OntoMath^{Edu} Educational Mathematical Ontology: Annotation of Concepts

Olga Nevzorova[a], Liliana Shakirova[a], Marina Falileeva[a], Alexander Kirillovich[b], Vladimir Nevzorov[c] and Evgeny Lipachev[a]

[a] *Kazan Federal University, Kazan, Russia*
[b] *Joint Supercomputer Center of the Russian Academy of Sciences, Kazan, Russia*
[c] *Kazan National Research Technical Universitys, Kazan, Russia*

Abstract
This work is dedicated to population of the OntoMath$^{\text{Edu}}$ ontology by definitions of mathematical concepts. OntoMath$^{\text{Edu}}$ is a new educational mathematical ontology, intended to be used as a Linked Open Data hub for mathematical education, a linguistic resource for intelligent mathematical language processing and an end-user reference educational database. We propose a template-based method for automatical extraction of definitions from educational mathematical texts in Russian. The method has been implemented on the base of the “OntoIntegrator” system and evaluated on a collection of educational texts from the yaklass.ru website. The obtained F-measure is 89.2%.

1. Introduction

This paper is dedicated to population of the OntoMath$^{\text{Edu}}$ ontology by definitions of mathematical concepts.

OntoMath$^{\text{Edu}}$ is a new educational mathematical ontology [1, 2], intended to be used as a Linked Open Data hub for mathematical education, a linguistic resource for intelligent mathematical language processing and an end-user reference educational database.

The ontology underlines the eduation platform of OntoMath digital ecosystem [3], an ecosystem of ontologies, text analytics tools, and applications for mathematical knowledge management, including semantic search for mathematical formulas [4] and a recommender system for mathematical papers [5].

OntoMath$^{\text{Edu}}$ is organized in three layers: a foundational ontology layer, a domain ontology layer and a linguistic layer. The domain ontology layer contains language-independent concepts, covering secondary school mathematics curriculum. The linguistic layer provides linguistic grounding for these concepts, and the foundation ontology layer provides them with meta-ontological annotations.

Russian Advances in Artificial Intelligence: selected contributions to the Russian Conference on Artificial intelligence (RCAI 2020), October 10-16, 2020, Moscow, Russia
liliana008@mail.ru (O. Nevzorova); onevzoro@gmail.com (L. Shakirova); mmwwff@yandex.ru (M. Falileeva); alik.kirillovich@gmail.com (A. Kirillovich); nevzorovvn@gmail.com (V. Nevzorov); elipachev@gmail.com (E. Lipachev)

The concepts are organized in two main hierarchies: the hierarchy of objects and the hierarchy of reified relationships. The description of concept contains its name in English, Russian and Tatar, axioms, and relations with other concepts. Figure 1 represents the *Altitude of a triangle* concept in the WebProtege ontology editor.

