An Ontology-Based Semantic Design of the Survey Questionnaires

Alexander V. Borodin, Yulia V. Zavyalova
Petrozavodsk State University (PetrSU)
Petrozavodsk, Russia
{aborod, yzavyalo}@cs.petrsu.ru

***Abstract*—Questionnaire is the most common tool of data collection for the purposes of either assessment of the trend of observed parameter of the individual, or statistical survey. Due to the prevalence of mobile devices, the cost of questionnaire-based data gathering from the respondents became extremely low, and thus, collecting of the information can be easily reiterated, providing the deep insight of the studied phenomenon. From the other hand, regular questionnaire-based data collection inevitably leads to the significant growth of the recorded information and to the necessity of automatic processing, e.g., the reasoning on the data to take a decisions. In this paper, the generic ontology for representing questionnaires in machine-readable form is proposed. The presented ontology is intended to use in decision support systems and in personalized smart environments.**

I. Introduction

Almost any subject domain may be described with ontology [1]. For example, GENE ontology refers to its knowledge representation of genomic area of knowledge. SSN ontology answers the need for a domain-independent and end-to-end model for sensing applications by merging sensor-focused, observation-focused and system-focused views.

Prospects using ontologies may be divided into the following categories: better cooperation, data exchange harmonization, specification process formalization, improving of reliability and provision of a multiplicity use.

Ontology approach for data modelling allows to decrease terminological and conceptual confusion and miss of understanding.

Subject domain of surveys was considered in the context of ontology appliance. Surveys are a method of information gathering from individuals. Surveys have a variety of purposes and may be conducted in many ways. Surveys may be conducted to gather information through a printed questionnaire, over the telephone, by mail, in person or on the web, etc.

In this paper structure of survey questionnaires and feedback is described in ontological terms. Questionnaire presentation in the form of ontology provides an opportunity not only to structure the survey data but also to analyse the responses.

II. Motivating use cases

One of applications is health questionnaires. Doctors often carry out surveys about physical and emotional patients health. A survey of the patient may include questions related to its physiological parameters such as age, height, weight, waist size. Due to such questions patient may be correlated with specific group, for example a risk group of obesity. Other questions may relate to symptoms that the patient feels. They allow to estimate its intensity and frequency of their occurrence.

Introduce an example of a survey compiled by World Health Organization [2]. The main objective of the survey is development of integrated response of health care systems to rapid population ageing. Questionnaire is intended to hypertensive patients and contains 23 questions.

Further the types of questions are considered in detail. The patient is required to enter answers in a different form. Choosing the answer to “Question 1” and not finding a suitable one patient may record his own version. “ Question 4” limits the person‘s choice only one statement from the list, because the answers are mutually exclusive. Questions from 10 to 15 suggest to answer positively or negatively. They concern complications and hospitalizations.

The questions are divided into blocks of topics: diagnosis of hypertension, complications and hospitalization, medications and adherence, knowledge and self care. Answers “no” on “Question 12” suggest to go to question from another block and does‘t read the next question.

Another common scope of the surveys is education. Teachers use quizzes to test the students knowledge. An example of the questionnaire ”Literacy and numeracy test for initial teacher education students” [3] shows a variety of possible questions.

In contrast to medical questionnaire, educational questionnaires contain questions with a graphic description especially section “numeracy sample questions”. Tables, charts, graphs and histograms are presented here. Each question is related to them and aimed at obtaining a certain image characteristics. Questions 6 and 8 are interesting because the answer is complex. Students need to evaluate each statement on some scale. Also, there are questions that need to enter the number. In this case, the unit changes and their limits are regulated. For educational purposes questions with 3-4 choice questions with single selection are also important, such as questions 1 to 5 in literacy section.

One more common scope of the surveys is marketing and customer satisfaction surveys [4]. Effective businesses focus on creating and reinforcing pleasurable experiences so that they might retain existing customers and add new customers. Properly constructed customer satisfaction surveys provide the insights that are the foundation to creating and reinforcing pleasurable customer experiences.

ISSN 2305-7254

The most useful questions in this area are measurement scales. They use to categorize customers (nominal scale) and differentiate customers along some continuum (interval scale). The ranking of companies is best accompanied by evidence that the ranking reflects meaningful, statistically significant differences between companies.

