Development of Lexico-Syntactic Ontology Design Patterns for Information Extraction of Scientific Data

Kristina Ovchinnikova[1[0000-0002-7400-0586]], Irina Kononenko[2[0000-0001-5057-6807]] and Elena Sidorova[2[0000-0001-8731-3058]]

[1] Novosibirsk State University, Novosibirsk, Russia
[2] A.P. Ershov Institute of Informatics Systems, Siberian Branch, Russian Academy of Sciences, Acad. Lavrentjev avenue 6, 630090 Novosibirsk, Russia
k.ovchinnikova2@g.nsu.ru, irina_k@cn.ru, lsidorova@iis.nsk.su

Abstract. The work considers an approach to information extraction based on lexico-syntactic patterns (LSPs). LSPs are built on the basis of knowledge about the scientific subject domain presented in the ontology and the corpus of scientific publications in different areas of knowledge. Two key tasks must be solved with the help of the LSPs: extracting object names and constructing objects in accordance with the structure of the ontology classes. In line with these tasks, terminological and informational LSPs are differentiated. Terminological patterns ensure the extraction of object names and properties based on indicators - marker words and phrases. Information patterns provide identification of ontology objects based on key attributes, description of actant structure for predicates expressing attributive relations and relations between ontology objects, as well as matching language constructions to values of attributes of ontology objects and their relations. Research is conducted on the basis of a corpus of scientific publications, which includes 100 articles from various fields of knowledge. The ways of expressing information about research method as the central concept of the ontology of scientific activity are investigated.

Keywords: Lexico-Syntactic Patterns, Ontology Population, Subject Dictionary, Ontology of the Scientific Activity, Ontology Design Patterns.

1 Introduction

The development of Semantic Web tools today is associated with the demand for ontologies as a means of unifying subject knowledge, storing, providing navigation and searching for well-structured data. The standardization of ontology representation instruments and the creation of a bank of ready-made solutions by the community [1] poses new tasks for researchers. There is a necessity of providing flexible mechanisms for using "samples" of ready-made solutions for the design and development of custom ontologies, as well as tools for their automated augmentation. For more than ten years, methods based on the application of Ontology Design Patterns, or ODP [2] documented descriptions of practical solutions to typical problems of ontological modeling [3], have been used.

In some existing works on the automated augmentation of ontologies and thesauri, a linguistic approach based on knowledge is distinguished using constructs, or templates. These templates are subdivided into grammatical [4], lexico-grammatical [5, 6] and lexico-syntactic [7, 8] ones depending on the type of linguistic information.

One of the linguistic approaches proposed in the work [9] to solving the problem of automated ontology population is using the idea of the possibility of automating the construction of semantic relations based on diagnostic contexts presented in the form of lexico-syntactic patterns (LSPs), structural samples of language constructions which represent lexical and surface syntactic properties [7]. A more detailed definition is given in the work [6], where LSP is defined as a model (or structural sample) of a linguistic structure, indicating the essential grammatical characteristics of a lexemes set included in linguistic expressions of this class, and syntactic conditions of the use of linguistic expression constructed according to pattern (for example, the rules for matching morphological features of tokens). In a number of works [10] the term lexical-semantic pattern is also used for LSP, emphasizing the possibility of describing lexical units in a sample using lexico-semantic classes (features). The technique proposed in the work [8], known as Hearst patterns, is intended for processing unstructured texts. It has been widely used for extracting hyponymy relations involving extracting ordered pairs of words from a collection of documents that match a set of predefined patterns. An approach of M. Hearst has been further developed by many researchers and applied to various languages and text genres [10-12] in the form of universal or domain-specific LSPs for extracting ontological entities and relations from texts.

For the formal representation of LSP, means of varying degrees of complexity are used, from lexical regular expressions (Hearst patterns) or regular expressions over a part-of-speech alphabet to non-semantic specialized languages, such as the Tomita parser [13], LSPL [7, 14], Diglex instrument [15], or those using semantic markup, such as Gate [16] and Faton [17] platforms.

The purpose of this work is to develop a methodology for constructing LSPs for the ontology of scientific activity in order to extract information from texts and enrich the ontology that already has an initial structure and content. The LSP representation model based on available template technologies provides tight integration with ontology design patterns. This technique is demonstrated by application to the research method, a key component of the scientific ontology, and is oriented at further automation.

