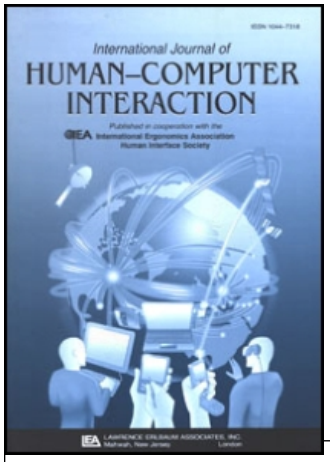


International Journal of Human-Computer Interaction
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t775653655

Challenges of Capturing Natural Web-Based User Behaviors

Melanie Kellar [a]; Kirstie Hawkey [b]; Kori M. Inkpen [c]; Carolyn Watters [c]
[a] Google Inc., Mountain View, CA, USA
[b] Department of Electrical and Computer Engineering, University of British Columbia,
[c] Faculty of Computer Science, Dalhousie University,

Online Publication Date: 01 May 2008

To cite this Article: Kellar, Melanie, Hawkey, Kirstie, Inkpen, Kori M. and Watters, Carolyn (2008) 'Challenges of Capturing Natural Web-Based User Behaviors', International Journal of Human-Computer Interaction, 24:4, 385 — 409

To link to this article: DOI: 10.1080/10447310801973739
URL: http://dx.doi.org/10.1080/10447310801973739

INTL. JOURNAL OF HUMAN–COMPUTER INTERACTION, *24*(4), 385–409, 2008

ISSN: 1044-7318 print / 1532-7590 online
DOI: 10.1080/10447310801973739

Taylor & Francis
Taylor & Francis Group

Challenges of Capturing Natural Web-Based User Behaviors

Melanie Kellar[1], Kirstie Hawkey[2], Kori M. Inkpen[3], and Carolyn Watters[3]
[1]Google Inc., Mountain View, CA
[2]Department of Electrical and Computer Engineering
University of British Columbia,
[3]Faculty of Computer Science, Dalhousie University

It can be difficult to properly understand aspects of user behavior on the Web without examining the behaviors in a realistic setting, such as through field studies. In this article, an overview of the experiences in augmenting logged data with contextual information over the course of two separate research projects conducted in the field is presented. One project investigated the privacy sensitivity of normal Web browsing, and the other examined user behavior during Web-based information-seeking tasks. Throughout both projects, the contextual information was collected through participant annotations of their Web usage. Based on experiences in conducting this research, implications of methodological decisions are considered, unanswered questions are highlighted, and considerations for other researchers are provided. These shared experiences and perspectives will assist future researchers planning similar field studies, allowing them to build upon the lessons learned.

1. INTRODUCTION

Web browsers are an example of an everyday technology used in both workplace and personal settings. Before developing new tools and techniques to support users in their Web tasks, it is important to fully understand how users currently engage with their Web browsers and related tools. Studying user behavior on the Web is complex because behaviors can be influenced by a number of factors, such as domain expertise (Hölscher & Strube, 2000), task (Kellar, Watters, & Shepherd, 2006b), motivation (Loeber & Cristea, 2003), and individual differences (Herder & Juvina, 2004). Web behavioral studies conducted in a field setting can often provide a more realistic picture of natural behaviors than what can be evoked in a

This research was conducted while Melanie Kellar and Kirstie Hawkey were PhD students in the Faculty of Computer Science at Dalhousie University. This research was supported by the Natural Sciences and Engineering Research Council of Canada and NECTAR.

Correspondence should be addressed to Melanie Kellar, Google Inc., 1600 Amphitheatre Parkway, Mountain View, CA 94043. E-mail: kellar@google.com

controlled laboratory setting. In the field, participants have access to their usual Web browser, are in their own physical environment, and are completing their own tasks (rather than tasks motivated by a researcher). However, it is often difficult to observe and record these natural behaviors in dynamic environments such as the Web. This is an ongoing challenge within the community as evidenced by recent workshops (e.g., Edmonds, Hawkey, Kellar, & Turnbull, 2006; Kort & de Poot, 2005).

One common method of studying user behavior in a field environment is through the collection of logged data. This method is unobtrusive to the user and provides researchers with an overview of the user's behavior. However, logged data by itself do not provide a full understanding of users' activities, goals, attitudes, and processes. Contextual information plays an important role in how we understand and interpret people's everyday behavior. Information that provides additional details about people, such as their location or task, can help us better understand and interpret their actions. In a Web environment, contextual information can be used to determine the activity in which a user is engaging; motivations for engaging in that activity; and perceptions about the current tool, the information being viewed, or privacy issues.

We recently conducted two separate research projects that necessitated the collection of contextual information while studying user behavior on the Web. Throughout both projects, contextual information was collected through participant annotations of their Web usage.

The first research project (Hawkey & Inkpen, 2005, 2006a) examined privacy concerns if traces of previous Web-browsing activity could be seen during collaboration. Through two field studies, participants were required to annotate their visited Web pages with privacy ratings (public, semipublic, private, don't save). The goal of the first field study was to determine whether participants would be willing to organize the incidental information visible in their Web browsers across a small number of privacy levels or gradients. We also wanted to explore the existence of privacy patterns on a per-browser window basis to evaluate the feasibility of different privacy management approaches. The goal of the second field study was to determine the relationship between the context of the browsing activity (location, page content) and the privacy comfort levels that participants applied to their Web browsing. We also wanted to examine the feasibility of using automated content categorization to assign privacy levels to visited Web pages.

The second research project (Kellar, Watters, & Shepherd, 2006a, 2007) examined how users interacted with their Web browsers during information-seeking tasks through a field study. In addition to Web browser usage logs, this study necessitated the collection of detailed task information from participants. The goal of this exploratory study was to examine how users interact with their Web browsers to accomplish different information-seeking tasks. We expected that users would interact differently with their Web browser across the following information-seeking tasks: fact finding, information gathering, browsing, transactions, and other. We also wanted to develop a high-level classification of information seeking tasks on the Web.

In this article, we present our recent approaches to Web-based field research and reflect upon the inherent challenges of research conducted in situ. We first present methodologies for studying user behavior on the Web and discuss the

benefits of differing methodological choices. This is followed by an overview of our two recent research projects and the challenges that guided our methodological decisions and tool development. We then reflect on our experiences with respect to participant annotation of contextual information, data collection, data transfer, and data analysis. We conclude with a summary of this work.

2. METHODS FOR STUDYING USER BEHAVIOR ON THE WEB

A wide variety of methodologies can be employed to study user behavior on the Web, and there are trade-offs inherent with each approach. Each methodology impacts the naturalness of the behavior being studied and the amount of contextual information that can be gathered. Surveys (e.g., Aula, Jhaveri, & Käki, 2005) and interviews (e.g., Teevan, Alvarado, Ackerman, & Karger, 2004) offer an in-depth exploration of participants' behavior; however, the data are self-reported and may be collected outside of the relevant context of use. Although surveys are a convenient way to study a large population sample, interviews allow researchers to follow-up on interesting subject matters not previously anticipated. Laboratory experiments (e.g., Hölscher & Strube, 2000) allow researchers to study a particular context, but the tasks are typically contrived and participants are not working in their natural environment. Observational (e.g., Byrne, John, Wehrle, & Crow, 1999; Thury, 1998) studies typically provide a high-level realistic view of user behavior. However, these methodologies discourage interruption of participants, which makes it difficult to solicit contextual information. Methods of collecting contextual information during observational studies include talk-aloud protocols (Byrne et al., 1999) and post hoc interviews where participants explain their actions using still pictures (Brown, Sellen, & O'Hara, 2000), recorded video (Allen, 1989), or logged data (Choo, Detlor, & Turnbull, 2000). Field studies (e.g., Kelly & Belkin, 2004) can offer a fairly natural view of user behavior, but care must be taken to ensure that the data collection does not reduce the naturalness of the environment.

The focus of this article is the collection of fine-grained information about users' activities in their natural environment. We first discuss the mechanics of logging user activity, including the location of the logging (i.e., client-side, server-side, or through a proxy). Further to the discussion of logging activity, we include an overview of methods of gathering contextual information, both as the primary measure and in conjunction with logged data of activities. We finish with a discussion of the privacy issues related to logging participant' activities.

