Ontology Modification Using Ontological-Semantic Rules

Anastasia Mochalova*, Victor Zacharov**, Vladimir Mochalov*

** Institute of Cosmophysical Research and Radio Wave Propagation FEB RAS , Mirnaia str. 7, 684034 Paratunka, Kamchatka region, Russia*

***Petersburg State University, Universitetskaya emb. 7-9., 199034 St Petersburg, Russia*

stark345@gmail.com, v.zakharov@spbu.ru, sensorlife@mail.ru

***Abstract*—In this work we consider different types of semantic dictionaries and describe the problems of their construction. We also describe the ontological-semantic rules proposed for ontology modification. We provide examples of such rules and describe the process to generate them. The software implementation of ontology modification using ontological-semantic rules is employed as a component of a question answering system integrated with the ontology.**

***Keyword*— ontology modification, semantic analyzer, basic ontological-semantic rules.**

I. INTRODUCTION

WITH development of information technologies, the challenges of automated processing of natural language text become more and more urgent. One of the problems of automated text processing is the problem of organizing data storage in a structured way, with various additional information about stored elements — such as relations between these elements, names of these relations, hierarchical dependencies between them, qualificative semantic information etc. Ontologies that are used in many problems of automated text processing (for example, in development of question answering systems ([1-3]) and information retrieval systems ([4]), in classification systems ([5]) and estimation of texts resemblance ([6], [7]), in plagiarism detection ([8]) and disambiguation ([9]), in problems of Semantic Matching in Search [10] and information retrieval [11] etc.), are storages of such kind. Also, ontologies are essential components of Semantic Web technology [12] that becomes more and more popular lately.

Ontology creation and modification is a separate problem that does not have a universal solution in our days. One can distinguish two approaches: manual input (which is a very effort-consuming task requiring contribution from highly qualified specialists who know the domain of the ontology well) and automated input. The authors of the work [13] divide automated data input to the ontology into two stages:

1) automatic or automated input using conventional lexicographic information (encyclopaedic, definition and other dictionaries as well as databases);
2) automatic or automated input using analysis of distributional vocabulary characteristics in corpus of texts.

In this work we propose a method for automatic of an object-oriented ontology modification using ontological-semantic rules of the latter type.

In this work, by ontology modification we mean any of the listed below changes of the ontology:

1) Name change of a Class, Object or relation of the ontology;
2) Removal of a Class, Object or relation from the ontology;
3) Addition of a new Class, Object or relation into the ontology;

We propose the following scheme of ontology modification (see Fig. 1):

1) User feeds the analyzed text in natural language into the system's input;
2) The text proceeds into the input of the ontological-semantic analyzer (the detailed work description of such analyzer is provided in work [2]). The result of the analyzer's work is an ontological-semantic graph (a semantic graph, each node of which has a corresponding class or an object of the ontology).
3) The resulting ontological-semantic graph is fed into the input of the module of automated ontology supplement, which, using special ontological-semantic rules, constructs a request for ontology modification (the details of such request construction and of the ontological-semantic rules will be provided further).

In its work, the ontological-semantic analyzer uses data from the ontology (see [2] for details), together with the module of automated ontology supplement (the ontological-semantic rules lying in the base of the module use information from the ontology).

In our work, we developed and implemented in software a

Manuscript received June 10, 2016. The work was implemented with financial support from the Russian Foundation for the Humanities as part of research project No.15-04-12029 ― Software development of an electronic resource with an online version of a Russian-language question answering system.

Mochalova A. V., is with the Institute of Cosmophysical Research and Radio Wave Propagation FEB RAS, Russia (corresponding author to provide phone:+7-924-794-21-82; fax: 8-(41531)-33718; e-mail: stark345@gmail.com).

Zakharov V. P, is with the Saint-Petersburg State University, Russia (e-mail: v.zakharov@spbu.ru).

Mochalov V. A., is with the Institute of Cosmophysical Research and Radio Wave Propagation FEB RAS, Russia (e-mail: sensorlife@mail.ru).

prototype of an ontology modification system using ontological-semantic rules. This system is employed as a component in a question answering system integrated with the ontology. The software implementation of the ontology modification system is based on an expert system (the program [14] has been registered) and a semantic analyzer (the program [15] has been registered).

