Recommendations for reporting context in studies of web browsing behaviour

Kirstie Hawkey
Melanie Kellar

Technical Report CS-2004-16

November 8, 2004

Faculty of Computer Science
6050 University Ave., Halifax, Nova Scotia, B3H 1W5, Canada

Recommendations for reporting context in studies of web browsing behaviour

Kirstie Hawkey and Melanie Kellar
Faculty of Computer Science
Dalhousie University
Halifax, Nova Scotia, Canada
{hawkey, melanie} @ cs.dal.ca

Abstract

The study of web browsing behaviour is a relatively young and developing researching domain. Research in this area is complicated by the fact that the Web and its user population are constantly growing and evolving. This can make it difficult for researchers to evaluate previous work in the proper context, particularly when detailed information about the user population, experimental methodology, and results is not presented. Therefore, we present eight categories of contextual information that, when not reported, make it difficult to replicate the work, make comparisons between current work and the reported work, and build upon this reported work to advance the field. We have conducted a survey of recently published papers that studied user behaviour on the Web and examined the contextual information currently reported.

Keywords

Web browsing behaviour, methodology, standardization, user study, survey.

1 Introduction

Over the past ten years there has been a wide variety of research conducted looking at an individual's behaviour on the Web, beginning most notably with Catledge and Pitkow's (1995) study of navigation strategies. This field of research has expanded to evaluate a variety of user behaviours on the Web such as information seeking behaviour, navigational behaviour, and general characteristics of web usage. Aspects of web browsing behaviour motivate the design of tools and interfaces for web applications. In this paper, we use the term "web browsing behaviour" to include any interactions between a user and a web browser, ranging from a user's use of the back button, to website revistation patterns, information seeking behaviour, and general web usage. We want to make the distinction between our use of the term "web browsing behaviour" and the serendipitous task of "browsing the Web". There are many applications of traces of web browsing behaviour, particularly for the validation and comparison of models, algorithms or techniques.

Although much research has been conducted since the early studies of Catledge and Pitkow (1995) and Tauscher and Greenberg (1997), researchers are still using the results of this early research as statements of fact and as a basis for conducting research that builds upon it without challenging its current validity. However, the web environment has changed along multiple dimensions since these early studies: it has become much larger, access speeds are much higher, and users have a wider variety of web browsers and search tools at their disposal. The user population has also changed considerably as web usage has become commonplace at work and at home; there is now a much greater variability in the demographics and reasons for web use.

It is important that researchers and practitioners currently developing algorithms, techniques, and applications based upon the behaviours of web users have an understanding of the current state of the browsing behaviour, including the amount of variability and trends over time. However, comparing

previous research studies is complex because the studies need interpretation within the context of the web environment at the time of the study, the population studied, the nature of the tasks performed, the browser or tools used, and the metrics recorded. Additionally, insufficient reporting of the study methodology and results in the literature can make this a challenging task. This inability to easily determine the design and procedure of studies and judge the validity of results makes it difficult to replicate and build upon previous work.

The study of web browsing behaviour is a research domain that is still growing and evolving (as the Web grows and evolves). It is important and worthwhile for the community to engage in discussions about 'good practice' to ensure that research contributions are placed correctly within the overall body of work. In this paper, we propose guidelines for reporting the contextual details of web browsing behavioural studies to enable more effective sharing of research results. We conducted a survey of recent research to demonstrate the nature and scope of the problem.

In the background, we present the seminal papers in this field, discuss methods of observing users' web browsing behaviour, and present previous work with standardized reporting. We then present a timeline to illustrate the changing nature of both the user population and the browsing environment that may impact web browsing behaviour. We propose a set of minimal guidelines for reporting the contextual information necessary for researchers and practitioners to be able to appropriately evaluate the applicability of the results. We follow these guidelines with a survey of recently published studies that demonstrates the variability in reporting of methodology, analysis and results. We conclude with a discussion of the impact of this variability in reporting on the ability to establish external and temporal validity.

2 Background

A variety of web browsing behaviours, have been studied since the mid 90's, starting most notably with the seminal works presented below. These works have studied web browsing behaviours using a variety of methodological approaches. However, although standard guidelines for the reporting of results are used in other research domains, none exist for reporting of the methodological details and results for studies of web browsing behaviour.

2.1.1 Seminal Works

One of the first studies examining user behaviour on the Web was conducted by Catledge and Pitkow (1995) for three weeks in 1994. Participants' behaviour was logged while they browsed the Web using a modified version of XMosaic that collected browsing activity. Two dominant methods of navigation were revealed by the participants: hyperlinks and the back button. Navigation strategies were categorized according to frequency.

Trace logs of web usage from March through May of 1995 were used, along with topology and textual similarity between nodes, by Pirolli, Pitkow and Rao (1996) to extract structures of websites. This work was one of the early applications of web usage logs.

Similar to Catledge & Pitkow (1995), Tauscher and Greenberg (1997) observed user behaviour with a modified version of XMosaic, in order to study revisitation patterns of users. Over a six week period in 1995, they observed that 58% of page visits were revisits and that the back button was used in 30% of navigations.

Byrne et al. (1999) conducted a task analysis of user web behaviour through a 1998 study. Participants were video taped in their offices, for a day, as they used the Web. Participants spent the majority of the time on the Web reading. The most common navigation method was hyperlinks, followed by the back button.

Choo et al. (2000) investigated information seeking behaviour on the Web in a two week study conducted circa 1998. Participants' web behaviour in the workplace was logged client-side during the

course of the study. Through the analysis of clickstream data, interviews, and questionnaires, four modes of information seeking behaviour were defined: formal search, informal search, conditioned viewing, and undirected viewing.

Cockburn and McKenzie (2001) conducted a four month retrospective observational study, from October 1999 to January 2000, of history and bookmark files retrieved from server backups. The authors found an average revisitation rate of 81%. Analysis of the bookmark files found that participants were either heavy or light users of bookmarks.

Sellen et al. (2002) studied the activities and characteristics of knowledge workers on the Web. Participants were interviewed in front of their history lists and described the web activities they had recently completed. Knowledge workers engaged in six types of activity on the Web: finding, browsing, information gathering, transacting, communicating, and housekeeping.

2.2 Methods of Observing User Behaviour

A variety of methods have been used for learning about a user's behaviour on the Web. One of the most common approaches is the use of web logging, which can be client-side, server-side, or through a proxy. Other approaches include direct researcher observations, diary studies, interviews and questionnaires. Each technique has its own set of advantages and disadvantages and its use is often dictated by the goal and setting of the research. Whichever method is selected, there are implications for the interpretation of results.

2.2.1 Client-Side Logging

Several research domains have used client-side logging to examine user behaviour on the Web. These include information seeking behaviour on the Web (Choo et al., 2000), usability evaluation (Hilbert and Redmiles, 2000), and the evaluation of implicit indicators of interest (Claypool et al., 2001). Approaches for client-side logging include commercial "spy-ware" tools (Kelly and Belkin, 2004; Kim and Allen, 2002), custom logging tools (Obendor et al., 2004; Reeder et al., 2001; Turnbull, 1998), and custom browsers (Claypool et al., 2001). Client-side logging offers the richest exploration of user behaviour. However, many client-side logging tools are designed to work with a specific browser and may be consuming and costly to update as new versions of the browser are introduced.

2.2.2 Server-Side Logging

Server-side logs do not capture the same level of detail as is possible with client-side logging; however, benefits include a reduction in cost and time of implementation. Server-side logging has proved useful in the study of search engine use (Anick, 2003; Spink et al., 2001), information seeking (Zhang et al., 2004), and general web behaviour (Huberman et al., 1998; Pitkow, 1997). This method is ideal for research with large populations, remote users, or for field studies.

The data recorded by server logs includes the IP address of users and the time and address of web page requests. The use of dynamic IP addresses makes it difficult to distinguish between distinct users; however, cookies can alleviate this problem (Anick, 2003). Caching is an issue as pages that are loaded from the web browser cache do not reach the server (Fenstermacher and Ginsburg, 2003).

2.2.3 Proxy Logging

Proxy-logging is a compromise between client-side logging and server-side logging and provides some of the advantages of each. By allowing participants to log into the system instead of downloading and installing software, proxy solutions such as WebQuilt (Hong et al., 2001) allow participants to work within their normal browsing environment. However, proxy servers do 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 et al., 1998). Proxy-side 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 data generated by the proxy-based logger was found to be neither reliable nor accurate.

2.2.4 Other Approaches

Several other approaches are used to capture user behaviour on the Web. Direct researcher observations have been used in laboratory settings (Card et al., 2001; Holscher and Strube, 2000) as well as in the field where researchers can observe participants in their natural setting (Teevan et al., 2004; Thury, 1998). The use of video cameras (Byrne et al., 1999) or video capture software (Hargittai, 2002; Jenkins et al., 2003) provides a record of a user's behaviour and its context but can be extremely time consuming to code. Furthermore, the user's motivation and thoughts may still be unclear. This technique is only feasible for small groups of participants and for studies that are qualitative in nature. Diary Studies (Rieh, 2003), surveys (Heinström, 2003; Schiano et al., 2001) and interviews (Jones et al., 2002) are other methods of self-reporting used to obtain a qualitative view of user behaviour on the Web.

2.2.5 Standardized Reporting

Standardization of reporting is an approach that has worked in other disciplines to ease meta-analyses. The Controlled Standards of Reporting Trials (CONSORT) (www.consort-statement.org) helps readers of randomized controlled trials understand the design and running of the study and the analysis and interpretation of the results. The Common Industry Format (CIF) (Laskowski et al., 2001) is an ANSI approved standard for reporting the results of usability studies. This standard was developed to aid organizations in making decisions based on usability when choosing new software.

