Research Technology Based on Open-Ended Questions: The Key Modules, Contribution of Software Discant

Galina Saganenko[1], Alexey Geger[1], Kirill Boyarsky[2], Victoria Dudina[3], Elena Stepanova[4]
[1]The Sociological Institute of the RAS – Branch of the Federal Center of Theoretical and Applied Sociology of the Russian Academy of Sciences (SI RAS-FCTAS RAS)
[2]ITMO University
[3]St. Petersburg State University
[4]Baltic State Technical University VOENMEH
St.Petersburg, Russia
saganenko.selina@yandex.ru, ageger@gmail.com, boyarin9@yandex.ru, victoria_dudina@mail.ru, Len.stepanova@gmail.com

Abstract–**This paper is of interdisciplinary matter: to explore subjective social areas and to develop digital instruments for them. We have two final objectives: To offer a new effective research technology based on a system of open-ended questions for complicate sociological area; To offer its key module – a computer systems DISCANT and VEGA for storing empirical data and analyzing them by effective methods of machine examination of texts. We present a special type of research, namely, full-scale Reflective Integrated Research Technology based on open-ended questions (RIRT).Our mission and tasks are to show the effectiveness of the developed resources, their high social, cognitive, technological and humanistic features to pay attention to the lack of such resources in the investigative field of sociology, to identify the causes of this poor situation and, creating an attractive format, make proper efforts to change the cognitive situation. Finally, to transform our developments into open public resources. Our developments at the moment contain the following units: Concept of research technology; multiplicity of open-ended question types; A mass of reflective interrogation techniques relating to many areas of personal and public life; A mass of implemented projects at home and abroad supported by Russian and foreign funds; The format of group surveys ensuring the anonymity of participants, the reliability of the results; Computer systems DISCANT and VEGA for the systematic support of reflective research; System for creating classifiers and classification of text arrays; Comparison of the basic modules of the three main types of survey research.. We would like to try on the possibilities of creating an information product for integrated description of our multidimensional, multicomponent technology, similar to what we met in the development of Dmitry Korzun: Semantic infrastructure of a smart museum: toward making cultural heritage knowledge [1]. As for our complicated situation it is related to corresponding presentation of research instruments, research methods and research results and, much important, in a number of global social areas.**

I. Introduction

At the moment, there are three types of sociological research: quantitative, qualitative, and mixed qualitative-quantitative (mixed methods research). In quantitative research, the sociologist deals with numerical data at the input and output. This type of research is the most common in sociology, not least due to the reliable computer support of, for example, products like SPSS, Statistics, and R. Classical qualitative research deals with large unstructured texts. Here, analyst is helped by such products as MAXqDA, Ethnograph, NVivo, and a number of other similar developments in functionality. The main goal of these programs is to help the analyst pass through unstructured data by creating special notes; the classification is mostly done manually and, of course, Natural Language Processing is out of the question here. Mixed methods research combines all the above mentioned approaches.

Our technology is significantly different from all three currently existing sociological research strategies. Our surveys are conducted in large samples, but they are based on open-ended questions, which allows us to neutralize the directive role of the researcher and get the respondent's relevant opinion on various issues of interest to us.

As a result, we get a large number of short textual judgments that have a more organized structure than in qualitative research. Thus, on the one hand, we have the relevant opinion of the respondent (advantage of the qualitative approach), and on the other hand, we have large statistics (advantage of the quantitative approach).

In this article, we will consider the main modules of our research technology, including copyright developments in the field of Natural Language Processing - DISCANT and VEGA software.

Open-ended questions are not systematically used in sociological research. One of the reasons for the refusal to work with them, the authors stress call encoding difficulties [2-4]. While analyst assistance programs (such as Nvivo and Atlas.ti) have long existed for “large”, qualitative research, open-ended questions, as Roberts claims, were almost universally encoded manually until recently [5]. In 2014, Roberts proposed his structural-thematic model for processing answers to open-ended questions, but it was based on a “bag-of-words” and was not applicable to the processing of Russian-language texts. If for English texts to assess the semantic affinity of sentences, tools such as the WordNet thesaurus are widely used, for the Russian language the corresponding resources are rather scarce. To solve these difficulties, we use the "hybrid" technology in VEGA and DISCANT systems, combining semantic-syntactic analysis of text and content analysis.

