

Multitasking Web Search on Alta Vista

Amanda Spink* &
Minsoo Park
School of Information
Sciences
University of Pittsburgh
610 IS Building, 135 N.
Bellefield Avenue
Pittsburgh PA 15260
Tel: (412) 624-9454
Fax: (412) 648-7001
E-mail: aspink@sis.pitt.edu

Bernard J. Jansen
School of Info Sciences and
Technology
The Pennsylvania State
University
4P Thomas Building
University Park PA 16802
Tel: (814) 865-6459
Fax: (814) 865-6424
E-mail: jjansen@ist.psu.edu

Jan Pedersen
Chief Scientist
Overture Web Search
Division
1070 Arastradero Road
Palo Alto, CA 94304
jan.pedersen@overture.com

* To whom all correspondence should be addressed.

Abstract

A user's single session with a Web search engine may consist of seeking information on single or multiple topics. Most Web search sessions consist of two queries of two words. We present findings from a study of two-query search sessions on the Alta Vista Web search engine to examine the degree of multitasking search by typical web searchers. A sample of two-query length search sessions were filtered from Alta Vista transaction logs from 2003. Findings include: (1) 81% of two-query sessions were multitasking searches, and (2) there are a broad variety of search topics in multitasking search sessions. The implications of our findings for Web design and further research are discussed.

1. Introduction

Cognitive psychologists have studied many aspects of multitasking or task switching [1, 2]. Rubinstein, Meyer and Evans [3] found that multitasking between different types of tasks can reduce productivity. Recent information retrieval (IR) studies suggest that users' searches may have multiple goals or topics and occur within the broader context of their information-seeking behaviors [4]. Spink, Bateman and Jansen [5] and Spink, Wilson, Ford, Foster and Ellis [6] show that users often conduct related searches over time on the same or evolving single topic or *successive search*.

A user's single session with a Web search engine may consist of seeking information on single or multiple topics. Most Web search sessions consist of two queries of two words [7]. Spink, Ozmutlu and

Ozmutlu [8] show that IR searches often include multiple topics, during a single search session or *multitasking search*. Spink, Batemen and Greisdorf [9] found that eleven (3.8%) of the 287 Excite users responding to a Web-based survey reported multitasking searches. However, limited knowledge exists on the characteristics and patterns of multitasking searches. Recent studies have examined multitasking searching on the Excite and AlltheWeb.com Web search engines [10, 11].

Ozmutlu, Ozmutlu and Spink [10] provide a detailed analysis of multitasking sessions on AlltheWeb.com. Findings include: (1) Almost one third of AlltheWeb.com users perform multitasking Web searching (2) multitasking sessions often included more than three topics per session (3) multitasking sessions are longer in duration than regular searching sessions (4) most of the topic in multitasking searches were on general information, computers and entertainment.

Ozmutlu, Ozmutlu and Spink [11] compared the incidence of multitasking search by AlltheWeb.com and Excite Web search engine users were filtered from transaction logs. Findings include: (1) multitasking Web searches are a noticeable user behavior, one tenth of Excite users and one third of AlltheWeb.com users conducted multitasking searches, (2) multitasking search sessions are longer than regular search sessions in terms of queries per session and duration, (3) both Excite and AlltheWeb.com users search for about three topics per multitasking session and submit about 4-5 queries per topic, and (4) there is a broad variety of search topics in multitasking search sessions.

2. Research Objective

The objective of the study reported in this paper is to further examine the prevalence and characteristics of multitasking Web searching during two-query sessions by Alta Vista users.

The broader goal of our study is to compare our findings with previous studies in order to identify overall changes that have occurred for Web searching on AltaVista, along with presenting the current state of Web searching by Alta Vista users.

3. Research Design

3.1 Alta vista

In 2002, Alta Vista was the 9th most popular search engine, had a content collection of 550 million Web pages [12], and approximately 5.6 million unique visitors per month. The drop in Web site traffic reflects the intense competition and consolidation that has occurred in the Web search engine industry. Alta Vista supported the same query operators as in 1998 [13]. We see from this information that AltaVista offers a full range of searching options, has an extremely large content collection, and millions of unique visitors per month. After being an independent company for several years, Overture Services purchased AltaVista in 2003 [14]. At the time of this study, data from other search engines (e.g., Google, Inktomi) was not available.

