A fun with data or how to obtain information in a simple way

Pavel Horovčák[1]

Zábava s dátami alebo ako dostat informácie jednoduchou cestou
The contribution deals with a specification and realisation of universal efficient script in the PHP language. The script is intended for survey results valuation, which are saved in the database table (MySQL). The proposed method enables solution of survey results valuation individually for each question from the questionnaire, and also enables the results decomposition by four another parameters (accordingly another questionnaire question). The method's functionality is illustrated on computer literacy poll and students' readiness for ICT (Information and Communication Technologies) at the start of academic study.

Statistics is like a bikini. What they reveal is suggestive but what they conceal is vital. Levenstein.

Introduction

The use of various questionnaires for finding information and respondents' thought in some matters is standard in many areas. The practice has two phases – data acquisition and data processing. The data acquisition brings some organizational problems or tasks. The data processing (especially in the case of a small survey) is rather difficult. At present, the data processing usually consists in a transfer of questionnaire results into the MS Excel environment and its consequent (manual) processing. At the faculty of BERG, Technical University in Kosice, we proceed in this way, i. e. we observe the computer literacy and the students' readiness to study of the information and communication technologies (Horovčák, 2002). The survey is carried out after finishing the first study year in the study branch Management of processes of retrieval and processing of raw materials. The survey was initially (2001-02) in a written form. In following years, the electronic form was used. The survey results were saved into a database table (MySQL). The creation of efficient (and at the same time universal) script (in the PHP language) for the purposes of survey valuation rises a possibility of simple obtaining not only survey results but also various information, dependencies and trends, which are included in the survey data. Simply – fun with data.

Questionnaire development and evaluation

The survey process conceptualization, organization, sampling, questionnaire design, data collection, data processing, and analysis affect the quality of a survey (AAPOR, 2005). Amount of information on the questionnaire evaluation standards can be found in QUEST (2003).

The problem of questionnaire design and evaluation has many theoretical and practical aspects. Many of them were discussed in frame of the first international meeting devoted to the methods used for the questionnaire development, evaluation, and testing (QDET, 2002). The benefits of iterative questionnaire-evaluation research documents Esposito (2002). The independent and dependent interviewing is compared and valuated by Hoogendorn (2002), who for the interviewing implementation used the Blaise system – (Weerman, 2001, also VIROS, 1999) in a combination with the CentERdata's C2B software. Blaise has historically featured several revolutionary paradigms that make it one of the leading computer assisted interviewing (CAI) systems – (Pierzchala, 2001). Information on the use of Blaise and associated topics is available in proceedings of Blaise user conferences (IBUC, 2001). Blaise is based on Visual Basic and complemented by other technologies, from Delphi pending to PHP and MySQL.

Other known project based on the utilization of Java technology is Decisionarium designed by Hämäläinen (2005). The decisionarium (www.decisionarium.hut.fi) is the first public site for an interactive multicriteria decision support with tools for the individual decision making as well as for the group collaboration and negotiation. It includes the Web-HIPRE (www.hipre.hut.fi) support for the value

[1] *doc. Ing. Pavel Horovčák, CSc.,* Department of Applied Informatics and Process control, Institute of Production Process Control, Technical University of Košice, Letná 9, Košice 04011, Slovak Republic, Pavel.Horovcak@tuke.sk
(Recenzovaná a revidovaná verzia dodaná 17. 4. 2007)

tree and the analytical hierarchy process analysis including group models – (Mustajoki and Hämäläinen, 2000), the Opinions-Online (www.opinion.hut.fi) as a platform for the surveys voting and the group collaboration with an interactive viewing of results – (Hämäläinen and Kalenius, 1999), and the RICH (Rank Inclusion in Criteria Hierarchies) methodology (www.rich.hut.fi) which allows the decision maker to provide incomplete ordinal preference statements when considering the relative importance of attributes in a value tree – (Liesiö, 2002; Salo and Punkka, 2005; as well as Mustajoki et al., 2003).

