Using the Web as a Survey Tool: Results from the Second WWW User Survey

James E. Pitkow
Graphics, Visualization, & Usability Center
Georgia Institute of Technology
Atlanta, GA 30332-0280

Margaret M. Recker
College of Computing
Georgia Institute of Technology
Atlanta, GA 30332-0280

Abstract:
This paper presents the initial results from the second World-Wide Web User Survey, which was advertised and made available to the Web user population for 38 days during October and November 1994. The survey is built on our architecture and Web technologies, which together offer a number of technical and surveying advantages. In particular, our architecture supports the use of adaptive questions, and supports methods for tracking users' responses across different surveys, allowing more in-depth analyses of survey responses. The present survey was composed of three question categories: general demographic questions, browsing usage, and questions for Web information authors. In addition, we added an additional, experimental category addressing users' attitudes toward commercial use of the Web and the Internet. In just over one month, we received over 18,000 total responses to the combined surveys. To the best of our knowledge, the number of respondents and range of questions make this survey the most reliable and comprehensive characterization of WWW users to date. It will be interesting to see if and how the user trends shown in our results change as the Web gains in global access and popularity.

Keywords:
new applications, surveys, demographics, tools.

INTRODUCTION

In the few years since its inception, the World-Wide Web, or WWW or Web, (Berners-Lee et al., 1994) has grown dramatically in the number of users, servers, and its geographical distribution (Merit NIC, 1994). These technologies for the first time hold the potential of ushering in the "Age of Information" to people of all ages, backgrounds, and economic status. Wide-spread networking coupled with the ease of publishing multimedia materials within the Web will support radical changes in areas such as medicine, education, business, and entertainment.

The universal accessibility of information technologies means that the user population will be extremely diverse in terms of skills, experiences, abilities, and backgrounds. As such, a crucial ingredient to the success of such endeavors is an understanding of its user population. One powerful method of characterizing the background, usage patterns, and preferences of users is via surveys. Coupled with other methods, such as log file analysis (e.g., Pitkow and Recker, 1994a), these results enable appropriate targeting of services, and the development of intelligent user-centered applications and interfaces.

In January of 1994, we conducted the first survey of World-Wide Web users (Pitkow and Recker, 1994b). This survey was advertised and made available to Web users for one month, and received over 4,800 responses to all questionnaires within the survey. Although quite successful, this survey suffered from a number of technical and design shortcomings that we wished to address. To this end, we modified the basic architecture in order to enhance the capabilities of the surveys. In addition, we expanded the range and focus of questions. These changes improved the robustness of the system, the reliability of the data, and the quality of the human-computer interaction.

In particular, we designed and implemented adaptive questions. With the use of adaptive questions, answers provided to certain questions are used to determine the next series of questions. In this way, respondents need not wade through a series of unrelated questions, and instead are only presented with relevant ones. Thus, adaptation serves to reduce the number and complexity of questions presented to each user. Secondly, we implemented methods for tracking respondents in a way that respects respondent privacy and guards their anonymity. This enables cross-tabulation of responses across survey sections, thus facilitation more in-depth analyses of survey responses. In addition, this method enables future longitudinal tracking of the Web user population.

As with the first survey, questions were presented in separate survey categories, which provides several advantages. First, by using categories, respondents were able to quickly finish each section of the overall survey. We note that one long survey containing all of the questions may discourage potential respondents, and adds considerably to the survey's complexity. Second, many Web browsers have difficulty managing documents with a large number of embedded forms. Third, categorizing questions allows users to decide a priori if the particular question category applies to them.

The first category asked general demographic questions about the respondent. Questions about the respondents browsing patterns, motivations, and usage comprised the second category. The third category asked questions of respondents who were information providers, about the nature of their information, and their opinions about existing tools. In addition, we added an additional, experimental category addressing users' attitudes toward commercial use of the Web and the Internet. This category was divided into a short and long version of the questionnaires, and respondents could choose which section to answer. We felt that this stratification was sufficient to help us characterize WWW users, their reasons for using the WWW, and their opinion of WWW tools and technologies.

The second survey was advertised and made available to the Web user population for 38 days during October and November 1994. During this period, we received over 18,000 total responses to the combined surveys for over 4,000 users. To the best of our knowledge, the number of respondents and range of questions made this survey the most reliable and comprehensive characterization of WWW users to date. This paper describes the technical details of the implementation and followed by a brief presentation the survey's results.

OVERVIEW

There are a variety of methods for surveying user populations via the Internet, though the effectiveness of WWW technologies presents many advantages. We define the term "effectiveness" from an overall measure of time and respondent complexity with respect to other survey methods. Though a thorough comparison of surveying techniques is beyond the scope of this paper, we will briefly overview several methods and the trade-offs involved.

Traditional e-mail based surveys require the user to perform text entry, usually by placing X's in boxes or typing numbers, then sending the message off to the surveyors. This scenario functions properly if the survey ends up in the mail boxes of respondents who are willing to respond, that is, if they self-select themselves, and expend the necessary time and effort. In other e-mail based surveys, the questions are posted to newsgroups, which then require the users to extract the message and proceed as above. Either way, once the responses have been submitted, the collation of the data can become problematic, since consistent structure within responses can only be suggested, not enforced. For example, if the question is posed "How old are you?" the answer may appear on the same line as the question, two lines below, may contain fractions, an integer, or even a floating point number. Phone-based surveys impose less of a task load on the user, but increase cognitive load by requiring the user to keep all the options in memory. Also, response data usually are entered by humans, an error-prone process. Furthermore, respondents cannot review their responses, and are typically subject to time constraints.

Use of Web technologies helps to minimize the above costs by: 1) enabling point-and-click responses, 2) providing structured responses, 3) using an electronic medium for data transfer and collation, 4) presenting the questions visually for re-inspection and review, 5) imposing very loose time constraints and finally, 6) utilizing adaptive questions to reduce the number and complexity of questions presented to users. For the purposes of this paper, complexity is defined as a metric of the visual and cognitive demands placed on a user when answering questions.

