

Dynamic Interviewing Program (DIP): Automatic Online Interviews via the Instant Messenger ICQ

STEFAN STIEGER, Ph.D.¹ and ULF-DIETRICH REIPS, Ph.D.²

ABSTRACT

A Dynamic Interviewing Program (DIP) is presented for automated online interviewing via the instant messaging service ICQ. DIP can conduct parallel structured interviews with several ICQ users at a time, 24 hours a day. The interview itself can be dynamic (branched testing), which means that each question depends on the answer given to the previous one. An online ICQ address book offers the possibility to validate demographic data in order to evaluate data quality. Experiences with DIP and a first empirical study open up new perspectives for Internet-based automatic interviewing.

INTRODUCTION

FOR MORE THAN A DECADE, the Internet has been used to collect quantitative data, and the advantages are manifold.¹ Access to many participants from rare populations and convincing anonymity in the Internet situation are major plusses for Internet-based research studies, particularly those conducted on sensitive topics.² With sensitive topics, online interviews are a sensible way of collecting quantitative data online. Interviews require synchronous communication; therefore, not all Internet applications are suitable for this. Up until now, online interviews have been mostly conducted in Internet Relay Chats and Multi-User Dungeons.³⁻⁴

Because an interviewer needs to be present for each interview, only one-to-one communication (in contrast to a one-to-group communication) is possible. Only one interview can be conducted at a time, and the presence of the investigator at the same time in the same channel or chat room is essential. The result of these restrictions is a low number of participants compared to the usually higher rates of participants in online questionnaire studies.

What is instant messaging?

In the last few years, another Internet application became very popular: instant messaging (IM). With IM programs, it is possible to send short text messages to other people who use the same IM program. An important feature is the so-called buddy list, which is comparable to an address book. The difference is the possibility to see the online status (online, offline, away, etc.) of people who are listed in the buddy list. Several IM programs are available that have seen double-digit growth rates in their numbers of users.⁵ Besides functionalities like sending files and SMS (Short Message Service) messages to mobile phones and pagers, some IM programs even offer speech and video, provided the users have the proper technical equipment.⁶

Despite the popularity of IM programs, they have rarely been used in social science research.⁷

What is ICQ ("I seek you") and vICQ?

ICQ (homophone for "I seek you") is one of the most popular IM programs on the Internet. In the last decade, the functionality of ICQ has been con-

¹Core Unit for Medical Education, Medical University of Vienna, Vienna, Austria.

²Department of Psychology, University of Zurich, Zurich, Switzerland.

module, and the administration module (see Figure 1). These modules are installed on the iScience server at the University of Zurich, <http://psych-science.unizh.ch/>. More information about DIP and its availability can be found at <http://home-page.univie.ac.at/Stefan.Stieger/dip/>.

Participant recruitment module (PRM)

As with online questionnaires, recruitment can be active or passive. In the active mode, DIP searches autonomously for potential participants. In the passive mode, DIP just waits for interested participants.

There are several methods for active recruiting: (1) Complete randomization is a method very similar to the random-digit method used for telephone interviews. UINs are generated randomly, and invitations for an online interview are sent to these UINs. If the user of an invited UIN is not online at the moment, the message will be delivered upon the next logon. About half of ICQ's accounts are no

longer valid (see Study 2 below); therefore, not all invitations reach their destination. (2) Single interview is the classic form of interviewing. The investigator types in the desired UIN, and DIP starts asking that user. The UIN has to be known in advance. (3) File interview enables searching for UINs of users who fit certain criteria, such as users who live in a specific country. A query is sent to the central ICQ server (see Figure 1), which immediately sends the results back. These results are stored in a file. Then DIP can begin interviewing all users who are listed in the file. Currently, the central ICQ server returns only 20 UINs per query. (4) System randomization is an ICQ service that picks a random chat partner via the central ICQ server (users can set their status from online to "free-for-chat"). (5) On the ICQ Web site, it is possible to search in an online address book for chat partners with certain characteristics and limit the search to users who are online at the moment of search. DIP can store these users' UINs in a database for later interviewing (see



FIG. 2. Sample interview with DIP.

Figure 1, right side). This feature is particularly useful for recruiting only users who are actually online; therefore response rate is higher.

