Emerging Trends in the WWW User Population

James E. Pitkow & Colleen M. Kehoe

Graphics, Visualization, & Usability (GVU) Center
Georgia Institute of Technology, Atlanta GA 30332-0280
{pitkow, colleen}@cc.gatech.edu

Background

Vast amounts of attention and resources have recently been devoted towards the World Wide Web (WWW) [Berners-Lee 94], but relatively little research has been conducted that examines Web usage and societal implications. With the goals of understanding the Web user population and promoting the Web as a viable surveying medium, GVU's WWW User Surveys were initially conducted during January 1994. Subsequent surveys were administered approximately every six months thereafter. The surveys employ non-random sampling techniques, which limit the ability of the results to generalize to the entire Web population. Each survey is conducted 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. Each survey is conducted for a one month period.

The First Survey [Pitkow & Recker 1994] was administered during January 1994 and received over 1,500 respondents, 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) [Pitkow & Recker 1995] employed an extended and refined question base which included a set of question regarding consumer attitudes towards electronic commerce developed by the University of Michigan's Hermes Team. The response rate continued to grow significantly, recording over 4,500 unique users. This tremendous growth continued through the Third and Fourth Surveys (April and October 1995) [Pitkow & Kehoe 1995a; Pitkow & Kehoe 1995b], 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 [CommerceNet 1995; O'Reilly 1995].

Walk-Through of the Survey Interaction

In order to convey the sense of interaction present while completing the surveys, a quick walk-through follows (see [Pitkow & Recker 94] for more details on survey execution and architecture). Essentially, the respondents are led through a series of *"question-answer-adapt/re-ask"* cycles. Upon selection of a questionnaire from the Main Launching Page that provides access to all the questionnaires, the surveying software generates the default set of questions from the question database. The user then responds to the questions displayed via their WWW browser by selecting options presented via radio buttons, pulldown 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.

Limitations of the Results & Methodology

Highly 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.

Basically, the surveys suffer two problems: self-selection and sampling. When people decide to participate in a survey, they select themselves. This decision may reflect some systematic selecting principle (or judgment) that affects the collected data. Just about all surveys, electronic or otherwise, suffer from self-selection problems. For example, when a potential respondent hangs up on a telephone based surveyor, self-selection has occurred. Likewise, when a potential respondent does not send back a direct mail survey, self-selection has occurred.

The other issue is sampling. There are essentially two types of sampling: random and non-probabilistic.

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 in a random manner, e.g., drawing numbers out of a hat. The surveys use a form of non-probabilistic sampling which 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 [CommerceNet 1995; O'Reilly 1995] 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 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 (Fourth) represent less and less the "technology developers/ pioneers" found in the First Survey (early 20's, computer savvy users) and more of what we refer to as the "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 and more users

from diverse segments of the population are participating in the surveys.

The remainder of this paper will focus on changes in user demographics, content providers, and usage patterns. The reader is referred to [Pitkow & Kehoe 1995a] 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.

[insert Figure 1 around here]

How has the gender ratio changed over time?

In the Fourth Survey, 29.3% of the respondents were female. Compared to the last survey, women represent a 15% increase and men a 12% decrease (the last survey had a "Rather not Say!" option). This is nearly double the growth rate of women users observed across the first three surveys (5% increase per survey) as shown in Figure 2. In the US, 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 US continues to integrate female users into the Web user population at a greater rate than Europe. However, the gender ratio of US

users is still far from the 52% female, 48% male composition of the entire US population (1995 Estimated US Census). As noted above, the increase in female users occurred largely in college students and K-12 educators.

[insert Figure 2 around here]

What is the average and median income?

The estimated average income across all users in the Fourth Survey was $63,000 (US 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 US income was $64,700 which is lower than the Third Survey of $67,600. Thus, we see that as more US users are joining the Web, fewer are ultra-elite computer-users, but still the estimated median income is quite above the national median—$36,950 as reported by the 1993 US 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 to 31.8% from 22.6% in the Third Survey. This is indeed alarming for those wishing to apply a subscription business model to the Web. This may also very well reflect the perceived value of the material and resources currently available on the Web by its users. Additionally, this may place the financial burden upon sites wishing to generate monies on 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 browser?

