Metadata Normalization Methods in the Digital Mathematical Library

P. O. Gafurova[1, 2][0000-0002-1544-155X], A. M. Elizarov[1, 2][0000-0003-2546-6897], E. K. Lipachev[1, 2][0000-0001-7789-2332] and D. M. Khammatova[1][0000-0001-5486-2325]

[1] N. I. Lobachevskii Institute of Mathematics and Mechanics, Kazan Federal University
[2] Higher School of Information Technologies and Intelligent Systems, Kazan Federal University
pogafurova@gmail.com, amelizarov@gmail.com, elipachev@gmail.com, dianalynx@rambler.ru

Abstract. Methods for the automatic generation of metadata for documents in digital mathematical collections in the formats of international resource aggregators in mathematics and Computer Science are proposed. Metadata normalization services for electronic collections of scientific documents in accordance with DTD rules and XML schemas Journal Archiving and Interchange Tag Suite (NISO JATS) V1.0, V1.1, V1.2 have been created. Algorithms for creating electronic collections and including them in the digital mathematical library are presented. Tools for generating metadata of collection documents in accordance with the syntactic rules of digital libraries have been developed. An algorithm for the automated preparation of metadata of electronic collections of the Lobachevskii DML library according to the rules of the dblp Computer Science Bibliography (DBLP) bibliographic database on computer sciences is given. An algorithm for converting metadata to the oai_dc format and generating the archive structure for import into DSpace digital storage has been created. Methods for integrating electronic mathematical collections of Kazan University into Russian and foreign digital mathematical libraries have been proposed and implemented.

Keywords: Digital Mathematical Library, Metadata Extraction, Metadata Normalization, Lobachevskii DML.

1 Introduction

With the development of information and communication technologies, for the first time, the opportunity has appeared to make available the scientific knowledge created over the entire printing period. Therefore, it is no coincidence that initiatives such as the World Digital Mathematics Library (WDML) and Global Digital Mathematics Library (GDML) appear. All of them are aimed at developing the basic principles of the integration of scientific knowledge in the field of mathematics [1, 2]. The goal of the project "The European Digital Mathematics Library" (EuDML, https://initiative.eudml.org/) is to integrate the mathematical resources of European

digital libraries [3, 4]. The Russian project MathNet.Ru (http://www.mathnet.ru/) made available archives of domestic journals and collections, proposed navigation and advanced search methods for mathematical content, as well as a system of links to bibliographic databases [5].

From the point of view of WDML program documents [1], the leading value in the integration of mathematical knowledge is given to digital mathematical libraries (see, for example, [6]). Within the framework of these libraries, methods for processing and managing mathematical documents based on semantic relationships not only between documents, but also with the objects contained in them are being developed [7–10].

In connection with a significant increase in the volume of scientific publications, it became necessary to create specialized methods for the automated processing of large amounts of documents [8, 11, 12].

In line with the WDML strategy, we are developing Lobachevskii DML (https://lobachevskii-dml.ru/) [13], a digital mathematical library of Kazan Federal University.

Lobachevskii DML is based on OntoMath digital ecosystem [14, 15], an ecosystem of ontologies, text analytics tools, and applications for mathematical knowledge management, including semantic search for mathematical formulas [16] and a recommender system for mathematical papers [17].

The core component of the OntoMath ecosystem is the semantic publishing platform [18]. 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, in turn, is based on the $\text{OntoMath}^{\text{PRO}}$ [19] and $\text{OntoMath}^{\text{Edu}}$ [20] ontologies, the ontologies of professional and educational mathematical knowledge respectively. These ontologies are fully integrated to the Linked Open Data (LOD) cloud. Concepts on these ontologies are interlinked to the external LOD-resources, including DBpedia [21] and ScienceWISE [22]. Moreover, the labels of the $\text{OntoMath}^{\text{Edu}}$ ontology are being interlinked with the external lexical resources from the Linguistic Linked Open Data cloud [23], including, WordNet [24], BabelNet [25], RuThes Cloud [26] and Russian-Tatar Thesaurus [27].

