TADEQ: A Tool for Analysing and Documenting Electronic Questionnaires

Jelke Bethlehem, Statistics Netherlands, and Tony Manners, Office for National Statistics (UK)

Introduction

National Statistical Institutes (NSIs), research institutes, and commercial marketing research organisations are increasingly using computer-assisted interview (CAI) systems for collecting survey data[i]. The paper questionnaire is replaced by a computer program that guides the respondent through the questionnaire and checks the answers on the spot. The growing possibilities of computer hardware and software have made it possible to develop very large and very complex electronic questionnaires. Unfortunately, it has become more and more difficult for developers, interviewers, supervisors, and managers to keep control of the content and structure of CAI instruments. Secondary analysts of the microdata - and even some of the analysts within the agencies collecting data by CAI - have found it difficult to use the questionnaire documentation, which looks completely different from the documentation they are used to (i.e. paper questionnaires). Sometimes, the CAI documentation is no more than the questionnaire program or a simple listing of the questions without routing information (so the user is forced back to the program). Some CAI software packages can produce a paper questionnaire version for simple surveys, but the social surveys typically carried out by NSIs and research institutes are much larger and more complex than these can cope with. Moreover, CAI instruments can include important information about the questions which paper questionnaires do not, such as edit checks and detailed interviewer instructions. This paper argues that CAI documentation ought to make all this information available, in a range of easily accessible options, for the different types of users with their varying needs.

With *4th Framework* funding from the European Commission, a research project has been set up under the leadership of Statistics Netherlands (see Section 5 below) to develop a tool for human-readable presentation of the electronic questionnaire. This tool is called TADEQ (Tool for the Analysis and Documentation of Electronic Questionnaires). The output of such a tool can serve to document (on paper, or electronically in hypertext form) an electronic questionnaire in a human-readable way. Such a tool should not only provide a useful documentation of the contents and structure, but also help to analyse the questionnaire, and report possible sources of problems in its structure.

This contribution contains a description of the TADEQ project. Section 2 gives some background with respect to the phenomenon of computer assisted interviewing. Section 3 explains the problem of documenting electronic questionnaires. Section 4 describes the possible analysis functions of the proposed tool.

2? Computer assisted interviewing

Our complex society experiences an increasing demand for statistical information. Such information enables policy makers and others to take informed decisions for a better future. Although administrative records in registers and other databases can sometimes provide the required information, more often they cannot. In this case, the sample survey is a powerful means to collect new information

Carrying out a survey is often a complex, costly and time-consuming process. The first step in the process is the survey design phase, in which the statistician specifies the population to be investigated, the data to be collected, and the characteristics to be estimated. Also a questionnaire has to be designed, containing the questions to be asked of the respondents. Furthermore, in the case of a sample survey, the sampling design must be specified, and the sample must be selected accordingly.

The second step in the process is data collection. Traditionally, in many surveys the questionnaires are completed in face-to-face interviews. The quality of the collected data tends to be good. However, since it typically requires a large number of interviewers, who may all have to do much travelling, it can be expensive and time-consuming. Therefore telephone interviewing is sometimes used as an alternative. However, there are limits. Telephone interviewing is not always feasible: only connected people can be contacted, and the questionnaire should not be too long or too complicated. A mail survey is cheaper still: no interviewers at all are needed. Questionnaires are mailed to potential respondents with the request to return the completed forms. Although reminders can be sent, the persuasive power of the interviewer is lacking, and therefore response tends to be lower in this type of survey, and so does the quality of collected data.

If the data are collected by means of paper forms, completed questionnaires have to undergo extensive treatment. In order to produce high quality statistics, it is vital to remove any errors. This step is called data editing. Routing, range and consistency errors have to be detected and corrected, but this can be very difficult if it has to be done afterwards, at the office. In many cases, particularly for household surveys, respondents cannot be contacted again, so other ways have to be found to do something about the problem. Sometimes it is possible to determine a reasonable approximation of a correct value by means of an imputation technique, but in other cases an incorrect value is replaced by the special code indicating the value is ‘missing’.

The computer has always played in important role in processing survey data. In the early days of the computer era, the computer was only used for sorting, tabulation, and checking. In the sixties, the computer was increasingly used for statistical analysis. The rapid development of information technology in recent decades made it possible to use microcomputers for computer-assisted interviewing (CAI). The paper questionnaire is replaced by a computer program containing the questions to be asked. The computer takes control of the interviewing process. It performs two important activities:

- *Route control.* The computer program determines which question is to be asked next and displays that question on the screen. Such a decision may depend on the answers to previous questions. Hence it relieves the interviewer of the task of taking care of the correct route through the questionnaire. As a result, it is not possible anymore to make route errors.

