ISSN 1064-5624, Doklady Mathematics, 2022, Vol. 106, No. 3, pp. 429–435. © Pleiades Publishing, Ltd., 2022.
ISSN 1064-5624, Doklady Mathematics, 2022. © Pleiades Publishing, Ltd., 2022.
Russian Text © The Author(s), 2022, published in Doklady Rossiiskoi Akademii Nauk. Matematika, Informatika, Protsessy Upravleniya, 2022, Vol. 507, pp. 29–35.

MATHEMATICS

OntoMath^{PRO}: An Ontology of Mathematical Knowledge

A. M. Elizarov[a,*], A. V. Kirillovich[a,], E. K. Lipachev[a,***], and O. A. Nevzorova[a,****]**

Presented by Academician of the RAS E.I. Moiseev

Received June 3, 2022; revised July 24, 2022; accepted September 6, 2022

Abstract—The article describes OntoMath^{PRO}, the first Linked Open Data ontology of professional mathematical knowledge. The ontology is designed to represent mathematical concepts. The concepts of the ontology are organized into two hierarchies: a hierarchy of mathematical objects and a hierarchy of reified relationships. OntoMath^{PRO} respects meta-ontological distinctions provided by a foundational ontology and annotates the concepts as kinds and roles. Relationships between mathematical objects are represented in a reified form, i.e., as instances of relationship concepts linked to their arguments that are instances of role concepts. The ontology defines multilingual lexicons that describe how the concepts are expressed in natural language text. The lexicons are represented as Linguistic Linked Open Data datasets. The OntoMath^{PRO} ontology is under development and will be enriched by new areas of mathematics..

Keywords: ontology, ontology development, mathematical fact extraction, reified relation, mathematical knowledge management

DOI: 10.1134/S1064562422700016

1. INTRODUCTION

Modern scientific research assumes the use of knowledge management technologies, including advanced search for related scientific results, which can be ensured, in particular, by applying techniques for semantic search in available digital collections of scientific documents. In this case, an important role is played by ontologies. It is well known (see, e.g., [1]) that an ontology is a conceptual model of a subject area represented in a formal language, which ensures the application of information-processing machine tools. Note that the concept "ontology" has a variety of meanings depending on the community (in the historically first community, it has a philosophical meaning). The above definition, which reflects a computational meaning, is characteristic primarily of knowledge engineering and has appeared relatively recently.

Since the appearance of the Semantic Web, ontologies have played a prominent role in the modeling of intelligent data management systems, and they take a central place in Tim Berners-Lee's well-known layer cake architecture (https://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html). The currently observed tendency demonstrates the expansion of ontology applications due to the development of ontologies for new domains, reusing concepts of the ontologies from related domains, and the use of various databases and other information resources. At the same time, all-encompassing ontology-based descriptions of domains are not available, because any ontology involves development and depends on the aims of its creation and application. Accordingly, the design of conceptual models of domain sections and, on this basis, the formation of corresponding thesauruses and ontologies make up a research issue of current interest. Important results in this direction related to the domain of mathematics have been obtained in a number of works.

Issues concerning the development of ontologies of scientific information space were addressed in [2]. A thesaurus of the domain "Mixed equations of mathematical physics" was presented in [3]. A way of describing terminology and related equations and formulas in a thesaurus of the domain "Equations of mathematical physics" was proposed in [4]. Approaches and methods for creating a semantic

[a] *Kazan Federal University, Kazan, Russia*
**e-mail: amelizarov@gmail.com*
***e-mail: alik.kirillovich@gmail.com*
****e-mail: elipachev@gmail.com*
*****e-mail: onevzoro@gmail.com*

library in the domain "Mathematics" were presented in [5]. Specifically, the LibMeta intelligence system and the LibMeta Ontology intended for the generation of semantic libraries were described. The importance of projects applying semantic technologies in mathematics according to the Linked Open Data (LOD) principles (https://lod-cloud.net/), including the OntoMath^{PRO} project, was noted in the study concerning the formation of the machine-readable Mathematics Subject Classification LOD collection [6].

