Know. of CGI by CGI Script Use Question 11 by 12	155.007	2	0.000	0.582
Know. of CGI by ISMAP Question 11 by 13	72.535	4	0.000	0.281
Know. of CGI by ISMAP Use Question 11 by 14	40.937	2	0.000	0.296
CGI Use by ISMAP Question 12 by 13	71.809	2	0.000	0.398
CGI Use by ISMAP Use Question 12 by 14	63.903	2	0.000	0.350
ISMAP by ISMAP Use Question 13 by 14	277.407	2	0.000	0.774

Table 12: Results of Chi-Square for HTTP Survey

Interaction	Chi-Square Value	Degrees of Freedom	Probability	Cramer V
Use of Help by On-Line Docs Helpful Question 4 by 5	124.135	6	0.000	0.244

Table 13: Results of Chi-Square for Mosaic Survey

Interaction	Chi-Square Value	Degrees of Freedom	Probability	Cramer V
Frequency of Use by Orientation Question 1 by 19	21.060	10	0.031	0.097
Frequency of Use by Hours Exploring Question 1 by 20	312.501	10	0.000	0.370
Orientation by Hours Exploring Question 19 by 20	23.518	4	0.000	0.103

Table 14: Results of Chi-Square for Usage Survey

Operate by ISMAP Use Question 1 by 14	24.838	1	0.000	0.224
Server by Knowledge of Server Script Question 2 by 9	41.596	6	0.000	0.236
Server by Knowledge of Module Question 2 by 10	145.472	6	0.000	0.444
Server by Knowledge of CGI Script Question 2 by 11	31.442	6	0.000	0.206
Server by CGI Script Use Question 2 by 12	23.121	3	0.000	0.252
Server by Knowledge of ISMAP Question 2 by 13	10.919	6	0.091	not significant
Server by ISMAP Use Question 2 by 14	8.859	3	0.031	0.155
Server Script by Knowledge of Module Question 9 by 10	53.810	4	0.000	0.239
Server Script by CGI Script Question 9 by 11	150.649	4	0.000	0.401
Server Script by CGI Script Use Question 9 by 12	112.718	2	0.000	0.494
Server Script by ISMAP Question 9 by 13	97.941	4	0.000	0.325
Server Script by ISMAP Use Question 9 by 14	54.503	2	0.000	0.340
Know. of Module by CGI Script Question 10 by 11	47.495	4	0.000	0.226
Know. of Module by CGI Script Use Question 10 by 12	31.997	2	0.000	0.264
Know. of Module by ISMAP Question 10 by 13	46.291	4	0.000	0.224
Know. of Module by ISMAP Use Question 10 by 14	24.985	2	0.000	0.231

Table 12: Results of Chi-Square for HTTP Survey

Interaction	Chi-Square Value	Degrees of Freedom	Probability	Cramer V
Documents Authored by ISO Standard Question 1 by 2	61.080	8	0.000	0.210
Documents Authored by Years Program. Question 1 by 3	30.766	16	0.014	0.105 *
Documents Authored by Languages Question 1 by 4	30.766	16	0.014	0.105 *
Standard by Years Programming Question 2 by 3	19.541	8	0.012	0.119
Standard by Languages Question 2 by 4	19.722	8	0.011	0.119
Years Programming by Languages Question 3 by 4	733.663	16	0.000	0.514

Table 11: Results of Chi-Square for HTML Survey

* Denotes a negative Kendall Tau-B result

Interaction	Chi-Square Value	Degrees of Freedom	Probability	Cramer V
Operate by Access to WWW Database Question 1 by 4	27.185	1	0.000	0.388
Operate by Knowledge of Server Script Question 1 by 9	69.383	2	0.000	0.383
Operate by Knowledge of Module Question 1 by 10	7.643	2	0.22	0.128
Operate by Knowledge of CGI Script Question 1 by 11	62.132	2	0.000	0.365
Operate by CGI Script Use Question 1 by 12	50.759	1	0.000	0.316
Operate by Knowledge of ISMAP Question 1 by 13	44.283	2	0.000	0.309

