

Wikidata in Metadata Formation Methods for Documents of Digital Mathematical Library

Alexander M. Elizarov, Polina O. Gafurova, Evgeny K. Lipachev

Institute of Information Technology and Intelligent Systems, Kazan (Volga Region) Federal University, Kremlyovskaya ul., 18, Kazan, 420008, Russia

Abstract

Methods for the formation of digital collections of a digital mathematical library are presented. With the help of tools for analyzing the structure of documents and their style features, the main set of document metadata has been formed. For each document, this set includes the title of the article, a list of authors, and a list of cited bibliography. To supplement the metadata, methods of extracting knowledge from Wikidata were used. With the help of the developed system of SPARQL-queries, the search and refinement of data on documents of the collections was carried out. In particular, information about the authors of the articles has been added (full spelling of surnames, first names, patronymics in various languages, place of work at the time of writing the article, etc.). In addition, methods are proposed for refining and supplementing bibliographic references given in the articles. When forming metadata of retro collections, a search was made in Wikidata for information about the years of life of authors of articles, URLs of web pages with information about articles and their authors. The results of the several digital collections formation, which are included in the digital library Lobachevskii-DML, are presented.

Keywords

Digital Mathematical Libraries, digital mathematical collection, retrodigitized mathematical collection, metadata, metadata factory, Wikidata, Lobachevskii-DML

1. Introduction

Metadata are the basis of communication in the information scientific space and it is used at all stages of the life cycle of a scientific publication (see, for example, [1]). Currently, all scientific publications are “born-digital” (see [2]) and contain at least a minimal metadata set. Modern rules for the scientific publications preparation contains requirements for the inclusion in documents of subject classifiers, keywords, ORCID authors and other information (for example, [3, 4]). Based on all this information, a metadata set for a scientific document is formed.

Scientific documents of the “pre-digital” period, usually, do not contain sufficient information for metadata formation. In such a situation, methods for analyzing the structure of a document allow at least to obtain a basic set of metadata, including the title of an article, a list of authors, and a bibliography [5–7].

Key words and subject classifiers, for example, UDC [8] and MSC [9], are mandatory attributes of a modern scientific publication. Methods of text analysis are used to create or expand the list of keywords (see, for example, [10]). The selection of subject classifiers for mathematical articles is carried out by the methods of automatic classification and categorization (for example, [11–14]). However, these methods are not sufficient to obtain a complete metadata set. For example, in the formation of scientific retro collections, problems arise even with obtaining complete information about the authors of documents. The most important problems of the metadata formation for scientific retro collections documents are presented in [15, 16]. In general, methods of forming metadata of mathematical documents are being developed in projects for creating digital mathematical libraries (see, for example, [15, 17–23]).

The main goal of the “Lobachevskii Digital Mathematical Library” project is to create a system of interconnected software services that ensure the formation, processing, storage and management of digital

SSI-2021: Scientific Services & Internet, September 20–23, 2021, Moscow (online)

EMAIL: amelizarov@gmail.com (A.M. Elizarov); pogafurova@gmsil.com (P.O. Gafurova); elipachev@gmail.com (E.K. Lipachev)

ORCID: 0000-0003-2546-6897 (A.M. Elizarov); 0000-0002-1544-155X (P.O. Gafurova); 0000-0001-7789-2332 (E.K. Lipachev)



© 2021 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

library objects, as well as the integration of the created collections into aggregating digital mathematical libraries. Within the framework of this project, the digital mathematical library Lobachevskii-DML (<https://lobachevskii-dml.ru/>) was designed [24]. When developing methods and implementing tools for managing metadata of digital collections documents, we applied xml-schemes used in the European Digital Mathematics Library (<https://initiative.eudml.org/>) [25, 26].

This paper describes a method for enriching a set of metadata documents in digital collections using knowledge extraction from Wikidata [27]. One of the results obtained is an algorithm for supplementing the metadata of the retro collection of the journal “Izvestia of the Physics and Mathematics Society at Kazan University” (“Bulletin de la Société Physico-Mathématique de Kasan”) (hereinafter – “Izvestia”) and the collection of collections of conference proceedings “Proceedings of Lobachevskii mathematical center” (hereinafter – “Proceedings”). Both of these collections are part of the Lobachevskii-DML digital library.

2. Digital math library collection metadata workflows

The Lobachevskii DML library includes a number of digital collections. During the creation of some of the collections it was necessary to complete the full cycle of their formation: from digitizing paper documents to loading metadata and the digital documents themselves into the library. These collections include “Proceedings” [28], as well as “Izvestia” [29]. “Proceedings” have been published since 1998, and until 2015, most of the issues were only in paper form. The archives of “Izvestia” journal were kept in the Scientific Library of Kazan University only in paper form and in single copies.