3.2 Data collection

To address our research questions, we obtained, and quantitatively analyzed, actual queries submitted to Alta Vista in 2002. The queries examined for this study were submitted to Alta Vista on Sunday, 8 September 2002 and span a 24-hour period. We checked news stories from this day to see if any looked as if they may influence the analysis, namely the term analysis. There did not appear to be a major news stories occurring on this date. However, the date is near the anniversary of the 9-11 terrorist's attacks.

We recorded the queries in a transaction log and represent a portion of the searches executed on the Web search engine on this particular date. The original general transaction log contains approximately 3,000,000 records. Each record contains three fields:

Time of Day: measured in hours, minutes, and seconds from midnight of each day as recorded by the Alta Vista server;

User Identification: an anonymous user code assigned by the Alta Vista server

Query Terms: terms exactly as entered by the given user.

3.3. Data analysis

The transaction log of the 2002 data is a flat ASCII file, which we imported into a relation database in order to conduct the analysis. We generated a unique identifier for each record. Using the three fields (*Time of Day*, *User Identification*, and *Query Terms*), we located the initial query and then recreated the chronological series of actions in a session.

A *term* is any series of characters separated by white space or other separator.

A *query* is the entire string of terms submitted by a searcher in a given instance

A *session* is the entire series of queries submitted by a user during one interaction with the Web search engine.

The transaction log contained searches from both human users and agents. We were interested in only those queries submitted by humans rather than by some automated process. Given that there is no way to accurately identify human from non-human searchers, most researchers utilizing transaction logs for data collection must either ignore it or assume some temporal or interaction cut-off.

We chose the latter approach by separating all sessions with 100 or less queries into an individual transaction log. We chose this cut-off because it is almost 50 times greater than the reported mean search session [7] for human Web searchers, assuring that we were not excluding any human searches. Although this cutoff probably introduced some agent queries, we were satisfied that we had retrieved a subset of the transaction log that contained queries submitted primarily by human searchers in a non-common user terminal, yet broad enough not to introduce bias by too low of a cut-off threshold.

When a searcher submits a query, then views a document, and returns to the search engine, the Alta Vista server logs this second visit with the identical user identification and query, but with a new time (i.e., the time of the second visit). This is beneficial information in determining how many of the retrieved results pages the searcher visited from the search engine, but unfortunately it also introduces duplicate queries.

3.4. Two-query sessions

We analyzed the sample of two-query sessions to examine: (1) the degree of multitasking search sessions (the proportion of two-query sessions that include two topics), and (2) the topics in multitasking Web search sessions. The multitasking search sessions were sifted manually from the two-query dataset. We qualitatively analyzed each search session to identify different topics in the multitasking search sessions, and determine if the topic switch was fairly distinct (e.g., a topic switch from black jaguar cars to Italian operas).

4. Results

4.1. Alta vista sessions

The Alta Vista Web query data set is shown in Table 1.

	Alta Vista 2003
Total sessions	254
Multitasking sessions	206
% of multitasking sessions	81.1%
Total queries	655
Queries in multitasking sessions	532
% of queries in multitasking search sessions	81.2%
Mean queries per session	2

Table 1: Alta Vista data set.

	Alta Vista 2003
Total topic changes	206
Mean topic changes per session	1
Topics per session	2
Mean queries per topic	1

Table 2: Topic changes in multitasking sessions.

For the Alta Vista dataset, there were 206 topic changes in 206 multi-topic sessions, yielding a mean of 1 topic change per session. Typical Alta Vista users engaged in multiple topic sessions and investigated more than one topic.

Ozmutlu, Ozmutlu and Spink [10, 11] found that most of the Excite and AlltheWeb.com users make 1-3 topic changes (cover 2-4 topics) per session. The mean queries per topic was 4.5 queries for the Alta Vista dataset, meaning that on average users making multitasking searches changed the search topic every 4.5 queries. The relevant figure for the Excite dataset

For the Alta Vista dataset, of 254 user sessions of two queries, 206 (81.1%) were multitasking Web search sessions. The 206 multitasking Alta Vista sessions included 532 queries.

