Big Math Methods in Lobachevskii-DML Digital Library

Alexander Elizarov[0000−0003−2546−6897] and Evgeny Lipachev[0000−0001−7789−2332]

Higher School of Information Technologies and Intelligent Systems; N.I. Lobachevskii Institute of Mathematics and Mechanics of Kazan (Volga region) Federal University,
35 Kremlevskaya ul., Kazan, Russia, 420008
amelizarov@gmail.com, elipachev@gmail.com

Abstract. We offer methods for the formation of digital collections from a set of documents (scientific articles, monographs, collections of reports), which are presented in various storage formats. Based on the analysis of the structure of documents and the stylistic features of their design, we have developed an algorithm for extracting the metadata of these documents. We present a software tool for dividing collections of articles into separate documents and the formation of their semantic presentation. On the example of a collection "Proceedings of N.I. Lobachevskii Mathematical Center", which have a different format and structure, we describe the algorithm for creating a digital collection and its inclusion in the Lobachevskii-DML.

Algorithms for replenishing the electronic collections of the Lobachevskii-DML digital library and generating metadata of documents of these collections in selected formats are presented. Services for the normalization of the Lobachevskii-DML digital library collection metadata were developed in accordance with the DTD rules and the NISO JATS and DBLP XML schemas. Algorithms for creating a mandatory and fundamental collection of metadata collections are presented in accordance with the rules of the European digital mathematical library EuDML.

Keywords: Electronic mathematical collections · Digital libraries · Formation and extraction of metadata · Semantic links of information objects · Metadata normalization services · Mandatory and fundamental metadata sets · Lobachevskii-DML digital library.

1 Introduction

The term "Big Data", which is currently widely used in various subject areas [1], in relation to mathematics requires certain clarifications: big data in mathematics is fundamentally different from big data in the current traditional understanding. In mathematics, all data is essential; moreover, in mathematical documents, many of their parts, especially formulas, are a peculiar code that requires decoding and special interpretation.

Information and communication technologies (ICT) have completely changed the life cycle of scientific documents both at the stages of their preparation and dissemination, and at the stage of obtaining results. The above fully applies to scientific activities in the field of mathematics. But when solving mathematical problems, the expectations from the use of ICT are significantly higher. Here we can draw an analogy with the way computers have completely eliminated manual calculations. The complexity of manual calculations, moreover, their routine, can be demonstrated by the example of numerous tables of function values. Such, for example, are the four-digit tables of V.M. Bradis [2], familiar to all mathematicians: they were created in 1921 and reprinted more than 60 times.

Computations have always required the use of specific methods and non-standard organizational solutions to cope with the volume (Volume is one of the characteristics of big data) and to overcome the barrier of the computing power of the individual. Riche de Prony (Gaspard Clair François Marie Riche de Prony) in 1791–1802 to compile "cadastral tables" containing logarithms (8 characters), created a "factory for calculation" ("usine à calculer"), dividing the calculators into sections: a section of theorists from five prominent mathematicians, including Legendre, a section of "calculators", the number which was 7–8 people familiar with numerical and analytical calculations, and a section of 60-80 people who were engaged only in addition and subtraction [3]–[6]. Speaking of Velocity as one of the characteristics of big data, the duration of manual calculations illustrates an example of calculating π: V. Shanks (William Shanks, 1873) spent 15 years calculating 707 characters of this number (but only 555 of them turned out to be true).

"Manual" calculations are a typical calculation practice used almost until the middle of the 20th century. After that, the created computers saved the scientists from tedious arithmetic operations. Today, on the simplest laptop, the calculation of Pi using the same algorithm that was used for manual counting will take less than a second. Humankind is expecting the same progress now not only in calculations. In the same way, intelligent computer tools should leave in the past time-consuming routine (and not only!) operations in Mathematics. In addition to computing and document preparation, intelligent search tools are needed, including recommendation systems for finding scientific articles that are close in content; terminological annotation services; personal information assistants and information platforms for publishing automation.

This article describes approaches to managing large collections of digital mathematical documents based on semantic methods and consistent with the principles of the World Digital Mathematical Library (WDML), as well as related to the areas constituting the Big Math concept. These approaches are being developed and already partially implemented in the framework of the project for creating the Lobachevskii–DML digital math library.