The structure of mathematical knowledge and the possibilities of its ontology-based representation on the Web were studied in [7]. A survey of semantic methods for solving fundamental problems in mathematical knowledge management can be found in [8] (where ontologies are described as formalisms for mathematical knowledge representation). A review of a broader class of languages for formalizing mathematics was given in [9].

The development of mathematical ontologies is associated with a number of important tasks. Fundamental are the tasks of classifying and systematizing mathematical knowledge concepts, semantic search, extraction of mathematical statements from texts, and their representation in the LOD cloud. Another important task is the construction of recommender systems as elements of decision support systems, including in mathematics. Application of ontologies in learning systems, in particular, for mathematical knowledge assessment is one more important goal of ontology development.

2. OntoMath^{PRO} ONTOLOGY: BASIC APPROACHES

OntoMath^{PRO} belongs to the class of domain-specific ontologies. It was designed to classify and systematize concepts of professional mathematical knowledge, and it covers several major fields of mathematics. An important feature of OntoMath^{PRO} is the broad coverage of concepts from various fields of mathematics, which makes it possible to reuse these concepts. This possibility is a fundamental property of all domain-specific ontologies, and it was used to a full extent in the development of OntoMath^{PRO}.

A major goal of developing the OntoMath^{PRO} mathematical ontology was its use in scientific research, including the global goals pursued in the global projects "Global Digital Mathematical Library" (GDML) and "World Digital Mathematical Library" (WDML) [10, 11].

With the aim of classifying concepts of mathematical knowledge, OntoMath^{PRO} can be used to construct modern classifiers in various field of mathematics and to create specialized mathematical knowledge bases. Incorporation of the ontology into text processing systems for mathematical documents makes it possible not only to extract new mathematical concepts for enriching the ontology itself, but also to form particular mathematical statements and to store them in knowledge bases. Databases and knowledge bases produced by applying methods for information extraction from mathematical texts can be published in the LOD cloud.

The development of OntoMath^{PRO} relied on our previous results concerning mathematical knowledge management and methods designed for structural and semantic analysis of mathematical documents [12].

The first version of the ontology was presented in [13]. It was intended for modeling mathematical entities and was organized in the form of hierarchies of mathematical objects and fields of mathematics. In the next version of the ontology, the constructs for modeling mathematical entities were supplemented with constructs for modeling mathematical statements. For this purpose, a hierarchy of reified relationships was introduced into the ontology and linguistic and foundation ontology layers were added to its structure (see Section 2). Additionally, OntoMath^{PRO} was used to develop a number of applications for mathematical knowledge management (see Section 3).

3. ARCHITECTURE OF THE OntoMath^{PRO} ONTOLOGY

The structure of the ontology consists of three layers (Fig. 1), which are organized according to different models of concept representation:

- *domain ontology layer* contains concepts of professional mathematical knowledge;
- *foundation ontology layer* provides the concepts from domain ontology layer with meta-ontological annotations;
- *linguistic layer* describes how the concepts from domain ontology layer are expressed in in scientific mathematical texts in the Russian and English languages.

In the *domain ontology layer*, a hierarchy of objects and a representation of reified relationships are specified.

The *hierarchy of objects* contains concepts covering a wide variety of fields of mathematics (mathematical logic, set theory, mathematical analysis, algebra, geometry, differential equations, etc.). Concepts of the first hierarchical level are of three types: basic mathematical objects (such as *Set*, *Operator*, *Mapping*), root elements of a corresponding field of mathematics (e.g., *Element of probability theory*), and general mathematical concepts (such as *Problem*, *Method*, *Statement*, *Formula*).

The description of a concept in the ontology contains its name and definition in Russian and English, meta-ontological annotations, relations to other concepts, and references to external resources. As an example, Fig. 2 presents the concept of the *Degree of a*

