

Application Action	Number of Occurrences	Interface Technique	Application Category	Description of Action
Add Current	207	M	Navigate	Add Current File to Hotlist
Anchor	16176	M	Navigate	Hyperlink in Document
Annotate	44	M K	Annotate	Spawn Annotate Window
Audio Annotate	4	M	Annotate	Spawn Audio Annotation
Back	12632	M K I	Navigate	Navigate Back
Binary Transfer Mode Off	62	M	Options	Load to Disk
Binary Transfer Mode On	68	M	Options	Don't Load to Disk
Clear Global History	6	M	Options	Clear Global History
Clone Window	25	M K I	File	Clone the Window
Close Window	481	M K I	File	Close the Window
Delay Image Loading Off	9	M	Options	Delay Loading of Images
Delay Image Loading On	12	M	Options	Don't Delay Loading of Images
Exit Program	840	M	File	Exit Mosaic
Fancy Selections Off	5	M	Options	Disable Fancy Selections
Fancy Selections On	14	M	Options	Enable Fancy Selections
Find In Current	235	M K	File	Search Current File
Flush Image Cache	2	M	Options	Flush the Image Cache
Forward	537	M K I	Navigate	Navigate Forward
Home Document	179	M I	Navigate	Navigate to Home Document
Hotlist	2336	M K	Navigate	Spawn Hotlist
Interrupt	464	I	File	Abort Loading of File
Load Images in Current	2	M	Options	Load Images in Current File
Mail To	49	M K	File	Mail a File
New Window	30	M K I	File	Open a New Window
Open Local	487	M K	File	Open Local File
Open URL	1753	M K I	File	Open File via a URL
Print	350	M K	File	Print File
Refresh Current	14	M K	File	Redrawn Current File
Reload Configuration Files	3	M	Options	Reset Configuration File
Reload Current	1507	M K I	File	Reload Current File
Reload Images	14	M	File	Reload Current File's Images
Save As	340	M I	File	Save Current File
Source Document	631	M K	File	View Source
Window History	203	M K	Navigate	Spawn Window History

Table 3. List of salient user interface events with number of occurrences and a brief description.

Note that the number of occurrences above differ slightly from Table 1. This discrepancy results from differences in tabulation methods. Specifically, in the above table, all subtasks related to the application action were included. For example, Table 3 reports 203 occurrences for Window History. This number includes all events from the Window History subwindow, e.g., Help, Mail To and Go\_To. The Interface technique is added to clarify certain event. For example, a menu item exists for Add current to Hotlist as well as a button in the Hotlist subwindow. Table 3 reports the former.