2 Lexico-Syntactic Ontology Design Patterns

LSPs considered as a tool for augmentation of the ontology allow solving the following key tasks: the extraction of object names (including “new” names not presented in the dictionary) and object attribute values and construction of objects according to the structure of the ontology classes. In accordance with these tasks, two types of LSPs were identified: terminological and informational ones.

Terminological lexico-syntactic patterns (T-LSP) provide extraction of the names of objects or their properties due to indicators - marker words and phrases identified during the analysis of ontology and language constructions found in the text corpus of the field of knowledge. All indicators are entered in the subject dictionary and marked with lexico-semantic features (classes and values).

The proposed methodology for the creation of T-LSP has three features. The first is the dictionary system of lexico-semantic classes formed in accordance with the description of the subject domain presented in the ontology. The second is the use of indicator terms obtained on the basis of the ontology (names of classes, attributes, relations) as well as their synonyms (or contextual synonyms). The third peculiarity concerns the inclusion of variables with given properties in the T-LSPs. The values of these variables are specified by the corresponding normalized text fragments.

[<Adj>*, <Method>, <N, gen>*] ⇒ Method.Name

This pattern[1] is focused on extracting the name of a method represented by the noun phrase Adj + N + N,gen, with the lexeme of the Method class as its syntactic head.

Informational lexico-syntactic patterns (I-LSP) are used to match language constructions against ontology objects. The formalism for I-LSP representation is a modification of the fact extraction schemes language proposed in [17] to extract facts from the text. Each I-LSP implements a model of the form:

<Arguments, Constraints, Result>,

where Arguments correspond either to terms extracted using T-LSPs, or to objects (it is assumed that objects have already been extracted earlier using some other I-LSPs), Constraints set semantic and / or syntactic conditions on Arguments, and Result describes a fragment of the ontology generated by the pattern.

In the process of constructing I-LSPs, the following tasks are to be solved: a) identification of ontology objects based on key attributes, b) description of the actant structure for predicates that express the ontological relations between objects and their attributes or between the objects themselves, c) formulation of syntactic, semantic and positional constraints on I-LSP arguments in the most generalized form, d) matching pattern arguments to values of attributes of ontology objects and their relations.

[Person()<nom>, описал ‘described’, Method()<acc>] ⇒ arg1::Method(Author: arg3)

This pattern allows linking an object of the *Person* class and the *Author* attribute of the *Method* class.

3 Information Extraction Approach based on LSPs

The process of extracting information with a given set of patterns goes through several stages: the preliminary stage, at which the text is tokenized, the stage of extracting subject terms presented in the dictionary, and generating objects and adding them to the ontology (see Fig.1).

---

1 In angle brackets, the lexical-semantic class and grammatical categories are indicated, the * symbol means the possibility of repeating an element in the structure, including optionality.

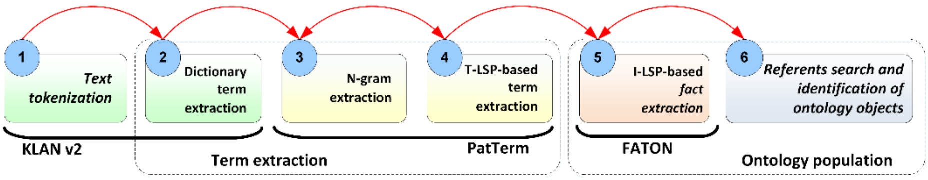


Fig. 1. Scheme of information extraction based on LSP.

This scheme demonstrates the sequence of steps for LSP-based population of an ontology. Take a closer look at this process.

Tokenization of the text (step 1) provides the transformation of the document to the required text format and the splitting of the text into elementary units - words, punctuation marks, separators.

The term extraction step includes search for dictionary terms (step 2), extracting N-grams (step 3), and T-LSP-based search for new terms and term-like N-grams (step 4). At this stage, morphological and surface syntactic analysis of the text is carried out, as well as matching template structures against the chain of found tokens with the necessary check of their grammatical features.

At the next stage, the found terms are sent to the input of the object search and formation subsystem. At this stage, for each I-LSP, a suitable set of arguments is searched for among the found terms and already generated objects, the match conditions are checked, and new objects or relations are created (step 5). All found objects go through the identification procedure – comparison with individuals of the domain ontology – and enrich the ontology with new facts (step 6).

