PG Lab ディスカッションペーパーシリーズ
第29号　2008年7月

ISSN 1880-9146

Discussion Paper Series
No.29 Juiy, 2008

Experimental Use of an Internet Survey: Fusing Social Science and Information Systems for Policy Research

Yang CAO , Taiyo MAEDA and Tadahiko MURATA

文部科学省私立大学社会連携研究推進拠点
関西大学政策グリッドコンピューティング実験センター

Policy Grid Computing Laboratory,
Kansai University
Suita, Osaka 564-8680 Japan
URL : https://www.pglab.kansai-u.ac.jp/
e-mail : pglab@jm.kansai-u.ac.jp
tel. 06-6368-1228
fax. 06-6330-3304

関西大学政策グリッドコンピューティング実験センターからのお願い

本ディスカッションペーパーシリーズを転載，引用，参照されたい場合には、
ご面倒ですが、弊センター（pglab@jm.kansai-u.ac.jp）宛に
ご連絡いただきますようお願い申しあげます。

Attention from Policy Grid Computing Laboratory, Kansai University

Please reprint, cite or quote WITH consulting Kansai University Policy Grid Computing Laboratory (pglab@jm.kansai-u.ac.jp).

Experimental Use of an Internet Survey: Fusing Social Science and Information Systems for Policy Research

Yang CAO[1], Taiyo MAEDA[1] and Tadahiko MURATA[1, 2]

Abstract

Developing Internet survey systems can be seen as a fusion of social science and information systems. An experimental Internet survey used in this paper will be discussed herein with regard to the methodology of Internet surveys based on social science, and electronic survey systems based on information science. Specifically, the sampling method used for the Internet survey will be explained first, then the manner in which data about the sample were collected and analyzed. Next, an outline will be given of the Internet-based questionnaire system and its manner of construction. Finally, two proposals will be made, based on the results of this study's experiment, regarding the conduct of Internet surveys.

Keywords: Conventional Sampling Methodology, Response's Character, Questionnaire System, Access Log.

[1] Policy Grid Computing Laboratory, Kansai University

[2] Faculty of Informatics, Kansai University

1 Introduction

The Kansai University Policy Grid Computing Laboratory (PG Lab) takes as its goal the construction of systems for enhancing policy making through the application of grid computing, based on specific research themes such as “Designing Investigations for the Accurate Assessment of Local Needs,” “Constructing and Testing Pre-Policy Evaluative Models Based on Survey Results,” and “Constructing Social Simulation Systems through the Application of Grid Computing Environments.” In order to design the agents of conventional multi-agent simulations more objectively, we collected micro data through two questionnaire surveys carried out between May and July 2006. One was a “Child-Rearing Questionnaire Survey,” taking mothers of children attending private kindergartens in S city in Japan as its subjects. Conducted by commission, the rate of response was 68.9% with 1,536 valid responses for analysis. The other was a “Questionnaire Survey Related to Health in Daily Life,” conducted with a stratified sample drawn from registered voters of S city. Conducted as a postal survey, the response rate was 29.3% with 586 valid responses for analysis.

In the course of planning, executing, and analyzing the above surveys, it became apparent that in order to secure an adequate response rate at low cost for surveys planned from 2008 on, it would be necessary to consider new social survey methods. From the 1990s, and particularly over the past decade, response rates for conventional social surveys have indeed been dropping in Japan. Yatsuhashi [1] points out that previously, for example, the response rate in the placement method of distribution, in which the researcher would visit the subject’s residence and leave the questionnaire, requesting cooperation and returning later to collect it, was about 70%, while for distribution by postal survey methodology a response rate of about 40% was formerly achieved, but currently the rate for the placement method is about 50%, and for postal surveys the percentage is in the 20s. This not only invites high survey costs, but also harbors the possibility of decreasing reliability of the results. In contrast to this trend, with the technological advances and increasing popularity of Internet use, surveys conducted over the Internet have been greatly increasing. The merits of this survey method are its swiftness and low cost. But as Hayashi observed [2], the greatest weakness of the Internet survey method is that the nature of the sample is unclear. Osumi [3] also points out the following problems in his research on Internet survey methods.

