

Emerging Trends in the WWW User Population

VAST amounts of attention and resources have recently been devoted toward the World-Wide Web (WWW) [1], but relatively little research has been conducted examining Web usage and societal implications. With the goals of understanding the Web user population and promoting the Web as a viable surveying medium, the WWW User Surveys were initially conducted by Georgia Institute of Technology's Graphics, Visualization, and Usability Center during January 1994. Subsequent surveys have been administered approximately every six months thereafter. Each survey is conducted for one month using the limited interactivity of the Web, where users point and click on responses within their Web browsers and submit results to a centralized server for processing.

The first survey [6] was administered during January 1994 and received over 1,500 responses, which was a considerable amount at that time. This response rate, along with tremendous positive feedback from the Web community, justified continuing the surveys. The second survey (October 1994) [7] employed an extended and refined question base, which included a set of questions developed by the University of Michigan's Hermes Team regarding consumer attitudes toward electronic commerce. The response rate continued to grow significantly, recording over 4,500 unique users. This tremendous growth has continued through the third and fourth surveys (April and October 1995) [4, 5], with 13,000 and 23,300 users responding, respectively. Based upon current estimates, the last two surveys were completed by nearly one out of every thousand Web users [2, 3]. We expect this trend to continue for the fifth survey, the results of which will be available in mid-June 1996.

Walk-Through of the Survey Interaction

Essentially, the respondents are led through a series of *question-answer-adapt/re-ask cycles* (see [6] for more details on survey execution and architecture). Upon selection of a questionnaire the surveying software generates the default set of questions from the question database. Users respond to the questions displayed via their WWW browser by selecting options presented in the form of radio buttons, pull-down menus, scrolling lists, and check boxes. (The surveys intentionally avoid the use of open-ended text entry, as this increases the complexity of response processing.) When the user is done answering a set of questions, the survey software *adapts*, generating a set of follow-up questions based on the user's answers. When the questionnaire is completed, the user

is given a list of the remaining questionnaires from which to select.

Limitations of the Results and Methodology

Widely distributed, heterogeneous, electronic surveying is an exciting new field, especially with respect to the Web. The adaptive WWW-based surveying techniques are pioneering and as such, they require conservative interpretation of collected data due to the absence of tested validation metrics.

The surveys suffer from two basic problems: self-selection and sampling. When people decide to participate in a survey, they select themselves. This decision may reflect

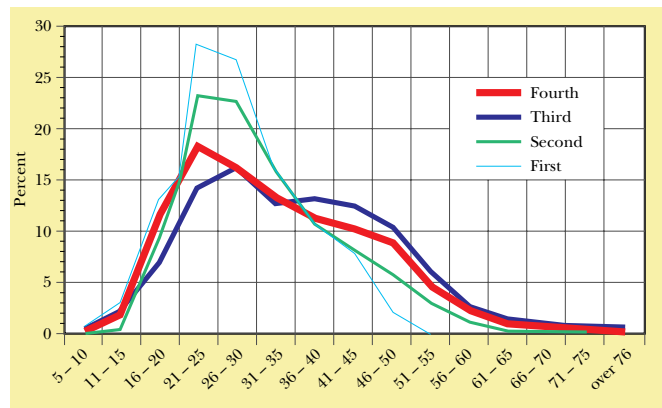


Figure 1. Age distributions across the four surveys

some systematic selecting principle (or judgment) that affects the collected data. Nearly all surveys, electronic or otherwise, suffer from self-selection problems. For example, when a potential respondent hangs up on a telephone-based surveyor or does not return a completed direct-mail survey, self-selection has occurred.

The other issue is sampling. There are essentially two types of sampling: random and non-random. Random selection is intended to minimize bias and make the sample as typical of the population as possible. To accomplish this, steps need to be taken to get respondents randomly (e.g., drawing numbers out of a hat). The surveys use a form of non-random sampling that relies on users' awareness of the surveys for participation. As a result, all portions of the user population may not be represented in the sample. This reduces the ability of the gathered data to generalize to the entire Web user population.

Until recently, no random-sample surveys had been publicly released which could be used to compare the WWW User Surveys. However, in the fall of 1995, the CommerceNet and O'Reilly surveys [2, 3] were published. Since these employed traditional surveying techniques, their results provide an important benchmark for results from non-traditional surveying. Surprisingly, an initial analysis between these random surveys and the WWW User Surveys reveals few differences between major characteristics (e.g., gender, age, income). The main differences were found in frequency of Web use and skill level in using the Web. Thus, while the WWW User Surveys do attract heavier and more experienced users than random phone-based surveys, it does not appear that these are differentiating characteristics within the population. Unfortunately, no international random-sample survey has taken place, so the biases and corrective metrics necessary are still undetermined.

