Mathematical Physics Problems: Thesaurus and Ontology

Olga Ataeva[1], Vladimir Serebryakov[1], Natalia Tuchkova[1]

[1]Dorodnicyn Computing Centre, Federal Research Centre "Computer Science and Control" of the Russian Academy of Sciences, 40 Vavilov St., Moscow

Abstract. The work is devoted to the study of knowledge representation in the subject area "Equations of mixed type in the sections of mathematical physics". Such comprehensive resources as Wikipedia, claim to be encyclopedic knowledge, but cannot provide informational support for in-depth research. This is a field of activity for specialists in specific areas of knowledge.

The thesaurus and ontology are considered as a formal means of describing the subject area. This approach takes into account the peculiarities of knowledge representation in the domain, namely the presence and use of formulas as independent objects. Consideration of the domain features leads to a reduction in search noise and a reduction in search time within the framework of the constructed library. The use of thesaurus and ontology in the design of a digital semantic library is considered.

Keywords: mathematical physics, thesaurus, ontology, formulas and texts, digital libraries

1 Introduction

The issues with information representation of mathematical knowledge in digital space are directly related to the logical framework organization set within the mathematical subject domains. The research is prompted by spreading digital information representation in mathematical sciences and explained by the effect mathematics now have on the developed countries' economy [1]. Many papers highlight that the way mathematical branches are presented on the Internet is significant, for the future of science in general as well [2]. The analysis of digital mathematical resources shows that it is essential to generate relevant thesauri to build digital information images of mathematical subject domains relying on scientific knowledge ontologies. Such comprehensive resources as Wikipedia, Wiktionary claim to be encyclopedic knowledge, but cannot provide informational support for in-depth research. This is a field of activity for specialists in specific areas of knowledge.

They do not reveal the essence and mathematical meanings of concepts, in contrast to a special tool – the thesaurus of the subject area, where emphasis is placed on semantics and the use of the term in linked publications. The originality of the proposed work is that the bibliographic base is associated with persons, and there is an opportunity for authors and experts to replenish information by linking it to existing terms and complement the list of terms, including in Russian.

The main condition for research information support is the provision of up-to-date information on achievements, confirmed by publications of professionals. Naturally, the mathematical resources on the Internet need such professional information support. Known databases of scientometric data of publishing houses partially perform this function in the section of actual publications. However, exponentially growing number of publications, it complicates the search, requires expensive and time consuming. In this case, every professional is interested in having a collection on a certain topic. It is possible to organize the creation of such a collection technologically by using thematic thesauri and the mechanism of their replenishment, as proposed in this paper. Feature of modern digital representation of the data makes it possible to move the centuries created thesauri and ontologies in thematic databases and thereby ensure their search for the completion and updating.

Another feature of mathematical subject domains is explained by the fact that mathematical statements in natural language are better expressed as mathematical equations. When building information images for branches of mathematical physics it is crucial to consider the listed aspects, specifically, rely on representative dictionaries which define terms and formulas as a background for information retrieval thesaurus for the subject domain.

In [12], a thesaurus was presented for the subject domain “ordinary differential equations” and now its extension to the domain of “partial differential equations” is being developed as part of a common mathematical resource on “equations of mathematical physics”. Numerous studies of specialists, such as V. A. Steklov [13], V. S. Vladimirov [14], R. Curant, D. Gilbert [15], A. N. Tihonov, A. A. Samarsky [19], A. G. Sveshnikov, A. N. Bogolyubov [20], M. M. Smirnov [16], A. V. Bitsadze [17], V. A. Ilyin, E. I. Moiseev [18] and other classics of mathematical analysis and differential calculus allow to establish paradigmatic connections of concepts and formulas to use them as lexico-semantic data arrays for presentation and search in mathematical information resources.

This paper describes the creation of the ontological model of a thesaurus for some problems of mathematical physics within the framework of the terminology of the semantic library LibMeta and its use in the tasks of searching and navigating through its resources. At the first stage, a series of related dictionaries for individual equations is combined into a thesaurus. It is incomplete and therefore a means is proposed for replenishing it with the inheritance of previous knowledge from available data from open sources. It turns out a non-standard resource, but it reflects the state of modern research.

2 Thesaurus Description

Mathematical physics deals with mathematical models of physical phenomena [3]. It relies on mathematical methods to build and study the models [3–5]. The methods of mathematical modelling enable us to solve mathematical problems applying equations of differential calculus [3], [4]. Each equation establishes a correlation between mathematical model and physical phenomena. The topics to be described are as follows: *problems of mathematical physics, modeling methods, equations, methods of solutions, solutions and their analysis*.