> If information (for example, bank account data) about persons who register with the survey is accurately obtained in order to reward all respondents, for example, it appears there are claims this would allow precise identification of the registrants. But this may be cleared if the names of

registrants' friends or family members are used. No matter what method is adopted, in order to secure the quality of the "group of registrants" (the resource, or panel), there is no alternative but to rely on the registrants' cooperation with the survey, the same as in conventional methods. But even if it would appear that anonymous participants or impersonators have been avoided, the situation is more complicated than in conventional surveys because the respondents' faces are not visible. In this manner, no matter how high the level of information processing technology employed, as long as the basic survey facts such as "who was selected in what manner, was the target sample selected with certainty, in what manner did they respond?" are not satisfied then discussion is useless. Further, although it is not possible to get by while avoiding discussion of these points, information about them is in fact almost nil[1].

According to Osumi, "Internet surveys are comprehensive activities conducted with the application of scientific survey methods over the entire survey process, starting with the basic devices of an electronic survey system, the construction of a group of registrants (resource, panel) and its use and management, the design of electronic survey forms and their questions, the sampling operations, the initial request for participation through the collection of responses, the survey report, and so forth"[2]. The needs of fusing of information systems and social science may be glimpsed from this definition.

In Japanese society, where the use of Internet surveys is on the rise in policy research, it is necessary to consider methods of utilization of Internet surveys through experiments in which they are put to practical use. This is the primary goal of the current research.

The PG Lab is engaged in the task of making multi-agent simulation models for the effect of interaction by word of mouth. Web-based social network services (SNS) may be considered a recent example word of mouth interaction. In Japan the SNS that has gained the most popularity is the website "mixi"[3]. According to statistics current as of 7 Nov 2007, membership had already reached the level of 1.24 million registrations. In order to make a comparison with our "Child-Rearing Questionnaire Survey" conducted in 2006, the second goal of this research is to take a "mothers of small children" sample among mixi users, and by using an Internet survey, test the hypothesis that "Social network service utilization may reduce child-rearing stress."

Next, we will explain the social science-based Internet survey method, and the information science-based electronic survey system equipment and structure that are used to achieve the first-mentioned goal of this ongoing research. Specifically, the sampling method used for the Internet

[1] See Osumi [3] p. 2.
[2] See Osumi [3] p. 8.
[3] In strict adherence with mixi's Rules of Use [4], we used only SNS services that are open to the public. The company was in no other way involved in providing the material reported in this contribution.

survey will be given in Section 2, from the perspective of social science, and the manner in which data about the sample was collected and analyzed will be explained. In Section 3, an outline of the Internet-based questionnaire system and its manner of construction will be given. In Section 4, two proposals will be made, based on the results of this study's experiment, regarding the conduct of Internet surveys.

2 An Internet survey based on sampling methodology

2.1 Sample selection and sample response

In comparison with conventional survey methodology (for example, postal surveys, telephone surveys, interview surveys), uncertainty about the nature of the sample, with regard to "who was selected in what manner, was the target sample selected with certainty," has been regarded as problematic in the case of Internet surveys. Accordingly, efforts were made in the current research to follow conventional sampling methodology as closely as possible for the Internet survey.

As a preliminary step for the sampling, a list of mixi user groups was compiled. The list consisted of the first 100 groups that were returned by searching the mixi site with the term "child-rearing" over the period from 19 June to 10 July 2007. From an examination of the data available on the website for this large and diverse list, three representative groups were finally selected. The characteristics of these three groups are given in Table 1.

Because the Internet survey conducted in this research was premised upon the lack of any incentives, it was anticipated that the response rate would probably be around 1%, based upon the researchers'previous experience. Accordingly, in order for the number of responses valid for analysis to reach 100, it would be necessary to send email requests for cooperation to 10,000 persons in the initial sample.