Thus, the result of the extraction process will be the objects found in the text and referring to existing or new instances of the domain ontology classes.

To implement the approach described above, a stack of technologies previously developed by our team for solving various AOT problems [18] is used. The Klan system is used for creating a subject dictionary and text analysis, carrying out morphological and surface syntactic analysis of the text and extraction of dictionary terms [19]. The PatTerm system [4] provides a search for text fragments according to a given pattern and is used for finding term-like N-grams in the text [20]. The Faton system [17] is designed for matching language constructions formulated as samples in I-LSP to ontology objects and extracting information on the basis of these patterns.

4 Subject Dictionary

For the study, a corpus including 100 Russian-language articles was compiled from the resource (https://cyberleninka.ru). Articles were selected from 5 scientific collections corresponding to the humanities, natural, technical, social and exact sciences. Each field has several subfields. The humanities includes linguistics, pedagogics and psychology; the natural sciences incorporate archeology, biology, geography, medicine, chemistry and ecology; the technical science consists of such directions as mechanics and physics; the social sciences enclose architecture, history, politics, law, sociology, philosophy , economics and energy, while information technology and mathematics are related to the exact science.

This breakdown was made in order to test (and further confirm) the hypothesis about the same or similar use of constructions in different scientific fields. The section of the humanities includes 21 articles, natural science - 14 articles, technical science - 12 articles, social science section consists of 36 articles and 17 articles relate to exact science. The total volume of the corpus is 370.8 thousand tokens.

Pre-processing of the texts was performed to convert them into a suitable format (txt), check them for technical errors and correct if necessary.

The next stage is creating a dictionary and its processing in order to remove incorrectly formed terms and phrases and carry out thematic classification of significant terms.

In the course of the analysis of scientific texts, constructions describing research method were identified and classified in terms of information to be extracted as follows:

- constructions representing basic information about the method: name, description (characteristic), task (purpose) of the method, object of research, activity in which it is used (e.g., project);
- constructions representing information about the creation of a method: author, date of appearance, geographical location, information resource;
- constructions representing additional information: publication describing the method, person applying it, the scientific result and the section of science.

Dictionary subsystem allows to form subject specific dictionaries based on the text corpus and provides a set of research tools (to recognise terminological phrases, construct concordances, calculate occurrence statistics, make semantic annotation for terms). These tools helped create the dictionary of terms required for extracting information about the research method. It contains 16918 words (terms) and 65407 terminological phrases (compound terminology entries).

Based on the structure of a specific ontology the system of lexico-semantic classes in the dictionary reflects the hierarchy of the ontology objects and relations. For analyzing the vocabulary, the concordance constructed by the word *метод 'method'* proved to be useful. The following is a table of lexico-semantic groups (LSG).

Table 1. Lexico-Semantic classes.

Class name	Lexemes list
comprehension	понимать 'understand', предполагать 'suppose', осознать 'realize', осознавать 'realize', воспринимать 'perceive', пониматься 'be understood', предполагаться 'be supposed', восприниматься 'be perceived'
mental-speech	объяснять 'explain', определять 'define', трактовать 'interp', расценивать 'regard', рассматривать 'consider', объясняться 'be explained', определяться 'be defined', трактоваться 'be interped', рассматриваться 'be considered'
existence	являться 'be', быть 'be', be considered 'считаться', выступать 'act as'

essence	заключаться ‘consist’, состоять ‘consist’, особенность ‘feature’, смысл ‘meaning’, принцип ‘principle’, идея ‘idea’, суть ‘point’
creation	предложить ‘offer’, ввести ‘introduce’, разработать ‘develop’, описать ‘describe’, создать ‘create’
intel-activity	решать ‘solve’, вычислять ‘calculate’, исследовать ‘investigate, изучать ‘study’, решаться ‘be solved’, изучаться ‘be studied’, вычисляться ‘be calculated’
result	получить ‘obtain’, получать ‘obtain’, решить ‘solve’, получаться ‘be obtained’, получиться ‘be obtained’, решиться ‘be solved’, изучить ‘be solved, вычислить ‘calculate’
use	применять ‘apply’, применяться ‘be applied’, использовать ‘use’, использоваться ‘be used’
purpose	задача ‘task’, цель ‘purpose’
method	метод ‘method’, метода ‘method’, способ ‘way’, прием ‘method’, технология ‘technology’, средство ‘means’
inceptive	начаться ‘start’

In the table 1 verbs stand for all verb forms: personal, infinitive, participles and gerunds. The division into groups presented in the table is based on the semantic meanings expressed in sentences. For example, LSG intel-activity includes imperfective verbs that define a series of intellectual actions (operations) taken in order to achieve a particular end and performed using a method, and LSG result includes an action with an emphasis on the result obtained by a method.

