The Semantic Context Models of Mathematical Formulas in Scientific Papers

© Olga Nevzorova
Kazan Federal University, Kazan, Russia
Research Institute of Applied Semiotics of the Tatarstan Academy of Sciences, Kazan, Russia
onevzoro@gmail.com

© Alexander Kirillovich
Kazan Federal University, Kazan, Russia
alik.kirillovich@gmail.com

© Vladimir Nevzorov
Kazan National Research Technical University named after A.N. Tupolev, Kazan, Russia
nevzorovvn@gmail.com

© Konstantin Nikolaev
Kazan Federal University, Kazan, Russia
konnikolaeff@yandex.ru

Abstract. This paper describes the results of semantic annotating and semantic search in a mathematical collection. We explored semantic models of contexts of a mathematical formula and applied the new knowledge to improving semantic search. The solutions are based on the application of the previously developed OntoMath^{Pro} ontology.

Keywords: ontology, mathematical paper, semantic annotation, semantic search

1 Introduction

In the field of professional mathematics, considerable experience has been accumulated in processing and use of electronic mathematical content in projects for developing mathematical electronic libraries. Being well-structured, mathematical texts are favorably distinguished by presence of standards and markup languages, availability of electronic mathematical libraries with highly developed search services, as well as of software for the automatic processing of individual components of mathematical texts. Works devoted to formalizing the level of representations of mathematical articles are rather widely spread today. For these purposes, specialized formal languages for representing mathematical texts, as well as software for converting these languages [1-3] have been developed. Semantic publication of articles implies an increase in the number of computational components of texts, i.e. components of a certain semantics that are extracted from the text for further processing. Mathematical search is also an actively developing field of research. Widely known are specialized mathematical search engines, for example, (uni)quation (http://uniquation.com/en/), Springer LaTEX Search (http://latexsearch.com/), and Wolfram Formula Search (http://functions.wolfram.com/).

Proceedings of the XX International Conference "Data Analytics and Management in Data Intensive Domains" (DAMDID/RCDL'2018), Moscow, Russia, October 9-12, 2018

The mathematical projects carried out by the world community paved the way for a new idea - creating the World Digital Mathematical Library (WDML) [4-7]. The main tasks of building WDML and the technologies needed to solve them, were discussed in 2014-2015 by a wide circle of mathematicians and are fixed in a number of documents adopted by the World Mathematical Union. In particular, it was suggested that the next step in the development of the WDML project would be to go beyond traditional mathematical publications and build a knowledge-based network of information contained in these publications. The WDML project is focused on the object system of organizing and storing mathematical knowledge. Unlike traditional electronic mathematical libraries, where the electronic document is the storage unit in the database, it is proposed to represent the mathematical knowledge of the document collection in the form of a specially organized repository of mathematical objects.

Theorems, axioms, proofs, mathematical definitions, etc are standard classes of mathematical objects. The semantic relationships between elements are important constituents of the object model. To represent the document's object model, it is proposed to use modern technologies of the semantic web. This way of representing the management of mathematical knowledge will allow creating management tools directly by objects of mathematical knowledge (means of aggregation, semantic search, search by formulas and identification of identical objects). In this paper, we will consider some solutions to the problem of semantic search by formulas that take into account the semantic context of mathematical formulae.

The paper is organized as follows. In Section 2, we present a semantic search service for math formulas. In Section 3, we discuss basic methods for semantic annotation and search, underlying this service. In Section 4 we discuss how these basic methods can be improved by context models of mathematical formulas. In Section 5 we made concluding remarks.

2 Semantic search of mathematical formulas

In this section, we present OntoMath Formula Search (https://lobachevskii-dml.ru/mathsearch) [8], a semantic search service for mathematical formulas (Fig. 1).

While there are currently many math formula search engines have been developed at present time, they are mostly syntactical, and allow only simple search by expression patterns. For example, to find formulas, containing “$(a + b)^2$”. Unlike them, our service is semantic, and, therefore, can find formulas with respect to a given math concept, regardless of its symbolic representation. For example, to find formulas that contain variables, expressing *open sets*.