These domains questionnaires are different in content but similar in structure. It enables to structure and classify the survey‘s questions. Ontological modelling technology allows to the most efficient submit these data as well as to provide their a semantic binding.

III. Review of questionnaires ontologies

Simple survey ontology is presented in [5]. A quiz is defined, as a set of questions designed in order to evaluate knowledge and pulls into Semantic Learning Content Management System (SLCMS). This quiz ontology is intended to build semantics metadata that express the question for the student profile and learning path.

The system described in [6] uses survey questionnaire (SQ) ontologies as knowledge base, the patient medical profile to be stored in a semantic way and an inference mechanism to infer data in the decision making process. The proposed ontology for decision support system model helps to predict the risk of hypertension and diabetics in related diseases.

These articles describe the structure of the questionnaire only in isolation from user responses, as well as questions are not classified. It doesn’t give the possibility an opportunity to provide a semantic linking of questions and user responses.

IV. Questions classification

Questionnaire is a special method of survey using form. Way of its presenting may be different but number of question‘s types is formalized.

Benefits of questionnaire are time and cost effective [7]. Researchers don‘t have to plan survey sessions with the software developers to manage them. The respondent may pass a survey at any time, regardless of the survey developers. Web questionnaires facilitate obtaining data received in electronic form. They may also easily collect data from a large number of respondents in different geographical locations.

Since there is no interviewer, ambiguous and poorly worded questions are problematic. This fact is a survey‘s disadvantage. Despite the fact that the questionnaires are relatively easy to create by a developer, it takes a time [7].

Social science has formulated a set of questions types: open-ended, closed-ended and combined.

A. Open-ended questions

Sociologists use open-ended questions presented at Fig. (a) 1, that don‘t envisage tips. Using open questions the respondents have an opportunity to express their views freely and fully, and sociologist may gather a wealth of information.

The main disadvantage of the open questions is the difficulty of formalizing and thus further processing. Respondent is guided solely by their own ideas. Naturally, the answers are unique and diverse. Open-ended questions are used in cases of complete information about the views of the respondent. Open questions may be asked on direct, but it may be difficult to obtain information about the motives of human behaviour, the causes of certain acts or certain views. The consumer may close and not express his opinions honestly. For example, he hesitates to speak negatively about something as subconsciously don‘t want to offend the researcher. Open-ended questions development based on the projective method allows to avoid this protection [8].

Projective method involves setting of unstructured and indirect questions, asks respondents to express their ulterior motives, beliefs or feelings about the problem under discussion. According to the classification proposed by V. Gordon and R. Langmaid [8], projective questions are divided into the following groups: associative, on completion of the task, design and expressive.

Associations are used to find the ratio of a particular product, to find out what associations are arise. This approach facilitates the release of the respondent by logical contradictions and directs its energy to find emotive and colourful models. Questions on completion of the task consist in asking to complete the pending proposals, drawings, stories. Respondents express more clearly they attitude to something, and the researcher may select the brightest and most common themes, therefore to infer about the hidden feelings and motives of consumer behaviour. Expressive questions are focused on the definition of consumer‘s emotional perception of something. They may detect not only the attitude to the last, but the images associated with this product in his minds.

B. Closed-ended questions

Closed-ended question asks the respondent to select the appropriate answer from offered or evaluate any specified parameters, characteristics, properties, etc. Differently closed-ended questions call multivariate. The closed-ended questions are single choice and multiple choice.

Single choice questions are used for allowing a single selection of an option from a pre-defined set of answers. A multiple choice question gives the participant a list of options, of which he may select one or more options. Having a clue, the respondent becomes easier to answer the question if a corresponding version of his opinion or position is offered. However, the desired variant absence leads to a distortion of information. For the construction of multivariate questions using a scales: nominal, ordinal, interval, ratio, semantic difference, Likert, Guttman and etc.

Stanley Stevens came up with terms nominal, ordinal, interval and ratio scales in 1946 [9]. This scales classification has become a major in the construction and processing of multivariate questions.

Nominal scale has only a description of the characteristics and represents the simple listing of options, without any hint of ordering and comparing. The numbers serve only as labels or tags for identifying and classifying objects. Such questions is assumed to be that as one answer at Fig. (b) 1 as well as some of the presented options at Fig. (c) 1 may be selected.

