [3]. We choose to explore queries through the lens of structured data, with an eye to understanding the user appetite for structured information on the web. We will show that more than half of all queries contain a direct reference to a structured object that is central to the query, in a sense to be made formal below. We also present a taxonomy of types of structured information, and give statistics of their occurrence frequencies in search logs. We suggest that the future of search will include increasing focus on gathering, understanding, and when appropriate serving structured information from these domains.

Based on the prevalence of product queries of different types, we explore this claim further through a small study of the connection between search and online commerce. We automatically identify the action of adding an item to a shopping cart, and then study how users arrive at these checkout pageviews. We show that 20% of checkout pageviews are reached through search. We also characterize behavior with respect to time spent before performing the final search of the session, then time spent between the search and entry to the e-commerce site, and finally time spent on the e-commerce site before reaching the checkout pageview. Pre-search, post-search, and on-site periods respectively represent on average about 1/2, 1/6, and 1/3 of total pageviews in the path.

The remainder of the paper is organized as follows. Section 2 covers related work. Then in Section 3 we present our taxonomy of online search, and study the connection between search and navigation. Section 4 presents a study of structured data on the web. Section 5 then studies specifically the relationship between search and e-commerce. Finally, in Section 6 we present our concluding thoughts.

2 Related work

There have been several approaches to model user interaction with search engines. For example, Lau and Horvitz [18] introduce a probabilistic Bayesian network in order to predict query-query topic transitions, by examining the context of the query along with inter-query time period. Radlinski and Joachims [23] identify sequences of queries on the same topic using features based on shared words in the queries; see Spink et al [24] for the problem of topic switching and multi-tasking in query sessions. Downey, Dumais, and Horvitz [9] introduce an expressive language for describing searching and browsing behavior on the client side. The language is based on a state-machine representation and provides a unified framework for analyzing general models of user behavior, including many server-side models that were studied earlier [7, 18, 15, 17, 23]. They also construct machine-learned models to predict the next action of the user using features derived from their behavior thus far. Very recently, Guo et al [11] and Chappelle and Zhang [6] propose Bayesian click models to infer relevance of search results.

Web browsing activity beyond search has also been extensively studied and modeled, e.g., [5, 19]. Bucklin and Sismeiro [4] use a server-side log to model the browsing behavior of visitors to a website; they focus on "stickiness" of the website and analyze repeat visit patterns. Park and Fader [22] develop a stochastic timing model of cross-site user visit behavior in order to use information from one site to explain the behavior at another. Johnson et al [14] study online search and browsing behavior across competing e-commerce sites. Recently, Dumais et al [10] study browsing behavior after the user departs the search engine and begins to follow an information thread through the web.

Search queries have been analyzed from a variety of viewpoints, including data mining query logs and analyzing user activities from query sessions. There is an extensive literature on mining query logs and user search activities for various search-related applications including query suggestions [8, 17], query expansion [7], and ranking [1]. We refer to the survey by Baeza-Yates [2] for further details. Identifying sessions and session boundaries in query streams has been a hot research topic, e.g., [13]. This is typically done by using an off-the-shelf classifier and developing features based on query terms and user activities, including dwell times and timeouts; see [12, 21, 20]. Jones and Klinkner [16] develop effective classifiers that go beyond timeout-based features in order to identify sessions and goals, and more fine-grained task and subtask boundaries.

3 Search and navigation

In this section we study the prevalence of different types of searches, and their impact on web navigation. We first introduce the toolbar dataset used to derive most of the results in this paper. The toolbar dataset contains a random sample of US users drawn from Yahoo! toolbar logs over a one week period from March 18, 2009 to March 24, 2009. The dataset has been scrubbed to remove information that may allow identification of any user in the set, and to remove any other sensitive material. The number of pageviews in the dataset is around 50M, with around 7.2M pageviews per day over the weekdays and around 6.5M per day over the weekends. Certain of our results also employ auxiliary datasets generated from Yahoo! search logs; we describe these datasets as we introduce them.

3.1 Search result pageviews

Pages containing search results may be broken into three key types: (1) *main search* represents pageviews on web search sites, restricted to pageviews that are part of web search, rather than other types of search, (2) *multimedia search* captures search for videos, images, music, or other forms of multimedia, and (3) *item search* captures search through a database of listings, as provided by the search boxes of Amazon, eBay, or Craigslist.