Table 12: Results of Chi-Square for HTTP Survey

APPENDIX TWO

Interaction	Chi-Square Value	Degrees of Freedom	Probability	Cramer V
Machine Shared by Hrs in Front of Cptr Question 3 by 4	43.834	6	0.000	0.183 *
Machine Shared by Domain Question 3 by 5	139.779	6	0.000	0.324
Machine Shared by Location Question 3 by 6	24.233	4	0.000	0.135 *
Machine Shared by Occupation Question 3 by 7	199.290	6	0.000	0.388
Machine Shared by Willing to Pay Fees Question 3 by 8	30.632	2	0.000	0.152
Machine Shared by Age Question 3 by 9	122.729	6	0.000	0.306 *
Hrs in Front of Computer by Occupation Question 4 by 8	169.460	36	0.000	0.147 **
Hours in Front of Computer by Fees Question 4 by 8	23.295	12	0.000	0.094 **
Hours in Front of Computer by Age Question 4 by 9	90.139	30	0.000	0.118
Occupation by Fees Question 7 by 8	52.816	12	0.000	0.141
Occupation by Age Question 7 by 9	835.564	36	0.000	0.326 **
Willing to Pay Fees by Age Question 8 by 9	26.628	12	0.009	0.101 *

Table 10: Results of Chi-Square for General Survey

NOTES: Since there were only two responses for the under 15 category in Question 9 (Age), the responses were not included into the Chi-Square calculations.

* Denotes a negative Kendall Tau-B result

** WARNING: More than one-fifth of the cells were sparse (Frequency < 5) significance suspect

Question 1.	Over 9/ day	5 to 8/ day	1 to 4 / day	Few times / week	Once / week	Once/ month	Total
Frequency of Browser Use / %	235 20.38	202 17.52	481 41.72	199 17.26	23 1.99	9 0.78	1153
Question 19.	Search	Text	Visual				Total
Search Orientation / %	319	432	372				1344
Question 20.	Under 5	6 to 10	Over 10				Total
Hours Exploring the Internet / %	484 41.98	407 35.30	252 21.86				1153

Table 9: Results from the Usage Survey

Question	Mean	Std. Dev.	Total
Question 3. < never (1) - always (9) > Use Browser to Retrieve Documents	5.81	3.39	1145
Question 4. < never (1) - always (9) > Use Browser Instead of Gopher	7.42	3.96	1048
Question 5. < never (1) - always (9) > Use Browser Instead of Wais	7.60	5.44	1007
Question 6. < never (1) - always (9) > Use Browser Instead of Archie	5.65	8.17	1026
Question 7. < never (1) - always (9) > Use Browser to View Research	4.20	5.15	1116
Question 8. < never (1) - always (9) > Use Browser to Explore Internet	7.30	2.23	1138
Question 9. < never (1) - always (9) > Use Browser to Retrieve Weather Information	4.10	6.32	1139
Question 10. < never (1) - always (9) > Use Browser to Access Reference Materials	4.97	4.41	1127
Question 11. < never (1) - always (9) > Use Browser to Access Publishers, Periodicals, etc.	4.16	4.72	1116
Question 12. < never (1) - always (9) > Use Browser to Access Call for Papers, Proposals, etc.	4.33	5.13	1122
Question 13. < love (1) - hate (9) > User ‘s Preference of Text	6.62	2.57	1140
Question 14. < love (1) - hate (9) > User ‘s Preference of Embedded Images	7.30	2.31	1140
Question 15. < love (1) - hate (9) > User ‘s Preference of Spawned Images	6.48	3.42	1124
Question 16. < love (1) - hate (9) > User ‘s Preference of Sounds	5.41	4.33	1089
Question 17. < love (1) - hate (9) > User ‘s Preference of Movies	5.83	4.19	1097
Question 18. < love (1) - hate (9) > User ‘s Preference of Search by Keyword	7.93	1.87	1130

Table 8: Results from the Usage Survey

Question	Mean	Std. Dev.	Total
Question 5. < horrid (1) - excellent (9) > Overall Rating of server	5.815	1.854	292
Question 6. < horrid (1) - excellent (9) > Overall Rating of GN's server	6.041	1.958	98
Question 7. < horrid (1) - excellent (9) > Overall Rating of NCSA's server	7.453	1.239	750
Question 8. < horrid (1) - excellent (9) > Overall Rating of Plexus's server	6.788	1.686	170