There has been some effort at creating standards for the reporting of on-line search behaviours. Jansen and Pooch proposed a framework for reporting Web searching studies to facilitate comparison of results. Their framework includes detailed descriptive information about the searchers, the information retrieval system (including the searching rules at the time of the study), the methods of data collection, and the transaction log. When presenting analysis, the authors recommend that analysis is attempted at the level of session, query, and term (these terms are defined). The authors also recommend that statistical analysis be reported and that data be reported at low levels of detail as well as when aggregated to improve the ability of other researchers to compare their results. Wildemuth et al. (2004) are currently conducting a Delphi study to investigate standard variables to collect for the study of online search behaviours. This research is still underway but after the second iteration, three main categories of variables have emerged: the search process, the search system, and the user.

The study of web browsing behaviour takes place in a rapidly changing and dynamic environment. Coupled with the fact that this is still a relatively young area of research, it is often difficult to conduct comparisons between studies or replicate previous work. Although many of the recommendations for reporting contextual information presented in this paper may appear to be common sense, they are routinely omitted from publications. We expect that the set of reporting guidelines presented in Contextual Information section can serve as a useful tool for both new and veteran researchers in the field as they endeavour to share their research findings effectively.

3 Motivation

The World Wide Web is relatively young and is continually evolving. Technological innovations have changed the way that people access the Web: as the state of hardware (Unix boxes, desktop PCs, laptop computers, handheld, cellular phones) and software (web browsers and search engines) progresses, the experience of the end user changes. With penetration of the market expanding from

technologists to general home and business users, the population has become more diverse. Internet access is no longer restricted to those with a high income and level of education. As a result, Internet usage patterns are continually changing and their study must be continual.

As discussed in the background, gathering data about a user's behaviour during web browsing is a difficult task and there are tradeoffs inherent with the methods selected. Depending on the method of data collection, there may be different metrics captured that must then be interpreted to gather an understanding of the user's actual behaviour. The task and goals of a user may effect their behaviour while web browsing. There are also individual differences that may have an impact.

It is therefore imperative that when reporting the results of a study, that the study design, methods, etc. be sufficiently described so that the reader can see the impact on the results of decisions made. The inclusion of these important details also allows researchers to compare and replicate previous work. Replication and extension studies have the potential to provide insight into how the ongoing growth of the Web is affecting its users.

3.1 Timeline

We present snapshots of the state of the World Wide Web and its users at the time some of the research in the seminal papers was conducted. It must be noted that the figures reported have been selected from a variety of sources with varying methodologies, populations, and metrics. Therefore, direct comparisons are not always appropriate. These snapshots have been provided to illustrate the changing nature of user behaviour on the Web that gives the context for the seminal papers in the area. We have indicated the dates of the studies appearing in the seminal papers. If a study date was not available, we note the likely date based upon the submission deadlines for the publication.

3.1.1 In the Beginning

- Catledge and Pitkow (1995): August 1994
- Pirolli, Pitkow, and Rao (1996): March-May 1995
- Tauscher and Greenberg (1997): October-December 1995

Fall/94: The typical user is a 31 year-old educated male who works with computers and has authored about 30 web documents (Pitkow and Recker, 1994). He uses a Mosaic browser 1-4 times a day for about 5 hours per week; however web browser choices are growing with the release of Netscape (Pitkow and Recker, 1994). He uses the Web to browse, for entertainment, for work or business, and for research (Pitkow and Recker, 1994) and has a choice of about 10,000 websites (Marsh, 2003)

Fall/95 (Pitkow and Kehoe, 1996): Worldwide web traffic has surpassed ftp data and search engines are now available (Marsh, 2003). Users are shifting towards "early adopters/seekers of technology" instead of the "technology developers/pioneers" of a year before (Pitkow and Kehoe, 1996) with the start of commercial internet providers such as Compuserve, AOL and Prodigy (Infoplease, 2004b). Women now account for about 30% of web users and there has been some increase in the number of younger and older users (Pitkow and Kehoe, 1996). Most users have 14.4 or 28.8 kbs modems (Pitkow and Kehoe, 1996).

3.1.2 Home Users and Browser Wars

- Byrne, John, Wehrle, and Crow (1999): circa 1998
- Choo, Detlor, and Turnbull (2000): circa 1998

Fall/98: Women now account for almost 40% of web users (GVUOnlineSurvey, 1997). About a third of users have a 56K modem (Kehoe et al., 1999) and 84% are interested in high speed internet access (Pastore, 1998b). Microsoft IE wins the browser wars, just surpassing last year's dominant browser, Netscape Navigator, to capture 50% of the market (Pastore, 1998a). More than 40% of the people between the ages of 9 and 49 are on-line (Infoplease, 2004c); the average age is 38 (Kehoe et al., 1999). Almost a third of users shop on-line (Pastore, 2000a). Google arrives and 10,000 searches are performed per day (Google, 2004).

3.1.3 Work and Home: The Need for Speed

- Cockburn and McKenzie (2001): Oct. 1999-Jan. 2000

Fall/99: The year 2000 is looming and the 150 million web users (Infoplease, 2004b) worldwide are looking for information about Y2K as the Lycos 50 listing of the top web searches debuts (although Pokemon and the Blair Witch Project top the list) (Lycos, 1999). Google performs 3 million searches per day (Google, 2004). Napster allows swapping of music and 'E-Commerce' is the new buzz word (Infoplease, 2004b). The 6% of users with high speed internet access view 130% more pages and surf the Web 83% more often than the 45% of users that still have a 28.8/33.6 K modem (Pastore, 2000c). According to Nielsen//NetRatings the average web user had 17 29-minute sessions each month, viewing an average of 32 pages per session (Pastore, 2000b).

3.1.4 In the Mainstream: Just Google it

- Sellen, Murphy and Shaw (2002): circa 2001

Fall/01: Google has become a verb: with over 3 billion web documents (Google, 2004) available to be searched and the Google toolbar to help them do it, users over the world are telling each other to Google it. Napster has lost its court case (Infoplease, 2004b) but other file sharing applications are quick to fill the void. The structure of the population on-line is much closer to that from census data than in previous years (Pastore, 2001). There is an equal split of male/female users, but household incomes for web users are still higher than for the general population ($49,800 vs. $40,800) and the web user population is still younger (75% of adults 18-49 are on-line vs. 63% of the population, 24% of adults 50+ are on-line vs. 37% of population) (Pastore, 2001). Our average web user now has 33 33-minutes sessions per month, viewing an average of 36 pages per session (ClickZStatsStaff, 2002). 72% of the population are now using the Internet (58% at home, 73% at school, 51% at work) (Kerner, 2004).

3.1.5 A Daily Tool:

Fall/04: The Internet has become a daily tool: 56% of those with access to the Internet go on-line daily, 48% send email, 27% get news, and 19% do research for a job (Infoplease, 2004a). Google has added Gmail and Desktop Search (Google, 2004) and the division between on-line and off-line blurs. Our average web user now has 31 web sessions per month at home during the almost 26 hours of home PC use (Nielsen//NetRatings, 2004a) and 65 sessions at work during the 76 hours of work PC use (Nielsen//NetRatings, 2004b).

4 Contextual Information

Contextual information consists of the set of information that provides context about a particular piece of research relating to user behaviour on the Web. We present eight categories of contextual information crucial to the reporting of results: population, dates, environment, nature of the browsing

task, measures, descriptive reporting, statistical analysis, and context of results. Without these details, researchers are unable to replicate the reported work, make comparisons between their current work and the reported work, and build upon this reported work to advance the field. We present each of the categories in turn and discuss how this information (or lack thereof) can impact interpretation of reported results. It is important to acknowledge that the space constraints of a given publication may not allow for our recommended level of detail. However, extra information can be provided in the form of a technical report or experimental design document made available on a publicly accessible website.

4.1 Population

When conducting research examining the behaviour of individual web users, details about the sample population provide important information about the significance of the results. These measures help to determine the external validity of a study and give insight into how well the results can generalize to other populations. The types of population details include: size and gender of the sample population, ages, background, occupation, and web experience.

The impact of individual differences on web browsing behaviour is a growing research area. For example, a recent study by Herder and Juvina (2004) collected extensive data on cognitive abilities (spatial, episodic memory, working memory), internet expertise, and affective disposition of participants and correlated this data with self-reports of satisfaction and perception of lost-ness during web-assisted personal finance tasks. They identified two navigation styles (flimsy and laborious) that predict the perceived disorientation of web users.

However, individual differences have not received a great deal of attention in previous research. Even in the cases where individual user behaviour is distinguishable from one another, it is typically aggregated in order to develop a general user model for general purposes (Grace-Martin and Gay, 2001). Issues arise when individuals' web behaviour exhibits large variability, as in (Cockburn and McKenzie, 2001). In this study participants were recruited from within the academic community, but one person was employed as a webmaster and had a much higher level of web usage. Web experience, age, occupation, and background play a role in a user's behaviour and can contribute to large differences between users. In order to facilitate interpretation of results in light of advances in understanding on the role of individual differences of web browsing behaviour, it is important that as much detail as possible is reported.

Details about the population may not be available to the researchers depending on the type of data collected. For instance, web usage studies that use server data typically cannot report much information about participants. However, in this case it is important to report that this information is not available, so the reader understands why the information was omitted.

4.1.1 Size and Gender

Due to the high degree of variability between individual web users, sample size is important. Large sample sizes can help lessen the effect of individual differences and provide enough data points to aid in the identification of outliers in the population.

Although it is not always possible or appropriate to balance a sample population by gender, it is still important to report the ratio of males to females to help place the results in context. A primarily male population was appropriate with older studies conducted when most web users were male; however, now that the population of web users more closely matches the general population, a sample balanced by gender is preferable. The exception is research within a specific domain (for instance, nursing) that may have an inherently uneven gender split.

4.1.2 Age

As the Web has evolved, so have the key characteristics of users of the Web. As demonstrated in the timeline, the 'typical' user of the Web has shifted from young computer professionals to a varied assortment of users, ranging from young children to grandparents. Although it may be convenient to conduct our research on 18-22 year old university students, this demographic is only a subset of the average web users. Sample populations must begin to reflect the actual age range of the general web user population to maintain external validity.