2.1. Logging Activity

Client-side logging takes place on the user's own computer. Researchers have used a number of tools to collect client-side logs, including commercial "spyware" tools (Kelly & Belkin, 2004; Kim & Allen, 2002), custom-built logging tools (Obendorf, Weinreich, & Hass, 2004; Reeder, Pirolli & Card, 2001; Turnbull, 1998), and custom Web browsers (Claypool, Le, Waseda, & Brown, 2001). Client-side logging has been used to capture a variety of behaviors and to study a

wide-range of research areas. For example, Catledge and Pitkow (1995) studied the navigation patterns of users on the Web through an instrumented Web browser, which logged all Web pages visited as well as user interface events within the browser. Client-side logging offers the richest exploration of user interactions with the Web browser. However, many client-side logging tools are designed to work with a specific browser and may be time-consuming and costly to update as new versions of the browser are introduced. There may also be performance issues with a custom Web browser because of lack of robustness.

Server-side logs do not capture the same level of detail as is possible with client-side logging; however, there is a lower cost of implementation. One benefit of server-side logging is that researchers can study a large sample population. This method has been successfully used to study search engine transaction logs (Jansen, Spink, & Pedersen, 2005; Spink, Wolfram, Jansen, & Saracevic, 2001) and library portals (Zhang, Zambrowicz, Zhou, & Roderer, 2004). The data recorded typically includes the IP address of users and the time and address of Web page requests.

With server-side logging, researchers have very little information about the participants being studied; users are typically anonymous. Jansen, Spink, and Saracevic (2000) analyzed Excite search engine logs and acknowledged that although the data reflect real search behavior, they can "report on artifactual behavior, but without a context." Zhang et al. (2004 page 209), who logged users of the library portal MyWelch, stated that additional studies in the form of interviews or surveys were needed to better understand users' behavior from a "mental or cognitive" perspective as well to get a sense of other applications and multitasking.

Logging conducted through a proxy sever is a compromise between client-side and server-side logging. Proxy servers act as an intermediary between a user's Web browser and a Web server and can log interactions between the client and the server. By allowing participants to log in to the proxy server instead of downloading and installing software, proxy solutions such as WebQuilt (Hong, Heer, Waterson, & Landay, 2001) allow participants to work within their normal browsing environment. However, proxy server logging does not capture the full spectrum of user interactions with the browser and may not capture access to pages that have been cached at the browser level (Barford, Bestavros, Bradley, & Crovella, 1999). Proxy sever logging may also be problematic when trying to collect fine-grained measurements. Kelly and Belkin (2004) found a large discrepancy between a client-side logging tool and a proxy-based logging tool while collecting Web page dwell times; the timing data generated by the proxy-based logger was found to be inaccurate when compared to the client-side logging tool.

2.2. Gathering Contextual Information

A further consideration when conducting field research is capturing the context in which users perform their activities. Depending on the research question, contextual information can consist of not only the environmental setting but also the users' motivations, intentions, goals, tasks, and perceptions about their Web behavior. Contextual information can be gathered at a high level, such as through interviews and surveys (Aula et al., 2005; Bruce, Jones, & Dumais, 2004;

Rieh, 2003; Teevan et al., 2004). However, to interpret logged activities, it is important that contextual information is collected at a more fine-grained level. We next present various approaches for augmenting logged data in the field with fine-grained contextual information.

Several researchers have used client-side logging methods supplemented with interviews after the fact to gather additional contextual information. For instance, Choo et al. (2000) studied critical incidents of information seeking on the Web. After logging participant Web usage over a 2-week period, they conducted interviews using preanalyzed Web usage logs to guide the discussions with participants. Sellen, Murphy, and Shaw (2002) studied the Web activities of knowledge workers by interviewing participants about their web activities from the previous 2 days, while seated in front of their browser history lists.

Other researchers have used periodic contextual annotations to supplement logged data. For example, Kelly and Belkin (2004) studied the use of display time as a measure of user interest across information-seeking tasks. They conducted a 14-week field study, during which all participants' Web usage was logged. Once a week, participants were asked to annotate their Web usage with task information and to indicate the usefulness of viewed Web pages.

A further consideration is the scope of data collected. Depending on the research question, it may be appropriate to study participants in a fixed context, such as a single physical location (e.g., home; Rieh, 2003) or within a single virtual environment (e.g., Web portals; Cothey, 2002). However, it may also be necessary to gain an overall view of user behaviors across contexts. It can be logistically difficult to observe users across multiple contexts (e.g., home and work, across all information-seeking tasks, viewing both private and public Web pages) as we must devise methods to collect contextual information without interfering with participants' natural behaviors.

2.3. Participant Privacy

Participant privacy is an issue when researchers log participants' Web-based activities outside of an experimental setting. Privacy supports social interaction and "healthy functioning by providing needed opportunities to relax, to be one's self, to emotionally vent, and to cope with loss, shock, and sorrow" (Margulis, 2003, p. 246). Increasingly the Internet has become a mechanism by which people can engage in activities to support these needs (e.g., surfing the Web, visiting personal support forums, blogging, investigating health concerns). Sensitive content within Web browsers may include items such as socially inappropriate activities, confidential business items, and personal activities conducted on company time. Olson, Grudin, and Horvitz (2005) examined the privacy of several types of shared electronic information. They found that personal activities (e.g., viewing nonwork-related Web sites) and transgressions (e.g., viewing erotic material) are considered more sensitive than content such as availability and contact information. Furthermore, knowledge of activities can be even more sensitive when a user's identity is known, because their hidden personae may be revealed (Lederer, Mankoff, & Dey, 2003). Privacy comfort levels when traces of

Web activity are visible depend on the perceived sensitivity of the captured browsing, the amount of control retained, the relationship to the viewer of the information, and inherent privacy concerns (Hawkey & Inkpen, 2006b).

Privacy and surveillance are aspects of the same concept, with privacy actions serving as a nullification mechanism against surveillance (Marx, 2003). Eleven methods of maintaining privacy in case of surveillance have been identified (Marx, 2003). Five of the methods are particularly applicable during field studies including *self-regulating*, *switching*, *blocking*, *masking*, and *refusal* moves. For example, participants may self-regulate their Web-browsing activities during the study period by avoiding activities they perceive to be socially undesirable (Fisher, 1993). A person's attitudes and perceptions about privacy, trust, and social relationships or norms will influence his behavior (Liu, Marchewka, Lu, & Yu, 2004). Partcipants may switch their activities to another computer or another program to avoid surveillance. The may also block the recording of visited sites by turning off the logging software. Alternatively, participants may be provided with methods to mask the activity rather than to block it completely. Finally, participants may refuse to take part in the study if they feel that the privacy violations are untenable.

3. OUR RESEARCH EXPERIENCES

This section describes two separate research projects that recently studied user behavior on the Web. Both research projects were conducted in a field setting and necessitated the collection of contextual information. Contextual information was collected through participants' annotation of their Web activity logs throughout both research projects. The first project consisted of two studies, which we refer to as IIP1 and IIP2, and investigated users' privacy concerns with respect to the incidental viewing of previous Web activity. The contextual information collected during this study consisted of participants' privacy perceptions if each visited page was later viewed. The second project, which we refer to as the ISB study, examined information-seeking behavior on the Web. The contextual information collected during this study consisted of task descriptions and categorizations of the information-seeking activity motivating the Web browser usage. In this section, we present a profile of the study participants, discuss the data collection challenges that influenced our methodological decisions and tool design, and provide a brief summary of the study findings.

3.1. Studying Incidental Information Privacy in Web Browsers (IIP1)

A weeklong field study was conducted to investigate participants' incidental information privacy concerns in Web browsers during collaboration around their personal display (Hawkey & Inkpen, 2005). We define incidental information as the information that is visible on a computer display that is not pertinent to the task at hand. Web browsers have several convenience features such as History and Auto Complete that capture traces of previous activity and may reveal this incidental information during later collaboration around a personal display. This

incidental information may or may not be appropriate depending on the viewing context. A four-tier privacy scheme (public, semipublic, private, don't save) was proposed to see if that level of granularity was appropriate to reflect the privacy needs between types of Web sites and potential viewing audience.

Participants. Twenty participants were recruited from the Dalhousie University community. To qualify for inclusion, participants needed to be Microsoft Internet Explorer (IE) users who performed the majority of their Web browsing on a laptop computer. This allowed us to capture most of participants' personal and work/school-related Web browsing as they moved between these locations with their laptop. Participants also needed to have had occasions in the past where their Web browser window was visible by others, so that the concept of privacy in this situation had some relevance.

Participants (16 male, 4 female) ranged in age from 19 to 47 (average = 26.1). Participants were primarily technical (14 computer science, 4 science) and highly educated, with 65% having completed at least an undergraduate degree. There were 18 students, 1 professor, and an information technology professional. Participants were generally experienced computer users (*Mdn* = 10 years) and spent a considerable amount of time each week using their computer (*Mdn* = 29–35 hr/week) and Web browsers (*Mdn* = 22–28 hr/week).