Table 5: Results from the HTTP Survey

Question 1.	UNIX	PC	Mac			Total
Which Mosaic Platform / %	1006 / 93.23	48 / 4.45	25 / 2.32			1079
Question 4.	Daily	Weekly	Monthly	Never		Total
Freq of Use of Help Menu / %	89 / 8.26	306 / 28.41	326 / 30.27	356 / 33.05		1077
Question 5.	Yes	No	Some			Total
Is On-line Help Useful / %	551 / 52.33	74 / 7.03	428 / 40.65			1053

Table 6: Results from the Mosaic Survey

Question	Mean	Std. Dev.	Total
Question 2. < horrid (1) - excellent (9) > Overall Rating of Mosaic	8.086	0.881	1081
Question 3. < horrid (1) - excellent (9) > Rating of Support Staff @ NCSA	7.458	1.520	334

Table 7: Results from the Mosaic Survey

Question 1.	Yes	No			Total
Operate HTTP Server / %	378 / 78.59	103 / 21.41			481
Question 2.	GN or Cern	NSCA	Plexus	Other	Total
If Yes, Which Server / %	28 / 7.43	271 / 71.88	48 / 12.73	30 / 7.96	377
Question 3.	80	8000	8001	8080	Total
Server Port Number / %	279 / 78.81	10 / 2.82	13 / 3.67	16 / 4.52	354
Question 4.	Yes	No			Total
If no, Able to Add to DB / %	122 / 79.22	32 / 20.78			154
Question 9.	Yes	No	Some		Total
Knowledge of Script / %	260 / 58.82	66 / 13.87	130 / 27.31		476
Question 10.	Yes	No	Some		Total
Knowledge of Modules / %	73 / 15.50	341 / 72.40	57 / 12.10		471
Question 11.	Yes	No	Some		Total
Knowledge of CGI / %	190 / 40.60	191 / 40.81	87 / 18.59		471
Question 12.	Yes	No			Total
Written CGI scripts / %	177 / 38.31	285 / 61.69			462
Question 13.	Yes	No	Some		Total
Knowledge of ISMAP/ %	138 / 29.61	217 / 46.57	111 / 23.82		466
Question 14.	Yes	No			Total
Used ISMAP scripts / %	128 / 27.12	344 / 72.88			472

Table 4: Results from the HTTP Survey

Question 1.	None	1 to 5	6 to 10	11 to 20	Over 20	Total
Num Docs Authored / %	52 / 7.42	189 / 26.96	102 / 14.55	86 / 12.27	270 / 38.52	701
Question 2.	Yes	No	Unsure			Total
Is HTML ISO Standard / %	21 / 3.00	211 / 30.10	463 / 66.05			701
Question 3.	None	1 to 3	4 to 6	7 to 12	Over 12	Total
Years of Programming / %	21 / 3.00	55 / 7.85	145 / 20.68	233 / 33.24	240 / 34.24	701
Question 3.	None	1 to 3	4 to 6	7 to 12	Over 12	Total
Num Programming Langs / %	16 / 2.28	111 / 15.83	292 / 41.65	172 / 24.11	106 / 15.12	701
Question 12.	Yes	No				Total
Prior SGML Use / %	31 / 8.52	333 / 91.48				364

Table 2: Results from the HTML Survey

Question	Mean	Std. Dev.	Total
Question 5. < 1 - 9 > Hours taken to learn HTML	2.258	1.377	654
Question 6. < easy (1) - hard (9) > Overall, learning of HTML	2.132	1.393	651
Question 7. < easy (1) - hard (9) > Learning Forms	3.458	1.749	295
Question 8. < easy (1) - hard (9) > Learning ISMAP	3.781	2.072	201
Question 10. < easy (1) - hard (9) > Finding up to the minute HTML documentation	3.951	2.360	551
Question 11. < easy (1) - hard (9) > Understanding of HTML documentation	3.064	1.681	626