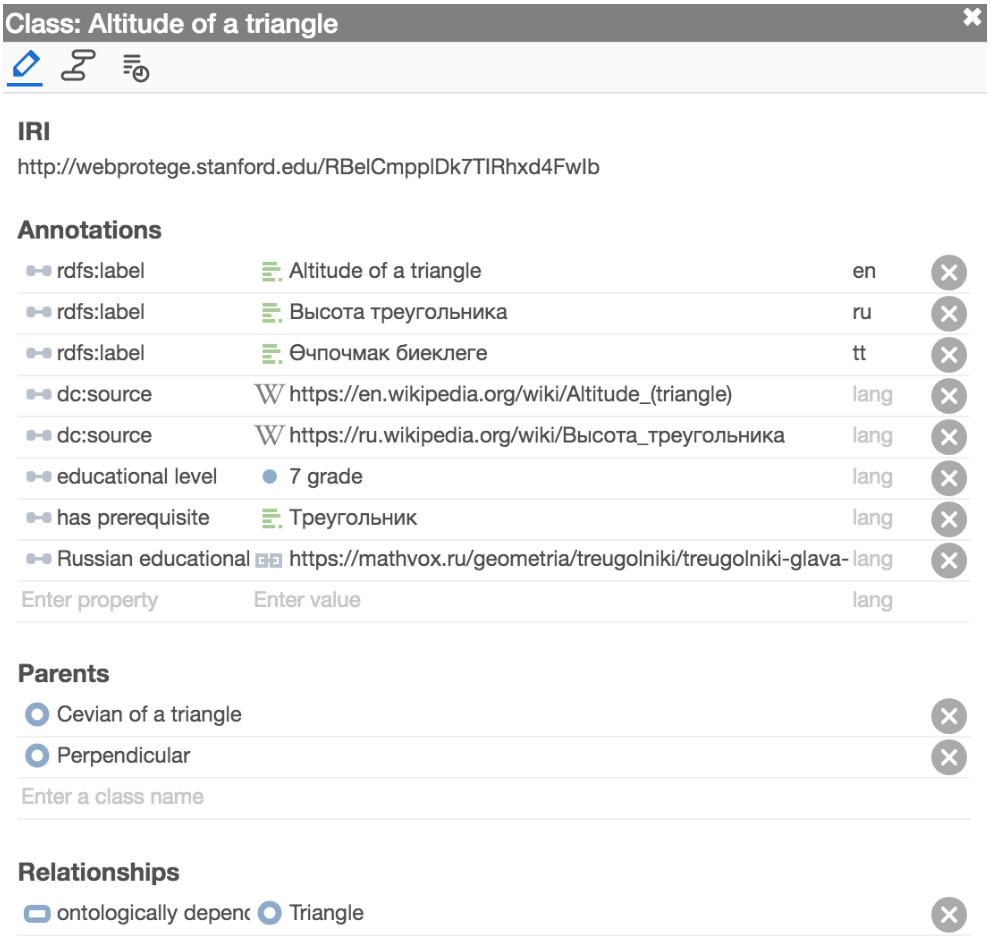


Figure 1: Altitude of a triangle concept

In order for using of the ontology for educational purposes, the description of the concepts has to be complemented by its definitions. The manual annotation of definition is time-consuming task, so some automatic method is needed. In this paper, we propose a method for automatic extraction of concepts definitions from educational texts in Russian.

The rest of the paper is organized as a following. In Sect. 2 we survey the related works. In Sect. 3 we propose the method for definitions extraction. In Sect. 4 we describe our experiments of application of this method. And in Conclusion we propose the future work.

2. Related Work

Definition extraction is a popular topic in NLP research. Relying to [6], we can determine three directions to definition extraction.

The first direction is the rule-based approach. The majority of these works use symbolic methods. These methods are based on lexico-syntactic patterns or features, which are manually crafted or semi-automatically learned [7]. Patterns are either very simple sequences of words (e.g. “refers to”, “is defined as”, “is a”) or more complex sequences of words, parts of speech, and chunks. A version of the language of lexical-syntactic patterns for Russian is proposed by E.I. Bolshakova et. al. [8, 9, 10, 11]. As well as they apply this method in various tasks, for example, for extracting the terms and their links for constructing a subject index for scientific text. The rule-based approach is intuitive and has high precision, and used for different languages [12, 13, 14]. Another related field is hypernym detection (see e.g., [15, 16] Because many hypernym definitions use the pattern “X is a (type of) Y”, and this form is indeed a common structure of the sentence with a definition.

Another fully automated method is proposed by C. Borg and colleagues [17, 18]. They apply genetic programming and rules to differ between definitions and non-definitions. But rules are learned for only one category of patterns, namely “is” patterns. However, most methods have both low recall and precision, because definitional sentences occur in highly variable and potentially complex syntactic structures.

The second direction is the feature engineering approach. This approach uses the statistical machine learning models (i.e., SVM, support vector machine and CRF, conditional random fields) [19]. However, this approach not be adapted to new domains efficiently as the designed features might be unavailable or less effective in the new domains.

And the third direction is connected to the deep learning approach which has been recently shown its ability to effectively exploit the word embedding via multiple layers of neural networks [20, 6]

3. Semantic Annotation of Definitions of Mathematical Objects

A mathematical text contains a set of structural elements, such as definition, theorem, lemma, proof, etc. Automatic methods for recognizing the structural elements of mathematical texts can help extract from the text relevant data that can be used to create specialized knowledge bases of mathematical objects, and, in particular, to replenish mathematical ontologies.

The OntoMath$^{\text{Edu}}$ ontology developed by the authors of this paper includes a hierarchy of mathematical objects whose elements should have corresponding definitions which are extracted, in a general case, from various sources of knowledge. Moreover, since the ontology is focused on educational applications and supports various educational levels, it becomes necessary to describe mathematical objects at different levels of abstraction depending on the educational level. Thus, we need to classify definitions by level of complexity. The developed method for annotating definitions of mathematical objects makes it possible to extract definitions of various structural complexities with graphic and text components from mathematical texts.