Our research technology is an interdisciplinary product, and we put forward the following range of tasks: 1) to prove the effectiveness and heuristic potential of RIRT in relation to sociological research; 2) to demonstrate effective algorithms to improve machine understanding of the text; 3) to interest the IT community in the application of its technological developments to solve promising tasks in sociological research. If the interaction of specialists from two different branches of knowledge in the field of statistical processing of sociological research data has been ongoing for a long time, then cooperation in the field of big data, data mining and social mining has begun relatively recently and we expect significant progress in the implementation of joint projects of sociologists and IT specialists to solve non-trivial interdisciplinary tasks.

II.The specifics of the source material inthe study based on open-ended questions

Let us single out the general characteristic features of the array of primary information based on open-ended questions:

- This is predominantly high-quality (textual) material, accompanied by numerical ratings;

- This textual material as a whole is quite simple - it is judgments or assessments of individuals regarding given social objects, a kind of “nomination” of objects; in each response of the respondent, usually several (up to 5-8) such judgments are presented; the complete answer to a separate question, as a rule, is fairly easily divided into independent judgments ("phrases"), while the general part is repeated in each proposition, the totality of such phrases is then processed;

- In relation to a separate open-ended question, this is a fairly large amount of statistical material in a sense of the same type, which can be processed statistically (phrases), which allows comparisons and hypotheses; accordingly, many open surveys in each questionnaire at times increase the material for analysis;

- This is a plastic material, self-adjusting to the specifics of the situation in society and in different social groups;

- This is material that carries the signs of a specific time and reflects the specifics of social groups and individuals, representing, in a certain sense, historical value - for real representations of real people related to real time are fixed in this study;

- This material allows one to compare different structural ideas about different objects (such a possibility is completely absent in the "closed" question).

- This material reveals a fundamentally different idea of the meaning of empirical data in sociological research. This information is not about the respondents as such, but a syndrome of several interpenetrating entities - the spirit of the existing era, the characteristics of the social object, the specifics of the mass consciousness, the nature of the audience, the specifics of the social group, the possibilities of the method, the professional experience of the researcher, etc.;

- This material has the advantages of data of both types of studies (qualitative and quantitative) and occupies an intermediate position between complex textual material, for example, an array of interviews or biographical narratives, and digital material of a standardized survey. This type of material provides arguments of both types, in particular, there is more “evidence-based” processing of “qualitative” textual data, there is less “research arbitrariness”; conclusions from this material are clearly visible and easily verifiable (since the study completely classifies the entire textual material, rather than selective use of utterances in the interpretations of the researcher). Textual judgments are supported by numerical estimates.

III.Relevance of the open-ended methodfor studying the social situation

Elaborating our earlier developments [6], [7], we systematize research situations in which, from our point of view, this method of open-ended questions is most relevant, where it should give adequate results, clarify the specifics of social realities. So, the method is relevant in the following circumstances.

1) The method of open-ended questions represents ahumanistic perspective in sociology, when participation in the study is also beneficial for the respondent. The method enables a person to understand the realities of his own life and the world around him, formulate his own ideas, gain experience in primary conceptualization and articulation of social phenomena in a significant area for him.

2) The method of open-ended questions has amulticomponent effect in teaching social disciplines and in communicating with classrooms. The student formulates his own independent representations and then moves more meaningfully along the subject and disciplinary field. The teacher gets the opportunity to understand the students' ability to reflect on the surrounding society and their own lives and formulate their own judgments on the essence of social conflict. The teacher also receives many primary conceptualizations and, as a result, the primary body of information to advance in an empirical clarification of the problem. There is a lot to discuss with students.

3) The method is relevant and effective if the study is aimedat a very broad object, the search for its principal components (for example, “the situation in Russian society”, “ecological situation in Europe”, “changes in the life of the respondent due to changes in the country”, etc.). Such collisions usually cannot be closed, and they should not be closed for various reasons: too many-dimensional object (listing here positions of different volumes and different significance could take more than one page), too much load on the researcher when searching for a list of positions; “closing” questions - this is, in effect,

imposing on respondents, as a rule, a narrow and biased “concept” of a social object.