Table 3: Results from the HTML Survey

APPENDIX ONE

Question 1.	Mac	PC	Unix	VMS	Other			Total
Primary Platform / %	48 / 3.60	66 / 4.94	1181 / 88.46	33 / 2.47	7 / 0.01			1335
Question 2.	Cello	Lynx	Mosaic	Other	Samba			Total
Primary Browser / %	1 / 0.08	26 / 1.95	1294 / 97.22	6 / 0.38	4 / 0.30			1331
Question 3.	No	Yes						Total
Shared Machine/ %	596 / 44.74	736 / 55.26						1332
Question 4.	Under 5	6 to 10	11 to 20	21 to 30	31 to 40	41 to 50	50 +	Total
Hrs in Front of Computer / %	2 0.15	40 7.59	102 7.59	245 18.23	295 21.95	273 2.98	360 26.79	1344
Question 6.	Asia	Australia	Europe	North Amer.	South Amer.			Total
Location / %	10 / 0.75	27 / 2.04	377 / 28.43	910 / 68.63	2 / 0.15			1326
Question 7.	UGrad. Student	Administration	Business	Faculty	Grad. Student	Professional	Other	Total
Occupation / %	188 14.15	38 / 2.86	53 / 3.99	94 / 7.07	287 / 21.60	594 / 44.70	75 / 5.64	1329
Question 8.	Maybe	No	Yes					Total
Willing to Pay Fees / %	727 54.87	532 40.15	66 4.98					1325
Question 9.	Under 20	21 - 25	26 - 30	31 - 35	36 - 40	41 - 50	50 +	Total
Age / %	85 / 6.32	380 / 28.27	362 / 26.93	218 / 16.22	144 / 10.71	103 / 7.66	26 / 1.93	1344
Question 10.	Female	Male						Total
Gender/ %	67 / 5.10	1247 / 94.90						1314

Table 1: Results from the General Survey

utilize the ability of HTTP servers to execute programs. Under this scheme, a program would produce the desired survey and attach a hidden key with the survey. Upon receiving the completed survey, the program would create another survey and, again, encode a key within the document. As with the previous method, cross-survey question analysis is enabled, but this time, we impose a potentially severe time constraint on the user, i.e., the time to complete all the surveys. If a user decides to stop participating in the currently returned survey, any future submissions will require complete re-entry of all surveys, which is undesirable. Finally, HTTP could be modified to make it into a state machine, though the likelihood of this happening is uncertain.

Similar server related problems occur if we accept only fully completed surveys. While we gain from having consistent number of responses to each question in each survey, we would have to develop software that checked each answer, and created an additional document containing the questions that were not answered. The software would also have to keep track of the previous answers and, once the other answers were successfully returned, merge the results into one entry. We will most likely implement this method in future surveys.

In addition, at the time of our survey, few browsers and platforms supported the HTML+ forms extension. Thus, our results are biased in favor of those that did support Forms (i.e., XMosaic for UNIX). Presently, more platforms and browsers support the extension, thus we expect to eliminate the bias with subsequent survey runs.

A final bias is due to a sampling problem, that is, the self-selected nature of survey respondents. Clearly, frequent Web users were more likely to answer our survey, and hence skew the results. We hope that future deployments of the survey will begin to attract responses from less-frequent users, and thereby allow us to characterize the changing Web user 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.

CONCLUSION

In this paper, we reported results from a survey of World Wide Web users. As shown by the high number of survey respondents, we demonstrated the viability of using the Web and HTML+ forms for conducting on-line surveys. In addition, our results served to begin to describe the range of Web users, their activities, and preferences. 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.

REFERENCES

"A Growing Internet is Trying to Take Care of Business", *New York Times*, December 12, 1993, page F7.

Berners-Lee, T. (1994) Hypertext Markup Language Plus. URL:http://info.cern.ch/hypertext/WWW/MarkUp/HTMLPlus/htmlplus_1.html

Berners-Lee. T. (1993) Hypertext Transfer Protocol. Internet Engineering Task Force Working Draft, 5 Nov 1993. URL:ftp://ds.internic.net/internet-drafts/draft-ietf-iiir-http-00.txt

Berners-Lee, T. (1994) Uniform Resource Locators. Internet Engineering Task Force Working Draft, 21 March 1994. URL:ftp://ds.internic.net/internet-drafts/draft-ietf-uri-url-03.txt

Ragget, D. (1994) Hypertext Markup Language Plus. URL:ftp://ds.internic.net/internet-drafts/draft-raggett-www-html-00.txt

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 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.