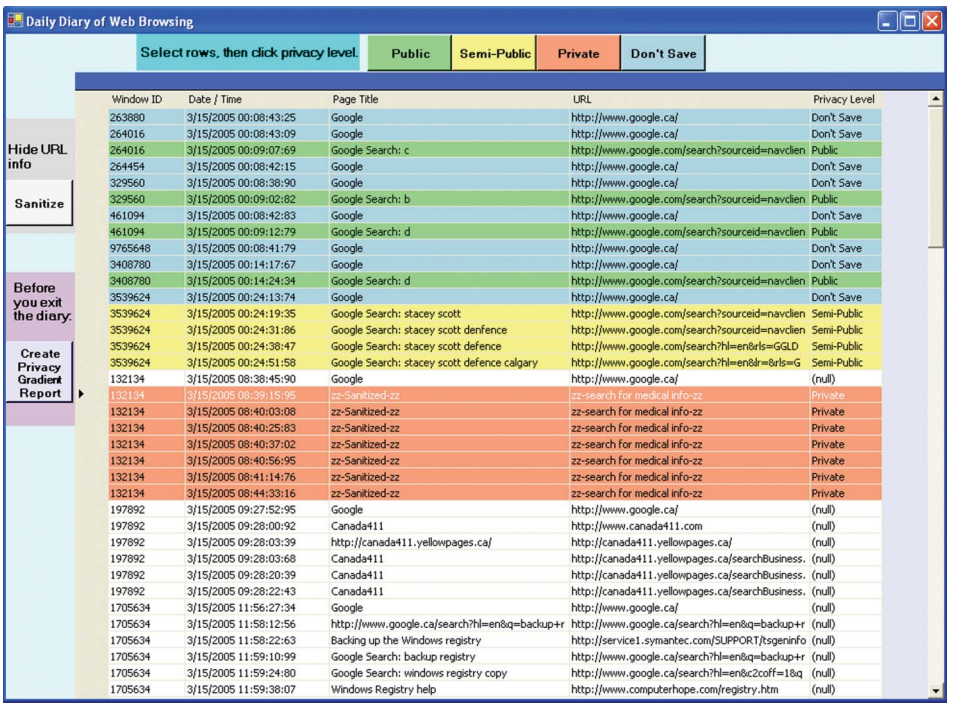
Data collection. There were two primary categories of data collected during this study. The first category consisted of a record of the Web page visits, including the date/time stamp, page title, and URL. Furthermore, to investigate patterns that may occur on a per-window basis, the browser window in which the page visit occurred needed to be recorded. The second category consisted of participants' perceived privacy of their Web usage. Standard logging tools did not support our data collection requirements. Although there are several research and commercial logging tools that record visited page data, none include the browser window ID. We therefore had to develop two client-side data collection tools: one to log users' Web activities and the other to allow participants to daily annotate their Web activity with a privacy rating.

The design of the data collection tools presented several challenges. First, we needed to explore normal Web-browsing activities to see if privacy patterns existed. Therefore, it was important that the experimental software not interrupt the flow of participants' Web browsing (Novak, Hoffman, & Yung, 2000). Second, we also wanted to maintain the participant's normal Web-browsing environment (i.e., their usual Web browser with all the convenience features and settings intact). Finally, we were also concerned about participants' privacy; we did not want the recording of the sites visited to impact their normal Web-browsing activity. Given the nature of the study, it was important that participants visit Web sites as they normally would, across their normal range of content sensitivity, regardless of the social desirability of the content.

The ability to maintain participant privacy (recording data locally) and to gather rich information about user activity on a per-window basis led us to a

client-side solution. To record the browsing activity of participants, a browser helper object (BHO) was implemented to work with IE. A BHO is a .dll file that loads every time IE loads. As each IE window opens, the BHO loads and logs all Web sites visited until the window closes. For this study, the visited Web page (URL and page title), time stamp, and ID number of the browser window were recorded. All pages viewed in the browsing process were logged, even if navigation continued before the document fully loaded. Individual frames or images loaded within a Web document were not logged. An advantage of the BHO was that the users' browsing environment did not change; they continued using IE with their normal settings intact.

An electronic diary was developed to allow participants to assign privacy gradients to their Web browsing on a daily basis (similar to that shown in Figure 1). The diary displayed all the logged data and required participants to indicate how they would classify the privacy level of each Web page they visited if others were able to view the history of this activity later. Participants could annotate individual entries with a privacy level or select multiple entries for annotation. The entries could be sorted by any field (time, URL, page title), allowing participants to easily classify groups of page visits (e.g., repeated visits to the same site). After



Window ID	Date / Time	Page Title	URL	Privacy Level
263880	3/15/2005 00:08:43:25	Google	http://www.google.ca/	Don't Save
264016	3/15/2005 00:08:43:09	Google	http://www.google.ca/	Don't Save
264016	3/15/2005 00:09:07:69	Google Search: c	http://www.google.com/search?sourceid=navclien	Public
264454	3/15/2005 00:08:42:15	Google	http://www.google.ca/	Don't Save
329560	3/15/2005 00:08:38:90	Google	http://www.google.ca/	Don't Save
329560	3/15/2005 00:09:02:82	Google Search: b	http://www.google.com/search?sourceid=navclien	Public
461094	3/15/2005 00:08:42:83	Google	http://www.google.ca/	Don't Save
461094	3/15/2005 00:09:12:79	Google Search: d	http://www.google.com/search?sourceid=navclien	Public
9765648	3/15/2005 00:08:41:79	Google	http://www.google.ca/	Don't Save
3408780	3/15/2005 00:14:17:67	Google	http://www.google.ca/	Don't Save
3408780	3/15/2005 00:14:24:34	Google Search: d	http://www.google.com/search?sourceid=navclien	Public
3539624	3/15/2005 00:24:13:74	Google	http://www.google.ca/	Don't Save
3539624	3/15/2005 00:24:19:35	Google Search: stacey scott	http://www.google.com/search?sourceid=navclien	Semi-Public
3539624	3/15/2005 00:24:31:86	Google Search: stacey scott denfence	http://www.google.com/search?sourceid=navclien	Semi-Public
3539624	3/15/2005 00:24:38:47	Google Search: stacey scott defence	http://www.google.com/search?hl=en&rls=GGLD	Semi-Public
3539624	3/15/2005 00:24:51:58	Google Search: stacey scott defence calgary	http://www.google.com/search?hl=en&lr=&rls=G	Semi-Public
132134	3/15/2005 08:38:45:90	Google	http://www.google.ca/	(null)
132134	3/15/2005 08:39:15:95	zz-Sanitized-zz	zz-search for medical info-zz	Private
132134	3/15/2005 08:40:03:08	zz-Sanitized-zz	zz-search for medical info-zz	Private
132134	3/15/2005 08:40:25:83	zz-Sanitized-zz	zz-search for medical info-zz	Private
132134	3/15/2005 08:40:37:02	zz-Sanitized-zz	zz-search for medical info-zz	Private
132134	3/15/2005 08:40:56:95	zz-Sanitized-zz	zz-search for medical info-zz	Private
132134	3/15/2005 08:41:14:76	zz-Sanitized-zz	zz-search for medical info-zz	Private
132134	3/15/2005 08:44:33:16	zz-Sanitized-zz	zz-search for medical info-zz	Private
197892	3/15/2005 09:27:52:95	Google	http://www.google.ca/	(null)
197892	3/15/2005 09:28:00:92	Canada411	http://www.canada411.com	(null)
197892	3/15/2005 09:28:03:39	http://canada411.yellowpages.ca/	http://canada411.yellowpages.ca/	(null)
197892	3/15/2005 09:28:03:68	Canada411	http://canada411.yellowpages.ca/searchBusiness.	(null)
197892	3/15/2005 09:28:20:39	Canada411	http://canada411.yellowpages.ca/searchBusiness.	(null)
197892	3/15/2005 09:28:22:43	Canada411	http://canada411.yellowpages.ca/searchBusiness.	(null)
1705634	3/15/2005 11:56:27:34	Google	http://www.google.ca/	(null)
1705634	3/15/2005 11:58:12:56	http://www.google.ca/search?hl=en&q=backup+r	http://www.google.ca/search?hl=en&q=backup+r	(null)
1705634	3/15/2005 11:58:22:63	Backing up the Windows registry	http://service1.symantec.com/SUPPORT/tsgeninfo	(null)
1705634	3/15/2005 11:59:10:99	Google Search: backup registry	http://www.google.ca/search?hl=en&q=backup+r	(null)
1705634	3/15/2005 11:59:24:80	Google Search: windows registry copy	http://www.google.ca/search?hl=en&c2coff=1&q	(null)
1705634	3/15/2005 11:59:38:07	Windows Registry help	http://www.computerhope.com/registry.htm	(null)

FIGURE 1 A screenshot of the electronic diary used in the second privacy field study. The diary allowed participants to annotate their visited pages with a privacy level. *Note*: This is similar to the diary used in the first field study with the addition of selective blinding of page titles and URLs.

classification, participants clicked a button to generate a report to e-mail to the researchers. In this report, the viewing history was sanitized so that the URL and page title were eliminated (to protect participant privacy). Participants could view the data about to be sent but could not modify it. Although this approach to maintaining privacy was designed to maximize the participants' willingness to engage in their usual browsing activities, the lack of URL information means that the number of unique Web sites visited or the extent of site revisitation is unknown.

