

Web Browsing Today: The Impact of Changing Contexts on User Activity

Kirstie Hawkey and Kori Inkpen

Faculty of Computer Science, Dalhousie University

Halifax, NS B3H 1W5

{hawkey, inkpen}@cs.dal.ca

ABSTRACT

Although web browsing behaviour was studied in detail in the mid-to-late 1990s, few recent results have been reported. The nature of web browsing has changed significantly since these early studies, both in the profile of the typical web user and in the context of their browsing (e.g. location, connection speed, web browser features). This paper reports on per-session and per-browser window usage, such as the number of pages visited and the speed of browsing. Some of our findings differ from previously published results that continue to motivate research in this area. Our research indicates that changes in user behaviour, such as the magnitude of web browsing activity, may place restrictions on web-browser related applications.

Author Keywords

Web browsing behaviour; field study; client-side logging

ACM Classification Keywords

H.5.4 [Information Interfaces and Presentation] Hypertext/Hypermedia: Navigation; User Issues

INTRODUCTION

The web browsing environment is continually evolving. In 1994, the typical web user was a young, technical male, using a browser with limited features over a slow connection. Today, users come from all segments of the population and are often using browsers equipped with advanced navigation and search features over a high-speed connection. A detailed discussion of contextual changes can be found in [5]. Web browsing increasingly occurs in mobile contexts as laptop computers and other devices accompany individuals to the workplace, home, and school. Despite these contextual changes, seminal works, such as Catledge and Pitkow's [1] 1994 study characterizing web browsing strategies, are still used as a motivation for innovative web navigation tools and techniques. This early research needs reevaluation against current contexts of use to see if the results still stand. We must continue to update our understanding of web browsing behaviour as the environment evolves so we can appropriately ground research of new tools and techniques.

Copyright is held by the author/owner(s).
CHI 2005, April 2-7, 2005, Portland, Oregon, USA.
ACM 1-59593-002-7/05/0004.

In a related project [4], we gathered data on how users perceive privacy needs related to incidental viewing of traces of their browsing activity. Beyond the privacy information gathered in this study, a great deal of data related to general browsing behaviours was collected. The goal of this paper is to disseminate these browsing statistics to update and augment earlier research.

We first describe the methodology used to gather our data. We then present descriptive statistics of web browsing behaviours observed, and contrast these findings with previous results in the literature. We then discuss challenges inherent in client-side logging of browsing behaviour, the difficulties of building upon research in this area, and the impact of web browsing behaviour on the development of web browsing solutions. We conclude with recommendations for further research.

METHOD

A week-long field study was conducted during August 2004 to elicit, normal web browsing behaviors of participants. Data were gathered on the number of web pages visited, the number of browser windows utilized, and temporal browsing patterns indicating bursts of activities, sessions, and transitions between browser windows. In order to capture the full picture of personal and work/school related browsing, we recruited participants who used laptops for the majority of their browsing.

Participants

A convenience sample was selected with twenty participants (16 males, 4 females), ranging in age from 19 to 47 years old. The minimum education level was some university, with 90% having completed at least an undergraduate degree, in primarily technical fields (14 computer science, 4 science). Eighteen of the participants were students. Participants were experienced computer users (10 years' use) and spent a considerable amount of time each week using computers (29-35 hrs./week) and web browsers (22-28 hrs./week).

Participants in our study represented a fairly homogenous group: highly educated, predominantly male laptop users. Despite the limited generalizability of this sample, it is similar in construct to that used in earlier related research, so comparisons with previous results may be valid. However, this group is not representative of the overall web

browsing population; and, therefore, the external validity of these results are limited.

Study Instruments

To record the browsing activity of participants, we built a browser helper object (BHO) to work with Microsoft’s Internet Explorer (IE). As each IE window opens, the BHO loads and logs all web sites visited until the window closes. For this study, we recorded the visited web page (URL and page title), a time stamp, and the ID number of the browser window. All pages viewed in the browsing process were logged, even if navigation continued before the document fully loaded. We did not record individual frames or images loaded within a web document, just the complete document.

We also developed an electronic diary to allow participants to assign a privacy level associated with each of their visited pages. Researchers only received a sanitized report (without the URL and page title). While we hoped this approach to maintaining privacy would contribute to participants’ willingness to engage in their usual browsing activities, it means we are unable to report on the number of unique web sites visited or the extent of site re-visitation.