MOSAIC PATTERNS

There were 1079 responses to the Mosaic survey. Not surprisingly, most users (93%) were using Mosaic for the UNIX platform (see Table 6), due to the unavailability of HTML+ forms support in Macintosh and DOS/Windows environments. Over a third reported never using the items contained in Mosaic's Help menu while roughly 30% used the Help menu on a weekly or monthly basis. Only 7% found the on-line help on Mosaic to be not useful, with the rest split between considering the help useful (52%) and somewhat useful (40%). The mean rating of Mosaic was 8.086 (same scale as above) with the support staff at NCSA receiving a mean rating of 7.458.

USER PATTERNS

The survey results allowed us to identify patterns or clusters among users' preferences and activities, using the likelihood ratio chi-square test and Cramer statistic (G^2 and C). From the General Survey (see Table 10), we note a strong interaction between shared machines and the occupation of the users ($G^2(6) = 199.290$, $p < .001$; $C(6) = .388$), with more users sharing machines in the educational environments and less in the professional environments. We also observe an interaction between shared machines and the users age ($G^2(6) = 122.729$, $p < .001$; $C(6) = .306$), with younger users more likely to share machines than older users. As can be expected for the above interactions, we find that occupation is associated with user age ($G^2(36) = 835.563$, $p < .001$, $C(36) = .326$), reflecting the younger responses from students. Mild interactions exist between shared machines and willingness to pay fees, with shared machine users less willing to pay for fees and single machine users willing to pay fees for WWW database accesses. Interestingly, we note a slight interaction between user occupations and fees ($G^2(12) = 52.816$, $p < .001$; $C(12) = .141$), where students would not pay for access and professionals would.

The analyses of the HTML survey offer less insight. Here, we note that the number of years programming interacts strongly with the number of programming languages known ($G^2(16) = 733.633$, $p < .001$; $C(16) = .514$). This follows intuition, as users with few years programming know few languages and users with over twelve years programming correspond to knowledge of over twelve languages known. A moderate interaction was observed between the number of HTML documents authored and uncertainty of whether HTML is an ISO standard ($G^2(8) = 61.080$, $p < .001$; $C(8) = .210$).

Numerous interactions exist between the questions in the HTTP survey, though, most are as expected (see Table 14). Of note, we found a strong interaction between knowledge of ISMAPs and ISMAP use ($G^2(2) = 277.407$, $p < .001$; $C(2) = .774$) along with knowledge of CGI scripts and CGI script use ($G^2(2) = 155.007$, $p < .001$; $C(2) = .582$). As expected then, knowledge of server scripts relates to knowledge and use of CGI scripts and ISMAPs, with the latter having a slightly weaker association. Similarly, we observe that operating a server interacts moderately with knowledge and use of CGI scripts and ISMAPs.

Analysis of the Mosaic survey (see Table 13) revealed that those users who used the Help menu on a weekly basis were more inclined to view the on-line help on Mosaic as useful ($G^2(6) = 124.135$, $p < .001$; $C(6) = .244$).

In terms of the Usage Survey (see Table 14), the most frequent Web users appeared to explore the Internet the most ($G^2(10) = 312.501$, $p < .001$; $C(10) = .370$). Although a mild interaction, we note that as the preference for using Web browsers for Internet exploration increased, users appeared to prefer text-based search; as preference for using Web browsers for Internet exploration decreased, users preferred keyword search ($G^2(4) = 312.501$, $p < .001$; $C(4) = .103$).

SURVEY SHORTCOMINGS AND PLANNED IMPROVEMENTS

There were several problem in our survey methods that we plan to correct in future runs. For example, we encountered a trade-off between keeping the questions concise and removing ambiguity. In the general survey, one question asked the name of the domain to which the user belongs. While our intention was to classify the users based upon educational, commercial, military, government, etc. categories. it became apparent that the respondents, especially those not from the United States, were unclear as to whether we were inquiring about their domain (i.e., DNS) or their work activity. Also, with our intention of creating easy to use, point-and-click surveys, we opted to use radio buttons instead of text-entry fields. While this permitted less user overhead, it constrained some user responses by forcing them to answer "other" or "somewhat," often times inappropriately. We plan to remedy these problems in future releases by working closely with survey experts.