Assumptions of the designed method – Electronic questionnaire

The method for an effective valuation of survey results is based on the data model structure according to Fig. 1.
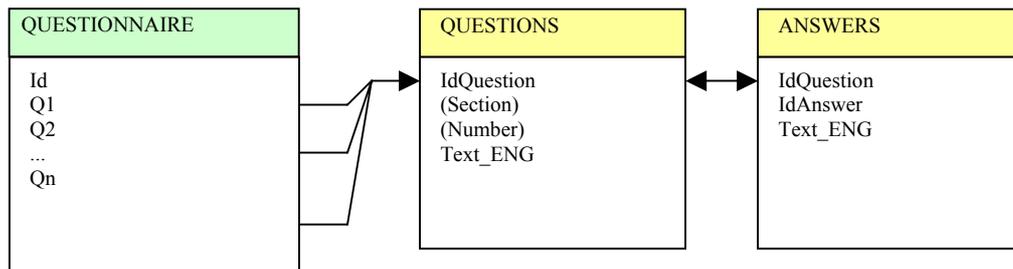

Fig. 1. Structure of the questionnaire data model.

The mentioned data model structure can be realized in two manners. The first one is an application of the mentioned structure to a specific questionnaire design (i.e. questions and appropriate answers), which needs to be patterned and scheduled (repeatedly for every particular survey) – it is a very laborious way. An advantage of such solution is the possibility of its detailed adaptation to requirements and individual questions, eventually answers. The second manner is a realization of interactive development environment (IDE) supporting the electronic questionnaire creation in an arbitrary area. IDE is based on the present data model utilization. Such IDE was created at our department by Horovčák and Orosz, 2004 (using the PHP script language and the MySQL database system) and it has been used successfully in several projects.

IDE enables to achieve the transmission of all questionnaire creation – design of sections, questions' text and type, particular answers as well as all repairs and supplements – straight to the author. The questionnaire creation in this way is essentially simpler and at the same time quicker. IDE hides to the user the whole scale of specialized technological problems, practices as well as solutions from the area of database systems, html technology, interactive scripts php, cascading styles css as well as their mutual communication, and enables him to concentrate on the content aspect of the electronic questionnaire.

The main IDE function – support of electronic questionnaire creation – is solved in the multi language realization. It means that it enables a creation of (the same) questions and answers in various languages. Among partial functions of IDE we can include an addition of further questions, answers list or answer type, a new section and description. Moreover, here belongs editing or deleting of mentioned questionnaire components as well as functions for a questionnaire layout modification. A part of IDE is also the specific group of functions for the standard questionnaire valuation. This group enables a modification of valuation manner for concrete conditions of given project. A meaningful feature of IDE is the possibility of remote project administration and valuation (that means that the location of the project and its database is virtually arbitrary within the internet). Besides the questionnaire creation, IDE supports a creation of the electronic test and electronic inquiry as special types of electronic form.

The operation of IDE is based on a communication with the user by means of the web site with several various forms, through which separate elements of the electronic questionnaire are entered. After their filling and submitting, the questionnaire is executed and specified data (related structure, sections, questions and possible answers) are saved into corresponding database tables.

The IDE enables to create more (whole group) electronic forms. For this purpose, every electronic form creator must sign for the IDE system (login and password). Generating and valuating tables of the form include in their names also particular login, by means of which the individual forms' projects are satisfactorily differentiated. By a suitable selection from the mentioned tables, a page of the electronic questionnaire will be generated.

Principle of valuation method

The designed method enables a solution of the survey results valuation individually for each question from the questionnaire, and also enables the results decomposition by four another parameters (according to another questionnaire question). The solution consists of two steps.

The first step is a selection of questionnaire questions combination, whose mutual dependence is relevant for the user. For this purpose, the user has available a form with five input elements in form of lists. Each input element is filled by all questionnaire questions list whereas the first position (implicit option) is empty by default. The empty position is transmitted as null value into the processing step. The first input element does not content the first empty position. The first input element with the empty first position terminates the specified questionnaire questions combination. In addition, the input form enables the selection of the results' decimal part separator (dot, comma).