The Second WWW User Survey itself was composed of three main questionnaires and two experimental questionnaires. Extending and refining upon the initial set of questions asked in the first survey, the three main questionnaires were: General Demographic Information, WWW Browser Usage, and HTML Authoring/Publishing. Additionally, two experimental consumer surveys, developed by Sunil Gupta at the University of Michigan, were included. These were deployed as two separate surveys, Part One and Part Two, with the latter containing more in-depth questions. We note that the inclusion of surveys and questions developed externally is consistent with our philosophy of working with other interested researchers in the community during question development and refinement.

In order to convey the sense of interaction present while completing the surveys, a quick walk-though follows. After entering a unique one word id (see Longitudinal Tracking section below for details), the user is presented with the survey home page. Access to each of the surveys is provided via radio buttons and a "Press Here to Proceed to Survey" button at the bottom of the page. Once the users selects a survey in which to participate, the Question Engine (see Architecture section below) generates the initial set of questions specific to the desired survey. The initial set of questions presented is the same for all users, i.e. no adaptation occurred at this stage of question presentation. The user then answers the questions and submits the responses by clicking on the "Submit Survey" button at the end of the page. The Question Engine then processes the submitted responses, with three possible results for each submitted response:

1. The response triggers an adaptive question to be added to the list of questi ons asked to the user during the next iteration of questioning
2. The Question Engine realizes that the question has not been answered and re-asks these question during the next iteration.
3. The response is fine, and no further action occurs.

1. The response triggers an adaptive question to be added to the list of questions asked to the user during the next iteration of questioning
The list of adapted and un-answered questions is returned to the user. This cycle of "question - answer - adapt/re-ask" repeats until all questions have been answered. At this point, the user is returned to the survey home page that lists the surveys that have not yet been completed.

This iterative cycle accomplishes several goals. Foremost, the adaptation of questions reduces the number and complexity of questions presented to each user. For example, an interesting question to developers as well as Web database managers is "Who uses what browsers?" Given the existence of seven or so major platforms (e.g., X/Unix, Macintosh, PC, etc.), with numerous browsers readily available on each, the space required to list all platforms and browsers would easily fill two screens. Clearly this is undesirable and inefficient. However, by staging the question in two parts, one that asks for the primary platform of the user's browser and the other that provides a list of known browsers for that specific platform, the amount of space required to pose the question is reduced as well as the cognitive overhead necessary for the user to answer the question correctly. Additionally, this method enables the acquisition of detailed responses, which facilitates a more in-depth understanding of the user population. For example, with only two questions, the region and state of the user can be obtained.

CLASSIFICATION OF ADAPTIVE QUESTIONS

During the course of question development, we observed a certain structure that existed within question adaptation (see Table One). As with most traditional surveys and for all on-line surveys we have seen that use the Web, most questions do not result in any adaptation, or inference. We refer to such questions as Standard questions. These are the building blocks upon which all other types of questions are built. Inferential questions, on the other hand, define a class of questions that are based upon answers to previously asked questions. For the Inferential class we found it helpful to base our taxonomy on a Single Question and a Multiple Question basis, with the latter being composed of more than one Single Question. In other words, a multiple question defines a question based upon the responses to more than one question.

Inferential Question Class:

Multiple Class Properties:

Number of Questions Used:

Single Class Properties:

Number of Responses Used:

Single Response, Multiple Response, Complete Response

Number of Questions Triggered:

Single Adaptation, Multiple Adaptation

Standard Question Class:

Properties:

Question, Valid Responses, Interaction Type

Table 1. Classification of the types of Adaptive Questions

Single Question adaptation is based on the following properties: the Number of Responses Used and the Number of Questions Triggered. The Number of Responses can be divided into three categories. Single Response adaptation occurs when only one response to a question results in further questioning. An example from the survey asked "Are you the sole/primary user of you machine?" with follow-up questions only for `No' responses. Multiple Response adaptation occurs when several responses results in adaptation. Our survey did not include any from this category. It naturally follows that Complete Response adaptation occurs when all responses to a question result in additional questions. A question that falls into this category from the survey was "Which browser do you primarily use?" All answers to this question were followed with lists of specific browers for each major computing platform.

Once adaptation is triggered, either Single Adaptation or Multiple Adaptation can occur. With Single Adaptation, the response triggers only one follow-up question. With Multiple Adaptation, more than one follow-up question is asked. For example, the question: "Do you operate a WWW server?" can be classified as Complete Response, since both `Yes' and `No' answers triggered adaptation. A `No' response results in Single Adaptation, "Can you add documents to a WWW database?" A `Yes" response results in several questions ranging from choice of servers, to the speed of the network connections to the server.

Multiple Question adaptation defines the set of questions that are triggered by the responses to multiple questions. Note that each question that triggers adaptation has the properties described above: the Number of Responses Used, and the Number of Questions Triggered. Though this survey did not include questions from this class, we are currently investigating questions of this type for future surveys.

ARCHITECTURE

The main architectural issue facing the survey was the infusion of state information into the stateless HTTP protocol. State information was necessary for supporting several aspects of the surveys. First, the user's id needed to be tracked between questionnaires in order to perform between-questionnaire analysis and longitudinal tracking of users. Second, access to the responses to previously asked questions were required in order to enforce question completion within individual questionnaires. This was also required to implement the use of adaptive questions, since these are based upon the responses from multiple answers. Third, information regarding which surveys the users had finished was required in order to keep track of the remaining surveys.

Note that all but the latter case contain information that can be written to disk and read into memory between each cycle of questioning. However, we chose not to take this approach, except for survey completion information, in attempts to minimize the number of requests to disk necessary on the server and to reduce server side CPU load.

