

Paper prepared for CSS2000 online conference (April 15 – May 15, 2000)

Online Survey of Personal Webpage Owners in Japan\*  
Method, System and Practice

Yasufumi SHIBANAI  
Doshisha University

and

Yasuyuki KAWAURA  
Yokohama City University

**Introduction**

The Internet has rapidly grown, especially since the latter half of 1990s, and computer-mediated communication that it provides has come to be an unnegligible part of daily lives. Many advertisements on television, magazines and even street billboards show URLs which provide detailed information, and WWW is an essential means of political campaigns nowadays. Email is no more a special communication medium for computer experts. Nielsen//Netratings, one of the major Internet measurement company, reported that at the end of 1999, there are 119 million Internet users in the U.S., just half of which consists of females (Nielsen//Netratings, 2000). Things are almost the same in Japan, although still on the rapid growing process. The number of Internet user at the end of 1999, surveyed by RDD method among Japanese households was estimated to be 15 million, 37% of which were female (Netratings Japan, 2000).

With the recent development and diffusion of the Internet, psychological, sociological and anthropological research on the behavior of the Internet user has come to receive considerable attention (Wallace, 1999; Gackenbach, 1998; Turkle, 1997; Hill and Hughes, 1997; Lindlof and Schatzer, 1998; for example). There appear researches on the social impact of the Internet usage among general samples, especially their email

---

\* This research is part of a joint project by authors and Ken'ichi Ikeda (University of Tokyo). Authors would like to express their gratitude to Okawa Foundation for Information and Telecommunications and NTT Communication Science Laboratories, for their financial supports that made the project possible.

communications and web-viewing (Kraut, et al., 1998; etc), but many researches on computer-mediated communication have focused on more “active” participation in the net, such as BBSs, Internet newsgroups or virtual communities like MUDs. These kinds of research adopt planned/natural experiments, surveys of participants, network analysis of interactions, content analysis of messages and participatory observation, to reveal how netters feel, behave, and relate one another in the “virtual community” (Rheingold, 1993).

These researches have an important contribution to the field, to interpret and predict social consequences of the technology and to develop and improve better communication systems. In the meantime, social scientific research of the World-Wide Web usage, especially that of webpage owners who are personally distributing information, have been very rare. These personal webpages are interesting in several senses. At first, the WWW enables everyone to broadcast information to the world-wide receivers, which was virtually impossible except among mass media industries. Reasons of publishing personal webpages, and gratification they receive from their activities, have not well known yet. The WWW is also interesting from a standpoint of communication models. Since Shannon and Weaver (1949) proposed their sender-receiver model of communications, many communication theories of social sciences, which were strongly influenced by the model, implicitly or explicitly assumed activeness of information senders and passiveness of receivers. On the other hand, the new WWW technology can be said to change the mode of communication. Information senders just put their information at a server, and wait until that information will be found by information receivers (seekers). Here the activeness of communication is exchanged between senders and receivers. What kinds of communication behavior are emerging here has also not well known enough to the date. The reason for small number of researches may be due to difficulties to make a observation of these webpages partly caused by the very nature of communication passiveness.

If active participants in the Internet (like in newsgroups) are research targets, observation and drawing of samples is relatively easy, because the whole population is properly defined when an appropriate (virtual) place and period is set. Making a complete log file and extracting data from it is sufficient. The records are everything because their action is active and explicit.

But in the case of webpage owners, things are quite different. They are usually under the surface, as they are passive communicator. They come to the surface when their URL is linked or introduced by someone, but it reflects the popularity of their webpage so that research based on such explicit cues may be biased. To grasp the whole image of personal webpage owners by survey method is quite difficult.

The present research aimed two goals. First one is to test authors' planned procedure for studying personal webpage owners. The present paper mostly focused on the description of the procedure in detail for the purpose of establishing web survey methodology. The second goal is to obtain basic descriptive data of Japanese webpage owners, to generate research hypotheses for further study. These basic data will be presented in the subsequent section.