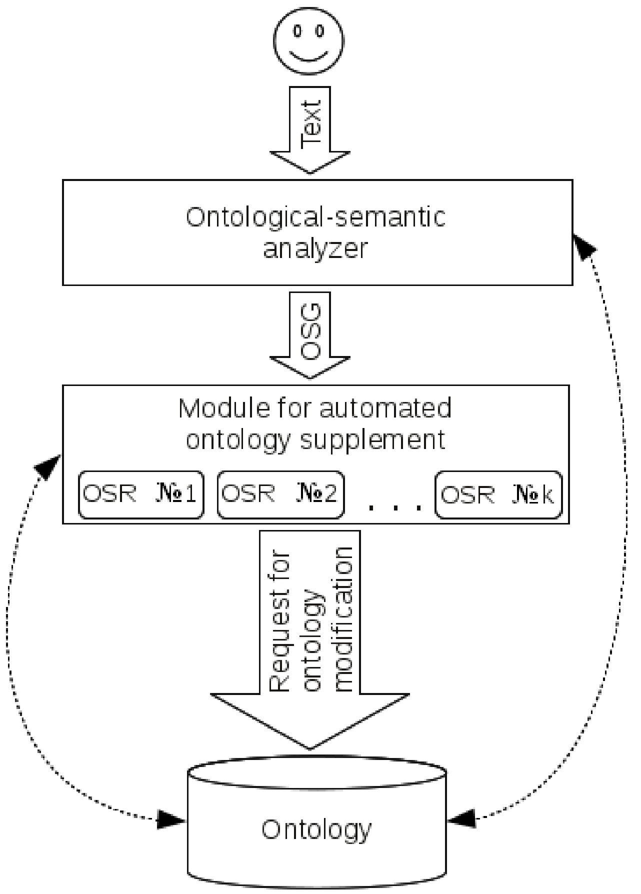

Fig. 1. Workflow of the system implementing automated ontology supplement.

II. Semantic Dictionaries

The work quality of a question answering system strongly depends on completeness and accuracy of the data stored in semantic dictionaries used by the system. These dictionaries are employed in different modules of the QAS, including the syntactic analyzer [16]. In this section we will consider approaches to construct the thesaurus and role models dictionaries; all of them are kinds of semantic dictionaries.

A. A general-purpose thesaurus

A thesaurus in its general sense is a dictionary with semantic relations between dictionary units. Since the end of 1950s, thesauri have been used in systems for machine translation and information retrieval systems (IRS).

In contrast with semantic dictionaries that are intended for detailed description of general lexis, thesauri are created to store and classify the ultimately concrete words and collocations. For example, the word вещество [substance] is in the RGPSD (Russian General-Purpose Semantic Dictionary), while all names of chemical substances are stored in a thesaurus.

Which relationships are described in a thesaurus? As a rule, the following:

1) AKO relationship (see examples in Fig. 2)
2) POF relationship
3) synonymy/antonomy
4) associative relationships.

Relationships stored in ontologies are much more numerous and various.

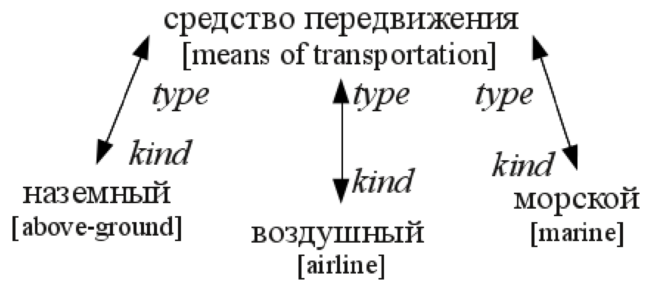

Fig. 2. Example of an AKO relationship

These are paradigmatic relationships (stable relationships between words in a language or in a text). Syntagmatic (textual) relationships are not presented in the thesaurus in an explicit form.

B. Role models dictionaries

Let us employ such approach to semantic text analysis that a sentence is considered as some predicate and a set of arguments. Usually a verb (or another predicate word, e.g., a verbal noun) describing an action acts as a predicate, while actants are the arguments.