2 Big Data in Mathematics and Big Math

Mathematicians, as well as scientists of other specialties, in recent decades have faced such volumes of scientific documents that require the involvement of new methods of working with information. These methods should be primarily based on the use of intelligent software tools. Estimates of the growth in the volume of scientific production made today are fairly approximate and take into account only articles in scientific journals. As an example, we present the results of a calculation carried out by the Center for Science and Technology Research at the University of Leiden (SBF 2007). According to this Center (see, for example, [7]), the number of scientific publications in professional journals worldwide increased from about 686 thousand in 1990 to about 1,260 thousand in 2006, which corresponds to an increase of 84%. The annual growth rate calculated on this basis was more than 5%. At the same time, the number of scientific publications is growing faster than the world economy. In addition to journal articles, scientific knowledge is being disseminated today through such new forms of publications as academic blogs, social networks, and dynamic publications. These forms have already become widespread on the Web (see, for example, [8]).

Specialized software services are being developed for working with scientific content. Currently, computer support is used at all stages of the life cycle of a scientific document. Mathematical content has features that do not always allow using general-purpose software tools to work with it. The specificity of mathematical documents is determined, first of all, by the logical structure of texts presented in the form of a strict sequence of objects – definitions, statements and proofs. They clearly indicate or are implicitly hidden links with objects from other documents that are understandable only to a specialist in mathematics.

The presence of specialized formulas is another feature of mathematical documents, which requires the use of specialized software tools at all stages of the document life cycle, in particular, for their input and display. Such tools, as a rule, are developed by mathematicians themselves (for example, [9]–[13]).

Documents that contain similar texts may differ significantly in terms of the content laid down in the formulas contained in these documents. Moreover, absolutely identical formulations of theorems can have qualitative differences on the results declared in them. Examples are theorems on the improvement of approximation estimates or reference books on special sections of mathematics (see, for example, [14]). Therefore, without methods that use the semantics of not only texts, but also formulas, effective work with mathematical documents is impossible [15]–[18].

Big data in mathematics also manifests itself in studies that require consideration and analysis of numerous cases. For example, the classification of finite simple groups required the long-term efforts of a large group of mathematicians and is presented on more than 10,000 journal pages. An overview of this grand study is given in [19, 20]. The well-known problem of four colors was reduced to 1936 configurations and to create an algorithm for checking them on a computer [21, 22]. The validity of the computer proof was confirmed by G. Gonthier by the formalization in the Coq language in 2005 [23].

J. Carette, W.M. Farmer, M. Kohlhase and F. Rabe [24] proposed to use, by analogy with the term Big Data, the term Big Math to denote the field of creating methods and developing software systems to support mathematical research. They highlighted 5 main aspects of Big Math:

- Inference (output of statements by deduction);
- Computation (algorithmic transformation of representations of mathematical objects into forms that are easier to understand);
- Tabulation (creating static, specific data related to mathematical objects and structures that can be easily stored, queried and shared);
- Narration (bringing the results into a form that people can assimilate);
- Organization (modular organization of mathematical knowledge).

The main task of mathematical software systems today is to integrate the aspects that make up Big Math.

3 Integrating Mathematical Knowledge with Digital Mathematical Libraries

The system of digital mathematical libraries currently being created is intended to consolidate and make accessible both modern mathematical knowledge and the knowledge contained in articles and books published in the pre-digital period. To achieve this goal, in the framework of digital libraries, methods for managing digital information are developed that take into account the characteristics of the presentation of mathematical content (see, for example, [25, 26]).

The most important tasks in the management of mathematical knowledge are highlighted in [17, 26, 27]. The defining part of these problems can be solved with the help of digital mathematical libraries built using semantic technologies [26].

An overview of digital mathematical libraries from the point of view of the DELOS Digital Library Reference Model is given in [25]. These libraries are mainly national and carry out the task of consolidating the mathematical documents of their countries, primarily books and journal articles. Examples of such libraries are The Numdam French digital mathematics library [28] and the All-Russian Mathematical Portal Math-Net.Ru [29].

In the field of integration of mathematical knowledge, the most significant is the Global Digital Mathematics Library (GDML) initiative [30, 31]. The World Digital Mathematics Library (WDML) project put forward the idea of combining the entire corpus of digital mathematical documents in the distributed system of electronic collections as the main task [26]. The European Digital Mathematics Library (EuDML, https://initiative.eudml.org/) [32] project is aimed at integrating European mathematical resources. This project is considered as one of the stages of building the World Digital Mathematical Library.

4 Lobachevskii Digital Mathematical Library

In accordance with the basic principles of WDML, a digital library Lobachevskii Digital Mathematics Library (Lobachevskii-DML, https://lobachevskii-dml.ru/)

is being created at the Kazan University [33]. The construction of this library involves the development of management tools for mathematical content that take into account not only the specifics of mathematical texts, but also the peculiarities of processing Russian-language texts. Another objective of this digital library is the integration of the mathematical resources of Kazan University and their inclusion in the global scientific infrastructure, in particular, Math-Net.Ru and EuDML. To solve this problem, methods for the normalization of metadata are being developed in accordance with the schemes of international scientometric databases.