## Bibliography

- [1] Bates, M.J., "The design of browsing and berrypicking techniques for the on-line search interface," *Online Review*, Vol 13. No. 5, 1989, pp. 407-431.
- [2] Berners-Lee, T., R. Cailliau, J.F. Groff and B. Pollermann. "World-Wide Web: The Information Universe." *Electronic Networking: Research, Applications and Policy*. 1992.
- [3] Bernstein, M., J.D. Bolter, M. Joyce and E. Mylonas (1991), "Architectures for Volatile Hypertext," Hypertext'91: Third ACM Conference. on Hypertext, ACM, 243-260.
- [4] Bieber, Michael, and Jiangling Wan. "Backtracking in a Multiple-window Hypertext Environment." ACM European Conference on Hypermedia Technology, 1994. pp. 158-166.
- [5] Caramel, Erran, Stephen Crawford and Hsinchun Chen. "Browsing in Hypertext: A Cognitive Study." *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 22, No. 5, Sept.-Oct 1992. pp 865-883.
- [6] Cove, J.F. and B.C. Walsh. "Online text retrieval via browsing," *Information Processing and Management*, Vol. 24, No. 1, 1988. pp. 31-37.
- [7] Crow, D. and B. Smith, in eds. Beale, R. & Finley, J. Neural Networks and Pattern Recognition in Human Computer Interaction. 1992.
- [8] Koster, M., 1994. Personal communication.
- [9] Laurel, Brenda. Computers as Theatre. Reading, MA. Addison-Wesley Publishing Co., 1991.
- [10] Lucarella, Dario. "A Model for Hypertext-Based Information Retrieval," Proceedings of the ECHT '90 European Conference on Hypertext. Cambridge University Press, 1990. pp. 81-94.
- [11] Marchionini, G. "Information-seeking strategies of novices using a full-text electronic encyclopedia," *Journal Amer. Soc. Inform. Sci.*, Vol. 40, No. 1, 1989. pp. 54-66.
- [12] Mukherjea, Sougata, James D. Foley, Scott E. Hudson. "Interactive Clustering for Navigating in Hypermedia Systems." ACM European Conference on Hypermedia Technology, 1994. pp. 136-145.
- [13] Nielsen, Jakob. Usability Engineering. Boston, MA. Academic Press, 1993.
- [14] Pitkow, James E. and Margaret M. Recker. "Integrating Bottom-Up and Top-Down Analysis for Intelligent Hypertext." *Conference on Intelligent Knowledge Management*. Intelligent Hypertext Workshop, Dec. 12, 1994, National Institute of Standard Technology.
- [15] Pitkow, J. and Recker, M. "Results from the First World-Wide Web Survey." Special issue of Journal of Computer Networks and ISDN systems, 1994 Vol. 27, no. 2.
- [16] Pitkow, J. and Recker, M. "Using the Web as a Survey Tool: Results from the Second WWW User Survey." Unpublished work, 1994.
- [17] Sproull, Lee and Sara Kiesler. Connections: New Ways of Working in the Networked Organization. Cambridge, MA: MIT Press, 1993.
- [18] Sukaviriya, Piyawadee "Noi", James D. Foley and

Todd Griffith, "A Second Generation User Interface Design Environment: The Model and the Runtime Architecture," Proceeding of the ACM INTERCHI '93 Conference on Human Factors in Computing Systems, 1993.

## Acknowledgments

Thanks to all members of the Graphics, Visualization, & Usability Center and its director, Dr. Jim Foley, for their support and help. James would also like to thank Dr. Jorge Vanegas for implicit funding throughout the development and testing of the prototype. Thanks to all who read and commented on our paper at the Human Computer Interface Consortium, especially Jonathan Grudin and Erik Altmann.

## Author Information

LARA CATLEDGE received her B.A. in Professional Writing from Carnegie Mellon University in 1992. She is currently pursuing a Master's degree in Information Design and Technology at the School of Literature, Communication and Culture at Georgia Institute of Technology. Research interests include interface design, usability, collaboration and CSCW.

JAMES PITKOW received his B.A. in Computer Science Applications in Psychology from the University of Colorado Boulder in 1993. He is a Graphics, Visualization, & Usability graduate student in the College of Computing at Georgia Institute of Technology. His research interests include user modelling, adaptive interfaces, and usability.

## *Directions for Design*

Since users accessed on average 10 pages per server, this would indicate that “must see” information must be accessible within two to three jumps of the initial home page (two/three navigations in, two/three out, performed three/two times). However, the placement of numerous links on one page can lead to increased search time by users to find relevant information as well as a cluttered screen layout. As such, information dense interface tactics that preserve screen space, such as using image maps, may be a more successful strategy for page design.

For rich information ecologies, the use of indexes throughout the document space supports hub and spoke observed usage patterns. Additionally, these pages help orient users, minimizing the “lost in hypertext” phenomenon. Since most users explored small regions at a time, this design recommendation can increase the exploration of clusters of related information.

Document designers need to be cognizant of the classification of expected visitors as serendipitous browser, general browser, or searcher. Granted, within a server collections of documents need to be targeted toward different users. Just the same, authors aware of the three classes of users can tailor documents to suit the intended use of the documents. When more than one class of visitor is expected, a separate document can be created for each class<sup>4</sup>, thus providing customized, alternative views of the information. Note that this already occurs with the stratification of users based upon graphics-based and text users as well as forms and nonforms-compliant Web clients.