When one has constructed a dictionary of verbal role models basing on usage of syntactic and morphological information, it is possible to define roles of nominal groups (arguments) by the predicate, as well as relations between them. For example, one can employ information about a preposition used with the nominal group and the case of the main word of the group. Nevertheless, syntactic information is not always sufficient. Consider an example: “Мы прибыли на автобусе на конференцию на пять дней” [We arrived by bus to a conference for five days]. An example of semantic parsing of this sentence is presented in Figure 3.

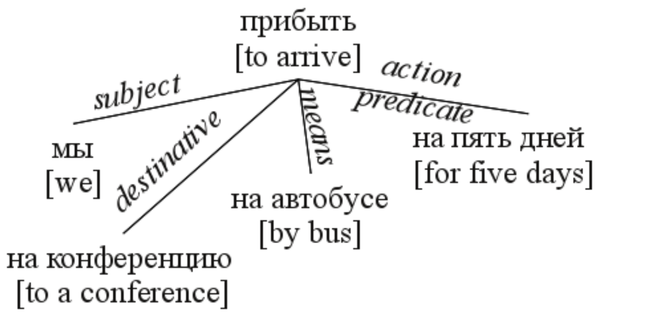

Fig.3. Example of semantic analysis.

The formal attributes (preposition+case) of some nominal groups coincide, so for correct interpretation of such sentence, one additionally needs a thesaurus. In such way, we obtain the following possible appearance of a dictionary entry presented in Table 1. (We assume the occurrence of a thesaurus containing categories “means of transportation” and “time interval”).

III. SEMANTIC RELATIONS

In this work we use the term semantic relation (defined below) as the relation defining a link of type "Class-Class", "Class-Object", "Object-Class" or "Object-Object" in the considered object-oriented ontology.

TABLE I
AN EXAMPLE OF A DICTIONARY ENTRY FOR THE VERB "ARRIVE".

Role of the verb	Preposition + case of the nominal group	Class of the nominal group
Mean	На [by] + Prepositional case	Means of transportation
Duration	На [for] + Accusative case	Time interval
Object localization	На [to] + Accusative case	Event

By semantic relation we mean a certain universal relation that a native speaker beholds in the language. This connection is binary: it connects two semantic nodes (each of which is a Class or an Object of the ontology) with each other [17]. By semantic nodes we mean syntaxems (syntaxem is an irreducible semantic-syntactic unit conveying primitive categorical meaning and acting as a structural component of a more complicated syntactic composition [18]). Let us say, that two different semantic nodes α and β are connected by the semantic relations R (R(α, β)) if there is a universal binary connection between α and β [17]. Direction of the connection is defined so that the formula R(α, β) would be equivalent to one of the following statements:

1) "β is R for α";
2) "question R can be asked from α to β".

Below you can find examples of the semantic relations equivalent to the first statement:

1) Description(вечер [evening], теплый [warm]);
2) Action(дети [children], пошли купаться [went for a swim);
3) Characteristic_of_action(разоделись [dressed], в пух и прах [to kill]);
4) Time(опоздать [be late], на час [for an hour]).

Below you can find examples of the semantic relations equivalent to the second statement:

1) With_who(прийти [come], с другом [with a friend]);
2) What_for(уронил [drop], нарочно [on purpose]);
3) Whose(мамин [mother's], шарф [scarf]).

It is obvious that these two types of relations are interdependent.

IV. ONTOLOGICAL-SEMANTIC RULES

By an ontological-semantic rule (OSR) we mean the rule of the form «if A, then B» according to which the expert system performs actions described in the right side of the rule in case the conditions described in the left side are held.

Let us put a fact of the expert system into correspondence with each syntaxeme allocated in the analyzed text. By the fact f_i of the expert system we mean a set of six elements: (1) the class or the object of the ontology the syntaxeme belongs to; (2) morphological characteristics of the syntaxeme; (3) ontological characteristics of the syntaxeme; (4) syntaxeme position in the analyzed text; (5) link to the previous fact of the expert system (prev); (6) link to the next fact of the expert system (next). The syntaxeme with the minimal position in the analyzed text (f_0) corresponds to the fact that has link to NULL as the link to the previous fact. The fact with the maximal position in the analyzed text (f_n) has link to NULL as the link to the next fact.