Procedure

Participants completed a background questionnaire while the logging BHO and electronic diary were installed on their laptops. A few web pages were loaded to verify the logging software was working and the logged data was used to demonstrate the functionality of the electronic diary. Participants then went about their normal browsing activities for the week, completing the diary daily when possible. After one week, participants completed post session questionnaires and had the software uninstalled.

RESULTS

The browsing behaviours that our participants exhibited were highly variable, both within each person’s browsing and across the group. As such, we report descriptive statistics, including the range of behaviours observed.

Number of Pages Visited

On average, the participants in our study visited 1808 pages during the seven days (~258/day) (Table 1). However, the volume of page visits was highly variable. The number of page visits ranged from 422 (~60/day) to 5127 pages (~732/day). This is a dramatic increase from earlier reports: 42 page visits per day (1999/2000) [2], 21 visits (1995) [8], and 14 visits (1994) [1].

Earlier research only captured browsing from a single location, not accounting for all browsing that users may have done during the day. Cockburn and Mackenzie [2] collected data via history files captured on university backup files. The data collection period included the holidays, which may account for the lower traffic levels. In addition, high-speed internet access is far more prevalent than during earlier studies. Previous statistics have shown that users with high-speed internet view more web pages

Quartile	Pages visited	Browser windows	Pages per window			Window Re-visits	
			Mean	Mode	Max	Total	Per window
0%	422	47	3	1	27	22	.27
25%	1064	134	5	2	55	90	.40
50%	1508	246	7	2	92	125	.53
75%	2133	441	9	2	170	216	.68
100%	5127	799	20	2	255	430	3.72
Mean	1808	289	8	1.85	108	167	.68

Table 1. Quartile and mean values for number of pages visited and browser window usage over the week

and surf the web more often than those with dial-up connections [7]. The popularity of web-mail, news sites, and the prevalence of pop-up windows may also account for increases.

Browser Window Usage

Overall, the participants in our study opened an average of 289 different browser windows during the seven days (Table 1). Again, this result was highly variable. Across participants, the number of different browser windows opened ranged from 47 to 799. Figure 1 shows the actual per window patterns of browsing for the first participant for the first hour of the study. This participant opened 7 windows and visited 78 pages during the hour. We will refer to this example throughout discussion of the results.

The number of pages loaded in a browser window varied widely within users. In most cases, only one or two pages were viewed within each window. The number of pages opened in a window had a mode of 2 for seventeen participants, while the remaining three had a mode of 1. This is not surprising given the number of windows automatically spawned while browsing. However, there

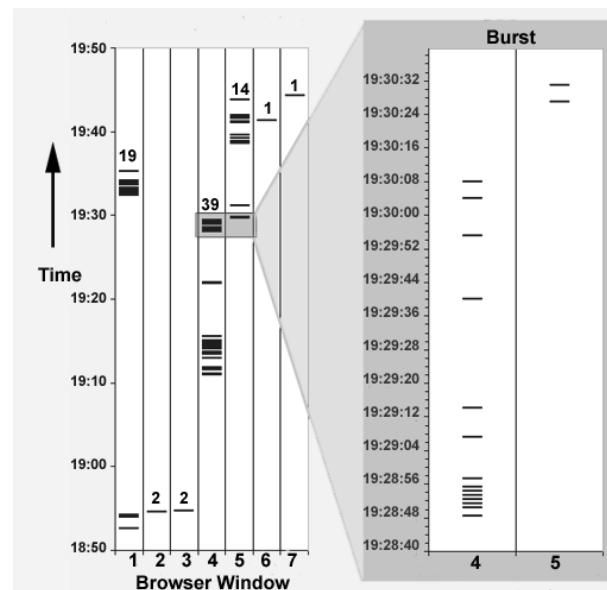


Figure 1. Example of temporal patterns of web browsing on a per window basis. A burst of activity is shown on the right.

	Burst (1 minute cutoff)						Session (10 minute cutoff)						Session (30 minute cutoff)					
	#	Speed	Duration		Length		#	Speed	Duration		Length		#	Speed	Duration		Length	
Quartile		Avg.	Avg.	Max.	Avg.	Max.		Avg.	Avg.	Max.	Avg.	Max.		Avg.	Avg.	Max.	Avg.	Max.
0%	43	8	0.8	5	4	22	14	15	7	38	12	58	9	24	17	60	16	72
25%	156	11	1.2	7	6	38	48	29	9	59	18	105	26	46	21	97	27	143
50%	264	12	1.4	11	7	51	66	33	12	75	25	143	40	53	34	174	38	226
75%	324	13	1.6	13	8	81	82	38	14	88	32	237	48	65	43	223	57	313
100%	510	17	2.1	36	17	172	123	59	23	177	57	394	62	88	54	295	102	805
Mean	258	12	1.4	12	7	63	66	35	13	79	28	174	38	56	33	171	46	258