Alta Vista users seem to be conduct multitasking searches, since they submitted more multitasking Web search sessions and queries than non-multitasking sessions.

Previous research by Ozmutlu, Ozmutlu and Spink [10, 11] examined multitasking during Excite and alltheWeb.com searching, not just two-query sessions. They showed that the mean queries per multitasking search session were 14.9 for Excite and 14.3 for AlltheWeb.com users. The mean queries per session for the entire Excite sample was 10, making Excite multitasking sessions about 50 percent longer than regular search sessions. The same statistics for the AlltheWeb.com dataset shows that the mean queries were 10.3 for the entire sample and 14.3 for multitasking sessions.

3.2. Topic changes

Table 2 shows the data relating to topic changes during the Alta Vista two-query data set.

For the Excite users, there were 246 topic changes in the 114 user query sessions, with a mean of 2.2 topic changes per session that can also be interpreted as a mean of 3.2 topics per session.

Both Excite and AlltheWeb.com users search more three different topics per search session. is 4.8 queries per topic. Excite and AlltheWeb.com users submit almost the same number of queries per topic.

3.3. User search topics

The queries in multitasking sessions were categorized with respect to the topics provided in Spink, Ozmutlu and Ozmutlu [8]. The categories

used in the study and the number of queries falling into each category for the Alta Vista dataset is provided in Table 3.

Topic Category	Number of Queries	%
General Information	100	19%
Arts/Humanities	11	2%
Business/Economy	27	5%
Computer/Internet	64	12%
Education	33	6%
Employment/Job	14	3%
Entertainment	37	7%
Government/Politics	9	2%
Hobbies	13	2%
Individual/Family	24	5%
Medical/Health	28	5%
News	11	2%
Science	13	2%
Sex	32	6%
Shopping	58	11%
Travel	12	2%
Inexplicit	46	9%
Total	532	100%

Table 3: Number and percentage of queries in each topic category.

The most preferred categories for the sample of two-query Alta Vista sessions we examined were general information, computers and shopping. These categories form about 35% of the queries in multitasking sessions. Ozmutlu, Ozmutlu and Spink [10, 11] show that Excite users preferred the categories of hobbies, shopping and business that form about 47% of all queries in multitasking sessions. The subject categories might reflect topic at the same time. They found out that multitasking search sessions included more than three topics per search session.

5. Discussion

We identified some characteristics of two-query Alta Vista search sessions. Many two-query sessions were multitasking sessions where users switched from a single query on one topic to a single query on another topic. Ozmutlu, Ozmutlu and Spink [10, 11] found that most multitasking search sessions were twice longer than regular search sessions in terms of duration and the mean queries per Web multitasking search session being 50% higher than that of regular search sessions. Although the reasons for multitasking behaviors were not directly tested in the studies discussed above, one can draw some insights from the data analysis.

First, it appears that users' Web searching behaviors, including typical Web search sessions of two queries, often includes a need to seek and search on more than one topic concurrently due to the complex nature of work or living tasks. Rubinstein, Meyer and Evans [4] highlight the need for people to multitask in work environments as they use the microprocessor at the same time they talk on the telephone. People have many tasks at hand at the same time, including information seeking tasks. In these cases, a person may pool their topics together and interact with the Web on more than one related or unrelated topics.

The findings from this study have implications for the design of Web systems, the development of information seeking and interactive IR models, and the training and searching practice of end-users.

On the theoretical side, current information seeking models and interactive IR models, and studies of IR system use (that focus on the single search assumption), do not address multitasking search. In practice, users should be trained to understand how to work concurrently and successively on multiple topics that are not resolved with one IR system search. Spink, Ozmutlu and Ozmutlu [8] suggest that on average users may need to conduct possibly 2-3 searches to clarify to resolve an information problem.

Some commercial IR systems, have a save search feature based on the assumption that many users come back to the IR system for more than one search on their topic over time. How might multitasking search sessions be supported by Web systems and interfaces? There exist many Web systems that support searching on one information task in a single database or support browsing within a single database or many Web systems, e.g., meta-search tools.