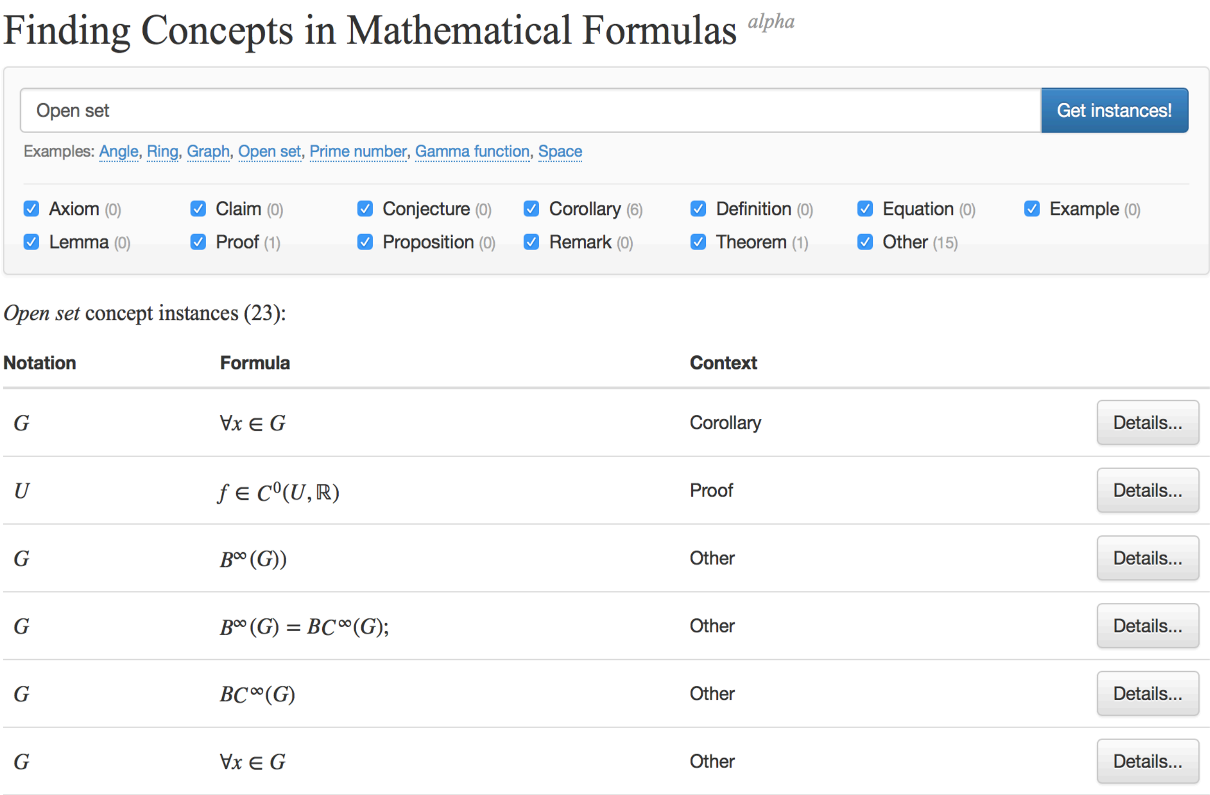

Fig.1. A search results window of OntoMath Formula Search

The search index stores journal publications in mathematics.

The user enters keywords, filtering suggestions of the system. The search interface supports the terminology of OntoMath[Pro] ontology [9] as well as parts of other sources aligned onto it. The search result table has three columns, such as notation, formula, and context. In each row, the first column contains a variable standing for the query concept in the relevant formula. The second column contains the relevant formula. The third column provides the document fragment, containing the formula. To limit search results, the user can filter the structural context.

Using logical reasoning with respect to the ontology relations, the service finds formulas that contain not only the given concept, but also subclasses of this concepts in the ontology hierarchy. For example, the search results for *polygon* query contains formulas with variables, denoting such concepts as *triangle*, *parallelogram*, *trapezoid*, and so on.