We selected a random sample of 1000 pageviews from the toolbar dataset, and manually classified each pageview into one of the three types of search pageviews above, or "Other." Search result pages overall represent 9% of total pageviews, larger than we anticipated. Search is dominated by main search (standard web search) at 6.2% of total pageviews, while multimedia and item search each represent 1.4% of pageviews.

Automatically recognizing search pageviews. To perform a larger-scale analysis of these pageviews, we develop simple automated recognizers for different types of searches based on hand-built rules. For main search, we construct recognizers for the five largest US search engines (as our toolbar dataset contains US pageviews): Yahoo, Google, MSN, Ask, and AOL. For multimedia search, we construct recognizers for multimedia search result pages from the above search engines, along with Youtube, Hulu, Flickr, and Picasa. For item search, we construct recognizers for the URLs corresponding to search results from the most popular providers of listings in the dataset: Amazon, eBay, Craigslist, Imdb, Singlesnet, Careerbuilder, and Leboncoin. Additionally, we employ a set of general rules matching any URL that has a search-like parameter embedded in the URL, such as `&q=madonna` or `&search=madonna`.

Using the automated recognizers, we observe that 5.1% of pageviews are classified as main search, 1.5% as multimedia search, 0.5% as item search, and 1.7% as other searches. Observe first that these recognizers overall capture 8.8% of pageviews, and our earlier manual classification study showed that the correct fraction is 9% of pageviews ($\pm 1.8\%$ at 95% confidence). Thus, our recognizers possibly capture almost all search pageviews. Second, note that the fractions of pages we capture for main search and multimedia search are quite similar to the fractions of the overall population as determined by our manual study. Item listings, however, have a longer tail, and thus we capture a smaller fraction of these in our explicit recognizer.

3.2 Referrals

To begin, we explore the interactions between search and web navigation by considering the source of the referrer link that took a user to a particular page. 34.4% of pageviews have no such referrer, because the user arrived at the page by selecting a bookmark or by directly typing into the browser's navigation bar or by clicking on links from other applications, and so forth. We call these pageviews *starting points,* and all other pageviews *referrals.* 5.3% of all pageviews have a referral from main search, 1.4% from multimedia search, 0.6% from item search, and 1.5% from other search. Thus, each category of search pageviews tends to result in around one follow-on click on average.

Of course, that follow-on click may result in yet more downstream clicks. Consider for example a user who visits Yahoo!, performs a search, explores several search results, discovers a high-quality hub page, and continues to browse from the hub page. In this scenario, the object of study should be the *referral chain*. As a user can potentially follow multiple links from a single page (either through multi-tab browsing, or through use of the back button), the chain is in fact a tree. And as there can be multiple starting points within a session, the tree might become a forest. We refer to this artifact as the *referral forest* of a session.

In the context of a referral forest, we ask the following question: given that a user has arrived at a certain page, was search involved in getting the user there? More specifically, is there a search page in the path from the current page to the root of its tree in the referral forest?

16.2% of all pageviews have a main search page as an ancestor. This value is 3.3% for multimedia search, 0.9% for item search, and 4.4% for other search. In aggregate, 21.4% of pageviews have some type of search result page as an ancestor. Thus, about one out of six pageviews are reached directly or indirectly through web search, and more than one out of five pageviews are reached directly or indirectly through some kind of search. These calculations count the pageview for the search result itself, but do not count the pageview to navigate to the search engine.

Search session statistics. Given that search is in some way responsible for a significant fraction of pageviews, we present some basic statistics of a *search session*, defined as follows. We say that a pageview is a *search root* if it is labeled as main search, and does not have any ancestor in the referral forest labeled as main search. A search session for a particular search root is then taken to be the set of pageviews in the subtree rooted at the given search root. Notice that, by definition, search sessions may overlap with one another in time, but each pageview belongs to at most one search session. The mean number of search sessions per user per day is 0.57. The average number of pageviews in a search session is 6.3, and the average depth of the subtree rooted at a search root is 4.2. This implies that users do not simply perform a search and then explore the immediate results; they typically explore longer paths from the initial search results page.