Instead, our approach was designed to leverage off of the hidden attributes of the TYPE field used in input forms in HTML. Initially, we opted to pass the data from the client to the server via the GET method(1). Since the URL contains the information passed to the server via GET, we designed the survey home page to uniquely identify each user by making it only accessible via a CGI front-end. Thus, users could add the survey home page to their hotlists and use this to re-access the next round of surveys without having to write or manually store their id. As it turns out, this decision had several interesting results. First, we discovered that several browsers had hard-coded limits to the length of URLs. Thus, once the limit was reached, these browsers failed to load the requested URLs. Second, it forced us to re-evaluate our use of GET and POST(2). In the end, we decided to keep the use of GET for access to the survey home page, but change the method for the questionnaires to POST.

One of the overall design goals was to implement the surveys with as generic an architecture as possible. We wanted the underlying code that generates and processes the surveys to only require minor adjustments between questionnaires. Towards this end, we decided to make each questionnaire a stand-alone executable that utilized a common set of library routines and structure. Figure One shows a diagram of the components of the architecture for one questionnaire.

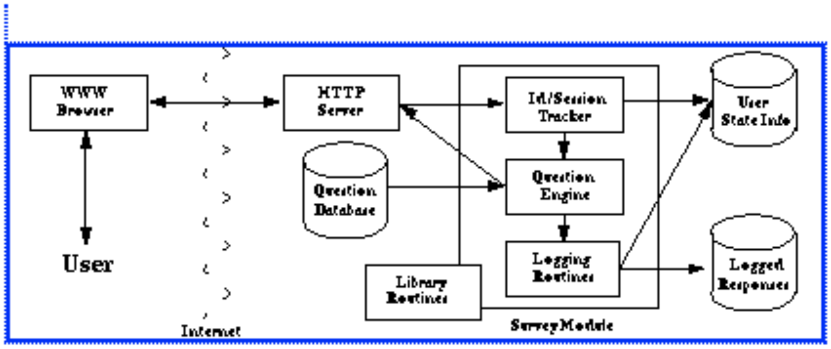


Figure 1. The above diagram overviews the architecture used for the implementation of one survey.

Integral to the design is the questionnaire database. The database is essentially an associative array of questions, which facilitates a direct mapping between the adaptable questions and the responses values of the questions they are contingent upon. Additional keys/value pairs were inserted into the database to parameterize iteration control and enforce question completion. The questions in the database are marked up in HTML.

The Id / Session Tracker manages id namespaces as well as access to questionnaires. The motivation behind tracking user ids within the survey was to: 1) allow for analysis between each questionnaire, (which the first survey did not do), 2) be flexible enough to manage users making submissions from multiple IP addresses with the same domain, 3) enable longitudinal analysis of the user population and 4) be quick and efficient from both a client and sever side perspective.

Given that the hostname and IP addresses are passed into the shell forked by the server, we chose to map namespaces to the class of the IP address. That is, class A IP numbers correspond to namespaces derived from the first octet, class B IP numbers to the first two octets, and class C IP numbers to the first three octets. This scheme permits users to fill out surveys from different machines within their organizations allocated IP numbers(3) and allows for quick conversion from IP address to the directory where the user information is stored. For example, a user whose IP address begins with 130.207 (Class B) must choose a unique id across all other users from the same domain. All subsequent information for the user is then stored in the directory /130/207.

A file exists within each namespace that keeps track of ids and the surveys that have been completed by the user. Every time a request is made for a page in the survey, the id passed to the Id Tracker is checked against the ids registered in this file. If the id is not found, the software reissues the id entry page. Similarly, upon reentry to the home page, the file is consulted to determine the remaining surveys to offer the user.

The Question Engine performs several tasks by exploiting the transparent use of associative arrays and database routines in Perl. First, it generates the initial set of questions which are returned to the user. This is accomplished by consulting the database for the total number of base questions in the survey and then looping through the associative pairs, and appending the questions (already in HTML) to the output stream. Second, the engine determines whether questions posed to the user have been answered. This task requires state information, which we handled by mangling the responses to questions into special hidden forms. Specifically, the value bound to NAME in the hidden input tag was prepended with `WWW_' and was appended to the output stream. Thus, the state of a question could be easily determined by inspecting the key of the key/value pairs passed back from the user. Finally, since the initial set of questions and their responses determine all subsequent adaptation, the state of all adapted question can be determined by evaluating simple boolean expressions, which cleanly map into the classification of questions outlined above.

The server used for the survey operates NCSA's http version 1.3 and runs on a Sparc Station 1000 running Solaris 5.3 with two co-processors, over 7 gigabytes of disk, and 175 megabytes RAM. The server resides on Georgia Tech's external FDDI ring with two T3 connections - one to NSFNET, and the other to SuraNET. The server also performs other functions like NNTP, Gopher, FTP, etc. The Survey Modules and library routines are written in Perl 4.36.

METHOD

Obviously a survey without respondents has marginal utility. Yet, the current state of WWW provides very little support for broadcasting and raising awareness of all Web users to timely or important events. As a result, cooperation and endorsement from both CERN and NCSA were obtained in publicizing the surveys. Both organizations placed links in highly visible places - CERN's Home page and NCSA's "What's New Page." Announcements and re-postings were also made to several Web related newsgroups and mailing lists including: comp.infosystems.announce, comp.infosystems.www.*, comp.internet.net-happenings, and www-talk. Additionally, several sites placed links to the survey (Dr. Fun, CUI Search Engines, EiNet's Search Page, etc.). Additionally, several trade magazines contained articles about the survey. We realize that this method of sampling is not random and are actively seeking other methods for widespread awareness of the surveys in hopes of minimizing judgement sampling bias. Furthermore, the very nature of survey methodology introduces self-selection confounds as well.

SURVEY QUESTIONS

The second survey was composed of three main questionnaires and two experimental questionnaires: General Demographic Information, WWW Browser Usage, and HTML Authoring/Publishing. The experimental questionnaires addressed commercial usage of the Web and the Internet.

The General Demographic category contained general background questions about respondents and their use of the Web (10 questions, 3 adaptive). For example, this questionnaire posed questions about the user's age, gender, geographical location, occupation, and level of education. In addition, we asked the user to identify the kind of Web browser employed. Users were also asked to estimate the amount of time spent working with computers per week. Finally, we asked the user to indicate their willingness to pay for accessing Web databases (see Appendix A for the full list of questions). As with all of the other questionnaires, a text-entry comment box was located at the end of the survey for users to contribute whatever additional information deemed relevant.