4) The method is adequate if the study is carried out duringthe period of social transformations - at such times, many social realities are either momentary, temporary, or currently not yet stable. So in many cases there is neither time nor sense to wait for certainty. Open-ended questions “self-adjust” to the situation that has changed.

5) If the study is panel-based, regular, especially carried outduring a period of social change - “imposing” hard identities on the social reality is completely illegal, moreover, it will not allow comparing different moments of changing historical time.

6) If different socio-cultural groups are compared, that havesignificantly different social statuses, specifics of life experience, different experience of social reflections, etc., this method will allow to obtain relevant social information, compare and detect differences, which, as a rule, are always practically significant on any social object and perception of the social situation. For example, only this type of research reveals that workers and intellectuals have fundamentally different, virtually disjoint systems of significant preferences in the field of work. You can track how the comfort / discomfort of the school or family environment changes as the child grows up, etc.

7) If the respondent is more qualified than the researcher inthe studied subject area, and then the respondent already acts as an “expert” and he “identifies” and “qualifies” the content of the studied social situation / conflict / object. The task of the researcher is to propose an adequate tool for the "presentation" of the statements of experts and then cope with their analysis. For example, readers of the National Russian Library in St. Petersburg, who once a year take part in our study, act precisely as such a collective expert in actualizing and identifying changes in the world, country, and St. Petersburg. Their aggregate representations are undoubtedly much richer and more penetrating than the representations of one researcher and even a research team;

8) If a standardized study on a “hard” questionnaire is beingprepared (especially a massive and responsible research), then using open questions will help to select questions and their “closure” for a future questionnaire more competently;

9) If you have trial ideas for future research, or even there isjust left a piece of free space in the questionnaire - two lines with a relevantly formulated question can give you good material for thinking about a future project.

There are other research situations where the use of open-ended questions is useful, appropriate, interesting, etc.

IV.THE PLACE OF OPEN QUESTIONS IN THEMETHODOLOGY OF SOCIOLOGICAL RESEARCH

Meanwhile, we regret to note that the method of open-ended questions has not yet received an independent research status in the methodology of sociological research; methodologists have significant problems with updating the research potential of this method. Let us denote a number of contradictory collisions that were discovered in studies on the possibilities of open-ended questions; we are based on copyright developments of reflexive technology of open-ended questions, on the results of many completed projects within the framework of this technology. All these problems are connected with each other and complement each other.

Let 's look at the first problem in relation to open-ended questions methodology. By this we mean the vagueness of the methodological status of an open question. On the one hand, researchers give an open question an independent place in the methodology, indicate that open questions can collect information about important respondents' problems in the form of qualitative data [8, 9]; on the other hand, when assessing its cognitive potential and data quality, the criteria commonly applied for closed questions are used. In this context, the problem of non-responses to an open-ended question is studied in detail [10].

In their work, L. Miller & Amber D. Dumford studied in detail the relationship of non-responses with the socio-demographic characteristics of respondents; among other recommendations for the use of open-ended questions, the authors advise not to start the questionnaire with an open-ended question; that is, an open question is considered in the framework of a standardized questionnaire, as an addition to closed questions [10]. But a single open-ended question among a total set of closed-ended questions, especially in a standardized interview, is a special methodological situation that prevents the respondent from seriously looking for and formulating his own ideas. Our questionnaires mainly consist of open-ended questions, and the respondent predominantly searches for his own answers to the systemic set of questions, without wasting the effort to switch from one format to another.

Let 's consider the another problem in relation to the heuristic potential of open-ended questions. By this we mean a consideration of the cognitive potential of open questions in terms of standardized research. An open question is compared to a closed one [2], while indicating that open-ended questions provide a “more diverse set of answers” - and this is their plus; at the same time, open-ended questions give smaller percentages and they are much more difficult to code - and this is their big minus [2]. Almost all authors point to the difficulties of coding open-ended questions as their weak side.

Perhaps the problems of coding, including automatic coding of responses, are the leading problem for methodologists; a number of researchers devote separate works to the problem of coding open-ended questions [3], [4], [11], [12]. However, our developments in the field of machine understanding of the text, as well as the general development trends of Natural Language Processing, give hope that such an effective tool as open-ended questions technology will finally take its deserved place in the research arsenal of sociologists.