4.1 Use in the Organization of Digital Collections of Semantic Analysis Methods

In the project WDML [26] in the organization of digital collections proposed to use an object approach. It involves the analysis and processing of not only the documents themselves included in the collections, but also the objects contained in these documents (in particular, definitions, mathematical statements and their proofs). This section presents a number of methods that have been developed within the framework of this approach and are implemented in the formation of Lobachevskii-DML's digital scientific collections. These collections were formed as a result of processing an array of unstructured digitized mathematical documents, presented in various formats (.pdf, .tex, .doc, .docx), using the developed special methods. Approbation of the methods is performed on the journal archive "Proceedings of N.I. Lobachevskii Mathematical Center" for 1998–2018, containing more than 60 volumes.

Note that the main purpose of the "Proceedings ..." is the publication of materials of mathematical conferences. As a result, the majority of the volumes of the "Proceedings ..." contain several dozen articles with a limited (from a modern point of view) composition of metadata. Since 1998 (since the release of the first volume), several style rules have been used to prepare materials, which influenced the choice of formats and the design of articles in the collected collections. The prerequisites for creating a digital collection from the array of files "Proceedings ..." were the division of volumes into separate articles, the selection of metadata describing each article, the generation of additional metadata containing, in particular, the bibliographic description of the article, a relation to the article file in the digital collection, as well as relations to the profiles of the authors of the article on academic portals and scientometric databases (kpfu.ru, MathNet.ru, Scopus, etc.). The main steps in creating this digital collection are as follows.

At the first stage, the processed archive was clustered: the volumes of "Proceedings ..." were divided into classes in accordance with the similarity of their structure and design. For each class, a set of regular expression patterns was developed that define the rules for searching information blocks. The basis of this algorithm is the approach proposed in [34, 35]. The algorithm is implemented in the form of programs in the C# language, allowing to process files in TEX, OpenXML (.docx) and .pdf formats. TEX files were processed using standard functions that implement text string operations. PDFLib (https://www.pdflib.com) and iTextSharp libraries (https://www.nuget.org/packages/iTextSharp/) were used to process PDF files. For documents presented in the form of .docx files, the word/document.xml file was extracted from the .docx archive in accordance with the Office OpenXML format (see, for example, [36]).

At the next stage, the metadata that describe both the volume as a whole and the articles included in it were selected from the array of files of the"Proceedings..." volumes. In particular, for all the articles of each volume were allocated their names, as well as the page numbers of their beginning and end. For this, an algorithm was developed that uses the structural homogeneity of each volume and the style uniqueness in the design of articles in it. In addition to the listed metadata, this algorithm allowed us to also highlight lists of authors, bibliography blocks and other metadata (for example, e-mail addresses and keywords), if they are present in the text.

Further, an XML-language was proposed for describing digital mathematical collections, which consists of a set of tags and XML-schemas based on the Journal Archiving and Interchange Tag Suite (https://jats.nlm.nih.gov/1.2d2/). In the notation of this language, on the basis of the data obtained at the stage of processing the initial array of files, a description of the collection "Proceedings..." was carried out.

Using the methods of text analysis [1, 37] from the documents of the digital collection, we have isolated the terms that make up the sets of keywords for inclusion in the metadata. The term extraction algorithm is a development of the approach proposed in [34, 35, 38].

The next step in creating a digital collection included the procedures for dividing each volume of "Proceedings..." into separate articles. To do this, from XML-files containing meta-descriptions of volumes, we read tags, whose attributes point to the starting and ending pages of articles. Next, we divide the files into separate documents, which are assigned names in accordance with the rules adopted in the digital collection. The process of selecting articles was organized using a program developed in Python using the functions of the PyPDF2 library (http://pybrary.net/pyPdf/).

Such metadata as authors' email addresses and their affiliation, we imported from authors' profiles that are presented on academic sites and in various scientific databases, and in parallel they were refined. In this procedure, the semantic links established in the process of forming a digital collection were applied. The corresponding algorithm is based on the method of [33, 35, 38].

The implementation of the algorithms described above allowed us to form a digital collection of the "Proceedings of N.I. Lobachevskii Mathematical Center" and together with the specified set of metadata to include it in the digital library Lobachevskii-DML.