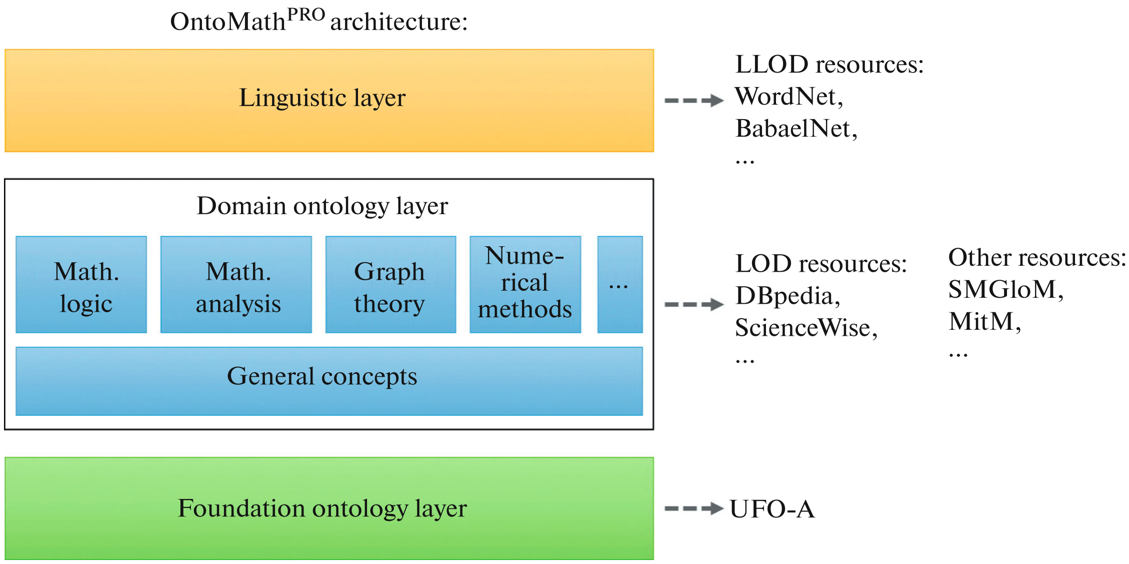

Fig. 1. General structure of the OntoMath^{PRO} ontology.

polynomial in the WebProtégé editor. Its description contains two variants of the concept name in Russian (property `rdfs:label`): "*Степень многочлена*" and "*Степень полинома*", the name of the concept in English (property `rdfs:label`), the definition of the concept (property `rdfs:comment`), the meta-ontological class *Role* (`gufo:Role`), and the parent concept *Number*.

OntoMath^{PRO} specifies the object properties connecting the concepts from the hierarchy of objects. The introduced properties represent the following relations: (1) the relation between an object of mathematical knowledge and a section of mathematics, (2) "*defined in terms of*" relation, (3) associative relation, and (4) the relation between a problem and its solution method.

The *representation of reified relationships* is aimed at modeling *n*-ary relations between mathematical objects expressed in terms of *n*-ary predicates. In the OWL language, such relations are represented in reified form, i.e., in the form of classes. The arguments of *n*-ary relations are represented by role classes, whose instances are related to instances of reified relations by object properties.

In natural language text, a reified relationship is usually expressed by a verb form, for example, *divide*, *belong*. As an example, Fig. 3 presents the divisibility relation. Its arguments are two role concepts, *Dividend* and *Divisor*, which are subclasses of the kind concept *Number*. Instances of the reified relationship are the relations between particular numbers, as indicated in the figure. Instances of the relation arguments are connected with the relation instance by the object property `omp:hasArgument`.

Foundation ontology layer. OntoMath^{PRO} respects meta-ontological distinctions, including distinctions between kind concepts and role concepts.

A kind is a concept that is semantically rigid and ontologically independent [14, 15]. For example, the concept of *Positive integer* is a kind, because any positive integer is a positive integer regardless of of its relations to other numbers.

A role is a concept that is semantically anti-rigid and ontologically dependent. An object can be an instance of a role concept only by virtue of the relation of this object to another object [14, 15]. For example, the concept of *Divisor* is a role of a number, that a number plays only in context of division operation. Role concepts are linked to corresponding kind concepts by relations of ontological dependence.

In the foundation ontology layer, concepts from the hierarchy of objects are annotated by meta-ontological classes (such as Kind, Role, Relator) defined in the Unified Foundation Ontology (UFO) [14].