Algorithms for the formation of document metadata in the collections of the digital mathematical library Lobachevskii-DML are presented in [30, 31].

The formation of a digital collection of mathematical documents consists of the following main stages:

- Digitization of documents;
- Division of journal issues into separate articles;
- Extraction of metadata from articles by methods of document structure analysis and NLP;
- Clarification of metadata;
- Supplementing metadata with information from Wikidata;
- Formation of metadata of articles on xml-schemes of the digital library;
- Integration of the digital collection into the digital mathematical library;
- Normalization according to xml-schemes of aggregating digital libraries.

The metadata of digital collections documents is created by the software services of the Lobachevskii-DML digital library metadata factory. These services implement methods based on the analysis of the structure of documents and the peculiarities of their styling [5, 6].

In Fig. 1 shows a fragment of a document of one of the issues of “Proceedings”, as well as a fragment of the generated metadata. Note that only the title of the article, the author’s names and the abstract can be extracted with standard methods of text analysis.

A feature of the documents of the digital collection of “Proceedings”, like many other collections of conference materials, is the lack of uniform requirements for the structure of scientific documents included in these editions (see Fig. 2). This circumstance complicates the process of extracting metadata using methods based on the analysis of the structure of the document and its style features.

Key words, annotations and subject classifiers are present only in a small number of collections, while this information is necessary for the formation of metadata sets according to the schemas of aggregators of mathematical documents [25, 32].

Using the tools of the metadata factory, we carried out a procedure for normalizing metadata in accordance with the DTD rules and XML schemas of the Journal Archiving and Interchange Tag Suite (NISO JATS) [33]. Thus, a set of metadata was formed in the form of an item-structure, which includes both the content of the metadata and information about the language of their presentation. This set of metadata makes it possible to include not only the names and surnames of the authors given in the article, but also to supplement them with alternative spellings indicating the language. As a result of the operation of the corresponding software application, a set of files in the JATS format was generated, which describe each article from the processed source [34, 35].

GEOMETRIC CONVERGENCE OF ITERATIVE METHODS FOR VARIATIONAL INEQUALITIES WITH M-MATRICES AND DIAGONAL MONOTONE OPERATORS¹

A.V. Lapin

Abstract

Iterative solution for the problem

$$Au + B\gamma + \delta = f, \quad \gamma \in Cu, \quad \delta \in Du \quad (1)$$

with M - matrices A, B and diagonal maximal monotone (multivalued) operators is C, D is studied. Geometric (or, linear) rate of convergence is proved for several classes of iterative algorithms under some additional assumptions imposed to input data. Applications of general results to mesh schemes for free boundary problems are considered.

1 Introduction

Finite element and finite difference schemes for the free boundary problems with several constraints and/or with several convex non-differentiable functionals lead to the algebraic problem (1). We can

```
<paper id="5">
<author> J. Karppinen, E. Laitinen </author>
<title-paper> ON FEM SOLUTION FOR 3D CONTINUOUS CASTING
PROBLEM </title-paper>
<start-page> 48 </start-page>
<end-page> 58 </end-page>
</paper>
<paper id="6">
<author> I.V. Konnov </author>
<title-paper> MIXED VARIATIONAL INEQUALITIES AND EXACT
PENALTY FUNCTIONS </title-paper>
<start-page> 59 </start-page>
<end-page> 63 </end-page>
</paper>
<paper id="7">
<author> A.V. Lapin </author>
<title-paper> GEOMETRIC CONVERGENCE OF ITERATIVE METHODS
FOR VARIATIONAL INEQUALITIES WITH M-MATRICES AND DIAGONAL
MONOTONE OPERATORS </title-paper>
<start-page> 63 </start-page>
<end-page> 72 </end-page>
</paper>
<paper id="8">
<author> S.A. Lapin </author>
<title-paper> THE FICTITIOUS DOMAIN METHOD FOR THE MIXED
FINITE ELEMENT APPROXIMATION OF THE WAVE EQUATION
</title-paper>
<start-page> 72 </start-page>
<end-page> 81 </end-page>
</paper>
<paper id="9">
<author> A.B. Mazo </author>
<title-paper> EXTERNAL PROBLEMS OF INCOMPRESSIBLE
```

Figure 1: Fragment of an article from the digital collection of “Proceedings” and metadata extracted by the tools of the metadata factory