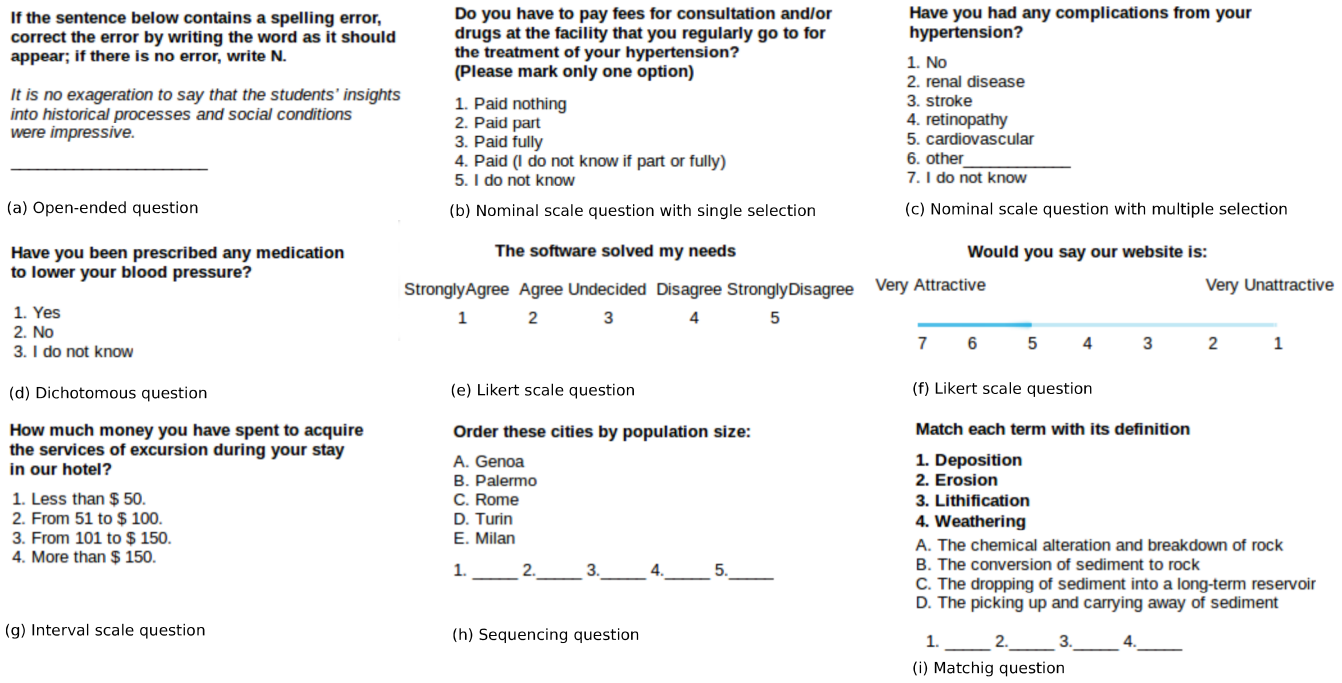

Fig. 1. Question examples

Ordinal scale consists of categories, which differ from one another by conventional notions or qualitative characteristics, in which the presence of ordering implied in any form. Located in order in the first place object is more strongly expressed response compared to the second, etc. Ordinal scales are commonly used in the formulation of questions about relationships and estimates (eg, from ”very good” to ”very bad”), the general satisfaction or dissatisfaction.

The semantic differential is a series of bipolar characteristics that determine the properties of the object being studied. The semantic range consists of a large number of pairs of opposites (such as ”bad - good”, ”comfortable - uncomfortable”, ”useful - useless”, ”cheap - expensive”, ”like - not like”, etc.). Boundary values are plotted on a scale, and the space between them is partitioned into seven ranges (from -3 to +3, or from 1 to 7) characterizing the degree of proximity of opinion to a particular definition.

If only two responses are possible, such question is called dichotomous at Fig. (d) 1. Surveys often use dichotomous questions that ask for a Yes/No, True/False or Agree/Disagree response. This type of questions has two opposite responses. One is necessarily true and the other must be false. Its basis is completed declarative sentence and answers are reduced to the substitution of the words ”yes” or negation ”no” instead of the interrogative particle.