For meta-ontological annotation of concepts in OntoMath^{PRO}, there are two approaches. According to one of them, annotation is made using the predicate `rdf:type`. This approach requires the OWL Punning mechanism (https://www.w3.org/2007/OWL/wiki/Punning), but allows one to use multilevel modeling principles, including the MLT Ontology (https://nemo.inf.ufes.br/projects/mlt/). In the other approach, the `omp2:hasMetaclass` property is used for annotation.

The *linguistic layer* of OntoMath^{PRO} consists of multilingual lexicons describing how the concepts from the domain ontology layer are expressed in Russian and English texts. A lexicon consists of

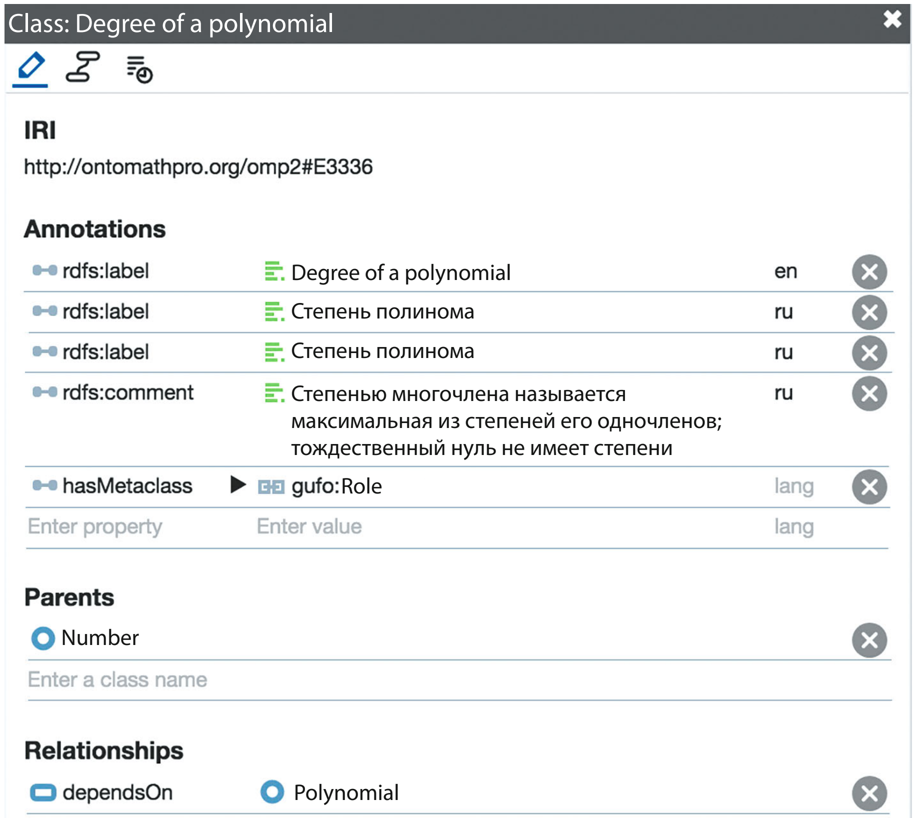

Fig. 2. Example of the "Degree of a polynomial" concept in the WebProtégé editor.