4.1.3 Background

Although the academic community was at one time representative of the average web user, this is no longer the case. However, the tendency is to use (as in many other fields) only academic sample populations. Previous research (e.g. (Holscher and Strube, 2000; Ihadjadene et al., 2003)) has found that domain expertise does have an effect on a users' web behaviour. Therefore, awareness of the background and occupation of the sample population is important. If this information is unknown, it can be difficult to interpret unexpected or surprising results.

It is important that researchers attempt to target populations other than the academic community. Small focused studies are also necessary to examine behavioural differences with respect to web browsing between populations of different backgrounds. Identification of explicit differences could assist other researchers in understanding the limitations resulting from selecting a homogenous sample. It may be the case that background of users or other differences do not impact web behaviour for a class of tasks, but without empirical validation this cannot be assumed.

4.1.4 Experience

In addition to domain experience, web experience also affects a user's behaviour on the Web. Assessing web experience is difficult as it is constantly evolving. In 1995, a year of web experience may have indicated an individual had a high level of web experience. Today, this is not the case. Furthermore, length of time of web usage does not always imply 'expert' web usage.

Web experience is not an easily quantifiable measure but reporting detailed information about web experience allows other researchers to make judgments. In addition to web experience, which often is general, researchers can define experience in terms of the task or device. For instance, if investigating web behaviour on a mobile device, it may be important to collect and report information regarding the participants' experience with a mobile device and with web browsing on a mobile device.

4.2 Dates

Studies examining user behaviour on the Web must be interpreted in the context of the state of the Web at the time of the study. Providing the date that a study was conducted is essential to allow future researchers to place the results of studies in context. Furthermore, the duration of the study should also be recorded. There are temporal patterns in web activity associated with work days, weekends, holidays, and leisure time. Knowing the duration of the activity logged aids in identifying what types of web usage and patterns may have been captured.

4.3 Environment

Details regarding a study's environment allow others to replicate previous work and properly interpret study results. This includes information such as the study's setting and the tools used. The setting of a study (lab/home/work/school) may influence a user's behaviour and provides information regarding the potential completeness of the user's data. For instance, a lab setting is a controlled

environment in which a participant may not be acting as they normally would. In a field situation, if a user is studied while only at work, then much of their personal web usage/behaviour may not be captured. These details are often omitted, but give great insight into the realism of the study and the type of browsing tasks in which users might be engaged.

One of the biggest challenges in studying web browsing behaviour is employing software that records the appropriate information, while not impeding the user's natural behaviour. Often, web browsers used for research purposes are augmented with new tools or the ability to track a user's behaviour. However, if the browser used in a study differs from a participant's usual browser, this may influence the user experience.

4.4 Web Browsing Task

It is crucial for researchers to provide a rich description of the tasks performed by participants. This helps give an understanding about the types of behaviours that might have occurred during the experiment and evaluate the realism of the task. It is also important to include whether the tasks were personally motivated or motivated by the experiment as user behaviour can be affected by motivation. Loeber and Cristea (2003) describe the effect of motivation on the extensiveness of searches for information and include motivation in their model of navigation on the Internet.

If specific web sites are used during the task, details such as the navigation system available within the site can lend insight into the performance of participants. Ahuja and Webster (2001) found that participants perceived disorientation in a web search task was correlated to whether the website had a simple or global navigation scheme. Danielson (2002) found that in addition to the confidence level of users correlating with the availability of a site overview, their patterns of behaviour also changed. Participants abandoned fewer information-seeking tasks, went deeper into the site structure and made large navigation leaps in the site hierarchy, and made less use of the web browser's back button.

4.5 Measures

The methods of capturing user data provide context for the interpretation of results. We use the types of measures defined by McGrath (1994) to frame our discussion. With self-reports, participants knowingly report on their own behaviour. Examples include questionnaires and semi-structured interviews. Observations consist of records of behaviour collected by a researcher (or software). Observations can be either overt or covert. Examples include client-side logging software and researcher field notes. Trace measures are records of behaviour inadvertently left by participants. Examples include data collected from web server logs. Archival Records are records of user behaviour collected not for the intended purpose of research. The records may be either public or private knowledge. Examples include blogs or web browser bookmarks.

There are inherent biases and limitations possible with each type of measure so it is important to provide information about the type of measures chosen. Explicitly discussing the biases and limitations, when describing the measures and interpreting the results, is encouraged. While these are overall categories of measures, a description of the study instruments selected or designed, and discussions of their validity should be included.

It is also important to define the metrics collected with the study instruments. Depending on the type of data collection, there may be subtle changes in the interpretation of the metric. For example, when counting the number of pages viewed, a different count may be recorded depending on if the data was collected at the client, proxy, or server due to caching issues. Frames within a web page are often handled differently by logging software and change the nature of what is counted. An explicit definition of each metric can greatly contribute to the ability to compare results between studies. The W3C has developed recommendations for web characterization terminology (W3C, 1999) but they are somewhat abstract and metrics defined at a finer-grained level would remove any ambiguity.

4.6 Descriptive Reporting

There are several methods of describing the data included in the results. Raw data allows other researchers to get a good sense of the data, view outliers and variability, ranges of "normal behaviour", and generally gain a better understanding of what was measured. In most studies of user behaviour on the Web, there may be raw data that was captured but not pertinent to the study at hand. However, this data may help to characterize general browsing behaviour (pages visited, page views, time, actions); and, if made available, could be valuable to other researchers. Additionally, this would also allow the community as a whole to monitor how behaviour is changing over time.

Aggregate data provides an overall picture of a sample population's behaviour. However, a danger exists in reporting aggregate data because there may be individual participants who skew the data. For this reason, variability measures, such as standard deviations and quartiles are highly informative. For measurements of frequency, it may be appropriate to normalize the data at the participant level before determining mean frequencies to minimize the impact of outliers. Participant quotes, obtained through talk aloud protocols, interviews, and general dialog provide valuable insight. Anecdotal stories can also provide this same insight.

4.7 Statistical Analysis

Presenting quantitative results without statistical analysis reduces the validity of the reported results. Without descriptions of the statistics used and the results found, it is impossible for other researchers to judge the suitability of the analysis. In the case of data analysis that is not straight-forward due to characteristics of the data, providing additional details about the selection of the statistical tests and the assumptions behind the tests will aid other researchers in understanding the analysis.

4.8 Results in Context

Finally, it is important to place the results found in the context of previous works. Where possible, compare and contrast results with previous studies. This may be difficult if previous researchers have not provided full details of the studies. Reflecting on current results in the context previous research, especially within research domains that are still evolving, is necessary in order to advance the field. It also helps other researchers understand the contribution of the work to the field.

5 Survey of Previous Web Behaviour Studies

In this section, we present general categorizations of the studies that we surveyed. Studies were examined according to purpose of research, experimental strategy and location of publication in order to get a general sense of their breadth. We first present the methodology of our survey followed by the location of the publications, the purpose of each study, and the experimental strategies of the studies.

5.1 Methodology

We collected 44 publications, published in either 2003 or early 2004, dealing with individual user behaviour on the Web. Of these 44 papers, 17 were solely applications of web usage data (e.g. validating a model) and did not present actual studies of user behaviour. These application papers (see Appendix 2) were not included in the survey, but are discussed within this section. The studies in the remaining 27 papers (see Appendix 1) were included in our survey. One study (S31) did collect user browsing behaviour to train a recommender system; however, the authors performed a study to collect the annotated user sessions and described the user behaviours observed, thus we included this as a survey paper rather than an application paper.

Location of Publications	Studies Surveyed	Application Papers
Journal		
Interacting with Computers	S24	
International Journal of Human-Computer Studies	S2*, S4*, S10, S21, S25	
Journal of the American Society for Information Science and Technology (JASIST)	S6, S22, S30	
IT & Society	[S2,S3,S4,S5], S12	
Marketing Science		A1, A10, A14
Conference		
ACM SIGIR	S13	
ACM Human Factors in Computing Systems (CHI)	S1, S3*, S26	A2
ACM Intelligent User Interfaces (IUI)		A4, A11, A12
Annual Meeting of the American Society for Information Science and Technology (ASIST)	S7, S9, S23	
World Wide Web (WWW)	S14, S16, [S17,S18], [S19,S20], S27, S29, S31	A3, A6, A13, A15, A16, A17
IEEE Web Intelligence (WI)	S15, S28	A5, A7, A8,
Australian User Interface (AUIC)	S3, S12	
Adaptive Hypermedia and Adaptive Web-based Systems (AH)	S8	
OZCHI	S5*	
Other		
Technical Reports		A9

Table 1. **Surveyed studies (indicated by 'S') and Application papers (indicated by "A") classified by publication. Studies that appear in square brackets represent multiple studies reported in a single paper. Supporting papers (not from 2003/2004) are indicated with a *.**

The 27 papers in our survey presented 31 distinct studies. We treated each distinct study as a single item in our survey, even if two or more appeared in the same publication. Three publications contained multiple, distinct studies. One study (S3) appeared in both a journal and conference proceedings in 2003. One journal article contained 4 studies (S2,S3,S4,S5) and referenced previous papers (not from the 2003/2004 publication years) for additional detailed information about the study. We retrieved these papers in order to gain as much information about each study as possible and have included them in Appendix 1.

This survey by no means attempts to be a comprehensive overview of all studies appearing in 2003/2004 publications related to user behaviour on the Web. Instead, we have selected a limited number of publications in order to conduct a thorough analysis of each study. In selecting the surveyed papers, we first included those papers that we were familiar with from previous literature reviews in this area. We then searched the publications in which those papers appeared and other major HCI,

Information Science, and WWW journals and conferences that publish research about user behaviour on the web. We included all relevant papers published in 2003 and early 2004 .

5.2 Location of Publication

We first categorize the studies and applications of data according to their location of publication in order to demonstrate the range of papers surveyed (see Table 1).

5.3 Purpose of Research

We identified five main categorizes of research that investigated or made use of individual user behaviour on the Web: information seeking; web navigation; general web use; evaluation of tools, techniques and interfaces; and applications of browsing behaviour data. Two studies (S9, S31) belonged to more than one category. Table 2 displays this categorization. The largest proportion of studies dealt with the evaluation of tools, techniques and interfaces of web browsers (13/31 – 41.9%), followed by information seeking behaviour on the Web (12/31 – 38.7%). Studies also examined web navigation behaviour (4/31 – 12.9%) and general web use (3/31 – 9.7%). As previously noted, one study (1/31 – 3.2%) with a purpose of creating data for a recommender system was included in the survey; the 17 studies that were found to be purely an application of the user data are discussed in the next section.