- *Error checking*. The computer program checks the answers to the questions which are entered. If an error is detected, the program gives a warning and one or more of the answers concerned can be modified. The program will not proceed to the next question until all detected errors have been corrected.

Application of computer-assisted data collection has three major advantages. In the first place it simplifies the work of interviewer (no more route control), in the second place it improves the quality of the collected data, and in the third place data is entered in the computer during the interview resulting in a clean record, so no more subsequent data entry and data editing is necessary.

The first mode of interviewing for which computers were used, was Computer Assisted Telephone Interviewing (CATI). This is a form of telephone interviewing in which the computer selects the proper question to be answered. This question is displayed on the computer screen, and thus can be asked by the interviewer. The answer is typed in and the computer checks it for range errors and consistency errors. If an error is detected, the computer warns the interviewer that something is wrong, and corrections can be made.

A more recent technique is Computer Assisted Personal Interviewing (CAPI). It is a form of face-to-face interviewing in which interviewers use a small laptop computer to ask the questions and to record the answers, instead of the traditional paper form. It is like CATI except that interviewers equipped with laptops visit respondents, usually at their home addresses. The laptops communicate with the central office by telephone, sending completed interviews and other data about the sample and receiving new CAPI instruments, new addresses, email and other electronic items.

Mail surveys are also becoming more and more automated. This form of computer assisted interviewing is called Computer Assisted Self Interviewing (CASI). In a CASI survey, the electronic questionnaire is sent to the respondents (on diskette, or by modem). The respondents run the interview program on their own computers, answer the questions, and the data is sent back in the same way it came.

CATI, CAPI and CASI are the main forms of CAI. New varieties are constantly under development, such as CASI via the Internet and audio-CASI[ii].

There are related forms of electronic data collection, such as Electronic Data Interchange (EDI) and passive observation methods such as linking supermarket bills with personal information on loyalty cards. This paper, like the project it describes, is confined to electronic data collection where the data to be collected do not already exist in accessible electronic form, i.e. where some form of interview or self-interview is required to obtain the information.

The growing possibilities of computer hardware and software have made it possible to develop very large and very complex electronic questionnaires. It is not uncommon for electronic questionnaires to have thousands of questions, although only a sub-set will apply to any particular respondent. Routing structures and filter questions see to it that respondents are only asked relevant questions for the particular sub-population to which they belong. Another aspect of

complexity concerns the structure of the instrument. Many survey questionnaires are of a hierarchical nature. A relatively simple but common example is a health survey, with questions at the household level, then at the level of each member of the household. Within each person there may be questions for each medical problem or each visit to the doctor.

Due to the increasing size and complexity of electronic questionnaires, it has become more and more difficult for developers, users and managers to keep control of their content and structure. Although several computer assisted interviewing systems have high level authoring languages to specify questionnaires, it takes a substantial amount of knowledge and experience to understand large and complex questionnaires. It has become more and more difficult to comprehend electronic questionnaires in their entirety, and to understand the process that leads to responses to each of the questions as they ultimately appear on data files.

This raises the question of the feasibility of a tool to represent the content and logic of an electronic questionnaire in a human-readable way. Such a tool should not only provide a useful documentation, but also help to analyse the questionnaire, and report possible sources of problems.

2? ? Documenting electronic questionnaires

When surveys were conducted with paper questionnaires, there was no need for separate documentation of questions and routing. Interviewers' instructions and edit checks carried out in the office were usually available in separate documents, though if these were computer programs they tended to be even less readable than CAI programs. Nevertheless, the paper questionnaire, developed for the interviewer's use, contained all the information available about the questions. Routing instructions were kept simple, in order to allow the interviewers to process the skip instructions themselves.

Computer assisted interviewing brought about two changes: the paper document was replaced by a computer program, and it became possible to implement more complex route logic. So, data collection instruments are more complex, and there is less documentation to understand them.

In the early days of computer assisted interviewing, when electronic questionnaires were developed for computers of limited memory size and speed, it was still possible to produce documentation by hand. There are ample examples of hand-made flow charts. The recent developments in information technology have made it possible to create such large and complex questionnaires that the cost of creating and maintaining documentation by hand has become prohibitive.