In this paper, we present a new version of the module of metadata extraction, customized for the Lobachevskii DML library. To solve the problem of integrating the created electronic collections into aggregating digital libraries, such as EuDML, MathNet.Ru, DBLP, methods for converting metadata according to the schemes adopted in these libraries are proposed. Methods for normalizing metadata of Lobachevskii DML digital library collections are described in accordance with the DTD-rules and XML schemas of the Journal Archiving and Interchange Tag Suite (NISO JATS, https://jats.nlm.nih.gov/archiving/) V1.0, V1 .1, V1.2 [28]. To denote the methods of generating and converting document metadata in accordance with the rules and XML schemes of digital libraries and scientometric databases, we use the term “normalization” (see also [8]). The NISO JATS metadata normalization method served as the basis for the formation of the mandatory and fundamental EuDML

metadata sets. An algorithm for the automated preparation of metadata of electronic collections of the Lobachevskii DML library according to the rules of the bibliographic base for computer science "dblp computer science bibliography" (DBLP, https://dblp.uni-trier.de/) is also presented.

2 Representation of Digital Mathematical Libraries Metadata

Currently, many scientometric databases index articles published in leading mathematical journals. These databases impose different requirements on the set of metadata of these documents, as well as on the schemes of their presentation (see, for example, [29]). Note that, as a rule, such new forms of publications as presentations, scientific blogs and video lectures are not indexed. However, these forms are important components of modern digital libraries.

Digital mathematical libraries use various metadata formats when forming the collections included in them. This is due to the fact that many such collections are formed from articles published in academic journals. In these cases, the relevant articles are made in accordance with the rules established in these journals and differ in the requirements for the metadata used. These differences relate primarily to the composition of the metadata and their format. Most of all, these differences are noticeable in archival collections of scientific journals.

2.1 Features of Representation Metadata

The metadata content of articles of even one journal, depending on the year of its publication, differs significantly. We indicate the archive of articles in the journal "Russian Mathematics (Izvestiya VUZ. Matematika)", https://kpfu.ru/science/ nauchnye-izdaniya/ivrm). His articles are one of the collections of the digital library Lobachevskii DML. The named journal has been published since 1957, and only articles published in this journal dating back to 2010 are accompanied by a relatively complete set of metadata. Articles published before 2008 lack keywords and abstracts (see Table 1).

With the transition of Russian journals to the international scientific space, the composition of affiliation changed. Affiliation was replenished with such new information about the authors as information about the author's place of work, business address, and email address.

The history of expanding the set of metadata used, described in the above example, is typical of almost all scientific journals. To replenish the set of metadata, methods are developed for extracting metadata from documents [8, 11, 29]. There is also a need for methods for normalizing metadata, which allow converting already created metadata into scientometric database formats. We also note that participation in projects such as EuDML involves the provision of sets of metadata formed according to schemes developed by aggregators of mathematical resources.

Table 1. Log Metadata Composition of the journal "Russian Mathematics"

Year	Annotation	City	University	Key words	UDC	Bibliography	English version
1957 – 1959	No	No	No	No	No	No	No
1960 – 1965	No	Yes	No	No	No	No	No
1965 – 1969	No	Yes	No	No	Yes	No	No
1970 – 1974	Yes	Yes	No	No	Yes	No	No
1975 – 1994	No	Yes	No	No	Yes	No	Yes
1994 – 1997	No	No	Yes	No	Yes	No	Yes
1998 – 2007	No	No	Yes	No	Yes	Yes	Yes
2008 – 2009	Yes	No	Yes	Yes	Yes	Yes	Yes
2010 – 2019	Yes	Yes	Yes	Yes	Yes	Yes	Yes

2.2 Normalization of metadata according to EuDML schemes

One of the stages of integration of electronic mathematical collections in EuDML is the normalization of the metadata of these collections according to the rules for the formation of a obligatory set of metadata. EuDML uses NISO JATS V1.0 XML schemas to describe articles from mathematical journals, and the general metadata schema of this digital library is described in [30]. Three sets of metadata were distinguished: obligatory metadata, fundamental metadata, supplemental metadata. The minimum of them in terms of composition is a mandatory set of metadata, which includes: title of the article in the original language, list of authors, bibliography, unique identifier of the article, for example, doi and URL of the full text of the article. The fundamental set of metadata in addition to the mandatory metadata includes annotation of the article and keywords [31].