In aggregate, then, a search session takes around 6 pageviews. Of these, one is the initial search, on average there is one additional search, and the remaining four pageviews are off-search. Figure 1 gives a more detailed breakout of this picture. It shows for each session length what fraction of pageviews is spent on the search engine, versus browsing on the rest of the web. As expected, longer sessions correspond strongly to more time spent engaging off the search engine with content on the rest of the web.

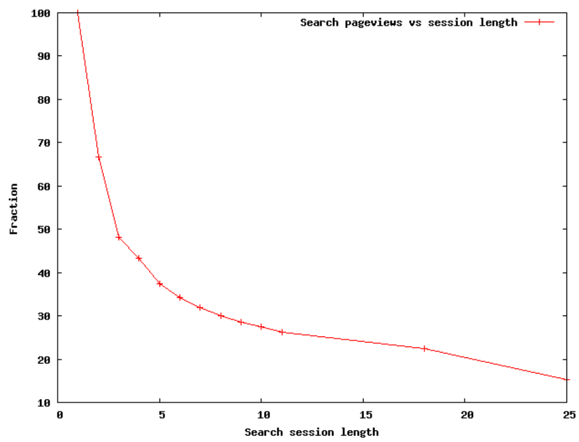


Figure 1: Fraction of search session pageviews spent at search engine, as a function of the total pageviews in session.

Assigning credit for pageviews. We have seen that search is a contributor to a significant fraction of pageviews (around one in five). However, any particular pageview may be reached by a complex chain of referrals of which search is only a piece. A different perspective may be gained by attempting to allocate credit for the pageview across the ancestors of the pageview. We consider the following two measures for assigning credit among the ancestors of a pageview. (1) *Root credit:* for each pageview, assign one unit of credit to the root of the referral tree in which the pageview appears. (2) *Amortized credit:* for each pageview, assign one unit credit, but spread this unit of credit across all ancestors of the pageview, with the root getting a small amount of credit, the next node getting twice as much, and so forth until the pageview itself is reached. Under the root credit measure, main search gets credit for 7.7% of pageviews, multimedia search for 0.5%, item search for 0.1%, and other search for 1.1%. For amortized credit, the corresponding numbers are 5.9%, 1.7%, 0.4%, and 1.9%.

Thus, the various approaches to assigning credit present a consistent picture: the fraction of search pageviews of each type are roughly akin to the fraction of direct referrals from search pageviews of each type, which are themselves roughly akin to the fraction of indirect referrals using both root and amortized methods of credit allocation. This suggests that, while there is likely to be a search pageview on the path to interesting content, such pageviews are a stepping stone along the way, rather than the sole attributable provider of access to key content.

4 Structured data in search

In this section we study the extent to which search queries include references to *structured objects*, or simply *SO*s. We will address this question using a dataset of search queries collected from the Yahoo! web search engine log in the third quarter of 2008. To form this dataset, queries were sampled randomly from the log, so the sample contains the appropriate fraction of head versus tail queries.

Examples of SOs include restaurants, products, cars, real-estate listings, cities, bands, sports teams, celebrities, sports players, and companies. To be considered an SO, a candidate should have non-trivial metadata available on the public web in a format that could be extracted for automated processing. Thus, it is possible that an entity might not be an SO today, but as more detailed metadata becomes available, it might become an SO per this definition at a later date.

One may draw a distinction between categories such as high-definition TVs versus instances such as the Samsung DLP8187, versus listings such as a particular Samsung DLP8187 being sold on eBay. The analysis below treats all three of these as SOs, as the intention is to provide some quantitative insights into the types of user needs that can be fulfilled more effectively based on repositories of SOs objects and technology for interpreting content through the lens of a universe of SOs. Understanding the particulars of listings versus instances or categories is an interesting area for future work.

We place each query into one of the following categories: (1) *Central SO.* The query directly references an SO, and furthermore, the SO is central to the context, rather than just a peripheral mention. For instance, in the query `south florida style long skirts`, there is a reference to the SO Florida, but the reference is peripheral; the primary object of interest to the user is a particular type of skirt. Likewise, in the query `disney coloring pages`, the reference to the SO Disney is peripheral. (2) *Peripheral SO.* The query directly references an SO, but the reference is peripheral to the central meaning of the query. (3) *Topic/Concept.* The query does not contain an SO, but contains a topic or concept of the type that might appear in Wikipedia. In some cases the distinction between a topic/concept versus an SO is subtle. Some examples of topics/concepts are Justice, Socialism, Hunger, I-80, the Vancouver Sky Train, the NFL Draft, and the Nuclear Test Ban Treaty. In time, some of these might become SOs as more and more metadata becomes available online. (4) *Other.* The query contains neither an SO nor a Topic/Concept.