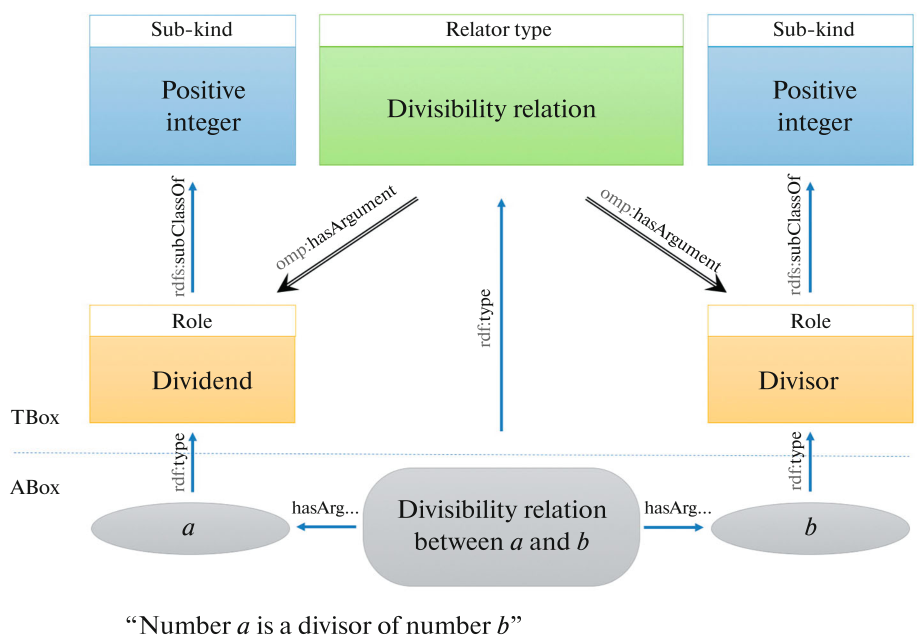

Fig. 3. Representation of reified relations.

• lexical entries (single-word and multiword), which denote mathematical concepts (e.g., the lexical entry “matrix” is used to denote the same-name concept *Matrix* from the domain ontology layer);

• forms of lexical entries (in different cases, singular and plural forms);

• syntactic trees for multiword lexical entries;

• syntactic frames for predicate lexical entries, which describe syntactic arguments of the predicates, grammatical features of the arguments, and their mappings to the concepts from domain ontology layer.

Multilingual lexicons are represented as Linguistic Linked Open Data datasets (Linguistic Linked Open Data, LLOD, http://linguistic-lod.org/) in terms of the ontologies OntoLex/Lemon (https://www.w3.org/community/ontolex/; https://www.w3.org/2016/05/ontolex/), LexInfo (https://lexinfo.net/), and PreMOn (Predicate Model for Ontologies, https://premon.fbk.eu/).

The OntoMath^{PRO} ontology is under active development. Its current version contains more than four thousand concepts in the hierarchy of objects, top layer relations in the hierarchy of reified relations, meta-ontological descriptions of 600 concepts, and linguistic descriptions of base concepts in the Russian and English languages. In the future, we intend that new sections of the ontology will be developed, the existing ones will be supplemented with new concepts, new reified relationships between concepts will be developed, and concepts will be equipped with new linguistic annotations.

4. APPLICATIONS OF OntoMath^{PRO}

An example of OntoMath^{PRO} applications is provided by the OntoMath digital ecosystem we have developed.

OntoMath is a digital ecosysystem of ontologies, text analytics tools, and applications for mathematical knowledge management [16]. A central component of OntoMath is a semantic publishing platform. It takes, as input, a collection of mathematical documents in LaT$_E$X format and automatically constructs their semantic representation in the form of an RDF dataset integrated into the LOD cloud. This dataset includes metadata, components of logical structures of mathematical publications, mathematical terminology, and mathematical formulas and representations specifying relations between terms and their symbolic notation in formulas. The generated RDF dataset underlies mathematical knowledge management services, in particular, semantic formula search [17] (https://lobachevskii-dml.ru/mathsearch), a recommender system for searching and analyzing mathematical papers [18], and a recommender system for choosing Universal Decimal Classification (UDC) codes for mathematical articles [19].

The principles of modeling the OntoMath^{PRO} ontology were used to develop the educational mathematical ontology OntoMath^{Edu} (https://github.com/CLLKazan/OntoMathEdu), which covers secondary school mathematics knowledge and can be used to design educational courses based on modern digital platforms [20]. The OntoMath^{PRO} ontology was also used to test the level of competence of students majoring in mathematics. The testing task was to link numerical analysis problems with appropriate solving methods. The conducted experiment demonstrated the effectiveness of the applied approach.

Other applications of OntoMath^{PRO} are associated with its use in related research areas.

In [21, 22] the OntoMath^{PRO} ontology was translated into Italian and was then used to develop recommender systems for mathematical learning objects.

In [23] OntoMath^{PRO} was used to design Scientific Knowledge Objects Ontology (SKOO) intended for scientific knowledge visualization.

In [24] OntoMath^{PRO} was applied as a component of a platform for searching for documents in digital library catalogs.

In [25] the OntoMath^{PRO} ontology was used as a component of a data quality control system in sensor networks.