V. Methodological modules

The following are explanations of the methodological modules providing RIIT.

(1) REFLECTIVE TYPE STUDY CONCEPT. It reflects all sections, in comparison with each other and in aggregate constituting a holistic and effective tool for empirical research (see paragraphs 2-10 below).

There is no idea of the type of empirical research in the domestic methodology, it is mainly discussed the idea of a research approach that represents a fairly broad unstructured methodological perspective, reducing it to the problem of a dichotomous choice between a qualitative and a quantitative approach. Another system into which the presentations of Russian methodologists fit into is the “process of sociological research”, but there are no developments under the name “type of sociological research”.

(2) TYPOLOGY OF OPEN-ENDED QUESTIONS: we have developed more than 25 types of open-ended questions, each of which has special cognitive properties, is attractive to the respondent, organizes a person’s efforts to sort out his own ideas about the essence of the object or problem being studied, and provides realistic and transparent primary information. And since systematic development of a typology of open-ended questions is still not available in publications, we are able to offer a description of a significant package of different types of open-ended questions and related methodological problems. We have no opportunity to present here a maximum of open-ended questions (which would require a lengthy article). So for this publication, we will focus on only one type of questions - mixed questions.

(3) OPEN-ENDED QUESTIONS OF A MIXED TYPE - this is one of the most effective types of questions that provides both the identification of the qualitative characteristics of the studied objects and their quantitative assessments, which makes a significant contribution to the development of the reflective capabilities of respondents and students. In recent years, researchers have been trying to combine the virtues of qualitative and quantitative research, developing the idea of mixed research. However, these are essentially two autonomous technologies that are superimposed on different sides of the object and thus, it turns out, are dealing with different objects.

It should be noted that researchers have no experience in developing formats for such questions, apparently without assuming that both tasks can be formulated in the structure of a separate open-ended question. Therefore, researchers turn to so-called "mixed studies" using different research methods - formalized and qualitative, however, more time-consuming and resource-intensive, and generally dealing with different objects.

(4) REFLECTIVE SURVEY METHODS. Since domestic and foreign methodologists discuss only individual open-ended questions, examples of holistic reflective techniques are still not presented in publications. Thus, the task of our project - to ensure a systematic acquaintance with the potential, structure and advantages of reflective survey methods - seems extremely relevant.

(5) ADVANTAGES OF GROUP SURVEYS. We will present the rationale for the relevance of the method of obtaining data through group surveys. To begin with, during the time of ideological pressure in the countries of communist ideology, Bulgarian colleagues were the first to study the potential of group polls, developed the technology of group polls, showed that this format provides sufficient anonymity for respondents and allows to get more relevant answers that are not loaded with fears of possible consequences for their frank judgments and diagnosis of social problems. But even in our time there are a lot of sensitive problems and situations of interest to society, in relation to which the respondent would prefer to maintain his anonymity, which in general is precisely what a group survey provides. Anyway, the researcher is interested in receiving “clean” unbiased data from respondents.

In addition, a group survey has a lot of other indisputable advantages: it ensures the uniformity of the sample of respondents for a variety of factors, comfortable survey conditions (usually a classroom or lecture hall), time, financial and other expenses for collecting information are reduced by several times, the unquestionably existing influence of the interviewer, and moreover of masses of various interviewers, is leveled. So group surveys generally provide higher quality and relevance of primary survey information, etc.

(6) POTENTIAL OF DISKANT / VEGA COMPUTER SYSTEMS. It is relevant to present the integrated computer system DISKANT developed within the framework of the technology, and VEGA system (developed for research in English). The system provides compact storage, retrieval and analysis of information for arrays with multiple textual, mixed and digital attributes. It allows to combine and separate arrays, switch from array to array, create autonomous classification systems for judgments from different surveys, use the classification modules of relevant questions from other studies, conduct a simultaneous analysis of the totality of the corresponding arrays, and compare the structure of text fields (see the next paragraph). In general, the DISCANT provides the researcher with comfortable and effective conditions for interacting with the mass of his empirical objects, while maintaining the database, identifying and refining the results. Undoubtedly, the presentation of an effective national analytical system DISKANT / VEGA is relevant for the development of science and solving applied problems.