### Method

#### **Sample**

To obtain a representative sample, a sampling frame of the target population is required. Although it might be ideal to set the research target to whole webpage owners (of Japan or in the world) in the Internet, it also is not realistic, even if web-robot based sample collection is possible. As a substitute, we decided to use a sufficiently large and diverse collection of personal webpages, and choose those registered in NIFTY (<http://www.nifty.com/>), which is one of the largest Internet service providers (ISP) in Japan. NIFTY has 2.3 million individual customers at the end of August, 1999, and provides to them number of own services, such as electronic community (forums and BBSs), electronic commerce and transaction. NIFTY also provides "members homepage service", which enables its customers to publish their own webpage at no surcharge. Official release of the company reported that there are about sixty-six thousands of personal webpages registered at that time.

To make a sampling frame of personal webpages in NIFTY, search system of member's webpage was used. By a search with no specification of a keyword, a list of URLs of all webpages registered in the system could be obtained. The list was split in 1,265 separate html pages, and a script written by perl was executed to automatically download the entire record. This procedure was conducted on August 20, 1999. After editing the list, 16,689 individual URLs were obtained finally, which covered about a quarter of officially reported number of webpages at that time. From this sampling

frame, 1,000 URLs were selected by simple random sampling.

Before launching the survey system, whole URLs were checked by authors and classified according to the types of publisher, to exclude inappropriate samples for this study, such as pages made by profit/non-profit organizations. Table 1 shows the result of this classification. Classification of “pseudo-personal” means that although they could be a personal webpage, most of their contents were occupied by advertisements (especially “banner advertisements”), and no substantial contents were found. Here first (personal) and second (co-opened) category at table 1 were selected for the sample of the study (743 samples).

Table 1 Classification of samples

1. personal/individual	721
2. co-opend (ex. friends/families)	22
3. pseudo-personal profit-making	4
4. commercial (with company's name)	112
5. non-profit organizerion	75
6. under construction	14
7. unable to access	52

These 743 webages were checked manually, to collect email addresses of their owners. , 729 email addresses were identified, and five addresses among them were not valid (rejected later by the mail system). As a result, 724 webpages and their owners were drawn as the survey targets.

### Survey system and questionnaire

The survey system was constructed by ordinary html components (radio buttons, check boxes and text boxes), and CGI programs written by perl language. Actually, the whole survey system consisted of one html page and two CGI programs. The html page is an “entrance”, which explains an aim of the study and asks targets to type a password allocated to them individually (described after). By clicking the submit button in the entrance page, the first CGI (CGI-1) was executed, to check the password with the password list. If a given password is invalid, or valid but the response of that password was already registered, CGI-1 outputs an error message and refuses participation. When a given password is valid and the response corresponding to the

password has not registered at that time, CGI-1 outputs one-page questionnaire. Responses to the questionnaire were parsed and checked by the second CGI (CGI-2), after clicking the submit button in the questionnaire. When a missing answer was found, CGI-2 reports that question and ask the respondent to go back and fill. Final message with appreciation was displayed if CGI found no missing answers. The time of registration is automatically recorded to the log by CGI-2.

The questionnaire was comprised of four parts, (1)Questions regarding to published webpage, (2)General Internet usage and psychological responses to the Internet, (3)General personality items, and (4)Demographic variables. Webpage-related questions contains the date and reasons of the publication, frequency of webpage update, behavior and psychological responses of webpage publication, such as how they make links to other pages, and what kind of responses they get from their readers. Seventeen question sections and four demographic variables were presented in a single page.