We also encountered difficulties with keeping track of users across surveys. This problem results mostly from the stateless nature of the HTTP protocol. That is, the server does not keep track of what machine accessed which documents; this is only logged to file. This meant that we were unable to take the answers submitted by a user for the General Survey and perform analyses against questions in the other surveys. Hence, we could not test for possible interactions like the age of the users and user media orientation. There are several ways which we can avoid this issue in the future. One method would be to have each user provide a unique key upon starting the first survey, and then use this key for subsequent submissions. Hence, all responses to the surveys could be logged with a unique key for each user. This would enable the cross-question analyses we desire. This method adds additional burden to users by making them responsible for remembering the key for future surveys (which may be completed several days later). Another method would extensively

RESULTS

Overall, there were 4,853 responses to all surveys combined. Because our design did *not* ensure that only one survey was submitted per user, we developed software to identify multiple submissions. With this list, we manually went through and compared the time of submissions and if necessary, the content of each submission. That is, if two surveys were received within a fifteen minute period, the content was inspected for differences in responses. This method is both time consuming and vulnerable to error. In all, there were 1006 submissions received from identical machines. From this, we discarded 76 responses, or seven and a half percent from duplicate machines (or one and a half percent of the 4,777 net responses)[2]. In addition, since our design accepted partially completed surveys, the number of responses gathered for each question frequently differ within surveys.

In the next section, we discuss the findings from each survey, followed by a discussion of these results.

GENERAL USAGE

There were over 1344 valid responses in this survey category, accounting for 28% of all the responses. It is important to note that the above number represents the total number of people who completed the survey, but as previously mentioned, this does not mean that there were that number of responses for each question.

The results indicate that 56% of respondents are between the ages of 21 and 30, 94% are male, and 69% are located in North America**.** Over 97% percent report using Mosaic as their browsing tool, 88% percent use UNIX as their platform, and 55% use single-user machines (though, as we discuss in the Shortcomings section, the latter results must be interpreted with care) (see Table 1)**.**

In terms of occupation, 45% of the respondents describe themselves as professionals, and 22% as graduate students (the two largest categories). Thirty-six percent of the respondents claim to spend over 10 hours/week exploring the Internet, while 42% claim to spend 0 to 5 hours/week. Of interest to enterprises contemplating commercial use of the Internet, 55% of the respondents say they might be willing to pay fees for access to WWW information repositories, while 40% say they would not (see Table 1).

BROWSER USAGE

There were over 1000 valid responses to questions regarding browser use and activities (see Tables 8 and 9). A majority of users access their browsers several times daily; 41% say they spend less 5 hours per week exploring the Internet, while 21% claim to spend more than 10 hours per week. Users seem to be fairly evenly divided in their orientation, with 28% preferring search mode, 38% preferring text mode, and 33% preferring visual mode (see Table 9).

We surveyed users as to how often they used their WWW browser, instead of accessing specific client services (e.g., Wais, FTP, etc.), where 1 = "never" and 9 = "always." The results indicate that, overall, users show a strong preference for using their WWW browser for general Internet exploration and instead of the standard Gopher and Wais clients. They show a slight preference for using a Web browser for text retrieval and Archie services. They are neutral in their preference for using WWW browsers to access publisher information, weather, and news of research activities and abstracts (see Table 8).

Finally, we asked users to rate their preferences of the various media supported by the Web. Users show a strong preference for keyword search and embedded images. They show a preference for text and spawned images, and a slight preference for movies and sounds (see Table 8).

HTML PATTERNS

The survey on HTML was answered by 701 people. Table 2. presents the actual number of responses per question as well as the percent per category within each question. Interestingly, slightly over 50% of the users have authored over ten HTML documents, while 7% have never authored a HTML document. Just the same, while a handful (3%) incorrectly answered that HTML was an ISO standard, most respondents (66%) were unsure. Most people (77%) had over ten years of programming experience and knew six to ten programming languages (41%). The late-entry question on prior SGML use indicates that 91% of the users had never used SGML before.

HTTP PATTERNS

The HTTP was the least answered survey (10% of the total responses), with only 481 people submitting responses (see Table 4 for the frequency and percent information) Most users operated a server (78%), with NCSA's server being the most widely operated (71%), and port 80 (78%) the most listened to port. Note that the total number of responses differ for each of these questions, so care must be taken in cross-question comparisons. Even though 58% knew what server scripts are (41% have knowledge of CGI scripts), most have never written a CGI script or used ISMAPs. Finally, out of the people who had used the different servers (CERN, Gn, NCSA, Plexus), the highest mean rating, 7.5 (on a 0 to 9 point scale), went to NCSA's server, though no server received a rating below 5.8 (same scale).