5 Extracting Information about Research Method based on LSPs

For methodology development the research method was taken as a target for information extraction due to its being one of the main concepts of the scientific activity ontology. The types of information extracted can be subdivided into named entities (persons, names of organizations and geographic objects), attributes of objects (position and place of work for a person), relations between objects (being an employee), facts and events (creating an organization).

5.1 Ontological Class Method

Consider the scheme of attributes and relations of objects of the *Method* class (see Fig.2).

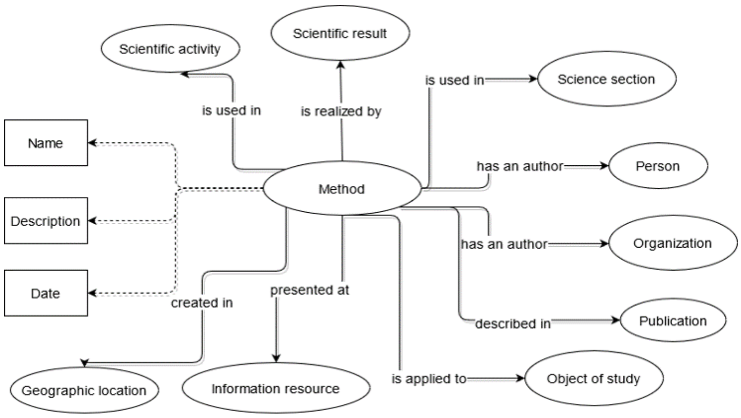


Fig. 2. Scheme of relations and attributes of objects of the Method class.

An object of the *Method* class can be associated with objects of such classes as *Scientific activity* in which it is used, *Scientific result* obtained by it, *Science section* using it, *Person* or *Organization* which are its authors, *Publication* describing it, the *Object of study* applying it, the *Information Resource* presenting it, and the *Geographical Place* where it was created. The attributes of an object of the *Method* class are *Name*, *Description* and creation *Date* having string values.

The formation of lexico-semantic classes is based on ontology and predicate classes allowing expression of relationships between classes. Lexemes of lexico-semantic groups are able to create rows of synonyms and be interchangeable in sentences.

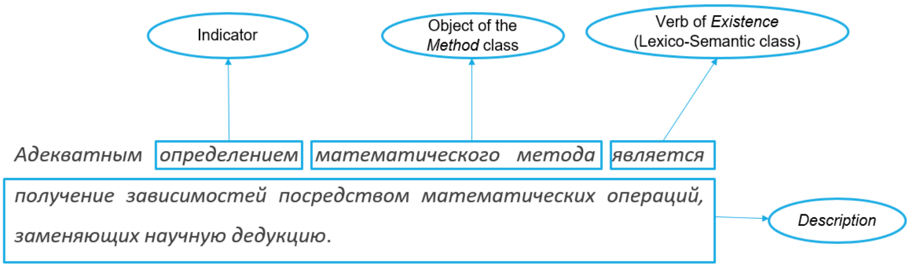
5.2 Representation of Patterns

The typification of constructions based on the ontology obeys the following principles. Class objects are written in constructions as the name of the class: *Method*, *Author*, *Task*, *Object*, *Result*, *Geographic Object* and *Information Resource*. The extracted attributes are written in constructions as the names of the attributes of the Method object: Name, Description and Date. Any lexeme of a lexico-semantic class is represented by the class name, the grammatical class and features of the lexeme being specified if necessary.

The sentence (see Fig. 3) is described by the construction:

[*определение ‘definition’*, Method(), <Existence><Verb>, <Method.Description>]

This construction includes four components: the lexeme *определение ‘definition’*, an instance of the Method class, a Verb of the Existence LSG and the Method attribute called Description.