5.4 Applications of Data

During our survey, 17 of the papers we encountered described applications of traces web browsing behaviour but did not study the behaviour itself. We present them briefly in this section. We found three main applications of web usage data: to validate a model (3/17 – 17.6%), to evaluate or compare algorithm(s) (7/17 – 41.2%), and to evaluate an architecture/application (7/17 – 41.2%). Existing traces were used in 11/17 (64.7%) of the studies while 6/17 (35.3%) generated their own data. Of the 6 who generated their own data, 4 did so because they required web traces of specific tasks while the other 2 studies collected general web browsing activity through a proxy server and client-side logging, presumably in order to capture current web browsing behaviour.

Purpose of Research	**Studies Surveyed**
Evaluation of Tools, Techniques & Interfaces	S1, S3, S4, S5, S9, S11, S16, S18, S19, S20, S21, S27, S29
Information Seeking Behaviour	S6, S7, S9, S12, S13, S14, S22, S23, S24, 26, S30, S31
Web Navigation Behaviour	S8, S15, S17, S25
General Web Use Behaviour	S2, S10, S28
Application	S31

Table 2. **Studies categorized by purpose of research.**

Web traces were collected using a variety of logging techniques. Server-side logging was used in 8/17 (47.1%) instances, followed by client side logging in 5/17 (29.4%), proxy logging 3/17 (17.6%) and 1/17 (5.9%) used traces from multiple locations. The dates of the trace(s) used appeared in 10/17 (58.8%) of the 2003/2004 application papers. Of the 12 dates known: 2 were from 1995, 4 were from 1996-1998, 4 were from 1999-2000, and 2 were from 2002. The duration of the trace(s) appeared in 13/17 (76.5%) of the papers. Of the 15 durations given: 4 traces captured 1 day or less, 1 captured 1-2 weeks, 2 were for 1-2 months, 4 took place for 4 months, and 4 traces were from 6-18 months long.

5.5 Experimental Strategy

We categorized each of the surveyed studies according to its experimental strategy, as defined by McGrath (1995) (see Table 3). Although some overlap exists between the different strategies and some studies may not fit precisely within a single strategy, we assigned each study to the category deemed most representative of the strategy used. The authors of this paper discussed each categorization until a consensus was reached. A brief description of 5 of the 8 strategies defined by McGrath (1995) follows. The strategies of *judgment study* (response to stimulus), *formal theory* and *computer simulation* do not fit the studies of web browsing behaviour included in our survey.

Experimental Strategy	Studies Surveyed
Laboratory Experiment	S3, S4, S5, S9, S11, S16, S16, S19, S20, S21, S24, S25, S27, S29
Field Experiment	S10, S13, S17, S18, S22, S23, S26
Experimental Simulation	S1, S6, S8, S12, S31
Field Study	S2, S14, S28, S30
Survey Sample	S7

Table 3. **Studies categorized by experimental strategy.**

In a *field study*, researchers directly observe participants in a natural setting, such that the behaviour would occur whether or not the study existed. There is little intrusion by the researchers. In a *field experiment*, researchers study participants through direct observations, in a natural setting, such that the behaviour would occur whether or not the study existed. Researchers typically manipulate some feature in the natural system resulting in higher level of obtrusiveness. During an *experimental simulation*, participants are studied in a contrived situation. In an effort to introduce a degree of realism, the situation is similar to some class of naturally occurring behaviour**.** In a *laboratory experiment***,** the situation is contrived; researchers control the elements of the environment. For a *survey sample*, responses are elicited from a sample population with respect to a subject of interest to obtain information about the respondents.

The most common experimental strategy was laboratory experiment (14/31 – 45.2%), followed by field experiment (7/31 – 22.6%), experimental simulation (5/31 – 16.1%), field study (4/31 – 12.9%) and survey sample (1/31 – 3.2%).

6 Analysis of Contextual Information

We surveyed the studies for the level of contextual information provided according to the 8 categories established in the Contextual Information section: population, dates, environment, nature of the browsing task, measures, descriptive reporting, statistical analysis, and results in context. For each of the elements of contextual information we present the results of our survey. In all cases, the authors of the paper made the judgments about the classification and interpretation of the contextual information. In the case of discrepancy, we achieved consensus through discussion.

6.1 Population

Population information was evaluated using binary judgements as to whether or not the specific population information had been provided. Where possible, we reported population characteristics of the participants taking part in the surveyed studies.

6.1.1 Size of Sample Population

Sample size is the most commonly reported piece of contextual information. We found that 29/31 (93.5%) of all the studies reported the size of the sample population. The mean sample size was 40.4 (median of 20; range of 4 to 305). In addition to sample size, reporting the male/female spread is also important. We found that 13/29 of those who reported the sample size also reported the gender of participants. The male/female spreads of these 13 studies were fairly balanced with a mean ratio of 55% male and median of 52% male.

6.1.2 Population Age

We found that 11/31 (35.5%) of the studies surveyed reported information relating to the age of the participants. The type of information provided was not consistent however and could not be compared across studies. The method of reporting for population age ranged from mean age, to a range of ages, to age categories.

6.1.3 Background/Occupation

Clear and informative background or occupation details were reported in 28/31 (90.3%) of the surveyed studies. Of those who did provide this information, 21/28 studies reported that the sample populations were recruited from within an academic institution (students, staff and faculty). Not surprisingly, many of these were computer science or information science students. This population is used for convenience as this is a demographic readily available to most researchers.

6.1.4 Web Experience

We found that 17/31 (54.8%) of the studies reported some description of the web experience of their participants. However, researchers described web experience quite differently, making it difficult to compare the experience of participants from different studies. In a study for which web experience is particularly important, more concrete measures of web experience may be necessary. For instance, GVU's WWW User Surveys (1997) included questions that measure web experience.

6.2 Dates

In the studies we surveyed, both the study dates and durations were evaluated using a binary evaluation scheme. We found that only 8/31 (25.8%) of the studies provided information regarding the date(s) during which the study was conducted. Although study dates can be inferred from publication dates, research is often published a year or two after the study is carried out. We found that for those papers published in 2003 that provided the dates of the studies, the range of study executions was from early 2000 to mid 2002. When examining the reported duration of the surveyed studies, we found that 23/31 (74.2%) reported some indication of the duration of the study. Of these 23 studies, 12/23 took place during a single session, 2/23 took place for 1-7 days, 2/23 took place for 1-4 weeks, 5/23 took place for 1-12 months, and 2 took place for 18 months.

6.3 Environment

Details regarding the environment of a study were evaluated using a binary evaluation scheme. We examined whether studies reported the setting of the experiment as well as the software used by the participants during the study. We also provide a breakdown of the settings and software reported by the surveyed studies. It was found that 10/31 (32.3%) of the studies provided concrete details regarding the setting of the study. Of these, 5/10 stated the study took place in a lab or researcher's office, 2/10 in the participant's home, 2/10 in the participant's workplace, and 1/10 studied some participants in the home

and others in the workplace. Of the remaining 21 studies, 15/21 appeared to take place in either a lab or researcher's office (although not explicitly stated), 1/21 was conducted in the user's natural environment (but it was not clear whether this was home, work, or school), and 4/21 did not provide any detail with respect to setting. Finally, 1 study stated that some aspects of the study took part in the researcher's office while the rest took place in the user's natural environment; but again this was not clear if this was work, home, or school.

Most studies reported some detail on the software used by the participants. Web browsers were used in 24/31 (77.4%) of the studies and the name of the particular web browser used was provided by 16/24 studies. Of the 24 studies that did use web browsers, 5/24 augmented the browser with new functionality, 3/24 used them to access library portals, and 2/24 used them in conjunction with other custom software. For the remaining 7 studies that did not report using a web browser, 4/7 used custom built software, 2/7 did not provide any detail at all regarding the type of software used by the participants, and in a single case there was no software used since it was only questionnaire based.

6.4 Web Browsing Task

We examined the surveys for the nature of the web browsing carried out by the participants and classified web browsing behaviour as: mandated, work-related, school-related, personal, and no browsing. Several studies included more than one type of browsing and these were typically combinations of personal/school/work related browsing.

We found 13 instances of mandated browsing, meaning that participants browsed pages mandated by the researchers involved in the study. We also found 6 instances of work-related browsing, 7 instances of school-related browsing, and 7 instances of personal browsing. There were 8 studies in which there was no browsing required of the users. Of these, 6/8 were mandated navigation, meaning participants manipulated navigation tools such as the history list, back button, or favourites. On several occasions, we found it was difficult to discern the type of browsing due to the lack of details in the methodology. In the cases where it was not explicitly stated, we made a "best guess" based on information that was implied or was visible through the quotes and anecdotes.

6.5 Measures

The surveyed papers were classified according to the following types of measures: observations, self-reports, trace measures, and archival records. In many cases, combinations of different measures were reported. Where appropriate, we have also provided information regarding specific tools and techniques reported by the surveyed studies.

Observations and self-reports were the two most common types of measures used to collect data in the studies we surveyed. Each of these measures were used by 24/31 (77.4%) of studies. We also found that 21/31 (67.7%) studies used a combination of both self-reports and visible observations. We classified all of the observations as visible (the subjects knew that their actions were being observed). Trace measures were used in only 3/31 (9.7%) of the studies and archival records in 2/31 (6.5%). We did encounter 1 study in which the measures used were not clear although from the data described, they appeared to consist of either visible observations or trace measures.

Visible observation data collection can further be broken down by method. Software logging was used in 16/24 (66.7%) of these studies. Screen capture was used in 6/24 (25.0%) of studies, with half of those using software and half using video recording of the screen. A combination of software logging and screen capture were used in 3/24 (12.5%) of studies. Automatic data collection was often augmented by researcher observations and in 2/24 (8.3%) of studies, research observations were the sole method of data collection. In 2 cases, it was unclear whether the data was collected manually by the researcher or through logging software.