<p>ПОСТУПАТЕЛЬНОЕ ДВИЖЕНИЕ КРУГОВОГО ЦИЛИНДРА В ОТКРЫТОМ КАНАЛЕ С ДВУХСЛОЙНОЙ ЖИДКОСТЬЮ</p> <p>А.М. Елизаров, А.О. Карамышева, С.И. Филиппов</p> <p><i>НИИ математики и механики им. Н.Г. Чеботарева Казанского государственного университета e-mail: Sergei.Filippov@ksu.ru</i></p>
<p>Ф.Г. Авхадиев, Л.А. Аксентьев, А.М. Елизаров, С.Р. Насыров (Казань)</p> <p>НАУЧНЫЙ СЕМИНАР ПО ГЕОМЕТРИЧЕСКОЙ ТЕОРИИ ФУНКЦИЙ: ОСНОВНЫЕ РЕЗУЛЬТАТЫ ДВУХ ПОСЛЕДНИХ ДЕСЯТИЛЕТИЙ</p>
<p style="text-align: right;"><i>Е.К. Липачёв</i></p> <hr style="width: 50%; margin: 0 auto;"/> <p>О КРАЕВЫХ ЗАДАЧАХ ДЛЯ УРАВНЕНИЯ ГЕЛЬМГОЛЬЦА В ОБЛАСТЯХ С “НЕРОВНОЙ” ГРАНИЦЕЙ ¹</p> <p>Е.К. Липачёв (Казань)</p>

Figure 2: Differences in the design of collections of conference materials on the example of volumes 12, 14 and 17 of “Proceedings” (2002) – the difference in the structure of documents and completeness of information about the authors

One of the structural features of the JATS metadata format is the need to choose the main language for presenting an article, and the rest of the languages are declared alternative. This creates difficulties in the formation of multilingual collections.

The choice of the main presentation language is one of the issues that have to be noticed when creating xml-presentation of documents. One of the options for solving this problem is to use the original article language, but this does not always allow organizing an effective search in the collections of the Lobachevskii-DML library. This is due to the fact that the digital collections of this library contain mainly articles in Russian, and most of the materials from the retro collections are documents in the pre-reform Russian language. When processing articles, in this case, difficulties arise in writing the titles of articles and names of authors, as well as additional information necessary for the formation of metadata.

Difficulties in meta-description of documents of retro collections in the JATS format arise with articles published in parts in various issues of the journal, as well as with articles that have continuation (this, as a rule, is said only in the text of an article).

3. What can be obtained from Wikidata

Wikidata is the Wikipedia knowledge base and central data management platform for Wikipedia and part of Wikipedia ecosystem (see, for example, [36, 37]). Since the launch of Wikidata in 2012, the site of this project, with the participation of more than 5 million registered users, has collected data on 96,228,512 items (as of December 1, 2021) [38]. The significant professional interest in the project is due to the fact that Wikidata covers a wide range of general and specialized knowledge, relevant in many areas of application. Most Wikidata claims are provided with information about their provenance, as well as additional contextual information such as time validity. In addition, the data is linked to external datasets in many areas of knowledge, and the information is duplicated in different languages.

Real world objects are represented in Wikidata by items. Each element is assigned a numeric identifier prefixed with “Q”. Items correspond to Wikipage in the Wikidata main namespace. The wikipage of each item is organized as properties and statements. Instances of properties and statements are also called entities and have their own identifiers (prefixed with “Q” for statements and “P” for properties), which serve as an important source of item metadata [39]. Both elements and properties have a label, description, and (multilingual) aliases. The Wikidata data model is described in [40, 41]. The peculiarities of working with named entities in Wikidata are highlighted in [41]. Formulas are present in all mathematical articles. Methods for representing formulas in Wikidata are given in [42].

The image shows a screenshot of the Wikidata page for the item 'Andrey Markov' (Q176659). The page is divided into several sections:

- birth name:** Андрей Андреевич Марков (Russian)
- given name:** (empty)
- family name:** (empty)
- date of birth:** (empty)
- place of birth:** (empty)
- date of death:** (empty)
- occupation:**
 - mathematician (0 references)
 - statistician (0 references)
 - university teacher (1 reference)
- field of work:**
 - Russian mathematician (Q176659)
 - Andrey Andreyevich Markov | A. A. Markov | Andrei Markov
 - In more languages table:

Language	Label	Description	Also known as
English	Andrey Markov	Russian mathematician	Andrey Andreyevich Markov A. A. Markov Andrei Markov
Russian	Андрей Андреевич Марков	русский математик (1856 – 1922)	Марков, Андрей Андреевич ... Андрей Марков старший старший Андрей Марков
Tatar	No label defined	No description defined	
Bashkir	No label defined	No description defined	
- employer:** (empty)

Figure 3: Page from Wikidata for the query “mathematician Andrey Markov”

In Fig. 3 shows the Wikidata page obtained at the request of “mathematician A. Markov” in the process

of forming the metadata of the “Izvestia” retro collection. The workflows for creating this digital collection and the specifics of generating metadata for retro documents are described in [31]. The most significant is the problem of identifying the authors of articles when filtering query results. The authors of the articles in this collection are indicated in the issues of the journal only by their surname and initials, sometimes even with one initial (for example, “А. Марковъ”). When processing the results of such queries, filtering by several criteria was required, as well as verification by experts.