In designing for all strategies and behaviors, there exists a tension between “volatile hypertexts” and efficiency (between the browser and the searcher) in all of these recommendations. However, as Sproull and Kiesler [Sproull & Kiesler, 1993] found in their study of the uses of electronic mail, efficiency may not always be the appropriate met-

---

4. Not that expensive resource-wise considering only three classes were observed.

ric for system evaluation. User satisfaction may provide a more accurate measure of the success of an interface.

In the future, servers may use the user classification to offer a “usual” view of a database. Additionally, servers could also offer a guided tour of a server based on the paths most travelled, or more excitingly, alter page design on the fly based on accesses by users.

## **Future Analysis**

Recent studies that correlate reading time with document relevancy for USENET news articles suggest that a similar correlation may exist with Web information spaces as well. That is, we hypothesize that browsers spend less time on pages and within sites than searchers.

Users who access a large number of documents in a fixed period of time will have higher y-intercepts in their individual frequency to path length plots. These users may well be prime candidates for macro suggestion. Furthermore, it would be interesting to run a correlation analysis on the y-intercepts and the total number of sites visited.

Finally, a cost function for browsing can be developed based on analysis of expected value to the user of particular information and the expected time to retrieve that information.

## **Conclusion**

This paper presented interface usage data for XMosaic and characterizations of user navigation patterns as serendipitous browsing, general browsing or searching based upon empirical analysis of user event log files. These characterizations were derived from existing hypertext research and seem to extend well into the realm of the Web.

(<http://www.cc.gatech.edu> and <http://www.gatech.edu>) due to abnormal access patterns.

## Discussion

Given the above relationship between frequency and depth, one can begin to characterize navigation strategies based on users' average slope. Using Cove and Walsh's characterizations, the following classifications can be made:

"Serendipitous Browser" (slope < -.24) These users avoid the repetition of long invocation sequences. This shallow browser may be reflective of a WWW repository structuring in that the databases visited by these users may be weakly connected.

"General Purpose Browser" (slope = -.24) Here users perform as expected. Probabilistically, they have roughly a one in four chance of repeating a more complex navigation sequence. This is the average inertia for all users sampled.

"Searcher" (slope > -.24) A user performs the same short navigation sequences relatively infrequently, but does perform long navigational sequences often.

Furthermore, the slope can be used to classify sets of documents according to their usage patterns. Table 2 displays the classification of several types of site visits as by frequency and length as supported by the data.

### *Within Site Navigation*

Overall, users tended to operate in one small area within a particular site. This structure resembles a spoke and hub structure due to the frequent use of backtracking. Backtracking occurs when a user issues the "Back" command to exit a server via the path used for entry. This "leave as you've entered" strategy was heavily used by all users. In contrast, the looping back strategy occurs when users return to the original point of entry after a path traversal by utilizing the history feature or by selecting a "Return to Home/Entry Page" link. Both navigation strategies can be visualized as a kind of spoke and hub structure. In the example below, the user orientated with <http://www.cc.gatech.edu/people> and <http://www.cc.gatech.edu/people/People.Faculty.html> as hubs.

<http://www.cc.gatech.edu/people/>  
<http://www.cc.gatech.edu/people/People.Faculty.html>  
<http://www.cc.gatech.edu/gvu/people/Faculty/Neff.Walker.html>  
<http://www.cc.gatech.edu/people/People.Faculty.html>  
<http://www.cc.gatech.edu/gvu/people/Faculty/Piyawadee.Sukaviriya.html>  
<http://www.cc.gatech.edu/people/People.Faculty.html>  
<http://www.cc.gatech.edu/gvu/people/Faculty/Michael.J.Sinclair.html>  
<http://www.cc.gatech.edu/people/People.Faculty.html>  
<http://www.cc.gatech.edu/people/>

The example above is very typical in that users rarely traverse more than two layers in the hypertext structure before returning to an entry point. Initial evidence suggests that this pattern occurs independent of hyperlink per page ratios.