Let us describe the main ideas of the approach developed to implement the method of extracting definitions from mathematical texts. Each structural element of the text is associated with a set of initial and final lexical-syntactic patterns through which the corresponding segments are recognized in the text. Recognizing the exact boundaries of segments is generally a very complicated procedure for automatic methods, so in some cases, only approximate estimates of the boundaries of segments of structural elements can be obtained, especially for structural elements with complex semantics (definition, statement, etc.). One of the solutions is generation of signal lexical-syntactic patterns which indicate that within the boundaries of a fixed fragment (most often a sentence) a given structural element is present.

To highlight the definition, we also used sets of start and end tags of the constructed definition model.

The developed method for annotating the structural elements of a mathematical text is based on lexical-syntactic laws and rules that describe the lexical-syntactic features of using the elements in scientific, technical, and educational texts in Russian.

Thus, to build a model of a structural element of a text, it is necessary:

- to highlight the set of lexical and syntactic patterns characterizing the structural elements of the text. For this purpose, we use collections of mathematical texts;
- to describe the lexical composition and syntactic models of the selected patterns;
- to develop methods for recognizing lexical-syntactic patterns in mathematical textbooks, taking into account potential ambiguity, which in some cases directly correlates with pattern type recognition.

To write lexical-syntactic patterns, we use a special language of syntactic patterns in which a specific pattern is represented as a finite sequence of tokens with specified grammatical characteristics and fixed semantics.

The semantic annotation of the structural elements of a text is represented in XML notation. The list of lexical and syntactic patterns in the current version is open and serves to refine the method.

Template syntax models include the following groups of elements:

- introductory words (for example: *therefore*, *thus*, *it means*, and others);
- conjunctions and particles (for example: *if*, *since*);
- collocations;
- syntactic models such as “NP_1 – *это* NP_1” and “NP_1 *называется* NP_5” (NP_1 *is* NP_1). These models include the morphological characteristics of each unit. In these examples, NP_1 is a nominal group in the Nominative case and NP_5 is a nominal group in the Instrumental case (specific word forms are highlighted in italics). For example: the syntactic model “$PP_5|$<под>$+NP_5+V_{p3n1t1v2}+NP_1$” (Preposition + Noun phrase in the Instrumental case + Verb, 3^{rd} person ($p3$), singular ($n1$), Present tense ($t1$), Passive voice ($v2$)+ Noun phrase in the Nominative case) is implemented in a Russian context such as “*под* NP_5 *понимается* NP_1” (NP_1 *is understood as* NP_5);

- a cliché (for example, <admits a form>, <satisfies a condition>);
- elements of paratext (heading units such as Definition, Theorem, Lemma, Proof; also special symbol "□" which is the sign of the end of a proof, etc.).

Consider the main problems of the method of extracting definitions from mathematical texts.

From the point of view of its syntactic structure, the definition in mathematical documents is defined by a standard set of structural schemes:

1. NP_1 Cop NP_1,
2. NP_5 $V_{p3n1t1v2}$ NP_1,
3. NP_1 $AbrV^-$ NP_5,
4. NP_1 $Cop_{n2f1t3}V^+inf$ NP_5,

where NP_1 is a noun phrase in the Nominative case and NP5 is a noun phrase in the Instrumental case; *Cop* is a copula; $V_{p3n1t1v2}$ is a verb in the form of 3^{rd} person ($p3$), singular ($n1$), Present tense ($t1$), Passive voice ($v2$); $AbrV^-$ is a short form of passive participle; $Cop_{n2f1t3}V^+inf$ is an analytical verb form with features such as person ($p1$), plural ($n2$), Future tense ($t3$), and Active voice.

The scientific style is characterized by the use of constructions with verbs in the form of 3^{rd} person, plural ("we give a definition", "we will use the definition"). In Russian, the word form definition has two homonymous forms in the Nominative and Accusative cases. In contexts with verbs *to give*, *to apply* etc. it is used in the Accusative case whereas in headings it appears in the Nominative case.

Recognition of the boundaries of a definition fragment is carried out by using templates of the initial and final tags of this structural element. The list of initial tags contains a heading unit (a definition) and a set of verbal formulas (e.g. "we give a definition", "give a definition", "we will use a definition"). The text segment linked with the definition contains, as a rule, one sentence. Therefore, if the initial tag of the definition is set, then the final tag is set at the end of the corresponding sentence. Signal tags for the definition are represented by a set of verbal formulas that are built into certain syntactic models describing definition segments. Signal tags are contained within the definition segment and can probabilistically establish its boundaries (within the boundaries of the sentence containing the signal tag).