4.2 Formats and Normalization of Metadata Documents of Digital Math Libraries

Metadata Formats. At present, publications on mathematics are indexed in many scientometric databases. These databases impose different requirements on the composition of the metadata of the documents included in them and the schemes for their presentation. On the other hand, digital math libraries also use various metadata formats when building their collections. This is partly due to the fact that the articles included in such collections, being published in journals in accordance with the rules established in them, differ in the requirements for the metadata used. These differences can be quite significant, primarily related to the composition of metadata and their format, and are most noticeable in the archival collections of scientific journals. For example, in many articles published before 2000, there are no keywords and annotations, and the affiliation of authors appeared only in articles of recent years. At the same time, the constantly expanding set of metadata used today testifies to their increasing role in the improvement of modern scientific communications. Thus, there is a need to develop both methods for extracting missing metadata from documents and methods for converting already created metadata into the formats of relevant scientometric databases. Note also that participation in such projects of integration of mathematical resources as EuDML (The European Digital Mathematics Library, https://initiative.eudml.org/) [32, 39], involves the provision of sets of metadata generated according to schemes of aggregators of mathematical resources.

Note that the metadata scheme of the digital mathematical library EuDML is described in [40]: the metadata is divided into basic, fundamental, and additional [41]. To describe journal articles in the EuDML project, XML schemas (NISO JATS V1.0) [42] are used. The mandatory set of EuDML metadata is minimal in composition and contains the title of the article in the original language, the names and surnames of the authors, the list of bibliographies, the unique identifier of the article (for example, doi) and the URL of its full text. The fundamental set of metadata, in addition to the required metadata, includes annotation of the article and keywords.

A number of electronic collections of the digital library Lobachevskii-DML are physically located in other digital libraries. Our tasks are to replenish such collections with additional metadata, as well as automatically selecting objects and establishing semantic links between them.

When forming the fundamental set of metadata of electronic collections stored on external resources, the metadata presented on these resources is initially imported. For this purpose, a program for extracting metadata from web pages and writing them in the XML-format of the digital library Lobachevskii-DML, as well as replenishing and subsequent conversion by EuDML schemas.

As an example, we will point out the archive of articles of the journal “Russian Mathematics (Izvestiya VUZ. Matematika)”. This journal collection is digitized, supplied with meta descriptions, presented on the portal MathNet.Ru (http://www.mathnet.ru/php/journal.phtml?jrnid=ivm) (see also [29]), and is also one of the collections digital library Lobachevskii-DML. The following steps are implemented for this collection.

Part of the metadata was imported from the “Citation in AMSBIB format” block of the MathNet.Ru portal. Then, keywords and a hyperlink to the

Springer Link portal page (https://link.springer.com/journal/11982) with the English version of the article were read from the web page. This information is included in the metadata, and a hyperlink is made.

The next step involves analyzing the web page of the English version of the article, extracting and recording metadata. Next, a personal identifier of this article was generated, which was proposed to be created as a string concatenation – journal identifier (attribute value “jrnid =”) and article identifier (attribute value “paperid =”) on the portal MathNet.Ru.

Normalization of Metadata. By normalization, we mean the use of methods for generating or transforming document metadata in accordance with the rules and XML-schemas of digital libraries and scientometric databases.

One of the most popular and respected computer science libraries is “Dblp Computer Science Bibliography” (DBLP, https://dblp.uni-trier.de/). A prerequisite for the inclusion of electronic collections in this library is the reorganization and normalization of the metadata of the relevant documents. Among the collections of the digital library Lobachevskii-DML, such is the collection of the “Russian Digital Libraries Journal” (https://elbib.ru/). An archive of articles published in this journal, starting in 2015, was chosen to prepare for indexing in DBLP. The necessary metadata are: publication identifier, the names and surnames of the authors, title of work, year of publication, volume, number, starting and ending pages of the article in the journal number and URL of the full text of the article.

Normalization to the DBLP format occurs in three stages: the extraction of the required metadata, the addition of metadata and their normalization into the desired format.

Using the program developed in C# and the System: XML extension tools, the collection files are processed sequentially and, as a result, a set of metadata is generated for each document. At the next stage, the metadata is updated with information about the article and its authors in English. This information is imported from the English version of the journal’s site using the HTMLAgilityPack extension tool. Since the English-language information about the authors is incomplete – only the names and initials are indicated – the names are translated from the Russian-language page. The result of this work was the inclusion of the Russian Digital Libraries Journal and articles published in it in 2015–2018 in the DBLP database (https://dblp.uni-trier.de/db/journals/rdlj/).