In order to select a sample of 10,000 persons, it was initially planned to select at random 7,000 from the 15,028 registered members of Group A, 1,000 from the 2,844 registered members of Group B, and 2,000 from the 3,938 registered members of Group C. But when email requests for cooperation were actually sent out, using mixi's messaging function, to the first 70 members of the sample selected from Group A, it was realized that two unanticipated problems had been encountered. The first comes from a limitation of mixi's messaging function. Specifically, this is set so that only 20 messages can be sent out within one hour's time. Accordingly it would not be possible to send out email requests for cooperation to the original sample within the time span of the experiment. The second problem relates to the possibility that the requests would be

disregarded as spam mail. While Group A had the most registered members, as it may also have been subject to frequent abuse, many members of the Group had messages like “No solicitations for work at home” inscribed on their member profiles. Accordingly, it was inferred that many would be highly wary of sudden email requests for cooperation with the survey. In order to increase the efficiency of the request procedure under these circumstances, a revised sample of 1000 persons was devised to include the 70 members already contacted from Group A, plus 430 from Group B, and 500 from Group C.

Table 1. Characteristics of the selected groups*

Name (pseudonym)	No. of members	Conditions of participa- tion; level of access	Category	Reasons for selection
Group A	15,028	Open participation; public access	Other	Largest number of participants. Group aim is to consult and share information about childrearing.
Group B	2,844	Open participation; public access	Same age group	Features interchanges among participants with children of the same age.
Group C	3,938	Open participation; public access	Interest group	Participants are housewives.

*As observed during the period 19 June to 10 July 2007

Email requests for survey cooperation were sent to the 1,000 persons thus selected between 8 Nov and 16 Nov 2007. As a result, there were a total of 27 responses to the 1,000 transmissions, for an overall response rate of 2.7%. Further, among the responses there 17 cases in which the reply was “will not cooperate,” so the number of responses of “will cooperate” was 10. Among these 10 respondents, 8 cooperated by using computers to complete the Web survey form, and the remaining 2 cooperated by using mobile phones to respond to an email survey.

The rate of the response for willingness to cooperate in this research thus reached 1%, matching the initial expectation. Further, of the remaining 99% who did not cooperate, 973

persons (97.3% of the total sample) did not even respond to the email request for cooperation. But among these a large number opened the email and accessed the profile of the sender. The reason this is known is first, because messaging service within mixi places a link on the sender's name (nickname), and that person's profile is accessed by clicking on the link. Also, mixi has what it calls a "footprint" (*ashiato*) function, whereby one user can look up which other users have accessed his or her profile, and when. Table 2 gives the correlations for whether the sender's profile was accessed, and whether responses were made to the email request.

Table 2. Correlations for access of sender's profile and email responses

Combinations	No. of persons	Percent
Profile access, email response	9	0.9
No profile access, email response	18	1.8
Profile access, no email response	201	20.1
No profile access, no email response	772	77.2
Total	1,000	100.0

Those who either did not read the email requesting survey cooperation, or who completely ignored the email, neither replying nor checking the sender's profile, totaled 772 persons, or 77.2% of the total. By contrast, among persons who did not respond, those who showed some interest in either the person making the request, or in the content of the request itself, by accessing the email sender's profile, numbered as many as 201 persons, or 20.1% of the total.

The difference between those who cooperate with a survey and those who do not is a problem of great importance which is linked with bias in Internet survey results. Data collected using mixi's "footprint" function, on persons who did not cooperate with the survey request yet accessed the profile of the request's sender, are analyzed in the following.

2.2 Users who examined the request sender's profile

As Wada and Kondo [5] argue, humankind's active nature may be glimpsed in the process of seeking out information, while surfing from one webpage to another, by clicking on a link with a mouse. Accordingly it is necessary to analyze the characteristics of the mixi users who were

intrigued by some element of the survey cooperation request email or its sender, and accessed the sender's profile in an attempt to get more information.

In comparison with the users registered in Groups B and C, it was thought that the users in Group A, perhaps because they frequently receive spam and other unwanted solicitations by email due to diversification of motivations for contacting the group, would be less suited for analysis of users who checked the request sender's profile. Accordingly, the following analysis was conducted with 61 members of Group B and 111 members of Group C, for a total of 172 users.