The syntactic models with a signal tag are (the signal tag is in italics):

- NP_1 - *<это/is>* NP_1,
- NP_1 *<названо/ was denoted>* NP_5,
- NP_1 *<называется/ is denoted>* NP_5,
- *<аналогично определяется/ is similarly denoted>* NP_1,
- *<назовем/ we denote>*,
- NP_1 *<будем называть/ we will denote>* NP_5.

Evaluation of the method of annotating the structural element “*Definition*” of a mathematical document using signal tags is based on the data of the experiment which was carried out using NLP models and methods implemented in the onto-logical-linguistic system “OntoIntegrator” [21, 22].

4. Experiment Description

For the experiment, a collection of educational texts was selected from the “I Class” website of a geometry course for grade 7 (https://www.yaklass.ru/p/geometria#program-7-klass). The educational texts relate to 4 studied sections: basic geometric knowledge, triangles, parallel lines, and relations between the sides and angles of a triangle. In each of the sections, there is a theory subsection containing basic definitions which are illustrated by drawings. Quite often, theoretical information at this level is given as a description of corresponding figures.

The purpose of the experiment was to extract all the definitions in the geometry course for grade 7 and to match each mathematical concept of a definition to the OntoMath$^{\mathrm{Edu}}$ ontology concept and also to annotate the ontology concept by a definition extracted from the text.

The annotation model of the structural element “*Definition*” was implemented using the “OntoIntegrator” ontological-linguistic system. Using the designed approach, we developed the conceptual model “*Definition*” implemented in the model ontology, which is the basic component of the “OntoIntegrator” system.

The graphic structure of the conceptual model “*Mathematical Definition*” is shown in Figure 2. The vertices of the graph are concept models from the model ontology, and the edges are relations defined in the model ontology. Here the “*Model Aggregation*” relation is used.

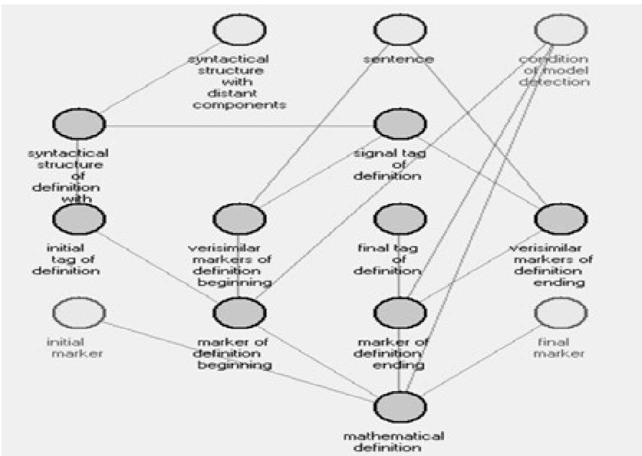


Figure 2: The graphic structure of the conceptual model “*Mathematical Definition*”

The “*Mathematical definition*” conceptual model allows us to detect any definitions in texts. The conceptual structure contains two main types of concept models: concept properties (light gray background) and concept models of the m-implementation type (dark gray background).

All the necessary parameters for a complete description of the mathematical definition model are transmitted by the parameters of the "*model aggregation*" relationship, to which all elements are linked. Among the concept models of the m-implementation type, we use syntactic models that are connected with the search procedures for discontinuous structures and sentence segmentation and semantic models that are connected with identification of text boundaries of a definition, as well as with verification of various semantic search conditions for the required entities. The "*Signal tag of the definition*", "*Initial tag of the definition*", "*Final tag of the definition*", "*Verisimilar markers of definition beginning*", and "*Verisimilar markers of definition ending*" conceptual models determine the conditions for detecting the beginning, the middle and the end of the mathematical definition by different procedures. Such detection is based both on the syntactic properties of the definition and on the ontological markup of the text by the OntoMath^{Edu} ontology concepts.

We have analyzed the results of the experiment and distinguished several types of definitions. Further, in the examples, the ontological object will be highlighted in italic.

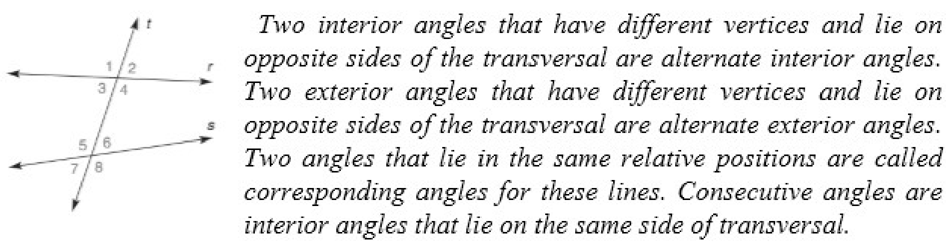
The first type includes classical definitions in which object X (an ontology object) is defined as "is denoted the Xth/called the Xth" (e.g. A circle is the set of all points in a plane equidistant from a fixed point in the plane called the center).