In addition to the diary portion of the study, participants completed pre- and poststudy questionnaires that included questions about demographics, their current browser convenience feature settings, and privacy management strategies. In addition, two theoretical privacy classification tasks were given to participants. One asked them to classify the privacy of categories of Web sites (based on content) into the four levels (public, semipublic, private, and don't save). The other asked them to classify categories of viewers at one of three levels: allowed to only view pages classified as public, allowed to view pages classified as both public and semipublic, and allowed to view all visited pages.

Summary of research findings. Results showed that the magnitude of incidental information complicates any privacy management approach to incidental information in Web browsers. The sheer number of pages that people visit (36,170 pages were viewed over the course of the study) and the speed at which browsing can occur (i.e., frequent short bursts of about five pages per minute) is staggering. Any manual solution would be overly arduous and therefore impractical.

Participants' behaviors varied considerably in terms of the number of pages visited, number of separate windows in use, and the overall application of the privacy levels. With the exception of one participant (who did not use the don't save category), participants utilized all four privacy categories when classifying their visited pages. Patterns in the application of privacy levels emerged from the data. Participants tended to partition their browsing so that private browsing occurred within a single window. Within a window, most browsing (85% of page visits) occurs within streaks (i.e., two or more consecutive pages) at a given privacy level. There were also relatively few transitions between levels, with an average of 0.9 transitions in a browser window. Given the per-window patterns of privacy streaks with minimal transitions, we believe that one management approach may be to allow browser windows of different privacy levels. These windows could not only filter what incidental information is displayed but also tag new sites visited in that window. However, such a scheme would require integration with a more proactive approach to be manageable for users.

3.2. Studying Incidental Information Privacy in Web Browsers (IIP2)

A second field study was conducted in March 2005 to extend the information learned in the first study (Hawkey & Inkpen, 2006a). We examining additional

contextual information that was not collected during the first field study, such as the page title, URL, browser window interactions, and location of the browsing

Participants. Fifteen participants were recruited from the general Dalhousie University community. The first field study consisted solely of laptop users with a primarily technical background. To determine if results were consistent across a broader population, this second study included participants with varying technical experience and computer use. Three different classes of participants were recruited: technical desktop users, nontechnical desktop users, and nontechnical laptop users. A screening process assessed participants' technical background and identified computers on which they conducted their Web browsing. Participants were required to use IE and to have logging software installed on all their computer(s) so that the full picture of their personal and work/school-related Web browsing could be captured.

Participants (5 male, 10 female) ranged in age from 18 to 44 (average = 27.8). Participants were highly educated, and 5 had a technical background. Eleven were students, and 4 were office or administrative staff. Participants were generally experienced computer users (average = 9.7 years) and spent a considerable amount of time each week using their computer (*Mdn* = 29–35 hr/week) and Web browsers (*Mdn* = 15–21 hr/week). Given the educational domain from which participants were recruited, browsing activities may include more educational and reference sites than if participants were from another domain.

Data collection. As in our previous study, data collection consisted of date/time stamp, page title and URL of visited pages, and the browser window ID. Two new categories of data collection were introduced in this study to capture the desired additional contextual information. First, the physical location of the Web browser was logged (e.g., home, work, school). Second, the window focus events were logged so that we could determine when participants moved between windows, not just when they moved between windows for the purpose of navigating to a new page. We also logged the IE window open and close events. The BHO was modified to record this additional contextual information. Participants' location was hard coded into the BHO installed on desktop computers. Laptop users indicated their current location with a radio button that appeared in a form as the browser window closed; options were home, work, school, and other (a text box was provided for entry of the specific location).

As we wanted to not only collect the URL and page title but also send that information to the researchers, we needed to readdress privacy issues. We did not want the transfer of this information to impact participants' willingness to visit sensitive sites (e.g., erotica, health-related sites, personal improvement forums). Therefore, we wanted to provide participants with the ability to selectively any sensitive data. The electronic diary (see Figure 1) was modified to allow participants to mask entries in the diary by removing the page title and URL after applying a privacy level. Participants were asked to give a general reason for the sanitized browsing (e.g., "looking for medical information"); the default label was "no reason given." After classification, participants generated a report to e-mail to the researchers.

We expected that the privacy afforded by allowing participants to selectively sanitize their browsing record would contribute to their willingness to engage in normal Web activities while still providing us with the context for most visited pages.

In addition to the diary portion of the study, participants completed pre- and poststudy questionnaires. As in the first field study, one of the questionnaires was a theoretical classification task where participants were asked to assign one of the four privacy levels to each of 55 Web site categories indicating their comfort if a site of this type appeared in their Web browser. The categories (e.g., online games, news/media) and their descriptions were based on those used in commercial products to filter and block Internet content. The same categories were used by researchers to classify all of the browser activity conducted by participants over the course of the week.

Summary of research findings. Findings were similar to the first study in terms of browsing activity and overall application of privacy levels. As in the previous study, we observed heavy Web usage by participants, with 31,160 Web pages viewed. With the collection of page title and URL data in this study, we expected that participants would sanitize their browsing records when viewing sensitive content. However, only 6 of 15 participants had occasions of masking, for a total of 433 of 31,160 page visits. It is interesting that the participant with the highest number of masked sites was a colleague at the university; those participants who had no prior acquaintance with the researchers exhibited less masking behavior.

Results of the content analysis revealed that the categories of Web pages clustered into five groups based on participants' overall application of privacy levels to their Web browsing. Inconsistencies between participants, both for their theoretical and actual privacy classifications, suggest that a general privacy management scheme is inappropriate. Although participants often applied different privacy levels from each other for categories, results showed that participants were personally consistent within most categories. This suggests that a personalized scheme may be feasible, but a more fine-grained approach to classification is required to improve results for Web sites that tend to be very general, have multiple task purposes, or have dynamic content. In addition, participants' overall poor accuracy at specifying theoretically how they will actually label the Web sites in a category indicates that better descriptions of the types of sites that may fall within a category, as well as the types of sensitive information that may be revealed, is required.

The results of the poststudy questionnaires from both studies (IIP1 and IIP2) were used to examined whether our participants found it difficult to categorize their Web usage using the given privacy categories. The majority of participants (32/35) reported that the four privacy categories fit well at least most of the time; however, many (17/35) reported difficulty classifying some of the visited sites (approximately 15% of visited sites). Reasons given for the difficulty included that it depended on the person they envisioned viewing a record of the page visit (10 of 17), that it depended on the viewing location (7 of 17), that the site had multiple purposes (5 of 17), or that there were other reasons (5 of 17; e.g., the time of day, variations in content).

3.3. Studying Information-Seeking Behavior on the Web (ISB)

We conducted a 1-week field study to study users' information seeking behavior on the Web (Kellar et al., 2007). Over the course of the study, participants were asked to use a custom Web browser, which logged all of their browser interactions and to annotate all their Web usage according to the following categories: fact finding, information gathering, browsing, transactions, and other.

Participants. Twenty-one university students from Dalhousie University took part in a 1-week field study in March 2005. To take part in the study, it was necessary that participants were both laptop and IE users. Laptop users were recruited because we could capture the bulk of their usage on a single machine and because it facilitated the installation of the logging software. Because the Web browser used during the study was a clone of IE, participants were also required to be current users of IE.

The academic background of the participants was divided among computer science (11 participants), health informatics (2), business (4), economics (2), kinesiology (1), and arts (1). The median age group category of the participants was 20 to 29, and the gender was almost evenly split with 11 male and 10 female participants. The median category of Web usage reported by the participants was between 30 and 39 hours of usage a week. Although computer science students are typically considered to be highly technical, all participants were experienced Web users. All participants were the primary users of their laptops, and 5 participants reported they used a desktop (either at home or work) for some of their Web usage.

Data collection. The two categories of data collected in this study consisted of interactions with the Web browser and detailed task information. It was necessary that the Web browser interactions were captured at a fine-grained level (e.g., use of bookmarks, auto-complete, and copy/paste/cut functions). Detailed task information consisted of a categorization of the task (fact finding, information gathering, browsing, transactions, and other) and a short textual description (e.g., reading news, checking stock prices).