Thus, the OntoMath^{PRO} project has found a wide variety of applications.

5. CONCLUSIONS

The first Linked Open Data ontology of professional mathematical knowledge OntoMath^{PRO} has been presented. It (1) respects ontological distinctions defined in the foundation ontology; (2) represents mathematical relations as first-order entities; and (3) contains a linguistic layer that provides a detailed description of how mathematical concepts are expressed in natural language texts.

The main research contribution of the work is that it bridges the gap between three different areas of research: mathematical knowledge management, ontological modeling, and Linguistic Linked Open Data (LLOD).

The OntoMath^{PRO} project is under development. We intend to enrich the OntoMath^{PRO} ontology with new branches of mathematics.

The OntoMath^{PRO} ontology is distributed under the Apache 2.0 license and is available at GitHub repository (https://github.com/CLLKazan/OntoMathPro/) with the possibility of its development by interested mathematics specialists.

ACKNOWLEDGMENTS

The authors are grateful to the reviewers for their remarks, which helped significantly improve the manuscript.

FUNDING

This work was supported by the Russian Science Foundation, project no. 21-11-00105.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

1. N. Guarino, D. Oberle, and S. Staab, in *Handbook on Ontologies*, 2nd ed., Ed. by S. Staab and R. Studer (Springer, New York, 2009), pp. 1–17. https://doi.org/10.1007/978-3-540-24750-0
2. E. I. Moiseev, A. A. Muromskii, and N. P. Tuchkova, *Ontology of Scientific Information Space* (Vychisl. Tsentr im. A.A. Dorodnitsyna Ross. Akad. Nauk, Moscow, 2013) [in Russian].
3. A. A. Muromskii and N. P. Tuchkova, "Representation of mathematical concepts in scientific knowledge ontology," Ontol. Proekt. **9** (1), 50–69 (2019). https://doi.org/10.18287/2223-9537-2019-9-1-50-69
4. N. P. Tuchkova, "On semantic model of the subject area 'Equations of mathematical physics'," Inf. Mat. Tekhnol. Nauke Upr., No. 4, 132–142 (2020). https://doi.org/10.38028/ESI.2020.20.4.012
5. V. A. Serebryakov and O. M. Ataeva, "Ontology based approach to modeling of the subject domain 'Mathematics' in the digital library," Lobachevskii J. Math. **42** (8), 1920–1934 (2021). https://doi.org/10.1134/S199508022108028X
6. S. Arndt, P. Ion, M. Runnwerth, M. Schubotz, and O. Teschke, "10 years later: The mathematics subject classification and linked open data," in *Intelligent Computer Mathematics, CICM 2021*, Ed. by F. Kamareddine and C. Sacerdoti Coen., Lect. Notes Comput. Sci. **12833**, 153–158 (2021). https://doi.org/10.1007/978-3-030-81097-9_12
7. C. Lange, "Ontologies and languages for representing mathematical knowledge on the Semantic Web," Semantic Web **4** (2), 119–158 (2013). https://doi.org/10.3233/SW-2012-0059
8. A. M. Elizarov, A. V. Kirillovich, E. K. Lipachev, O. A. Nevzorova, V. D. Solovyev, and N. G. Zhiltsov, "Mathematical knowledge representation: Semantic models and formalisms," Lobachevskii J. Math. **35** (4), 347–353 (2014). https://doi.org/10.1134/S1995080214040143
9. C. Kaliszyk and F. Rabe, "A survey of languages for formalizing mathematics," Lect. Notes Artif. Intell. **12236**, 138–156 (2020). https://doi.org/10.1007/978-3-030-53518-6_9
10. *Developing a 21st Century Global Library for Mathematics Research* (Academies, Washington, DC, 2014). https://doi.org/10.17226/18619
11. P. D. F. Ion and S. M. Watt, "The global digital mathematics library and the international mathematical knowledge trust," Lect. Notes Artif. Intell. **10383**, 56–69 (2017). https://doi.org/10.1007/978-3-319-62075-6_5