(7) ANALYSIS OF TEXT ARRAYS**.** The identification of stable generalized characteristics of objects occurs through the analysis of primary textual judgments using various dictionaries, and most importantly, the classification system - an effective, verifiable, correctable and, as a result, highly optimized judgment classification system. It is developed in an iterative mode - allowing one to move in portions towards understanding the studied object, presented in the judgments of the respondents and from simple and widespread ideas, leaving the most complex formulations and the most complex of the raised ideas “for a snack”.

In each reflective technique there are several studied objects, regarding which the mass of respondents express their opinions. Their analysis is initially carried out in a simpler way - using dictionaries, or by developing a more subtle and evidence-based classification system that provides a lot of additional features. Thus, the implemented classification system allows you to then make comparisons of the structures of text fields for different objects and objective factors. For example, to compare 6 sets of judgments in answers to questions about positive and negative changes in the World, in Europe, in Russia.

Or, for example, one can get structures of judgments: *what for* each of parents usually praise / abuse boys and *what for* each of parents usually praise / abuse girls in the family - that is, compare the structures 4 and 4 of the array of judgments.

One can create any "secondary variables" and compare them if their judgments are classified using a single classifier, adding any number of differentiating bases. For example, we can compare the rubrics of judgments “Why (what for) parents (usually) scold children of elite schools” and “Why (what for) parents (usually) scold children of ordinary schools” and see this in tables or on bar charts of DISCANT.

(8) The relevance of the proposed systematic presentation of different methods of analyzing text arrays is determined by the fact that the principles, methods and results of the analysis of information obtained in answers to open-ended questions are minimally discussed by methodologists and seem to be simple. Large survey centers sometimes lay out unsophisticated “lines” of percentages on primitively grouped answers to a separate question.

(9) The quality and evidence of the RESULTS should be noted. Using the analytical resources of the DISCANT program, one can trace all the stages of obtaining the final results, their transparency, evidence, and verifiability. Moreover, thanks to the group survey, we are well aware of the nature of the audience who shared with us, the researchers, their ideas about the problem / object or the conflicts of their own lives. What cannot be said about the reliability of the results of mass, so-called representative surveys, - this problem was carefully studied by us [13].

One can also note such an important quality of group surveys as the “targeting” of the audience, which provides the researcher with the opportunity of repeated meetings with the audience. So if we interviewed, for example, the 8th grade of the 11th gymnasium, then we can easily find this group and either tell them about the results, or conduct a survey on other topics, or interview the group regularly, for example, six months later, to identify changes .

According to RIRT, we have implemented more than 30 research projects, for lack of space we will not describe the most remarkable results for each of them, but focus on one thing - the study of life values.

We have been conducting life value research for over 15 years. During this time, many surveys were conducted: students of the "techies" and humanities were interviewed; pupils of prestigious schools and ordinary schools; French and Russian students; residents of megacities and small towns. In all cases, we found confirmation that the open-ended methodology responds very sensitively to differentiating factors: gender, age, university specialization and school rating, country of residence and the size of the settlement; as a result, it was always possible to identify any new collisions. More detailed results of our first research on values by the method of open-ended questions can be found in [14].

In addition, studies are longitudinal in nature. Open questions are very sensitive to time, to the era. So, in 2005 there was no cheap Internet available. Therefore, these conflicts are found in the answers, but rather rarely. The next poll was in 2010, when the whole country found out who Pavel Durov was and many became users of the social network Vkontakte. There was a staggering bias towards social networks in the responses of the respondents. This was no longer in the 2015 survey because it had become commonplace. In 2015, many respondents declare democratic values, the values of civil society, which was not so clearly seen fifteen years ago simply because, for example, there was no figure of Alexei Navalny. Thus, the methodology constantly reveals changes in society, and often those that are difficult to predict for the researcher in the office when developing survey tools. For example, the World Values Survey, the largest in sociology of values, which now covers more than 100 countries of the world, offers the same list of 8 values for evaluation for people of different countries and cultures; it has been offered unchanged over the past 30 years. As it is not difficult to guess, it is rather difficult to identify any new significant phenomena with such an approach.