### *Other Navigation Techniques*

One supplemental navigation method often observed was use of home pages as indexes to interesting places. For instance, a typical session begins with the "College of Computing Home Page" followed by a traversal to a user's personal home page. Once there, jumps to other sites, or other parts of the local database ensue. While providing similar functionality to "Hotlist" commands, the use of personal home pages as indexes allows for better layout control and customization and therefore is a natural, yet crafty adaptation to an impaired interface.

### *What's worth Saving?*

Surprisingly, only 2% of retrieved documents were either saved to file or printed. Furthermore, "Window History" and "Hotlist" based document accesses accounted for less than 3% of all accesses. The minimal use of such archival interface commands may be indicative one or more of the following: the quality of Web documents, the temporal nature of certain documents, the design of these archival interfaces, or reliance on other navigation techniques like personal home pages.

This also implies that there is minimal potential copyright infringements by this population. If material retrieved by users was printed or saved to disk, unauthorized local copies of information could potentially violate certain copyright restrictions, although legal precedence remains to be set.

The sites map to user document testing, Georgia Tech's home page server, a digital archive in Nederland, NCSA, and CERN. Users accessed a total of 1222 unique sites outside of Georgia Tech. Thus, given the estimate of Web servers during the observation period was 7,300 by SG-Scout, roughly 16% of all available sites were accessed during the study. Interestingly, items put on peoples hotlists did not match the most popular sites. The sites most accessed through the hotlist were:

1. <http://www.secapl.com>
2. <file://localhost>
3. <http://info.cern.ch>
4. <http://akebono.stanford.edu>
5. <http://www.cs.ubc.edu>

### Site Analysis

1222 sites outside of Georgia Tech were accessed by College of Computing users. A modified version of the Pattern Detection Module (PDM) algorithm [Crow & Smith, 1991] identified the frequency of repeating sequences of site and document accesses. Specifically, the program tallied the number of occurrences of sequences of accesses, or *paths*. Paths of length two through fifty were computed. For example, suppose a user went from [www.gatech.edu](http://www.gatech.edu) to [www.ici.edu](http://www.ici.edu) to [www.ncsa.uiuc.edu](http://www.ncsa.uiuc.edu) a total of seven times throughout the study, the PDM would identify a path of length three (three sites) with a frequency of seven (repeated seven times). Stated differently, the length of a path is the number of successive document requests, which are to be viewed as user navigation.

	High Frequency	Low Frequency
<b>Short Path Length</b>	home pages orientation pages meta indexes	sporadic visits dead ends un-useful pages
<b>Long Path Length</b>	source of reference sites, like NCSA or CERN	one shot resources directed searching

Table 2.Characterization of sites based on frequency and path length relations.

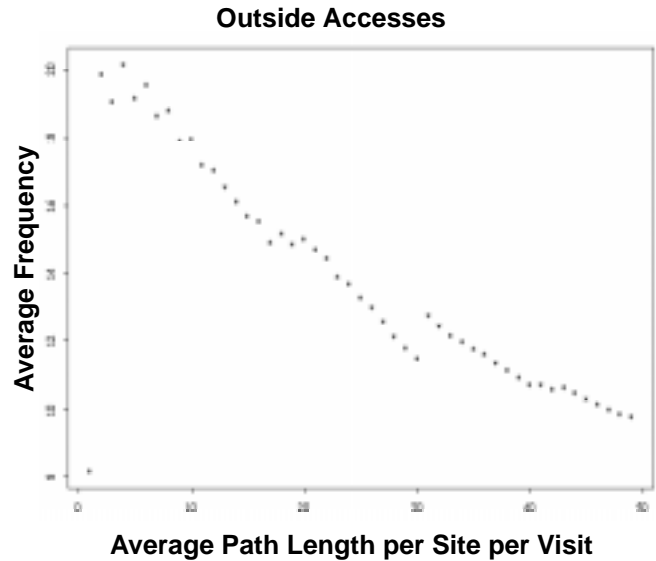


Figure 1: Average slope of all users in all sites (except [www.gatech.edu](http://www.gatech.edu), [www.cc.gatech.edu](http://www.cc.gatech.edu) and [www.cnam.fr](http://www.cnam.fr))

The PDM analysis revealed long sequences of between-site access patterns on a per-session and a per-user basis. By “per-session” we refer to patterns within a session by a single user. Likewise, by “per-user” we refer to all sessions by a user, thus allowing for the identification of between-session patterns. For the per-session analysis, paths including seven different sites occurred with a frequency of five times. On a per-user basis, the PDM algorithm identified sequences of length eight with a frequency of nine. Furthermore, numerous shorter sequences were discovered with higher frequencies with a maximum frequency of seventeen [Pitkow and Recker, 1994b].