The second step consists in a determination of counts of all answers for a specified questions combination, their conversion to the percentage and the output in a table form. This process has as many levels (subqueries) as non-empty values of input elements specified. The subqueries are not supported in MySQL (before 4.1 versions), so that the calculation is solved in the form of nested levels. The calculation algorithm is illustrated by a part of the script code in the php language for one (the first) level. It begins with the determination of the total respondents number (variable $pocet), then follows the statement of selected question's name ($n) by means of a selection criterion ($vyber0) and the introductory part finishes by the specification of selection attribute ($hlp) of the questionnaire table. Afterward follows the calculation of total counts of particular answers to chosen questions, their conversion to the percentage and the output in the table form. The table has two columns which consist of the answer's text and its percentage contribution. In the case of multiple levels (after the calculation and answers' values) the output for the first parameter ($vyber0) is repeated by the same technique for the second level ($vyber1) etc.

```php
$id="Id";                                    //key of table Questionnaire ($tbd)
$sql="select count($Id) from ".$tbd;         //$tb – table Questions
$result = mysql_query($sql);
$poceta=mysql_fetch_array($result);
$pocet=$poceta[0];                           //total records number
$co="Text_ENG";          //attribute name with question's text – table Answers ($tb1)
    $sql="select $co from $tb where ID=$vyber0";
    $result = mysql_query($sql);
    $poceta=mysql_fetch_array($result);
    $n=$poceta[0];                           //name of selected question

$hlp='Q'.$vyber0;                            //attribute name in table Questionnaire
if($vyber0 && !$vyber1 && !$vyber2 && !$vyber3 && !$vyber4){     //only 1. choise
    $q1="select $co, count($id) from $tb1,$tbd";
    $q1.=" where $tb1.IdAnswer = $tbd.$hlp and IdQuestion = ".$vyber0." group by $co";
    @$r = mysql_query($q1);
    print "<table border='1'>";
    print "<tr align='center'><td><b>".$n."</b></td><td>%</td></tr>";
    while ($l=mysql_fetch_array($r)){
        $perc=$l[1]/$pocet*100;
        $cpocet1+=$perc;
        print "<tr><td>".$l[0]."</td><td align='right'>".$perc."</td></tr>";

    //area for selection by next non-empty input elements

    }
    if($cpocet1<100.){                       //questionnaire's data consistence proof
        print "<tr><td> Total ($hlp)</td><td>".$cpocet1."</td></tr></table>";
    }
    else
        print "<tr><td> Total </td><td>".$cpocet1."</td></tr></table>";
    mysql_free_result($r);
}
```

The correct method operation is conditioned by a consistency of the input data. It is highly important to take into consideration this fact already in the phase of questionnaire realization (each input element needs to have a present value – selected). When some question enables an incorrect input (usually the vacant input value), this fact is showed in the results' valuation by the total sum less then 100. In such situation there is together with the question specified also the identifier of "wrong" table's column. Following this, it is possible to achieve an appropriate correction (either to fill up the missing entry or to delete the whole record, which means the deletion of one respondent).

Realization of the method

The method of survey results valuation is realized in the PHP language in the form of two files. The first file realizes the algorithm of multiple selections, their processing and output in the form of table, the second file (db.php) ensures the connection to appropriate database (authentication), operation with corresponding tables and columns of these tables. The second file is the implicit part of the first file (require). In the case of multilingual design of electronic questionnaire, one should only modify the language identification (instead of Text_SVK it will be for example Text_ENG) in the table's attribute name.

For the method realization there are two approaches possible. The first approach creates n input lists (in our case n=5). The second approach creates only one input questions list with a possibility of multiple questions' selection (Ctrl C). From the necessary window surface point of view, the second approach is preferable (more compact). An advantage of the first approach is the possibility of arbitrary questions order selection, which is not enabled by the second approach.

The script of the method has a modular structure and it consists of 9 subsidiary functions, by combination of which the algorithm is realized.