Lobachevskii-DML Metadata Factory. As a rule, the term “metadata factory” refers to a set of software tools for managing metadata in digital libraries (see, for example, [28]). These tools are aimed at performing operations such as extracting metadata from digital documents, improving metadata, refining metadata, updating metadata and normalizing metadata into digital library formats and formats of scientometric databases. The structure of the metadata factory of the digital library Lobachevskii-DML also includes semantic transliteration services and a recommendatory system for refining scientific classifiers.

4.3 Digital Mathematical Ecosystem

On the Lobachevskii-DML digital library portal, the OntoMath digital ecosystem is presented, which is an essential part of this digital library [43]. The main components of this ecosystem are: mathematical ontologies Mocassin, OntoMath$^{\mathrm{Pro}}$ and OntoMath$^{\mathrm{Edu}}$, the semantic publishing platform, the semantic search service OntoMathSearch, recommender systems for the selection of scientific classifiers, search for related articles and terminological annotation.

Mathematical Ontologies. The concept of the Semantic Web assumes the semantic structuring of the Inter-net data space for its use by software agents, and the main tasks are the unification (compatibility) and binding of data from different sources. Most relevant to applying Linked Data principles is the LOD project. Its main advantage is in a standardized approach to the structuring and storage of integrated data that is loaded and presented in the form of RDF, that is, triplets of the "subject – predicate – object" type.

An important direction in the development of the Semantic Web domain was the development of ontologies of subject domains, including ontologies of the presentation of mathematical knowledge [44].

The representation and exchange of knowledge in any subject area is based on its conceptualization (see, for example, [17]). The communication process (both between people and between machines) uses a language with a dictionary containing a set of terms to denote elements of conceptualization. Successful communication requires that all its participants, first, share a common conceptualization and, second, use a common vocabulary. A means of solving this problem, as is known, is ontology. Ontology defines the basic concepts of a certain subject area and the relationship between them. The main components of ontology are classes, relations and axioms.

Mocassin Ontology [45] is an ontology of the logical structure of mathematical documents, designed for automatic analysis of mathematical publications in the LaTeX format. This ontology formally (in the OWL language) describes the semantics of the structural elements of mathematical documents (for example, theorems, lemmas, proofs, definitions, etc.) expressed in the form of classes and properties. In addition, the ontology contains the axioms of cardinality and transitivity.

The ontology of professional mathematics OntoMath$^{\mathrm{Pro}}$ [46, 47] is the ontology of mathematical knowledge, which is organized in the form of two hierarchies:

– hierarchies of areas of mathematics: mathematical logic, set theory, algebra, geometry, topology, and so on;
– hierarchies of mathematical objects: set, function, integral, elementary event, Lagrange polynomial, etc.

The OntoMath$^{\text{Pro}}$ ontology is developed in OWL-DL/RDFS and contains 3450 classes, 6 types of object properties, 3630 instances of the IS-A property, and 1140 instances of the remaining properties. It contains five types of relationships: Class → Subclass, Defined with the help, Associative relationship, Task → Solution method and Area of Mathematics → Mathematical object. Ontology concepts contain their name in Russian and English, definition, links to external resources from the Linked Open Data cloud, and links to other concepts. Objects of semantic annotation are also formulas associated with formulas, fragments of text that specify the descriptions of variable formulas.

Ontology of educational mathematics OntoMath$^{\text{Edu}}$. In the current version, this ontology is developed for the system description of the educational aspect of mathematical knowledge. The initial ontology design of OntoMath$^{\text{Edu}}$ is based on the OntoMath$^{\text{Pro}}$ ontology developed by us earlier and described above. A new conceptualization has been created, reflecting the conceptual system of mathematics that corresponds to school education. Professional terminology has been adapted to educational activities, in particular, the language of school mathematics. Relationships reflecting the didactic dependence between the concepts have been added to OntoMath$^{\text{Edu}}$. Ontology concepts contain their names in English, Russian, and Tatar languages, as well as basic definitions, relationships with other ontology concepts (associative relationships), and links to concepts from external data sets. The OntoMath$^{\text{Edu}}$ ontology is built on a set of OntoMath$^{\text{Pro}}$ basic ontology relationships such as taxonomic relation (ISA); the relationship between the mathematical object and the field of mathematics; the relationship between mathematical objects is "determined by"; the relationship between the task and the method of solving it; a new set of didactic relations was also introduced [49].

When creating the top level of ontology OntoMath$^{\text{Edu}}$, the planimetry section of the school mathematics course was selected as a pilot: the current version of the ontology contains 585 concepts related to the planimetry course of 5–9 classes of secondary school. The ontology structure contains type hierarchies; hierarchies of materialized relationships; hierarchy of roles and network of points of view. The specificity of school geometric knowledge was taken into account, therefore, when designing ontology, a number of relations between the concepts were singled out: "whole–part", "determined", relation of ontological dependence, "theorem–property", "theorem–characteristic", "found by formula" (see also [50, 51]).