On the Wikidata page (Fig. 3), the information that was used in the formation of the metadata of the documents of the digital collection “Izvestia” is presented. We indicate only the main properties, the values of which were included in the metadata: *birth name* (“Андрей Андреевич Марков (Russian)”), *given name* (“Andrey”), *family name* (“Markov”), *date of birth* (“2 June 1856^{Julian}”, “14 June 1856^{Gregorian}”), *date of death* (“20 July 1922^{Gregorian}”), *occupation* (“mathematician”, “statistician”, “university teacher”), *field of work* (“probability theory”, “mathematical analysis”, “number theory”), *employer* (“Saint Petersburg Academy of Sciences”, “Saint Petersburg State University”). The *unique ID* (“Q176659”) stored in the metadata is used later to retrieve updated information from Wikidata.

4. Algorithm for enriching metadata from Wikidata

As a source of metadata replenishment, we used the open resources of the Semantic Network. The software tools of the metadata factory of the digital mathematical library Lobachevskii-DML, based on the text analysis of documents from digital retro collections, made it possible to extract such metadata as the title of the article, bibliographic references, page ranges, the names of authors in the original language (Russian, pre-reform Russian, German, French or English). Currently, there is information on the Internet about the authors of most of the articles that was absent in the articles themselves. This makes it possible to extract from the network resources the missing information about the authors of the articles, in particular, the spelling of the surname in different languages, the first and middle name, the place of work at the time of writing the article.

Note that there are a number of data binding services that contain objects of mathematical knowledge. Most of them have a connection point (SPARQL endpoint) [43].

The screenshot shows the Wikidata page for Q16648192. At the top, it says "No label defined (Q16648192)". Below this, there is a table with columns for Language, Label, Description, and Also known as. The table lists labels in English, Russian, Tatar, and Bashkir. The Russian label is "Елизаров, Александр Михайлович". Below the table, there are two statements: "instance of" with the value "human" and "sex or gender" with the value "male".

Language	Label	Description	Also known as
English	No label defined	No description defined	
Russian	Елизаров, Александр Михайлович	No description defined	Александр Михайлович Елиза...
Tatar	No label defined	No description defined	
Bashkir	No label defined	No description defined	

Statements:

- instance of: human (1 reference)
- sex or gender: male (1 reference)

Figure 4: Person's Wikidata page. For this person, there is only a Russian-language page in Wikipedia, therefore there is no label on the English-language page

In Fig. 4 shows an example of processing the result obtained by request in Wikidata and subsequent

filtering by a number of features. Wikidata found 229 items in a query containing the name of the author of the article. After the refinement procedure, an element with the label Q16648192 is highlighted, containing information about the author of the article (Fig. 3).

Table 1 lists the main properties, information from which is used in the formation of additional metadata.

Table 1

The main Wikidata properties used in the metadata replenishment algorithm (for example, an item with ID Q570859)

Property	ID	Example content	Jats_Tag
<i>family name</i>	P734	Chebotaryov	<surname>
<i>given name</i>	P735	Nikolai	<given-names>
<i>name in native language</i>	P1559	Николай Григорьевич Чеботарёв (Russian)	<string-name>
<i>birth name</i>	P1477	Николай Григорьевич Чеботарёв (Russian)	<string-name>
<i>date of birth</i>	P569	3 June 1894 ^{Julian} , 15 June 1894 ^{Gregorian} , 1894	<def-list>
<i>date of death</i>	P570	2 July 1947, 1947	<def-list>
<i>occupation</i>	P106	Mathematician (Q170790), university teacher (Q1622272)	<def-list>
<i>employer</i>	P108	Kazan Federal University (Q113788)	<aff>
<i>member of</i>	P463	Academy of Sciences of the USSR (Q2370801)	<aff-alternatives>
<i>academic degree</i>	P512	Doctor of Sciences in Physics and Mathematics (Q17281097)	<degrees>
<i>field of work</i>	P101	number theory (Q12479), algebra (Q3968), function theory (Q4455174)	<def-list>
<i>notable work</i>	P800	Chebotarev's density theorem (Q1425529), Chebotarev theorem on roots of unity (Q17007435)	<def-list>

An important property used in SPARQL-queries against Wikidata is the occupation property (P106). It can be used to filter the query results, leaving only the pages of documents of those persons who are associated with scientific activities (see Table 2).

