Parallel Browsing Behavior on the Web

Jeff Huang
University of Washington
Information School
ht@jeffhuang.com

Ryen W. White
Microsoft Research
Redmond, WA 98052
ryenw@microsoft.com

ABSTRACT

Parallel browsing describes a behavior where users visit Web pages in multiple concurrent threads. Web browsers explicitly support this by providing tabs. Although parallel browsing is more prevalent than linear browsing online, little is known about how users perform this activity. We study the use of parallel browsing through a log-based study of millions of Web users and present findings on their behavior. We identify a power law distribution in browser metrics comprising "outclicks" and tab switches, which signify the degree of parallel browsing. We find that users switch tabs at least 57.4% of the time, but user activity, measured in pageviews, is split among tabs rather than increasing overall activity. Finally, analysis of a subset of the logs focused on Web search shows that while the majority of users do not branch from search engine result pages, the degree of branching is higher for non-navigational queries. Our findings have design implications for Web sites and browsers, search interfaces, and log analysis.

Categories and Subject Descriptors

H.5.4 Information interfaces and presentation (e.g., HCI): Hypertext/Hypermedia—Navigation.

General Terms

Human Factors, Design, Measurement

Keywords

Parallel browsing, tabs, log mining.

1. INTRODUCTION

Browsers such as Internet Explorer and Firefox are the primary means by which people access Web content. The growth in information and services available online has led to more time spent using Web browsers. Recent estimates suggest that the average U.S. individual spends 12 hours per week on the Web, almost as much time as they spend watching television [1]. In activities engaged online, Web search is a close second in popularity behind email. It is therefore critically important to understand how people navigate and search the Web and how their browsing strategies can be enhanced.

Until recently, the prevalent browsing paradigm for the Web involved visiting a sequence of pages in the same browser window. However, through the provision of browser tabs, Web

browsers have started supporting *parallel browsing*, allowing users to engage multiple concurrent tabs simultaneously. These are fundamental behaviors for interacting with hyperlinks in a hypertext system, which is the underlying structure of the Web. Figure 1 illustrates a parallel browsing session within a browser window (Window 1) starting with a single page (page 1) and involving four browser tabs. In the figure, time flows from left to right and the numbers assigned to pages depict the order in which they are visited within the session. One or more pages stemming from a single page (e.g., visits from page 2 to both page 3 and page 5) are called outclicks, from which branching occurs when new tabs are opened. Browsing flow within a single tab may then be interrupted by tab switches to visit pages in other tabs.
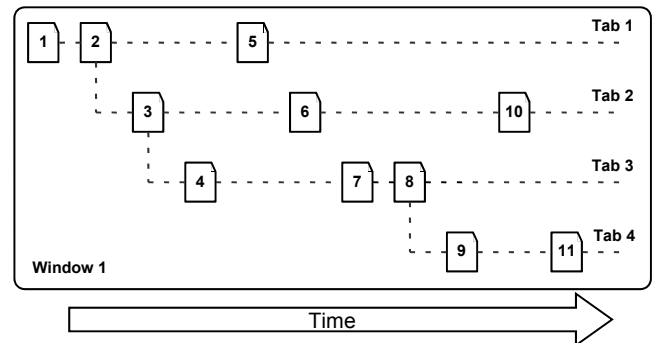

Figure 1: An example parallel browsing session.

While parallel browsing was previously possible by opening multiple browser windows, browser tabs are a lightweight interface mechanism specifically designed for this purpose. Qualitative studies have observed users using multiple browser windows and tabs to browse in parallel. One survey of experienced Web users (reported in [2]) noted that using multiple windows and tabs was a key strategy in supporting backtracking or engaging elsewhere while one page was loading. Clearly, tabs alter browsing behaviors, but a large-scale study of these behaviors has been lacking due to the difficulty of obtaining usage data at sufficient volume. Fortunately, we were able to obtain log data containing Web usage information (including tabbed browsing behavior) from millions of consenting users of a popular Web browser. In this paper we offer findings and implications from our data analysis.