The second category contained questions about the respondents' browser use (20 questions, 0 adaptive). We asked users how often they launch their browser, the amount of time spent browsing, and their primary motivation behind browsing. Since WWW browsers allow access to almost all Internet resources, we were interested in the degree to which these browsers are replacing the client software designed for each individual resource. Hence, we asked questions on browser use to access of Gopher, FTP, etc., as well as questions on Web use for exploration and accessing other resources (e.g., weather).

Since a benefit of the Web is as a multimedia publishing environment, the third category addressed questions to users who are Web information providers (11 questions, 8 adaptive). We were interested in determining how document publishing is managed and therefore asked the user to estimate the number of documents authored, the kinds of information provided, and the nature of the organization served. We asked providers to rate their computer expertise, and how difficult they found it to become a Web information provider. Providers were also asked whether they also operate a HTTP server, and if so, the network connectivity, and platform, hardware, and software used.

Increasingly, the Internet and the Web are being considered by the commercial sector. For this reason, we added a category that addressed users' attitudes toward commercial use of the Internet. Since these issues are complex, we presented these questions within two survey sections, a short and long version of the questionnaire. Users chose which version they wished to answer. The short version contained questions about respondent's use and planned use of the Internet for product information and purchasing. In addition, we were interested in determining users' attitudes toward the purchase of information via the Internet. The long version of the questionnaire addresses the same issues, but in considerable more depth.(4)

RESULTS

Figure Two shows the daily number of visits to the survey home page and the number of respondents of each survey category during the days the survey was available. As can be seen, response started slowly, but built up as awareness of the survey spread. Several spikes are evident that correspond to when the survey was announced on highly visible pages (for example, "What's New" at NCSA). Lows are evident during weekends.

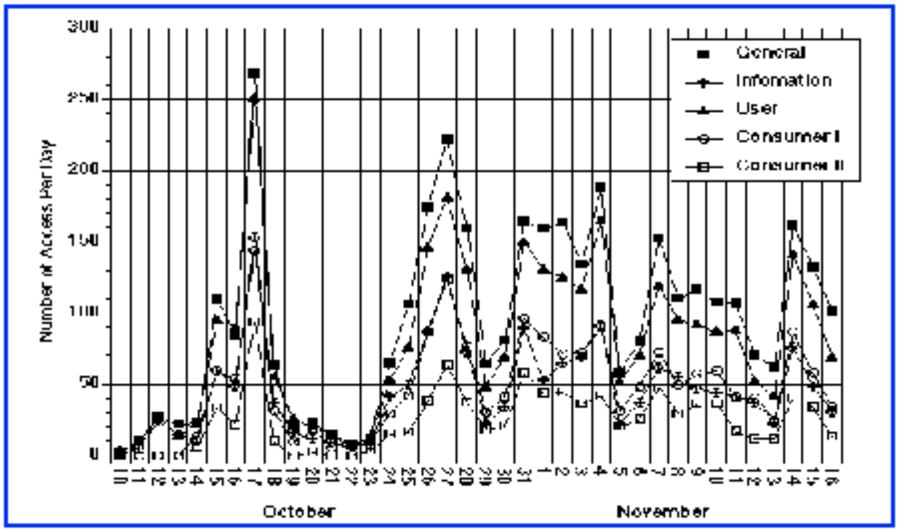


Figure 2. The number of successfully completed accesses per survey on a per day basis. Note the drop in activity during the week of October 18th (2nd International WWW Conference).

Overall, there were a total of 18,503 responses to all survey questionnaires combined. From this, 709, or 3.8%, duplicate submissions were removed. Duplicates were identified using software to detect multiple submissions of a survey by a user in the same namespace. Two thirds of the duplicates detected occurred on the experimental consumer surveys, which caused some technical problems on certain browsers due to the number of questions asked. In all cases, the last entry was used with all others being discarded from the dataset. Invalid submissions, 0.05%, resulted from browsers that mangled the response data during submission. Exactly 17,804 records (462,001 data points) were collated into the final datasets.(5)

One area of difficulty that occurred in the preprocessing stage was related to the use of text entry fields on three questions. As mentioned previously, unstructured responses are a problem with the data preprocessing of traditional surveying methods. We experienced similar problems in transforming respondent entries into uniform structured data. The two questions that enabled the user to type a number will be replaced in future surveys with ranges for the initial question and adaptation on this response in order to determine the exact number. We can, however, justify the costs incurred in one instance, where acquiring the name of the user's primary browser (as entered by the user) will assist in determining the range of options listed for each platform during subsequent surveys.

In the next section, we discuss the findings from each survey, followed by a discussion of these results. Please refer to Figures 3-8 for a graphical representation of some of the results and Appendix A for complete results.

GENERAL DEMOGRAPHICS

There were over 3522 valid responses (indicating an equal number of respondents) in this survey category, accounting for 20% of all the responses. The results indicate that the mean age of the respondents is 31, and that 44% are between the ages of 26 and 30. The youngest respondent was 12, while the oldest was 73. Figure 3 shows a histogram of users' age. Interestingly, the age does not follow a normal distribution, instead showing a rapid rise around the age of 20. We suspect this results from the relative inaccessibility of Web tools for people younger than university age.

Over 90% are male, and 87% describe their race as white. 94% do not suffer any disabilities. More than 71% of the respondents came from North America, 23% from Europe, and 3% from Australia. A more detailed breakdown shows that 12% are from California, 8% from the U.K., and 6% from Canada (Figure 7). In terms of occupation, 27% describe themselves as working in a technical field and 26% as university students (the two largest categories) (Figure 4). In terms of highest level of education completed, over 33% have university-level degrees, while 23% have completed post-graduate work, and 18% describe themselves as having "some" university-level education.