Table 2

The main criteria for selecting item by title and the number of corresponding pages

occupation (P106)	ID	Number of results
<i>scientist</i>	Q901	444354
<i>mathematician</i>	Q170790	35182
<i>researcher</i>	Q1650915	148544
<i>university teacher</i>	Q1622272	164512

Note that synonymous properties are taken into account when creating queries to Wikidata. They provide different ways of getting the same data, for example, the properties “name in native language” and “birth name” give the same results.

Let us now give an algorithm for enriching metadata using SPARQL queries to Wikidata.

Algorithm 1: Enriching the metadata of a digital collection document

```
1: read metadata_set
2: List<string> authors_result = selected content from tag <authors>
   #List of metadata in xml format
3: List <XElement> metadata
4: foreach authors_str in authors_results
5:     List <string> authors = Split(authors_str)
6:     foreach author in authors
7:         #author's search in Wikidata,
           form SPARQL requests for Wikidata by family name (P734)
           #example of the request in fig 5 and 6
8:         get list Request_list from request
9:         filter out by initials (from birth name or name in native
           Language), occupation set (from Table 2)
10:        if Request_list.Length>1 then expert verification required
11:        else
           #Attributes of class Table is from Table 1, and have the same
           names, also it have a list, example of the list of requests
           in fig 7
12:        List<Table> Props = new List<Table>
13:        fill in the attributes ID, Jats_Tag, Property for each class
           instance
14:        foreach Prop in Props
15:            form SPARQL requests for Wikidata: property is Prop.ID
16:            get content for Prop.Content
           #Formation of metadata set
17:            form a metadata_set using list Props
18:        metadata.Add(metadata_set)
19: form new metadata_set
20: save new metadata_set
```

The search is done using the MediaWiki API service. It allows you to call the MediaWiki API from SPARQL and get results from a SPARQL-query. Below are some of the queries that are used in the algorithm.

Fig. 5 presents a standard request for searching in Wikidata for additional information on the author of the article (corresponds to step 7 of Algorithm 1; searches in Wikidata for pages of documents with the name of the author of the article).

```
select ?item where {
  ?item rdfs:label "Елизаров"@ru.
  ?item wdt:P31 wd:Q101352.
}
```

Figure 5: Request with the *instance of property* (P31) with an explicit indication of *the entity family name* (Q101352)

Now let's search for the entities obtained in step 7 of Algorithm 1. By filtering by profession ("scientist" or another value from Table 2), in most cases, the results are narrowed down to links to article pages of the desired author (Fig. 6).

```
SELECT DISTINCT ?item ?itemLabel WHERE {
  SERVICE wikibase:label { bd:serviceParam wikibase:language "ru". }
  {
    SELECT DISTINCT ?item WHERE {
      ?item p:P734 ?statement0.
```

```

        ?statement0 (ps:P734/(wdt:P279*)) wd:Q21507140.
        ?item p:P106 ?statement1.
        ?statement1 (ps:P106/(wdt:P279*)) wd:Q901.
    }
    LIMIT 100
}
}

```

Figure 6: Search query for the entity obtained in the previous step of the algorithm, filtered by the value “scientist” (Q901) of the *occupation property* (P106)

The request to get all the metadata specified in Table 1 is presented in Fig. 7. The result includes not only a reference to an entity, but also the value of that entity.

```

select * where {
  wd:Q570859 wdt:P734 ?family_name_id.
  ?family_name_id rdfs:label ?family_name filter(lang(?family_name) =
'ru')
  wd:Q570859 wdt:P735 ?given_name_id.
  ?given_name_id rdfs:label ?given_name filter(lang(?given_name) =
'ru')
  wd:Q570859 wdt:P1559 ?name_in_native_language.
  wd:Q570859 wdt:P1477 ?birth_name.
  wd:Q570859 wdt:P569 ?date_of_birth.
  wd:Q570859 wdt:P570 ?date_of_death.
  wd:Q570859 wdt:P106 ?occupation_id.
  ?occupation_id rdfs:label ?occupation filter(lang(?occupation) =
'ru')
  wd:Q570859 wdt:P108 ?employer_id.
  ?employer_id rdfs:label ?employer filter(lang(?employer) = 'ru')
  wd:Q570859 wdt:P463 ?member_of_id.
  ?member_of_id rdfs:label ?member_of filter(lang(?member_of) = 'ru')
  wd:Q570859 wdt:P512 ?academic_degree_id.
  ?academic_degree_id rdfs:label ?academic_degree filter(lang(?aca-
demic_degree) = 'ru')
  wd:Q570859 wdt:P101 ?field_of_work_id.
  ?field_of_work_id rdfs:label ?field_of_work fil-
ter(lang(?field_of_work) = 'ru')
  wd:Q570859 wdt:P800 ?notable_work_id.
  ?notable_work_id rdfs:label ?notable_work filter(lang(?notable_work)
= 'ru')
}
}