Computer literacy survey at the BERG faculty

The questionnaire consists of three sections of questions. The questions of the first section the specify respondents' sex, region, district, possibilities of access to computers and the secondary school type. The second section of questionnaire is devoted to the issue of ICT knowledge from the secondary school. The questions are oriented to the ICT education length, active working with computer, email using, operating system and programming language knowledge, secondary school connection to the internet as well as the internet services use. In the third section, there is rated an interest in the ICT study at the university, students' sense on the informatics course needful as well as the achieved valuation from this course. The questionnaire questions are illustrated in Fig. 2.

A few results

The functionality and possibilities of the designed method of survey results valuation are demonstrated by a few results of the computer literacy survey at the FBERG in the academic year 2004-05. The following tables show in successive steps the results of the access to computers
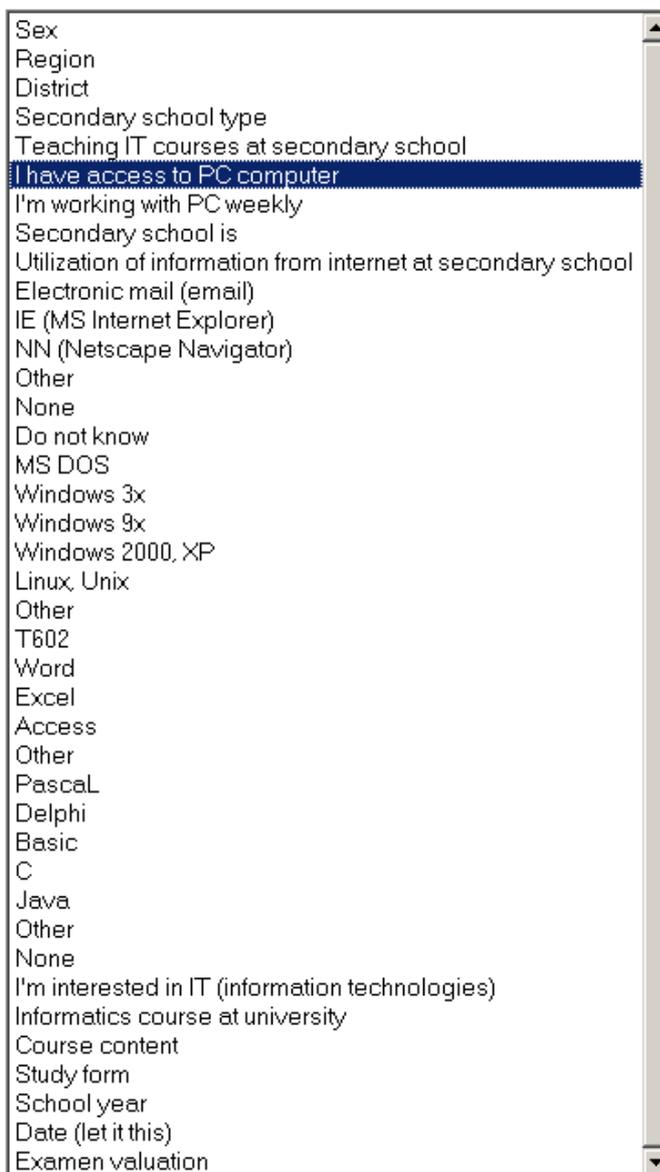

Fig. 2. List of FBERG questionnaire questions.

,and the sex (Vab. **1**), the electronic mail use (Tab. 2) and study form (Tab. 3). Finally, there is a comparison of the students' access to computer during the last four academic years (Tab. 4).

Tab. 1. Access to computers and to computers by sex.

I have access to PC	%
at home	91.58
else	6.32
I have not	2.11
Total	100.00

I have access to PC computer	%	Sex	%
at home	91.58	Female	46.32
		Male	45.26
else	6.32	Female	3.16
		Male	3.16
I have not	2.11	Male	2.11
Total	100.00		100.00

Tab. 2. Access to computers by sex and email using.

