Journal of Modern Applied Statistical Methods

Volume 10 | Issue 1 Article 19

5-1-2011

Information Technology for Increasing Qualitative Information Processing Efficiency

S. N. Martyshenko
Vladivostok State University of Economics and Services, Vladivostok, Russia, sergey.martishenko@vvsu.ru

E. A. Egorov
Vladivostok State University of Economics and Services, Vladivostok, Russia, egorov.e.a@gmail.com

Journal of Modern Applied Statistical Methods
May 2011, Vol. 10, No. 1, 207-213

1538 – 9472/11/$95.00

Information Technology for Increasing Qualitative Information Processing Efficiency

S. N. Martyshenko E. A. Egorov
Vladivostok State University of Economics and Services
Vladivostok, Russia

The problem of qualitative information processing in questionnaires is considered and a solution for this problem is offered. The computer technology developed by the authors to automate the offered decision is described.

Key words: Open question, typology consumers, qualitative information processing technology.

Introduction

Questionnaire survey methods are becoming more widespread. This is facilitated in part by the development of democratic processes and a market economy in Russia and other countries. Moreover, institutions of power have shown increased interest in establishing public feedback using the Internet.

Most researchers use simple methods for questionnaire data collection and processing. However, qualitative and poorly structured information processing techniques are common. Martyshenko and Kustov have been engaged in questionnaire data processing computer technology development for more than six years. In their previous work (Martyshenko & Kustov, 2007) they envisaged a computer technology for open questionnaire processing. The technology is based on a type-design practice.

Researchers have achieved considerable results in qualitative information processing technology. Some new elements of the technology are considered here. Qualitative information processing efficiency has been improved through the introduction of a new block, namely, knowledge base represented by three types of dictionaries.

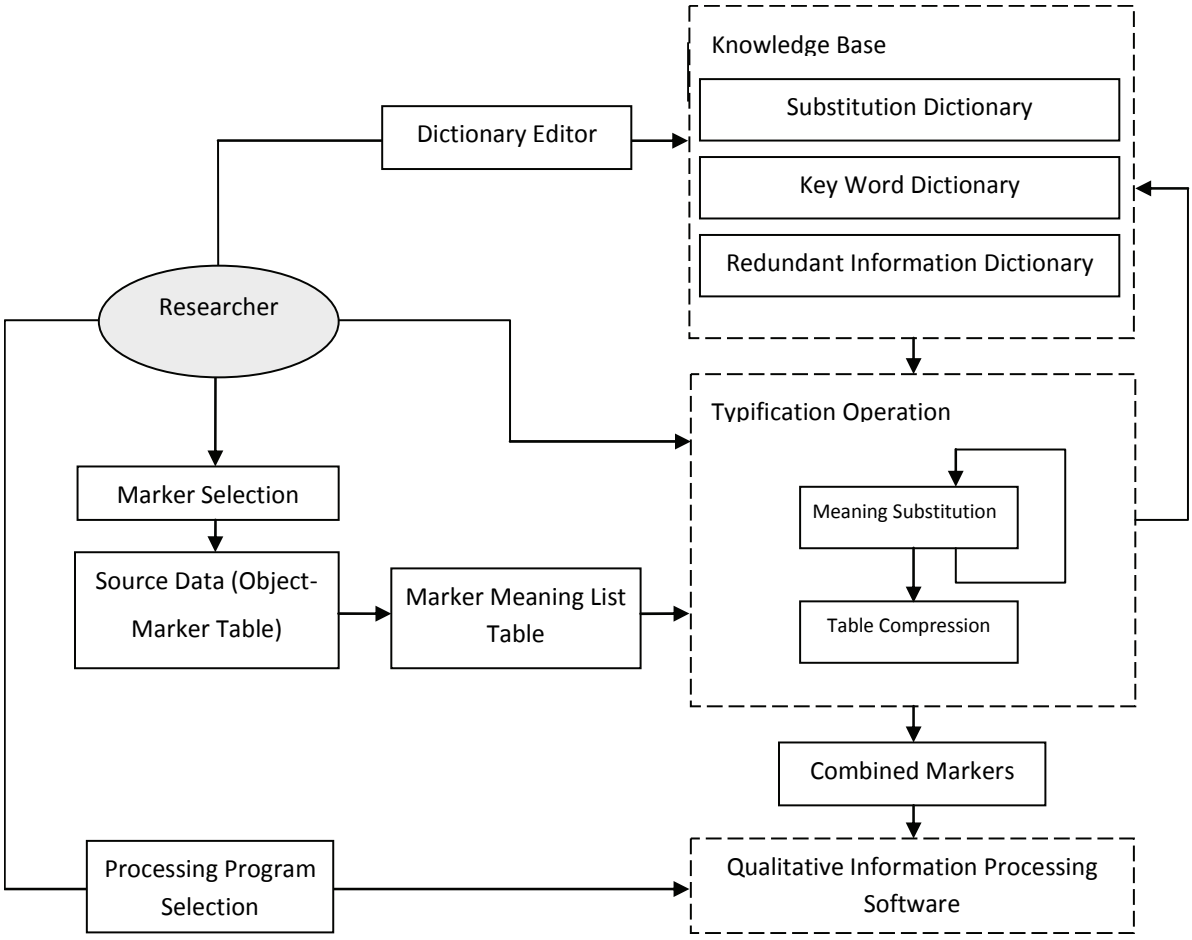
S. N. Martyshenko is a Professor and Department Chair in Mathematics and Modeling. Email: sergey.martishenko@vvsu.ru. E. A. Egorov is a post-graduate student in Mathematics and modelling. Email: egorov.e.a@gmail.com.