We examined various commercial and academic logging programs and found that they typically collected a small subset of interactions with the browser, such as the time and URLs of visited pages, and did not log how each Web page was accessed (e.g., bookmarks, typed in) or any other interactions with the Web browser. We also examined instrumented Web browsers but found they were out of date and missing standard browser functionality. Another alternative was the use of a BHO; however, BHOs cannot track fine-grained user interactions with the browser. We also explored the use of screen capture software. This approach presented two problems. First, the software created a delay on older systems, and we did not want to limit our recruitment to participants with high-powered machines. Second, all captured video would then have to be coded by hand, which would be extremely time-consuming.

We also encountered several methodological challenges while undertaking this research. One of the most challenging aspects of this work stemmed from our data collection needs. In particular, we were concerned with how to collect detailed data without changing users' natural behaviors. It was also important that users maintained their normal Web browsing environment, including navigation and information-seeking tools such as history, bookmarks, and search toolbars. Furthermore, we were concerned with how to protect participants' privacy, as we were asking them to submit logs of their Web usage. In particular, we did not want the logging to influence their behavior.

In response to these problems, we developed our own custom logging browser (shown in Figure 2a) using Microsoft's browser control object. This allowed us to build a browser that mimicked IE and provide users with access to their IE bookmarks, settings, and history. Participants were asked to use the custom Web browser for the duration of the study. Two types of Web browser events were logged: browser functions and document complete events. Browser functions consisted of all button, text, and menu interactions with the Web browser (e.g., opening and closing a window, printing or saving a document, edit functions such as cut/copy/paste). Document complete events captured page navigation and logged URL and page titles, date/time, window ID, and the navigation tool used to access the page (e.g., Back button, auto-complete, bookmarks, history).

We also had to provide a mechanism for collecting the task categorizations (fact finding, information gathering, browsing, transactions, and other) and descriptions in a way that was as nonobtrusive as possible. We iterated on several different data collection interfaces and techniques, including pop-up windows, toolbars, and electronic diaries. Two interfaces were pilot tested: a task toolbar

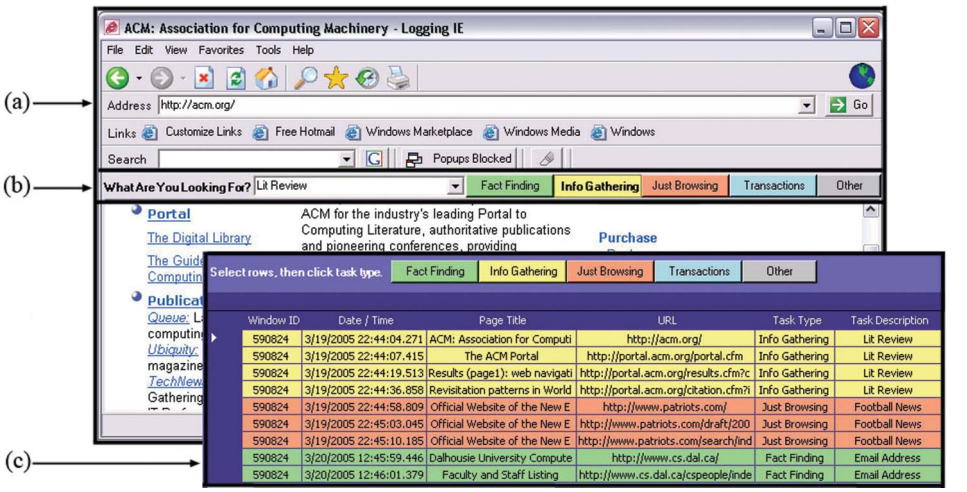


FIGURE 2 The custom Web browser (a) was built to mimic IE and provided a task toolbar (b) for participants to record their task information in real time. Participations could also use the task diary (c) to record their task information at the end of each day.

embedded in the Web browser and a stand-alone electronic diary (similar to the tool developed for the privacy study). During the pilot study, participants were evenly split on their preference for the two techniques. Therefore, participants were provided with access to both methods.

The task toolbar (shown in Figure 2b) was used by participants to provide task information in real time. Participants who used the toolbar method were instructed to fill in the toolbar before beginning a new task. Tool tips displaying task definitions were displayed when a participant hovered over one of the task buttons. An auto-complete function was implemented for the textual description based on feedback received during the pilot study. Participants quickly built a small library of tasks to choose from when assigning textual descriptions.

The task diary (shown in Figure 2c) was used by participants to provide task information at the end of each day. Using the task diary, participants could mass assign task information to multiple URLs at once, allowing them to quickly annotate their day's usage. As with the task toolbar, an auto-complete function was implemented that allowed participants to quickly assign task information. The items in the auto-complete function were shared between the toolbar and the task diary. Tool tips displaying task definitions were displayed when a participant hovered over one of the task buttons.

To protect participant privacy, we had to allow participants to remove any URLs that they did not want to share with researchers. Therefore, regardless of whether participants chose to use the task diary or toolbar to annotate their data, before e-mailing a daily task report to the researchers, participants could use the task diary to blind any Web site addresses they were uncomfortable sharing. We expected that this would encourage participants to engage in their normal activities on the Web.

In addition to the logged data collection, participants completed three separate questionnaires over the course of the study. Before beginning the study, a demographic questionnaire was used to collect participants' demographic information and current Web usage. An inventory questionnaire of the Web browser navigation mechanisms used was also completed by participants at this time. Upon completion of the study, participants completed a poststudy questionnaire, which examined any difficulties they encountered during the study.

Summary of findings. In this study, participants viewed 13,498 pages and reported 1,192 tasks. We found several differences in how users interacted with their Web browser during different tasks. Information-gathering tasks were the longest in duration and were search based with a heavy use of the Google search engine. They were also characterized by a large number of pages viewed and the greatest use of browser functions (e.g., copy, print, save). Fact-finding tasks were short lived and search based with a heavy use of Google. Browsing tasks were also short lived, and we observed a high level of revisitation. Transactions consisted mostly of e-mail, were short in duration, and were the most often repeated task.

Based on results of this study and previous research (Choo et al., 2000; Morrison, Pirolli, & Card, 2001; Sellen et al., 2002), we developed a Web information task

classification, which consists of three top-level information goals: information seeking, information exchange, and information maintenance. Information seeking consist of tasks in which users are trying to change their state of knowledge. Information exchange consist of online actions including transactions (e.g., banking) and communications (e.g., e-mail). Information maintenance consists of visits to Web pages to maintain the information (e.g., Web page development).

Through the poststudy questionnaires, participants were asked to rate the ease of categorizing their tasks using a 5-point Likert scale from 1 (*very hard*) to 5 (*very easy*). They reported that they found information-gathering tasks the most difficult to categorize (*Mdn* rating = 3 [*neutral*]), followed by browsing and fact finding (*Mdn* rating = 4, respectively [*easy*]), and then transactions (*Mdn* rating = 5 [*very easy*]). The majority of participants (17) stated that the two most difficult categories to distinguish between were the fact-finding and information-gathering tasks. Participants were also asked to rate the accuracy of their task categorizations and descriptions on a 5-point Likert scale, from 1 (*strongly disagree*) to 5 (*strongly agree*). Almost all participants reported that they agreed (i.e., a rating of 4 or 5) that their task categorizations (21 of 21) and task descriptions (19 of 21) were accurate.

We were also interested in how (a) requiring users to annotate their Web usage and (b) using a custom Web browser impacted how participants usually work on the Web. Using a 5-point Likert scale from 1 (*not at all*) to 5 (*a great deal*), participants reported a median rating of *a little* (rating of 2) for both measures. Some participants reported that using the task toolbar to annotate their Web information influenced their usual task-switching habits. For instance, one participant reported that instead of switching between multiple tasks during the study, she would sometimes fully complete one task before beginning a new task because this would minimize the number of task information updates required. Most participants (17) reported that they preferred to provide their annotations in real time using the toolbar; some indicated it was difficult to provide detailed task information after the fact, even with the provided convenience features (e.g., sorting capabilities, auto-complete). Although these data are subjective, they do allow us to gain some insight into how our methodological decisions may have impacted participants' behavior.

4. REFLECTIONS

In this section we reflect on our experiences in collecting contextual information on the Web. We discuss our approaches to several methodological challenges, including methods for enabling participant annotation of data, data collection, data transfer, and data analysis. Many of our reflections are based on our choice of methodologies and our result related to the collection of contextual information, a summary of which is provided in Table 1. Although Table 1 highlights the commonalities across studies, it is important to note that the IIP studies (1 and 2) and the ISB study were designed to address different research questions, and therefore different poststudy questionnaires were administered.