Our study addresses two research questions: (i) when and to what extent are users parallel browsing on the Web? and (ii) what affects parallel browsing behavior during interaction with Web search results? To answer these questions, we analyzed browser logs, which describe natural user behaviors at scale. We collected these logs from a popular Web browser plug-in and used the data to analyze parallel browsing behavior through metrics such as pageviews, outclicks, and tab switches. We also study parallel browsing in search results to analyze user branching behavior. We conclude by discussing the implications of our findings for Web sites and browsers, search interfaces, and log analysis.

1.1 Parallel Browsing Terminology

To help the reader understand the terminology used in the paper, we provide definitions and relate them to Figure 1 as appropriate.

The latest browser interfaces blur the line between window and tab. In our paper, we refer to *tabs* as the components which display Web pages, and *windows* as the containers for tabs. Figure 1 illustrates four tabs within one window.

When a user visits a page in the browser, our plug-in records information about the visit in a central Web server's logs. Each visit is counted as a *pageview*. Each numbered page in Figure 1 is an example of a pageview.

Branching occurs when a user opens a hyperlink in a new tab. This is primarily achieved through the right-click menu, middle-mouse click, or by holding the CTRL key before clicking a link. In Figure 1, the transition between page 2 and page 3 is an example of branching behavior.

Outclicks can be either direct clicking or branching from a hyperlink. In other words, navigating to a link in the same tab or opening that link in a new tab. All transitions between pages in Figure 1 are examples of outclicks.

We refer to *tab sessions* as browser interactions occurring entirely within one tab, comprising one or more pageviews. There are four tab sessions in Figure 1, each comprising multiple pages (e.g., the tab session for Tab 1 contains pages 1, 2, and 5).

Tab switches involve visits to other tabs between pageviews in a tab. While the user is active in one tab, they may open or switch to another tab to view pages, and then switch back to the original tab. This is common during simultaneous use of multiple tabs. In Figure 1, the visit to page 3 (on Tab 2) between page 2 and page 5 (both on Tab 1) is an example of a tab switch.

2. RELATED WORK

Several qualitative studies have reported on the use of multiple windows in Web search. Thatcher et al. observed users opening multiple browser windows to simultaneously conduct multiple searches [15]. A survey of experienced Web users by Aula et al. [2] found that multiple windows and tabs were used to support backtracking or multitasking while one page was loading. For backtracking, they observed users branching and eventually switching back to the originating tab rather than using the Back button in the browser. For multitasking, they observed users interacting with another tab while waiting for the original one to load, i.e., tab switching. More recently, a Web navigation study by Weinreich et al. [18] found their participants using multiple windows frequently, enabling them to compare search results side-by-side, load pages in the background while they continued browsing, and retain important pages in their sessions for back-tracking. One participant in that study explained that they used "new tabs for closely related tasks and new windows for parallel tasks." Weinreich et al. also found that the participants who used tabs backtracked less than those who did not. Similarly, those who created more windows tended to backtrack less. These studies describe instances of parallel browsing. We extend this work and study *when* and *what* are the extents of this behavior on the Web.

Viermetz et al. [17] showed how tabbed browsing affects the validity of search log studies. However, they analyze server side logs which do not directly contain information about tabs or windows since they only record incoming network traffic. Using those data, they construct a "clicktree" of possible paths from the pageviews in the logs. They find that parallel browsing occurs 4-85% of the time, a fairly wide range. In contrast, our browser logs contain information about tabs, allowing us to study parallel browsing behaviors directly.