Qualitative Information Processing

Consider a general scheme of qualitative information processing technology (see Figure 1). Source data is considered to be represented in an object-marker schedule, regardless of the collection technique used. Besides common properties with only one meaning of a marker under investigation, it is assumed that combined properties can be included into the object-marker table. A combined marker appears when a respondent can choose several answers for the same question. For example, a respondent may mention several cities when asked: *What large cities have you visited within the last three years?* Thus, a combined marker consists of at least more than one simple answer. To identify a combined marker it is necessary to introduce a single divider symbol because a simple answer may consist of several words or even an entire sentence.

Typification, which allows for the quantitative processing, is a practice used for the shift from non-structured data representation to structured representation. A typification operation is a substitution of a source simple statement (in the form of text) by a generalizing statement (in the form of text) with similar or close meaning. The typification operation is performed with the assistance of a tabulated marker meaning list, that is, one of the columns in the table contains all the unique meanings of the source marker.

Figure 1: Qualitative Data Processing Computer Technology Scheme



Although a combined marker is subject to a typification operation, all simple statements in a complex or combined statement are included into the marker meaning list. A marker meaning list table has an automatic filter and contains a column in which meaning frequency is calculated; a typification operation is used for marker meaning list table data rather than object-marker table data. In a typification operation, simple situations are processed first. For example, different spellings of the same word or word order. The best (or correct) form of the statement is selected from among several similar statements and is then copied into marker meaning list table cells containing similar statements. By substituting some unique statements for those existing in the meaning list the number of lines in the marker meaning list table are reduced. After several substitutions have been made, a compression operation can be reasonably used to recalculate the marker meaning list table. Gradually, using this process, the marker meaning list table shrinks to become more demonstrative.

After simple situations have been processed, more complex statements are considered. A group of low-frequency statements related to the same topic should be found in a marker meaning list table. For this group of simple statements a researcher can select a generalizing statement or, if no statement can be found, a new generalizing statement may be introduced to render the general meaning or topic of a group of simple statements. For example, in response to the question, *what else do you like to do during a*

vacation at the seaside besides sunbathing and swimming?, respondents mentioned *raising children*, *raising grandchildren*, *playing with children* and/or *teaching children to swim*. However, these statements did not account for a high percentage of response (less than 0.1%), thus, the generalizing statement *taking care of children* found in the marker meaning list table was substituted. in general, this substitution preserved the meaning of the original statements.

However, it is important not to lose information, particularly when a repeated questionnaire survey is used, for this reason, similar statements for the generalizing ones using a specification in parenthesis are substituted. Specification or nuance is indicated in brackets. For example, in the above case the original meanings were substituted for:

- taking care of children (raising grandchildren),
- taking care of children (raising children),
- taking care of children (playing with grandchildren),and
- taking care of children (teaching children to swim).

The nature of the response, which helps to determine the respondent's (consumer's) personality, is more important than the word-by-word content of the answer. An initial marker meaning list table can contain several thousand meanings, but after processing (typification operation) the table normally has up to three hundred meanings, including specifications. Creation of the table is the final stage in typification (level one). Even if automated, the process is time-consuming and requires the skills and concentration of a researcher. After each session the results are stored and the next session starts from the point where the previous session was stopped.

If a newly designed marker contains two or more meanings it is subject to analysis and requires additional processing (level two). At this stage the specification is excluded and another column, called a subclass, is placed in the marker meaning list table with the number of unique statements reduced to 30 or 50.