The majority of users who participated in the Fourth Survey (78.4%) reported using their browsers daily. Users in Europe spent slightly less time using their browsers than user in the US, 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 US 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 browser?

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 lack of use of the Web for shopping was also found in the other surveys, where 10.0% (Third) and 8.0% (Second) 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 towards 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 14.4 or 28.8 kbs modems. The other problem areas identified by users were: not being able to find a page that they know is out there (34.5%), not being able to organize the pages & 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, users did not report that being able to determine where they are (the classical "lost in hypertext" problem) as being problematic (6.5%). Only 3.2% reported that their Web browser was either poorly designed or did not work well as being a problem. 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 in the document offline (59.7%). Other responses were: reading the document offline (50.9%), distribute 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.

Conclusions

Clearly, today's Web is not the same Web of January 1994. The infusion of 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 flowed from the originators of the technology to the initial users in the educational and research settings to the users who are provided connectivity at work and school to those who actively seek out Web connectivity. The WWW User Surveys are able to keep pace with the fluidity by identifying and quantifying real changes in the adaptation of what may very well be the most important technological revolution since Guttenberg.

References

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.

CommerceNet/Nielsen. (1995). Internet Demographics Survey. <*URL:http://www.commerce.net/information/surveys/*>.

O'Reilly & Associates/Trish Information Services. (1995) Defining the Internet Opportunity. <*URL:http://www.ora.com/gnn/bus/ora/survey/index.html*>.

Pitkow, J. and Kehoe, C. (1995a). Results from the Fourth World Wide Web User Survey. <*URL:http//www.cc.gatech.edu/gvu/user_surveys/survey-10-1995*>.

Pitkow, J. and Kehoe, C. (1995b). Results from the Third World Wide Web User Survey. *The WWW Journal*, Vol. 1, no. 1.

Pitkow, J. and Recker, M. (1994). Results from the First World-Wide Web Survey. *Journal of Computer Networks and ISDN systems*, Vol. 27, no. 2.

Pitkow, J. and Recker, M. (1995). Using the Web as a Survey Tool: Results from the Second World-Wide Web User Survey. *Journal of Computer Networks and ISDN systems*, Vol. 27, no. 6.

Acknowledgments

Georgia Tech's Graphics, Visualization, & Usability (GVU) Center operates the surveys as a pubic service as part of its commitment towards the Web and Internet communities. This material is based upon work supported under a National Science Foundation Graduate Research Fellowship. Thanks to all members of the GVU, its director Dr. Jim Foley, and staff for their help and support.