5 Conclusion

This paper describes approaches to managing large collections of digital mathematical documents that are based on semantic methods and are consistent with the principles of the World Digital Mathematical Library (WDML). These approaches and methods fully relate to the areas that make up the new concept of Big Math. They are being developed and practically implemented as part of the creation of the Lobachevskii-DML digital math library. The main results mentioned are as follows.

Methods for the formation of digital collections from a set of documents – scientific articles, monographs, reports presented in various storage formats are proposed. Based on the analysis of the structure of documents and the stylistic features of their design, an algorithm for extracting their metadata has been developed.

In connection with the increasing role of metadata in the improvement of modern scientific communications, both methods for extracting missing metadata from documents and methods for converting already created metadata into the formats of relevant scientometric databases have been developed and described.

A software tool has been developed for dividing collections of articles into separate documents and forming their semantic presentation. For example, the set of "Proceedings of N.I. Lobachevskii Mathematical Center", which have a different format and structure, describes an algorithm for creating a digital collection and its inclusion in the Lobachevskii-DML digital mathematical library.

Algorithms for enriching the electronic collections of the Lobachevskii-DML digital library and generating metadata of documents of these collections in selected formats are presented.

Services for the normalization of the collection metadata of the Lobachevskii-DML digital library have been developed in accordance with the DTD rules and NISO JATS and DBLP XML schemas. By normalization, we mean the use of methods for generating or transforming document metadata in accordance with the rules and XML schemas of digital libraries and scientometric databases.

Algorithms for creating a mandatory and fundamental collection of metadata collections are presented in accordance with the rules of the European digital mathematical library EuDML.

The digital ecosystem OntoMath, which is the most important part of the Lobachevskii-DML digital library, is described. The main components of this ecosystem are: mathematical ontologies Mocassin, OntoMath$^{\mathrm{Pro}}$ and OntoMath$^{\mathrm{Edu}}$, the semantic publishing platform, the semantic search service OntoMathSearch, recommender systems for the selection of scientific classifiers, search for related articles and terminological annotation.

Acknowledgments

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. 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. 7/1251/2 019. The work was also carried out with the partial support of the Russian Fund for Basic Researches and the Government of the Republic of Tatarstan within the framework of scientific projects 18-29-03086, 18-47-160012.

References

1. Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data. EMC. Education Services (ed.), Wiley (2015).
2. Bradis, V.M.: Four-digit mathematical tables. Moscow: Drofa, 2019.
3. Riche de Prony: Tables des logarithmes, sinus et tangentes pour la division décimale du quart de cercle calcules avec 8 ou 9 décimales pour être imprimées avec 7 d écimales exactes au bureau du Cadas-tre, https://patrimoine.enpc.fr/exhibits/show/dataincognita/item/1817. Last ac-cessed 16 May 2019
4. Bulletin de bibliographie, d'histoire et de biographie mathématiques. Notice sur la découverte des logarithmes. Nouvelles annales de mathématiques. Journal des candidats aux écoles polytechnique et normale, Serie 1, vol. 14, pp. 1–204 (1855) (Additional pages), http://www.numdam.org/item/NAM 1855 1 14 S1 0/. Last accessed 16 May 2019
5. Peaucelle, J.L.: Le détail du calendrier de calcul des tables de Prony de 1791 à 1802. http://rybn.org/human computers/articles/calcul des tables de prony.pdf. Last accessed 16 May 2019
6. Roegel, D.: A reconstruction of the "Tables des logarithmes à huit decimals" from the French "Service géographique de l'armée" (1891). [Research Report] 2010. inria-00543952. https://hal.inria.fr/inria-00543952. Last accessed 16 May 2019
7. Binswanger, M.: Excellence by Nonsense: The Competition for Publications in Modern Science. In: Bartling, S., Friesike, S. (eds). Opening Science. The Evolving Guide on How the Internet is Changing Research, Collaboration and Scholarly Publishing, pp. 49–72. Springer International Publishing (2014). https://doi.org/10.1007/978-3-319-00026-8 3
8. Heller, L., The, R., and Bartling, S.: Dynamic Publication Formats and Collaborative Authoring. In: Bartling, S., Friesike, S. (eds). Opening Science. The Evolving Guide on How the Internet is Changing Research, Collaboration and Scholarly Publishing, pp. 191–211. Springer International Publishing (2014). https://doi.org/10.1007/978-3-319-00026-8 13
9. Knuth, D.E.: The TeX book. Addison-Wesley Publishing Company (1984, 1986, 1991).
10. Cervone, D.: Math Jax: A Platform for Mathematics on the Web. Notices of the AMS **59**, 312–316 (2012).
11. Tantau, T.: The TikZ and PGF Packages. Manual for version 3.1.4a (2019). https://pgf-tikz.github.io/pgf/pgfmanual.pdf. Last accessed 16 May 2019
12. Tools & Technical Specifications. EuDML Enhancer toolset demos. https://initiative.eudml.org/tools-technical-specifications. Last accessed 16 May 2019
13. OpenDreamKit. https://kwarc.info/projects/odk/. Last accessed 16 May 2019
14. Polyanin, A.D. and Zaitsev, V.F.: Handbook of Ordinary Differential Equations. Exact Solutions, Methods, and Problems. CRC Press. Taylor & Francis Group (2018).
15. Kohlhase, M.: Semantic Markup in TeX/LaTeX (2019). http://ctan.altspu.ru/macros/latex/ contrib/stex/sty/stex/stex.pdf. Last accessed 16 May 2019
16. Kohlhase, M.: OMDoc – an open markup format for mathematical documents [Version 1.2]. Springer, Berlin (2006).
17. Elizarov, A.M., Kirillovich, A.V., Lipachev, E.K., Nevzorova, O.A., Solovyev, V.D., and Zhiltsov, N.G.: Mathematical knowledge representation: semantic models and for-malisms. Lobachevskii Journal of Mathematics **35** (4), 348–354 (2014). https://doi.org/10.1134/S1995080214040143
18. Elizarov, A., Kirillovich, A., Lipachev, E., and Nevzorova, O.: Semantic formula search in digital mathematical libraries. Proc. of the 2nd Russia and Pacific Conf.