OntoMath Formula Search is a part of OntoMath digital ecosystem, an ecosystem of ontologies, text analytics tools, and applications for math knowledge management [10-11].

The core component of the OntoMath ecosystem is its semantic publishing platform [12-13]. This platform takes as an input a collection of mathematical papers in LaTex format and builds their ontology-based Linked Open Data representation. The generated mathematical dataset includes metadata, the logical structure of documents, terminology, and mathematical formulas, bound to terms.

The semantic publishing platform pipeline consists of several steps, one of which is semantic annotation of mathematical texts. In the next two sections, we describe the current implementation of this steep, based on simple context-agnostic models and discuss how this implementation can be improved by context models.

3 Semantic annotation of mathematical texts

The developed model of semantic annotation of mathematical texts is based on the professional mathematics ontology OntoMath^{Pro} [5] which is described in OWL-DL / RDFS language. The current version consists of 3450 classes, 6 types of object properties, 3630 instances of IS-A properties, and 1140 instances of remaining properties. The objects of semantic annotation are also formulas and formula-related fragments of text that define descriptions of variables in formulas.

Semantic annotation of mathematical texts is carried out by the system "OntoIntegrator" [14] which belongs to the class of ontology-linguistic systems. Annotation is carried out in several consecutive stages: tokenization, segmentation of sentences, morphological analysis, and identification of nominal groups.

Annotation is performed for mathematical documents in XML format, where the <Math> tag marks mathematical expressions. Mathematical expressions can be included in a noun phrase, and can also be used as prefixes in constructions with hyphens. At the tokenization stage, structures with the <Math> tag are treated as separate objects. In a mathematical text, there are four types of noun phrases (NP) which are marked as TERM1-TERM4 tags (Table 1).

Table 1 Markup of noun phrases

Tag	Noun phrase	Markup example
TERM1	NP without prepositions, contains no formulas	<TERM1 Form="теорема" HeadBegin="1" HeadEnd="7"> Теоремы существования </TERM1>
TERM2	NP without prepositions, contains formulas	<TERM2 Form="блок" HeadBegin="1" HeadEnd="4"> блок размерности <Math mode="inline" tex="k" text="k"> <XMath> <XMTok role="UNKNOWN" font="italic">k</XMTok> </XMath> </Math> </TERM2>
TERM3	Prepositional NP, contains no formulas	<TERM3 Form="ядро" HeadBegin="3" HeadEnd="7"> с ядром оператора системы </TERM3>
TERM4	Prepositional NP, contains formulas	<TERM4 Form="область" HeadBegin="13" HeadEnd="19"> в замкнутой области <Math mode="inline" tex="U" text="U"> <XMath> <XMTok role="UNKNOWN">U</XMTok> </XMath> </Math> </TERM4>

The construction of nominal groups is described by rules that take into account the internal structure of the nominal group. The noun phrases allocation method works within the bounds of the sentence. When a noun phrase is annotated, the borders of the head word are highlighted with HeadBegin and HeadEnd attributes, its grammatical form is normalized (reduced to the canonical form), and the normal form of NP appears in the Form attribute. Annotated noun phrases are used in the following stages of processing mathematical texts - when processing extended formula contexts.

Among all the noun phrases extracted from the texts on the basis of syntactic models, noun groups containing ontological concepts from the OntoMath^{Pro} ontology are selected. Such nominal groups will be denoted as NP (O_i) and PP (O_i), where O_i stands for an element of the ontology.

The distribution of nominal groups for the above-mentioned types in the experimental collection of 40 mathematical articles is shown in Figure 2. A total of 11865 registered groups are identified in the collection.