Table 1: A Summary of the Methods and Results Related to Collection of Contextual Information Across All Three Studies

	Incidental Information Privacy		
	IIP1	*IIP2*	*Information Seeking Behavior (ISB)*
Participants	20	15	21
Study duration	7 days	7 days	7 days
Total Web pages viewed	36,170[a]	31,160[b]	13,498[c]
Masked URLs	NA	< 1%	< 1%
Annotations	Post hoc	Post hoc	81% preferred real time
Data collection tool	Browser helper object (IE)	Browser helper object (IE)	Custom Web browser (IE)
Logged measures	Date & time, URL & page title, Window ID	Date & time, URL & page title, Window ID, Participant Location, Window focus events	Date & time, URL & page title, Window ID, Navigation mechanisms, Browser functions (e.g., print, copy, find)
Participants' poststudy questionnaire reponses			
Difficulty with categorizations	90% reported that categories fit well most of the time. 40% reported some difficulty classifying Web sites	93% reported that categories fit well most of the time. 60% reported some difficulty classifying Web sites	Information gathering was the most difficult task to categorize. 81% reported difficulty in distinguishing between fact finding and information gathering.
Accuracy of categorizations	NA	NA	100% agreed that task categorizations were accurate. 90% agreed that task descriptions were accurate.
Impact of categorizations	NA	NA	76% reported that having to categorize their web usage only impacted their behavior a little.

Note. NA = not available; IE = Microsoft Internet Explorer Internet browser.
[a]M = 1,081. [b]M = 2,077. [c]M = 643.

4.1. Participant Annotation of Data

When requiring participants to annotate their behavior, a number of factors impact the quality of the data, including the categorization schema provided to participants, the duration of the study, and the time of annotation (real-time vs. post hoc). In this section, we reflect on our experiences in dealing with these factors.

Categorization of behavior. Often when collecting contextual information in the field, participants are required not only to describe their actions or intent but also to characterize their own behavior within a previously defined schema. For participants to do this, they must be trained so that they properly understand the categorization scheme. Furthermore, if participants must assign categories as was done in our studies, it is important that the categories are obvious and easily distinguishable. Depending on the characterization scheme, some types of annotations may be more difficult for participants than others.

In the ISB study, we iterated several times on the task categorization, relying on previous research (Choo et al., 2000; Morrison et al., 2001; Sellen et al., 2002), a pilot study, and a focus group. Despite our efforts to clarify the categorizations through training tasks (e.g., classify "looking for a pizza dough recipe"), handouts, and reminder tool tips, participants still encountered some difficulties in categorizing their behavior (i.e., fact finding, information gathering, browsing, transactions, and other). In particular, participants struggled most with differentiating between fact finding and information gathering, which are two search-based tasks.

Participants in the IIP1 and IIP2 studies were trained on the classification scheme (e.g., a job search might be inappropriate for a boss to see but fine for a close friend, so that may be considered semipublic) during the installation session and provided with a reference handout. Most participants found the four privacy category schemes fit well; however, approximately half of the participants did experience difficulties classifying some of the visited sites, as the privacy concerns were dependent on contextual factors such as the viewing location or who they envisioned seeing the page.

The willingness of the participants to carefully and thoughtfully annotate their data, as well as the required frequency of the annotations, must be considered when evaluating the accuracy of the annotations. For example, we encountered one participant during the ISB study whose data were not usable because their task information was simply too vague and the participant was unwilling to provide any clarification. Participant fatigue may cause accuracy to decline over time; however, it may also improve as participants become more skilled and comfortable with the categorization schema. An open question remains as to whether there is a temporal relationship between study duration and accuracy.

To ensure accuracy during the ISB study, we inspected all task annotations by hand. We encountered some situations where the task information (either the description or the categorization) did not appear to match the URLs recorded. In cases where the behavior was habitual and obvious, we modified the task

information ourselves. In all other cases, we contacted the participants to clarify the task information. Through the postsession questionnaires, participants seemed to agree that their annotations were accurate. It is interesting to note that participants felt that they were more accurate providing task categorizations than task descriptions. We hypothesize that this may be partly because of task drift, whereby the type of task remains constant (e.g., information gathering) but the description of the task may change over time. Also, unlike the task descriptions, the task categorization did not require participants to articulate succinct descriptions of their behavior.

Duration. In comparison to previous research that has collected Web usage logs on the Web for extended periods of time (e.g., Cockburn & McKenzie, 2001; Tauscher & Greenberg, 1997), there is a limit to how long participants will be willing to provide contextual information. Depending on the type of information being collected, the participant overhead may be simply too heavy to allow sustained involvement. In both projects (IIP1/IIP2 and ISB), some participants expressed relief at the conclusion of the 1-week study, as they began to find it tiresome to annotate their Web usage on a daily basis.

There are instances of previous research that have successfully collected contextual information for extended periods of time. Participants in Kelly and Belkin's (2004) field study provided contextual information, such as task descriptions and measures of usefulness for viewed Web pages, on a weekly basis for 14 weeks. Therefore, participants may be willing to take part in a longer duration field study if the frequency of annotations is minimal (e.g., once a week vs. once a day). However, as the annotations become more fine grained, it becomes more important for the annotations to be provided in a timely matter (e.g., on a daily basis). Therefore, weekly or monthly annotations may be possible only with higher level contextual information.

Real-time versus post hoc annotation. We must also consider whether participants should provide their annotations in a real-time or post hoc basis. When the collection of participant annotations occurs in real time, the characteristics of the activity are fresh in the participant's mind. However, the normal flow of Web usage is interrupted. Alternatively, annotations collected at a later time are less intrusive; however, participants may not be able to accurately recall their activities. Researchers must carefully weigh the trade-offs involved in real-time annotations versus annotations collected at a later time. The decision of which method to use depends on factors such as the complexity of the data being collected and the required frequency of the data collection.

In the IIP1 and IIP2 studies, participants provided privacy ratings at the end of each day using the task diary. Privacy ratings may change from one page to the next, so it would not have been feasible to interrupt the flow for each and every page to assign privacy ratings. All participants assigned privacy ratings to all visited pages captured during the study. During the uninstall session, none of the participants indicated that they found it problematic to assign their privacy ratings at the end of the day. The electronic diary also allowed them to return to their

annotations at a later time if they were unable to complete their daily classification. In the IIP2 study, location information was provided by laptop users in real time through a browser pop-up window. We did not expect that participants would be able to accurately assign location information at the end of the day for all of their Web usage, especially if they accessed the Web from several locations. We were therefore willing to accept occasional interruption of flow for the benefit of more accurate location information. To minimize the disruption, the pop-up window appeared when a browser window was closing. No participants commented that this was bothersome.

In the ISB study, participants were given a choice as to which method they preferred to use: real-time annotations using the task toolbar or end-of-day annotations using the task diary. Most participants preferred to use the task toolbar to annotate their Web usage in real time, as it was difficult to provide task information after the fact. In contrast with the IIP1 and IIP2 studies where the annotations may change page by page, participants had to change their task information only when they began a new task, which made real-time annotations more viable.

4.2. Data Collection

As previously discussed, the choice of a data collection tool is strongly influenced by the type of data and the level of detail to be collected. Two different Web logging tools were used in our studies, and each choice of tool came with a set of benefits and drawbacks. During the ISB study, a custom logging Web browser that mimicked the look and functionality of IE was used. During the IIP1 and IIP2 studies, a logging BHO was launched simultaneously with IE. The custom Web browser was necessary to have full control over the user interactions logged; the BHO could log only more limited navigation events, such as Web page URLs and document events. Both the custom Web browser and BHO allow for the installation of custom toolbars, which is useful for real-time annotations.

One of the main reasons for selecting field studies as a methodology was to capture natural user behavior. The focus of our research included the tools participants used, the sites they visited, and their normal patterns of activity. In both projects participants were presented with a familiar browser and had access to all of their usual IE features, such as bookmarks, history, and the Google toolbar. Upon completion of all studies, it was important for us to reflect on the perceived naturalness of our participants' behavior through inspection of the data as well as participant responses. For instance, we observed only a small number of blinded URLs and several instances of (nonblinded) adult content during the IIP2 study. The proportion of participants in the field study with instances of adult content was comparable to frequency reports of erotica viewing in a previous survey (Hawkey & Inkpen, 2006b). This result may indicate that we have captured participants' normal Web usage, including those activities not considered to be socially desirable (Fisher, 1993). During the ISB study, the number of visited Web pages was comparatively lower (approximately one third) than the IIP1 and IIP2 studies. This leads us to believe we received only snapshots of participants' usage on the Web in the information-seeking study. This may be

partly because participants had to remember to launch the custom Web browser, whereas the BHO automatically loaded with each IE window opened. We also expect that having to provide details about the task, especially those of a personal nature, may have been a barrier.