"An adequate definition of the mathematical method is obtaining dependencies through mathematical operations that replace scientific deduction."

Fig. 3. Structural representation of the Research Method Description in the sentence.

5.3 Representation of Patterns

Some grammatical categories (GC) and syntactic relations appeared to be irrelevant for the description of certain constructions. The following are examples.

(1) *Тянь чжан предложил метод агломерационной иерархической кластеризации под названием birch* (Tian Zhang proposed a method of agglomerative hierarchical clustering called birch).

(2) *Взаимодействие МК с МЭА параллельно исследовали методами изомолярных серий и молярных отношений* (The interaction of MC with MEA was investigated in parallel by the methods of isomolar series and molar ratio).

In (1) the 3rd person verb is used with the subject expressed by a proper name. The component *Author* implies the presence of a proper name, therefore, in this case, for simplification of the pattern the indication of the third person of the verb can be omitted. So, (1) is covered by the sample

[Author(), <Creation>, Method()]

The sentence (2) is indefinitely personal one-member sentence, however, changing the number to a singular and replacing the verb *исследовать 'investigate'* with a reflexive *исследоваться 'be investigated'* will make the sentence definitely personal or two-member. Hence, the two verbs (*исследовать 'investigate'* and *исследоваться 'be investigated'*) can reasonably be combined in a single pattern, for which the syntactic constraint on the subject-predicate relation is optional:

[Object(), <Intel-Activity><Verb>, Method()]

Some grammatical categories and syntactic constraints (such as noun case and verb tense) appeared to be relevant for constructing samples in certain contexts. These contexts have been revealed using the concordance for the word 'method'.

The relations between the members of the sentence can be divided into subject-predicate agreement, coordinative and subordinate relations [21, 22]. The following examples illustrate three types of the first relation.

(3) *Математический метод определялся как использование математических символов для выражения исходных посылок и выводов и математические операции с этими символами для получения новых зависимостей* (The mathematical method was defined as the use of mathematical symbols to express the initial premises and conclusions and mathematical operations with these symbols to obtain new dependencies).

This type of agreement takes into account the nominative case of the subject and the subject-predicate grammatical agreement in number and gender forms.

(4) *В работе общегистологическим, люминесцентно-гистохимическим и иммуногистохимическим методами исследованы надпочечники 42 крыс-самок.* (The adrenal glands of 42 female rats were studied using general histological, luminescent-histochemical and immunohistochemical methods).

The second type of agreement is an agreement between the subject and the nominal part of the predicate in the forms of gender and number.

(5) *Математический метод - это получение зависимостей посредством математических операций, заменяющих научную дедукцию.* (The mathematical method is the derivation of dependencies through mathematical operations that replace scientific deduction).

The third type is the relation between the main members of the sentence, in which the grammatical forms of the subject and predicate are mutually independent.

Subordinate relations accounted for in the patterns are concord (*статистический метод 'statistical method'*), strong government (*расценивать метод 'evaluate method'*, *метод объяснения 'explain method'*) and weak government (*начать использование в 50-х годах 'start using in the 50s' / к 50-м годам 'by the 50s'*).

A coordination is a relation between homogeneous members of a sentence, which are equal components that do not depend on each other. There are two ways to describe several research methods. The first one is use of the plural of 'method' combined with coordinated method names (*сопоставительный и описательный методы 'comparative and descriptive methods'*, *методы контекстуального анализа и интерпретации 'methods of contextual analysis and interpretation'*). The second way is the repetition of lexeme *метод* '*method*' to represent each of the research methods (*метод контекстуального анализа и метод интерпретации 'contextual analysis method and interpretation method'*). In both cases, homogeneous members can be separated by commas or conjunction *and*.

The grammatical number of the predicate is plural in sentences with several subjects.

(6) *Диденко В. Н., Фахразиев И.И. и Мартынов А.И. разработали метод расчетного определения границ нестабильной детонации природного газа.* (Didenko

V.N., Fakhraziev I.I. and Martynov A.I. developed a method for calculating the boundaries of unstable detonation of natural gas).

In this case, the verb 'developed' is plural. The agreement constraint requires plural or multiplicity of the sentence subject.