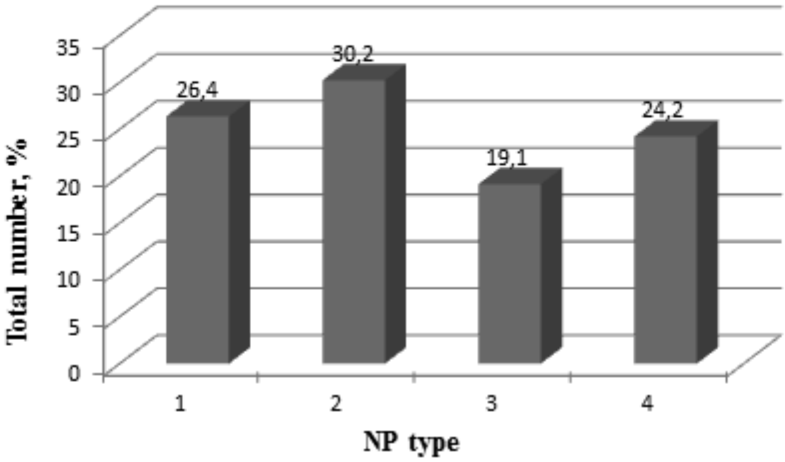

Fig.2. The distribution of noun phrases by type in mathematical texts

4 Improving search based on the models of the formula's semantic context

Analysis of the semantic context of the mathematical formula is aimed at improving search results, primarily in finding a formula using a textual description of formula variables. To solve the problem, the following models of the semantic context of the formula were proposed.

1) **Model of structural classification of formulas**. According to this model, all mathematical formulas are divided into two classes: the class of "basic" formulas F1 and the class of "auxiliary" formulas F2. Informally, a "basic" formula is a formula, that express significant relationship between mathematical concepts. This informal notion is formalized as follows: mathematical formulas from F1 class have to be longer than 10 characters and include the basic mathematical operators, as well as binary equality and inequality operators.

The distinction between "basic" and "auxiliary" formulas is crucial in the two other context models (see below). Additionally, this distinction is in use in the semantic search service, where search results are restricted to "basic" formulas.

We have developed a tool for automatic classification and have evaluated it on the collection of mathematical papers (50 items) from the "Russian Mathematics (Iz. VUZ)" journal. To evaluate this tool, we just compare a count of formulas classified as F1 automatically by this tool, with a count of formulas classified as F1 manually by a human assessor. The result of evaluation for ten papers is presented in Table 2.

Table 2. The results of evaluation of the classification tool

Paper number	All formulas	$F1_{auto}$ count	$F1_{manual}$ count
1	220	69	59
2	103	50	60
3	150	42	30
4	383	69	60
5	135	70	50
6	236	73	61
7	223	81	71
8	154	73	65
9	191	80	75
10	161	62	48

2) **The expanded syntactic context of the variable formula description,** allowing to take into account the compatibility (grouping) of the types of noun groups shown above. Various combinations of noun phrases were studied. The most informative was found to be the construction of the form $NP_1(O_1) + PP_2(O_2)$, which makes it possible to form a hypothesis about the connection between the two classes of ontology and determine the stable relation between noun phrases.

If NP and PP are found in the neighborhood of a formula from class F1, and they follow each other directly and contain ontological concepts in their composition, that is, they satisfy the pattern $NP_{12}(O_1, O_2) = NP_1(O_1) + PP_2(O_2)$, where symbols O_1 and O_2 are OntoMath^{Pro} ontology concepts, then the compound noun phrase NP_{12} is allocated. A comparison of text strings was carried out using the Jaccard measure (the threshold value was set to 0.7).

The selected pairs of O_1 and O_2 concepts will undergo an expert analysis to confirm the existence of a significant link between these concepts. These are examples of pairs found by the NP + PP model associated with ontological concepts (Table 3).

Table 3. Examples of noun phrases (NP + PP)