We further examined the 20/31 (64.5%) of the studies where it was clear that some form of computerized data collection was used beyond screen captures. Of those, 15/20 (75.0%) were identified as software logging, 3/20 (15.0%) as trace measures, 1/20 (5.0%) as archival records, and 1/20 (5.0%) as a combination of archival and software logging. The descriptions of the software used varied widely. While 10/20 (50.0%) gave full details that clearly identified the specific software used, 7/20 (35.0%) only gave partial details such as the general type of software used, and 3/20 (15.0%) gave no details at all. We further examined these studies for details of the data recorded. We found that only 5/20 (25.0%) of the studies reported full details of the data that was recorded. The remaining studies (15/20 – 75.0%) did not specifically address what was actually recorded or reported only partial details. We often needed to infer what data was collected through the results that were reported. We also examined whether or not these studies addressed the impact of the method of data collection on the metrics being reported. We found that 8/20 (40.0%) did fully report on how the data collection impacted the metrics they reported, 7/20 (35.0%) gave partial discussion, while 5/20 (25.0%) did not address this issue at all.

6.6 Descriptive Reporting

We examined the surveys in terms of the types of the descriptive results reported: raw data, aggregate data, measures of variability, quotes, and anecdotes. Many papers reported multiple data types. Aggregate data was most common (26/31 – 83.9%), followed by raw data (15/31 – 48.4%), measures of variability (15/31 – 48.4%), quotes (9/31 – 29%) and anecdotes (6/31 – 19.3%).

We observed instances of aggregate data reporting in which a single participant accounted for a large percentage of web activity (between 25% to 40%). Simple aggregation without normalization of the data can give a false sense of the general user patterns. We found several instances where studies that included quotes of user dialog and anecdotal passages enabled us to gain more information about the methodology and characteristics of the sample population.

6.7 Statistical Analysis

In order to examine the statistical analyses described by each study, a four point scale was defined as follows: 1) *No statistical analysis* given; 2) *Significance stated*, but no descriptive statistics given; 3) *Statistics and descriptions* given, and 4) *Fully explained statistics*. Each surveyed paper was assigned a single rating on the four point scale based on the detail of statistical analysis provided. .

Based on our analysis, we classified the studies in the survey as: no statistical analysis (10/31 – 32.2%), significance stated (2/31 – 6.5%), statistics and descriptions (11/31 – 35.5%) and fully explained statistics (8/31 – 25.8%). In the case of studies that provided fully explained statistics, in addition to reporting the statistics, authors carefully described the analyses used and the motivation for using them. For the remaining 10 studies that did not report any statistical analysis, 8 were purely qualitative while 2 presented their findings in terms of raw data and percentages. Many of these studies had a small sample size or were exploratory in nature. In general, most studies provided an adequate description of the statistical analyses performed and the results obtained

6.8 Results in Context

We examined the studies in order to determine if the results were discussed in context of previous work. Discussions of the results in the context of previous work were judged on a three point scale that is defined as follows: 1) Provided *no discussion* of the results in context of previous work, 2) Provided a *brief discussion* of the results in context of previous work, and 3) Provided an *in-depth discussion* of the results in context of previous work.

Based on our analysis, we found the following breakdown: no discussion (12/31 – 38.7%), brief discussion (5/31 – 16.1%) and in-depth discussion (14/31 – 45.2%). The results of our analysis seem to

indicate that most studies either provided no discussion of the results in context of previous work or they provided an in-depth discussion.

6.9 Overall Scores

As a general basis for comparison, we computed scores for the results presented in the previous sections (population, dates, environment, web browsing task, measures, descriptive reporting, statistical analysis and results in context). We assigned binary (0-not reported, 1-reported) or tertiary (0-not reported, 1-partial details, 2-full details) scores as appropriate. We then summed the points to arrive at a total score for each study. We normalized this total score by computing the ratio of points earned to potential maximum score, taking into consideration those points that were not applicable for the study. While this is a rough metric and does not include all possible contextual detail, it serves as a useful tool to get a more general sense of the overall scope of reporting. We found a wide range in the overall level of contextual data reported in the studies surveyed, with a mean normalized total score of 0.65 (range of 0.38 to 0.93).

7 Recommendations/Discussion

We now discuss our findings and make recommendations in three areas. First, we discuss issues regarding the temporal context of web browsing behaviour research. We then examine the reporting of studies in our survey and give recommendations for increasing the contextual information reported. Finally, we look at the design of future studies with particular attention to the implications of experimental strategies, individual differences and population characteristics.

7.1 Temporal Context

As previously discussed, the state of the Web has changed quickly and drastically since its inception. It is important that seminal works are acknowledged; but, given the ever-changing state of the Web, there is a concern that data that is no longer relevant is being used to support current research. Care must be taken to ensure the context in which the data was recorded does not differ significantly from the current context with respect to the aspects of web browsing behaviour under study. These seminal works do however provide us with a baseline from which we can measure the changes in user behaviour through the evolution of the Web.

The reliance on older data sets was noted in the studies that used an application of web browsing data (Appendix 2). Often, older trace data sets were used; traces were usually pre-2001 (83% of the studies that reported a date). In some cases, this seemed to be due to a lack of newer trace data available to the public. There may be some benefit in evaluating new algorithms with the same data sets used in previous evaluations, but there also needs to be some validation that algorithms scale to the much busier and larger Web environment of today. We encourage those needing to use existing web traces to examine the temporal contexts of the existing traces and the current environment. This would allow researchers to evaluate whether or not it is reasonable to assume that the older traces are still representative of the behaviour of current web users.

One example of changing user patterns is research about Back button usage. Catledge & Pitkow (1995) reported the Back button was used in 41% of all navigation, while one year later Tauscher (1997) reported the Back button was used in only 30% of all navigation. In the two studies reported in the Smartback paper (Milic-Frayling et al., 2004) (dates unknown, approximately 12 months apart, and published in 2004), back button usage was down to 22% (exploratory study) and approximately 8% (back button and Smartback button equivalent, evaluation study) of all navigation. In a recently study (completed in March 2005), we found that the back button was responsible for 18% of all navigation. However, each of these studies had a relatively small number of participants and there may be individual

differences or population differences that account for the decrease in usage in addition to the increased in navigation aids such as auto-completion of URLs and enhanced History and Favorites features.

This does not imply that no results from different contexts are relevant, but the relevance has to be challenged by evaluating the context of the state of the Web, the web browsing environment and the characteristics of the user population studied. There are aspects of web browsing behaviour that may be relatively stable. During the construction of the timeline, it was noted that the page view time has remained fairly constant at about 55-60 seconds per page with no large variations in the 1999-2004 monthly reports from Nielsen//Net ratings.

7.2 Reporting of Context

The results of our survey highlighted the areas of reporting done well and those that need some improvement. Information was reported well with respect to details about user population, sample size, and background information; but the ratio of male to female participants and their age was reported infrequently. The reporting of web experience is an area that needs improvement in both including the information and providing detail as to the classification of users. In many other research domains, the date of the study may not bear much importance. However, in the domain of user behaviour on the Web, dates and duration of studies are crucial to give temporal context to the study, but were lacking in the majority of studies surveyed.

The level of reporting of methodological details was highly variable. Space limitations may account for some of the lack of detail. However, if we examine the normalized total scores by publication venue (Figures 1 & 2), we can see that, as expected, there are fewer contextual details in short conference papers than in long conference papers or journals. However, there is still a great deal of

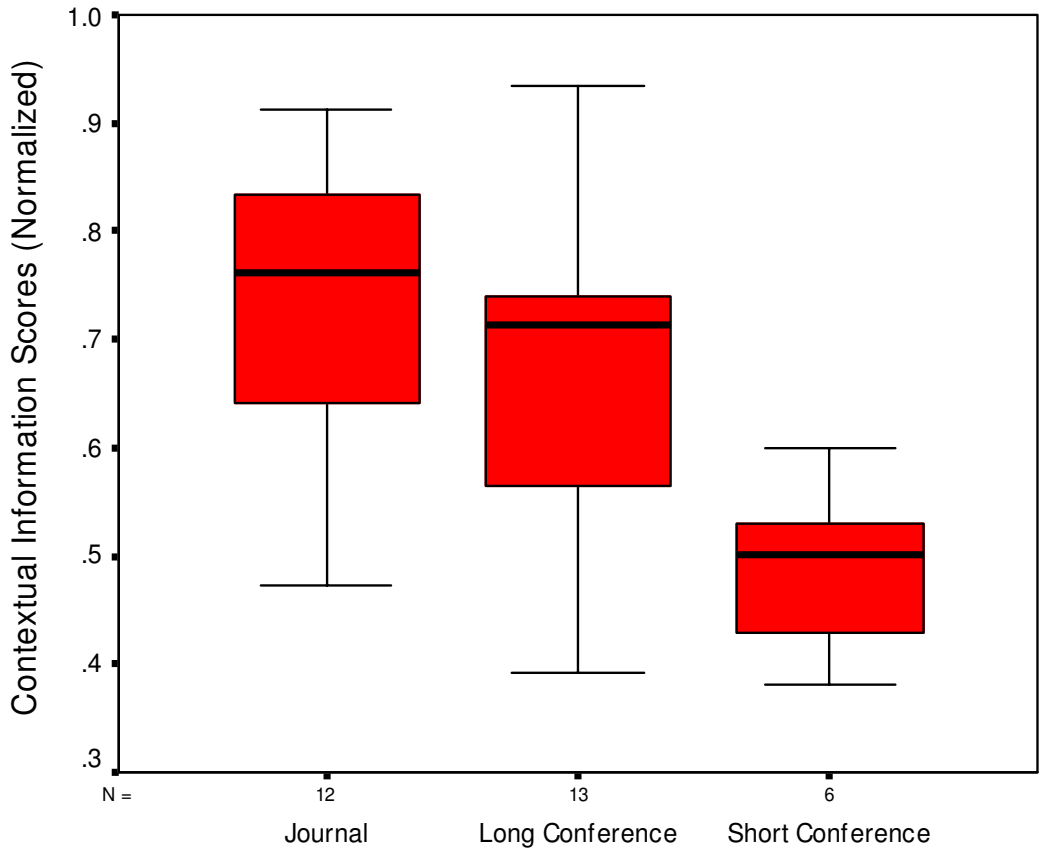

***Figure 1*. Box plot shows the variability of contextual information scores by publication venue**.

variability in the level of context reported within each venue. It is clear that space limitations do not fully account for the lack of contextual detail reported. In cases where space limitations are an issue, we recommend that authors create technical reports from experimental design documents to divulge methodological details including task descriptions, study instruments, and well-defined metrics. This detail will ensure that the reported results can be used as a basis for, and compared with, future works.