The second type, in contrast to the first one, gives distant construction of the object name in Russian (e.g. "Если обе стороны угла лежат на одной прямой, угол называют развёрнутым / If both sides of the angle are on one straight line, the angle is called the straight [angle]") or more complex example as "the triangle side ... opposite to the corner". In this case we need to reformulate (to bundle) the parts of the object name ("in our example as *сторона, противолежащая угл*"(RU)/ "a side of a triangle that subtends *the opposite angle*"). We developed a special procedure for processing distant names to recognize the names of an ontology object.

The third type is also a classic type of definition in which the defined object X (an ontology object) is specified in the syntactical construction "*X is this Y*" (e.g. "*A line segment*, or a segment, is a set of points consisting of two points on a line, called endpoints, and all of the points on the line between the endpoints"). This model allowed us to identify some set of geometric objects that are missed in the current version of the ontology (e.g. "*A dimension* is a comparison of a measurement object with a selected unit of measure").

The fourth type includes mathematical expressions in the definition body (e.g. "AB is the side opposite $\angle C$ and that $\angle C$ is the angle opposite side AB").

The fifth type includes a drawing in the definition body as in the example below.



Two interior angles that have different vertices and lie on opposite sides of the transversal are alternate interior angles. Two exterior angles that have different vertices and lie on opposite sides of the transversal are alternate exterior angles. Two angles that lie in the same relative positions are called corresponding angles for these lines. Consecutive angles are interior angles that lie on the same side of transversal.

Figure 3: The definition contains a picture

The sixth type is a complex definition consisting of several sentences (e.g. "*A circle* is the set of all points in a plane equidistant from a fixed point in the plane called *the center. A radius of a circle* is a line segment from the center of the circle to any point of the circle"). The main difficulty is that the boundaries of such a definition are not clear. The current example of automatic detection of such cases gives an error in the placement of the beginning tag, due to the fact that the first sentence does not include the appropriate signal tag. Automatic analysis of this type of definition requires additional semantic analysis, which is our future research.

In total, 56 definitions for various geometric objects and their properties are extracted from 4 studied sections from the collection of educational texts in our experiment. Table 1 at Appendix contains examples of definitions extracted automatically from school texts. Statistics on extracted definitions are given in Table 2 (if a definition was extracted only partially, we count it as 0.5 definition).

Thus, the method precision of extracting definitions in a text on geometry is 91% (5 erroneous definitions, including incomplete ones) and the method recall is 87.5% (7 definitions were not extracted from texts). F-measure is 89.2%.

The sufficiently high precision of extracting definitions in texts allows us to automatically annotate definitions in a text body, and also to autocomplete annotations of ontology concepts.

5. Conclusion

In this paper we presened a template-based method for populating the OntoMath$^{\mathrm{Edu}}$ ontology by definitions, automatically extracted from educational mathematical texts in Russian. The method has been implemented on the base of the "OntoIntegrator" system and evaluated on a small collection of educational texts from the yaklass.ru website. The obtained F-measure is 89.2%. As a future work we are going to:

1. apply the proposed method to other text collections;
2. complement a template-based approach by deep learning ones;
3. extend the developed method for extracting definitions from professional mathematical texts and apply it for population the ontology of professional mathematics OntoMath$^{\mathrm{PRO}}$, starting with the *Computability theory* domain.

Acknowledgments

The work was funded by Russian Foundation of Basic Research according to the research projects no. 19-29-14084 and 20-31-70012.

References

[1] A. Kirillovich, O. Nevzorova, M. Falileeva, E. Lipachev, L. Shakirova, OntoMath$^{\mathrm{Edu}}$: Towards an educational mathematical ontology, in: C. Kaliszyk, et al. (Eds.), Workshop Papers at 12th Conference on Intelligent Computer Mathematics (CICM-WS 2019), CEUR Workshop Proceedings (forthcoming), ????

[2] A. Kirillovich, O. Nevzorova, M. Falileeva, E. Lipachev, L. Shakirova, OntoMath^{Edu}: a new linguistically grounded educational mathematical ontology, in: C. Benzmüller, B. Miller (Eds.), Proceedings of the 13th International Conference on Intelligent Computer Mathematics (CICM 2020), Lecture Notes in Artificial Intelligence, vol. 12236, Springer, 2020. https://doi.org/10.1007/978-3-030-53518-6_10.