In addition, an analysis of the length of paths within each site visited per user was performed. Figure 1 shows the average frequency per path length. This corresponds to the mean path of length x, for all x between 2 and 50. Exploratory data analysis revealed a slightly negative linear relationship between frequency and path length, with the slope across all users equalling -0.24. Thus,

$$\text{frequency} = -0.24 (\text{path length})$$

This equation was derived from all sites except <http://www.cnam.fr>, and Georgia Tech servers

## Analysis and Results

The original log file<sup>1</sup> comprised over 43,000 events, with each record uniquely identifiable by user id and time of occurrence. The file was sorted by user id and secondarily by event time. This file includes all user interface events.

Since users will often leave XMosaic running for extended periods of time without interacting with it, determining session boundaries artificially was necessary. With the intent of identifying these boundaries, the time between each event for all events across users was calculated. The mean between each user interface event was 9.3 minutes. In order to determine session boundaries, all events that occurred over 25.5 minutes apart were delineated as a new session. This means that most statistically significant events occurred within 1-1/2 standard deviations (25.5 minutes) from the mean. Thus, a new log file was derived that indicated sessions for each user. Interestingly, a consistent third quartile was observed across all users, though we note no clear explanation for this effect.

Users averaged 9.4 sessions each, or approximately one session every other day. For subsequent analyses, navigational related events were extracted<sup>2</sup>, which brought the total number of events to 31,134 representing 73% of all generated events (for a complete list of all events, see Table 3).

Document requests were distinguished by protocol. Eighty percent of the document requests were of type http (i.e. requests for a document from a WWW servers). Four percent of these were generated by "cgi" scripts. Files accounted for 8%, followed by ftp and gopher both at 4%. All other accesses combined (including news, wais, telnet, etc.) totalled 4%.

---

1. The datasets are freely available. Interested researchers should contact the authors.  
2. Location transparent commands like "Back" and "Home" were substituted with the corresponding URLs.

## *Methods of Interaction*

Hyperlinks were by far the preferred method of traversal, accounting for 52% of all document requests. Second, accounting for about 41%, was the "Back" command. Following in order of popularity were "Open URL," "Hotlist," "Forward," "Open Local," "Home Document" and "Window History" (see Table 1). This indicates that users typically did not know the location of documents a priori, or relied on other heuristics to navigate to a specific document. Furthermore, most users did not select items in the hotlist and window history. It seems that they either preferred using "Go\_To" or did not know how to employ this interface technique.

While all menu items have corresponding keyboard equivalents only 4272 events were instantiated via the keyboard, though this may be due to the lack of display of keyboard equivalents next to menu items, as is done on Macintosh applications. Finally, 486 or 1% interrupts/asynchronous aborts (hitting the spinning globe) occurred during file transfer. This indicates that the population as a whole was insensitive to retrieval latency, although there may be a difference for users using modems or slower connections.

## *Within Site Navigation*

Average successive document requests<sup>3</sup> within a single site across all users was 12.64. Outlier removal resulted in a mean of 10.31 (min=1, max=403) with a standard deviation of 28.56.

## *Popularity of Sites*

The five most popular sites were:

1. file://localhost
2. http://www.gatech.edu
3. http://w3.eeb.ele.tue.nl
4. http://www.ncsa.uiuc.edu
5. http://info.cern.ch

---

3. For the purposes of this analysis, document request and document access are used interchangeably. The terms page and document are also used interchangeably.