It is important that when interfering with the participants' natural environment, even if only through the collection of contextual information, we attempt to obtain a sense of how this interference has impacted participants' behaviors. For instance, in the ISB study, the poststudy questionnaire asked participants to reflect on how the annotations and use of a custom browser impacted their Web browser usage. Responses revealed that our choice of methodology may have impacted participants' task switching behavior.

Over time it can be difficult to continually update and refine tools to work with new versions and features of commercial software. For instance, a custom Web browser can quickly become obsolete as new versions of IE and new Web browsers are introduced. We had many former IE users, who now use Firefox, volunteer to take part in the information-seeking study until they realized the custom browser did not support tabbed browsing (consistent with the version of IE at the time of the study). Although versioning is still somewhat problematic for the BHO approach, it could potentially be upgraded to function with a new version of IE.

Although the custom Web browser used during the ISB study was designed to mimic IE and contained most of the same features, it is difficult to match the robustness of a commercial Web browser. There were a small number of bugs (e.g., occasional problems loading pages containing JavaScript) and issues with the custom Web browser, such as speed (i.e., slow to launch the application and first launch of bookmarks), that were not encountered when IE was augmented with the BHO. Participants also seemed concerned about the security offered by the custom Web browser even though security is provided through the browser control object.

4.3. Data Transfer

One question that arises during field research is how to transfer the data from participant to researcher. Although logged data during a laboratory experiment is typically stored directly on a research computer, researchers conducting research in the field must make decisions regarding where to store the data, when to transfer the data, and how to transfer the data. There are trade-offs inherent to each approach. For instance, storing the data on the participant's machine for the duration of the study may simplify the participants' duties; however, researchers run the risk of data loss if the participant's machine crashes. If data are transferred more frequently, the participant may be inconvenienced.

In both projects, we chose to build a custom application in which participants could e-mail the researchers a daily data report. This process allowed us to review the data on a daily basis to ensure that participants were properly annotating their data and to quickly spot problems with the data collection tools. For instance, in the IIP1 study, we observed that the BHO was not formatting Chinese characters properly (in page titles), and we were able to quickly issue a fix for the problem. The absence of data can also indicate that participants are encountering

difficulties. During all studies (IIP1, IIP2, ISB), when we failed to see an e-mail report from a participant for a 2-day period, we contacted the participant to inquire if there were any problems.

Although there were several advantages to participants e-mailing their data on a daily basis, problems did arise for some participants. To successfully use the custom e-mail program, many participants had to temporarily disable their virus scanner. Less technically included participants often failed to do so. In particular, some participants during the ISB study became perplexed when their e-mail report failed to send and stopped the study temporarily. As computer scientists, we often forget that not all participants are comfortable with a high level of technology and it is important that participants are not asked to engage in activities that are technically demanding.

Backups were also kept on the participants' machines for those cases when there were problems with the e-mailed data transfer or study software. The backups were created each time the data were accessed by the software (e.g., when opening the electronic diary). If problems were encountered with the e-mailed data reports, the data were recovered from the participants' computers during the uninstall sessions. This backup system ensured that no data were lost over the course of the research projects.

4.4. Data Analysis

Although the collection of contextual information can be problematic in itself, it is also challenging to analyze the data (Hilbert & Redmiles, 2000). Techniques are required for synchronization of various data sources and for transforming the low-level captured events into meaningful instances of activity. For example, in the ISB study we combined two log files and transformed sequences of Web pages visits into data-rich task instances. Once transformed, techniques were needed for analysis of the data (e.g., summary statistics, pattern detection, and visualization).

One problem we encountered during the ISB study illustrates the importance of piloting data analysis processes, from data collection through to data transformation and analysis. During the study, we collected two log files. The first log file consisted of page title, URL, timestamps, window ID, task description, and task categorization. The second log file consisted of timestamps, window ID, and browser functions. This meant that to know how each Web page loaded, two corresponding log entries (one from each log file) needed to be merged into a single entry. However, the two timestamps were not synchronized, meaning that scripts could be not used to merge the two files and much of it had to be conducted by hand—a very time-consuming process. In the IIP1 and IIP2 studies, we assigned a unique timestamp to related events to avoid the problem of unmatched timestamps, which results from delayed file writes.

Logging events can result in extremely large data sets, which can be difficult to manipulate and analyze. The number of Web pages viewed in our study ranged from a low of 13,498 in the ISB study to 36,170 Web pages during the IIP1 study. Therefore, it is important to be cognizant about how the data transformation processes will be affected by very large data sets.

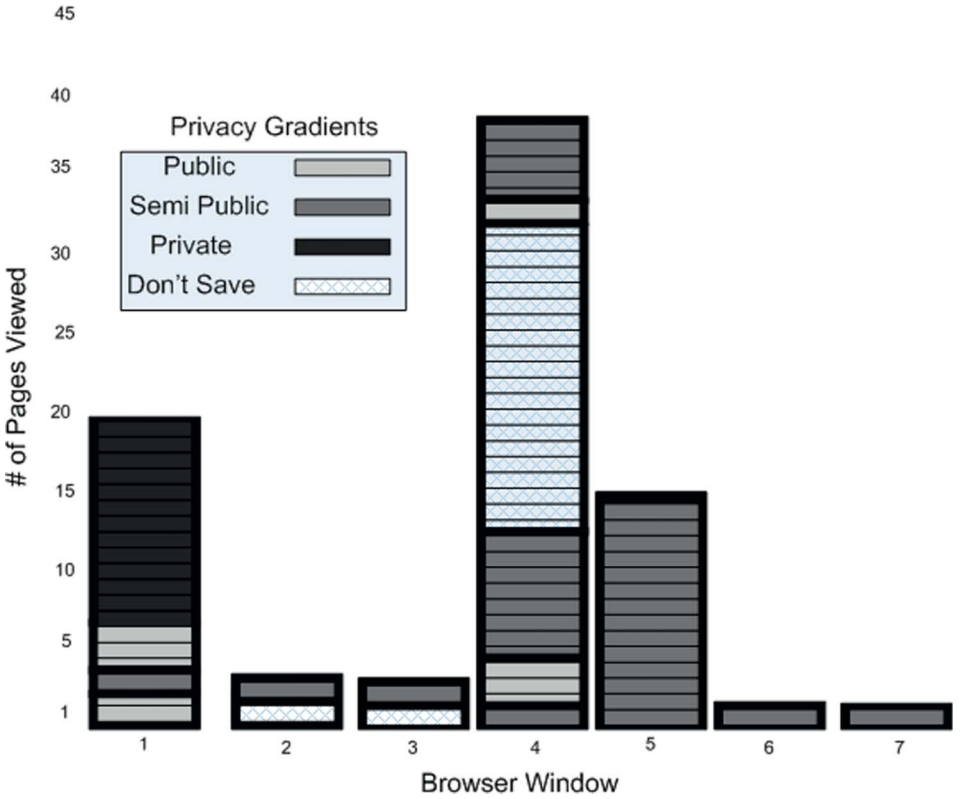


FIGURE 3 A visual representation of streaks of browsing on a per-browser-window basis (by 1 participant over the course of 1 hr).

Visualization tools can be effective for understanding user behavior, such as finding trends and patterns within the data. It can be difficult to pick out meaningful patterns from textual data. Figure 3 shows a visual representation of 1 hr of logged data that was handcrafted during data analysis for the IIP1 study. Visualizations such as this can help researchers gain a better sense of which behavioral patterns should be further investigated (e.g., streaks of browsing at a particular privacy level). Tools that can be customized to reflect the needs of different researchers are needed to view combinations of logged and contextual data.

5. SUMMARY

Studying user behavior on the Web is a difficult area of research. It can be challenging to capture realistic behaviors when users are not studied in their natural environment, engaging in intrinsically motivated everyday activities, and using their normal tools. However, an understanding of these realistic behaviors is required to appropriately ground development of new Web-based tools and techniques. It is therefore important that focused laboratory studies and attitudinal

surveys are augmented with field research. In our research, we have found field study methodologies to be effective at capturing a rich set of behavioral data. In particular, we found that contextual information provided in the form of participant annotations for privacy and task information, coupled with logs of Web usage, afforded valuable insight into our participants' behavior.

In this article, we have presented an overview of our experiences in collecting contextual information in a Web-based field environment over the course of two separate research projects in which we developed our own data collection tools. Based on these experiences, we have shared the implications of our methodological decisions, highlighted unanswered questions, and provided considerations for other researchers. We hope that by sharing our experiences and perspectives that others contemplating research in similar areas are able to build on the lessons that we have learned.

REFERENCES

Allen, C. (1989). The use of video in organizational studies. *ACM SIGCHI Bulletin*, *21*(2), 115–117.