Likert scale was developed in 1932 by D. Likert [10] for assessment the extent of agreement or disagreement with each judgement at Fig. (e,f) 1. Likert scale is, in fact, a type of ordinal scale, but it is usually considered separately. In practice, five-digit Likert scale is widely used. Two intermediate approval ”agree” and ”disagree” are introduced in addition to the two extreme (”strongly agree,” ”strongly disagree”) and neutral (”I may not say”, ”difficult to answer”). Statements numerical designation is applied. Its value corresponds to the degree of relationship assessment. Thus, absolute agreement (”strongly agree”) estimated the value of ”5”, and the radical opposition (”strongly disagree”) corresponds to the assessment of ”1”. Often, each item of the questionnaire is accompanied by a visual analogue scale, on which the response options are painted.

Interval scale consists of numerical values that are amenable to physical measurement at Fig. (g) 1. A classic example of such a temperature range (between the numerical values of the interval is known and has a constant size, for example, the temperature in degrees Celsius). Similarly, the questionnaire on the age of respondents, income, expenses, prices are usually presented pre-breaking measurements on some areas.

Interval scales are good because they are easy to statistical analysis of data. Processing of measurement results on an interval scale allows you to determine ”how much more”, but not suggests that one measured in so many times larger or smaller than the other [11].

The peculiarity of interval scale is that the zero point is chosen arbitrarily [11]. The estimated property of the object or event does not disappear when the measurement result is equal to zero: 1) calendar time (the beginning of each calendar in chronology set differently), 2) the zero temperature, etc.

This scale has a well defined null point, and therefore, at some time the measured quality may be zero. Through this relationship the scale doesn’t impose any restrictions on the use of mathematical apparatus. In assessing the results of measurements may be determined by ”how many times” one object more than another. Ratio scales provide a wealth

of possibilities when it comes to statistical analysis. These variables may be meaningfully added, subtracted, multiplied, divided (ratios). Central tendency may be measured by mode, median, or mean; measures of dispersion, such as standard deviation and coefficient of variation may also be calculated from ratio scales.

Answering Guttman scale question the respondent checks each option with which they agree. The options themselves are constructed so that they are cumulative – if respondent agree to one, he probably agree to all of the ones.

C. Sequencing question

A sequencing question on the other hand asks survey respondents to compare a list of different objects to one another by placing them in order of preference at Fig. (h) 1. It may effectively define knowledge in a logical sequence construction, process flows, the execution of any algorithms procedures, the chronological sequence construction. The structure of the sequencing question includes the instructions "Set the correct sequence", introduction part of the question, a list of items that need to be streamlined. Man should put a serial number to each element of the list according to the logic of the question content. A single matching as well as multiple selection may be provided in such questions.

D. Matching question

The essence of matching question is the following at Fig. (i) 1. This type of question is intended for establish a correspondence between one element of the plurality of properties, events, objects, definitions, etc. and one member of another set. Question examines the ability to find connections and associations between phenomena, events, processes, structural units, etc.

The correspondence between the elements of two columns may be one to one, when each element of the right corresponds to only one element left. If the number of elements in the two columns is the same, the selection isn't made for the last element of the left column. Determines by the subject's specificity content cases occur when elements number on the left column are selected the same items on the right, so they may be less than the left. Finally, the question is optimal if the right set contains more elements. In such case, each element is chosen only once.

Apart from the structural approach of classification questions may also be classified according to the content, form, function. Concepts for such classifications may also be included into ontology.

V. The Proposed Questionnaire Ontology

A. Modelling the structure of the question

The design of survey questionnaires have two objectives. The first objective is the development of a survey questionnaire structure. Such a universal representation of the data questionnaire allows to develop a variety of its visual representations.

The modules contain the classes and properties that may be used to represent particular aspects of questionnaire: for example answer and its type, connection between question ans response.

Fig 2 provides an overview of the ontology structure at the generalized level of description: the questionnaires themselves may have particular properties, such as an language and identification code of questionnaire on the remote storage. *Questionnaire* class is the root element of the ontology that identifies a particular questionnaire in the intellectual space.