Results and Trends

Throughout the four surveys, substantial shifts in the characteristics of Web users have occurred. While certain attributes of the users sampled in the surveys have remained the same or changed slightly, others have changed dramatically. More than ever, the users in the latest survey (the fourth) represent less and less the *technology developers/pioneers* found in the first survey (average age is early 20s, computer-savvy users) and more of what we refer to as *early adopters/seekers of technology*. The adopters do not typically have access to the Web through work or school, but actively seek out local or major Internet access providers. As the Web continues to expand its horizon of users, we expect, and indeed find, that more users from diverse segments of the population are participating in the surveys.

The remainder of this report will focus on changes in user demographics, content providers, and usage patterns. The reader is referred to [4] for complete coverage of the results. Where not explicitly indicated, figures refer to the most recent (fourth) survey results.

■ How Has the Age Distribution Changed?

As shown in Figure 1, which depicts the age profiles across surveys, more older and younger users have participated with each survey. The average age across all users in the fourth survey (32.7 years old) was down two years from the third survey (35.0 years old). This downward shift brings the average age of the Web users sampled closer to the average measured in the second survey (31.7 years old). The fourth survey shows a significant change in age distributions between genders: there are more female users between the ages of 16 and 20 years old (14.2% female vs. 11.5% male) and fewer above the age of 46 years old (13.82% female vs. 17.87% male). This trend toward younger females is consistent with the observed increase in female users in college and K-12 education occupations.

■ How has the Gender Ratio Changed?

In the fourth survey, 29.3% of the respondents were female. Compared to the third survey, women represent a 15% increase and men a 12% decrease (the third survey had a *Rather not Say!* option). This is nearly dou-

ble the growth rate of female users observed across the first three surveys (a 5% increase per survey) as shown in Figure 2. In the U.S., 32.5% of the users were female, 67.5% male. This represents a strong shift in the increased acceptance and use of the Web by women. The U.S. continues to integrate female users into the Web user population at a faster rate than Europe. However, the gender ratio of U.S. users is still far from the 52% female, 48% male composition of the entire U.S. population (1995 Estimated U.S. Census).

■ What is the Average and Median Income?

The estimated average income across all users in the

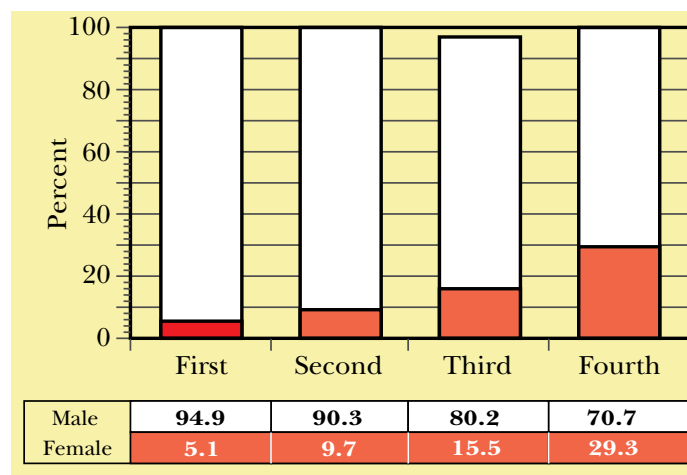


Figure 2. Gender distributions across the four surveys. The Third Survey included a "Rather Not Say!" option, which is why the genders do not sum to 100%.

fourth survey was \$63,000 (U.S. dollars), which is slightly lower than the third survey, which had an estimated income of \$69,000. The median income was in the range of \$50,000 and \$60,000, which is the same as the third survey. The estimated average U.S. income was \$64,700 which is lower than the third survey's \$67,600. Thus, we see that as more U.S. users are joining the Web, fewer are ultra-elite computer-users, but still the estimated median income is well above the national median—\$36,950 as estimated by the 1993 U.S. Census.

■ How Willing Are Users to Pay for Access to Web Sites?

One of the most stable characteristics of previous surveys has been that one of five users stated outright that they would not pay for access to WWW sites. For the fourth survey, this segment of the population increased from 22.6% in the third survey to 31.8%. This is indeed alarming for those wishing to apply a subscription business model to the Web. This may also very well reflect the user's perceived value of the material and resources currently available on the Web. Additionally, this may place the financial burden upon sites wishing to generate revenue from advertising rather than user subscriptions. Most users reported that their willingness would depend on both the cost of access as well as the quality of the material provided (58.7%).

■ How Often Do People Use their Web Browsers?

The majority of users who participated in the fourth sur-

vey (78.4%) reported using their browsers daily. Users in Europe spent slightly less time using their browsers than users in the U.S., which is the reverse of the third survey. 65.8% of European users spent 6 hours or less per week compared to only 59.4% of U.S. users. There also seems to be evidence that those who use computers for work are less likely to spend casual and recreational time computing. One hypothesis is that users are only willing to spend a fixed amount of time per week in front of a computer. More research into this issue is warranted before any conclusions can be drawn.

■ What Do People Do with their Web Browsers?