Spink et al. [14] discuss multitasking in Web search, which they define as searching for multiple topics in a single session. They conclude that users frequently multitask searches according to this definition. However, parallel browsing, when a user has simultaneous ongoing tab sessions, seems closer to the classical definition of multitasking. From a cognitive psychology perspective, Miyata and Norman [11] explain that humans have foreground and background tasks, where foreground tasks have the current focus and background tasks do not require conscious supervision. Parallel browsing follows this model because the user is not interacting with multiple tabs at the same time. Instead, the current active tab is the foreground task and has the user's attention, while other tabs may be loading in the background or contain information that is not yet needed.

3. METHOD

To study parallel browsing behavior, we gathered logs from a widely-distributed browser plug-in. This plug-in is installed alongside a popular Web browser and data was collected from users who consented to send information about their browser interactions. No geographic filtering of the logs were performed prior to this analysis; this means that users were geographically diverse and representative of browser usage in many countries throughout the world. Each pageview by a user triggered the recording of the following fields in logs on our servers,

{UserId, Timestamp, TabId, ReferrerUrl, Url, IsQuery}

UserId uniquely identifies each user while the TabId field uniquely identifies tab sessions. Timestamp is the time when the pageview occurred. When a user clicks on a hyperlink, the URL of the page containing the link is the ReferrerUrl field and the destination page is the Url field. Finally, IsQuery flags whether the URL is a Web search engine results page.

Outclicks are quantified as the number of times the Url of a pageview is the same as the ReferrerUrl for later pageviews (determined by Timestamp). Branching is detected as outclicks that lead to a new TabId. Finally, tab switches can be counted as the number of unique TabIds that occur between pairs of pageviews from a specific TabId.

We used the entire set of logs from June 2009 containing approximately 60 billion pageviews from over 50 million users, giving us high coverage of Web browsing behavior. Although we focus on tab usage, parallel browsing can also be studied via analyzing the creation of new browser windows. However, our data did not contain reliable indicators of window creation events, precluding that analysis. Findings related to tabs have more direct implications for browser design; tabs are explicitly included in browsers to support parallel browsing and we study their use for that task.

Parallel browsing sessions can be spawned by two types of behavior: creating a new blank tab by clicking on a button in the browser interface, or opening a link in a new tab[1] (i.e., branching). We studied three metrics in instances of parallel browsing: (i) *number of pageviews per tab*; (ii) *number of tab switches*; and (iii)

---

[1] This includes opening links in new windows, since a new window will contain a new tab according to our definitions.

number of outclicks from a pageview, effectively the total number of tab sessions leading from that visit. These three metrics help us understand parallel browsing behavior on the Web by quantifying navigation and tab interactions. Next, we investigated branching behavior on a subset of the data. Since branching may originate in situations beyond user control (e.g., from page popups) we elect to study branching behavior within the controlled setting of Web search results. This limits the generalizability of our findings to search result pages, but make the findings more meaningful.

4. RESULTS AND DISCUSSION

We first report results from general Web browsing data (Section 4.1), then focus on results specific to branching on the search result hyperlinks (Section 4.2).

4.1 Browsing Behavior

Figure 2 shows three plots on a log-log scale: (1) outclicks and (2) tab switches for each pageview, and (3) the number of pageviews in each tab. The x-axis represents the value of the metric, and the y-axis represents the frequency of occurrence in the logs.
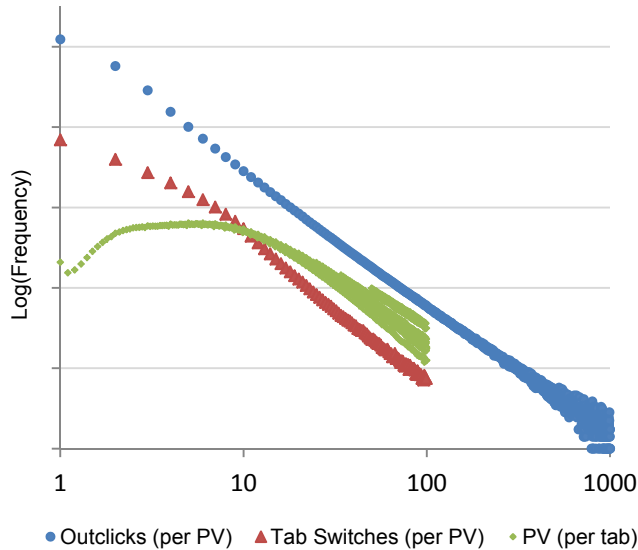

Figure 2: Number of outclicks per pageview (PV), tab switches per pageview, and pageviews per tab plotted on a log-log scale.