VI DIGITAL OPPROTUNITIES AND CHALLENGES

The situation with the lack of widespread use of open-ended questions in sociological research has led to the lack of adequate technological solutions. If there are a lot of computer programs for analyzing qualitative data, that is, “large texts”, then analysis / coding of short textual judgments, as Roberts states, is almost always done with the help of a person [5].

It should be noted that programs for the analysis of quality data practically do not use automatic text classification systems. The program acts as an assistant to the researcher; roughly speaking, this is the analyst’s “electronic notebook”. Most often, the functional of the programs is created as follows: databases are created with the transcript of the interview (most often, although audio- and video-graphic files can be downloaded), and the analyst goes through these "large", lengthy, unstructured texts and manually sets certain notes. This process is presented more clearly in the German MAXqDA program: the analyst goes through the text and manually assigns codes, which then can also be manually assigned to classes or subclasses; the program has many "chips" that facilitate the work of the researcher: for example, the function "color attribute", which contributes to greater visibility and visualization of the analyzed material. However, there is no question of transferring the main work of coding to a machine. There is an option for semi-automatic classification in MAXqDA, but classes must be set by the researcher first.

The AutoCoding function is also present in the popular Nvivo program, the essence of this function is to put the various answers to one question into one “catalog box”, and then view all the answers for a specific question in one node.

Atlas.ti developers position the program as an ideally equipped workspace for the researcher. It allows you to work with a huge amount of different format material - you can choose from more than 25 formats for the downloaded file, including even geographical maps. At the moment, this is the most “modernized” program, it is also released as a mobile application in the AppStore. However, working with the classification of text attributes in this program is a rather laborious process: our experiment showed that automatic coding and manual coding takes the same amount of time [15].

Not earlier than in 2014, Roberts and Stewart proposed their “Structural-thematic model” for analyzing answers to open questions [5]. At the same time, in Russia at SPbEMI and ITMO University (Boyarsky K.K., Kanevsky E.A.) such developments have been underway since 1994. The development of Roberts and Stuart is based on the technology of the "bag-of-words", which is not applicable to answers in Russian and irrelevant to our tasks.

An automatic analysis of the content of open-ended questions in Russian runs into serious difficulties due to the variability of the word order in sentences and a high degree of homonymy. Widely used technologies such as "bag-of-words", in which the word order is not taken into account at all, are applicable for the statistical processing of large volumes of texts. Obviously, based on the requests of sociologists using open-ended questions, this approach does not satisfy, and therefore requires the use of more subtle methods of machine linguistic analysis. At the same time, respondents' answers to questions are often not comprehensive Russian language sentences. For example, the answer to the question “For what / whom are you responsible” may look like “I am responsible in life for myself, my family, friends,” or just like “For myself and for my cat.” This makes it difficult for the parser to build a statement model, for example, a subordination tree.

The second problem is the semantic classification of responses. In this case, the lexical closeness of statements determined on the basis of any vector models is not enough [16]. If for English texts to assess the semantic closeness of sentences, tools such as WordNet are widely used, then for the Russian language the corresponding resources are rather scarce. Recently, Russian-language information resources that allow you to “attach” semantic information to terms and phrases begin to be created [16-19]. However, the task is significantly complicated by the fact that the fundamental differences in the word formation mechanisms of the Russian and English languages make “tracing” English-language resources impossible and require considerable work to clarify the structure and content of thesauruses, ontologies of subject areas, etc. The efficiency of using semantic tools and choosing the optimal structure for the diverse text mining tasks in specific subject areas require further research.

To solve the problems of analysis and classification of open answers, we use the "hybrid" technology in the VEGA [20] and DISCANT systems. At the first stage, a morphological and, if possible, syntactic analysis of the answer is carried out with the aim of removing homonymy as much as possible. At the same time, a semantic class is assigned to each word according to the classifier, which is an extended version of the Tuzov classifier [21]. Although, as was shown in [22], this allows one to successfully find even low-frequency terms in specialized texts, this is still not enough for high-quality processing of sociological materials.