In such way, the analyzed text may be presented as a doubly linked list of facts F={ $f_0, f_1, f_2, \ldots, f_n$} (see Fig. 4).

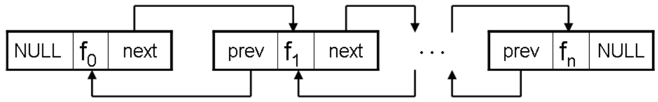

Fig. 4. Doubly linked list of facts of the expert system.

The left side of the rule can contain a doubly linked list of facts and/or Boolean functions having these facts as their arguments.

The right side of the rule contains the list of the actions each of which can modify the ontology of elements corresponding to syntaxemes allocated in the analyzed text. A particular case of the actions performed in the right side of the rule are the functions for ontology modification. The full description of these functions as well as the functions used in the left side of the expert system rules is provided in work [2].

V. ONTOLOGY MODIFICATION USING ONTOLOGICAL-SEMANTIC RULES

In the Explanatory Dictionary of the Russian Language by Ozhegov [19], the verb is defined as "the part of speech defining an action or a state, expressing this definition in forms of tense, person, number (in the present tense), gender (in the past tense) and forming participles and adverbial participles". The examples from this section of ontology modification are taken from work [20]. We suggest to create OSRs with left sides containing the facts corresponding to the syntaxemes which are verbs. In Fig. 5 we show some functions for work with the ontology (the arrow drawn from the ontology to a function means that the function extracts information from the ontology; the arrow drawn in the opposite direction means that the function modifies data stored in the ontology).

Below we consider an example of modification of an ontology part formed using the following functions (in square brackets we provide functions explanation):

1) CreateClass(172, "страна [country]"); [Create a class "country" with id = 172]
2) CreateClass(1023, "город [city]"); [Create a class "city" with id = 1023]
3) CreateObject(462, 172, "Россия [Russia]"); [Create an object "Russia" with id = 462 belonging to the class with id = 172 (i. e. the class "country")]
4) CreateObject(4017, 1023, "Санкт-Петербург [Saint Petersburg]"); [Create an object " Saint Petersburg" with id = 4017, belonging to the class with id = 10234 (i. e. the class "city")]
5) CreateRelation(7, "Принадлежит [belongs]"); [Create a relation "Belongs" with id = 7]
6) CreateRelation(1023, 172, 7); [Create a relation with id = 7 (i. e. the relation "Belongs" linking the class with id = 1023 (i. e. the class "city") and the class with id = 172 (i. e. the class "country"))]
7) CreateRelation(4017, 462, 7); [Create a relation with id = 7 (i. e. the relation "Belongs" linking the object with id =

4017 (i. e. the object “Saint Petersburg”) and the object with id = 462 (i. e. the object “Russia”))]

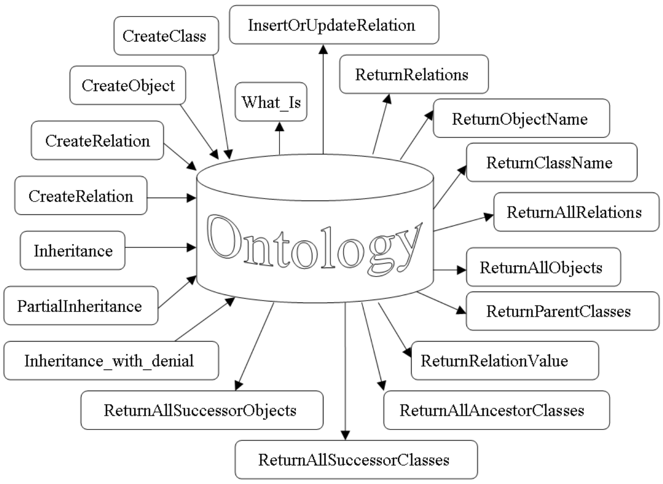

Fig. 5. Functions for work with the ontology.

In Fig. 6 we present the ontology part using the above-described functions. Let us consider the modification process for the ontology a part of which is presented in Fig. 6.