In our search dataset, central SOs account for 52.9% of the queries, peripheral SOs account for 4.7%, and Topic/Concept account for 8.5%. Of the remaining 33.8% (i.e., other), about 10% are URL queries, and another roughly 12% are non-URL navigational queries. Thus, of queries that do not contain an SO or a Topic/Concept, about 2/3 are navigational in nature. The remainder are various types of informational and transactional queries. The most striking aspect of this part of the study is that an actual majority of search queries contain central references to an SO, suggesting that an increasingly sophisticated understanding of structured data is key to meeting the needs of search users.

For queries that are labeled SO central, we also break out a few sub-categories as follows: (1) *Event.* Generally something that takes place at a point in space and time, although we also consider recurring events and distributed events to belong to this category. Examples include concerts, conventions, elections, movie releases, and so forth. (2) *Games.* Board games, computer games, multi-player online games, etc. (3) *Notable person.* The SO is a notable person, such as a celebrity, an actor or actress, a sports figure, a well-known blogger, a politician or world leader, a prominent businessperson or scientist, and so forth. (4) *Ordinary person.* The SO is a person, but not one of broad interest. An ordinary person is expected to be of interest primarily to people who are personally acquainted with him or her. (5) *Specific product* and *general product.* The SO is a product, either referring to it specifically (e.g., Samsung DLP8187) or generically (e.g., HDTV). (6) *Places.* The SO is a geographical location, at any level of granularity. (7) *Business Cat/Service.* The SO refers to businesses or services (e.g., laundromat). (8) *Health.* SOs related to medical issues, such as conditions, symptoms, drugs, and treatments. (9) *Real estate.* SOs related to real estate, including apartments and rentals. (10) *Media title.* This SO refers to CD albums, movies, and other media releases. (11) *Organization.* The SO refers to any organization such as a company, non-profit organization, local or national government.

In measuring occurrences of these categories, we may encounter queries that contain multiple SOs. In this case, we assign weight $1/k$ to each of k SOs occurring in the query. Figure 2 shows the breakdown of SOs in terms these categories. There are four categories that occur in more than 10% of queries: Organization is the dominant category (occurring in fully 1/3 of queries) followed by notable person, specific product, and media title. Not coincidentally, these are areas in which search engines already invest significant effort to develop top-of-page informational boxes that present appropriate structured information allowing users to either fulfill their goal without leaving the page, or to engage with the object directly in order to pursue a particular need.

5 Connections between search and e-commerce

We have focused above on the interactions between search and navigation, and on the nature of search queries. General and specific product queries represent around 20% of overall search queries, and are of great interest as they represent a monetizable user intent. In this section we dig a little more deeply into the impact of search on online purchasing, using the toolbar dataset described earlier.

Category	% queries
Event	2.31
Games	1.15
Notable person	13.08
Ordinary person	4.42
Specific product	11.35
General product	8.56
Places	4.90
Business Cat/Service	7.69
Health	1.35
Real estate	1.15
Media title	10.10
Organization	33.94

Figure 2: Breakdown of SO central into object categories.

Due to concerns around sensitive information, we do not consider URLs that represent purchase behavior, but instead, we recognize the activity of adding an item to an online shopping cart. To perform this task, we construct a simple recognizer that identifies terms such as `shopping cart`, `checkout`, etc., that are present in the directory path/cgi-parameter of the pageview URL; we call these the *checkout pageviews*. Note that this is only an approximation since there is no sure way to ascertain if the transaction actually happened.

We will consider the role that search plays in reaching a checkout pageview. This analysis is necessarily incomplete—for instance, we do not capture the significant impact search may have on so-called ROBO purchase (research online, buy offline), in which users read reviews, compare models, check prices online, but consummate the purchase in person at a local store.