Thirty to fifty variants of meaning is a large number for analysis on a nominal scale. Therefore, upon formation of an acceptable list of clearly different meaning variants, a researcher must group the answers in order to consider them properties of separate classes, types or topics depending on the informative meaning of the markers and tasks set developed for the typification operation. In our example determination of personality type would suit better. Combining simple statements into classes is a third level of typification practice. A researcher will introduce the name for each class in accordance with the nature of the statements to be combined. In practice, grouping results achieved by different researchers are very similar. Any differences may be explained by the transitory nature of some statements, which can be attributed to several classes and different researchers can give different names to each class.

Thus, as a result of open question data processing the following output is obtained:

- Three new representations of marker (property) included into the source data table, which can be subject to subsequent processing for obtaining informative conclusions;
- A marker meaning list table which can be used for a repeated questionnaire survey or typification practice with any other questionnaires developed for the process investigation; and
- Knowledge base in the form of three dictionaries: a substitution dictionary, a key word dictionary and a redundant information dictionary.

Dictionaries

It should be noted that, as a result of a typification operation with combined markers, other combined markers are formed. Special processing techniques have been developed for analysis of the latter. The efficiency of qualitative data processing computer technology can be increased by knowledge base creation and use. Computer technologies allowing for knowledge base use belong to an expert system class. The main distinctive feature of an expert system is that it is capable of correct forecasting.

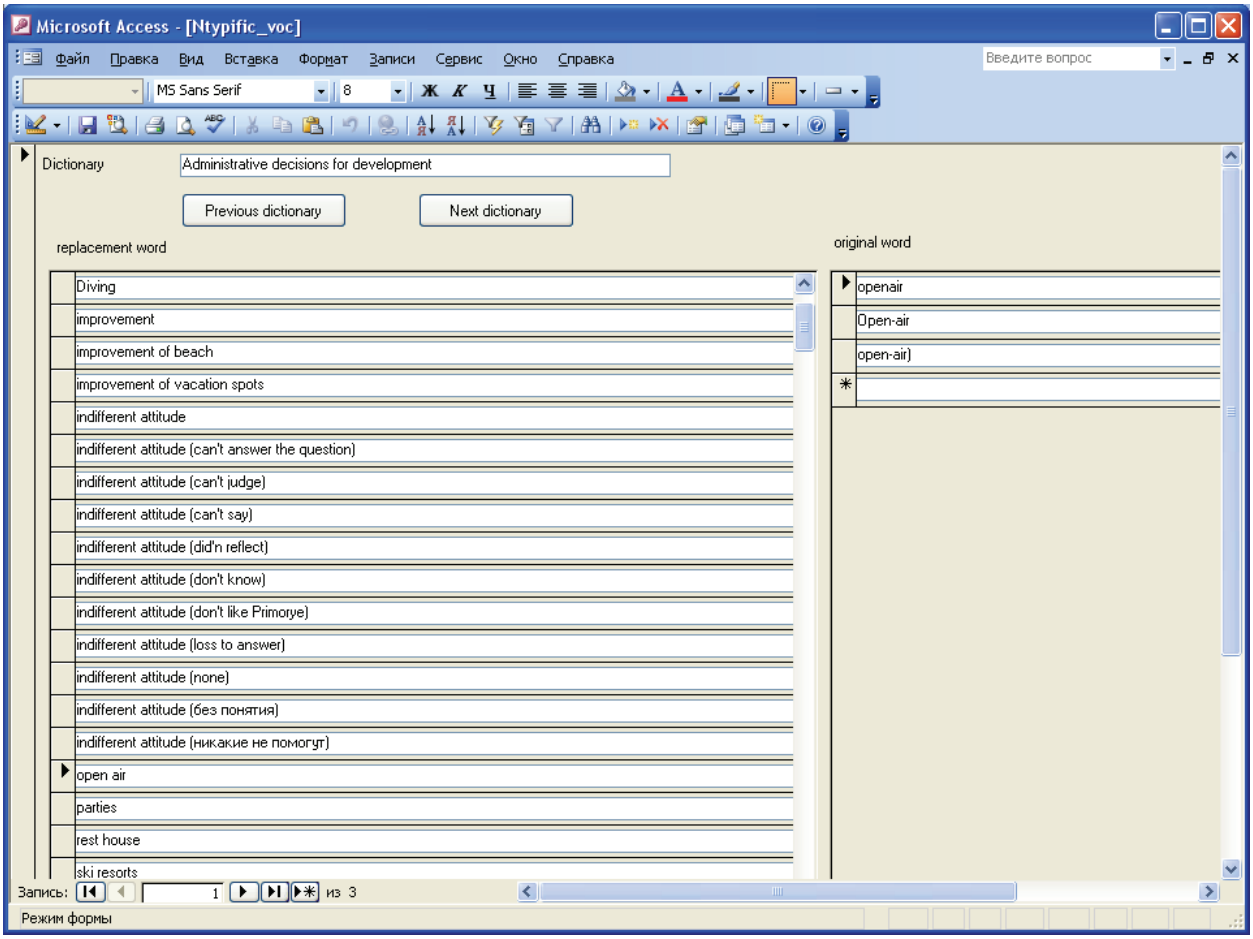
By giving various hints to a user during operation, specialized software can save much time. User hints are generated by special dictionaries, which are formed during typification operations. These dictionaries store user experiences gained through solving qualitative marker typification tasks. Consider the structure and functions of these dictionaries.