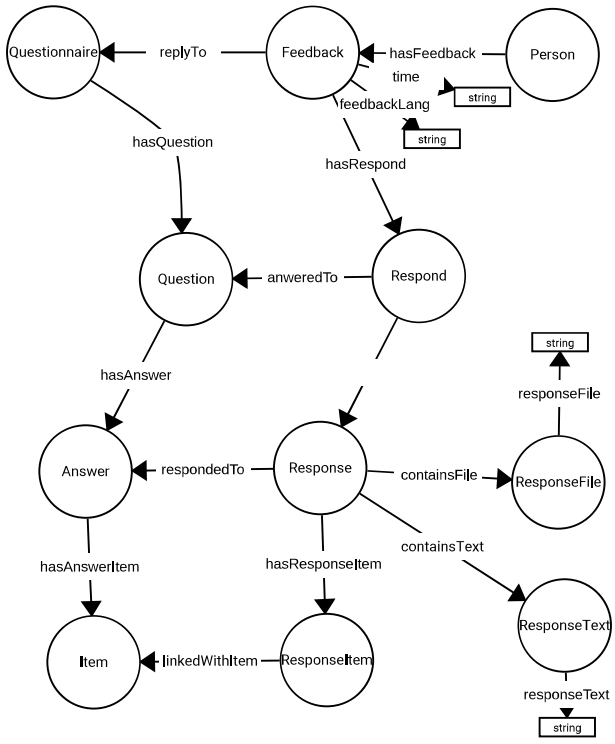

Fig. 2. Questionnaire ontology fragment

Questionnaire consists of a set of questions, each of them is represented by an instance of the *Question* class and associated with questionnaires by object property *hasQuestion* while the first question is determined by object property *hasFirstQuestion*. Order of the following questions is described by *nextQuestion* object property. The question itself is characterized by *description* property which describes the question's task.

Class *Answer* is an abstract entity, through which it possible to determine what the user's response should be and in what form variants of the answer (if set) should be described. As it was previously convicted in Section IV, questions may be classified as open-ended, closed-ended and combined. In the case of the combined question, it may contain two or more types of answers. If the type of question is defined at the level of class *Question* , the described above case is became impossible. Therefore, the question's type is described on the *Answer* class through its binding by object property *hasAnswerType* to *AnswerType* class. All versions of the question's classification may be determined by the *AnswerType* class.

Item class is required to determine answer's options for closed-ended questions. It contains the *itemText* data property to display and the *itemScore* data property which may be used

in different ways for various types. Text description of options is written in the *itemText* property for each type of *Item*.

To avoid ambiguities in questions types determining it is need to be specify in the ontology. Ontology is the proposed in Section IV classification. First type is divided into closed-ended and open-ended types. As mentioned earlier, the combined type is not necessary to allocate separately. Several different types of *Answer* class is necessary to define to a single instance of *Question* class.

Using the classification proposed in the Section IV question's separation by type presented in Figure 3. *Open* class defines an open type of answer. It is presented separately, that makes impossible to add an *Item*. The other types presuppose the existence of an instance of *Item* class. *Sequencing* class represents sequencing question. *SequencingItem* class associated with it defines one element of the sequence. Inherited from an *Item* class *itemScore* property specifies the ordinal number of the element in the sequence.