I have access to PC	%	Sex	%	Electronic mail (email)	%
at home	91.58	Female	46.32	I do not use	3.16
				I use	36.84
				I use rarely	6.32
		Male	45.26	I do not use	4.21
				I use	36.84
				I use rarely	4.21
else	6.32	Female	3.16	I use	3.16
		Male	3.16	I do not use	1.05
				I use	1.05
				I use rarely	1.05
I have not	2.11	Male	2.11	I use	2.11
Total	100.00		100.00		100.00

Tab. 3. Access to computers by sex, email using and study form.

I have access to PC	%	Sex	%	Electronic mail (email)	%	Study form	%
at home	91.58	Female	46.32	I do not use	3.16	internal (daily)	3.16
				I use	36.84	external	6.32
						internal (daily)	30.53
				I use rarely	6.32	external	1.05
						internal (daily)	5.26
		Male	45.26	I do not use	4.21	external	1.05
						internal (daily)	3.16
				I use	36.84	external	4.21
						internal (daily)	32.63
				I use rarely	4.21	internal (daily)	4.21
else	6.32	Female	3.16	I use	3.16	external	2.11
						internal (daily)	1.05
		Male	3.16	I do not use	1.05	internal (daily)	1.05
				I use	1.05	internal (daily)	1.05
				I use rarely	1.05	internal (daily)	1.05
I have not	2.11	Male	2.11	I use	2.11	external	1.05
						internal (daily)	1.05
Total	100.00		100.00		100.00		100.00

Tab. 4. Access to computers in the last four academic years.

I have access to PC	02-03 %	03-04 %	04-05 %	05-06 %
at home	66.67	84.43	91.58	90.22
else	21.28	8.20	6.32	7.61
I have not	12.06	7.38	2.11	2.17
Total	100.00	100.00	100.00	100.00

Conclusion

The designed and realized algorithm enables a very quick, simple and comfortable acquisition of survey results up to five levels. One can only choose the required dependencies in the input form and the results will be immediately available in a table format. One can also expand the mentioned algorithm to (arbitrarily) higher level numbers. Considering standard (and comparatively laborious) practices of results processed in the MS Excel environment the use of presented method provides a meaningful time saving, simplification of the result acquisition especially for higher amount of levels as well as a segregation of possible error states (at manual practices in MS Excel). In order to transfer the results in to the MS Excel environment, the possibility of a choice of the decimal part separator is also solved (dot, comma). The algorithm has a secondary function – it enables a consistency control of input data. The non-consistency is manifested as a summary percentage value less then 100. The consequent way of input data modification then depends on the user resolution. The usage of built-up script enables a tendency determination, considering one has available the survey results for a few time periods (for example academic years).

The built-up script pair is free of charge and it is available at the author address.

References

AAPOR Best Practices for Survey and Public Opinion Research. [online] [13.12.2005] Available at http://www.aapor.org/default.asp?page=survey_methods/standards_and_best_practices/best_practices_for_survey_and_public_opinion_research.

Esposito, J. L.: Iterative, Multiple-Method Questionnaire Evaluation Research: A Case Study. International Conference on Questionnaire Development, Evaluation and Testing (QDET) Methods, 14 - 17 November 2002, Charleston, SC. [online] [13.12.2005] Available at http://www.jpsm.umd.edu/qdet/final_pdf_papers/Esposito.pdf.

Hämäläinen, R. P.: Decisionarium – Aiding Decisions, Negotiating and Collecting Opinions on the Web. [online] [13.12.2005] Available at http://www.sal.hut.fi/Publications/pdffiles/mham03.pdf.

Hämäläinen, R. P., Kalenius, R.: Opinions-Online - Platform for global participation, voting, surveys and group decisions, v. 1.0 Computer software. Systems Analysis Laboratory, Helsinki University of Technology, 1999. (www.opinion.hut.fi).

Hoogendoorn, A: Evaluation of a Questionnaire Design for Dependent Interviewing in a Web Survey. the International Conference on Questionnaire Development, Evaluation and Testing (QDET) Methods, 14 - 17 November 2002, Charleston, SC. [online] [13.12.2005] Available at http://www.jpsm.umd.edu/qdet/final_pdf_papers/round%20three/Hoogendoorn.pdf.