Aula, A., Jhaveri, N., & Käki, M. (2005). Information search and re-access strategies of experienced Web users. *Proceedings of the 14th International Conference on World Wide Web (WWW 2005)*, 583–592.

Barford, P., Bestavros, A., Bradley, A., & Crovella, M. (1999). Changes in Web client access patterns: Characteristics and caching implications. *World Wide Web*, *2*(1–2), 15–28.

Brown, B., Sellen, A. J., & O'Hara, K. P. (2000). A diary study of information capture in working life. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'00)*, 438–445.

Bruce, H., Jones, W., & Dumais, S. (2004). Keeping and re-finding information on the Web: What do people do and what do they need? *Proceedings of the American Society for Information Science and Technology (ASIS&T 2004)*, 129–137.

Byrne, M. D., John, B. E., Wehrle, N. S., & Crow, D. (1999). The tangled Web we wove: A taskonomy of WWW use. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'99)*, 544–551.

Catledge, L. D., & Pitkow, J. E. (1995). Characterizing browsing strategies in the World-Wide Web. *Proceedings of the Third International World-Wide Web Conference*, 1065–1073.

Choo, C. W., Detlor, B., & Turnbull, D. (2000). Information seeking on the Web: An integrated model of browsing and searching. *First Monday*, *5*(2). Retrieved from http://firstmonday.org/issues/issues5_2/choo/index.html

Claypool, M., Le, P., Waseda, M., & Brown, D. (2001). Implicit interest indicators. *Proceedings of Proceedings of the 6th International Conference on Intelligent User Interfaces (IUI 2001)*, 33–40.

Cockburn, A., & McKenzie, B. (2001). What do Web users do? An empirical analysis of Web use. *International Journal of Human-Computer Studies*, *54*(6), 903–922.

Cothey, V. (2002). A longitudinal study of World Wide Web users' information-searching behavior. *Journal of the American Society for Information Science and Technology*, *53*(2), 67–78.

Edmonds, A., Hawkey, K., Kellar, M., & Turnbull, D. (2006). Logging traces of Web activity: The mechanics of data collection. *Workshop presented at the 15th International Conference on World Wide Web (WWW 2006)*.

Fisher, R. J. (1993). Social desirability bias and the validity of indirect questioning. *Journal of Consumer Research, 20*(2), 303–315.

Hawkey, K., & Inkpen, K. (2005). Privacy gradients: Exploring ways to manage incidental information during co-located collaboration. *Proceedings of the CHI'05 Extended Abstracts on Human Factors in Computing Systems*, 1431–1434.

Hawkey, K., & Inkpen, K. (2006a). Examining the content and privacy of Web browsing incidental information. *Proceedings of the 15th International Conference on World Wide Web (WWW 2006)*, 123–132.

Hawkey, K., & Inkpen, K. (2006b). Keeping up appearances: Understanding the dimensions of incidental information privacy. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'06)*, 821–830.

Herder, E., & Juvina, I. (2004). Discovery of individual user navigation styles. *Proceedings of the Workshop on Individual Differences in Adaptive Hypermedia (Adaptive Hypermedia 2004)* pp. 40–49.

Hilbert, D., & Redmiles, D. (2000). Extracting usability information from user interface Events. *ACM Computing Surveys, 32*(4), 384–421.

Hölscher, C., & Strube, G. (2000). Web search behavior of Internet experts and newbies. *Proceedings of the 10th International Conference on World Wide Web (WWW 2000)*, 337–346.

Hong, J. I., Heer, J., Waterson, S., & Landay, J. A. (2001). Webquilt: A proxy-based approach to remote Web usability testing. *ACM Transactions on Information Systems, 19*(3), 263–285.

Jansen, B. J., Spink, A., & Pedersen, J. (2005). A temporal comparison of Altavista Web searching. *Journal of the American Society for Information Science and Technology, 56*(6), 559–570.

Jansen, B. J., Spink, A., & Saracevic, T. (2000). Real life, real users, and real needs: A study and analysis of user queries on the Web. *Information Processing and Management, 36*(2), 207–227.

Kellar, M., Watters, C., & Shepherd, M. (2006a). A goal-based classification of Web information tasks. *Proceedings of the American Society for Information Science and Technology*.

Kellar, M., Watters, C., & Shepherd, M. (2006b). The impact of task on the usage of Web browser navigation mechanisms. *Proceedings of Graphics Interface 2006*, 235–242.

Kellar, M., Watters, C., & Shepherd, M. (2007). A field study characterizing Web-based information seeking tasks. *Journal of the American Society for Information Science and Technology, 58*(7), 999–1018.

Kelly, D., & Belkin, N. (2004). Display time as implicit feedback: Understanding task effects. *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2004)*, 377–384.

Kim, K.-S., & Allen, B. (2002). Cognitive and task influences on Web searching behavior. *Journal of the American Society for Information Science and Technology, 53*(2), 109–119.

Kort, J., & de Poot, H. (2005). Usage analysis: Combining logging and qualitative methods. Proceedings of the CHI'05 Extended Abstracts on Human Factors in Computing Systems, 2121–2122.

Lederer, S., Mankoff, J., & Dey, A. K. (2003). Towards a deconstruction of the privacy space. *Workshop on Ubicomp Communities: Privacy as Boundary Negotiation, UBICOMP 2003.* Retrieved from http://guir.berkeley.edu/pubs/ubicomp2003/privacyworkshop/papers/lederer-privacyspace.pdf

Liu, C., Marchewka, J. T., Lu, J., & Yu, C.-S. (2004). Beyond concern: A privacy-trust-behavioral intention model of electronic commerce. *Information & Management*, *42*(1), 127–142.

Loeber, S. C., & Cristea, A. (2003). A WWW information-seeking process model. *Educational Technology & Society, 6*(3), 43–52.

Margulis, S. T. (2003). Privacy as a social issue and behavioral concept. *Journal of Social Issues, 59*(2), 243–261.

Marx, G. T. (2003). A tack in the shoe: Neutralizing and resisting the new surveillance. *Journal of Social Issues, 59*(2), 369–390.

Morrison, J. B., Pirolli, P., & Card, S. K. (2001). A taxonomic analysis of what World Wide Web activities significantly impact people's decisions and actions. *Proceedings of the CHI'01 Extended Abstracts on Human Factors in Computing Systems*, 163–164.

Novak, T. P., Hoffman, D. L., & Yung, Y. F. (2000). Measuring the customer experience in online environments: A structural modeling approach. *Marketing Science*, *19*(1), 22–44.

Obendorf, H., Weinreich, H., & Hass, T. (2004). Automatic support for Web user studies with scone and tea. *Proceedings of the CHI'04 Extended Abstracts on Human Factors in Computing Systems*, 1135–1138.

Olson, J. S., Grudin, J., & Horvitz, E. (2005). A study of preferences for sharing and privacy. *Proceedings of the CHI'05 Extended Abstracts on Human Factors in Computing Systems*, 1985–1988.

Reeder, R., Pirolli, P., & Card, S. (2001). Webeyemapper and Weblogger: Tools for analyzing eye tracking data collected in Web-use studies. *Proceedings of CHI'00 Extended Abstracts on Human Factors in Computing Systems*, 19–20.

Rieh, S. Y. (2003). Investigating Web searching behavior in home environments. *Proceedings of the American Society for Information Science and Technology (ASIS&T 2003),* 255–264.

Sellen, A. J., Murphy, R., & Shaw, K. L. (2002). How knowledge workers use the Web. *Proceedings of Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'02)*, 227–234.

Spink, A., Wolfram, D., Jansen, M. B., & Saracevic, T. (2001). Searching the Web: The public and their queries. *Journal of the American Society for Information Science and Technology, 52*(3), 226–234.

Tauscher, L., & Greenberg, S. (1997). How people revisit Web pages: Empirical findings and implications for the design of history systems. *International Journal of Human-Computer Studies, 47*(1), 97–137.

Teevan, J., Alvarado, C., Ackerman, M. S., & Karger, D. R. (2004). The perfect search engine is not enough: A study of orienteering behavior in directed search. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'04)*, 415–422.

Thury, E. M. (1998). Analysis of student Web browsing behavior: Implications for designing and evaluating Web sites. *Proceedings of 1998 Conference on Computer Documentation*, 265–270.

Turnbull, D. (1998). *Webtracker: A tool for understanding Web use*. Retrieved from http://www.ischool.utexas.edu/~donturn/research/webtracker/ (Unpublished report)

Zhang, D., Zambrowicz, C., Zhou, H., & Roderer, N. (2004). User information seeking behavior in a medical Web portal environment: A preliminary study. *Journal of the American Society for Information Science and Technology, 55*(8), 670–684.