An additional problem of manual documentation is that it can be a source of errors. Whatever means are used to express the contents and logic of the questionnaire in a way accessible to a broad group of users, the information is conveyed in a form essentially different form the authoring language of the CAI system. Human translators only can perform their documentation task if they completely understand the electronic questionnaire specification. This is a subjective and error-prone task. Subtle errors are easily introduced. Moreover, it is difficult in practice to ensure that updates to the CAI program are carried over into the documentation, particularly as

these updates tend to be made under extreme time pressures. The errors from manual translation will usually go by unnoticed, causing the user of the documentation to form an erroneous picture.

Consistent and error-free documentation can only be obtained if it is generated automatically. All information about the content and structure is in essence available in the electronic questionnaire. What is needed is a software tool capable of automatically translating this questionnaire specification into a human-readable format.

4? ? A documentation tool

A documentation tool for electronic questionnaire should be able to produce human-readable documentation both in paper and electronic form. On the one hand, this tool must be able to make clear what the global structure of the questionnaire is, and on the other, it must provide means to focus on the details of parts of the questionnaire.

The global structure of the questionnaire can be represented by a routing graph. This is a graph in which each vertex represents a question, and each edge a possible transition to a next question. The challenge of the project is to display the routing graph of large and complex questionnaires. Due to the limited size of a sheet of paper and a computer screen, this is not a simple task. It must be accomplished without affecting the readability. It means a lot of attention has to be paid to layout issues.

The documentation of the questionnaire will be used by different kinds of people involved in the survey process. For many of them, the documentation will need to be in a more accessible form than even the highest level CAI language can provide. Examples of the different types of users include:

- The questionnaire developer, who wants accessible documentation for a variety or reasons, such as the important role it may play in independent testing that the CAI program meets the original specification (it may be cost-effective to employ for this testing staff who lack the skills to understand the CAI program). Proper documentation may also reduce the efforts to be spent on subsequent versions of the survey.

- The customers and managers who have to give formal approval for carrying out the survey. It is not surprising that customers and managers, while welcoming the cost-effectiveness of the transition to CAI, have often expressed their concern at the disappearance of the paper questionnaire. They perceive the paper questionnaire as a readily accessible means to understand what is going in the survey, and are dismayed by the absence of an acceptable substitute in CAI systems.

- The interviewers. It can be difficult for interviewers to get an overall picture of the interview and all its possible routes from running the CAI instrument. A paper document of the questionnaire can be of great help in the interviewer's preparation for fieldwork.

- The data analysts, who need a well-documented codebook if they are to work efficiently. Before CAI, the paper questionnaire was usually the starting point for producing such a codebook.

Different users may require different formats of the questionnaire documentation. Customers and senior managers may wish to concentrate on the questions asked, without getting lost in the detail of the edit checks and interviewers' instructions. Other users, such as secondary analysts working on the microdata, may need more detail. The same users may need more than one format: for example, interviewers may need a summary view of the main topics and their order in the interview and also a detailed listing of the questions. So, the proposed documentation tool must be flexible. The users of this tool must have some control on adjusting the documentation. Research must show what is required and by whom.

The documentation tool must not only display the routing graph, but also additional information. Since each vertex of the routing graph corresponds to a question, information about the question can be display closed to each vertex. Depending on the CAI system used, all kinds of question information can be displayed: the identification of the questions, the text of the question (possibly in different languages), the specification of the type of accepted answers, etc. The available amount of space in the graph is too limited to display all this information. Moreover it might affect readability. Therefore, a documentation tool must provide the means to select the information shown, and possibly also means to display information in different ways, e.g. adjacent to the graph.

The same problems apply to the routing information. It is important to display the conditions determining transitions from one question to another. Such conditions may be quite complex. Solutions have to be found to show this information without affecting readability and interpretability.

A questionnaire documentation tools must be able to generate output in at least two different formats:

- Paper documentation. This is static documentation. Once printed it is not possible any more to manipulate the information. The challenge of the paper format is that two, almost conflicting, goals must achieved. On the one hand, it must be able to show the global structure of the questionnaire in a simple way without being distracted by a wealth of detailed information. On the other, the user must be able to find detailed information about a specific question or route instruction.

- Electronic documentation. Although the small size of a computer screen causes even more limitations then a piece of paper, electronic documentation has the advantage that it can be dynamic. It provides possibilities like zooming in to specific parts of the questionnaire, clicking on vertices or edges to obtain more information, and using hypertext techniques to navigate through the documentation.