[3] A. Elizarov, A. Kirillovich, E. Lipachev, O. Nevzorova, Digital ecosystem OntoMath: Mathematical knowledge analytics and management, in: L. Kalinichenko, S. Kuznetsov, Y. Manolopoulos (Eds.), XVIII International Conference on Data Analytics and Management in Data Intensive Domains (DAMDID/RCDL 2016), Communications in Computer and Information Science, vol. 706, Springer, 2017, pp. 33–46. https://doi.org/10.1007/978-3-319-57135-5_3.

[4] A. Elizarov, A. Kirillovich, E. Lipachev, O. Nevzorova, Semantic formula search in digital mathematical libraries, in: Proceedings of the 2nd Russia and Pacific Conference on Computer Technology and Applications (RPC 2017), IEEE, 2017, pp. 39–43. https://doi.org/10.1109/RPC.2017.8168063.

[5] A. M. Elizarov, A. V. Kirillovich, E. K. Lipachev, A. B. Zhizhchenko, N. G. Zhil'tsov, Mathematical knowledge ontologies and recommender systems for collections of documents in physics and mathematics, Doklady Mathematics 93 (2016) 231–233. https://doi.org/10.1134/S1064562416020174.

[6] A. P. B. Veyseh, F. Dernoncourt, D. Dou, T. H. Nguyen, Definition extraction using linguistic and structural features, in: M. Walker, H. Ji, A. Stent (Eds.), Proceedings of the 34th AAAI Conference on Artificial Intelligence (AAAI 2020), vol. 34, no. 05: AAAI-20 Technical Tracks, AAAI, 2020, pp. 9098–9105. https://doi.org/10.1609/aaai.v34i05.6444.

[7] J. L. Klavans, S. Muresan, Evaluation of the DEFINDER system for fully automatic glossary construction, Proceedings of the American Medical Informatics Association Symposium (2001) 324–328. https://www.ncbi.nlm.nih.gov/pubmed/11825204.

[8] E. Bolshakova, N. Efremova, A heuristic strategy for extracting terms from scientific texts, in: M. Y. Khachay, et al. (Eds.), Revised Selected Papers of the 4th International Conference on Analysis of Images, Social Networks and Texts (AIST 2015), Communications in Computer and Information Science, vol. 542, Springer, 2015, pp. 297–307. https://doi.org/10.1007/978-3-319-26123-2_29.

[9] E. Bolshakova, K. Ivanov, Term extraction for constructing subject index of educational scientific text, in: Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference "Dialogue", 2018, pp. 143–149. http://www.dialog-21.ru/media/4291/bolshakovaei_ivanovkm.pdf.

[10] E. Bolshakova, N. Efremova, K. Ivanov, Terminological information extraction from Russian scientific texts: Methods and applications, in: G. Wohlgenannt, et al. (Eds.), Proceedings of 3rd Workshop on Computational linguistics and language science (CLLS 2018), EPiC Series in Language and Linguistics, vol. 4, EasyChair, 2019, pp. 95–106. https://doi.org/10.29007/k93q.

[11] Lexico-syntactic pattern language, ???? URL: http://lspl.ru, http://lspl.ru, last accessed 2020/05/08.

[12] A. Przepiórkowski, Ł. Degórski, M. Spousta, K. Simov, P. Osenova, L. Lemnitzer, V. Kuboň, B. Wójtowicz, Towards the automatic extraction of definitions in Slavic, in: J. Piskorski,

H. Tanev (Eds.), Proceedings of the Workshop on Balto-Slavonic Natural Language Processing, ACL, 2007, pp. 43–50. https://www.aclweb.org/anthology/W07-1706.

[13] A. Storrer, S. Wellinghoff, Automated detection and annotation of term definitions in german text corpora, in: N. Calzolari, et al. (Eds.), Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC'06), ELRA, 2006, pp. 2373–2376. https://www.aclweb.org/anthology/L06-1066/.

[14] I. Fahmi, G. Bouma, Learning to identify definitions using syntactic features, in: R. Basili, A. Moschitti (Eds.), Proceedings of the Workshop on Learning Structured Information in Natural Language Applications, ACL, 2006, pp. 64–71. https://www.aclweb.org/anthology/W06-2609/.