The digital library Lobachevskii DML is created on the basis of the principles of WDML, according to which the leading role is given to the relationships between documents and objects in them. In this case, the documents themselves can be physically placed outside a specific digital library. A number of electronic collections of the Lobachevskii DML library are physically hosted in other digital libraries. For example, the journal collection “Russian Mathematics” is digitized, equipped with meta-descriptions and presented on the MathNet.Ru portal (http://www.mathnet.ru/php/journal.phtml?Jrnid=ivm). Our tasks are to replenish

such collections with additional metadata, as well as to automatically select objects and establish semantic links between them.

When forming a fundamental set of metadata for electronic collections stored on external resources, the metadata presented on these resources is initially imported. For this purpose and using the package functions HtmlAgilityPack (https://html-agility-pack.net/) in C#, a program has been developed for extracting metadata from web pages and writing them in XML-format of the digital library Lobachevskii DML, replenishment and subsequent conversions according to EuDML schemes. For example, for the collection "Russian Mathematics" the following steps have been completed (see Algorithm 1).

It is proposed to create an article identifier as a combination of lines: a journal identifier (attribute value "`jrnid =`") and an article identifier (attribute value "`paperid =`") on the MathNet.Ru portal.

Algorithm 1: Extraction and normalization metadata of the journal collection "Russian Mathematics"

1: **load** article's page in Russian from journal web cite
2: **split** article's page, read AMSBib citation block
3: **read** from AMSBib citation block: author's name, article's title, begin and end pages, journal' name, volume, urls.
4: **read** from article's page: key words, annotation, affiliation, received date, UDC.
5: **load** article's page in English
6: **read** from AMSBibcitation block: author's name, article's title, begin and end pages, journal' name, volume, urls.
7: **read** from article's page: key words, annotation, affiliation, received date, UDC.
8: **form** article's identifier
9: **form** all metadata in EuDML xml format
10: **write** to file

One of the features of articles in Russian journals is that they can be translated, that is, the author creates an article in Russian, then it is published in the English version of the journal. Such articles cannot be considered as different articles, however, at present, the schemes proposed by EuDML do not allow combining an article published in Russian and its translated version in English within a single meta description. Therefore, in the fundamental set of EuDML, one has to describe the translated articles as different articles in different journals.

Note that when using Russian-language literature, the link is translated. However, as presented in Table 2, the translated and transliterated bibliographic differ in the names of articles, the name of the journal, the issue number and page of the article.

It must also be borne in mind that the same journal may have more than one name. For example, the journal "Izvestiya Vysshikh Uchebnykh Zavedenii. Mathematics" has the original (it is given), transliterated ("Izvestiya Vysshikh Uchebnykh Zavedenii. Matematika"), translated old ("Soviet Mathematics") and new ("Russian Mathematics") names. In the collections of the digital library Lobachevskii DML, as

well as in the eLibrary.ru and MathNet.ru libraries, such articles are presented as duplicates of one document.

Table 2. The difference between the descriptions of the same article in the original language, transliteration and translation into English

Original paper citation	А. М. Елизаров, А. Б. Жижченко, Н. Г. Жильцов, А. В. Кириллович, Е. К. Липачёв, «Онтологии математического знания и рекомендательная система для коллекций физико-математических документов», Докл. РАН, 467:4 (2016), 392–395
Transliterated paper citation	A. M. Elizarov, A. B. Zhizhchenko, N. G. Zhiltsov, A. V. Kirillovich, E. K. Lipachev, "Ontologii matematicheskogo znaniya i rekomendatelnaya sistema dlya kollektsiy fiziko-matematicheskikh dokumentov", Dokl. RAN, 467:4 (2016), 392–395
Translated paper citation	A. M. Elizarov, A. B. Zhizhchenko, N. G. Zhiltsov, A. V. Kirillovich, E. K. Lipachev, "Mathematical knowledge ontologies and recommender systems for collections of documents in physics and mathematics", Dokl. Math., 93:2 (2016), 231–233

Note that the process of preparing metadata in the eLibrary.ru format is automated (see [32]) and is successfully used by us in the Russian Digital Libraries Journal (https: elbib.kpfu.ru).