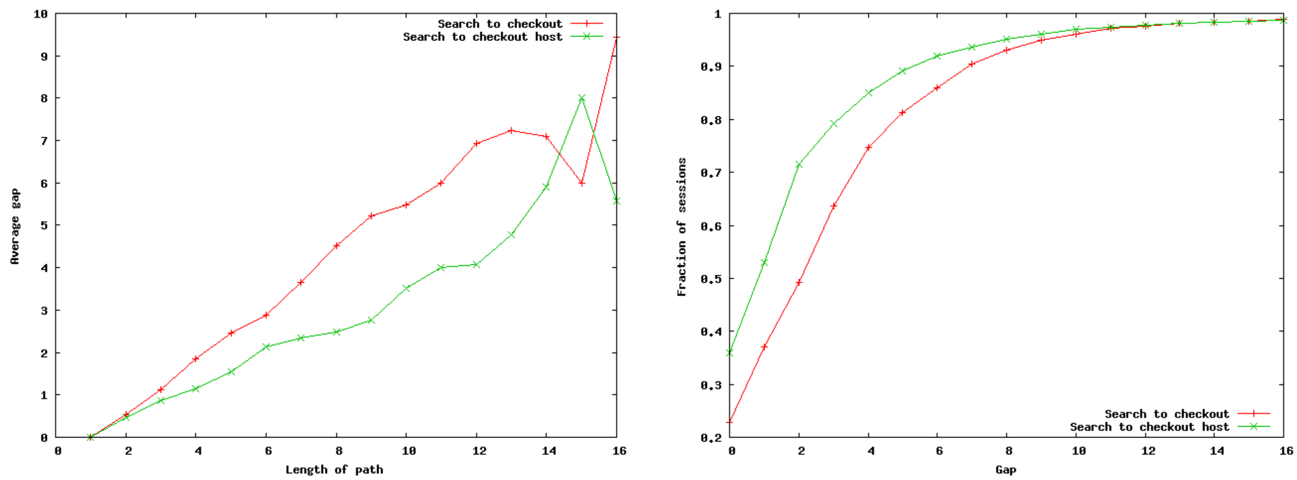
Search contributions to checkout pageviews. We consider each checkout pageview, and follow the referral chain backward to identify the first occurrence of a search pageview of any type. 20.1% of all checkout pageviews have a search ancestor, so about one in five checkout pageviews are reached directly or indirectly through search. 11.2% of checkout pageviews in fact have a search pageview as the root of the path.

Of the roughly 20% of checkout pageviews that have a search ancestor, 36% of these result from a search on the same host as the checkout, and the remainder represent an offsite search in which the user arrives at the e-commerce site through use of an external search engine.

Properties of paths to checkout pageviews. For checkout pages with an offsite search ancestor, we now consider the number of hops to get from the search page to the checkout page. The top, red curve of Figure 3(a) shows these values as a function of the overall path length from root to checkout pageview. The curve grows roughly linear, but with slope slightly above 1/2, indicating that on average a user encounters a search page slightly more than halfway through a session that culminates in a checkout pageview.

The lower, green curve in Figure 3(a) shows the gap from the search pageview to the first pageview on the host at which the checkout pageview occurs. The pattern is similar, but the slope is closer to 1/3. The gap between these curves represents the time spent on the host. The picture emerging is that conditional on performing an offsite search and then later a checkout pageview, the user typically spends approximately 2/3 of the intervening hops getting to the site, and the remaining 1/3 of hops getting to the page hosting the checkout pageview. These values are all based on averages, so a more detailed perspective would emerge from a study of the individual sessions.

Equipped with this view of behavior as a function of pathlength, we now turn in Figure 3(b) to the cumulative distribution of the gaps themselves. The lower, red curve shows the distribution of number of hops from search to the checkout pageview. The upper, green curve shows the distribution of number of hops from search to the host upon which the checkout pageview occurs. In 85% of sessions, the user reaches the host in four hops or fewer, and in 75% of sessions, the user in four hops or fewer reaches the checkout pageview itself.



(a) Average gaps as a function of the path length.

(b) Distribution of gaps.

Figure 3: Analysis of search to buy behavior.

6 Conclusions

We have shown that the search paradigm drives significant direct and indirect traffic. Roughly 90% of search queries are either navigational, or contain reference to some structured object, topic, or concept. These are dominated by queries that contain a direct central reference to an object. Our exploration of the paths taken by users from search to checkout pageviews, in agreement with other work (e.g., [10]), shows that a significant amount of effort ensues between the search query and the fulfillment. We conclude that the impact of search on online activity is more significant than we had anticipated on undertaking this work, and that the impact of structured data on search, while already large, leaves significant headroom for future extensions of the search product.