Substitution Dictionary

A substitution dictionary is formed automatically when simple statements are substituted for by other simple statements. The substitution dictionary is supplemented when a user works with a typification program; the dictionary stores all substitutions made by the user. All dictionaries are stored in a single Access data base and new questionnaire data requires new a typification operation. This means that a user must choose analogues for the new data. Situations processed by users at previous stages prove to repeat in most cases for the repeated data collection. By activating the substitution dictionary the user can acquire hints for substitution operations; this leaves only situations that have not been recorded before to be processed.

Dictionary support does not require an excess of time, but the dictionary should be reviewed occasionally because, over time, irrelevant variants of substitution collect. Any irrelevant records which accumulate in the dictionary should be deleted because excessive volume of the dictionary can reduce software operation speed. An example of a substitution dictionary is shown in Figure 2.

Figure 2: Substitution Dictionary Example



Key Word Dictionary

Next consider the functions of a key word dictionary. This dictionary is very useful for phrases containing many words. With long phrases the search for proper synonyms in the marker meaning list table can be difficult due to the large number of meanings in the beginning of operation. Because exact matching is a rare occasion, EXCEL filters are often used to identify shorter lists containing certain combinations of phrase elements (see Figure 3). Such phrases are termed key words, although phrase elements can be conditional key words only. The list shown in Figure 3 contains phrases identified by the key word exc. In the reduced list the phrase, *organize tours to preserved areas*, can be substituted for the phrase, *develop a preserved area excursion tour*. The substitution does not alter the meaning of the phrase.

Dictionary operation is based on the extended filter principle, and a key word dictionary includes high-frequency phrases from a marker meaning list table. These phases are provided with a list of key words for a similar phrase search. Users can perform these operations with the help of commands in a special dialog window titled show key word dictionaries (see Figure 4), which represents all functions for key word dictionary creation, supplement and review. Figures 5 and 6 show dialog windows used for supplementing key word dictionaries.

The difference between the dictionary and the extended filter operation is that the dictionary stores key words previously used for a synonym search. The dictionary can be created and supplemented by user only and is intended for future operations facilitation. Once the dictionary contains enough data it can be used as a database. When the dictionary is activated before a typification program is started an additional column is formed in the marker meaning list table, this is called a substitutions from key word dictionary. This column offers phrase substitutions from a key word dictionary that contains phrases chosen by maximum match between source phrase key words and key word dictionary phrases.

Figure 3: Example of the EXCEL User Filter Application for a Typification Operation