5.4 Representation of Patterns

Consider examples of extracting information from scientific and technical texts with the help of I-LSPs. They are written in the FATON language and based on T-LSPs. I-LSPs can be subdivided into initializing ones (creating an object of a class) and those linking to an object of another class or an attribute.

```
(1) Scheme Method
    arg1: Term::Method
        ⇨ Object::Method (Name: arg1.Name)
```

The scheme (1) shows the creation of an object of the Method class. The pattern identifies the terms marked up in the subject dictionary and creates an object with the corresponding name.

```
(2) Scheme Method_Author
    arg1: Object::Person (Case: 'instr')
    arg2: Term::Create (ps: Verb)
    arg3: Object::Method (Case: 'nom')
    Condition  Contact (arg1, arg2) = Contact_Object,
               Contact (arg2, arg3) = Contact_Object
    ⇨ arg3::Method (Author: arg1.Name)
```

The scheme (2) shows a simple variant of binding an object of the Method class with an object of the Author class. Contact in this case means the possibility for terms to be separated by insignificant terms of other classes and function words. This pattern will handle cases like:

(7) *Метод «мозгового штурма» был разработан в 1953 г. американским консультантом Осборном* (The brainstorming method was developed in 1953 by the American consultant Osborne).

```
(3) Scheme Method_Description
    arg1: Term::( Norm: 'definition', Case: 'instr')
    arg2: Object::Method (Case: 'gen')
    arg3: Term::Existence (ps: Verb)
    // noun phrase with subordinate parts
    arg4: TERMIN_ALEX::Description (Case = 'nom')
    Condition  Contact (arg1, arg2) = Contact_Absolute,
               Contact (arg2, arg3) = Contact_Object,
               Contact (arg3, arg4) = Contact_Absolute
```

```
⇨ arg2::Method (Description: arg4.Name)
```

Scheme (3) illustrates a complex binding option – adding the *Description* attribute. It will process cases like the example in Fig. 3.

In the first phase, 32 patterns with an emphasis on high precision were designed. They extract 159 sentences from 100 articles. At the next phase of the work, increasing recall by means of weakening the restrictions in the patterns is planned.

Conclusion

The article discusses the methodology for constructing the LSPs for population of the ontology of scientific activity. Research is conducted on the basis of a corpus of scientific publications, which includes articles that equally represent various fields of knowledge. The analysis of typical structures expressing ways of presenting information about research methods is carried out. The described patterns concern research method name, description, connections with the task and the object of research.

The peculiarity of this work consists in analyzing the possibility of automatic generation of patterns due to consideration of certain types of linguistic information: 1) composition of patterns based exclusively on ontological relations; 2) use of basic linguistic constructions presented either in the ontology or in the general scientific vocabulary, and 3) reliance on the hierarchy and properties of the ontology classes when developing a system of lexico-semantic meanings for dictionary terms. Special attention is paid to syntactic constraints binding the pattern elements, identified on the basis of the analysis of a representative sample of their occurrences in the text corpus. 32 patterns for extraction information about research methods were designed. On average, 3-7 sentences are extracted for each pattern from 100 articles.

During the study, the incompleteness of the ontological description of the Method class was revealed. Such attributes as disadvantages and advantages of method, conditions of application (requirements), and relations to other methods (includes, develops) can be added there. Further development of the methodology involves the automation of the process of constructing the LSP.

Acknowledgments. The work was carried out with the financial support of the Russian Foundation for Basic Research (grant №19-07-00762).

References

1. Ontology design pattern bank: http://ontologydesignpatterns.org/wiki/Main_Page, last accessed 2021/04/29.
2. Hitzler P., Gangemi A., Janowicz K., Krisnadhi A., Presutti V.: Ontology Engineering with Ontology Design Patterns: Foundations and Applications. Studies on the Semantic Web, vol. 25. IOS Press/AKA (2016).
3. Borovikova, O.I., Zagorulko, Y.A., Kononenko, I.S.: On application of ontology design patterns for extraction of information from scientific texts. In: Informacionnye i matematicheskie tekhnologii v nauke i upravlenii, № 4 (12), pp. 18-29 (2018). (in Russian)