The figure shows that 5-10 pageviews per tab is common, meaning users often visit a handful of pages in each tab. The plot for pageviews per tab resembles a log-normal distribution with parameters $\sigma = 0.76$, $\mu = 2.16$ found using maximum likelihood estimation. In other words, if we were to take the log of the number of pageviews, then we would find a close fit to a normal distribution with the above parameters, i.e., $\ln P \sim N(\mu, \sigma^2)$ where P is the sampled pageviews. The log-normal model of pageviews passes a Kolmogorov-Smirnov goodness-of-fit test at a significance level of $\alpha = 0.01$. Log-normal distributions occur when independent factors are innately multiplied; here, the number of pageviews occurring within one tab may be a product of user habits, information need, or viewed content. These distributions on the Web are uncommon but have been observed before [4].

Outclicks and tab switches follow a different model. Outclicks decrease sharply, suggesting that most page visits do not result in branching. We observe a similar distribution for tab switches. Both outclicks and tab switches have good linear fits ($R^2 > 0.99$) in the log-log plot; they follow a power law distribution conform-

ing to the general equation $y = ax^{-k}$, with exponents $k = 3.2$ and $k = 3.5$ respectively. A small k value and R^2 value close to 1 supports the premise that outclicks and tab switches conform to a power distribution rather than an exponential distribution. This is not surprising, considering the power law is present in many properties on the Web, including session length [10], PageRank of homepages [4], and hyperlink in-degree and out-degree [3][6]. Together, these distributions show regularities in Web characteristics, allowing us to concisely describe user behavior.

Our logs showed that 88.7% of pageviews and 42.6% of tab sessions did not involve tab switches. That is, users would remain on the same tab for the next pageview in 88.7% of cases, and continuous pageviews comprised entire tab sessions in 42.6% of cases. Conversely, 57.4% of tab sessions had a least one tab switch. This was surprising because this indicates that parallel browsing is even more common than linear browsing.

Since parallel browsing is prevalent, it is important to understand whether it is benefiting users in terms of their activity levels. Are parallel browsing users working on multiple tasks at the same time, interacting with one tab while other tabs load? Or are they splitting one task into multiple tabs? Or is the answer a combination of the two? These are complex questions requiring long and careful observations of large user groups to answer. Indeed, studies of multitasking e.g., [11] have investigated this topic from a cognitive psychology perspective. However, using browser logs and simple metrics, we can compute the general relationship between a user's average number of pageviews per tab and the average number of tab switches per pageview in their browsing history. This lets us estimate the effect of parallel browsing on the amount of activity within a tab. Figure 3 is a scatterplot of two parallel browsing metrics for users: tab switches and pageviews.
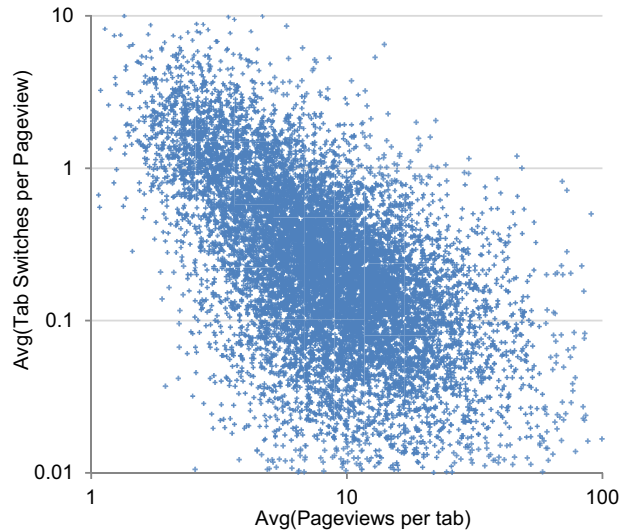