Largely, current search systems are based on the assumption that users will (or need to) engage in only one search task (topic) at a time during one search. However, in practice, many users now routinely engage in multitasking searches in the course of a single or multiple search episodes. Users at different information seeking and problem solving stages may conduct different search behaviors. Multitasking is a major systems design research topic. Windowing features provide support for interactive multitasking behaviors within operating systems. However, few interactive systems provide effective support for managing multitasking behaviors.

6. Conclusion

Currently, Web systems and interfaces provide limited support for multitasking search. As the complexity of information structures and problems increases, more complex human information processes and more effective Web technologies are required to sustain effective human information behaviors. We are currently continuing to examine: (1) how multitasking searches differ from non-multitasking sessions (2) if a user's information-seeking stage affects the number and performance of multitasking searches and (3) the relationship between the nature of the user's information problems and how they are coordinated. Further research is also required to compare and characterize information problems that lead to multitasking.

7. References

- [1] R.A. Carlson and M-Y, Sohn. "Cognitive Control of Multistep Routines: Information Processing and Conscious Intentions". In: *Stephen Monsell and Jon Driver (Eds.), Control of Cognitive Processes: Attention and Performance XVIII*. Cambridge, MA: The MIT Press, 2000 (pp. 443-464).
- [2] Y. Miyata and D.A. Norman. "Psychological Issues in Support of Multiple Activities". In *User Centered Design, D. A Norman and S. W. Draper, eds., Lawrence Erlbaum, NJ, 1986* (pp. 265-284).
- [3] J. Rubinstein, D. Meyer and J. Evans. "Executive Control of Cognitive Processes in Task Switching". *Journal of Experimental Psychology*, 2001y,
- [4] M. Miwa, "User Situations and Multiple Levels of Users Goals in Information Problem Solving Processes of AskERIC Users", *Proceedings of the 2001 Annual Meeting of the American Society for Information Sciences and Technology*, 2001, 38, 355-371.
- [5] A. Spink, J. Bateman and B.J. Jansen. "Searching Heterogeneous Collections on the Web: A Survey of Excite Users". *Internet Research: Electronic Networking Applications and Policy*, 1999, 9(2), 117-128.
- [6] A. Spink, T.D. Wilson, N. Ford, A. Foster and D. Ellis. "Information Seeking and Mediated Searching. Part 3: Successive Searching". *Journal of the American Society for Information Science and Technology*, 2002, 53(9), 716-727.
- [7] 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*, 2000b, 36(2), 207-227.
- [8] A. Spink, H.C. Ozmutlu and S. Ozmutlu. "Multitasking Information Seeking and Searching Processes". *Journal of the American Society for Information Science and Technology*, 2002, 53(8), 639-652
- [9] A. Spink, J. Bateman and H. Greisdorf. "Successive Searching Behavior During Information Seeking: An Exploratory Study". *Journal of Information Science*, 1999, 25(6), 439-449.
- [10] S. Ozmutlu, H.C. Ozmutlu and A. Spink. "Multitasking Web Searching: Implications for Design." *ASIST'03: Annual Meeting of the American Society for Information Science and Technology, Oct 18-22, 2003. Long Beach, CA*.
- [11] S. Ozmutlu, H.C. Ozmutlu and A. Spink. "A Study of Multitasking Web Searching". *IEEE ITCC'03: International Conference on Information Technology: Coding and Computing, April 2003, Las Vegas, NV*.
- [12] D. Sullivan, Nielsen /Netratings Search Engine Ratings [website]. SearchEngineWatch.com. Retrieved 6 January, 2002, from the World Wide Web: <http://www.searchenginewatch.com/reports/netratings.html>
- [13] AltaVista. *Special Search Terms* [Web site]. AltaVista. Retrieved 16 May, 2003, from the World Wide Web: http://www.altavista.com/help/adv_search/syntax.
- [14] B. Morrissey, *Overture to buy AltaVista* [Web site]. Internet Advertising Report. Retrieved 16 May, 2003, from the World Wide Web: <http://www.internetnews.com/IAR/article.php/1587171>