3 Normalization of metadata according to DBLP schemes

One of the authoritative libraries in computer science is "dblp computer science bibliography" [28]. A prerequisite for including electronic collections in this library is the reorganization and normalization of the metadata of digital library documents.

Among the collections of Lobachevskii DML, DBLP requirements are satisfied by the content of the "Russian Digital Libraries Journal". Since 2015, a new model of document submission has been used in this journal and the Open Journal Systems (OJS) publishing system has been introduced [30]. Metadata sets are now automatically generated using software tools developed by the editors of this journal (http://ojs.kpfu.ru/index.php/elbib). An archive of articles published since 2015 was selected to prepare for indexing in DBLP. The necessary metadata are: publication identifier, surnames and names of authors, title of work, year of publication, volume, number, start and end pages of the article in the journal number and URL of the full text of the article.

One of the problems in the preparation of metadata that can be encountered when describing Russian-language scientific collections is the following question: in what language should the metadata of a Russian-language article be presented in DBLP if the journal has the title and abstract of the article in English. On the one hand, it is

desirable to present the document in a form that will be understood by most users of this database, that is, in English. On the other hand, we note that in the early versions of OJS only one language was used to represent the authors, and the main language we use is Russian. Therefore, when choosing English for the presentation of article metadata, it is necessary to develop tools for the translation and transliteration of article metadata.

Normalization to DBLP format takes place in three main stages: extraction of the required metadata, addition of metadata and their normalization to this format. The corresponding algorithm is presented at Algorithm 2**.** It is implemented using a program developed in C#. Using the System: XML extension tools, parsing xml files is performed, and the html page is read from the NuGet functions of the HTMLAgilityPack package. As a result, an xml-file with metadata loaded into the program is generated. This file is fully compliant with DBLP rules (https://dblp.uni-trier.de/db/journals/rdlj/).

Algorithm 2. Normalization of articles in DBLP format of the journal "Russian Digital Libraries Journal"

1: **load** VolCollection //Set of xml files
2: **for each** volume **in** VolCollection **do**
3: **for each** paper **in** volume **do**
4: **read** from paper: author's names, title, page numbers, year of ussue, url in ojs.kpfu.ru, volume.
5: **read** cite page https://elbib.ru/en/year/+year \\volume page
6: **split** cite page, read metadata: author's names in English, url in elbib.kpfu.ru.
7: **split** author's names,
8: `Answer:=Form(author's name, Transliteration(name), title, page numbers, url, volume);`
9: **write** `Answer` in file dblp.xml
10: **end for**
11: **end for**

4 DSpace-based Digital Storage of Electronic Collections

One of the important tasks when working with digital mathematical libraries is the automated integration of repositories of mathematical documents into other information systems. This process is based on a model of aggregation and dissemination of metadata. Such a model (OAI Protocol for Metadata Harvesting (hereinafter OAI-PMH) [34]) is supported by most systems designed to store information resources. This system is supported, for example, by digital libraries such as EuDML and NumDam. Some digital libraries use specialized methods for harvesting metadata from other repositories. In this case, it is necessary that the data providers have tools and services that allow the dissemination of metadata.

The Open Archives Initiative (OAI) develops and promotes interoperability standards to effectively disseminate electronic resources, as well as to increase the availability of scientific information exchange. The corresponding OAI-PMH protocol requires the inclusion of a Dublin Core metadata set (Dublin Core, http://dublincore.org/) in the resource description. For this, the oai_dc format was developed, which is based on Dublin Core and uses only a limited number of Dublin Core tags [35]. The application of the OAI-PMH protocol requires the exchange of information within the framework of well-established data schemes. As a rule, such schemes are not implemented in specific information systems; therefore, dynamic conversion of metadata or automatic preparation of metadata in a format suitable for OAI-PMH is required.