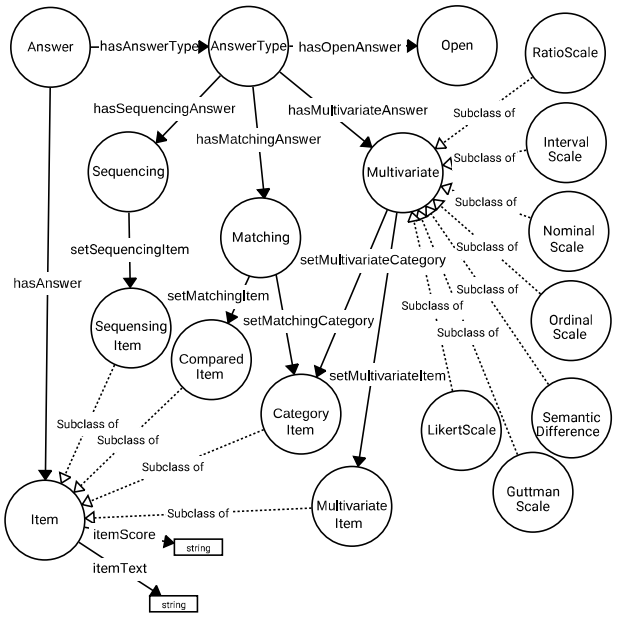

Fig. 3. Answer types ontology fragment

Multivariate class differs from other types of questions by the presence of two types of *Item*. *MultivariateItem* class describes the options which are necessary to evaluate. Scale's type is inherited from *Multivariate* class. Scale's elements are described by *CategoryItem* class. Its order determined by the *itemScore* property. Comparative question (*Matching* class) is described by analogy with multivariate question. Only in this case *ComparedItems* class means a set of category options, which are compared and *CategoryItem* class - comparative options.

This set of classes is sufficient to describe the questionnaire in general terms. Using *AnswerType* class classification of questions is defined.

Feedback ontological representation is necessary to define for the semantic processing of responses. This part of the ontology is designed to generate responses written by the respondent. *Feedback*, *Respond* and *Response* classes are associated with the *Questionnaire*, *Question* and *Answer* classes respectively to determine what type and what question is. *Response* class describes what information the respondent has introduced. This may be a file (*ResponceFile* class), or/and the text (*ResponceText* class), or/and selected or mapped answers (*ResponceItem* class). *ResponceItem* class instance is join with a corresponding instance of *Item* class.

The second objective is to provide a questionnaire and user responses semantic binding for further analysis of answers according to the subject domain.

B. Modelling order of the questions

From the point of view of the survey respondent, the questions usually form a sequence and simply follow one after another. In the ontology, the sequence of questions could be expressed with direct linkage such as *hasNext* object property of the *Question* class. But this approach possesses a number of shortcomings. In our model, the question itself has no knowledge of other questions. In addition, many surveys involve branches to navigate the respondent through the questionnaire in an efficient way [12]. The example of questionnaire for hypertensive patients, discussed in Section II, also contains branching, namely, the respondent skips several questions about admission to the hospital in case it didn't happen.

Difficulties in describing sequences in OWL are discussed by Drummond et. al. [13]. They proposed a general design pattern for modelling sequences using OWL-DL. In terms of questionnaire design, the pattern is shown in Fig. 4.

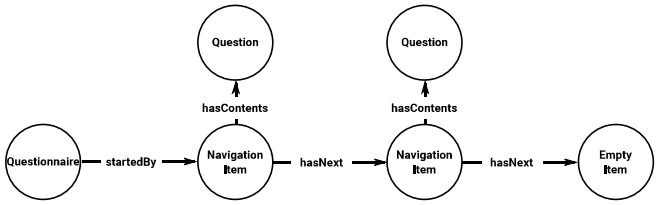

Fig. 4. Modelling a sequence of questions with no branching

We have adopted this pattern with several improvements, as it is shown in Fig. 5.

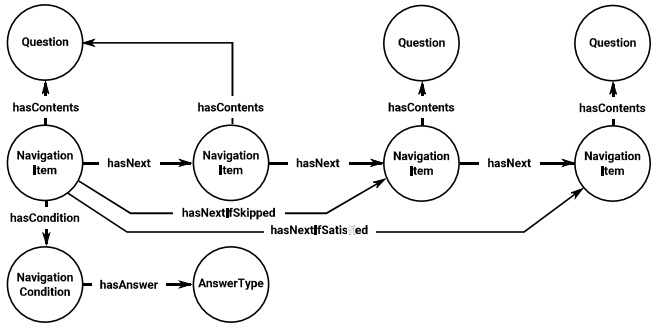

Fig. 5. Modelling a sequence of questions with conditional branching and skipping

Firstly, the question can be marked to be skippable. The respondent is has the right to omit answering it and to proceed to the question specified with *hasNextIfSkippable* object

property. Secondly, we introduced the *NavigationCondition* class to provide branching support. If the *NavigationItem* has a connected *NavigationCondition* then it is checked and, if satisfied, the question specified with *hasNextIfSatisfied* is selected as the next one. Otherwise, the *NavigationItem* pointed by *hasNext* is processed. Since the next *NavigationItem* can be linked with the same question, the cascade of condition checks can be described by the model. The simplest and most common example of *NavigationCondition* is the choice of a certain possible *AnswerItem* in a closed (most often – dichotomous) question, e.g., "if the answer is positive, then omit the next question". The concept of separate *NavigationCondition* provides the possibility for further extensions.