[15] R. Snow, D. Jurafsky, Y. N. Andrew, Learning syntactic patterns for automatic hypernym discovery, in: L. Saul, Y. Weiss, L. Bottou (Eds.), Proceedings of the 17th International Conference on Neural Information Processing Systems (NIPS 2004), MIT Press, 2005, pp. 1297–1304. https://papers.nips.cc/paper/2659-learning-syntactic-patterns-for-automatic-hypernym-discovery.

[16] V. Shwartz, E. Santus, D. Schlechtweg, Hypernyms under siege: Linguistically-motivated artillery for hypernymy detection, in: M. Lapata, P. Blunsom, A. Koller (Eds.), Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2017). Volume 1: Long Papers, ACL, 2017, pp. 65–75. https://www.aclweb.org/anthology/E17-1007/.

[17] C. Borg, M. Rosner, G. Pace, Evolutionary algorithms for definition extraction, in: G. Sierra, M. Pozzi, J.-M. Torres (Eds.), Proceedings of the 1st Workshop on Definition Extraction (WDE '09), ACL, 2009, p. 26–32. https://www.aclweb.org/anthology/W09-4405/.

[18] C. Borg, Automatic definition extraction using evolutionary algorithms. Master's thesis, University of Malta, 2009. http://staff.um.edu.mt/cbor7/publications/2009Thesis.pdf.

[19] E. Westerhout, Definition extraction using linguistic and structural features, in: G. Sierra, M. Pozzi, J.-M. Torres (Eds.), Proceedings of the 1st Workshop on Definition Extraction (WDE '09), ACL, 2009, pp. 61–67. https://www.aclweb.org/anthology/W09-4410/.

[20] L. Espinosa-Anke, S. Schockaert, Syntactically aware neural architectures for definition extraction, in: M. Walker, H. Ji, A. Stent (Eds.), Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 2018). Volume 2: Short Papers, ACL, 2018, pp. 378–385. https://doi.org/10.18653/v1/N18-2061.

[21] O. Nevzorova, V. Nevzorov, Ontology-driven processing of unstructured text, in: S. Kuznetsov, A. Panov (Eds.), Proceedings of the 17th Russian Conference on Artificial Intelligence (RCAI 2019), Communications in Computer and Information Science, vol. 1093, Springer, 2019, pp. 129–142. https://doi.org/10.1007/978-3-030-30763-9_11.

[22] O. Nevzorova, V. Nevzorov, A. Kirillovich, A syntactic method of extracting terms from special texts for replenishing domain ontologies, in: Proceedings of the 2nd Russia and Pacific Conference on Computer Technology and Applications (RPC 2017), IEEE, 2017, pp. 127–131. https://doi.org/10.1109/RPC.2017.8168083.

Appendix

Table 1
Examples of extracted definitions

#	Concept	Definition (Russian)	Definition (English)
1	Отрезок ‘Line segment’	Часть прямой, ограниченная двумя точками, называется отрезком	The part of the line limited by two points is called a segment
2	Развернутый угол ‘Straight angle’	Если обе стороны угла лежат на одной прямой, угол называют развернутым	If both sides of the angle are on one straight line, the angle is called the straight [angle]
3	Противолежащая сторона треугольника ‘Opposite side of a triangle’, Противолежащий угол треугольника ‘Opposite angle of a triangle’	Сторону, которая лежит напротив угла, называют противолежащей углу, и угол называют противолежащим стороне.	The side that lies opposite to the corner is called the opposite to the corner, and the corner is called the opposite to the side.
4	Треугольник ‘Triangle’	Треугольник – это геометрическая фигура, образованная тремя отрезками, которые соединяют три не лежащие на одной прямой точки.	A triangle is a geometric figure formed by three segments that connect three points not lying on one straight line.
5	Параллельные прямые ‘Parallel lines’	На плоскости две прямые a и b, которые не пересекаются, называются параллельными и обозначаются $a \parallel b$.	On the plane, two lines a and b that do not intersect are called parallel and denoted by $a \parallel b$.
6	Накрест-лежащие углы ‘Alternate interior angles’, Соответственные углы ‘Corresponding angles’, Односторонние углы ‘Consecutive angles’	Если две прямые пересекает третья прямая, то углы называются так <Image>: накрест лежащие углы: $\angle 3$ и $\angle 5$; $\angle 2$ и $\angle 8$; соответственные углы: $\angle 1$ и $\angle 5$, $\angle 4$ и $\angle 8$, $\angle 2$ и $\angle 6$, $\angle 3$ и $\angle 7$; односторонние углы: $\angle 3$ и $\angle 8$, $\angle 2$ и $\angle 5$.	If two lines intersect the third line, then the angles are called like this <Image>: angles lying crosswise: $\angle 3$ and $\angle 5$, $\angle 2$ and $\angle 8$, corresponding angles: $\angle 1$ and $\angle 5$, $\angle 4$ and $\angle 8$, $\angle 2$ and $\angle 6$, $\angle 3$ and $\angle 7$; one-sided angles: $\angle 3$ and $\angle 8$, $\angle 2$ and $\angle 4$.
7	Хорда ‘Chord’	Отрезок, который соединяет две точки на окружности, называют хордой	A line that connects two points on a circle is called a chord.