In addition, to organize work with OAI-PMH it is necessary to use a digital storage support system. An overview of various digital repositories is given in [36]. The most popular of these are DSpace, Eprints, Fedora, and Greenstone. We use the DSpace system. This is an open source application (BSD license) that is cross-platform and based on Java. To store metadata, Oracle or PostgreSQL DBMSs are used. For the basic data organization, a data model based on the Dublin Core scheme has been fixed. It is also possible to upload your metadata formats and converters. This makes this system the most attractive for use. Thus, we have implemented the ability to automatically convert various data formats to Dublin Core, which allows us to harvest metadata. However, to download custom formats, it is necessary to develop specialized metadata conversion systems.

Uploading metadata to DSpace is as follows. A table file is generated in csv format (Comma-Separated Values). It records metadata prepared according to the Dublin Core scheme. A method for converting the archive to the Simple Archive Format is also used. In addition, it is possible to download metadata through the console or use manual input of metadata on the site. It is most rational to use archive downloads. The main advantages of this option are the ease of downloading all files in one archive and the ability to download not only metadata, but also files. Since DSpace digital storage can be used together with OJS, you can get the “storage + system" model for working with any journal or collection of articles. Data exchange occurs through the OAI-PHM server, which allows you to automatically harvest metadata.

One of the important tasks is to create a service that would allow the conversion of metadata into the oai_dc format. Here, the difference from the classic Dublin Core format is that in oai_dc, Dublin Core tags are not elements of a metadata scheme, but are placed in the attributes of the <dcvalue> tag. When organizing an archive in Simple Archive Format, we need to create a dublin_core.xml file in oai_dc format for each article. This file contains all the necessary information about the article from the publication being processed. So you can prepare for downloading the data of the processed set of articles in one archive. Note that optimal file loading in DSpace requires the addition of data files whose names are written in the content file.

As a test collection for testing the method described above, we used files from a number of collections of the “Proceedings of the N.I. Lobachevskii Mathematical Center” published by Kazan Federal University. These files are a collection of articles from these collections and contain a description of each of them with the listed

metadata about the relevant articles. Each article's description contains metadata such as authors, title, start and end pages (Fig. 1). Therefore, it is necessary to add metadata such as volume, year of publication, publisher to the description of each article, as well as convert the description of each volume to a multitude of descriptions of individual articles. Information about the year and the publisher is compiled and presented as a csv-file (Fig. 2).

```
<book>

<paper id="1">
 <author> R.Z. Dautov, M.A. Ignatieva </author>
<title-paper> MULTIGRID SOLUTION OF A
 TWO-PHASE STEFAN PROBLEM WITH PRESCRIBED CONVECTION </title-paper>
 <start-page> 3 </start-page>
 <end-page> 11 </end-page>
</paper>

<paper id="2">
```

Fig. 1. Description of the article in internal format

Thus, the input data of the program are: xml-files containing information about articles, as well as a csv-file with information about the volume, year and publishing house. File names contain the volume number. As a result, we get the description files for each article in the oai_dc format, sorted in the order accepted for uploading to DSpace.

5	2000	УНИПРЕСС						
6	2000	УНИПРЕСС						
7	2000	УНИПРЕСС						
8	2001	Издательство ДАС"						
9	2001	Издательство ДАС"						
10	2001	УНИПРЕСС						
11	2001	Издательство Казанского математического общества						
12	2001	Издательство Казанского математического общества						
13	2002	Издательство Казанского математического общества						
14	2002	Издательство Казанского математического общества						
15	2002	Издательство Казанского математического общества						
16	2002	Издательство Казанского математического общества						

Fig. 2. Fragment of metadata about volumes (in the form of a csv-table)

The algorithm is implemented using the C# language using the System: XML extension (Algorithm 3).

Algorithm 3. Normalizing of metadata of the collection of "Proceedings of the Mathematical Center" in oai_dc format

1: **load** VolCollection\\ collection of xml files
2: **for each** volume **from** VolCollection **do**

```
3:    read volume number from file's name
4:    read from info.csv file: publisher, issue year
5:    Papers:=new string list
6.    for each paper from volume do
7.      read from paper: author's names, title, page numbers
8.      split author's names
9.      Paper:=Formoai_dc(authors's names, title, issue year, page numbers, pub-
    lisher);
10.     Papers.Add(Paper)
11.   end for
12.   create volume folder
13.   for each paper from Papers do
14.     create paper's folder
15.     create file dublin_core.xml in paper's folder
16.     write paper in file dublin_core.xml
17.     create content file
18.     copy data files
19.   end for
20. end for
```