Over 51% say that their Internet access comes from the educational sector, while 30% access the Internet from the commercial sector. Only 30% report sharing their primary machine with other users. For those sharing a machine, the number shared with varied widely, with a mean of 539, a median of 20, and a maximum of 60,000. Twenty-nine percent say they use a computer over 50 hours per week, and over 19% use it between 41 and 50 hours. The most common platfrom is X (43%) followed by PC (29% and Macintosh (19%). Similarly, the most used browser is Xmosaic (40%), followed by WinMosaic (18%), and - the released middle of the survey -Netscape (18% counting all X/PC/Mac versions). Of interest to enterprises contemplating commercial use of the Internet, 71% of the respondents answered as willing to pay fees for access to WWW repositories, depending both on quality and cost. Only 21% say they would not.

BROWSER USAGE

There were 2921 valid responses to questions regarding browser use and activities. Many users access their browsers 1 to 4 times daily (40%); 38% say they spend 0 to

5 hours per week using their browser, while 35% claim to spend 6 to 10 hours per week (see Appendix A).

We surveyed users as to how often they use their WWW browser, instead of accessing specific client services , where 1 = "never" and 9 = "always." The results indicate that, overall, users show a strong preference for using their WWW browser instead of the standard Gopher and Wais clients (mean = 6.5), and to find reference and research materials. Users do not frequently use their browser to access conference information, government documents, Newsgroups, and weather information. Users report the following reasons for using the Web: browsing (79%), entertainment (65%), education (59%), work and business (47%), academic research (42%), business research (27%), other (10%), and shopping (9%).

INFORMATION PROVIDING

The survey for information providers was answered by 1669 people. As expected, given the question category, over 97% of the respondents have authored HTML documents, with people, on average, authoring 31.5 documents.

Sixty percent of the authors have authored documents for other people. Of these, 36% report authoring document for 1 to 10 people, 24% for 11 to 50 people, and 21% for over 100 people. Among authors, 61% operate a HTTP server, 58% run NCSA's server, 19% run CERN or GN (a bug in our logging software resulted in answers for each to get tallied together), 10% run MacHTTP. As mirrored by other measurements of port activity, the majority of the servers (88%) listen on port 80. Of those who do not operate a server, 81% can still add documents directly to the server area. A majority of server administrators (52%) report network connectivity of 1 megabyte/sec.

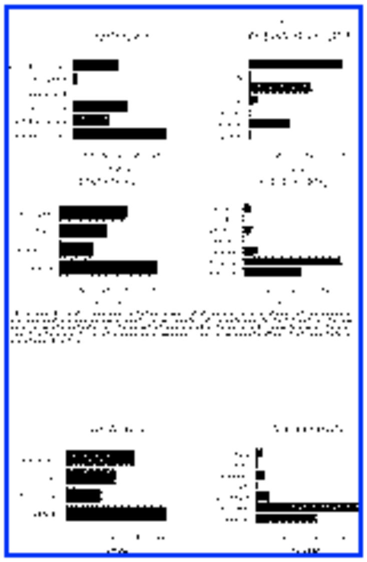
In terms of page topic, 81% report authoring documents on work, 77% on biographical information, 44% on research, 35% on entertainment, 26% on other, 25% on meta-indexes, 20% on news, 18% on product information, 14% on ads, 13% on art, 11% on conferences, and 6% on sports. Most providers (91%) know how to program, and 60% have over seven years of programming experience. Over 60% report learning HTML in 1 to 3 hours, with 89% saying it was "easy to learn" and most saying the HTML documentation was easy to understand. Fifty-seven percent have learned FORMS, and most (84%) found it easy to learn. Forty-five percent have learned ISMAP, and most (82%) found it easy to learn.

In terms of use of media, 91% report the use of images, 34% sounds, and 24% movies. In terms of links to other documents, 99% report using links, 91% use links to other Web servers, 65% use FTP links, and 50% use Gopher links. Finally, 47% report the use of FORMS, 42% report authoring interactive documents, and 25% report the use of ISMAP.

CONSUMER SURVEYS

While a more in-depth analysis of the results from the consumer commercialization sections is forthcoming, we present some interesting preliminary results. Most people "Disagree Somewhat" with the assertion that it is safe to use credit cards when making purchases from Web vendors. Similarly, nearly 82% of the users view the security of sensitive information as "Very Important," with an additional 15% regarding this issue as "Somewhat Important." Interestingly, the quality of information about purchasing choices and the reliability of Internet vendors rank even higher than the above security issues - 89% and 84% respectively. In terms of marital status, 53% of the users responded as single, and 42% responded as married.

Most users currently use the following products (listed in decreasing of order of use): compact disc (CD) players, VRC/video players, modems, and CD-ROMS. Slightly more than one out of ten users gain access to the Internet via work or school, with 28% paying for Internet access personally (note these two are not mutually exclusive). Over 42% of the users report their income as between $35,000 and $75,000, though 15% choose not to report their income. Thirteen percent report their income as below $15,000.



Figures 3 through 8. Contrary to popular belief, the distribution of ages of WWWW users does not fit a normal distribution. Technical professionals and university students together comprise the majority of the user population, with most users utilizing their WWW browser one to four times a day. The most widely used platform is X/UNIX, followed by PCs and Macintoshes. The graph of location represents the top four locations of uses, with California accounting for nearly one seventhe of all respondents. Finally, most users are either affiliated with educational institutions or commercial organizations.

CONCLUSION

In this paper, we reported results from a survey of World-Wide Web users. The survey is based upon a set of Web tools that allows the use of adaptive questions, and enables the tracking of users for longitudinal analysis. As demonstrated by the high number of survey respondents, the Web provides an easy-to-use, reliable, and low overhead survey medium. The results from our survey provide, to the best of knowledge, the most complete characterization of Web users to date. They suggest that the typical user is a 30-year old educated male from North America who works with computers. It will be interesting to see if and how these trends change as the Web gains in popularity.