Table 3 shows the number of days separating the transmission of the survey cooperation request and the accessing of the request sender's profile. It was found that 89% of the users accessed the profile in one day or less from the day the request was sent.

Table 3. Interval between transmission of request email and accessing of sender's profile

Interval	No. of persons	Percent
Same day	115	66.9
1 day later	38	22.1
2 days later	9	5.2
3 days later	2	1.2
4 days later	6	3.5
6 days later	1	0.6
8 days later	1	0.6
Total	172	100.0

To obtain further data regarding users who accessed the sender's profile, the profiles of those users posted on mixi were examined and their contents categorized.

First, looking at the types of photos or images posted on the top page of the profiles, it was found that roughly eight tenths of the users posted photographs of their children (Table 4).

With regard to the user's number of children, as given in the self introduction on the profile, the largest group, nearly seven tenths of the total, had but one child (Table 5).

As for the age of the user's children (taken to be the age of the youngest child in families with two or more children), also as given in the self-introduction, the range was from 2 months to 4 years 3 months, with an average age of 1.34 years. In the 2006 "Child-Rearing Questionnaire Survey" (hereafter "the prior survey"), respondents with only 1 child comprised 20.4% of the sample, while 65% had 2 children. The average age for the oldest child was 6.03 years, for the

second child it was 3.43 years, for the third it was 2.91 years, and for the fourth it was 2.80 years[4]. This contrast between the two samples agrees with statements frequently seen in the self-introductions of the mixi users, declaring their desire "to exchange information about childrearing with others," "to make friends with mothers of children the same age." In other words, the relative lack of experience and knowledge about childrearing was what motivated the mixi users to join the groups focusing on that theme.

Table 4. Photos/images posted on the top pages of users' profiles

Type	No. of persons	Percent
Own children	140	81.4
Character (from anime, etc.)	7	4.1
Dog, cat	4	4.1
Scenery	3	1.7
Celebrity	2	1.2
Food	2	1.2
Doll	1	0.6
Hand	1	0.6
Book	1	0.6
No image	7	4.1
Total	172	100.0

Table 5. Number of children indicated in user's profile

Number	No. of persons	Percent
1	117	68.0
2	36	20.9
3	1	0.6
No indication	18	10.5
Total	172	100.0

[4] See Kansai University Policy Grid Computing Laboratory [6], and Cao, Matsumoto, Murata [7].

2.3 Responses to the Questionnaire survey

As explained above in “2.1 Sample selection and sample response,” as a result of sending email requests for cooperation between 8 Nov and 16 Nov 2007, for cooperation to 1,000 persons selected from the three mixi communities, there were 10 responses of “will cooperate.” Among these, 8 completed the Web survey form from computers, and the remaining 2 responded to an email survey from cell phones. The following discussion of their responses is based mainly on the Web survey.

A second email, describing how to access the survey, was sent on 26 Nov 2007 to the 8 persons cooperating through the Web. Only 5 of these responded by accessing the survey form during survey period, from 26 Nov through 30 Nov 2007. The access log for each respondent will be described in the following section.

3 Questionnaire System

3.1 Overview of Questionnaire System

From the point of view of system architects, requirements of the system should be considered carefully when questionnaire system are developed [8]. The questionnaire system has both the subject needs and the researcher needs. For subjects, they like to answer anonymously at their convenient time. For researchers, they like to acquire additional information about the subjects such as the time when the subject answered the questionnaire besides results of the questionnaire.

In order to realize these needs, we constructed web-based questionnaire system, and we analyzed web server log files. We utilized “Bluemoon.Multi-Survey” [9] to manage questionnaire contents and its results. The Bluemoon.Multi-Survey is based on the phpESP [10] that provides an easy to use web-based interface to designing surveys, deploying them on web site constructed by Xoops. We employ these tools to realize the web-based system construction efficiently.