4. Kononenko, I.S., Akhmadeeva, I.R., Sidorova, E.A., Shestakov, V.K.: Problems of extracting terminological core of the subject domain from electronic encyclopedic dictionaries. In: System Informatics, №13, pp. 49-76 (2018). (in Russian)
5. Zakharov, V.P., Khokhlova, M.V.: Automatic extracting of terminological phrases. Strukturnaia i pricladnaia lingvistika, 10, 182–200. Saint-Petersburg State University (2014). (in Russian)
6. Mitrofanova, O., Zakharov, V.: Automatic Analysis of Terminology in the Russian Corpus on Corpus Linguistics. In: Proceedings of Fifth International Conference Slovko-2009: NLP, Corpus Linguistics, Corpus Based Grammar Research, pp. 249-255. Tribun, Smolenice, Slovakia (2009). (in Russian)
7. Bolshakova, E., Baeva N., Bordachenkova E., Vasilieva N., Morozov S.: Lexico-syntactic Patterns for Automatic Processing of Scientific and Technical Texts. In: Proc. of 10th National Conference on Artificial Intelligence with International Participation 2006, Vol 2, pp. 506-514. Moscow, Fizmatlit (2006). (in Russian)
8. Rabchevsky, E.: Automatic ontology construction based on lexical-syntactic patterns for information retrieval. In: RCDL'2009, pp. 69-77. Petrozavodsk (2009). (in Russian)
9. Hearst M.: Automatic Acquisition of Hyponyms from Large Text Corpora. In: Proceedings of the 14th International Conference on Computational Linguistics, pp. 539–545 (1992).
10. Saeeda, L., Med, M., Ledvinka, M., Blaško, M., Křemen, P.: Entity Linking and Lexico-Semantic Patterns for Ontology Learning. In: Harth A. et al. (eds) The Semantic Web. ESWC 2020. Lecture Notes in Computer Science, vol. 12123. Springer, Cham (2020).
11. Klaussner, C., Zhekova, D.: Lexico-Syntactic Patterns for Automatic Ontology Building. In: Proceedings of the Student Research Workshop associated with RANLP, pp. 109–114 (2011).
12. Maynard, D., Funk, A., Peters, W.: Using Lexico-Syntactic Ontology Design Patterns for ontology creation and population. In: Proceedings of the International Conference on Ontology Patterns WOP'09, vol. 516, pp. 39–52 (2009).
13. Tomita, M.: LR parsers for natural languages. In: 10th International Conference on Computational Linguistics COLING, pp. 354-357 (1984).
14. Bolshakova, E.I., Vorontsov, K.V., Efremova, N.E., Klyshinsky, E.S., Lukashevich, N.V., Sayapin, A.S.: Automatic text processing in natural language and data analysis. HSE, Moscow (2017). (in Russian)
15. Timofeev, P., Sidorova, E.: Lexico-semantic templates as a tool for declarative description of language constructions and linguistic analysis of the text. In: System Informatics, №13, pp. 35-48 (2018). (in Russian)
16. Cunningham, H., Tablan, V., Roberts, A., Bontcheva, K.: Getting More Out of Biomedical Documents with GATE's Full Lifecycle Open Source Text Analytics. In: PLoS Comput Biol 9(2): e1002854 (2013).
17. Garanina, N., Sidorova, E., Bodin E.: A Multi-agent Text Analysis Based on Ontology of Subject Domain. In: Perspectives of System Informatics PSI 2014, vol. 8974, pp. 102-110. LNCS (2015).
18. Sidorova, E.A.: Approach to modeling the process of information extracting based on the ontology. Ontology of designing, 8(1), pp. 131-151 (2018). (in Russian)
19. Sidorova, E., Akhmadeeva, I.: The Software Environment for Multi-Aspect Study of Lexical Characteristics of Text. In: Selected Papers of the XX International Conference on Data Analytics and Management in Data Intensive Domains DAMDID-19, vol. 2523, pp.306-315. CEUR, Kazan, Russia (2019).
20. Rosenberg, G., Salomaa, F.: Handbook of Formal Language. Vol. 1, Ch. 4. Springer-Verlag Berlin Heidelberg (1997).

21. Hudson, R.: Coordination and grammatical relations. Journal of Linguistics, vol. 24, pp. 303–342 (1988).
22. Moravcsik, E.A.: Agreement. In: Greenberg, J.H., Ferguson, C.A. (eds) Universals of Human Language: Syntax, vol. 4, pp. 331–374. Stanford University Press, Stanford, CA (1978).