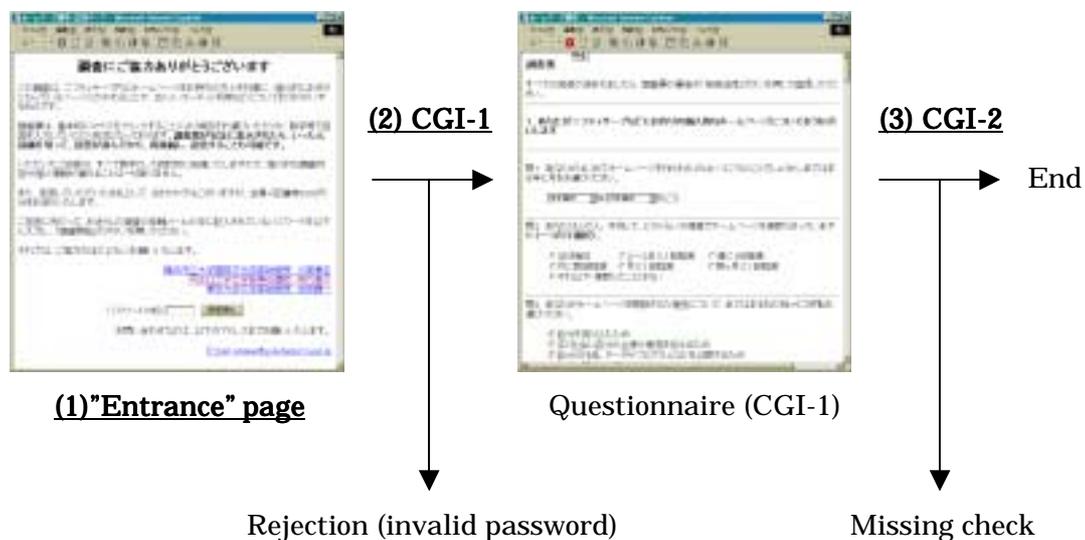


Figure 1 Structure of the survey system

The survey system was setup in a server computer at Yokohama City University, which runs BSD Unix and Apache server software. The machine was allocated a unique Internet address with domain name, and the URL was informed to the targets in the requesting email.

### Survey procedure

Request Emails were sent, on September 20, 1999 (in the morning, Japan standard time) to 724 targets whose email address is found in their webpage. The message explains an aim of the study, and asks them to participate to the survey. The URL address of the survey system was presented, with a unique password given to each target. The password consists of seven alpha-numeric characters randomly assigned (ex. "AXF26RT"), and the total number of combinations is up to 156 million. The possibility of guessing a password was quite low, and no attacks were recorded in the system log. The correspondence list of password and the targets' mail address (and URL of their homepage) was created when the passwords were generated, which made the identification of respondents possible, and further analyses, like combination of content analysis of webpages and the survey responses is now on process. In the email, it was informed that the participant could receive a reward of a five-hundred-yen book gift coupon (approximately worth about five dollars.) To send a requesting message to the targets, free software "Direct Email" was used, which can transmit multiple-recipient email message with automatic insertion of recipient-specific information from datafile (so individual passwords could be properly sent).<sup>1</sup>

The survey term was set to one week, so that the deadline of participation was September 26. In the evening of September 24th (fifth day of one-week survey term), a reminder email was sent to targets whose responses had not recorded to the system at that time. Rewarding policy was changed, at the time of the reminder, to do a lottery (three coupons for fifty winners) in addition to flat one coupon for everyone, for the purpose of motivating participation. This changed policy was announced in the reminder.

At the beginning of the survey term, authors received several emails inquiring whether the survey was "real" or malicious one with fake names. To make the survey more reliable, authors put links to their official homepages on the survey entrance, and also added a small notice of the survey on their official homepages, so that the survey's reliability could be cross-checked by targets.

---

<sup>1</sup> This software can be obtained from "<http://hp.vector.co.jp/authors/VA009907/demail/>". The page and software are both in Japanese language, and Japanese Windows system should be required.

After the term of the survey, a gift coupon was sent to the respondents who wish to receive a reward and registered their postal address at the end of the questionnaire. Result of the survey, including a brief description of the procedure and simple frequency tables was released to the respondents, at the same URL of the survey system. This announcement was made in the mail sent to them with the reward.

### Basic Result of the Survey

#### **Response rate and its time series change**

Two hundred and ninety responses were recorded at the end of the survey term, and response rate against approaches was 40.05%. Figure 2 shows the daily number of responses. As the graph indicates, over one-third of responses were recorded in the evening of the first day. The number declined along survey term, except the day when reminders were sent. Small number of increase is observed in the evening of the last day, but 70% of total responses had been registered before the reminder.