In the future, we plan to deploy our survey every six months. We believe that this will be a useful means for tracking the growth and changes in Web uses and population. Given the dynamic nature of WWW use and technologies, we believe that surveys run twice a year ought to provide an optimal trade-off between maintaining respondents from survey to survey and charting the Web's growth and changes. In addition, we hope that the WWW community will allow us to remain the sole Web surveyors in this domain. We fear that if other researchers clutter the Web with similar surveys, the overall utility of such surveys will be greatly diminished. In light of such a request to the community, we gladly open ourself to suggestions and specific research agendas of other researchers.

REFERENCES

(Berners-Lee et al., 1994)
Berners-Lee, T., Cailliau, R., Luotonen, A., Nielsen, H., and Secret, A. (1994). The World-Wide Web.Communications of the ACM, 37(8):76--82.

(Merit NIC, 1994)
Merit NIC. (1994). NSFNET Statistics. Available via: URL.

(Pitkow and Recker, 1994a)
Pitkow, J. and Recker, M. (1994a). Bottom-up and top-down analysis for intelligent hypertext. In Proceedings of the Third International Conference on Information and Knowledge Management. Maryland: NIST.

(Pitkow and Recker, 1994b)
Pitkow, J. and Recker, M. (1994b). Results from the first World-Wide Web survey. Special issue of Journal of Computer Networks and ISDN systems, Vol. 27, no. 2.

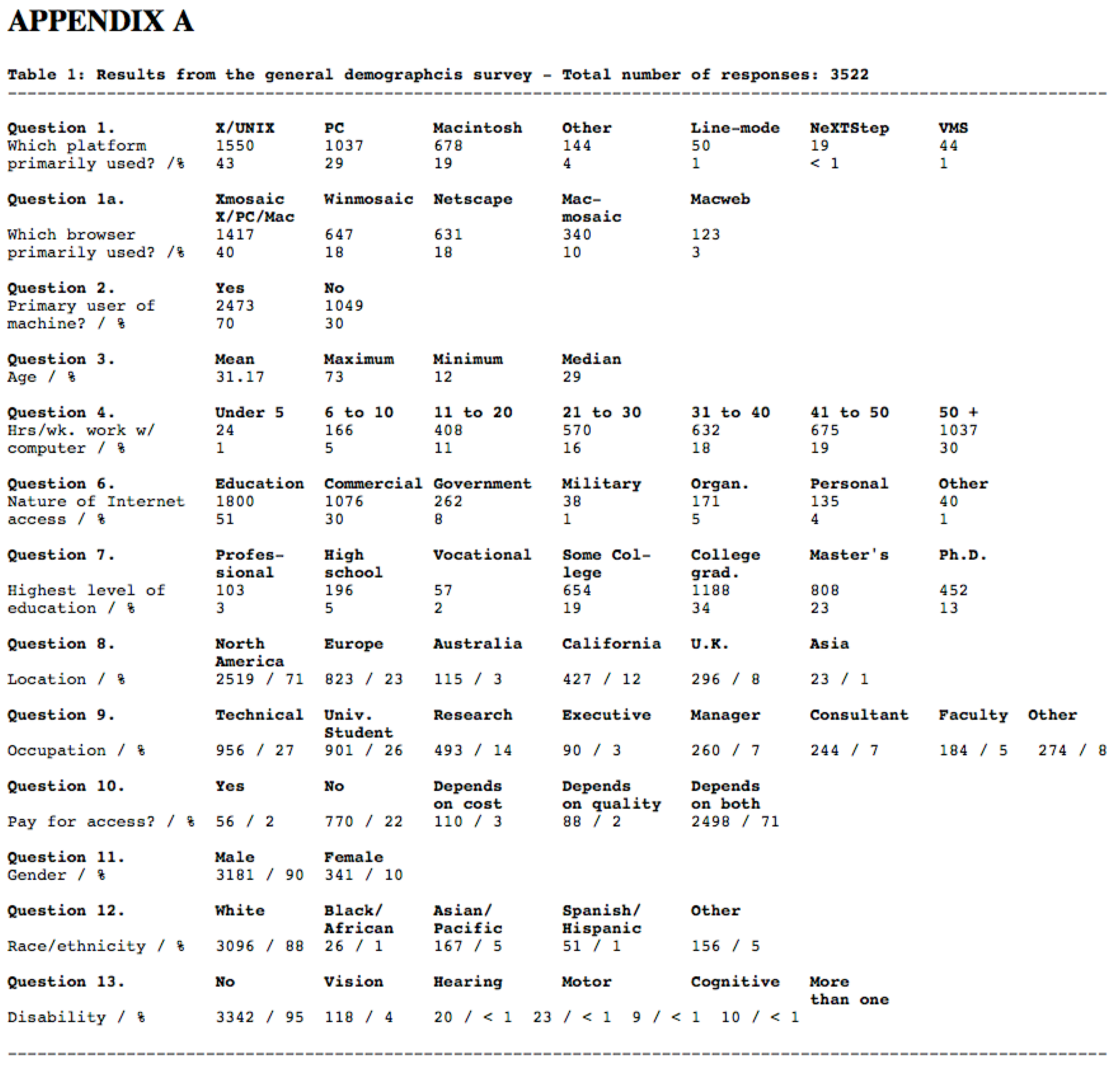
ACKNOWLEDGEMENTS

Thanks to all members of the Graphics, Visualization, & Usability Center and its director, Dr. Jim Foley, for their support and help. Special thanks extend to Michael Mealling and the entire Office of Information and Technology for their technical support during the surveys and their generous donation of equipment. Additional thanks go to Laurie Hodges and Colleen Kehoe for her help with data collation and analysis. Finally, James would like to thank Dr. Jorge Vanegas for implicit funding throughout the surveys.

AUTHOR INFORMATION

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.

MIMI RECKER received her Ph.D. from the University of California, Berkeley, in 1992. She is currently a Research Scientist in the College of Computing at the Georgia Institute of Technology. Her research interests include cognitive science approaches to learning, metacognition, instruction, interactive learning environments, human-computer interaction, cognitive modelling and multimedia.