```

Figure 7: Request to get all metadata listed in Table 1

Next, we will process the results of SPARQL-queries. It includes the transformation into a metadata set in JATS format. A fragment of the obtained metadata is shown in Fig. 8. Note that the Wikidata entity id is used internally to represent digital collection documents.

4. Conclusion

The further direction of our work involves the organization of the refinement of search results when adding metadata by adding the main topic of the article using ontologies, as well as solving the issues of submitting the article in various languages. Future work is to include such a semantic graph as Dbpedia in queries, as well as replenishment of these semantic networks.

5. Acknowledgements

The research was funded by RSF according to the project No. 21-11-00105.

```
<contrib-group>
  <contrib contrib-type="author">
    <name-alternatives>
      <name>
        <surname id="Q21493235" xml:lang="ru">Чеботарёв</surname>
        <surname id="Q21493235" xml:lang="en">Chebotaryov</surname>
        <given-names id="Q5486169" xml:lang="ru">Николай</given-names>
        <given-names id="Q5486169" xml:lang="en">Nikolai</given-names>
        <string-name id="P1559" xml:lang="ru">Николай Григорьевич Чеботарёв</string-name>
        <string-name id="P1477" xml:lang="ru">Николай Григорьевич Чеботарёв</string-name>
      </name>
    </name-alternatives>
    <bio>
      <def-list id="P569">
        <def-item>3 June 1894 Julian</def-item>
        <def-item>15 June 1894 Gregorian, </def-item>
        <def-item>1894</def-item>
      </def-list>
      <def-list id="P570">
        <def-item>2 July 1947</def-item>
        <def-item>1947</def-item>
      </def-list>
      <def-list id="P106">
        <def-item id="Q170790">mathematician</def-item>
        <def-item id="Q1622272">university teacher</def-item>
      </def-list>
      <def-list id="P101">
        <def-item id="Q12479">number theory</def-item>
        <def-item id="Q3968">algebra</def-item>
        <def-item id="Q4455174">function theory</def-item>
      </def-list>
      <def-list id="P800">
        <def-item id="Q1425529">Chebotarev's density theorem</def-item>
        <def-item id="Q17007435">Chebotarev theorem on roots of unity</def-item>
      </def-list>
    </bio>
    <aff id="Q113788">Kazan Federal University</aff>
    <aff-alternatives id="Q2370801">Academy of Sciences of the USSR </aff-alternatives>
    <degrees id="Q17281097">Doctor of Sciences in Physics and Mathematics </degrees>
  </contrib>
</contrib-group>
```

Figure 8: Fragment of the Jats-representation of the document with the meta description of the author of the article based on information obtained from Wikidata

References

- [1] I. Xie, K. K. Matusiak, Discover Digital Libraries: Theory and Practice. Elsevier Inc. 388 p. (2016).
- [2] Born-digital. URL: <https://en.wikipedia.org/wiki/Born-digital>.
- [3] Author Guide – ScholarOne Manuscripts. Clarivate Analytics. 2019. pp. 1–70.
URL: https://clarivate.com/webofsciencegroup/wp-content/uploads/sites/2/dlm_uploads/2019/10/ScholarOne-Manuscripts-Author-Guide.pdf.
- [4] Author tutorials. Writing a journal manuscript. Springer Nature Switzerland AG, 2021.
URL: <https://www.springernature.com/gp/authors/campaigns/writing-a-manuscript>.
- [5] E. Biryal'tsev, A. Elizarov, N. Zhil'tsov, E. Lipachev, O. Nevzorova, V. Solov'ev, Methods for Analyzing Semantic Data of Electronic Collections in Mathematics, Automatic Documentation and Mathematical Linguistics, 48 (2) (2014) 81–85. <https://doi.org/10.3103/S000510551402006X>.
- [6] A. M. Elizarov, E. K. Lipachev, S. M. Khaydarov, Automated system of services for processing of large collections of scientific documents, CEUR Workshop Proceedings 1752 (2016) 58–64.
- [7] D. Tkaczyk, New Methods for Metadata Extraction from Scientific Literature, arXiv:1710.10201v1. 2017. URL: <https://arxiv.org/pdf/1710.10201v1.pdf>.
- [8] Universal Decimal Classification. URL: <https://udcc.org/index.php>.