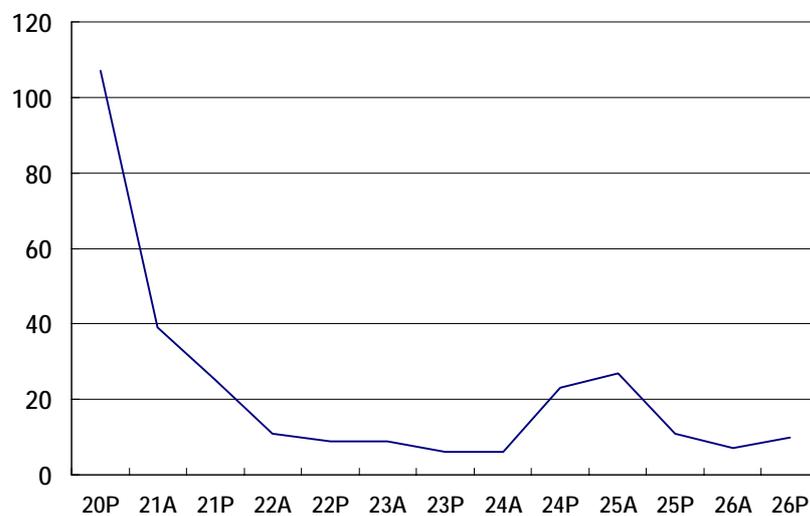


Figure 2 Daily number of survey responses

Note: A/P denotes AM/PM of each day

#### **Demographic characteristics**

About 80.1% of the respondents were dominated by male, which was somewhat higher rate than overall Japanese Internet demographics. The Mean age was 32.81 (SD=8.67, range from 14 to 61, median 32, N=289). From the standpoint of residential area,

20.34% respondents lived in Tokyo metropolitan district, which were overrepresented twice as large as the actual population. 47.58% lived in Kanto area, which is extended metropolitan area including Tokyo metropolis. 39.31% of respondents were engaged in some engineering profession. Students occupied only 9.65% of the total respondents.

**Year of webpage opening and update frequency**

Roughly a quarter of respondents opened their webpage before 1998 (Figure 3). There seems to be an outbreak in 1998, when nearly half of respondents reported to open their webpage. It was just in 1998 when the Internet user in Japan was also rapidly increased, and total number of user exceeded ten millions in the year (Internet Association of Japan, 1998).

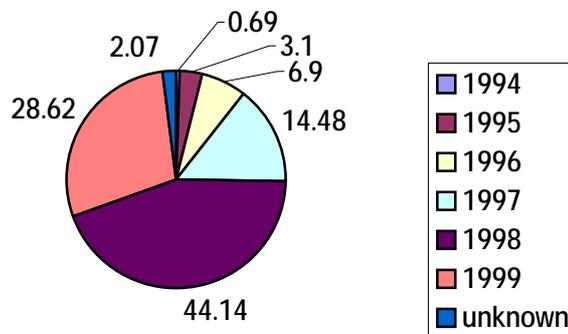


Figure 3 Year of webpage opening

Many respondents update their webpage frequently. One-third of respondents update their webpage more than once in a week, and 7.59 % reported even daily update (These may keep a diary on their webpage. See Kawaura et al. (1998) for detail). Forty percent respondents reported to update at least once in a month.

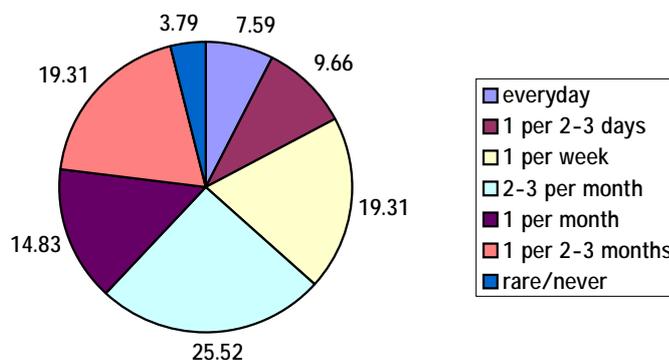


Figure 4 Frequency of webpage update

**Reasons to open personal webpage**

Reasons to open webpage were asked, by seventeen MA items. Table 2 shows items which were chosen by at least twenty percent respondents. Top three reasons regarded to the distribution and exchange of information about respondents' interested fields. Subsequent reasons were related to interpersonal communication, with acquaintances and newly encountered people in the net. Thus, main reasons of opening personal webpage can be said to be information distribution and communication.