Another way to get in contact with potential participants is passive recruiting, as follows: (6) As a registered ICQ user, DIP can be put into the "free-for-chat" mode in which it signals that it is accepting chat requests. As soon as a message is received, DIP starts interviewing this user. (7) Similar to the recruiting for a Web experiment or survey,¹⁰ the UIN of DIP can be posted in appropriate newsgroups, e-mail lists, forums, or on recruitment Web sites like the Web survey list.¹¹

Interview module

The interview module is the core of DIP. With this module, it is possible to conduct a structured interview where the following options are available (for a sample interview, see Figure 2): (1) Adaptive testing in which questions depend on the preceding answer(s). This is comparable to branched testing in diagnostics. (2) Forced response in which closed questions can be set to require an answer (i.e., respondents have to choose one of the answers provided). (3) Individual help text can be defined for each question to compensate for the missing interaction with a real interviewer who can clarify interviewees' questions. The interviewee simply writes "INFO" to get the proper information. This information has to be formulated by the interviewer in advance. (4) Address book data is saved from the online address book in order to validate demographic data. Whenever an interview is started, DIP automatically sends a request to the central ICQ server asking for details about the interviewee associated with the UIN. The central ICQ server then sends the data the interviewee provided in the online address book back to DIP, which stores the information in a file. This option makes it possible to validate data the interviewee provided in the online interview.

Administration module

In order to make changes to DIP's settings without stopping it, an administration module was programmed. With this module, it is possible to administer DIP from any ICQ program anywhere. Thus, an ongoing study will not be interrupted by any change in configuration. In DIP's configuration, the administrator's UIN is included. If any message is received from this particular UIN, DIP automatically switches itself into the administration mode and treats the message as a command.

Currently, the following commands are available:

STOP: DIP will be stopped.

RANDOM *count filename*: Randomized generation of *count* UINs (see participant recruitment module [PRM] method 1). The result will be stored in the file *filename*.

SINGLE *uin*: An invitation for participation will be sent to the stated UIN (see PRM method 2).

SEARCH_LASTNAME *name filename*, SEARCH_FIRSTNAME *name filename*, and SEARCH_MIN_AGE *name filename*: A search request will be sent to the central ICQ server with the option to search only for people with the stated name or age. The server sends a list of UINs back to DIP, which stores the results locally in the file *filename* (see PRM method 3).

INTERVIEW *filename*: UINs in the stated file *filename* will be invited for an interview (see PRM method 3).

FILE *filename*: The UINs in the given file *filename* will be sent to the administrator's UIN as text.

WAIT_FOR_RANDOM *count*: If DIP is in its active mode, the administrator can set how many seconds DIP should wait before inviting another UIN for an interview (see PRM methods 4 and 5).

Questioning file

The questioning file includes all data necessary to conduct the interview fully and automatically (see Figure 3). It includes the questions to be asked as well as commands in order to guide the interview. The easy setup makes it possible to develop a questioning file in a very short time. HTML programming, such as with online questionnaires, is not necessary.

The buildup shall be demonstrated on the basis of an example line 20 in the questioning file used in the Paine et al.¹² study:

```
{20 = (1/9) What is your gender? [ MALE /
  FEMALE // DECLINE / EXIT / INFO ]
 | FORCED = 1 | MALE = 30 | FEMALE = 30
 | DECLINE = 30 | EXIT = 120 | INFO = 210}.
```