A tool for the automatic documentation of electronic questionnaires must obtain its information about the questionnaire from the CAI system used to design the questionnaire. There are several

general CAI systems in use in the world. Examples are Autoquest, Blaise and Cases. Each of these systems uses a different authoring language. Therefore, it is difficult to build a documentation tool capable of reading electronic questionnaire definitions of any CAI system. To solve this problem, a two-step procedure is proposed.

The first step is to design a neutral file format for the documentation tool. The file structure should be such that all information required for documenting the questionnaire is contained in the file. So, it must contain information about both the questions and the routing. The documentation tool will acquire its information from this type of file.

The second step is to create conversion tools for CAI systems. It depends on the structure of the questionnaire specification of the CAI system how complex this job is. Within the framework of this project it is proposed to develop a prototype of one such tool. It is the conversion tool for the Blaise system. The reasons for this choice are that there is extensive Blaise knowledge among the participants in this project, and moreover, the use of Blaise is widespread among the member states of the EU.

4? ? A diagnostic tool for analysing the CAI instrument

A tool like the one described in the previous section could have much more potential then just documenting a questionnaire. It could also play an important role as an instrument to measure the quality and performance of the questionnaire. As a diagnostic instrument, it could be used in the design process. The tool can provide information like

- The number of different possible routes through the questionnaire
- Longest route, shortest route, average route
- The conditions under which specific questions are answered

The last item in the list may provide particularly useful information. It helps to make clear whether questions are indeed answered by the intended respondents. It may also show errors in the questionnaire design, like routes that never will be followed, and questions that will never be answered.

There can also be a role for the documentation tool after the fieldwork has been completed. The tool's importance to data analysts has already been mentioned. One example of a new kind of way that the tool could help analysts to check their data quickly (as an improved way of examining frequencies) could be to compute how many respondents followed certain edges in the routings graphs. Such information can be represented in a numerical way, but also in a graphical way (e.g. by taking the width of the line in the graph proportional to the number of trespassers).

In view of the interesting possibilities, it is proposed to extend the documentation tool with diagnostic functions for analysing the CAI instrument. Hence, it becomes a Tool for Analysis and Documentation of Electronic Questionnaires. The acronym TADEQ will be used for this tool.

5. The project

The TADEQ project is being carried out by a consortium consisting of five partners. The main contractor is Statistics Netherlands. The second partner is the Max Planck Institute in Saarbrücken in Germany. It will be responsible for providing the necessary mathematical knowledge and experience in handling, analysing, and displaying graph-like structures. It is expected to deliver software modules required for graphically displaying the routing structure of an electronic questionnaire. The third partner is the Office for National Statistics in the United Kingdom. This national statistical institute is long time user of CAI techniques. With the help of Statistics Netherlands, it has produced a Blaise Automatic Documentation tool which goes part of the way to answer users' needs; however, much remains to be done. Also, ONS has many contacts with CAI users in Europe and elsewhere in the world. It will be responsible for surveying and defining user requirements, and will co-ordinate evaluation and testing of prototypes. The fourth partner is Statistics Finland. This national statistical institute has substantial experience with CAI techniques, and the development of CAI management systems. Statistics Finland will contribute user requirements and test prototypes. The fifth partner is Instituto Nacional de Estatística in Portugal. This national statistical institute has experience in using CAI for major surveys. It also participates in other European projects in the field of automated data collection (EDI).

To create the TADEQ, a stepwise approach has been implemented:

1? Evaluation of user's experiences and wishes with respect to documentation of electronic questionnaire. This means consultation with several users of CAI systems in Europe.

2? ? Development of the TADEQ neutral file format. A number of major CAI packages will be evaluated in order to make an inventory of what they produce with respect to question and route information.

3? Development of a module capable of producing paper documentation. Initial exploration has shown that this function has priority. A key requirement of Blaise users, particularly the large corporate licence holders, has been for rapid development of a means for producing paper documentation automatically. Early development of a prototype within the TADEQ project will be an efficient means to meet this need and avoid the requirement for a separate project to provide a short-term solution.

1? Development of a module capable of producing electronic documentation.

2? Incorporating analysis functions.

It is important that users of the CAI are involved in the development process. At some points a selection of users will be offered prototypes for evaluation purposes. Their comments and remarks will be taken care of in the subsequent phases of the development process. The project aims for a prototype to be available for testing by the end of the millennium.