2. Specifically, 29 responses from the General Survey, 5 from the HTML Survey, 27 from the HTTP Survey, 11 from the Mosaic Survey, and 4 from the WWW Browsers Usage Survey were removed from further processing.

complete each survey in three to five minutes. If we had created one survey with all the questions, we felt that users might have been less likely to take the time necessary to complete the entire survey. Second, WWW browsers like Mosaic have difficulty managing documents with large number of embedded forms. We intentionally designed the surveys to minimize the number of embedded forms (but still received comments from users that display problems interfered with the overall aesthetics and effectiveness of the survey). Finally, categorizing questions facilitated clarity of intention. That is, the users knew before-hand the area each survey was intended to cover and therefore did not end up midway through surveys finding themselves unable to answer questions.

The logging software parsed the user's responses and append the results to file. Each entry in the logs was time coded, but machine names were not recorded. While this method ensured a moderate degree of animity (correlations between the time logged and the times recorded by the server log can still be derived) it primarily resulted from the use of National Center for Supecomputing Applications' (NCSA) httpd_1.0a5 server, which does not pass remote host information into the sub-shell's environment.

As stated above, the surveys were divided into five categories: 1) general background questions, 2) HTML questions, 3) HTTP questions, 4) Mosaic questions, and 5) WWW Browser/Internet usage. 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 surveys primarily used HTML+ forms. Specifically, radio button (buttons that only allow one answer to be selected per question) were used to list available choices or corresponded to numerical ratings. We intentionally designed as few questions as possible that required text entry as this imposes greater demands on the end-user and decreases control over of the content of the responses.

The survey was officially announced on January 17th, 1994, on the *comp.infosystems.www* newsgroup. The posting simply stated the impetus behind the survey (i.e., to provide the community with a characterization of WWW users) and the location via the Uniform Resource Locator [URL] of the survey. Subsequent notices were mailed to NCSA's *What New Page* maintainers and the maintainers of the WWW server at CERN due to the high visibility of these servers. Several pointers from other documents on the Web were also observed.

SURVEY QUESTIONS

The General Information Survey (11 questions) asked questions that involved no prior knowledge of WWW technologies. We mainly asked questions such as user age, occupation, domain, geographical location, and computer platform and usage. As with all the other surveys, a comment box was located at the end of the survey for users to contribute whatever information they thought relevant.

The HTML Survey (12 questions) was intended to characterize how difficult HTML is for users to learn and use. Question asked respondents about the number of HTML documents authored, whether or not they knew if HTML was an International Standards Organization (ISO) standard (HTML is not an ISO standard), how difficult/easy they found it to learn FORM and ISMAP, and if the documentation on HTML was current and easy to understand. We also asked background questions inquiring about the number of years of programming experience, and the number of languages known. An additional question was added on Jan 21, 1994. It asked whether or not users had prior experience with Standard Generalized markup Language (SGML) (from which HTML is derived).

The HTTP Survey (14 questions) was primarily designed to determine the extent of knowledge people have about HTTP. Hence, respondents found questions asking about whether they knew about server scripts, modules, CGI scripts, and ISMAPs and if they had used any of these methods. We also asked if respondents operated a server (and if so, which one and what port it listened to). Users were also asked to rate, if applicable, the existing servers at the time (CERN, Gn, NCSA, and Plexus) on a scale from one to nine.

The Mosaic Survey (5 questions) was intended to gain feedback from users of NCSA's Mosaic. This questionnaire was the shortest of all the surveys, (we had asked the staff at NCSA to submit additional questions). Basically, we wished to know what platforms users primarily used (this question was heavily biased - see Discussion below), what rating they would assign to Mosaic, how they perceived the support staff at NCSA, and how helpful they found the on-line help.