When there are large individual differences in behaviour, as exemplified by highly variable data, it is important to get a sense of the underlying data. Technical reports or public web sites can be a means of disclosing raw data collected in the study. Only half of the studies surveyed gave a sense of the raw data and reported the amount of variance in the data. However, the nature of web usage data, especially when collected in a field situation, often means it may contain personal or identifying information. It may be possible to blind the data sufficiently so that privacy is protected. If the data cannot be blinded and must remain private, providing measures of variance can be helpful

Much of the detail found lacking in our survey should not require a great deal of space to report. The date of the study, for instance, is crucial for research in this area, yet was omitted in almost 75% of the studies surveyed. Clarifications about the population being studied, the environment of the study and the nature of the web browsing task are all aspects that can be addressed briefly and make a great deal of difference in the ability of the audience to understand the research and compare and contrast it with other results.

More complete descriptions of the logging software employed and its impact on the metrics reported would assist other researchers in their determination of what type of logging software is appropriate for studies under design. Only half of the studies that used logging software gave full identifying details about the software. Commercial products often do not log all the aspects of web browsing that need to be captured, so custom software is often required. However, it can be challenging

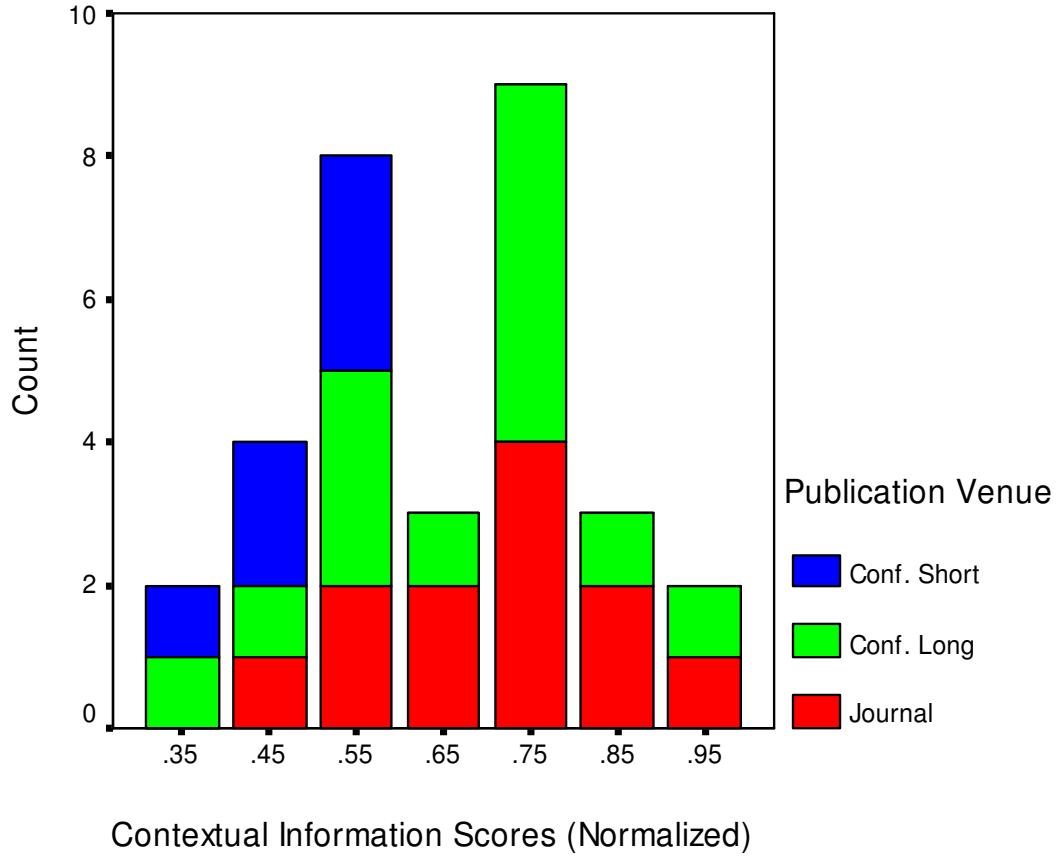

Figure 2. **The bar chart shows the distribution of the normalized scores by publication venue.**

to build custom logging software, as it must work with existing web browsers and tools or provide equivalent functionality. This is particularly challenging when the software must be robust enough to be employed client-side on multiple user computers in a field study. With current reporting practices, it can sometimes be difficult to determine the approach taken, let alone the specific methods of capturing the data.

Current reporting practices can make it difficult to compare new results with previous results in this area due to inadequate reporting of context, both temporal and methodological. We recently presented results from a field-study (Hawkey and Inkpen, 2005) that updated per-session and per-browser window usage (e.g. the number of pages visited, the speed of browsing, sessions) from those figures previously reported. However, we experienced difficulty finding previous studies with enough contextual details to allow meaningful comparisons of results. Those publications that did contain sufficient details allowed us to know when comparisons were inappropriate and to reflect upon changes in the context of browsing that may account for the differences noted. This ability to place results in the context of previous work is crucial and the responsibility lies with the research community to make sure that sufficient details are presented. This will ensure that the contextual details that are important for allowing comparison of web browsing behaviour are addressed.

7.3 Design of Studies

We noted during our survey that in the majority of studies, participants tended to be convenience samples recruited through the academic community. These users no longer represent typical web users and care must be taken when generalizing results to the general community. We may still need to conduct tightly constrained studies that, although they may not be reflective of the population, offer valuable insight, especially for exploratory studies. However, there is a need for follow-up research that explores the generalizability of the results. This can be either through a large study of a heterogeneous population or a series of smaller, more focussed studies comparing the web browsing behaviour of different types of participants.

There is also a need for complementary experimental strategies. Broad field studies are necessary to discover natural browsing behaviours, but more closely controlled experiments are also needed to isolate specific aspects of browsing and effects of task and environment. If metrics from each study are fully defined, complementary strategies employed within the research community can be more easily assimilated to advance the overall state of research.

8 Conclusions

We have presented the need for providing contextual information, including temporal information, related to web browsing behaviour studies. We detailed eight categories of contextual information that we deem to be crucial for readers to gain a full understanding of the research being presented and have discussed the impact on results for each of these categories if the details are not reported. Our survey of recently published papers highlighted this need for a set of guidelines to provide structure in reporting as currently there is a lack of contextual information being reported in studies of web browsing behaviour.

Although the community as a whole may argue that these suggestions are obvious and elementary guidelines on conducting research, the fact of the matter is that they are not being followed. As a result it can be difficult to learn the current knowledge about web browsing behaviours and judge if and how these behaviours are changing over time. In order to advance research in the field, it is important that researchers are able to find areas for investigation and that is difficult when the research previously done is not well defined. It is also difficult for researchers and practitioners to use the results to guide the development of algorithms, tools, or applications in this area.

9 Acknowledgements

Thanks to our supervisors, Dr. Kori Inkpen, Dr. Carolyn Watters, and Dr. Michael Shepherd for their guidance and the EDGE Lab and WIFL Group for their support. This work is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

Appendix 1 - Surveyed Studies

The set of all publications that described the *surveyed* 2003/2004 studies of user behaviour on the Web are listed below. Each reference is preceded by 'S', indicating it is a surveyed study.

[S1] Bhavnani, S., Bichakjiam, C., Johnson, T., Little, R., Peck,F., Schwartz, J., et al. (2003). Strategy Hubs: Next-Generation Domain Portals with Search Procedures. *Proc. of CHI 2003*, Ft. Lauderdale, FL. 393-400.

[S2,S3,S4,S5] Cockburn, A., Greenberg, S., Jones, S., McKenzie, B., and Moyle, M. (2003). Improving Web Page Revisitation: Analysis, Design and Evaluation. *IT&Society, 1*(3): 159-183.

[S2*] Cockburn, A. and McKenzie B. (2001). What Do Web Users Do? An Empirical Analysis of Web Use. *Int. J. of Human-Computer Studies*, 54:903-922.

[S3] Moyle, M. and Cockburn, A. (2003). The Design and Evaluation of a Flick Gesture for 'Back' and 'Forward' in Web Browsers. *Proc. of AUIC2003*, Adelaide, Australia. 39-46.

[S3*] Moyle, M. and Cockburn, A. (2002). Gesture Navigation: An Alternative 'Back' for the Future. *Proc of CHI 2002 Extended Abstracts*, Minneapolis, MN. 822-823.

[S4*] Cockburn, A. and McKenzie, B. (2002). Pushing Back: Evaluating a new Behavior for the Back and Forward Buttons. *Int. J of Human-Computer Studies*. 57: 397:414

[S5*] Andy Cockburn, Saul Greenberg, Bruce McKenzie, Michael Jasonsmith aund Shaun Kaasten. (1999). WebView: A Graphical Aid for Revisiting Web Pages. *Proc. OZCHI'99 Australian Conf. on HCI.*

[S6] Drabenstott, K. (2003). Do Nondomain Experts Enlist the Strategies of Domain Experts. *JASIST*, 59(9): 836-854.

[S7] Heinström, J. (2003). Fast Surfers, Broad Scanners and Deep Divers as Users of Information Technology - Relating Information Preferences to Personality Traits. *Proc. of ASIST 2003*, Long Beach, CA. 247-253.