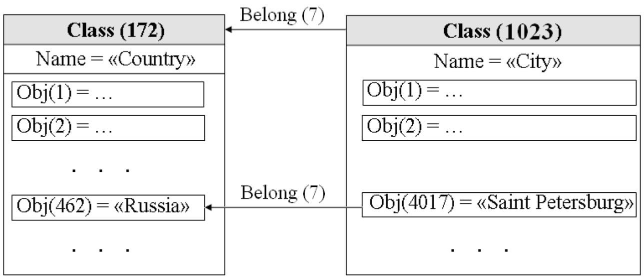

Fig. 6. An example of an ontology part.

Ontological-semantic rules will be applied to the analyzed text presented in one sentence “В августе 1914 года Николай II переименовал Санкт-Петербурга в Петроград [In August 1914, Nikolay II renamed Saint Petersburg into Petrograd]”. With use of the ontological-semantic analyzer described in work [21], we construct an ontological-semantic graph of this sentence (see Fig. 7). Each node of the ontological-semantic graph is a syntaxeme of the analyzed text, written in the lemmatized form. In curly brackets we provide the key ontological information about the corresponding graph node.

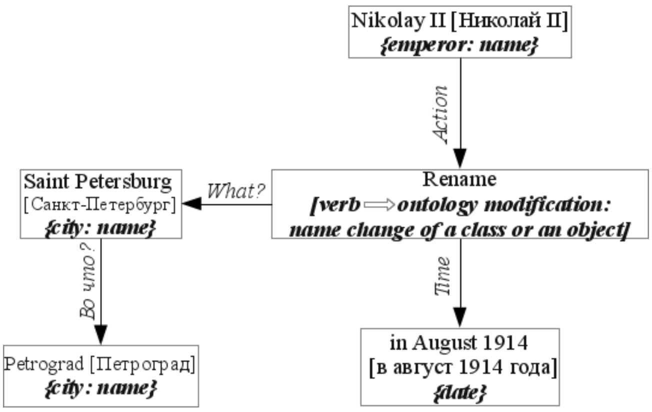

Fig. 7. Ontological-semantic graph.

As a result of the expert system operation using OSRs, modification of the above-decribed ontology will take place: upon expert system performing the OSR containing in its left part the verb “rename”, the name of the object formed previously by the function CreateObject(4017, 1023, "Санкт-Петербург [Saint Petersburg]") will be replaced with “Petrograd”. The information about exact modifications performed, together with their date, will be stored in a log file.

Let us consider another example of the work of the expert system which takes the sentence “В 1878 году по Сан-Стефанскому мирному договору город Батум перешел от Османской Империи к России [In 1878, by the Treaty of San Stefano, the Batum city was seded from Ottoman Empire to Russia]” as its input. Using the ontological-semantic analyzer, we will construct the corresponding ontological-semantic graph on this sentence and modify the existing ontology using the following functions:

1) RemoveRelation(552, 7645, 7); [Remove the relation with id = 7 (i. e. the relation “Belongs”, linking the object with id = 552 (i. e. the object “Batum”) and the object with id = 7645 (i. e. the object “Ottoman Empire”))]
2) CreateRelation(552, 462, 7); [Create a relation with id = 7 (i. e. the relation “Belongs”, linking the object with id = 552 (i. e. object “Batum”) and the object with id = 462 (i. e. the object “Russia”))]

Prepositions play a special role when constructing ontological-semantic rules for ontology modification. A significant part of semantic relations of verbs with other parts of speech is formed using prepositions. Though prepositions have abstract meaning, they manage to organize meaningful context when connecting meaningful parts of speech.

Prepositional constructions used to be described from the grammatical point of view and their semantics used to be neglected. One can hardly mention any corpus-based works dedicated to the Russian prepositions except for the paper by Klyshinsky [22], and a couple of others. It is also difficult to transform a set of constructions into a construction-based dictionary or grammar. To solve this task, one should pay attention to synonymy and variability of the constructions, variability of their grammatical features, and so on. For example, different constructions with the verb прятаться [to hide] differ in dynamical-statical aspect (in Russian meanings of such constructions would depend on the preposition chosen and on the case of the dependent component), while different constructions with the verb ударять [to strike] differ in manner of action (you can strike someone or you can strike the bell: in Russian, these constructions would include different prepositions). Treating constructions this way, we can grasp and describe normal “behavior” of constructions as well as abnormal cases (like the classical Goldberg's example to sneeze the napkin off the table [23]).