Figure 3: Average pageviews vs. average tab switches for 10,000 randomly sampled users on a log-log scale. The best fit regression is an inverse relation $y \approx 2/x$, $R^2 = 0.34$.

Users who view more pages in each tab tend to be less likely to switch tabs during a tab session. Figure 3 shows a linear inverse correlation between the tab switches and pageviews per tab, i.e.,

$$\text{Avg}(TabSwitchesPerPageview) \approx \frac{2}{\text{Avg}(PageviewsPerTab)}$$

This implies that on average, users view the same number of pages overall, but divide their browsing across multiple tabs. This preliminary finding shows that in the aggregate, tabs do not result in users viewing more pages; it simply leads to multitasking. In addition to analyzing user activity, more analysis is required of the quality and quantity of outcomes achieved when tabs are used.

4.2 Branching in Search Results

Query and user actions solely on search engine results pages are a subset of our logs, totaling around 10% of the pageviews analyzed in the previous subsection. These pages are an ideal setting for studying branching because they offer users a candidate list of navigation choices, nearly equally accessible. They avoid confounds due to the implementation of the Web page: some pages automatically open new pages (i.e., popups) in windows or tabs when loaded[2]; others force branching by having links that always open in new windows or tabs. Furthermore, the type of Web page, such as it being a portal or link hub, can encourage different browsing behaviors. Narrowing our analysis of branching to search result pages helps control these variables. Not only do we know the actual query and that the goal is to satisfy an information need, but we can be certain that users are in control of their browsing behavior, making the results more meaningful.

Table 1 shows the relative number of users who branch from the search results page compared to those who simply click the result hyperlink. For these data, we only considered users who had initiated at least 100 queries in the 30-day duration of our study since we needed enough interactions per user to determine their branching habits. The data shows that while a small percentage of users branch frequently, the majority always click results directly.

Table 1: The percentage of result clicks that are branches.

Branching Rate	% Users
Usually (76-100%)	5.50
Often (51-75%)	2.78
Sometimes (26-50%)	2.80
Occasionally (1-25%)	24.95
Never (0%)	63.98

In addition to analyzing branching behavior at the user level, we also study branching at the query level. We assembled a list of queries by combining the top 100 queries from both Google Zeitgeist [7] and Search Engine Guide [19], which are lists tracking the most popular search queries. From this list, we removed queries issued by less than 1,000 unique users in order to obtain significant results for each query. From analyzing the branching percentage for the resultant 97 queries (Table 2), we found that the query type has an effect on parallel browsing use. Specifically, users are unlikely to branch for navigational and transactional queries (e.g., [mapquest], [home depot]), but are more likely to branch for ambiguous and informational queries (e.g., [obama], [cancer]). Still, users only branch 6-15% of the time for queries that are the most likely candidates for branching. So while non-navigational queries generally have higher branching rates, users who branch often are a minority.

---

[2] The default popup-blocker in the browser we study limits popups' effect on counts of tab switches.

Table 2: Popular queries ranked in descending order of percent queries branched. Only queries with the highest and lowest branch % are listed, due to space constraints.