[S8] Herder, E. and Juvina, I. (2004). Discovery of Individual User Navigation Styles. In *Proc. of Workshop on Individual Differences, AH2004*, Eindhoven, The Netherlands, 40-49.

[S9] Ihadjadene, M., Chaudiron, S., and Martins, D. (2003). The Effect of Individual Differences on Searching the Web. *Proc. of ASIST 2003*, Long Beach, CA. 240-246.

[S10] Jackson, L. A., Eye, A. v., Barbatsis, G., Biocca, F., Zhao, Y., and Fitzgerald, H. E. (2003). Internet Attitudes and Internet Use: Some Surprising Findings from the Homenettoo Project. *Int. J. Human-Computer Studies, 59*: 355-382.

[S11] JasonSmith, M. and Cockburn, A. (2003). Get a Way Back: Evaluating Retrieval from History Lists. *Proc. of AUIC2003*, Adelaide, Australia. 33-38.

[S12] Jenkins, C., Corritore, C., and Wiedenbeck, S. (2003). Patterns of Information Seeking on the Web: A Qualitative Study of Domain Expertise and Web Expertise. *IT & Society, 1*(3): 64-89.

[S13] Kelly, D. and Belkin, N. (2004). Display Time as Implicit Feedback: Understanding Task Effects. *Proc. of SIGIR 2004*, Sheffield, UK. 377-384.

[S14] Mat-Hassan, M. and Levene, M. (2003). Understanding Web Searching & Navigation Patterns. *Proc. of WWW 2003*, Budapest, Hungary.

[S15] Milic-Frayling, N., Sommerer, R., and Rodden, K. (2003). Webscout: Support for Revisitation of Web Pages within a Navigation Session. *Proc. of WI'03*, Halifax, Canada. 689-693.

[S16] Milic-Frayling, N., Sommerer, R., Rodden, K., and Blackwell, A. (2003). Searchmobil: Web Viewing and Search for Mobile Devices. *Proc. of WWW 2003*, Budapest, Hungary.

[S17,S18] Milic-Frayling, N., Jones, R., Rodden, K., Smyth, G., Blackwell, A., and Sommerer, R. (2004). Smartback: Supporting Users in Back Navigation. *Proc. of WWW 2004*, New York, NY. 63-71.

[S19,S20] Obendorf, H. and Weinrich, H. (2003). Comparing Link Marker Visualization Techniques -- Changes in Reading Behavior. *Proc. of WWW 2003*, Budapest, Hungary.

[S21] Pearson, R. and Schaik, P. v. (2003). The Effect of Spatial Layout of and Link Colour in Web Pages on Performance in a Visual Search Task and an Interactive Search Task. *Int. J. Human-Computer Studies, 59*(2003): 327-353.

[S22] Pennanen, M. and Vakkari, P. (2003). Students' Conceptual Structure, Search Process, and Outcome While Preparing a Research Proposal: A Longitudinal Case Study. *JASIST*, 54(8): 759-770.

[S23] Rieh, S. Y. (2003). Investigating Web Searching Behavior in Home Environments. *Proc. ASIST 2003*, Long Beach, CA. 255-264.

[S24] Rouet, J.-F. (2003). What was I looking for? The influence of task specificity and prior knowledge on students' search strategies in hypertext. *Interacting with Computers*, Vol. 15, 409-428.

[S25] Song, G. and Salvendy, G. (2003). Effectiveness of Automatic and Expert Generated Narrative and Guided Instructions for Task-Oriented Web Browsing. *Int. J. Human-Computer Studies*, 59(2003): 777-795.

[S26] Teevan, J., Alvarado, C., Ackerman, M., and Karger, D. (2004). The Perfect Search Engine Is Not Enough: A Study of Orienteering Behavior in Directed Search. *Proc. of CHI 2004*, Vienna, Austria. 415-422.

[S27] White, R. W., Jose, J. M., and Ruthven, I. (2003). Using Top-Ranking Sentences for Web Search Result Presentation. *Proc. of WWW 2003*, Budapest, Hungary.

[S28] Yamakami, T. (2003). Towards Know-When Technology in the Mobile Information Space: Long-Term User Trace Log Analysis in the Mobile Internet. *Proc. of WI'03*, Halifax, Canada. 545-549.

[S29] Yip, A. (2004). The Effects of Different Types of Site Maps on User's Performance in an Information-Searching Task. *Proc. of WWW 2004*, New York, NY. 368-369.

[S30] Zhang, D., Zambrowicz, C., Zhou, H., and Roderer, N. (2004). User Information Seeking Behavior in a Medical Web Portal Environment: A Preliminary Study. *JASIST, 55*(8): 670-684.

[S31] Zhu, T., Greiner, R., and Haubl, G. (2003). An Effective Complete-Web Recommender System. *Proc. of WWW 2003*, Budapest, Hungary.

Appendix 2 - Applications

The set of all surveyed 2003/2004 publications of applications of user behaviour on the Web are listed below. Each reference is preceded by 'A', indicating it is an application paper.

[A1] Chatterjee, P. Hoffman, D.L., and Novak, T.P. (2003). Modeling the Clickstream: Implications for Web-Based Advertising Efforts. *Marketing Science*. Vol. 22, No. 4, 520-541

[A2] Chi, E., Rosien, A., Supattanasiri, G., Williams, A., Royer, C., Chow, C., et al. (2003). The Bloodhound Project: Automating Discovery of Web Usability Issues Using the Infoscent Simulator. *Proc of CHI 2003*, Ft. Lauderdale, FL. 505-512.

[A3] Iamnitchi, A. (2003). Data-Sharing Relationships in the Web. *Proc. of WWW 2003*, Budapest, Hungary.

[A4] Kim, H. R. and Chan, P. K. (2003). Learning Implicit User Interest Hierarchy for Context in Personalization. *Proc. of IUI'03*, Miami, FL. 101-108.

[A5] Lu, J., Xu, B., and Yang, H. (2003). Matrix Dimensionality Reduction for Mining Web Logs. In *Proc. of WI'03*, Halifax, Canada. 405-408.

[A6] Mazza, R. and Dimitrova, V. (2004). Visualising Student Tracking Data to Support Instructors in Web-Based Distance Education. *Proc. of WWW 2004*, New York, NY. 154-161.

[A7] Nagino, N. and Yamada, S. (2003). Future View: Web Navigation Based on Learning User's Browsing Patterns. *Proc. of WI'03*, Halifax, Canada. 541-544.

[A8] Niu, Y., Zheng, T., Chen, J., and Goebel, R. (2003). Webkiv: Visualizing Structure and Navigation for Web Mining Applications. *Proc. of WI'03*. 207-213.

[A9] Oztekin, B. U., Ertoz, L., Kumar, V., and Srivastava, J. (2003). Usage Aware Pagerank. *Technical Report No. TR 03-010*. Minneapolis, MN. University of Minnesota.

[A10] Park, Y.-H. and Fader, P. S. (2004). Modeling Browsing Behavior at Multiple Websites. *Marketing Science*, 23(3): 280-303.

[A11] Ruvini, J.-D. (2003). Adapting to the User's Internet Search Strategy on Small Devices. *Proc. of IUI'03*, Miami, FL. 284-286.

[A12] Sharon, T., Lieberman, H., and Selker, T. (2003). A Zero-Input Interface for Leveraging Group Experience in Web Browsing. *Proc. of IUI'03*, Miami, FL. 290-292.

[A13] Sugiyama, K., Hatano, K., and Yoshikawa, M. (2004). Adaptive Web Search Based on User Profile Constructed without Any Effort from Users. *Proc of WWW 2004*, New York, NY. 675-684.

[A14] Wu, J. and Rangaswamy, A. (2003). A Fuzzy Set Model of Search and Consideration with an Application to an Online Market. *Marketing Science*. Vol. 22, No. 3, 411-434.

[A15] Wu, T., Ahuja, S., and Dixit, S. (2003). Efficient Mobile Content Delivery by Exploiting User Interest Correlation. *Proc. of WWW 2003*, Budapest, Hungary.

[A16] Xue, G.-R., Zeng, H.-J., Chen, Z., Ma, W.-Y., Zhang, H.-J., and Lu, C.-J. (2003). User Access Pattern Enhanced Small Web Search. *Proc. of WWW 2003*, Budapest, Hungary.

[A17]Yoshida, S., Yukawa, T., and Kuwabara, K. (2003). Constructing and Examining Personalized Cooccurrence -Based Thesauri on Web Pages. *Proc. of WWW 2003*, Budapest, Hungary.

References

Ahuja, J. S. and Webster, J. (2001). Perceived disorientation: an examination of a new measure to assess web design effectiveness. *Interacting with Computers, 14*: 15-29.

Anick, P. (2003). Using Terminological Feedback for Web Search Refinement - A Log-based Study. *Proc. of SIGIR 2003*, Toronto, Canada. 88-95.

Barford, P., Bestavros, A., Bradley, A., and Crovella, M. (1998). Changes in Web Client Access Patterns: Characteristics and Caching Implications. *Special Issue on WWW Characterization and Performace, WWW Journal*.

Byrne, M., John, B., Wehrle, N., and Crow, D. (1999). The Tangled Web We Wove: A Taskonomy of WWW Use. *Proc. of CHI '99*, Pittsburgh, PA. 544-551.

Card, S., Pirolli, P., Van Der Wege, M., Morrison, J., Reeder, R., Schraedley, P., et al. (2001). Information Scent as a Driver of Web Behavior Graphs: Results of a Protocol Analysis Method for Web Usability. *Proc. of CHI 2001*, Seattle, WA. 498-505.

Catledge, L. and Pitkow, J. (1995). Characterizing Browsing Strategies in the World-Wide Web. *Proc. of WWW 1995*, Darmstadt, Germany. 1065 - 1073.

Choo, C. W., Detlor, B., and Turnbull, D. (2000). Information Seeking on the Web: An Integrated Model of Browsing and Searching. *First Monday, 5*(2): http://firstmonday.org/issues/issue5_2/choo/index.html.