When the questionnaire system is employed, recording the access log like footprint can be a matter of subjects who do not want to have their personal information recorded. In this survey, though, we utilize only the time information that can tell the starting time and the duration to answer. We obtain these information from a web server log file. This method of analyzing a log file is mentioned in Subsection 3.3.

The system overview is described in Figure 1. To realize the questionnaire system for the survey of a specific community, we develop a cooperating system between SNS site and our web server.

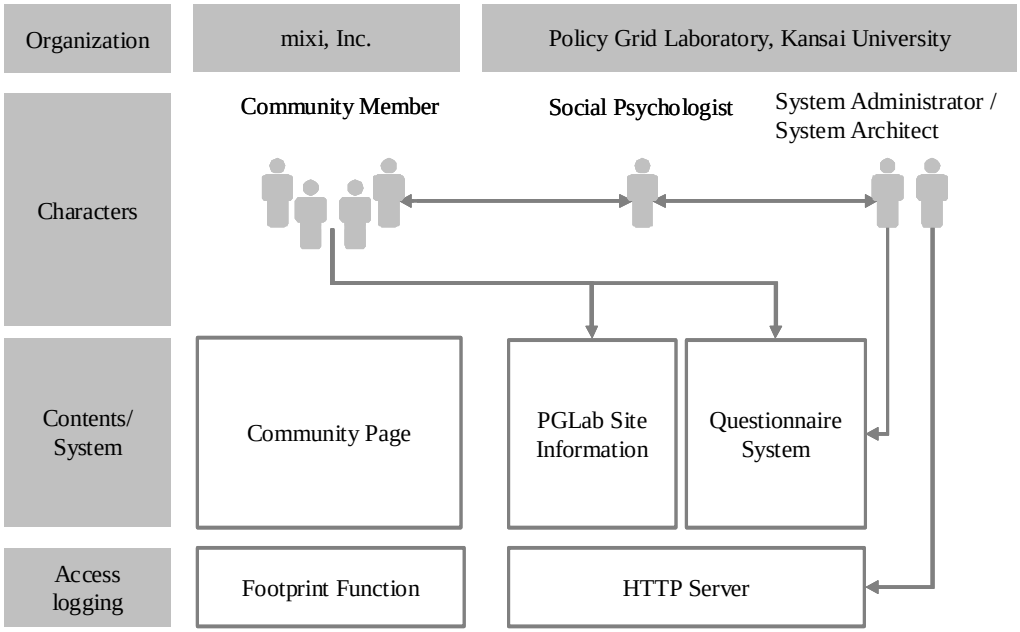

Figure 1. Overview of Web-based Questionnaire System

The specifications of the system server is Xeon3.2GHz, 1GB Memory, 80GB SATA HDD, Gigabit Ethernet. As for the software configuration, we employed the followings: FreeBSD 5.3, Apache2.0, PHP4.3, MySQL4.0, XOOPS2.0, and Bluemoon.Multi-Survey 0.8.

In order to increase the credibility of the questionnaire system, we place it on the same domain of the webpage to introduce the research institute.

3.2 Procedure of Questionnaire

Figure 2 shows a flow of execution procedure of questionnaire. Firstly, a social psychologist designs the investigation, composes the content of the questionnaire, and inputs the content through the system. Secondary, A system administrator and a system architect open the system. Then, the social psychologist send mails to some community members in mixi to ask their cooperations, the social psychologist (as researcher) obtains some answers from mixi community members (as subject), and sends the URL of the questionnaire system.

The subjects answer the questionnaire using the questionnaire system shown in Figure 3. The system administrator and the system architect manage accounts, systems, results of questionnaire and access log files, and transfer the results and the access log to the social psychologist.