Table 2. Quartile and mean values for the number of episodes, speed (seconds per page), duration (minutes), and length (pages) of bursts (1 minute cutoff) and sessions (10 minute and 30 minute cutoff) over the course of the week.

were also several instances where large numbers of page views occurred within a browser window. The average maximum number of pages viewed in one window was 108 (from 27 to 255). The browsing shown in Figure 1 is typical: three browser windows of 14, 19, and 39 pages and four windows of 1-2 pages.

People frequently moved between open browser windows. Analyzing the number of switches between browser windows, we find that participants ranged between 22 and 430 window revisitations (average 158) over the course of the week. Figure 1 shows three browser window revisitations (browser window 1 once and window 5 twice).

A thorough literature review has revealed little study of user behaviour with multiple browser windows. Tabbed browsers, such as Mozilla Firefox, allow users to organize multiple open pages within the browser. We have noticed that users may have windows open for different purposes: literature search in one window, email in another, news in a third; and we have found privacy patterns that exist on a per window basis [4]. Per-window behaviours may be useful when incorporating the concept of task into web tools.

Speed of Browsing

Participants frequently (~37 times per day) exhibited rapid bursts of browsing with several pages loaded per minute (Table 2). We define a burst to be a rapid sequence of web visits with less than one minute's elapsed time between web pages loading. Overall, the average duration of a burst was 82 seconds, although the longest burst was over 36 minutes. The average length of a burst was 7 pages, with bursts of up to 172 pages. The average speed during a burst was 12 seconds per page. Figure 1 shows several examples of bursts, including one that runs from 19:28:48-19:30:31 with 16 pages opened in 104 seconds (6.5 seconds/page).

In a study of the history files of 17 users over 4 months (1999-2000) [2], rapid browsing was noted with few gaps longer than 10 seconds per page loading. However, this picture is incomplete as the history files included frames loading within a page. This overestimation of speed is mitigated somewhat as pages were sorted daily by the last time of access and the time of some visits were lost. The authors speculate on the amount of rapid browsing done within a single and multiple windows. Our results show that rapid bursts do indeed occur across windows.

Sessions

We calculated sessions in the same manner as bursts, but with 10 and 30 minutes delimiting sessions. With the 10 minute cutoff, our participants averaged 66 sessions per week (9.4 per day), with an average duration of 13 minutes and length of 28 pages (Table 2). At the 30 minute cutoff, the number of sessions dropped to 38 per week (5.4 per day) with an average duration of 33 minutes and length of 46 pages. Again, there is a large variability in per-session behaviour that may impact any per-session solutions.

In a 1994 client-side study in an academic setting [1], 25.5 minutes was used as the session delimiter. Participants averaged 9.4 sessions over 3 weeks (~1 session every other day). During a longitudinal field study of home internet use in low-income families (circa 2001-2002) [6], participants logged in to the system an average of 0.6 sessions per day. A study examining laptop use by university students (circa 2000) [3] used a 10 minute cut-off and found an average of 3 sessions per day. Because the URL visits in [3] were defined differently from our page visits (counting each element downloaded within the page), we were unable to make any other comparisons.

DISCUSSION

Client-side Logging

The choice of data capture techniques for web browsing behaviour impacts the naturalness of the environment for participants, the ease of developing and supporting tools, and the type of data available. The ability to maintain participant privacy (not recording visited pages externally) and to gather rich information about user activity on a per-window basis led us to a client-side solution. An advantage of the BHO was that the users' browsing environment did not change: they were able to continue using IE with all their normal features and settings intact.

Available logging software does not always include all items of interest to the researcher. We needed to associate a browser window ID with activity, so we developed the software ourselves. It is difficult to develop experimental software that is sufficiently robust in the field for multiple versions of the operating system and web browser. Five participants had difficulties with their software, their hardware, or their internet connections during our study.

These participants did complete seven days of the study, but the days were not all consecutive. Maintaining frequent backups on the participants' laptops of the raw data log and the sanitized report allowed recovery of data when problems arose. Asking participants to email the data daily allowed us to question gaps in received data, then investigate and fix problems as they arose.