on Comp. Technology and Applications (RPC 2017). IEEE, pp. 39-43 (2017). https://doi.org/10.1109/RPC.2017.8168063

19. Gorenstein, D.: The Enormous Theorem. Scientific American **253** (6), 104–115 (1985).
20. Solomon, R.: A brief history of the classification of the finite simple groups. Bulletin of the AMS. New Series **38** (3), 315–352 (2001).
21. Appel, K. and Haken, W.: Every map is four Colourable. Bulletin of the AMS **82**, 711–712 (1986).
22. Appel, K. and Haken, W.: Every map is four Colourable. Contemporary Mathematics **98** (1989).
23. Gonthier, G.: Formal Proof – The Four-Color Theorem. Notices of the AMS **55** (11), 1382–1393 (2008).
24. Carette, J., Farmer, W.M., Kohlhase, M., and Rabe, F.: Big Math and the One-Brain Barrier. A Position Paper and Architecture Proposal. arXiv:1904.10405v1 [cs.MS] 23 Apr 2019.
25. Elizarov, A.M., Lipachev, E.K., and Zuev, D.S.: Digital mathematical libraries: Overview of implementations and content management services. CEUR Workshop Proceedings **2022**, 317–325 (2017).
26. Developing a 21st Century Global Library for Mathematics Research. The National Academies Press,Washington (2014). https://doi.org/10.17226/18619
27. Elizarov, A.M., Kirillovich, A.V., Lipachev, E.K., and Nevzorova, O.A.: Mathematical Knowledge Management: Ontological Models and Digital Technology. CEUR Workshop Proceedings **1752**, 44–50 (2016).
28. Bouche, T. and Labbe, O.: The New Numdam platform. CICM 2017: Intelligent Computer Mathematics, 70–82 (2017).
29. Chebukov, D.E., Izaak, A.D., Misyurina, O.G., Pupyrev, Yu.A., and Zhizhchenko, A.B.: Math-Net.Ru as a Digital Archive of the Russian Mathematical Knowledge from the XIX Century to Today. Intelligent Computer Mathematics. LNCS **7961**, 344–348 (2013). https://doi.org/10.1007/978-3-642-39320-4_26
30. Ion, P.: The Effort to Realize a Global Digital Mathematics Library. In: Greuel, G.-M. et al. (eds.). ICMS 2016, LNCS, vol. 9725, pp. 458–466. Springer (2016). https://doi.org/10.1007/978-3-319-42432-3 59
31. Ion, P.D.F. and Watt, S.M.: The Global Digital Mathematics Library and the International Mathematical Knowledge Trust. ICM 2017: Intelligent Computer Mathematics, 2017. LNAI, vol. 10383, pp. 56–69. Springer, 2017. https://doi.org/10.1007/978-3-319-62075-6_5
32. Ion, P.D.F. and Watt, S.M.: The Global Digital Mathematics Library and the International Mathematical Knowledge Trust. ICM 2017: Intelligent Computer Mathematics, 2017. LNAI, vol. 10383, pp. 56–69. Springer, 2017. https://doi.org/10.1007/978-3-319-62075-6_5
33. 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. FIZ Karlsruhe, pp. 57–80 (2013). https://www.emis.de/proceedings/TIEP2013/05bouche.pdf. Last accessed 16 May 2019
34. Elizarov, A., Khaydarov, S., and Lipachev, E.: Scientific documents ontologies for semantic representation of digital libraries. Second Russia and Pacific Conf. on Computer Technology and Applications (RPC). Vladivostok, Russky Island, Russia 25–29 September, pp. 1–5 (2017). https://doi.org/10.1109/RPC.2017.8168064