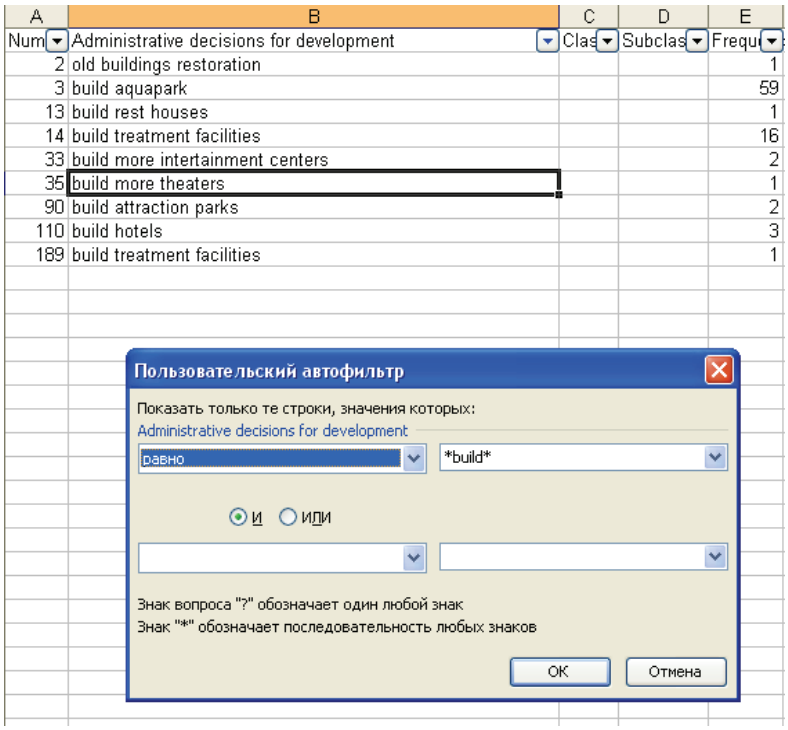


Figure 4: Toolbox for Operating a Key Word Dictionary

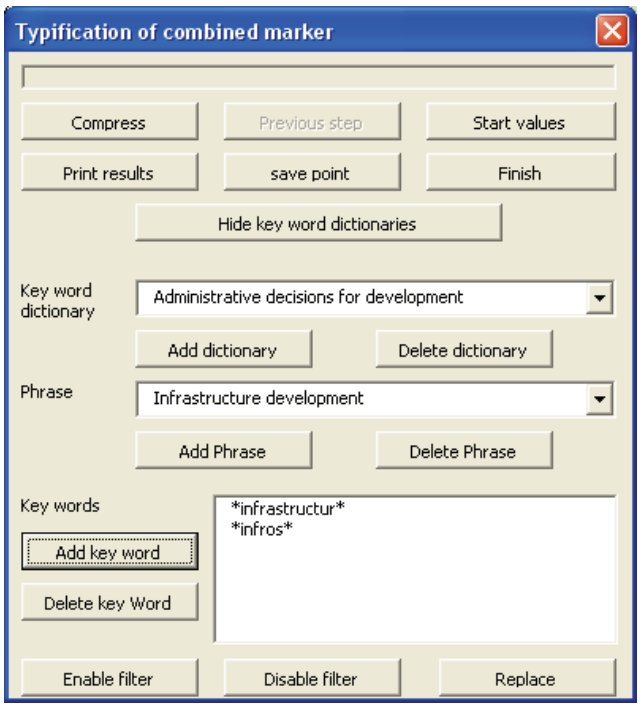
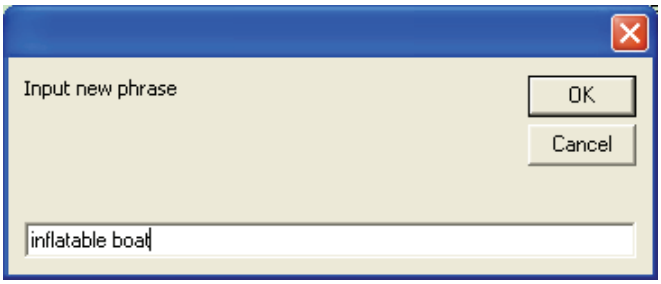


Figure 5: New Phrase Input for a Key Word Dictionary



Redundant Information Dictionary

The third dictionary in the database is the redundant information dictionary. Unlike the two previously mentioned dictionaries developed for each qualitative marker, the redundant information dictionary works with all the qualitative markers and also with different questionnaires. This dictionary is used at qualitative text information processing level one and it helps to delete or edit any statements containing redundant and/or irrelevant information.

For example, the dictionary can be used to exclude such phrases as *I think that*, *in* with typical generalizations, for instance, *isl* or *isl*. in data can be substituted with *island*. The redundant information dictionary has proven to be very useful for long phrase and sentence processing. Using the dictionary the marker meaning list table used in the first stage of operation can be reduced considerably.

Conclusion

The above technology has been tested for over a dozen different questionnaires and has received approval for use. Some practical results of qualitative marker processing can be found in the works of Martyshenko, Martyshenko, and Starkov (2007) and in Martyshenko (2008).

References

Martyshenko, S. N., Martyshenko, N. S., & Kustov, D. A. (2007). EXCEL toolbox for questionnaire data processing. *Information Technologies in Modeling and Management*, *1*, 112-120.

Martyshenko, N. S., & Starkov, A. S. (2007). Analysis of regional consumer tourist market structure. *Territory of Science*, *4*(*5*), 468-478.

Martyshenko, N. S. (2008). Tourist cluster formation and development management in Primorsky region. *Region: Systems, Economy, & Management*, *2*, 122-132.