12. A. M. Elizarov, E. K. Lipachev, O. A. Nevzorova, and V. D. Solov'ev, "Methods and means for semantic structuring of electronic mathematical documents," Dokl. Math. **90** (1), 521–524 (2014). https://doi.org/10.7868/S0869565214240049
13. O. Nevzorova, N. Zhiltsov, A. Kirillovich, and E. Lipachev, "OntoMath^{PRO} ontology: A linked data hub for mathematics," in *Knowledge Engineering and Semantic Web: 5th International Conference KESW 2014*, Ed. by P. Klinov and D. Mouromstev, *Communications in Computer and Information Science* (Springer, Cham, 2014), Vol. 468, pp. 105–119. https://doi.org/10.1007/978-3-319-11716-4_9
14. G. Guizzardi et al., "UFO: Unified foundational ontology," Appl. Ontol. **17** (1), 167–210 (2022). https://doi.org/10.3233/AO-210256
15. N. V. Lukashevich and B. V. Dobrov, "Designing linguistic ontologies for information systems in broad subject areas," Ontol. Proekt. **5** (1), 47–69 (2015).
16. A. Elizarov, A. Kirillovich, E. Lipachev, and O. Nevzorova, "Digital Ecosystem OntoMath: Mathematical knowledge analytics and management," Commun. Comput. Inf. Sci. **706**, 33–46 (2017). https://doi.org/10.1007/978-3-319-57135-5_3
17. A. Elizarov, A. Kirillovich, E. Lipachev, and O. Nevzorova, "Semantic formula search in digital mathematical libraries," *Proceedings of the 2nd Russian–Pacific Conference on Computer Technology and Applications RPC 2017* (2017), pp. 39–43. https://doi.org/10.1109/RPC.2017.8168063
18. A. M. Elizarov, A. B. Zhizhchenko, N. G. Zhil'tsov, A. V. Kirillovich, and E. K. Lipachev, "Mathematical knowledge ontologies and recommender systems for collections of documents in physics and mathematics," Dokl. Math. **93** (2), 231–237 (2016). https://doi.org/10.7868/S0869565216100042
19. O. Nevzorova and D. Almukhametov, "Towards a recommender system for the choice of UDC code for mathematical articles," CEUR Workshop Proc. **3036**, 54–62 (2021). http://ceur-ws.org/Vol-3036/paper04.pdf
20. A. Kirillovich, O. Nevzorova, M. Falileeva, E. Lipachev, and L. Shakirova, "OntoMath^{Edu}: Towards an educational mathematical ontology," CEUR Workshop Proc. **2634**, 1–10 (2020). http://ceur-ws.org/Vol-2634/WiP1.pdf
21. A. Barana, L. Di Caro, M. Fioravera, M. Marchisio, and S. Rabellino, "Ontology development for competence assessment in virtual communities of practice," in *Proceedings of the 19th International Conference Artificial Intelligence in Education* (*AIED 2018*), Part II, Ed. by C. Penstein Rosé et al., Lect. Notes Comput. Sci. **10948**, 94–98 (2018). https://doi.org/10.1007/978-3-319-93846-2_18

22. L. Di Caro, S. Rabellino, M. Fioravera, and M. Marchisio, “A model for enriching automatic assessment resources with free-text annotations,” *15th International Conference on Cognition and Exploratory Learning in Digital Age* (*CELDA 2018*), *October 21–23, 2018, Budapest, Hungary* (2018), pp. 186–193.

23. V. Daponte and G. Falquet, “Une ontologie pour la formalisation et la visualisation des connaissances scientifiques,” *29es Journées Francophones d’Ingénierie des Connaissances, IC 2018, AFIA, Nancy, France* (2018), pp. 129–136.
https://hal.archives-ouvertes.fr/hal-01839572

24. P. Intarapaiboon and C. Kesamoon, “Applying domain knowledge and academic information to enhance unknown-item search in OPAC,” Malays. J. Library Inf. Sci. **24** (1), 45–58 (2019).
https://doi.org/10.22452/mjlis.vol24no1.3

25. A. P. Vedurmudi, J. Neumann, M. Gruber, and S. Eichstädt, “Semantic description of quality of data in sensor networks,” Sensors **21**, 6462 (2021).
https://doi.org/10.3390/s21196462

Translated by I. Ruzanova