First, the text of the question has to be defined: "(1/9) What is your gender?" Possible answers to this question are listed between the square brackets: [MALE / FEMALE // DECLINE / EXIT / INFO]. The question is defined as forced response (FORCED = 1), meaning that only the options between the brackets are acceptable in order to proceed in the interview. If the interviewee sends an unacceptable answer, DIP generates an error message ("I am sorry, I only understand answers which are listed between the brackets. Please try again and use the options in the brackets."). If *male*, *female*, or

```

# Interview
1=Hello I am DIP. I am a "dynamic interviewing program". I would like to ask you about some demographic data. There are just 9 short
questions. Would you like to participate?
Please use the options in the brackets. [ YES / NO / INFO ]|FORCED=1|YES=20|NO=120|INFO=200
20=(1/9) What is your gender? [ MALE / FEMALE // DECLINE / EXIT / INFO ]|FORCED=1|MALE=30|FEMALE=30|DECLINE=30|EXIT=120|
INFO=210
30=(2/9) What is your age? [ DECLINE / EXIT / INFO ]|FORCED=0|GOTO=40|DECLINE=40|EXIT=120|INFO=220
40=(3/9) What is your citizenship? [ DECLINE / EXIT / INFO ]|FORCED=0|GOTO=50|DECLINE=50|EXIT=120|INFO=230
50=(4/9) What is your marital status? [ SINGLE / MARRIED / PARTNER / DIVORCED / WIDOWED // DECLINE / EXIT / INFO ]|FORCED=1|
SINGLE=60|MARRIED=60|PARTNER=60|DIVORCED=60|WIDOWED=60|DECLINE=60|EXIT=120|INFO=240
60=(5/9) What is your employment status? [ DECLINE / EXIT / INFO ]|FORCED=0|GOTO=70|DECLINE=70|EXIT=120|INFO=250
70=(6/9) What is your current occupation? [ DECLINE / EXIT / INFO ]|FORCED=0|GOTO=80|DECLINE=80|EXIT=120|INFO=260
80=(7/9) What is your highest education? [ DECLINE / EXIT / INFO ]|FORCED=0|GOTO=90|DECLINE=90|EXIT=120|INFO=270
90=(8/9) What is your own ICQ-Account? [ YES / NO / DECLINE / EXIT / INFO ]|FORCED=1|YES=100|NO=100|DECLINE=100|EXIT=120|INFO=280
100=(9/9) If you have any comment on that survey, please tell me now. [ DECLINE / EXIT / INFO ]|FORCED=0|GOTO=120|DECLINE=120|
EXIT=120|INFO=290
110=Thank you for your time.[ENDE=1
120=Would you like to be contacted again? [ YES / NO / INFO ]|FORCED=1|YES=130|NO=130|INFO=300
130=Thank you for your time.[QUIT=1

# End of interview
500=If you want more information about me just visit my homepage at http://mailbox.univie.ac.at/stefan.stieger/dip/|FORCED=0|GOTO=510
510=Sorry that's it. I have no more questions to ask.|FORCED=0|GOTO=520
520=I have to stop. Other people also want to hear my questions. See you next time.|FORCED=0|GOTO=530
530=-----|FORCED=0|GOTO=530

# Info messages
200=Please select one option of the words in brackets if appropriate or visit my homepage http://mailbox.univie.ac.at/stefan.stieger/dip/.
|BACK=1
210=Please tell me your gender. Just use the options between the brackets.|BACK=1
220=Please tell me how old you are. If you don't want to answer enter DECLINE.|BACK=1
230=Please tell me your citizenship.|BACK=1
240=What's your marital status. SINGLE, MARRIED, living with PARTNER, DIVORCED OR WIDOWED.|BACK=1
250=Possible answers can be FULL-TIME, PART-TIME, UNEMPLOYED, in MILITARY SERVICE, HOMEMAKER, RETIRED, NOT WORKING
and not looking for a job.|BACK=1
260=Please tell me what your current job is.|BACK=1
270=Possible is: less than Highschool degree, Highschool degree, Associate's degree, Bachelor's degree and Graduate degree.|BACK=1
280=Are you just using your own ICQ-Account or a different one.|BACK=1
290=Please tell me how you think about the survey and the way it's done.|BACK=1
300=If you say YES you will be added to my opt-in List. You can remove yourself from the list whenever you want on my Web site at
http://mailbox.univie.ac.at/stefan.stieger/dip/.|BACK=1

# Error message
1000=I am sorry, I only understand answers which are listed between the brackets. Please try again and use the options in the brackets.
|BACK=1

```

FIG. 3. Example of questioning file.

decline is given as the answer, DIP proceeds to the next question (line 30: "What is your age?"). If *exit* is entered, DIP stops the interview after a final question (line 120: "Would you like to be contacted again?"). If the interviewee types "INFO," DIP displays a customized help page. This means that DIP jumps to line 210, and the text "Please tell me your gender. Just use the options between the brackets." will be displayed. Directly after that, DIP will ask the original question again.

If a question is not defined as forced response (FORCED = 0), then DIP will move on to the next question regardless of the input. This is necessary because not every question can be presented in a closed format where the list of possible answers is short and clear (e.g., country of origin).

The adaptive nature of DIP can be seen in line 20: MALE = 30 | FEMALE = 30 |. If the interviewee enters *male* or *female*, DIP jumps to line 30. It is

also possible to jump to different lines depending on what answer is given (e.g., MALE = 40 | FEMALE = 30 |). Therefore, the interview process can be branched in order to ask different questions depending on the answer given to the previous question.