35. Batyrshina, R.R.: Method for extracting terms in digital mathematical collections. Proc. of the N.I. Lobachevskii Math. Center. Kazan: Kazan Math. Soc. Publ. **55**, 24–26 (2017).
36. Standard ECMA-376 Office Open XML File Formats. http://www.ecmainternational.org/publications/standards/Ecma-376.htm. Last accessed 16 May 2019
37. Ingersoll, G.S., Morton, T.S., and Farris, A.L.: Taming Text. How to Find, Organize, and Manipulate It. Manning Publications Co. (2013).
38. Sabitova, E.M.: Algorithm for extracting connections in scientific digital collections. Proc. of the N.I. Lobachevskii Math. Center. Kazan: Kazan Math. Soc. Publ. **55**, 123–126 (2017).
39. Bouche, T. and 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, San Diego, 99–108 (2013).
40. Jost, M., Bouche, T., Goutorbe, C., and Jorda, J.P.: D3.2: The EuDML metadata schema. http://www.mathdoc.fr/publis/d3.2-v1.6.pdf. Last accessed 16 May 2019
41. EuDML metadata schema specification (v2.0-final). https://initiative.eudml.org/eudml-metadata-schema-specification-v20-final. Last accessed 16 May 2019
42. Journal Article Tag Suite. NISO JATS V1.0. https://jats.nlm.nih.gov/1.0/. Last accessed 16 May 2019
43. Khaydarov, S. and Yamalutdinova, G.: Recommender System of Physical and Mathematical Documents Classification. CEUR Workshop Proceedings **2260**, 480–486 (2018).
44. Elizarov, A., Kirillovich, A., Lipachev, E., and Nevzorova, O.: Digital Ecosystem OntoMath: Mathematical Knowledge Analytics and Management. CCIS **706**, 33–46. Springer (2017). https://doi.org/10.1007/978-3-319-57135-5_3
45. 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
46. Solovyev, V. and Zhiltsov, N.: Logical Structure Analysis of Scientific Publications in Mathematics. Proc. of the Int. Conf. on Web Intelligence, Mining and Semantics (WIMS'11). ACM **21**, 1–9 (2011)
47. Elizarov, A.M., Zhizhchenko, A.B., Zhil'tsov, N.G., Kirillovich, A.V., and Lipachev, E.K.: Mathematical Knowledge Ontologies and Recommender Systems for Collections of Documents in Physics and Mathematics. Doklady Mathematics **93** (2), 231–233 (2016). https://doi.org/10.1134/S1064562416020174
48. Nevzorova, O., Zhiltsov, N., Kirillovich, A., and Lipachev, E.: OntoMathPRO Ontology: A Linked Data Hub for Mathematics. CCIS **468**, 105–119. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-11716-4_9
49. Elizarov, A., Kirillovich, A., Lipachev, E., Nevzorova, O., and Shakirova, L.: Open Linked Data and Ontologies in Mathematics Education. CEUR Workshop Proceedings **2260**, 186–196 (2018).
50. Kirillovich, A., Shakirova, L., Falileeva, M., and Lipachev, E.: Towards an Educational Mathematical Ontology. L. Gómez Chova, et al. (eds). 13th International Technology, Education and Development Conference (INTED2019), Valencia, Spain, March 11-13, 2019. IATED, 6823–6829 (2019).
51. Elizarov, A.M., Lipachev, E.K., and Khaydarov, S.M.: Method of automated selection of reviewers of scientific articles, implemented in the scientific journal information system. Proceedings of the 21th Conference Scientific Services & Internet (SSI 2019), Novorossiysk-Abrau, Russia, September 23-28, 2019.