Claypool, M., Le, P., Waseda, M., and Brown, D. (2001). Implicit Interest Indicators. *Proc. of IUI '01*, Santa Fe, NM. 33-40.

ClickZStatsStaff. (2002). Internet Usage Stats. *Online at: www.clickz.com/stats/big_picture/traffic_patterns/article.php/960101.*

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.

Danielson, D. R. (2002). Web navigation and the behavioral effects of constantly visible site maps. *Interacting with Computers, 14*(601-618).

Fenstermacher, K. and Ginsburg, M. (2003). Client-Side Monitoring for Web Mining. *JASIST, 54*(7): 625-637.

Google. (2004). Google Corporate Information: Timeline. *Online at: http://www.google.ca/corporate/timeline.html.*

Grace-Martin, M. and Gay, G. (2001). Web Browsing, Mobile Computing and Academic Performance. *Educational Technology & Society, 4*(3): 95-107.

GVUOnlineSurvey. (1997). GVU's 8th WWW User Survey. *Online at: http://www.cc.gatech.edu/gvu/user_surveys/survey-1997-10.*

Hargittai, E. (2002). Beyond Logs and Surveys: In-Depth Measures of People's Web Use Skills. *JASIST, 53*(14): 1239-1244.

Hawkey, K. and Inkpen, K. (2005). Web Browsing Today: The impact of changing contexts on user activity. *Proc. of Extended Abstracts of CHI*, Portland, Oregon. 1443-1446.

Heinström, J. (2003). Fast Surfers, Broad Scanners and Deep Divers as Users of Information Technology - Relating Information Preferences to Personality Traits. *Proc. of ASIST 2003*, Long Beach, CA. 247-253.

Herder, E. and Juvina, I. (2004). Discovery of Individual User Navigation Styles. *Proc. of Workshop on Individual Differences, AH2004*.

Hilbert, D. and Redmiles, D. (2000). Extracting Usability Information from User Interface Events. *ACM Computing Surveys, 32*(4): 384-421.

Holscher, C. and Strube, G. (2000). Web Search Behavior of Internet Experts and Newbies. *Proc. of WWW 2000*, Amsterdam, The Netherlands. 337-346.

Hong, J. I., Heer, J., Waterson, S., and Landay, J. A. (2001). WebQuilt: A Proxy-based Approach to Remote Web Usability Testing. *ACM Transactions on Information Systems, 19*(3): 263-285.

Huberman, B., Pirolli, P., Pitkow, J., and Lukose, R. (1998). Strong Regularities in World Wide Web Surfing. *Science, 280*: 95-97.

Ihadjadene, M., Chaudiron, S., and Martins, D. (2003). The Effect of Individual Differences on Searching the Web. *Proc. of ASIST 2003*, Long Beach, CA.

Infoplease. (2004a). Daily Internet Activities. *Online at: http://www.infoplease.com/ipa/A0921860.html. Retrieved November 4, 2004.*

Infoplease. (2004b). Internet Timeline. *Online at: http://www.infoplease.com/ipa/A0193167.html. Retrieved November 4, 2004.*

Infoplease. (2004c). Internet use from Any Location by Individuals Age Three and Older. *Online at: http://www.infoplease.com/ipa/A0901651.html. Retrieved November 2,2004.*

Jenkins, C., Corritore, C., and Wiedenbeck, S. (2003). Patterns of Information Seeking on the Web: A Qualitative Study of Domain Expertise and Web Expertise. *IT & Society, 1*(3): 64-89.

Jones, W., Dumais, S., and Bruce, H. (2002). Once Found, What Then?: A Study of "Keeping" Behaviors in the Personal Use of Web Information. *Proc. of ASIST 2002*. 391-402.

Kehoe, C. M., Pitkow, J., Sutton, K., Aggarwal, G., and Rogers, J. D. (1999). Results of GVU's Tenth World Wide Web User Survey: Online at: http://www.cc.gtech.edu/gvu/user_surveys/survey-1998-10/tenthreport.html.

Kelly, D. and Belkin, N. (2004). Display Time as Implicit Feedback: Understanding Task Effects. *Proc. of SIGIR 2004*, Sheffield, UK. 377-384.

Kerner, S. M. (2004). More Broadband Usage Means More Online Spending. *Online at: www.clickz.com/stats/markets/broadband/article.php/3419281.*

Kim, K.-S. and Allen, B. (2002). Cognitive and Task Influences on Web Searching Behavior. *JASIST, 53*(2): 109-119.

Laskowski, S., Morse, E., and Gray, W. (2001). CIFter Project Main Page. *Online at: http://zing.ncsl.nist.gov/cifter/. Retrieved April 18, 2005.*

Loeber, S. C. and Cristea, A. (2003). A WWW Information Seeking Process Model. *Educational Technology & Society, 6*(3): 43-52.

Lycos. (1999). The Lycos 50 Daily Report. *Online at: http://50.lycos.com/083099.html.*

Marsh, D. (2003). History of the Internet. *Online at: http://www.internetvalley.com/archives/mirrors/ davemarsh-timeline-1.htm.*

McGrath, J. E. (1995). Methodology matters: doing research in the behavioral and social sciences. In J. G. R. Baeker, W. Buxton, and S. Greenberg (Ed.), *Human-Computer Interaction: Toward the Year 2000* (pp. 152-169).

Milic-Frayling, N., Jones, R., Rodden, K., Smyth, G., Blackwell, A., and Sommerer, R. (2004). SmartBack: Supporting Users in Back Navigation. *Proc. of WWW 2004*, New York, NY. 63-71.

Nielsen//NetRatings. (2004a). United States: Average Web Usage, Month of September 2004, Home Panel. *Online at: www.nielsen-netratings.com.*

Nielsen//NetRatings. (2004b). United States: Average Web Usage, Month of September 2004, Work Panel. *Online at: www.nielsen-netratings.com.*

Obendor, H., Weinreich, H., and Hass, T. (2004). Automatic Support for Web User Studies with SCONE and TEA. *Proc. of CHI 2004*, Vienna, Austria. 1135-1138.

Pastore, M. (1998a). Microsoft Leads Browser Race. *Online at: www.clickz.com/stats/big_picture /hardware/article.php/151351.*

Pastore, M. (1998b). Online Users Need Speed. *Online at: www.clickz.com/stats/markets/ broadband/article.php/151701.*

Pastore, M. (2000a). E-Commerce, Mobile Access Drawing Interest from Net Users. *Online at: www.clickz.com/stats/big_picture/geographics/article.php/5911_494701.*

Pastore, M. (2000b). Internet Usage Stats. *Online at: www.clickz.com/stats/big_picture /traffic_patterns/article.php/291211.*

Pastore, M. (2000c). Slow Modems Still Dominate Home Internet Scene. *Online at: www.clickz.com/stats/big_picture/hardware/article.php/277191.*

Pastore, M. (2001). Online Consumers Now the Average Consumer. *Online at: www.clickz.com/stats/big_picture/demographics/article.php/5901_800201.*

Pirolli, P., Pitkow, J., and Rao, R. (1996). Silk from a Sow's Ear: Extracting Usable Structures from the Web. *Proc. of CHI '96*, Vancouver, Canada. 118 - 125.

Pitkow, J. (1997). In Search of Reliable Usage Data on the WWW. *Proc. of WWW 1997*, Santa Clara, CA. 1343-1355.

Pitkow, J. and Kehoe, C. (1996). Emerging Trends in the WWW User Population. *Communications of the ACM, 39*(6): 106-108.

Pitkow, J. and Recker, M. M. (1994). Using the Web as a Survey Tool: Results from the Second WWW Survey. *Computer Networks and ISDN Systems, 27*(6): 809-822.

Reeder, R., Pirolli, P., and Card, S. (2001). WebEyeMapper and WebLogger: tools for analyzing eye tracking data collected in Web-use studies. *Proc. of the Extended Abstracts of CHI 2000*, Seattle, WA. 19-20.

Rieh, S. Y. (2003). Investigating Web Searching Behavior in Home Environments. *Proc. of ASIST 2003*, Long Beach, CA. 255-264.

Schiano, D., Stone, M., and Bectarte, R. (2001). Search and the Subjective Web. *Proc. of the Extended Abstracts of CHI 2001*. 165-166.

Sellen, A. J., Murphy, R., and Shaw, K. L. (2002). How Knowledge Workers Use the Web. *Proc. of CHI '99*, Minneapolis, MN. 227-234.

Spink, A., Wolfram, D., Jansen, M. B. J., and Saracevic, T. (2001). Searching the Web: the Public and their Queries. *JASIST, 52*(3): 226-234.

Tauscher, L. and Greenberg, S. (1997). How People Revisit Web Pages: Empirical Findings and Implications for the Design of History Systems. *Int. J. Human-Computer Studies, 47*: 97-137.

Teevan, J., Alvarado, C., Ackerman, M., and Karger, D. (2004). The Perfect Search Engine is Not Enough: A Study of Orienteering Behavior in Directed Search. *Proc. of CHI 2004*, Vienna, Austria. 415-422.

Thury, E. M. (1998). Analysis of student Web browsing behavior: implications for designing and evaluating Web sites. *Proc. of 1998 Conf. on Computer Documentation*, Quebec City, Canada. 265-270.

Turnbull, D. (1998). WebTracker: A Tool for Understanding Web Use. *Online at: http://www.ischool.utexas.edu/~donturn/research/webtracker/.*

W3C. (1999). Web Characterization Terminology & Definitions Sheet. *Available online at: http://www.w3.org/1999/05/WCA-terms/01.*

Wildemuth, B., Barry, C., Luo, L., Crystal, A., and Oh, S. (2004). Establishing a Research Agenda for Studies of Online Search Behaviors: A Delphi Study. *Online at: http://ils.unc.edu/sig_use_delphi/. Retrieved April 18, 2005.*

Zhang, D., Zambrowicz, C., Zhou, H., and Roderer, N. (2004). User Information Seeking Behavior in a Medical Web Portal Environment: A Preliminary Study. *JASIST, 55*(8): 670-684.