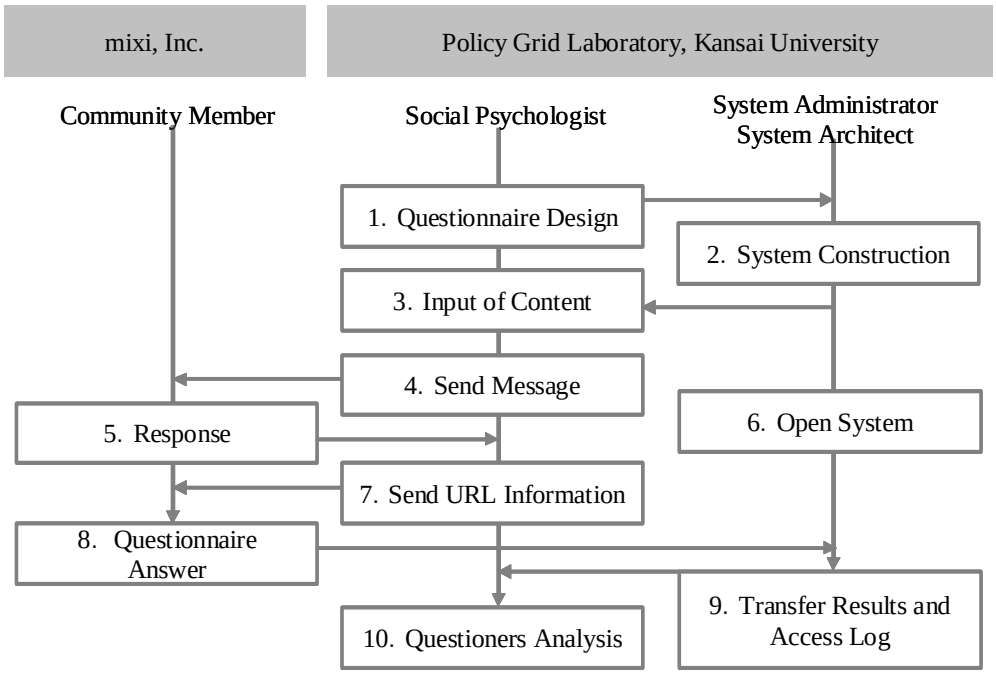

Figure 2. Flow of Questionnaire Execution and Roles of Characters

Figure 3. A Snapshot of Web-based Questionnaire System

3.3 Method of Analyzing Access Log

The contents obtained from access log files are, time zone which a subject answered (Answer Time Zone), elapsed time when a subject answered (Elapsed Time). Each method is mentioned as following.

［Answer Time Zone］

The access log file shows access time, access destination page, connection status when user access a specific page. We defined answer time zone is the access time.

［Elapsed Time］

Questionnaire is composed first question web page, other question pages, and submit answers page. It can know time when a subject access on each page. As elapsed time, we defined time difference between the time of accessing the first question page and the time of submitting the final answer.

3.4 Results from the Access Log

All subjects answered within five days after the questionnaire system was made accessible. The time required for completing the questionnaire is shown in Table 6.

Table 6. Behavior of Subject Access

Respondent	Date	Answer Start Time	Submit Time	Elapsed Time
A	26 Nov	13:24:38	13:29:57	0:05:19
B	26 Nov	22:40:26	22:43:59	0:03:33
C	27 Nov	10:45:44	10:50:55	0:05:11
D	27 Nov	15:03:28	15:22:01	0:18:33
E	30 Nov	15:17:32	15:21:39	0:04:07

From the perspective of social psychology, the difference in decision making between respondents who take, for example, 1 hour to reply versus those who take 5 minutes is a research topic of great interest. In conventional surveys, measuring the time taken to answer is difficult, but with Internet surveys this can be easily resolved. Low cost and swiftness have long been considered the merits of Internet surveys. In the current research, however, it is thought that

utilization of the access log, as a tool supporting the analysis and understanding of the respondents' characteristics, may be the Internet survey's greatest merit.

4 Conclusion

When Internet surveys are commissioned to survey research firms, whereas basic attributes of the respondents are obtained along with the questionnaire results, at present it is not possible to get information on questions such as "Were the respondents' basic attributes reported accurately by the registrants?", or "Who was selected in what manner, was the target sample selected with certainty, in what manner did they respond?" As long as these matters cannot be verified, there will be no common standard for comparing previously accumulated knowledge based on conventional social survey data, and new observations made from Internet surveys. Because of this problem in the current situation, we believe it highly significant for the PG Lab to design and carry out Internet surveys on its own. Accordingly, for the purpose of survey designs from 2008 on, the following proposals are made based on this research.