Therefore, at the next stage, the technology of content analysis is connected, in which not words are considered as an element of content, but specially highlighted normative phrases, which can consist of a whole sentence, as well as several words and even one word. Each such phrase is an expression of one judgment, one thought. The formation of a set of normative phrases can be performed both in manual and in automatic mode. And, finally, the answers are linked to normative phrases with subsequent classification. In this case, not only lexical, but also semantic (classes) information is used.

A feature of this technique is that the structure of the classifier is not compiled in advance, but is formed directly in the process of analyzing the array of answers. In this way, it differs from the currently existing developments: in MAXqDA, classes are initially defined by the analyst. The primary version of the classifier in VEGA can be formed completely automatically. In the future, the structure of the classifier is specified in iterative mode in the process of interaction between the researcher and the computer.

Thus, the essence of the proposed method is that instead of 100% classification of all answers, only normative phrases are classified. This gives a double advantage. Firstly, the amount of classification work is reduced. Secondly, with any change in the classifier — and this is not an exception, but the rule when analyzing texts — it is enough to change the class and group of a number of normative phrases so that corresponding changes automatically occur in all answers associated with them.

As indicated above, a sociological study consists of questions of various types, suggesting numerical answers (age), textual (your attitude to ...) or mixed types. The tools we use allow us to carry out a comprehensive analysis of the compatibility of answers of different types, and to provide a clear visualization of the results to the researcher.

VII.Conclusion

So, to summarize - what we reported in our article, what we own: (1) an integrated RIRT research system, equipped with all the necessary functional subsystems; (2) a host of methodological results; (3) the mass of research-covered spheres of private and public life and implemented empirical projects; (4) a comfortable and productive learning environment for students and their results.

At one time, we had significant developments in the field of standardized research, development of scales, widely used the psychological methods of Rokich, Jenkins, Likert, Thurstone and others, developed a lot of causal, regression models, etc. We published two monographs [23], [24], some significant sections in collective monographs. However, this intensive work in the standardized paradigm did not bring satisfaction, first of all, because of the questionable base of primary information (we always checked the data by different methods for stability with the test – re-test and it creeped significantly). However, we note that the content and methodological fields of Russian sociology are not sufficiently developed. A significant proportion of the reasons we see in the total orientation of researchers to the two technologies mentioned. Therefore, our mission is to introduce into the scientific and educational turnover? the developed and "fully packaged", comfortable, transparent, evidence-based RIRT technology.

In the second semantic module of our paper, we proposed a "hybrid" technology for coding answers to open-ended questions, implemented in DISKANT and Vega computer systems; Our technology takes into account the peculiarities of working with Russian-language texts and is used in two stages - semantic-syntactic and content analysis. In general, automatic coding procedures built on this "hybrid" technology have worked well, and DISCANT and Vega are invaluable assistants in the work of the analyst, reducing the routine work, leveling technical errors in coding, thereby increasing the reliability of the results of a sociological study.

ACKNOWLEDGMENT

The reported study was funded by RFBR according to the research project №19-011-00905.

REFERENCES

[1] Korzun, D., Varfolomeyev, A., Yalovitsyna S., Volokhova V. "Semantic infrastructure of a smart museum: toward making cultural heritage knowledge usable and creatable by visitors and professionals", *Personal and Ubiquitous Computing*, vol. 21, issue 2, pp 345–354, 2017.

[2] Reja U., Manfreda K.L., Hlebec V., Vehovar V. "Open-ended vs. Close-ended Questions in Web Questionnaires", *Developments in Applied Statistics*. Anuška Ferligoj and Andrej Mrvar (Editors). Metodološki zvezki, 19, Ljubljana: FDV, 2003.

[3] Popping R. "Human or Machine Coding of Open-ended Questions", *Bulletin of Sociological Methodology, Bulletin de Méthodologie Sociologique*. Vol. 115, pp. 79–88, 2012.

[4] Popping R. "Analyzing Open-ended Questions by Means of Text Analysis Procedures". *Bulletin of Sociological Methodology, Bulletin de Me´thodologie Sociologi*que, Vol. 128, pp. 23–39, 2015.

[5] Roberts E.& el. "Structural Topic Models for Open-Ended Survey Responses", *American Journal of Political Science*, Vol. 58, No. 4, pp. 1064–1082, 2014.