- [9] MSC2020 – Mathematics Subject Classification System.
URL: <https://mathscinet.ams.org/msnhtml/msc2020.pdf>.
- [10] H. Lane, H. Hapke, C. Howard, *Natural Language Processing in Action: Understanding, analyzing, and generating text with Python*. Manning Publications, 2019.
- [11] R. Řehůřek, P. Sojka, Automated Classification and Categorization of Mathematical Knowledge, In: S. Autexier, J. Campbell, J. Rubio, V. Sorge, M. Suzuki, F. Wiedijk (Eds.), *Intelligent Computer Mathematics. CICM 2008. Lecture Notes in Computer Science*. Springer, Berlin, Heidelberg. 2008, vol. 5144, pp. 543–557. https://doi.org/10.1007/978-3-540-85110-3_44
- [12] S. M. Khaydarov, G. S. Yamalutdinova, Recommender System of Physical and Mathematical Documents Classification, *CEUR Workshop Proceedings 2260 (2018)* 480–486. URL: http://ceur-ws.org/Vol-2260/57_480-486.pdf.
- [13] M. Schubotz, P. Scharpf, O. Teschke, A. Kühnemund, C. Breiting, B. Gipp, AutoMSC: Automatic Assignment of Mathematics Subject Classification Labels, *Proceedings of the 13th Conference on Intelligent Computer Mathematics*. 2020. arXiv:2005.12099v1. 25 May 2020.
- [14] Olga Nevzorova, Damir Almukhametov, Towards a Recommender System for the Choice of UDC Code for Mathematical Articles, *CEUR Workshop Proceedings 3036 (2021)* 54–62. URL: <http://ceur-ws.org/Vol-3036/paper04.pdf>.
- [15] E. M. Rocha, J. F. Rodrigues, Disseminating and preserving mathematical knowledge, in: J. M. Borwein, E. M. Rocha, J. F. Rodrigues (Eds.), *Communicating Mathematics in the Digital Era*. A K Peters, Ltd., 2008. P. 3–21.
- [16] P. Gafurova, A. Elizarov, E. Lipachev, Algorithms for Integration of Unstructured Mathematical Documents into the Common Digital Space of Scientific Knowledge, *CEUR Workshop Proceedings 2990 (2021)* 39–49. URL: <http://ceur-ws.org/Vol-2990/rpaper4.pdf>.
- [17] A. B. Zhizhchenko, A. D. Izaak, The Information System Math-Net.Ru. Application of Contemporary Technologies in the Scientific Work of Mathematicians, *Russian Math. Surveys* 62:5 (2007) 943–966. <https://doi.org/10.1070/RM2007v062n05ABEH004455>
- [18] T. Bouche, Toward a Digital Mathematics Library? A French Pedestrian Overview in: J. M. Borwein, E. M. Rocha, J. F. Rodrigues (Eds.), *Communicating Mathematics in the Digital Era*, A K Peters, Ltd., 2008, pp. 47–73.
- [19] D. E. Chebukov, A. D. Izaak, O. G. Misyurina, Yu. A. Pupyrev, A. B. Zhizhchenko, Math-Net.Ru as a Digital Archive of the Russian Mathematical Knowledge from the XIX Century to Today, *CICM'13: Proceedings of the 2013 International Conference on Intelligent Computer Mathematics*. July 2013, pp. 344–348. https://doi.org/10.1007/978-3-642-39320-4_26.
- [20] M. Bartošek, J. Rákosník, DML-CZ: The Experience of a Medium-Sized Digital Mathematics Library. *Notices of the AMS* 60:8 (2013) 1028–1033. URL: <http://dx.doi.org/10.1090/noti1031>.
- [21] A. M. Elizarov, E. K. Lipachev, D. S. Zuev, Digital Mathematical Libraries: Overview of Implementations and Content Management Services, *CEUR Workshop Proceedings 2022 (2017)* 317–325.
- [22] P. D. F. Ion, S. M. Watt, The Global Digital Mathematics Library and the International Mathematical Knowledge Trust, in: H. Geuvers, M. England, O. Hasan, F. Rabe, O. Teschke (Eds.), *Intelligent Computer Mathematics – CICM 2017, Lecture Notes in Computer Science*, Springer, Cham. 2017, vol. 10383, pp. 56–69. https://doi.org/10.1007/978-3-319-62075-6_5
- [23] A. Elizarov, E. Lipachev, Digital Libraries and the Common Digital Space of Mathematical Knowledge, *CEUR Workshop Proceedings 2990 (2021)* 25–38, URL: <http://ceur-ws.org/Vol-2990/rpaper3.pdf>.
- [24] A. M. Elizarov, E. K. Lipachev, Lobachevskii DML: Towards a Semantic Digital Mathematical Library of Kazan University, *CEUR Workshop Proceedings 2022 (2017)* 326–333. URL: <http://ceur-ws.org/Vol-2022/paper50.pdf>.
- [25] EuDML metadata schema specification (v2.0–final). URL: <https://initiative.eudml.org/eudml-metadata-schema-specification-v20-final>.
- [26] T. Bouche, J. Rákosník, Report on the EuDML External Cooperation Model, in: K. Kaiser, S. G. Krantz, B. Wegner (Eds.), *Topics and Issues in Electronic Publishing, JMM, Special Session*. San Diego, 2013, pp. 99–10. URL: https://www.emis.de/proceedings/TIEP2013/07bouche_rakosnik.pdf, last accessed 2021/11/07.
- [27] Wikidata: Main_Page. URL: https://www.wikidata.org/wiki/Wikidata:Main_Page.
- [28] Digital Collection: Proceedings of Lobachevskii mathematical center.
URL: <https://lobachevskii-dml.ru/journal/tmt>.
- [29] Digital Collection: “Izvestia of the Physics and Mathematics Society at Kazan University”. URL:

- <https://lobachevskii-dml.ru/journal/izfmo2>, <https://lobachevskii-dml.ru/journal/izfmo3>
- [30] A. Elizarov, E. Lipachev, Methods of Processing Large Collections of Scientific Documents and the Formation of Digital Mathematical Library, CEUR Workshop Proceedings 2543 (2020) 354–360. URL: <http://ceur-ws.org/Vol-2543/spaper05.pdf>.
- [31] , A. M. Elizarov, P. O. Gafurova, E. K. Lipachev, Metadata Extraction Methods for Organizing a Retro-Collection in the Lobachevskii Digital Mathematical Library, CEUR Workshop Proceedings 2784 (2020) 62–71. URL: <http://ceur-ws.org/Vol-2784/rpaper06.pdf>.
- [32] dblp computer science bibliography. URL: <https://dblp.uni-trier.de/>.
- [33] Journal Article Tag Suite. URL: <https://jats.nlm.nih.gov/about.html>.
- [34] P. O. Gafurova, A. M. Elizarov, E. K. Lipachev, D. M. Khammatova, Metadata Normalization Methods in the Digital Mathematical Library, CEUR Workshop Proceedings 2543 (2020) 136–148, URL: <http://ceur-ws.org/Vol-2543/rpaper13.pdf>.
- [35] A. Elizarov, E. Lipachev, Digital Library Metadata Factories, CEUR Workshop Proceedings 2813 (2021) 13–21, URL: <http://ceur-ws.org/Vol-2813/rpaper01.pdf>.
- [36] D. Vrandečić, M. Krötzsch, Wikidata: a free collaborative knowledgebase, Communications of the ACM 57:10 (2014) 78–85. <https://doi.org/10.1145/2629489>.
- [37] Wikipedia: Wikidata (2021). URL: <https://en.wikipedia.org/wiki/Wikidata>, last accessed 2021/11/07.
- [38] Statistics – Wikidata. URL: <https://www.Wikidata.org/wiki/Special:Statistics>.
- [39] Wikidata: Glossary. URL: <https://www.Wikidata.org/wiki/Wikidata:Glossary>.
- [40] F. Erxleben, M. Günther, M. Krötzsch, J. Mendez, D. Vrandečić, Introducing Wikidata to the Linked Data Web, in: P. Mika et al. (Eds.), The Semantic Web – ISWC 2014. ISWC 2014. Lecture Notes in Computer Science. Springer, Cham. 2014, vol. 8796, pp. 50–65. https://doi.org/10.1007/978-3-319-11964-9_4.
- [41] J. Geiß, A. Spitz, M. Gertz, NECKAr: A Named Entity Classifier for Wikidata, in: G. Rehm, T. Declerck (Eds.), Language Technologies for the Challenges of the Digital Age. GSCL 2017. Lecture Notes in Computer Science, Springer, Cham. 2018, vol. 10713, pp 115–129. https://doi.org/10.1007/978-3-319-73706-5_10.
- [42] Ph. Scharpf, M. Schubotz, B. Gipp, Mathematics in Wikidata, CEUR Workshop Proceedings 2982 (2021) 1–14. URL: <http://ceur-ws.org/Vol-2982/paper-1.pdf>.
- [43] SPARQL Query Language for RDF/W3C. URL: <https://www.w3.org/TR/rdf-sparql-query/>.