Query	Query Type	% Branched
bikini	erotic	15.28
nudist	erotic	12.84
megan fox	celebrity	10.99
kim kardashian	celebrity	9.04
obama	celebrity	7.96
search engines	search	7.67
girls	erotic	7.27
how to kiss	informational	7.11
cancer	informational	6.45
hot	erotic	6.42
Skip 77 queries…		
fox news	navigational	0.96
sears	navigational, store	0.94
white pages	navigational	0.93
home depot	navigational, store	0.88
dogpile	navigational, search	0.80
lowes	navigational, store	0.79
weather.com	navigational	0.78
mapquest	navigational	0.76
ask.com	navigational, search	0.69
map quest	navigational	0.65

4.3 Limitations

Our work has limitations typical of search log analysis studies. One is the inherent bias from the collection of the logs. The findings are derived only from data from consenting users of a popular browser. While the sample size is large, the browser's interface features and characteristics of its user base may affect some behaviors. The logs also cannot record browser events that do not result in a pageview, such as opening a new blank tab or client-side interactions with Websites using JavaScript or Ajax. Missing pageviews resulting from client-side interactions may cause us to overestimate the prevalence of parallel browsing.

5. IMPLICATIONS

Web site and browser design: A number of Websites force the decision of branching on users. Our findings suggest that this is not prudent because users have different branching preferences. If a Website designer wants to force branching, they should do so in an informational context (sites providing different information in links) rather than a navigational context (sites which list many links). However, most users do not branch even in a close-to-ideal setting such as on the search engine result page of an informational query. While a Web designer may believe opening multiple windows from a page containing multiple informational links is helpful, it is unlikely to be a user's preference.

Web browsers can return control of branching to users by disallowing Web pages to force branching. Browsers have already taken the first step with popup blockers, and while pages that create a new tab when clicked are not popups, users may be unprepared for the shift from linear browsing to parallel browsing.

Understanding the Web: We found that the frequency of out-clicks and tab switches follow the power law, while the number of pageviews per tab follow a log-normal distribution. These regularities can be used to predict and simulate Web interaction behavior. Parallel browsing behavior can be expected to be exponentially proportional to the pages users visit. This can guide design decisions such as the number of hyperlinks to place on a page. Power laws are well-studied distributions, from which summary statistics such as expected value and skewness can also help us find average case behavior, outliers, spread, and shape. Websites that observe deviations from these distributions may want to investigate whether this is caused by spammers or bots, as shown in [4]; currently, it is unlikely that these unwanted visitors would model their clicks to follow the power laws.

Log analysis: In search log analysis, where click data helps rank search results, frequent branching may indicate that there are multiple enticing search results on that page. Incorporating the expectation of more evenly distributed clicks across results may help search relevance by adjusting click distribution models.

Studies that consider sessions in search logs as sequential pageviews may be limited because of the prevalence of parallel browsing behavior. Many prior studies in search log mining use sequential timestamp to identify sessions [8][9][12]. This ignores the possibility that users may have branched their session, and now have multiple browser tabs which interleaves the click data in server-side logs. Server-side logs used in studies such as [9][17], which only capture events visible to the Web server, do not capture the parallel sessions since they do not contain tab data.

Search interfaces: A user's search habits and the type of information need significantly affects branching behavior in search results. Our finding that informational queries lead to branching more often than navigational queries suggests that a query's branching percentage may even predict query type. Search engines can apply this by providing different interfaces depending on the user and query. For example, if a query is a good branching candidate, a search results page with an embedded tabbed panel may allow the user to tab through the results without leaving the page. Showing tours and trails (e.g., [5][13][16]) on the results page may help users decide which links to visit via example paths.

6. CONCLUSION

Our study produces several significant findings. We discover a power law distribution in browser pageviews for outclicks and tab switches, which signify the degree of parallel browsing. Similarly, we find a log-normal distribution in pageviews per tab. We find that 57.4% of tab sessions involve parallel browsing, but we observe that users are multitasking by splitting their browsing activity into different tabs rather than viewing more pages overall. Finally, we focus on search result pages and note that the majority of users do not branch on that type of page, but the degree of branching increases for non-navigational queries. While each pageview has only a small likelihood of being branched, the tab session as a whole frequently contains parallel browsing.