The thesaurus was formed by analyzing the original works of classics of mathematical analysis and differential calculus, and a representative list of articles was organized for that purpose. The problem of defining paradigmatic relations between definitions of certain fields in mathematical physics is brought to attention along with outlining the hierarchical relations between the terms that can be used when searching on the mathematical resources along with additional classification parameters set in secondary documents.

It is possible to study a streamlined scheme (Fig. 1) to offer a step-by-step description to problems of mathematical physics starting with the name of physical/technical process and ending with solutions to develop data layout in this domain.

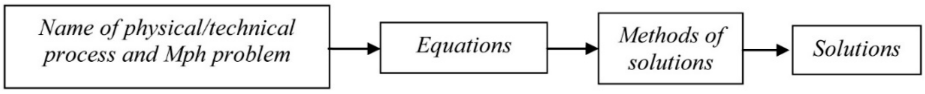

Fig. 1. The relation scheme in mathematical physics subject domain

Given that equations of mathematical physics, as a subject domain, cover a huge amount of research, the paper focuses on *physical processes identification*, as the pillar of mathematical models, and *partial differential equations* terminology with examples of *mixed type equations*.

2.1 MPh Problems

When describing the mathematical set of MPh problems, we consider it a hierarchy as provided in Fig. 2 that follow the logic of the domain. The graphical representation is one of the ways to describe hierarchical relation of problems in mathematical physics [6].

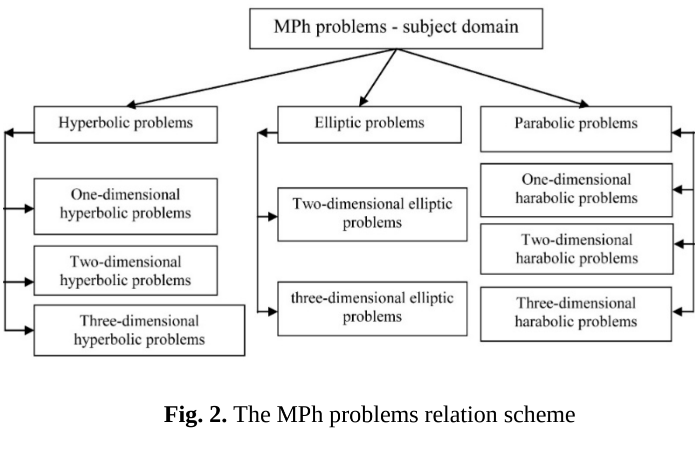

Fig. 2. The MPh problems relation scheme

Such structure provides a topic-related distribution within the section describing the problems of the MPh.

2.2 Partial Differential Equation

Let us note the features related to partial differential equations that should be added to the thesaurus:

- equation scope, as well as the material object of the physical process;
- summary on equation properties;
- researchers' surnames, authorship, named equations;
- specific and associative equation formulas;
- synonyms for the terms.

In different domains mixed type equations can be classified as hyperbolic, parabolic and elliptic [7]. Fig. 3 graphically shows hierarchical links of second order PDE with two independent variables.

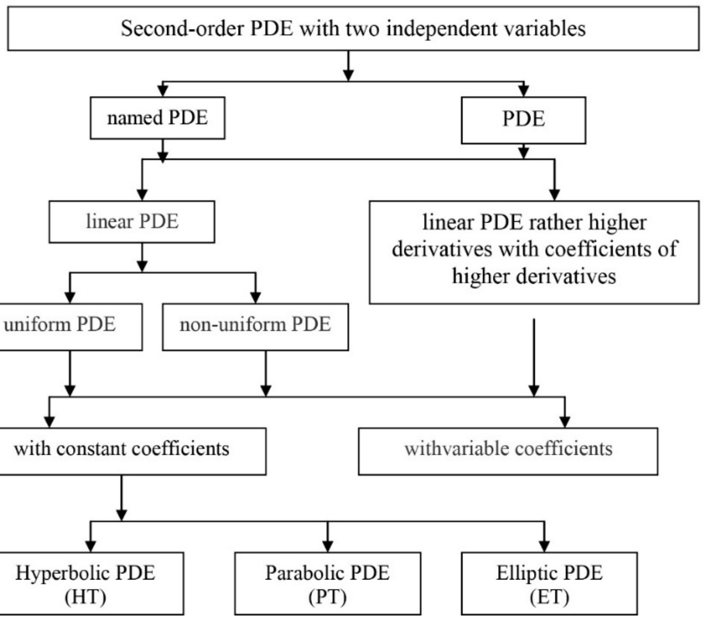

Fig. 3. Graphic chart representing linear/linear rather higher derivatives second order PDE with two independent variables

2.3 Thesaurus Structure

While developing the thesaurus one of the main objectives is to develop its structure considering the characteristics of the domain. The structure consists of thematic sections, sets of links between the elements of the thesaurus, the structure of the thesaurus articles.