The WWW Browser/Internet Usage Survey (20 questions) sought to find out more about how and why people use their WWW browser. This category was broken down into five subcategories: frequency, motivation, typical activities, media type assessment, and general usage traits. The first two categories are self-explanatory, while the others might require some explanation. 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, archie, FTP, wais, etc., as well as questions on use for exploration and accessing other non-traditional resources (e.g., weather). For the media type assessment, we wished to determine media preferences. The surveys used radio and check box buttons for questions addressing motivation, frequency, and general usage traits, and used scalar radio buttons for the typical activities and media assessment questions.[1]

1. The actual questionnaires in HTML+ form are available via: http://www.gatech.edu/pitkow/survey/survey-1-1994/index.html. Postscript and ASCII versions are accessible via anonymous ftp from: ftp.cc.gatech.edu in /pub/gvu/www/pitkow/survey/survey-1-1994. All the comments collected from the surveys are accessible via: http://www.gatech.edu/pitkow/survey/survey-1-1994/comments/comments.html.

RESULTS FROM THE FIRST WORLD-WIDE WEB USER SURVEY

James E. Pitkow & Margaret M. Recker

Graphics, Visualization, & Usability Center
College of Computing
Georgia Institute of Technology
Atlanta, GA 30332-0280
E-mail {pitkow, mimi}@cc.gatech.edu

ABSTRACT

The explosion of World-Wide Web (WWW) across the Internet is staggering, both in terms of number of users and the amount of activity. However, to date, no reliable characterization exists of WWW users. In this paper, we report results from a survey that was posted on the Web for a month, in January of 1994. There were several goals motivating our survey. First, we wished to demonstrate a proof of concept for WWW technologies as a useful survey medium. Second, we wanted to beta-test the design and content of surveys dealing with the Web. Third, as mentioned, we hoped to begin to describe the range of Web users. In one month, we had over 4,700 respondents to our survey. Their responses helped us to begin to characterize WWW users, their reasons for using the WWW, and their opinions of WWW tools and technologies.

KEYWORDS

statistics, surveys, demographics, tools, administration

INTRODUCTION

According to recent estimates, the Internet is gaining roughly 150,000 new users per month, joining 20 million existing Internet users [New York Times, 1993]. At the same time, the number of WWW packets is growing exponentially. While these numbers provide useful approximations on the amount of users and their Web-related activities, little is known about their backgrounds, interests, and characteristics.

In this paper, we report results from a survey that was posted on the Web for a month, in January of 1994. There were several goals motivating our survey. First, we wished to demonstrate a proof of concept for WWW technologies as a useful survey medium. Second, we wanted to beta-test the design and content of surveys dealing with the Web. Third, as mentioned, we hoped to begin to describe the range of Web users.

In one month, we had 4,777 respondents to our survey. Their responses helped us to begin to characterize WWW users, their reasons for using the WWW, and their opinions of WWW tools and technologies. However, as we will discuss, these responses are representative of a particular class of WWW users, and therefore cannot yet be generalized to the general WWW population. For this reason, we plan to improve the survey and run it periodically in the upcoming years. This will allow us to track changes in the population as WWW evolves and matures. Furthermore, the survey results will help us extrapolate to the general Internet population.

METHOD

Using the Internet to survey users is not new. However, the problem with most of these surveys has typically been an inadequate level of participation. We believe that the lack of respondents is primarily due to the overhead incurred. That is, most surveys ask the users to report their responses via electronic mail. For example, a survey on WWW technologies that required e-mail responses was posted to *comp.info-systems.www* in late January 1994. The results of the survey revealed that only fifty-five responses were gathered. While surveys of this type are more effective than conventional mail or phone, they still require considerable end-user effort, which ought to be avoided if possible.

Fortunately, the level of client interactivity supported by Hypertext Markup Language Plus [HTML+], i.e. forms, is sufficient to enable low-overhead, point and click responses, as well as text entry. Further enhancing the attractiveness of using WWWW technologies is the Hypertext Transfer Protocol [HTTP] server's ability to retrieve documents (the actual surveys) *and* execute programs (response logging software). Thus, not only are users able to respond to surveys in an intuitive, non-labor intensive manner, but the results can be logged automatically via the HTTP server.

The design of our survey attempted to exploit WWW technologies. Specifically, we created five separate surveys (with questions on general background, HTML, HTTP, Mosaic, and WWW Browser/Internet Usage) and rudimentary logging software. The decision to create separate surveys was made for several reasons. First, we wanted users to be able to