Difficulty Comparing Results

It is difficult to compare results between studies due to differences in methodology such as user task and logging software location (client, proxy, server). When comparing quantitative information such as the number of pages visited or session length it is crucial to understand the context of the study [5]. Was all the browsing of the user captured or just that in a certain environment? When was the study conducted? Did the pages visited include cached pages, all pages navigated to, all pages fully loaded, frames, or other page elements such as images etc.? How were sessions discriminated? We found it difficult in many cases to determine pertinent details and to then relate our results to those obtained in a different context. Researchers need to take care to consider fully the usage context when building upon previous results. This is particularly important given the evolving browsing environment.

Impact of Behaviour on Web Browsing Tools

It is important to understand these patterns as web browsing is such a frequent activity. There is a continuing need to research the daily activities and gain an understanding of users' tasks and behaviours [9]. The results from our study clearly demonstrate that variability and magnitude of browsing behaviours complicate the development of any tool or technique for web browsing. The sheer number of pages that people visit while browsing means that manual tools, that operate on a per-page level, will be overly arduous and therefore impractical. Beyond just the number of pages visited, the speed with which users browsed was at times staggering. The high volume of web sites visited and the rapid browsing indicate the need for seamless interactions between user and tool.

Behaviours vary considerably in terms of the number of pages visited, number of separate windows in use, and the session length and speed of browsing. This variability may make it difficult to arrive at standard solutions for web browsing tools and techniques. Furthermore, there is a high amount of variability both across users and within the browsing of a single user. We chose to report descriptive statistics as averages of user behaviour do not provide the full picture. Solutions must be sensitive to the changing needs and behaviours of users and allow users flexibility.

CONCLUSION AND FUTURE WORK

This study advances the understanding of patterns of web browsing activity of highly educated and experienced laptop users. To understand web browsing activity for a wider spectrum of users in different settings, the next stages

of this project will investigate a wider population base, at home, work, and at school to see how browsing patterns may change under different contexts of use. In this first study, we did not elect to record when a browser window closed or when it was in focus, so were unable to learn how many browser windows were concurrently open and how often window contents were viewed. In future studies, we will record this information to build upon our understanding of per-window behaviours during web browsing.

The potential impact that the web browsing behaviours we observed can have on potential web browsing solutions is substantial. Researchers and practitioners must consider the effect of individual differences in user behaviour, high volume of web sites visited, and rapid bursts of browsing when developing solutions and techniques in this area.

ACKNOWLEDGMENTS

Thanks to Melanie Kellar, co-developer of the BHO logging tool, and the other members of the EDGE Lab for their support. Funding provided in part by NSERC.

REFERENCES

1. Catledge, L. and Pitkow, J. (1995). Characterizing Browsing Strategies in the World-Wide Web. In *Proceedings of WWW 1995*, Darmstadt, Germany. 1065 - 1073.
2. Cockburn, A. and McKenzie, B. (2001). What Do Web Users Do? An Empirical Analysis of Web Use. *Int. J. Human-Computer Studies*, 54: 903-922.
3. Grace-Martin, M. and Gay, G. (2001). Web Browsing, Mobile Computing and Academic Performance. *Educational Technology & Society*, 4(3): 95-107.
4. Hawkey, K. and Inkpen, K. (2005). Privacy Gradients: Exploring Ways to Manage Incidental Information During Co-Located Collaboration. In *Extended Abstracts of CHI '05*, Portland, Oregon.
5. Hawkey, K. and Kellar, M. (2004). *Recommendations for Reporting Context in Studies of Web Browsing Behaviour* (No. CS-2004-16). Halifax, NS. Dalhousie University.
6. Jackson, L. A., von Eye, A., Barbatsis, G., Biocca, F., Zhao, Y., and Fitzgerald, H. E. (2003). Internet Attitudes and Internet Use: Some Surprising Findings from the Homenetoo Project. *Int. J. Human-Computer Studies*, 59: 355-382.
7. Pastore, M. (2000). Slow Modems Still Dominate Home Internet Scene. Online at: www.clickz.com/stats/big_picture/hardware/article.php/277191.
8. Tauscher, L. and Greenberg, S. (1997). Revisitation Patterns in World Wide Web Navigation. In *Proceedings of CHI '97*, Atlanta, GA.
9. Whittaker, S., Terveen, L., and Nardi, B. A. (2000). Let's Stop Pushing the Envelope and Start Addressing It: A Reference Task Agenda for HCI. *Human Computer Interaction*, 15: 75-106.