[6] Saganenko, G.I. "General Methodology", *Bulletin of Sociological Methodology, Bulletin de Me´thodologie Sociologique*, no.68, pp. 79-80, 2000.

[7] Saganenko, G.I. "Centre: Course Syllabus - Empirical Knowledge in Sociology: Studying Research Principles and Approaches, Cognitive Opportunities of Methods and Specificity of Obtained Results" | [Schéma de cours - Conaaissance empirique en sociologie: L'étude des principes de recherche et les approches, les possibilités cognitives des méthodes, spécificité des résultatsobtenus], *Bulletin of Sociological Methodology, Bulletin de Me´thodologie Sociologique*, no. 59, pp. 62-80, 1998.

[8] Geer, J. G.. "Do open-ended questions measure "salient" issues? " *Public Opinion Quarterly*, no. 55, pp. 360-370, 1991.

[9] Krosnick, J. A. "Survey Research." *Annual Review of Psychology*, no. 50, pp. 537–67, 1999.

[10] Millar, M.M. and D.A. Dillman. "Do mail and internet surveys produce different item nonresponse rates? An experiment using random mode assignment", *Survey Practice* vol. 5, no. 2, 2012.

[11] Popping R. *Online tools for content analysis*. In: Fielding, N.G., Lee, R. & Blank, G. (eds.), Handbook of Online Research Methods. London: Sage, 2017, pp. 329-343, 2017.

[12] Klein S.P. "Characteristics of hand and machine-assigned scores to college students' answers to open-ended tasks", *Probability and Statistics: Essays in Honor of David A. Freedman* Vol. 2, pp. 76–89, 2008.

[13] Saganenko G.I. *Reliability of the results of sociological studies*, Moscow, 2017.

[14] Geger, A.E. "The Detection of Individual and Group Values in Young People. Relevant Methodological Solutions", *Russian Education and Society*, Vol. 53, no. 1, pp. 60-79, 2011.

[15] Geger A.E., Chupakhina Yu.A., Geger S.A. "Computer programs for the analysis of qualitative and mixed data", *Petersburg Sociology today. Collection of scientific works of the Sociological Institute of the Russian Academy of Sciences* - St. Petersburg: Nestor - History, 2015. - P.374-383.

[16] Malykh, V. "Robust word vector for Russian language", *Proceeding of AINL FRUCT 2016 Conference*, Saint-Petersburg, Russia. FRUCT Oy, Finland, pp. 95-98, 2016.

[17] Thesaurus RuTez // URL: https://www.labinform.ru/pub/ruthes/index.htm (available: 9.07.2019).

[18] N. Loukachevitch and B. Dobrov. "The sociopolitical thesaurus as a resource for automatic document processing in Russian Terminology". *International Journal of Theoretical and Applied Issues in Specialized Communication*, vol. 21, no. 2, pp. 237–262, 2015.

[19] I. Shchitov et al. "A Survey On Thesauri Application In Automatic Natural Language Processing", *Proceeding of the 21st conference of FRUCT association* Saint-Petersburg, Russia. FRUCT Oy, Finland, 2017, pp. 296-303.

[20] Saganenko G.I., Kanevsky E.A., Boyarsky K.K. "The contexts of empirical knowledge in sociology and the possibilities of the VEGA program", *Telescope:The Journal of Sociological and Marketing Research*. SPb. 2008, No.6. pp. 43–55.

[21] Tuzov V.A. *Computer semantics of the Russian language*. SPb: Publishing House S. Petersburg. University, 2004.

[22] Archakova N., Boyarsky K., Kanevsky E. "Extraction of low-frequent terms from domain-specific texts by cluster semantic analysis*", Proceedings of the ISMW-FRUCT 2016*. Saint-Petersburg, Russia. FRUCT Oy, Finland, 2016, p. 86–89.

[23] *Man and his work in the USSR and after*: 2nd ed., Rev. and add. // under. Ed. Zdravomyslov A.G., Yadov V.A. - M .: Aspect Press, 2003.

[24] *Self-regulation and forecasting of social behavior of an individual: A dispositional concept*. 2nd extended ed. - M.: TsSPiM, 2013.