Below we provide an example showing how, depending on the context in two different sentences, two different semantic dependencies could be discovered (which means the ontology modification should also differ in the first and in the second case) with equal arguments:

1) Из-за огромных сугробов, наметенных в последнюю метель, экспедиция вышла на неделю позже. [Due to the huge snowbanks drifted by the recent blizzard, the expedition started a week later.] → REASON(выходить [start], из-за сугроб [due to snowbank]))
2) Из-за сугробов вышла маленькая девочка в сером

пальтишке. [A little girl in a gray coat appeared from behind the snowbanks] PLACE(выходить [appear], из-за сугроб [from behind snowbank])

VI. Conclusion

In this work we developed and implemented in software a prototype of a system for ontology modification using a database of ontological-semantic rules. This system is employed as a component of a question answering system using data from the ontology. The ontological-semantic rules for ontology modification were constructed, primarily, considering peculiarities of verbs and prepositions of the Russian language. In their description the rules use both morphological and ontological information about described objects. In the future we plan to extend the base of ontological-semantic rules, and employ context information as well as morphological and ontological information about syntaxemes.

References

[1] R. Sun, J. Jiang, Y. Fan, T. Hang, C. Tat-seng, C. M. Yen Kan, "Using syntactic and semantic relation analysis in question answering", *In Proceedings of TREC*, 2005.

[2] V.A. Kuznetsov, V.A. Mochalov, A.V. Mochalova, "Ontological-semantic text analysis and the question answering system using data from ontology", *ICACT Transactions on Advanced Communications Technology (TACT)*, vol. 4, Issue 4, pp.651-658, July 2015.

[3] A. Mochalova, "Search for answers in ontological-semantic graph", *Proceedings of the AINL-ISMW FRUCT*, Saint-Petersburg, Russia, ITMO University, FRUCT, pp. 174-180, 9-14 November 2015.

[4] M.-H. Hsu, M.-F. Tsai, H.-H. Chen, "Query expansion with conceptnet and wordnet: An intrinsic comparison", *Information Retrieval Technology*, pp. 1–13, 2006.

[5] A. Panchenko, R. Beaufort, H. Naets, C. Fairon, "Towards Detection of Child Sexual Abuse Media: Classification of the Associated Filenames", *In Proceedings of the 35th European Conference on Information Retrieval (ECIR 2013), Lecture Notes in Computer Science*, vol.7814, Moscow, Russia, 2013.

[6] R. Mihalcea, C. Corley, C. Strapparava, "Corpus-based and knowledge-based measures of text semantic similarity", *In AAAI'06*, pp. 775–780, 2006.

[7] G. Tsatsaronis, I. Varlamis, M. Vazirgiannis, "Text relatedness based on a word thesaurus", *Journal of Artificial Intelligence Research*, vol. 37, pp.1–39, 2010.

[8] M. Mozgovoy, V. Tusov, V. Klyuev, "Using measures of semantic relatedness for word sense disambiguation", *Computational Linguistics and Intelligent Text Processing*, vol. 2588 of LNCS, pp. 241–257, Springer Berlin, 2003.

[9] S. Patwardhan, S. Banerjee, T. Pedersen, "Using measures of semantic relatedness for word sense disambiguation", *Computational Linguistics and Intelligent Text Processing*, vol. 2588 of LNCS, pp. 241–257, Springer Berlin, 2003.

[10] H. Li, J. Xu, "Semantic Matching in Search", *Foundations and Trends in Information Retrieval*, vol. 7: No. 5, pp 343-469, 2014.

[11] M. Gupta, M. Bendersky, "Information Retrieval with Verbose Queries", *Foundations and Trends in Information Retrieval*, vol. 9, No. 3-4, pp. 209 – 354, 2015.

[12] G. Antoniu, P. Groth, F. Harmelen, R. Hoekstra, *A Semantic Web Primer*, The MIT Press, Cambridge, Massachusetts, London, England, 288 p., 2012.