#	$NP_1(O_1)$	$PP_2(O_2)$
1	пространство линейных операторов $L : \mathbb{R}\, n \to \mathbb{R}\, n$	с нормой $\vert L \vert \doteq \max \vert x \vert \leq 1 \vert L x \vert$
	the space of linear operators $L : \mathbb{R}\, n \to \mathbb{R}\, n$	with the norm $\vert L \vert \doteq \max \vert x \vert \leq 1 \vert L x \vert$
2	пространство Марцинкевича $M_q(I^d)$	с нормой $\Vert f \Vert_{q,\infty} = \sup\limits_{t\in[0,1]} t^{\frac{1}{q}-1} \int_0^t f^*(\tau)\, d\tau$
	Martsinkevich space $M_q(I^d)$	with the norm $\Vert f \Vert_{q,\infty} = \sup\limits_{t\in[0,1]} t^{\frac{1}{q}-1} \int_0^t f^*(\tau)\, d\tau$
3	производную $(\partial/\,\partial\xi_2)U^S(\xi) \equiv U_1^S(\xi)$	в представлении $U(\xi) = U^R(\xi) + U^S(\xi), \xi \in \bar{\bar{D}}^1$;
	a derivative $(\partial/\,\partial\xi_2)U^S(\xi) \equiv U_1^S(\xi)$	in the view $U(\xi) = U^R(\xi) + U^S(\xi), \xi \in \bar{\bar{D}}^1$;
4	множество точек пересечения прямых $x_s = x_s^{i_s}$	с границей γ
	the set of points of intersection of lines $x_s = x_s^{i_s}$	with the bound γ

We evaluated the experiments on extracting expanded syntactic contexts of variable formulas which were performed using the collection of 50 papers under analysis.

To compare the labels of ontological concepts and the textual description of formula variables, we used the Jaccard metric in the form k = (k1 + k2) / 2, where symbols k1 and k2 are the Jaccard coefficients obtained by comparing pairs (NP, O1), (PP, O2). The results are represented in Table 4. The analysis of the results allows us to conclude that the selected pairs demonstrate stable links between concepts in the ontology, which thus makes it possible to improve the OntoMath^{Pro} ontology.

Table 4. Examples of semantically related noun phrases

NP+PP	O1+O2	K1	K2	K
(круг F_1 + на область F_2) / *circle F_1 + on the domain F_2*	(круг F_1 +область F_2) / *circle F_1+ domain F_2*	1	1	1
(область F_1 + на область F_2) / *domain F_1+ on the domain F_2*	(область F_1 + область F_2) / *domain F_1 + domain F_2*	1	1	1
(проекция F_1 + на касательное пространство F_2) / *projection F_1+ onto the tangent space F_2*	(проекция F_1 + касательное пространство F_2) / *projection F_1 + tangent space F_2*	1	1	1
(Векторное поле F_1 + на многообразии F_2) / *A vector field F_1 + on a manifold F_2*	(векторное поле F_1 + многообразие F_2) / *vector field F_1 + manifold F_2*	1	0,91	0,96
(подмногообразие F_1 + в пространстве F_2) / *submanifold F_1 + in space F_2*	(подмногообразие F_1 + пространство F_2) / *submanifold F_1 + space F_2*	1	0,91	0,95
(множество F_1 + в пространстве F_2) / *set F_1 + in space F_2*	(множество F_1 + пространство F_2) / *set F_1 + space F_2*	1	0,91	0,95
(оператор F_1 + По определению F_2) / *operator F_1 + By definition F_2*	(оператор F_1 + определение F_2) / *operator F_1 + definition F_2*	1	0,9	0,95
(оценка F_1 + На множестве F_2) / *bound F_1 + on the set F_2*	(оценка F_1 + множество F_2) / *bound F_1 + set F_2*	1	0,88	0,94