Acknowledgments

We gratefully acknowledge ConnieAlice Hungate, Amy Holzer, Anneliese Eisentraut, Sarah Ellinger, and Teresa Block for providing the data used in Section 4.

References

[1] E. Agichtein, E. Brill, S. Dumais, and R. Ragno. Learning user interaction models for predicting web search result preferences. In *Proc. of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 19–26, 2006.

[2] R. Baeza-Yates. Applications of web query mining. In *Proc. of the 27th European Conference on Information Retrieval*, pages 7–22, 2005.

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

[4] R. E. Bucklin and C. Sismeiro. A model of web site browsing behavior estimated on clickstream data. *Journal of Marketing Research*, 11:249–267, 2003.

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

[6] O. Chappelle and Y. Zhang. A dynamic Bayesian network click model for web search ranking. In *Proc. of the 18th International Conference on World Wide Web*, pages 1–10, 2009.

[7] H. Cui, J.-R. Wen, J.-Y. Nie, and W.-Y. Ma. Probabilistic query expansion using query logs. In *Proc. of the 11th International Conference on World Wide Web*, pages 325–332, 2002.

[8] H. Daume and E. Brill. Web search intent induction via automatic query reformulation. In *Proc. of the North American Chapter of the Association for Computational Linguistics – Human Language Technologies*, pages 49–52, 2004.

[9] D. Downey, S. Dumais, and E. Horvitz. Models of searching and browsing: Languages, studies, and applications. *Journal of the American Society for Information Science and Technology*, 58(6):862–871, 2007.

[10] D. Downey, S. Dumais, D. Liebling, and E. Horvitz. Understanding the relationship between searchers' queries and information goals. In *Proc. of the 17th ACM Conference on Information and Knowledge Management*, pages 449–458, 2008.

[11] F. Guo, C. Liu, A. Kannan, T. Minka, M. Taylor, Y.-M. Wang, and C. Faloutsos. Click chain model in web search. In *Proc. of the 18th International Conference on World Wide Web*, pages 11–20, 2009.

[12] D. He, A. Goker, and D. J. Harper. Combining evidence for automatic web session identification. *Information Processing and Management*, 38:727–742, 2002.

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

[14] E. J. Johnson, W. M. Moe, P. S. Fader, S. Bellman, and G. L. Lohse. On the depth and dynamics of online search behavior. *Management Science*, 50(3):299–308, 2004.

[15] R. Jones and D. Fain. Query word deletion prediction. In *Proc. of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 435–436, 2003.

[16] R. Jones and K. L. Klinkner. Beyond the session timeout: Automatic hierarchical segmentation of search topics in query logs. In *Proc. of the 17th ACM Conference on Information and Knowledge Management*, pages 699–708, 2008.

[17] R. Jones, B. Rey, O. Madani, and W. Greiner. Generating query substitutions. In *Proc. of the 15th International Conference on World Wide Web*, pages 387–396, 2006.

[18] T. Lau and E. Horvitz. Patterns of search: Analyzing and modeling web query refinement. In *Proc. of the 7th International Conference on User Modeling*, pages 119–128, 1999.

[19] A. L. Montgomery and C. Faloutsos. Identifying web browsing trends and patterns. *IEEE Computer*, 34(7):94–95, 2001.

[20] H. C. Ozmutlu. Automatic new topic identification using multiple linear regression. *Information Processing and Management*, 42(4):934–950, 2006.

[21] H. C. Ozmutlu and F. Cavdur. Application of automatic topic identification to Excite web search engine data logs. *Information Processing and Management*, 41(5):1243–1262, 2005.

[22] Y.-H. Park and P. S. Fader. Modeling browsing behavior at multiple websites. *Marketing Science*, 23(3):280–303, 2004.

[23] F. Radlinski and T. Joachims. Query chains: Learning to rank from implicit feedback. In *Proc. of the 11th International Conference on Knowledge Discovery and Data Mining*, pages 239–248, 2005.

[24] A. Spink, M. Park, B. J. Jansen, and J. Pedersen. Multitasking during web search sessions. *Information Processing and Management*, 42(1):264–275, 2006.