VI. Conclusion

In this paper, we focused on the problem of semantic representation of questionnaires for the purpose of further automatic processing. We constructed the generic ontological model of questionnaire, which provides a possibility of question structure description including complex questions with a set of answers of different kind. We also proposed the ontological representation of question order, including skipping and branching.

Currently, the simplified version of the proposed model is implemented as a part of digital assistance service in medical emergencies for recording of the symptoms by the patient itself or bystanders.

Acknowledgement

This research is financially supported by the Ministry of Education and Science of the Russian Federation within project # 14.574.21.0060 (RFMEFI57414X0060) of Federal Target Program "Research and development on priority directions of scientific-technological complex of Russia for 2014–2020".

The study is financially supported by Russian Foundation for Basic Research 16-07-01289.

References

[1] B. Chandrasekaran, J. R. Josephson, and V. Benjamins, "What are ontologies, and why do we need them?" *IEEE Intelligent systems*, no. 1, pp. 20–26, 1999.

[2] World Health Organization, "World health organization developing integrated response of health care systems to rapid population ageing," [retrieved: Oct. 2016]. [Online]. Available: http://www.who.int/ageing/projects/intra/phase_one/alc_intra1_questionnaire_hypertensive.pdf

[3] ACER, "Literacy and numeracy test for initial teacher education students," [retrieved: Oct. 2016]. [Online]. Available: https://teacheredtest.acer.edu.au/files/Literacy_and_Numeracy_Test_for_Initial_Teacher_Education_students_-_Sample_Questions.pdf

[4] B. E. Hayes, *Measuring Customer Satisfaction: Survey Design, Use, and Statistical Analysis Methods*. ASQC Quality Press, 1998.

[5] R. Balog-Crisan, I. Roxin, and I. Szilagyi, "Ontologies for a semantic quiz architecture," in *2009 Ninth IEEE International Conference on Advanced Learning Technologies*, July 2009, pp. 492–494.

[6] P. C. Sherimon and R. Krishnan, "Ontodiabetic: An ontology-based clinical decision support system for diabetic patients," *Arabian Journal for Science and Engineering*, vol. 41, no. 3, pp. 1145–1160, 2016.

[7] T. C. Lethbridge, S. E. Sim, and J. Singer, "Studying software engineers: Data collection techniques for software field studies," *Empirical Software Engineering*, vol. 10, no. 3, pp. 311–341, 2005.

[8] W. Gordon and R. Langmaid, *Qualitative Market Research: A Practitioner's and Buyer's Guide*. Gower, 1988.

[9] S. S. Stevens, "On the theory of scales of measurement," *Science*, vol. 103, no. 2684, pp. 677–680, 1946, [retrieved: Oct. 2016]. [Online]. Available: http://science.sciencemag.org/content/103/2684/677

[10] R. Likert, "A technique for the measurement of attitudes." *Archives of Psychology*, vol. 22, no. 140, pp. 1–55, 1932.

[11] A. P. Shcherbakov, *Mathematical and statistical processing of materials of scientific and methodical work*. Foresight-Russia, Rybinsk: YSPU named after K.D.Ushinsky, 2007.

[12] K. Norman, "Implementation of conditional branching in computerized self-administered questionnaires," University of Maryland, Laboratory for Automation Psychology Technical Report (LAP-TR-2001-1), 2001, [retrieved: Oct. 2016]. [Online]. Available: http://hcil2.cs.umd.edu/trs/2001-26/2001-26.pdf

[13] N. Drummond *et al.*, "Putting owl in order: Patterns for sequences in owl," in *OWL Experiences and Directions (OWLEd 2006)*, Athens Georgia, 2006, [retrieved: Oct. 2016]. [Online]. Available: http://ceur-ws.org/Vol-216/submission_12.pdf