APPENDIX A

Table 1: Results from the general demographcis survey - Total number of responses: 3522

Question 1.	**X/UNIX**	**PC**	**Macintosh**	**Other**	**Line-mode**	**NeXTStep**	**VMS**	
Which platform	1550	1037	678	144	50	19	44	
primarily used? /%	43	29	19	4	1	< 1	1	
Question 1a.	**Xmosaic**	**Winmosaic**	**Netscape**	**Mac-**	**Macweb**			
	X/PC/Mac			**mosaic**				
Which browser	1417	647	631	340	123			
primarily used? /%	40	18	18	10	3			
Question 2.	**Yes**	**No**						
Primary user of	2473	1049						
machine? / %	70	30						
Question 3.	**Mean**	**Maximum**	**Minimum**	**Median**				
Age / %	31.17	73	12	29				
Question 4.	**Under 5**	**6 to 10**	**11 to 20**	**21 to 30**	**31 to 40**	**41 to 50**	**50 +**	
Hrs/wk. work w/	24	166	408	570	632	675	1037	
computer / %	1	5	11	16	18	19	30	
Question 6.	**Education**	**Commercial**	**Government**	**Military**	**Organ.**	**Personal**	**Other**	
Nature of Internet	1800	1076	262	38	171	135	40	
access / %	51	30	8	1	5	4	1	
Question 7.	**Profes-**	**High**	**Vocational**	**Some Col-**	**College**	**Master's**	**Ph.D.**	
	sional	**school**		**lege**	**grad.**			
Highest level of	103	196	57	654	1188	808	452	
education / %	3	5	2	19	34	23	13	
Question 8.	**North**	**Europe**	**Australia**	**California**	**U.K.**	**Asia**		
	America							
Location / %	2519 / 71	823 / 23	115 / 3	427 / 12	296 / 8	23 / 1		
Question 9.	**Technical**	**Univ.**	**Research**	**Executive**	**Manager**	**Consultant**	**Faculty**	**Other**
		Student						
Occupation / %	956 / 27	901 / 26	493 / 14	90 / 3	260 / 7	244 / 7	184 / 5	274 / 8
Question 10.	**Yes**	**No**	**Depends**	**Depends**	**Depends**			
			on cost	**on quality**	**on both**			
Pay for access? / %	56 / 2	770 / 22	110 / 3	88 / 2	2498 / 71			
Question 11.	**Male**	**Female**						
Gender / %	3181 / 90	341 / 10						
Question 12.	**White**	**Black/**	**Asian/**	**Spanish/**	**Other**			
		African	**Pacific**	**Hispanic**				
Race/ethnicity / %	3096 / 88	26 / 1	167 / 5	51 / 1	156 / 5			
Question 13.	**No**	**Vision**	**Hearing**	**Motor**	**Cognitive**	**More**		
						than one		
Disability / %	3342 / 95	118 / 4	20 / < 1	23 / < 1	9 / < 1	10 / < 1		

Table 2: Results from the WWW browser usage survey - Total number of responses: 2921

Question 1.	**Over 9 / day**	**5-8 /day**	**1-4 /day**	**Few/week**	**Once/wk.**
How often use browser / %	563	455	1154	670	69
	19	16	39	23	3
Question 2.	**0-5 hours**	**6-10 hours**	**11-20 hours**	**20+ hours**	
Hours/week use browser / %	1119	1032	466	304	
	38	35	16	11	
Question 3.	**Browsing**	**Entertainment**	**Education**	**Work/Business**	**Academic Research**
Primary use of browser / % (checkbox)	2311	1912	1731	1379	1230
	79	65	59	47	42

Question	**Mean**	**St.Dev.**	**Median**
Question 4. < never (1) - always (9) >	**6.53**	**2.08**	**7**
Use Web browser for Gopher and FTP			
Question 5. < never (1) - always (9) >	**2.24**	**1.97**	**1**
Use Web browser for Newsgroups			
Question 6. < never (1) - always (9) >	**3.23**	**2.40**	**2**
Use Web browser for weather information			
Question 7. < never (1) - always (9) >	**4.53**	**2.08**	**5**
Use Web browser for reference materials			
Question 8. < never (1) - always (9) >	**4.98**	**2.30**	**5**
Use Web browser for research reports			
Question 9. < never (1) - always (9) >	**3.33**	**2.32**	**3**
Use Web browser for conference announcements			
Question 10. < never (1) - always (9) >	**3.94**	**2.37**	**3**
Use Web browser for government documents			