(Кокасательное пространство F_1 + к проективному пространству F_2) / *The cotangent space F_1 + to the projective space F_2*	(кокасательное пространство F_1 + проективное пространство F_2) / *cotangent space F_1 + projective space F_2*	1	0,87	0,93
(базис F_1 + с условием F_2) / *basis F_1 + with condition F_2*	(базис F_1 + условие F_2) / *basis F_1 + condition F_2*	1	0,87	0,93
(продолжение F_1 + до функции из F_2) / *continuation F_1 + to a function of F_2*	(продолжение F_1 + функция F_2) / *continuation F_1 + function F_2*	1	0,85	0,92
(матрица F_1 + в формуле F_2) / *matrix F_1 + in the formula F_2*	(матрица F_1 + формула F_2) / *matrix F_1 + formula F_2*	1	0,85	0,92
(оценка F_1 + при условии F_2) */ bound F_1 + on condition F_2*	(оценка F_1 + условие F_2) / *bound F_1 + condition F_2*	1	0,85	0,92
(кривая F_1 + в точках F_2) / *curve F_1 + at points F_2*	(кривая F_1 + точка F_2) / *curve F_1 + point F_2*	1	0,83	0,91
(оператор из F_1 + в подпространство F_2) / *operator from F_1 + to a subspace F_2*	(оператор F_1 + пространство F_2) / *operator F_1 + space F_2*	1	0,8	0,9
(предел F_1 + в норме F_2) / *the limit F_1 + in the norm F_2*	(предел F_1 + норма F_2) / *the limit F_1 + norm F_2*	1	0,8	0,9
(значения функции F_1 + в точках F_2) / *function values F_1 + at points F_2*	(значение функции F_1 + точка F_2) / *function values F_1 + point F_2*	0,93	0,83	0,88
(оператору F_1 + с условием F_2) / *operator F_1 + with condition F_2*	(оператор F_1 + условие F_2) / *operator F_1 + condition F_2*	0,88	0,87	0,88
(подмногообразием F_1 + со структурными уравнениями F_2) / *submanifold F_1 + with structural equations F_2*	(подмногообразие F_1 + структурные уравнения F_2) / *submanifold F_1 + structural equations F_2*	0,93	0,82	0,88
(уравнение F_1 + при условиях F_2) / *equation F_1 + under conditions F_2*	(уравнение F_1 + условие F_2) / *equation F_1 + conditions F_2*	1	0,75	0,87
(функции F_1 + с множества F_2) / *Functions F_1 + from the set F_2*	(функция F_1 + множество F_2) / *function F_1 + set F_2*	0,85	0,88	0,87
(функции F_1 + в области F_2) / *functions F_1 + in the domain F_2*	(функция F_1 + область F_2) / *function F_1 + domain F_2*	0,85	0,85	0,85
(оператор умножения F_1 + на функцию F_2) / *operator of multiplication F_1 + by a function F_2*	(оператор вложения F_1 + функция F_2) / *embedding operator F_1 + function F_2*	0,83	0,85	0,84
(поверхность F_1 + по кривым F_2) / *surface F_1 + by curves F_2*	(поверхность F_1 + кривая F_2) / *surface F_1 + curve F_2*	1	0,66	0,83
(пространство Лоренца F_1 + с нормой F_2) / *Lorentz space F_1 + with the norm F_2*	(пространство лоренца F_1 + норма F_2) / *Lorentz space F_1 + norm F_2*	1	0,66	0,83
(Главное подрасслоение F_1 + со структурной группой F_2) / *The main subfibration F_1 + with the structural group F_2*	(главное расслоение F_1 + структурная группа F_2) / *main subfibration F_1 + structural group F_2*	0,88	0,76	0,82
(любая проективная связность F_1 + на многообразии F_2) / *any projective connection F_1 + on a manifold F_2*	(проективная связность F_1 + многообразие F_2) / *projective connection F_1 + manifold F_2*	0,72	0,91	0,82
(слой F_1 + в локальную алгебру Вейля F_2) / *layer F_1 + into the local Weyl algebra F_2*	(слой F_1 + алгебра вейля F_2) / *layer F_1 + Weyl algebra F_2*	1	0,61	0,80
(условиям F_1 + для точки F_2) / *conditions F_1 + for a point F_2*	(условие F_1 + точка F_2) / *condition F_1 + point F_2*	0,75	0,8	0,77

(Векторное произведение F_1 + в новом базисе F_2) / *A vector product F_1 + in a new basis F_2*	(векторное произведение F_1 + голономный базис F_2) / *vector product F_1 + in a holonomic basis F_2*	1	0,51	0,75
(расстояние от точки F_1 + до множества F_2) / *distance F_1 + from point to set F_2*	(расстояние F_1 + множество F_2) / *distance F_1 + set F_2*	0,6	0,88	0,74
(Оператору F_1 + на сетке F_2) / *To the operator F_1 + on the net F_2*	(оператор F_1 + сеть F_2) / *operator F_1 + net F_2*	0,88	0,6	0,74

3) **Models of stable context patterns.** We automatically extracted stable context patterns containing mathematical formulas from the collection of mathematical papers, and then manually selected patterns with fixed semantics among them. Examples of such patterns are mathematical concept definitions, formula descriptions, or patterns with auxiliary words such as "в виде" ("in the form of"). Some of these patterns are presented in Table 5.