A file describing an article in oai_dc format is represented at Fig. 3.

```
<?xml version="1.0" encoding="UTF-8"?>
 <dublin_core>
  <dcvalue element = "contributor" qualifier = "author"> R.Z. Dautov</dcvalue>
 <dcvalue element = "contributor" qualifier = "author"> M.A. Ignatieva </dcvalue>
<dcvalue element = "title" qualifier = "none"> MULTIGRID SOLUTION OF A
  TWO-PHASE STEFAN PROBLEM WITH PRESCRIBED CONVECTION </dcvalue>
 <dcvalue element = "description" qualifier = "none"> P.  3 - 11 </dcvalue>
 <dcvalue element = "description" qualifier = "none"> Vol. 9</dcvalue>
 <dcvalue element = "publisher" qualifier = "none"> DAS </dcvalue>
  <dcvalue element = "date" qualifier = "issued">2001</dcvalue>
</dublin_core>
```

Fig. 3. Generated xml-file in oai_dc format

5 Conclusion

In order to integrate electronic mathematical collections of Kazan University into the international scientific space, algorithms have been developed for the formation of metadata of these collections, as well as the documents included in them, in accordance with the formats of digital mathematical libraries and scientometric databases. Methods of normalizing metadata of electronic mathematical collections in accordance with the XML-schemes NISO JATS and DBLP are presented.

Acknowledgements. The work partially contains the results of the project "Monitoring and standardization of the development and use of technologies for storing and analyzing big data in the digital economy of the Russian Federation", carried out as

part of the program of competence of the National Technological Initiative “Center for storing and analyzing big data”, supported by the Ministry of Science and Higher Education of the Russian Federation under the Treaty of Moscow State University named after M.V. Lomonosov with the Project Support Fund of the National Technological Initiative dated 08/15/2019 No. 7/1251/2019. The work was also carried out with the partial support of the Russian Fund for Basic Researches (project 18-29-03086) and the Russian Fund for Basic Researches and the Government of the Republic of Tatarstan within the framework of scientific project 18-47-160012.

References

1. Developing a 21st Century Global Library for Mathematics Research. The National Academies Press, Washington (2014).
2. Ion, P.D.F., Watt, S.M.: The Global Digital Mathematics Library and the International Mathematical Knowledge Trust. In: ICM 2017: Intelligent Computer Mathematics, 2017. Lecture Notes in Artificial Intelligence, vol. 10383, pp. 56–69. Springer (2017), https://doi.org/10.1007/978-3-319-62075-6_5.
3. Bouche, T.: Reviving the free public scientific library in the digital age? The EuDML project. In: Kaiser, K., Krantz, S.G., Wegner, B. (eds.) Topics and Issues in Electronic Publishing JMM/AMS Special Session, pp. 57–80. FIZ Karlsruhe (2013), https://www.emis.de/proceedings/TIEP2013/ 05bouche.pdf, last accessed 2019/11/21.
4. Bouche, T., Rákosník, J.: Report on the EuDML External Cooperation Model. In: Kaiser K., Krantz, S.G., Wegner, B. (eds.) Topics and Issues in Electronic Publishing, JMM, Special Session, pp. 99–108. San Diego (2013), https://www.emis.de/proceedings/TIEP2013/07bouche_rakosnik.pdf, last accessed 2019/11/21.
5. Chebukov, D.E., Izaak, A.D., Misyurina, O.G., Pupyrev, Yu.A., Zhizhchenko, A.B.: MathNet.Ru as a Digital Archive of the Russian Mathematical Knowledge from the XIX Century to Today. Intelligent Computer Mathematics. In: LNCS, vol. 7961, pp. 344–348 (2013), https://doi.org/10.1007/978-3-642-39320-4_26.
6. Elizarov, A.M., Lipachev, E.K., Zuev, D.S.: Digital Mathematical Libraries: Overview of Implementations and Content Management Services. In: CEUR Workshop Proceedings, vol. 2022, pp. 317–325 (2017).
7. Bartošek, M., Rákosník, J.: DML-CZ: The Experience of a Medium-Sized Digital Mathematics Library. Notices of the AMS 60(8), 1028–1033 (2013), http://dx.doi.org/10.1090/noti1031.
8. Bouche, T., Labbe, O.: The New Numdam Platform. In: CICM 2017: Intelligent Computer Mathematics, pp. 70–82 (2017). https://doi.org/10.1007/978-3-319-62075-6_6.
9. Sadegh, A., Lange, C., Vidal, M.-E., Auer, S.: Integration of Scholarly Communication Metadata using Knowledge Graphs. In: International Conference on Theory and Practice of Digital Libraries, pp. 328–341 (2017).
10. Lange, C.: 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.
11. Elizarov, A.M., Lipachev, E.K., Khaidarov, Sh.M.: Automated Processing Service System of Large Collections of Scientific Documents. In: CEUR Workshop Proceedings, vol. 1752, pp. 58–64 (2016).