Application Action	Interface Technique	Instances	Percentage	Category of Action	Description of Action
Anchor	M	16140	51.9	Navigate	Selection of Hyperlink in Document
Back	M K	12633	40.6	Navigate	Go Back One Document
Open URL	M K	707	2	Navigate	Open File via a URL
Hotlist - Go To	M	636	2	Navigate	Go to Document via Hotlist
Forward	M K	537	2	Navigate	Go Forward One Document
Open Local	M K	221	.7	Navigate	Open Local File
Home Document	M K	179	.5	Navigate	Go to the Home Document
Window History	M K	39	.1	Navigate	Go to Document via Window History

Table 1. Occurrence of X Mosaic navigation-related user events mapped to UIDE- like representation, where M = mouse click; K = keyboard entry (after Sukaviriya et al., 1993)

minimize the potential for data loss resulting from network and/or system failures, all captured events were processed and forwarded to a secure disk via the syslog daemon.

Equally important was infusing a meaningful representation into the data of user events. This allows not only a clear understanding of the extent and functionality of the interface, but also allowed for clear extraction of task specific data during analysis. Accordingly, we recorded events according to the User Interface Design Environment (UIDE) [Sukaviriya, et. al, 1993] guidelines for task representation. This permits all actions to be viewed on three levels: an Application Action (high-level task, e.g. Open File), an Interface Action (mid-level task, e.g. select item from pull-down menu), and an Interface Technique (low level action, e.g. Mouse Click). In the example below, a user clicked on a hyperlink in the document window that pointed to <http://www.somewhere/>. The user is identified as participant number 123, and the event was generated from machine foo.gatech.edu on August 3rd, 1994 at 12:21:10 a.m.

```
Aug 3 00:21:10 foo.cc.gatech.edu uel: 775887872 123 1
Mouse Navigate Anchor:: http://www.somewhere/
```

The study was conducted for a three week period that commenced August 3, 1994. Participation was solicited through a consent window that

informed users of the experimental procedures employed as well as of their rights as human subjects. The intent of the consent window was both informative and to minimize the “Big Brother” effect [Nielsen, 1993]. This window appeared the first time XMosaic was executed by each user during the sampling period. One hundred and seven users, or sixty-three percent, chose to participate in the study.

The selection of XMosaic was made for several reasons. According to some estimates at the time [Kostner, 1994], XMosaic accounted for roughly 53% of all WWW related accesses to HTTP servers. Furthermore, XMosaic was one of the only UNIX based GUI browsers available. Still, since the computing environment studied also included several other platforms that supported non-logging WWW browsers, certain portions of the computing population were not able to participate. Another confound of the experimental design exists in that it was possible for users to compute on multiple platforms during the sampling period, which may have resulted in the users running the specialized Sun OS version of XMosaic in tandem with other non-logging versions of WWW browsers.

This continuum provides a nice middle ground to distinguish between browsing as a method of completing a task and open ended browsing with no particular goal in mind. Marchionini [Marchionini, 1989] further develops this distinction in designating open and closed tasks. Closed tasks have a specific answer and often integrate sub-goals. Open tasks are much more subject oriented and less specific. Browsing can be used as a method of fulfilling either open or closed tasks.

Intuitively, it would seem that browsing and searching are not mutually exclusive activities. In Bates's [Bates, 1989] work on berrypicking, a user's search strategy is constantly evolving through browsing. Users often move back and forth between strategies. Similarly, Bieber and Wan [Bieber & Wan, 1994] discuss the use of backtracking within a multi-windowed hypertext environment. They introduce the concept of "task-based backtracking," in which a user backtracks to compare information from different sources for the same task or to operate two tasks simultaneously. A similar technique, in a Web environment, would be backtracking to review previously retrieved pages.

All of these studies were performed on closed, single-author systems. The WWW however, is an open, collaborative and exceedingly dynamic hypermedia system. These previous findings provide the basis and structure for the describing the ways a user population behaves in a dynamic information ecology, like the WWW.

Given that we expect to find the same kinds of strategies used in the WWW, supporting both the browser and the searcher in designing WWW pages and servers is necessary, although difficult. Furthermore, supporting the kind of task switching described by Bates and Bieber and Wan adds another level of complexity because the work implies that a user should be able to switch strategies at any time.