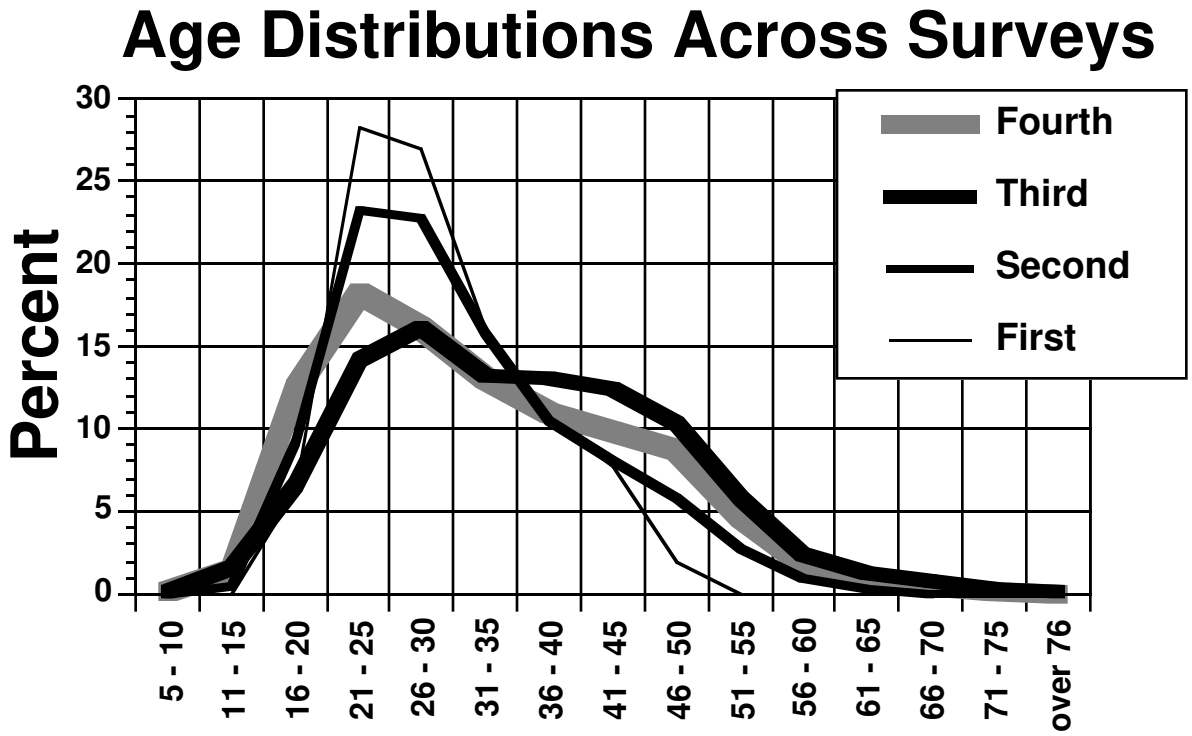
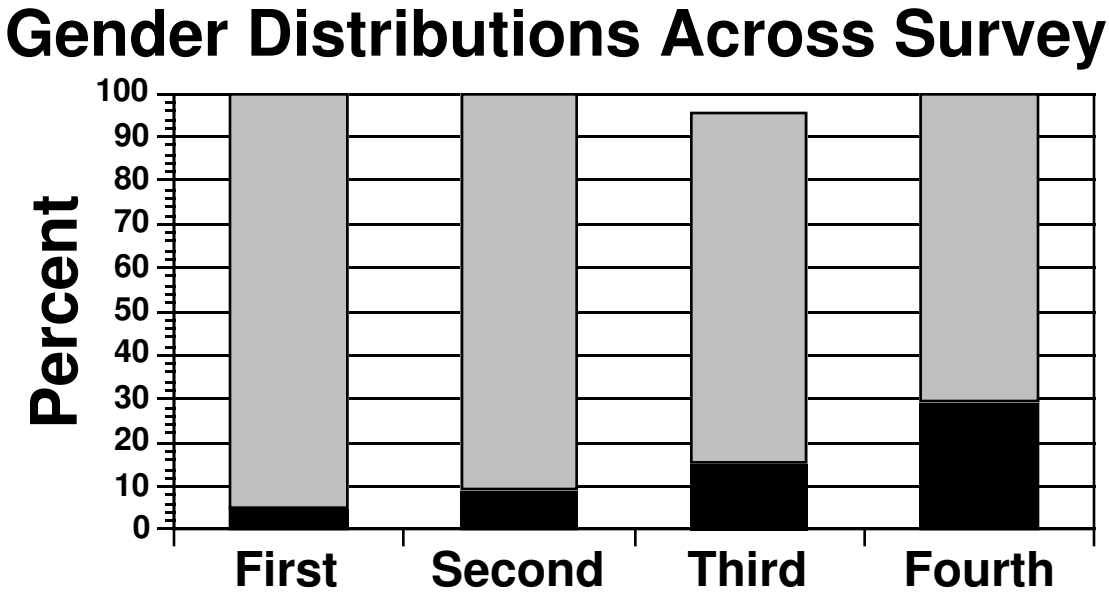


Figure 1. Age distributions across the four Surveys.



Female	**5.1**	**9.7**	**15.5**	**29.3**
Male	**94.9**	**90.3**	**80.2**	**70.7**

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%.