This approach was inspired by corpus-driven and empirically well-founded methodologies, such as Corpus Pattern Analysis (CPA) [15] and Chain Clarifying Relationship (CCR) [16].

We have implemented a customized semantic annotation tool for the pattern with auxiliary words "в виде" ("in the form of"), and we are going to implement such customized tools for the other frequent patterns.

Table 5. Models of semantic contexts of formulas

Model number	Context Description	Comment	Context Model	Example
1	<text + formula>	A definition preceding formula	NP + F1	Множество точек пересечения прямых $x_s = x_s^{i_s}$
				The set of points of intersection of lines $\mathrm{x_s} = \mathrm{x_s^{i_s}}$
2	<text + formula> + <text + formula>	A sequence of formulas of different types with definitions	NP(F_1) + PP(F_2)	Производная $(\partial/\partial\xi_2)U^S(\xi) \equiv U_1^S(\xi)$ в представлении $\mathrm{U}(\xi) = U^R(\xi) + U^S(\xi), \xi \in \bar{\bar{D}}^1$
				Derivative $(\partial/\partial\xi_2)U^S(\xi) \equiv U_1^S(\xi)$ in the view $\mathrm{U}(\xi) = \mathrm{U^R}(\xi) + \mathrm{U^S}(\xi), \xi \in \bar{\bar{\mathrm{D}}}^1$
3	<formula> + <separating character (dash)> + <text>	A formula with subsequent definition	F1 + '–' + NP	$D_h = D \cap R_h^2$ – множество внутренних узлов.
				$D_h = D \cap R_h^2$ is set of internal nodes
4	<text + formula> + {<formula + text>}	A formula with its definition (within NP) and a list of variables included in this formula	NP(F1) + "*где*" + {NP_1(F2), NP_2(F2), …, NP_n(F2)}	Метод механических квадратур линейных интегродифференциальных уравнений вида $x(t) + \int_0^1 h_0(t,s)x(s)\,ds + \int_0^1 h_1(t,s)x'(s)\,ds = y(t), 0 \le t \le 1,$ где $y(t)$ и $h_0(t,s), h_1(t,s)$ – известные функции.
			NP(F1) + "w*here*" + {NP_1(F2), NP_2(F2), …, NP_n(F2)}	The method of mechanical quadratures of linear integro-differential equations $x(t) + \int_0^1 h_0(t,s)x(s)\,ds + \int_0^1 h_1(t,s)x'(s)\,ds = y(t), 0 \le t \le 1,$ where $y(t)$ и $h_0(t,s), h_1(t,s)$ are known.

5 Conclusion

This paper describes the results of semantic annotation and semantic search in a mathematical collection, as well as new solutions for improving semantic search related to building semantic models of contexts of a mathematical formula. These solutions are based on the OntoMath^{Pro} ontology and allow enriching the ontology with new concept links.

These studies are in line with the project of developing the World Digital Mathematical Library, the purpose of which is to unite digitized versions of the entire body of mathematical scientific literature in a distributed system of interlinked repositories.

The developed technologies have been tested on digital math collections of Kazan Federal University.

Acknowledgments. A. Kirillovich would like to thank Sergio Pinillos Prado and Mikhail Malakhaltsev from the University of Los Andes, as well as Zaira Alexandra Gil Vargas for their assistance. This work was funded by the subsidy allocated to Kazan Federal University for the state assignment in the sphere of scientific activities, grant agreement no. 1.2368.2017, and Russian Foundation for Basic Research and the government of the region of the Russian Federation, grant № 18-41-000001.