It has long been recognized that methods for supporting directed searching are needed. As a response to this, certain WWW servers are completely searchable and there are World-Wide Web search engines available.

Supporting browsing, though, may be a more difficult task. Both Laurel [Laurel, 1991] and Bernstein approach the topic of how to assess and design hypertexts for the browsing user. Laurel considers interactivity to be the primary goal. She defines a continuum for interactivity along three variables: frequency (frequency of choices), range (number of possible choices) and significance (implication of choices). Laurel contends that users will pay the price "often enthusiastically -- in order to gain a kind of lifelikeness, including the possibility of surprise and delight." Bernstein takes a slightly different approach with his "volatile hypertexts" [Bernstein, 1991]. He argues that the value of hypertext lies in its ability to create serendipitous connections between unexpected ideas.

There is a tension between designing for a browser and designing for a searcher. The logical hierarchy of a file structure or a searchable database may work fine for a closed-task, goal oriented user. But a user looking for the unexpected element or a serendipitous connection may be frustrated by the precision required by these methods. The first step in balancing this problem is to determine what strategies are being used by the population. In order to do this, we collected log files of users interacting with the Web.

## Methodology

We sought to capture all events generated by consenting Georgia Institute of Technology's College of Computing staff, faculty and student populations who operate NCSA's XMosaic running Sun OS 4.1.3. Towards this end, a version of XMosaic was coded to trap all user interface level events. The computing environment of the study consisted of over 250 Sun OS 4.1.3 machines connected via a 100 Megabit/sec CDDI LAN. To



# Characterizing Browsing Strategies in the World-Wide Web

Lara D. Catledge<sup>1</sup> & James E. Pitkow<sup>2</sup>

<sup>1</sup>School of Literature, Communication and Culture

<sup>2</sup>Graphics, Visualization, & Usability Center

Georgia Institute of Technology

Atlanta, GA 30332-0280

E-mail {lara, pitkow}@cc.gatech.edu

## Abstract

This paper presents the results of a study conducted at Georgia Institute of Technology that captured client-side user events of NCSA's XMosaic. Actual user behavior, as determined from client-side log file analysis, supplemented our understanding of user navigation strategies as well as provided real interface usage data. Log file analysis also yielded design and usability suggestions for WWW pages, sites and browsers. The methodology of the study and findings are discussed along with future research directions.

## Keywords

Hypertext Navigation, Log Files, User Modeling

## Introduction

With the prolific growth of the World-Wide Web (WWW) [Berners-Lee et.al, 1992] in the past year there has been an increased demand for an understanding of the WWW audience. Several studies exist that determine demographics and some behavioral characteristics of WWW users via self-selection [Pitkow and Recker 1994a & 1994b]. Though highly informative, such studies only provide high level trends in Web use (e.g. frequency of Web browser usage to access research reports, weather information, etc). Other areas of audience analysis, such as navigation strategies and interface usage remain unstudied. Thus, the surveys provide estimations of who is using the WWW, but

fail to provide detailed information on exactly how the Web is being used. Actual user behavior, as determined from client-side log file analysis, can supplement the understanding of Web users with more concrete data. Log file analysis also yields design and usability guidelines for WWW pages, sites and browsers.

This paper presents the results of a three week study conducted at Georgia Institute of Technology that captured client-side user events of NCSA's XMosaic. Specifically, the paper will first present a review of related hypertext browsing and searching literature and how it's related to the Web, followed by a description of the study's methodology. An analysis of user navigation patterns ensues. Lastly, a discussion and recommendations for document design are presented.

## Literature Review

Many studies have addressed user strategies and usability of closed hypermedia systems, databases and library information systems [Caramel et. al., 1992]. Most distinguish between browsing and searching. Cove and Walsh [Cove et. al. 1988] include a third browsing strategy:

1. Search browsing; directed search; where the goal is known
2. General purpose browsing; consulting sources that have a high likelihood of items of interest
3. Serendipitous browsing; purely random