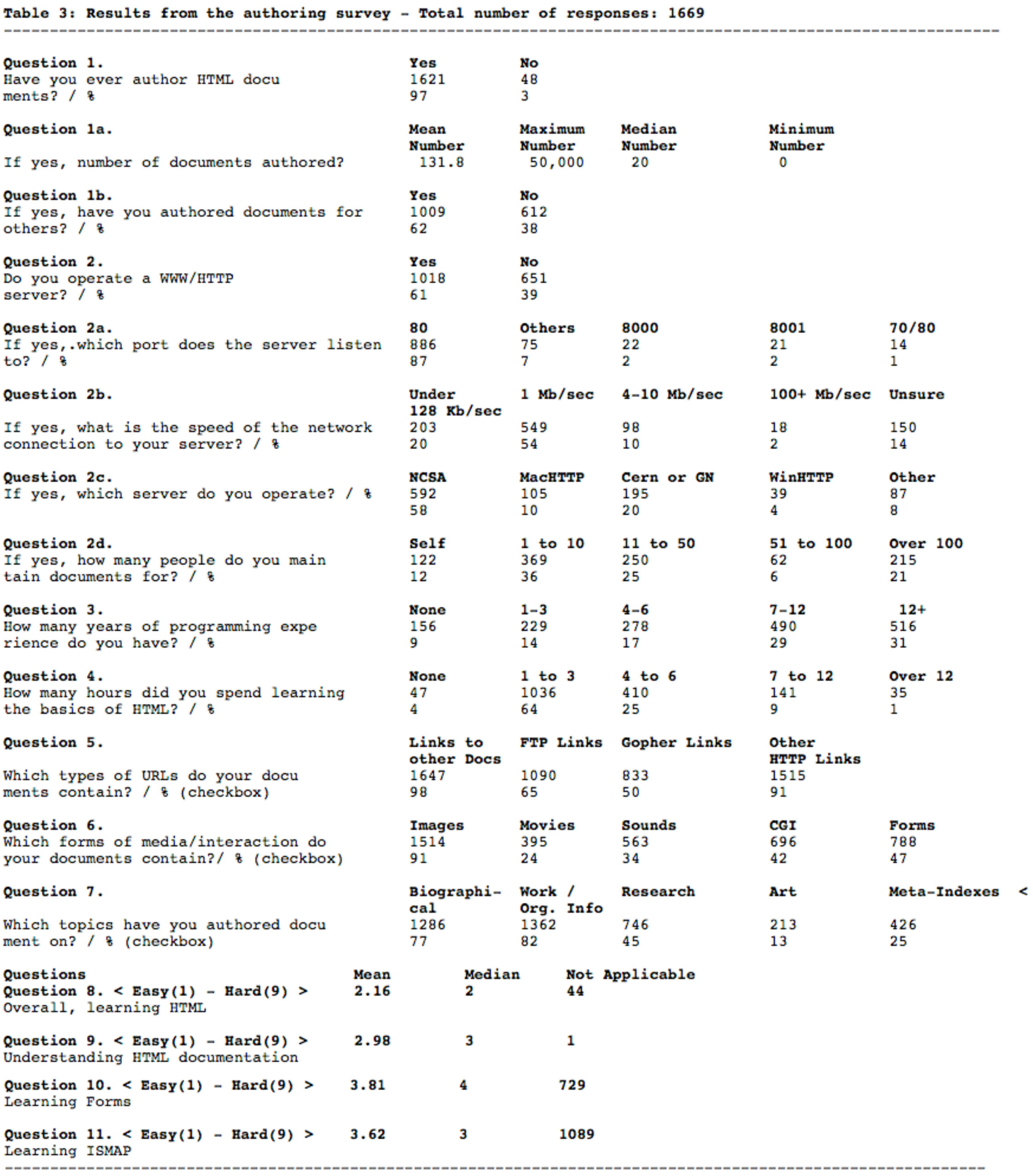


Table 3: Results from the authoring survey - Total number of responses: 1669

Question 1.	**Yes**	**No**			
Have you ever author HTML docu	1621	48			
ments? / %	97	3			
Question 1a.	**Mean Number**	**Maximum Number**	**Median Number**	**Minimum Number**	
If yes, number of documents authored?	131.8	50,000	20	0	
Question 1b.	**Yes**	**No**			
If yes, have you authored documents for	1009	612			
others? / %	62	38			
Question 2.	**Yes**	**No**			
Do you operate a WWW/HTTP	1018	651			
server? / %	61	39			
Question 2a.	**80**	**Others**	**8000**	**8001**	**70/80**
If yes,.which port does the server listen	886	75	22	21	14
to? / %	87	7	2	2	1
Question 2b.	**Under 128 Kb/sec**	**1 Mb/sec**	**4-10 Mb/sec**	**100+ Mb/sec**	**Unsure**
If yes, what is the speed of the network	203	549	98	18	150
connection to your server? / %	20	54	10	2	14
Question 2c.	**NCSA**	**MacHTTP**	**Cern or GN**	**WinHTTP**	**Other**
If yes, which server do you operate? / %	592	105	195	39	87
	58	10	20	4	8
Question 2d.	**Self**	**1 to 10**	**11 to 50**	**51 to 100**	**Over 100**
If yes, how many people do you main	122	369	250	62	215
tain documents for? / %	12	36	25	6	21
Question 3.	**None**	**1-3**	**4-6**	**7-12**	**12+**
How many years of programming expe	156	229	278	490	516
rience do you have? / %	9	14	17	29	31
Question 4.	**None**	**1 to 3**	**4 to 6**	**7 to 12**	**Over 12**
How many hours did you spend learning	47	1036	410	141	35
the basics of HTML? / %	4	64	25	9	1
Question 5.	**Links to other Docs**	**FTP Links**	**Gopher Links**	**Other HTTP Links**	
Which types of URLs do your docu	1647	1090	833	1515	
ments contain? / % (checkbox)	98	65	50	91	
Question 6.	**Images**	**Movies**	**Sounds**	**CGI**	**Forms**
Which forms of media/interaction do	1514	395	563	696	788
your documents contain?/ % (checkbox)	91	24	34	42	47
Question 7.	**Biographi-cal**	**Work / Org. Info**	**Research**	**Art**	**Meta-Indexes <**
Which topics have you authored docu	1286	1362	746	213	426
ment on? / % (checkbox)	77	82	45	13	25

Questions	**Mean**	**Median**	**Not Applicable**
Question 8. < Easy(1) - Hard(9) >	**2.16**	**2**	**44**
Overall, learning HTML			
Question 9. < Easy(1) - Hard(9) >	**2.98**	**3**	**1**
Understanding HTML documentation			
Question 10. < Easy(1) - Hard(9) >	**3.81**	**4**	**729**
Learning Forms			
Question 11. < Easy(1) - Hard(9) >	**3.62**	**3**	**1089**
Learning ISMAP			

Footnotes

[1] Essentially, the GET method appends the data being passed from the client to the server onto the URL, where as the POST method passes the data to the server without altering the URL of the requested document/program.

[2] In tandem with dialogs on www-talk, we concluded 1) use GET for logical independent tasks/retrieval and 2) currently POSTs should not be hotlisted/cached,

though this is doable via restructuring the representation of hotlist objects by the client.

(3)

This scheme does not handle certain organizations who own multiple ips within the same class, like Georgia Tech.

(4)

A version of the all questionnaires with adaptation are available via: http://www.cc.gatech.edu/gvu/user_surveys/survey-09-1994/.

(5)

All datasets are publicly available via ftp://ftp.cc.gatech.edu/gvu/www/survey/survey-09-1994/datasets and the above URL (footnote 4).