References

[1] Elizarov A.M., Lipachev E.K., Malakhaltsev M.A.: Web Technologies for Mathematicians: The Basics of MathML. Fizmatlit (2010)

[2] Kohlhase M.: An Open Markup Format for Mathematical Documents (Version 1.2). Lecture Notes in Computer Science, vol. 4180. Springer Verlag (2006). doi:10.1007/11826095

[3] Kohlhase M.: Semantic Markup in TeX/LaTeX. CTAN (2016). Url: http://ctan.altspu.ru/macros/latex/contrib/stex/sty/stex/stex.pdf

[4] Pitman J., and Lynch C.: Planning a 21st Century Global Library for Mathematics Research. Notices of the AMS, vol. 61, num. 7, pp. 776–777 (2014). doi:10.1090/noti1143

[5] Daubechies I., et al. Developing a 21st Century Global Library for Mathematics Research. Washington, D.C.: The National Academies Press (2014). doi:10.17226/18619

[6] Olver P.J.: The World Digital Mathematics Library: report of a panel discussion. Proceedings of the International Congress of Mathematicians, August 13–21, 2014, Seoul, Korea, vol. 1, pp. 773–785. Kyung Moon SA (2014)

[7] Ion P., et al. White Paper of the Semantic Representation of Mathematical Knowledge Workshop (2016). Url: http://mathontheweb.org/gdml/srmw/whitepaper.pdf

[8] Elizarov A., et al.: Semantic Formula Search in Digital Mathematical Libraries. Proceedings of the 2nd Russia and Pacific Conference on Computer Technology and Applications (RPC 2017), pp. 39-43. IEEE (2017). doi:10.1109/RPC.2017.8168063

[9] Nevzorova O., et al.: OntoMath^{Pro} Ontology: a Linked Data Hub for Mathematics. Klinov P., Mouromstev D. (eds.) Proceedings of the 5th International Conference on Knowledge Engineering and Semantic Web (KESW 2014). Communications in Computer and Information Science, vol. 468, pp. 105–119. Springer, Cham (2014). doi: 10.1007/978-3-319-11716-4_9

[10] Elizarov A., et al.: Digital Ecosystem OntoMath: Mathematical Knowledge Analytics and Management. Kalinichenko L., et al. (eds.) XVIII International Conference on Data Analytics and Management in Data Intensive Domains (DAMDID/RCDL 2016). Communications in Computer and Information Science, vol. 706, pp. 33-46. Springer (2017). doi:10.1007/978-3-319-57135-5_3

[11] Elizarov A., et al.: Mathematical Knowledge Management: Ontological Models and Digital Technology. Kalinichenko L. et al. (eds.) Selected Papers of the XVIII International Conference on Data Analytics and Management in Data Intensive Domains (DAMDID/RCDL 2016). CEUR Workshop Proceedings, vol. 1752, pp. 44–50. CEUR-WS (2016)

[12] Nevzorova O., et al.: Bringing Math to LOD: A Semantic Publishing Platform Prototype for Scientific Collections in Mathematics. Alani H., et al. (eds.) Proceedings of the 12th International Semantic Web Conference (ISWC 2013). Lecture Notes in Computer Science, vol. 8218, pp. 379-394. Springer (2013). doi:978-3-642-41335-3_24

[13] Elizarov A.M., et al.: Methods and means for semantic structuring of electronic mathematical documents. Doklady Mathematics, vol. 90, Is. 1, pp. 521-524 (2014). doi:10.1134/S1064562414050275

[14] Nevzorova O., Nevzorov V.: Terminological annotation of the document in a retrieval context on the basis of technologies of system "OntoIntegrator". Int. J. Information Technologies & Knowledge, 2011, vol. 5, n. 2, pp. 110-118

[15] El Maarouf I., et al.: Disambiguating Verbs by Collocation: Corpus Lexicography meets Natural Language Processing. Calzolari N., et al. (eds). Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014), pp. 1001-1006. ELRA (2014)

[16] Popescu P. Buildind a Resource of Patterns Using Semantic Types. Calzolari N., et al. (eds). Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012), pp. 2999-3006. ELRA (2012)