The most common use of Web browsers reported in the fourth survey was simply for browsing (79.0%) followed by entertainment (63.6%) and work (51.8%). The category with the least number of responses was *other uses* (10.8%) followed closely by shopping (11.1%). The low response rate for shopping was also found in the other surveys, where 10.0% (in the third survey) and 8.0% (in the second survey) reported using the Web for this purpose. Over 60% of the users cited security concerns as the primary reason for not buying merchandise on the Web. This profile of use has remained fairly consistent since the second survey, where browsing and entertainment dominated. Hence, while the Web is often used for academic and business purposes, the main uses are still recreational.

■ How Easy Was it for People to Learn HTML?

Just over half of the respondents to the fourth survey who have learned HTML mastered the basics in under 3 hours, with 79.4% learning in under 6 hours. In the second survey, 62% of the users learned HTML in under 3 hours and in the third survey, the number dropped to 55.2%—the trend is clearly toward users taking longer to learn HTML. Yet, taking into account the shifts in the demographics towards less experienced computer users, this increase is extremely small. The ease with which the Web enables many people to provide content remains one of its primary strengths.

■ What Are the Main Problems with Using the Web?

This was a new question in the fourth survey. The most widely cited problem was that it takes too long to view/download pages (69.1%). This problem is supported by the finding that 60.3% of the users report using either 14.4Kbs or 28.8Kbs modems. The other problem areas identified by users were: not being able to find a page that they know exists (34.5%), not being able to organize the pages and information they gather (25.8%) and not being able to find a page once visited (23.7%). Additionally, 14.3% of the users reported that not being able to visualize where they have been and where they can go was a problem. Surprisingly, most users did not report that being able to determine where they are (the classic *lost in hypertext* problem) was problematic (6.5%). Only 3.2% reported that their Web browsers were either poorly designed or did not work well. Thus, some of the traditional problems of hypertext do not seem to be as pertinent for the Web users surveyed.

■ Why Do People Save Documents on the Web?

The most widely cited reason for saving documents was to use the information offline (59.7%). Other respons-

es were: reading the document offline (50.9%), distribution to others not online (45.2%), and archiving the content (30.9%). These findings correlate fundamentally to the types and regularity of information being accessed, where people are most likely to read and use the content offline when it is reference, news, or product information. Saving documents in fear that the item would no longer be available was only reported by 18.8% of the users.

Conclusion

Clearly, today's Web is not the same Web of January 1994. The direction provided by the National and Global Information Infrastructure focus combined with easily acquired interfaces to the Web has left its trail across the surveys. The surveyed Web user populations have rapidly changed from the originators of the technology to the initial users in the educational and research settings to the users who are provided with Web access at work and school to those who actively seek out Web connectivity. The WWW User Surveys are able to keep pace with this fluidity by identifying and quantifying real changes in the adoption of what may very well be the most important technological revolution since Gutenberg. ■

Acknowledgments

Georgia Tech's Graphics, Visualization, and Usability (GVU) Center conducts the surveys as a public service as part of its commitment toward the Web and Internet communities. Thanks to all members of the GVU, its director Jim Foley, and staff for their help and support.

References

1. Berners-Lee, T., Cailliau, R., Luotonen, A., Nielsen, H., and Secret, A. The World-Wide Web. *Commun. ACM* 37, 8 (Aug. 1993), 76–82.
2. CommerceNet/Nielsen. *Internet Demographics Survey*, 1995. URL: <http://www.commerce.net/information/surveys/>.
3. O'Reilly & Associates/Trish Information Services. *Defining the Internet Opportunity*, 1995. URL: <http://www.ora.com/gnn/bus/ora/survey/index.html>
4. Pitkow, J. and Kehoe, C. *Results from the Fourth World-Wide Web User Survey*, 1995. URL: http://www.cc.gatech.edu/gvu/user_surveys/survey-10-1995.
5. Pitkow, J. and Kehoe, C. Results from the Third World-Wide Web User Survey. *The WWW Journal*, 1, 1 (1995).
6. Pitkow, J. and Recker, M. Results from the First World-Wide Web Survey. *J. of Computer Networks and ISDN Syst.* 27, 2 (1994).
7. Pitkow, J. and Recker, M. Using the Web as a survey tool: Results from the Second World-Wide Web User Survey. *J. of Computer Networks and ISDN Syst.* 27, 6 (1995).

About the Authors:

JAMES PITKOW is a Ph.D. candidate at the Graphics, Visualization, & Usability (GVU) Center at Georgia Institute of Technology.

COLLEEN KEHOE is a Ph.D. candidate at the Graphics, Visualization, & Usability (GVU) Center at Georgia Institute of Technology.

Author's Present Address: Graphics, Visualization and Usability (GVU) Center, Georgia Institute of Technology, Atlanta GA 30332-0280. email: {pitkow, colleen}@cc.gatech.edu

This material is based upon work supported under a National Science Foundation Graduate Research Fellowship.

Permission to make digital/hard copy of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage, the copyright notice, the title of the publication and its date appear, and notice is given that copying is by permission of ACM, Inc. To copy otherwise, to republish, to post on servers, or to redistribute to lists requires prior specific permission and/or a fee.

© ACM 0002-0782/96/0600 \$3.50