Thus, the basic version of the thesaurus includes the following main thematic sections:

- problems of mathematical physics
- equations of mathematical physics,
  - partial differential equations,
  - equations of mixed type.

The analysis of the domain/subject area revealed the need to allocate the following categories for mentioned sections:

- Type of problem (elliptic, hyperbolic, parabolic);
- Dimension of the problem (one-dimensional, two-dimensional, three-dimensional);
- Type of equations (named, nominal);
- Homogeneity of equations (linear uniform, linear non-uniform, …);
- Types of equation coefficients (with variable coefficients, with constant coefficients, …);
- Types of equations (elliptic, hyperbolic, parabolic).

Based on these categories, the following link are established:

- Task type – task dimension;
- Type of equations – type of equations;
- Uniformity of equations – types of coefficients in equations.

The following thesaurus terms are also reflected:

- Hierarchical: genus, species;
- Horizontal: synonyms, associations.

In addition to the main term categories in the thesaurus, it is necessary to introduce additional categories that support generation of various links tied to objects that are not explicitly reflected in the thesaurus but are necessary for the completeness of description. Such objects include Authors and References. In order to implement these features, thesaurus conceptual structure provides relevant set of links to describe references, authors, etc.

- References – introduced to describe references to literature that contains in-depth information about a concept of the thesaurus;
- Author – introduced for designating the author and the author's term for a concept.

The mentioned hierarchically and horizontally linked categories form a conceptual model of the domain.

The conceptual model of the thesaurus is thus reflecting the following:

- means to define concepts;
- method for defining concept synonyms;
- a list of conceptual properties and attributes;

- object category;
- composition of objects of each category.

Thus, structurally, the concept of the thesaurus includes the following elements:
- alphanumeric code of the concept;
- concept descriptor;
- non-descriptor – concepts' synonyms;
- thematic section of the concept;
- symbolic representation of the concept formula;
- list of links to other concepts;
- text additions (comments, notes, help);
- a list of references for the concept;
- authors of the concept.

Given the essence of the structural description there is a need to include various lexico-semantic categories as follows:
- type of equation: one-dimensional, two-dimensional, three-dimensional;
- type of equation: hyperbolic, parabolic, elliptic;
- types of coefficients: variables, constants;
- etc.

Thesaurus Ontological Model

Basics of LibMeta Information Model

The paper develops the resource for the topic in math on "mathematical physics problems", index of "MPh mixed type equations" and its integration in LibMeta [8].

LibMeta is an information system that implements a set of features that are necessary to work with the content of a prospective semantic library. LibMeta is a special electronic library management system (ELMS). LibMeta library is a storage of structured and diverse data with the ability to integrate it with other data sources that meet the requirements for sources within Linked Open Data [21]. Also the possibility of specifying its content by defining the subject area is presented.

The versatility of the system's content is based on the set of concepts which represents the LibMeta informational content model: *information resource* and *information object* that define a resource instance. The *information resource* is the basic descriptive unit of library content, and the *information object* represents instance of *information resources*. Each of them has its own unique LOD identifier. In fact, the semantic meaning of the *information resource* is equivalent to the concept of *ontology class* with discrepancies in description. The structure of the description of information objects is determined by the concepts of an *attribute* and a *set of attributes* that are defined in the description of the corresponding resource. The *attribute* is an element of a resource property description, and the *set of attributes* is defined as a collection of different attributes. Attribute types are as follows: *file*, *object*, *numeric*, *text*, *string*.

The model of the thesaurus at full compliance with ISO 25964 [9]. The model described by this standard supports multilingual thesauri and other types of dictionaries.

The standard contains recommendations for establishing and maintaining mutual correspondence between several thesauri or between thesauruses and other types of dictionaries used in information retrieval. This standard is also compatible with the SKOS model, which offers a way to present thesauri in the Internet.

The standard suggest rationale for using the following concepts: hierarchical relationship, horizontal relationship, term, thesaurus, concept, thematic group, terms, descriptor (or preferred term for a concept), non-descriptor (a set of terms that are synonymous with a descriptor). Earlier in [10] we provide analysis of the basic entities of the ontology of the model, which forms the base of the ontological thesaurus model [10], presented in current publication.

The base model of thesaurus is designed such way that the concepts of this thesaurus are related to the concept of an information object from the LibMeta content description model and allow for concept association with any type of resource present in the library. Description of the ontological content model of the library allows you to describe additional types of resources like *Authors* and *References* and link them with the thesaurus to support them.