STUDY 1: A FIRST EMPIRICAL STUDY

In May 2002, 541 ICQ users were selected randomly (system randomization: see PRM method 4) and invited actively to a short online interview via ICQ about sociodemographic data like sex, age, and profession. Answering these questions only took 1 to 2 minutes. Of the invitees, 28.3% ($n = 153$) showed some reaction to the invitation (e.g., asking questions like "who are you?" "???" or typing nonsense like "§A-1/2Ö," which could sometimes also be a product of a non-Western typeset); 8.1% ($n =$

44) answered at least the first question; and 7.2% ($n = 39$) answered all questions. By comparing interview data to the information stated in the online address book, additional demographic information and false answers could be detected. Mean age of respondents was $M = 21.8$ years ($n = 39$) and of nonrespondents $M = 23.4$ years ($n = 306$). For the remaining 196 users, no information about age could be retrieved. Of the participants, 78.6% ($n = 33$) stated to be male. Also, 69.5% ($n = 267$) of the nonrespondents were male, based on information provided in the online address book. Only 4.5% ($n = 2$) masked their sex, and in 8.1% ($n = 3$) of cases the stated age and age in the address book differed by +plus or minus 2 years or more. Twenty-three interviewees answered the "country of origin" question. Interviewees came from 17 different countries (Austria, Australia, Brazil, Bulgaria, Canada, China, Egypt, Estonia, Germany, Great Britain, Indonesia, Iran, Netherlands, Poland, Thailand, Turkey, and United States).

STUDY 2: ONLINE ADDRESS BOOK-A RESOURCE FOR VALIDATING DATA

In order to estimate what kind of information is available in the online address book and how much of this information is provided by users, an analysis of the ICQ online address book was performed. To get a representative sample, 2,758 ICQ UINs were randomly generated; 1,177 (42.7%) of them were valid accounts. An analysis of the public address book showed that age was stated in 63.2% of accounts, the exact date of birth in 49.1%, sex in 16.9%, and country of origin in 16.5%. Languages, city, and occupation were each provided in less than 15% of the accounts. The low rate of stated sex doesn't correspond to the results of Study 1, in which 53.2% of users had stated their sex in the online address book. A reason may be the significant number of dead accounts (i.e., accounts that are opened only to test the IM program or for an e-learning lesson). In these cases, users may not bother to fill in all the demographic data because they likely wouldn't use the program for long. In Study 1, only users who were actually online (i.e., actively using their accounts) were recruited; therefore, the amount of data that can be used for validation was higher.

An online questionnaire study about IM in which the participants' usernames were requested showed that demographic data from the online questionnaire and the data in the online address book of the

corresponding IM program differed only marginally (age: 6 out of 70; sex: 4 out of 83).¹³ An additional online interview with the same participants revealed an even lower rate of false data (age: 1 out of 13; sex 0 out of 16).

CONCLUSION

DIP was developed to circumvent some disadvantages of traditional online interviewing. With DIP it is possible to conduct several interviews at a time. The questioning file—which includes the order of presentation, the questions themselves, as well as an information text for each question—makes it possible to create a structured interview. This file can be easily edited with any text editor and requires no programming skills. DIP features adaptive testing (i.e., different questions can be asked depending on the interviewee's answer). Closed questions can be defined as forced response, where an acceptable answer has to be given in order to move on in the interview. It is not possible to contact a "real" interviewer during the interview, but an individual information text can be defined for each question.

DIP can be administered from any ICQ client. DIP offers the possibility to contact and interview only users who are actually online in ICQ. This is a big advantage compared with other online data-collection methods like online questionnaires. Our experience shows that the response rate to interview invitations sent by DIP is not as high as with online questionnaires.¹² The reason for this basically lies in ICQ's rigorous antispam policy. New clients come with a spam filter that is activated by default and only lets messages pass from people who are listed in the user's buddy list. Consequently, a large number of invitations for participation never reach their destinations. Furthermore, DIP contacted the respondents actively (i.e. participation was not sought by the respondents). Probably many of those contacted did not have time for participation, and some may have not responded in reactance.

ICQ's online address book can be searched by country, which offers the possibility to conduct cross-cultural studies. Furthermore, data from this online address book can be used to validate demographic data from the online interview. An analysis of this address book showed that, apart from dead accounts, the amount of data available for validation is high.