Table 2
Statistics on extracted definitions

#	URL	Extracted definitions	
		manually	automatically
1	https://www.yaklass.ru/p/geometria/7-klass/nachalnye-geometricheskie-svedeniia-14930/priamaia-i-otrezok-9703/re-18f77739-2ab6-4f1a-b5c0-049e88127967	1	1
2	https://www.yaklass.ru/p/geometria/7-klass/nachalnye-geometricheskie-svedeniia-14930/luch-i-ugol-9658/re-ac00706b-b905-490e-9e79-4d4c1566de6a	7	5.5
3	https://www.yaklass.ru/p/geometria/7-klass/nachalnye-geometricheskie-svedeniia-14930/sravnenie-otrezkov-i-uglov-12147/re-dbeceeb6-0f52-403c-b561-71bbaa8eafc5	4	3
4	https://www.yaklass.ru/p/geometria/7-klass/nachalnye-geometricheskie-svedeniia-14930/izmerenie-otrezkov-i-uglov-9704/re-8118f3d0-7a8f-4f3a-91cc-9e12cff98c74	6	6
5	https://www.yaklass.ru/p/geometria/7-klass/nachalnye-geometricheskie-svedeniia-14930/perpendikuliarnye-priamye-9886/re-3cce9aa8-9bff-4fa4-b214-017612e69d4a	3	3
6	https://www.yaklass.ru/p/geometria/7-klass/treugolniki-9112/pervyi-priznak-ravenstva-treugolnikov-9122/re-27c5cb9c-c428-473d-924c-17cb95d18acc	7	7
7	https://www.yaklass.ru/p/geometria/7-klass/treugolniki-9112/mediany-bissektrisy-i-vysoty-treugolnika-9481/re-56c524c8-9727-48db-9926-95988d203d40	13	12
8	https://www.yaklass.ru/p/geometria/7-klass/treugolniki-9112/vtoroi-i-tretii-priznaki-ravenstva-treugolnikov-9739/re-8a326c61-77a4-4f4c-8c5e-26f90695a4fa	1	1
9	https://www.yaklass.ru/p/geometria/7-klass/treugolniki-9112/zadachi-na-postroenie-10433/re-b5a2c2a4-5b38-4bef-b8f0-3ebb5cae946f	6	5.5
10	https://www.yaklass.ru/p/geometria/7-klass/parallelnye-priamye-9124/priznaki-parallelnosti-dvukh-priamykh-aksioma-parallelnykh-priamykh-9228/re-1e38c190-6fee-47d7-9380-d1e0d2858c37	2	2
11	https://www.yaklass.ru/p/geometria/7-klass/parallelnye-priamye-9124/priznaki-parallelnosti-dvukh-priamykh-aksioma-parallelnykh-priamykh-9228/re-4ba7ee5b-3478-495b-b7eb-3e4eeb2d9b4c	3	2
12	https://www.yaklass.ru/p/geometria/7-klass/sootnoshenie-mezhdu-storonami-i-uglami-treugolnika-9155/summa-uglov-treugolnika-9171/re-b78850d5-a0e0-4093-bad3-7e82a520e7d7	1	1
13	https://www.yaklass.ru/p/geometria/7-klass/sootnoshenie-mezhdu-storonami-i-uglami-treugolnika-9155/sootnosheniia-mezhdu-storonami-i-uglami-treugolnika-9738/re-8ff8415c-958d-4520-9f48-54b6707fe2c9	0	0
14	https://www.yaklass.ru/p/geometria/7-klass/sootnoshenie-mezhdu-storonami-i-uglami-treugolnika-9155/priamougolnye-treugolniki-9175/re-cef42b35-127b-4350-ac33-e249179f4160	0	0
15	https://www.yaklass.ru/p/geometria/7-klass/sootnoshenie-mezhdu-storonami-i-uglami-treugolnika-9155/postroenie-treugolnikov-po-trem-elementam-12420/re-c4d19cfc-02e9-45ed-a6bd-1921b7bbfd92	2	2
Total		56	51