Table 2 Major reasons to open webpage

1. to keep a record of hobbies	49.66%
2. to provide information to interest-shared people	48.62%
3. to publish works, data and programs	40.00%
4. to tell one's personal experiences	36.90%
5. to get to know various people	29.66%
6. to inform one's information and present situation to acquaintance	28.97%
7. to learn html and other languages	26.90%

**Discussion: for Further Research and Analysis****From analytical standpoint**

Based on basic statistics described in above section, applied analyses are now going to be performed. Table 3 shows a preliminary example, which was intended to investigate what kind of behavior can activate communication in the WWW based environment. A regression analysis was performed, to predict reported number of feedback emails which were received by respondents in a month. The result reveals that active communication style, such as sending a email to people who is encountered on the net (WWW or BBSs) could promote communication, while behaviors to put links actively in their webpage were not related to the number of received feedback.

Content analysis of respondents' actual webpages is also now on process. All contents of respondents' site were downloaded, and components are now being categorized and counted by coders. The dataset created here, with the combination of the survey

dataset, makes more detailed research possible, such as an investigation of the relationship between web contents and psychological/behavioral characteristics of webpage owners. Panel research, which traces time series change of the current survey respondents is now also planned, so that what kind of people continue/quit the publication of personal webpages can be revealed.

Table 3 Predicted number of reported feedbacks in a month (OLS model)

	beta
<b>[email exchange in general]</b>	
Q9x1 number of sent email in a week	-.033
Q9x2 number of received email in a week	.134*
<b>[link behavior]</b>	
Q4x1 links to most of friends who own page	.022
Q4x3 select carefully when link	-.002
<b>[communication behavior in the net]</b>	
Q10x3 send a email to person encountered on the net	.145*
Q10x6 registering one's URL to search site	-.105+
<b>[page history/update freq]</b>	
Q1x1 years from opening	-.069
Qq2 update frequency	.161*
<b>[demographic variables]</b>	
F1 sex(male=0; female=1)	-.071
F2 age	-.070
R-squared	.080*
N=282 +; p<.1 *; p<.05	
Note: all behavior items were measured by four point scale.	

### From methodological standpoint

There are two major limitations in the present sampling procedure. At first, one specific ISP was chosen to make a sampling frame. Second, the search service of this ISP, which was self-registration based, was used to make an entire URL list. These two procedural limitations may be the source of two biases. The ISP may not be representing Japanese users, although this was the largest and oldest ISP in Japan. More serious self-selection bias may be caused by the latter limitation, because there is the possibility that active webpage owners are overly sampled. Agent based sampling procedure, which automatically traces the web of links and makes a huge lists of webpages could be examined for further research. This procedure also enables network analysis of hyperlinks, which can widen the range of analyses with the combination of survey responses.

Response rate of the current survey was forty percent, which was relatively high result,

even compared to postal mail surveys against general samples. This may be the effect of incentives and the reminder email. But in general, electronic survey procedures are suffering from low response rates. How to achieve acceptable response rates should be also examined in further researches.

### References

- Gackenbach, J.(ed.) 1998 *Psychology and the Internet*. Academic Press.
- Hill, K .A. and Hughes, J. E. 1997 Computer-mediated political communication: The USENET and political communities. *Political Communication*, **14(1)**, 3-27.
- Internet Association of Japan 1998 *Internet white book '98* Impress corporation. (in Japanese)
- Kawaura, Y., Kawakami, Y. and Yamashita, K. 1998 Keeping a diary in cyberspace. *Japansese Psychological Research*, **40**, 234-245.
- Lindlof, T. R. and Shatzer, M. J. 1998 Media ethnography in virtual space: Strategies, limits, and possibilities. *Journal of Broadcasting and Electronic Media*, **42**, 170-189.
- Netratings Japan 2000 *Survey report of Internet demograohics in November/December, 1999* (Press Release, in Japanese)  
URL: [http://www.netratings.co.jp/press\\_releases/pr\\_010200.htm](http://www.netratings.co.jp/press_releases/pr_010200.htm)
- Nielsen//Netratings 2000 *Nielsen//Netratings reports on Internet year1999 in review* (Press Release)  
URL:[http://www.nielsen-netratings.com/press\\_releases/pr\\_000120\\_review.htm](http://www.nielsen-netratings.com/press_releases/pr_000120_review.htm)
- Rheingold, H. 1993 *Virtual community*. Addison-Wesley.
- Shannon, C. E. and Weaver W. 1949 *The mathematical theory of communication*. University of Illinois Press.
- Turkle, S. 1997 *Life on the screen :Identity in the age of the Internet*. Simon and Schuster.
- Wallace, P. 1999 *The psychology of the Internet*. Cambridge University Press.