DIP also has some restrictions. If open-ended questions are used, DIP cannot identify if the given

answer is really an answer to the question asked or some other comment. Therefore, DIP is most appropriate for closed questions and some short open-ended questions. Another restriction is ICQ's proprietary protocol, which means that the source code of ICQ is not open to the public. If AOL, which owns ICQ, were to change the protocol, it is possible that some functions of ICQ wouldn't work properly anymore.

Potential of DIP and IM in the future

For the first time, qualitative structured interviews could be conducted automatically via the IM program ICQ. A first study showed that response rate was not as high as with online questionnaires because of ICQ's rigorous antispam policy. The online address book of ICQ was successfully used to validate demographic data. Participants were mostly young males. The diversity of countries interviewees came from is noticeable. Thus, DIP and similar programs have a high potential in Internet-based research.

Current problems with DIP are that vICQ, DIP's underlying program, will not be further developed and that the ICQ protocol is proprietary. In the future, DIP will be ported to an open-source IM program called Jabber. Jabber is an independent IM solution that offers gateways to other IM programs. People who use, for instance, AOL Instant Messenger (AIM) or Microsoft Messenger (MSN) can therefore also be interviewed. Additionally, Jabber offers encryption of text messages, and its protocol is open to the public.

Interviewees in our empirical studies sometimes had difficulties with the fixed format of the forced response design. If the possible answers to a question were *yes* and *no*, exact wording was essential. If an interviewee typed *yeah*, *yep*, or just *y* instead of *yes*, DIP didn't recognize it as a valid answer. A natural language-processing approach could be very useful here. The given answers revealed that it wasn't always clear to interviewees that the interview was conducted by a program, not a human being.

Despite these solvable problems, DIP can be seen as a first draft of an automatically acting online interviewing program. It offers, on the basis of the results obtained from a first empirical study as well as an analysis of the ICQ online address book, a good basis for further developments.

REFERENCES

1. Birnbaum MH. Human research and data collection via the Internet. *Annual Review of Psychology* 2004; 55:803–32.
2. Mangan M, Reips U-D. Sleep, sex, and the Web: surveying the difficult-to-reach clinical population suffering from sexomnia. *Behavior Research Methods* 2007; 39:233–6.
3. Döring N. (2002). Studying online love and cyber romance. In Batinic B, Reips U-D, Bosnjak M, eds., *Online social sciences*. Seattle: Hogrefe, pp. 333–56.
4. Markham A. (1998). *Life online: researching real experience in virtual space*. London: Altamira Press.
5. Woods B. (2002). U.S. in-home IM usage hits 41M. http://www.instantmessagingplanet.com/enterprise/article.php/10816_1366931 (accessed March 30, 2007).
6. Riva G. The sociocognitive psychology of computer-mediated communication: the present and future of technology-based interactions. *CyberPsychology & Behavior* 2002; 5:581–98.
7. Lenhard A, Rainie L, Lewis O. (2001). *Teenage life online. the rise of the instant message generation and the internet's impact on friendships and family relationships*. Washington, DC: Pew Internet & American Life Project.
8. Gonzo. (2007). vICQ. http://gonzo.kiev.ua/?page_id=136 (accessed March 30, 2007).
9. Fischer R. (2002). Project: NET::ICQ2000. <http://sourceforge.net/projects/neticq2k/> (accessed March 30, 2007).
10. Birnbaum MH, Reips U-D. (2005). Behavioral research and data collection via the Internet. In Proctor RW, Vu K-PL, eds., *The handbook of human factors in Web design*. Mahwah, NJ: Erlbaum, pp. 471–92.
11. Reips U-D, Lengler R. The Web experiment list: a Web service for the recruitment of participants and archiving of Internet-based experiments. *Behavior Research Methods* 2005; 37:287–92.
12. Paine C, Reips U-D, Stieger S, Joinson A, Buchanan T. Internet users' perceptions of privacy concerns and privacy actions. *International Journal of Human Computer Studies* 2007; 65:526–36.
13. Stieger S, Göritz A. Using instant messaging for Internet-based interviews. *CyberPsychology & Behavior* 2006; 9:552–9.

Address reprint requests to:

Dr. Stefan Stieger
Core Unit for Medical Education
Medical University of Vienna
Spitalgasse 23, BT87, P.O. 10
A-1097 Vienna, Austria

E-mail: stefan.stieger@meduniwien.ac.at