Our overall contribution is an analysis of metrics relating to parallel browsing on the Web. This type of behavior has been growing recently and we foresee this continuing to gain popularity. Studies of this type of navigation behavior have been lacking in Web and hypertext communities, despite parallel browsing capturing fundamental interactions with hyperlinks. Our study is an initial exploration of this new and important user behavior.

7. REFERENCES

[1] Anderson, J. 2009. *Consumer Behavior Online: A 2009 Deep Dive.* Forrester Market Report. Retrieved from www.forrester.com/go?docid=54327, September 2009.

[2] Aula, A., Jhaveri, N., and Käki, M. 2005. Information search and re-access strategies of experienced web users. In *Proceedings of WWW '05*, 583-592.

[3] Barabási, A.-L., Albert, R., and Jeong, H. 2000. Scale-free characteristics of random networks: the topology of the world-wide web. *Physica A*, 272, 69-77.

[4] Becchetti, L., Castillo, C., Donato, D., Leonardi, S., and Baeza-Yates, R. 2006. Link-Based Characterization and Detection of Web Spam. In *Proceedings of AIRWeb '06*, 1-8.

[5] Bilenko, M. and White, R.W. 2008. Mining the search trails of surfing crowds: identifying relevant websites from user activity. In *Proceedings of WWW '08*, 51-60.

[6] Faloutsos, M., Faloutsos, P., and Faloutsos, C. 1999. On power-law relationships of the internet topology. In *Proceedings of SIGCOMM '99*, 251-262.

[7] Google Zeigeist 2008. Top of Mind. Retrieved from http://www.google.com/intl/en/press/zeitgeist2008/mind.html on July 8, 2009.

[8] He, D., Göker, A., and Harper, D. J. 2002. Combining evidence for automatic web session identification. *Information Processing & Management*, 38, 5, 727-742.

[9] Huang, J. and Efthimiadis, E. 2009. Analyzing and evaluating query reformulation strategies in web search logs. In *Proceedings of CIKM '09*, 77-86.

[10] Huberman, B.A., Pirolli, P., Pitkow, J.E., and Lukose, R.M. 2008. Strong regularities in world wide web surfing. *Science*, 280, 5360, 95-97.

[11] Miyata, Y. and Norman, D. 1986. Psychological issues in support of multiple activities. In: D.A. Norman and S.W. Draper, Editors, *User Centered Design*, 265-284.

[12] Ozmutlu, H. C. and Çavdur, F. 2005. Application of automatic topic identification on excite web search engine data logs. *Information Processing & Management*, 41, 5, 1243-1262.

[13] Reich, S.F., Carr, L., De Roure, D. and Hall, W. 1999. Where have you been from here? Trails in hypertext systems. *ACM Computing Surveys*, 31(4).

[14] Spink, A., Park, M., Jansen, B.J., and Pedersen, J. 2006. Multitasking during web search sessions. *Information Processing & Management*, 42, 1, 264-275.

[15] Thatcher, A. 2008. Web search strategies: The influence of Web experience and task type. *Information Processing & Management*, 44, 3, 1308-1329.

[16] Trigg, R.H. 1988. Guided tours and tabletops: tools for communicating in a hypertext environment. *ACM TOIS*, 6(4).

[17] Viermetz, M., Stolz, C., Gedov, V., and Skubacz, M. 2006. Relevance and impact of tabbed browsing behavior on web usage mining. In *Proceedings of Web Intelligence*, 262-269.

[18] Weinreich, H., Obendorf, H., Herder, E., and Mayer, M. 2006. Off the beaten tracks: exploring three aspects of web navigation. In *Proceedings of WWW '06*, 133-142.

[19] Wordtracker. 2009. Top 500 Search Engine Keywords Of The Week - June 10, 2009. Retrieved from http:// searchengineguide.com/wordtracker/top-500-search-engine-keywords-of-the-we-7.php on July 8, 2009.