[13] V.Sh. Roubashkin, V.A. Kapustin, "Usage of term definitions in encyclopaedic dictionaries for automated ontologies supplement", *XI All-Russian united conference "Internet and the modern society"*, Saint Petersburg, 2008.

[14] A.V. Mochalova, *Certificate of registration of a computer program 'Expert system for search of semantic relations in a Russian-language text using basic semantic rules with removal'*, No. 2016612038, Russia, 17.02.2016

[15] A.V. Mochalova, *Certificate of registration of a computer program 'Program for semantic text analysis basing on basic semantic patterns with removal'*, No. 2015613430, Russia, 28.01.2015.

[16] V.P. Zakharov, A.V. Mochalova, V.A. Mochalov, *Question answering systems. Some problems of automated text processing*, Petrozavodsk: PIN, 40 p., 2015.

[17] A.V. Sokirko, *Semantic Dictionaries and Natural Language Processing*, PhD thesis, Moscow, Russia, 2001.

[18] G.A. Zolotova, *Syntactic Dictionary. The repertoire of the elementary units of Russian syntax*, Moscow, Russia: Nauka, 1988.

[19] S. Ozhegov, N. Shvedova, *Definition dictionary of the Russian language: 80 000 words and phraseological collocations*, Russian Academy of Sciences. Institute of the Russian language named after V.V. Vinogradov. 4th edition, enlarged, Moscow, Azbukovnik, 944 p., 1999.

[20] V.A. Mochalov, A.V. Mochalova, "Ontology modification using basic ontological-semantic rules containing verbs", *Theory and practice of modern humanitarian and natural sciences. Issue 6: collection of research papers of the annual transregional research and application conference*, Petropavlovsk-Kamchatsky, 08-12 February 2016, pp.192-195, 2016.

[21] A.V. Mochalova, "Algorithm for semantic text analysis based on basic semantic patterns with removel", *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, No. 5. pp. 126-132, 2014.

[22] E.S. Klyshinsky, N.A. Kochetkova, M.I. Litvinov, Maximov V.Yu., "Automatic construction of collocation database on the base of the big corpus. Computational Linguistics and Intellectual Technologies", *Proceedings of the International Conference "Dialog 2010"*, vol. 9 (16), Moscow, pp. 181–185, June 2010.

[23] A.E. Goldberg, *Constructions at Work: the Nature of Generalization in Language*, Oxford, England: Oxford University Press, 2006.

Anastasia Mochalova was born in Petrozavodsk, Russia, in 1987. She received the bachelor's degree at Petrozavodsk State University, the master's degree in St. Petersburg State University of Aerospace Instrumentation. She is an external PhD student in technical sciences at Petrozavodsk State University. Her research interests include automated processing of natural language texts, development of question-answering systems, automation of ontologies creation, and development of the semantic analyzer.

Victor Zakharov – born Leningradskaya region, USSR, 17.07.1947. Graduated from Leningrad State University (Specialist in Structural and Applied Lingustics, 1970). PhD (Saint-Petersburg State University, Applied and Mathematical Liguistics, 1997). Major field of scientific research is Corpus Linguistics.

He is an Associate Professor, Saint-Petersburg State University. Previous positions ncluded Deputy Director of the Leningrad Center for Scientific and Technical Information, Automation Department Chief in th Russian Academy of Sciences Library. The main publications are as follows: "Corpora of the Russian Language", Text, Speech and Dialogue: Proceedings of the 16th International Conference (TSD 2013, Plzen, Czech Republic), Springer-Verlag (Lecture Notes in Artificial Intelligence, 8082), Berlin-Heidelberg, pp. 1-13, 2013. "Set phrases: a view through corpora", Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialog 2009" , vol. 14 (21). Moscow, pp. 667-682, June 2015. Current and previous research interests include information retrieval, natural language processing, and computational lexicography.

Dr. Zakharov is a member of the Russian Society of Information Specialists and a member of the Special Interest Group on Slavic Natural Language Processing.

Vladimir Mochalov was born in Lyubertsy, Russia in 1985. He received the Ph.D. degree in electronic engineering from Moscow Technical University of Communications and Informatics. His research interests include networks structure synthesis, artificial intelligence, bio-inspired algorithms, query answering systems, and Big Data.