Horovčák, P., Orosz M.: Web questionnaire – how to do it (in slovak)? Informatika v škole č. 27, 2004, ÚIPŠ Bratislava, ISSN 1335-616X, str. 11 – 19.

Horovčák, P.: The BERG faculty students readiness for information technologies study (in slovak). Acta Montanistica Slovaca, 4, 2002. ISSN 1335 – 1788, pp. 306 – 310 [online] [13.12.2005] Available at http://actamont.tuke.sk/pdf/2002/n4/23horovcak.pdf.

IBUC Proceedings of the 7th International Blaise Users Conference, September 12 - 14, 2001, Washington Marriott Hotel Washington, DC USA [online] [13.12.2005] Available at http://www.blaiseusers.org/ibucpdfs/2001/IBUC_2001-10-9-01.pdf.

Levenstein, A.: Quote [online] [27.7.2005] Available at http://www.global-investor.com/quote/5431/Aaron-Levenstein.

Liesiö, Juuso: RICH Decision – A Decision Support Software. [online] [13.12.2005] Available at http://www.sal.tkk.fi/Opinnot/Mat-2.108/pdf-files/elie02.pdf, 2002.

Mathiowetz, N. A.: Behavior Coding: Tool for Questionnaire Evaluation. Short course presented at the International Conference on Questionnaire Development, Evaluation, and Testing Methods (QDET), 14-17 November, 2002, Charleston, South Carolina. [online] [13.12.2005] Available at <http://www.amstat.org/sections/srms/jsmprg/QDET2002.pdf>

Mustajoki, J, Hämäläinen, R. P, Salo, A.: Decision Support by Interval SMART/SWING - A Method to Incorporate Uncertainty into Multiattribute Analysis. (Manuscript). [online] [13.12.2005] Available at http://www.sal.hut.fi/Publications/pdffiles/mmus03.pdf, 2003.

Mustajoki, J., Hämäläinen, R. P.: Web-HIPRE: Global decision support by value tree and AHP analysis, Journal of Information Systems and Operational Research (INFOR), Vol. 38, No. 3, Aug. 2000, pp. 208-220. [online] [13.12.2005] Available at http://www.sal.hut.fi/Publications/pdffiles/mmus03.pdf.

Pierzchala, M.: Revolutionary Paradigms of Blaise. In Proceedings of the 7th International Blaise Users Conference September 12 - 14, 2001, Washington Marriott Hotel Washington, DC USA [online] [13.12.2005] Available at http://www.blaiseusers.org/ibucpdfs/2001/IBUC_2001-10-9-01.pdf.

QDET International Conference on Questionnaire Development, Evaluation and Testing Methods, 14 - 17 November 2002, Charleston, SC. [online] [13.12.2005] Available at http://www.jpsm.umd.edu/qdet/ qdet-set.html.

QUEST Questionnaire Evaluation Standards. Mannheim: ZUMA 2004, 216 Seiten, ISBN 3-924220-27-1 [online] [13.12.2005] Available at http://www.gesis.org/publikationen/ zuma_nachrichten_spezial/ .http://193.175.239.210/Publikationen/Zeitschriften/ZUMA_Nachrichten_spezial/documents/znspezial9 /gesamt.pdf.

Salo, A, Punkka, A: Rank Inclusion in Criteria Hierarchies, European Journal of Operational Research 163/2, pp. 338-356. [online] [13.12.2005] Available at http://www.sal.hut.fi/Publications/pdf-files/psal05.pdf.

VIROS Virtual Institute for Research in Official Statistics (Eurostat) ETK'99 – Agenda. [online] [13.12.2005] Available at http://europa.eu.int/comm/eurostat/research/index.htm?http://europa.eu.int/en/comm/eurostat/research/c onferences/etk-99/agenda.htm&1.

Weerman, Bas: Internet interviewing using Blaise API. [online] [13.12.2005] 2001, ailable at http://www.blaiseusers.org/ibucpdfs/2001/Weerman--IBUC_Paper.pdf.