12. Elizarov, A.M., Khaydarov, Sh.M., Lipachev, E.K.: Scientific Documents Ontologies for Semantic Representation of Digital Libraries. In: Proc. of the 2nd Russia and Pacific Conf. on Computer Technology and Applications, pp. 1–5 (2017), https://doi.org/10.1109/RPC.2017.8168064.
13. Elizarov, A.M., Lipachev, E.K.: Lobachevskii DML: Towards a Semantic Digital Mathematical Library of Kazan University. In: CEUR Workshop Proceedings, vol. 2022, pp. 326–333 (2017).
14. Elizarov, A.M., Kirillovich, A.V., Lipachev, E.K., Nevzorova, O.A.: Digital Ecosystem OntoMath: Mathematical Knowledge Analytics and Management. In: Communications in Computer and Information Science, vol. 706, pp. 33–46. Springer (2017), https://doi.org/10.1007/978-3-319-57135-5_3.
15. Elizarov, A.M., Kirilovich, A.V., Lipachev, E.K., Nevzorova, O.A.: Mathematical knowledge management: ontological models and digital technology. In: CEUR Workshop Proceedings, vol. 1752, pp. 44–50 (2016).
16. Elizarov, A., Kirillovich, A., Lipachev, E., Nevzorova, O.: Semantic Formula Search in Digital Mathematical Libraries. In: In: Proc. of the 2nd Russia and Pacific Conf. on Computer Technology and Applications, 39–43 (2017). https://doi.org/10.1109/RPC.2017.8168063.
17. Elizarov, A.M., Kirillovich, A.V., Lipachev, E.K., Zhizhchenko, A.B., Zhil'tsov, N.G.: Mathematical knowledge ontologies and recommender systems for collections of documents in physics and mathematics. Doklady Math. 93(2), 231–233 (2016). doi: 10.1134/S1064562416020174.
18. Nevzorova, O., Zhiltsov, N., Zaikin, D., Zhibrik, O., Kirillovich, A., Nevzorov, V., and Birialtsev, E.: Bringing Math to LOD: A Semantic Publishing Platform Prototype for Scientific Collections in Mathematics. In: Lecture Notes in Computer Science, vol. 8218, pp. 379-394. Springer (2013). https://doi.org/10.1007/978-3-642-41335-3_24.
19. Nevzorova, O., Zhiltsov, N., Kirillovich, A., and Lipachev, E.: OntoMath^{PRO} Ontology: A Linked Data Hub for Mathematics. In: Communications in Computer and Information Science, vol. 468, pp. 105–119. Springer (2014). http://doi.org/10.1007/978-3-319-11716-4_9.
20. Kirillovich, A., Nevzorova, O., Falileeva, M., Lipachev, E., Shakirova, L.: OntoMath^{Edu}: Towards an Educational Mathematical Ontology. In: CEUR Workshop Proceedings (forthcoming).
21. Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P. N., Hellmann, S., Morsey, M., van Kleef, P., Auer, S., and Bizer, C.: DBpedia: A Large-scale, Multilingual Knowledge Base Extracted from Wikipedia. Semantic Web Journal, 6(2), 167–195 (2015). https://doi.org/10.3233/SW-140134.
22. Astafiev, A., Prokofyev, R., Guéret, C., Boyarsky, A., and Ruchayskiy, O.: ScienceWISE: A Web-based Interactive Semantic Platform for Paper Annotation and Ontology Editing. In: Lecture Notes in Computer Science, vol. 7540, pp. 392–396. Springer (2012). https://doi.org/10.1007/978-3-662-46641-4_33.
23. McCrae, J. P., Chiarcos, C., Bond, F., Cimiano, P., Declerck, T., de Melo, G., Gracia, J., Hellmann, S., Klimek, B., Moran, S., Osenova, P., Pareja-Lora, A., and Pool, J.: The Open Linguistics Working Group: Developing the Linguistic Linked Open Data Cloud. In: Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016), pp. 2435-2441. ELRA (2016).
24. McCrae, J. P., Fellbaum, C., and Cimiano, P.: Publishing and Linking WordNet using lemon and RDF. In: Proceedings of the 3rd Workshop on Linked Data in Linguistics (LDL-2014), pp. 13–16. ELRA (2014).