First, as it may not be possible to raise the response rate, it is suggested that designs which reduce survey costs will be necessary. For example, in a case where a group of 2,000 persons is sampled, and the response rate is between 20% and 30%, an Internet survey may reduce the cost to about one half that of a conventional survey. In recent years, the response rate for conventional surveys has remained at the level of 20% to 30%. Nearly all of those cooperating with the survey are motivated either by the reward, or a desire to have their opinions or views reflected in society. From the results of the Internet survey introduced in this contribution, those who accessed the profile of the person requesting cooperation, due to some interest in obtaining more information, reached 20% of the sample. From hereon, if the provision of incentives and other measures are taken when the PG Lab conducts full-scaled Internet surveys, it may not be too difficult to reach a response level in the 20% to 30% range, but the prospects are not good for research designs aiming for higher levels of response. Accordingly, conducting low cost surveys while maintaining a response rate at the level of 20% to 30% is thought to be a preferable strategy.

Next, in order to examine empirically the effects of time limits on the process and results of human decision making, the accumulation and analysis of data from access logs should be actively pursued. According to Murata [9], in social simulations there are two conceivable methods, that of inferring a future situation from a statistic representing a particular regional group, and the multi-agent simulation based on a summation of the individual decisions made by the persons living in a region who are regarded as autonomous agents. As the PG Lab is involved in researching the latter method, empirical research on the decision making of agents under time

constraints is indispensable. It is proposed that access logs be utilized more actively as a means to this end.

Acknowledgment

We are grateful to Shogo Ichihara （formerly with the Faculty of Sociology, Kansai University） and Kinya Ikenomoto (formerly with the PG Lab research support staff, Kansai University), for cooperation provided in the execution of this Internet survey. We also express our gratitude to mixi, Inc., for providing SNS service.

References

1. Yatsuhashi, T.: Internet Survey is Easy but Accompanied by Risk. Bunkyo University Graduate School of Communication and Information, IT News Letter 2(2) (2006) (in Japanese) 11-12 (http://www.bunkyo.ac.jp/faculty/gs-info/ITNL/2-2/200604-2.pdf)
2. Hayashi, C.:. Changes in Survey Conditions and Problems of New Survey Methods. Proceedings of the Institute of Statistical Mathematics 49(1), (2001) (in Japanese) 199
3. Osumi, N.: What is the Problem with Internet Surveys?: Current Problems and Aspects Needing Solution. Shin Joho 91 (2004) (in Japanese) 1-24
4. mixi (http://mixi.jp/)
5. Wada, S., Kondo, S.: Internet Communication: Pitfalls of the Digital Life. Baifukan, Tokyo(1999) (in Japanese) 52
6. Child-Rearing Questionnaire Survey Report (https://www.pglab.kansai-u.ac.jp/uploads/kosodate_report.pdf/)
7. Cao, Y., Matsumoto, S., Murata, T.; Characteristics of Consciousness about Private Time and Space among Mothers with Small Children: Interdisciplinary Research Related to Model Construction for a Multi-Agent Simulation(1), Proceedings of the 48th Annual Conference on Japanese Society of Social Psychology, (2007) 562-563
8. Kitagaki, I., Tomita, T., Hikita, A.: Designing and development of electronic questionnaire system, Proceedings of the Fifth International Conference on Information Technology Based Higher Education and Training , (2004) 309-311
9. bluemoon.XOOPS(c)-Japanese (http://www.bluemooninc.biz/~xoops/)
10. Kaskalis, T. H.: Localizing and Experiencing Electronic Questionnaires in an Educational Website. Proceedings of World Academy of Science, Engineering And Technology 3 (2005) 267-270
11. Murata, T.; Research and Development by Policy Grid Computing Laboratory, Kansai University. System, Control and Information 51(8) (2007) (in Japanese) 366-367