25. Ehrmann, M., Cecconi, F., Vannella, D., McCrae, J., Cimiano, P., and Navigli, R.: Representing Multilingual Data as Linked Data: the Case of BabelNet 2.0. In: Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014), pp. 401–408. ELRA (2014).
26. Kirillovich, A., Nevzorova, O., Gimadiev. E., and Loukachevitch, N.: RuThes Cloud: Towards a Multilevel Linguistic Linked Open Data Resource for Russian. In: Communications in Computer and Information Science, vol. 786, pp. 38-52. Springer (2017). http://doi.org/10.1007/978-3-319-69548-8_4.
27. Galieva, A., Kirillovich, A., Khakimov, B., Loukachevitch, N., Nevzorova, O., and Suleymanov, D.: Toward Domain-Specific Russian-Tatar Thesaurus Construction. In: Proceedings of the International Conference IMS-2017, pp. 120–124. ACM (2017). http://doi.org/10.1145/3143699.3143716.
28. "ANSI/NISO Z39.96-2019, JATS: Journal Article Tag Suite". National Information Standards Organization. 8 February 2019. 652 p. https://groups.niso.org/apps/group_public/download.php/21030/ANSI-NISO-Z39.96-2019.pdf, last accessed 2019/11/21.
29. Elizarov, A.M., Zaitseva, N.V., Zuev, D.S., Lipachev, E.K., Khaidarov, S.M.: Services for Formation of Digital Documents Metadata in the Formats of International Science-based Databases. In: CEUR Workshop Proceedings, vol. 2260, pp. 175–185 (2018).
30. Jost, M., Bouche, T., Goutorbe, C., Jorda, J.P.: D3.2: The EuDML metadata schema, http://www.mathdoc.fr/publis/d3.2-v1.6.pdf, last accessed 2019/11/21.
31. EuDML metadata schema specification (v2.0–final), https://initiative.eudml.org/eudml-metadata-schema-specification-v20-final, last accessed 2019/11/21.
32. Gerasimov, A.N., Elizarov, A.M., Lipachev, E.K.: Subsystem of Formation Metadata for Science Index Databases on Management Platform Electronic Scientific Journals. Russian Digital Libraries Journal 18(1–2), 6–31 (2015).
33. Akhmetov, D., Elizarov, A., Lipachev, E.: Service-oriented information system of "Russian Digital Libraries Journal". Russian Digital Libraries Journal 19(1), 2–39 (2016).
34. Open Archives Initiative Protocol for Metadata Harvesting, http://www.openarchives.org/OAI/openarchivesprotocol.html, last accessed 2019/11/21.
35. Expressing Dublin Core metadata using XML https://www.dublincore.org/specifications/dublin-core/dc-xml/, last accessed 2019/11/21.
36. Fedotov, A.M., Baidavletov, A.T., Zhizhimov, O.L., Sambetbayeva, M.A., Fedotova, O.A.: Digital Repository of Scientific and Educational Information System. Vestn. Novosib. gos. un-ta. Serija: Informacionnye tehnologii 13(3), 68–86 (2015).