2.4 Basic Thesaurus Model Expansion

In order to add the basic thesaurus version description [10], namely the structure of the thesaurus concepts, the system supports a class hierarchy for additional concept attributes. It includes subclasses of the ResourceAttribute superclass that add the description of the concept structure that corresponds to a certain thesaurus with the following concept values:

- ThesaurusAttributeText – presented as a text;
- ThesaurusAttributeTaxonomy – presented in the form of item of a particular classifier or dictionary;
- ThesaurusAttributeString – presented as a string;
- ThesaurusAttributeObject – presented as a certain content library information object;
- ThesaurusAttributeNumber –presented as a number;
- ThesaurusAttributeHref – presented as a link;
- ThesaurusAttributeFile – presented as a file;
- ThesaurusAttributeConcept – presented as other thesaurus concepts (it defines relationships between concepts implicitly supported in the system).

Each of these classes is compliant with OWL-supported inheritance paradigm, that contains the properties ascribed to the ThesaurusAttribute superclass.

The ThesaurusAttributeSet class contains the thesaurusAttributes property, which in turn contains many instances of the above listed classes that define the *Concept* class structure of a thesaurus.

The *Thesaurus* class is linked to ThesaurusAttributeSet through properties mediator thesaurusAttributeSet.

Such standardized modeling allows LibMeta to easily adjust the system to any subject area.

The description of the thesaurus “The MPh Mixed Type Problems” based on initial ontology version can be broadened by terms of an extended model, in order to further extend the article structure of this thesaurus, provided the following attributes are added:

- *comment;*
- *note;*
- *help;*
- *references;*
- *authors.*

2.5 Three Levels of Thesaurus Model

The *Comment* and *Note* attributes represent **ThesaurusAttributeText** class attributes, with Help being **ThesaurusAttributeString** class attribute, while References and Authors are instances of the **ThesaurusAttributeObject** class. Combined they are a set of thesaurus attributes.

Next, we analyze a three-level representation of the subject area thesaurus within the LibMeta library.

In order to be able to use the thesaurus of a specific subject area, the following sequence should be followed when constructing a semantic library within LibMeta system.

1. Based on the introduced model, a set of information resources used in the library is given. It is necessary to provide descriptions of the content of the future library in terms of the proposed model.
2. The structure of the thesaurus is finally set up. On the basis of certain classes the respective links between terms are set, the term description is expanded if necessary, the links with the system resources are determined as well.
3. According to the definition of collections, a module is implemented, within which collections are created and filled.

After completing these steps, we form a domain information model described in terms of the ontology of the semantic library introduced above. At the same time, if the newly introduced concepts are instances of the designated resources at the first level, then when filling the library, we use them as classes to describe the data. Dividing instances into classes is *metamodeling*. Even though the semantics of OWL 2 ontologies that is used to describe ontologies does not allow such metamodeling, this language limitation is bypassed with a syntax trick known as *punning*. This means that when an instance identifier is present in a class axiom, it is seen as a class, and when the same identifier seen in a separate statement, it is treated as an instance.

While describing a specific subject area in terms of the proposed semantic library ontology, we, in fact, construct a three-level ontology, in which instances of the first level are high-level concepts, with the second level containing concepts of a specific subject area. When uploading data to the ontology we use the first level terms to define the third level classes.

3 Searching for Mathematical Publications through Thesaurus Links

The use of mathematical formulas is novelty, yet due to the recent software development progress they see their use. The LibMeta system [8, 10] implements a comparison of formulas starting with its denotation. This renders possible to include symbolic expressions in search queries.

Using the domain/subject area of mathematical physics and related fields as an example, we can see how expanding a thesaurus-based query can improve search results. Let us consider Tricomi Equation as an example and highlight the advantages of adjusting the query with formulas.

The concept of "Tricomi Equation" is associated with synonymously tied to similar equations of different kind. For each of these equations there is a symbolic notation and a "Tricomi Equation" TEX formula notation. The formulas in this example are actually synonymous. For each of the formulas are also references to the source materials with a specific mathematical record. These "Tricomi Equations" are used for indexing and searching publications in databases on mathematical subject areas. Thus, if you select one of the entries you will find works about the "Tricomi Equation" for the other variants of different scientific schools and paradigms.

If necessary, you can expand the references section and expand the LibMeta library thesaurus. The formulae-synonyms links will provide references and respective authors data, search for which in this case was not directly undergone. Thus, the process of updating the thesaurus for the subject area is realized through the links. As a result, the LibMeta library will have new data on publications, and a user will receive a new list of publications for the "Tricomi Equations". By requesting this topic, the user will also receive complete information about the semantic links for the searched formula, which will include links to the formula-synonyms, which is especially important for the experts. Thus, we are talking about how far the search result corresponds to the information need of the user (pertinence property). The scope of the search and query can be quite narrow, but can be supplemented with a formula and the result will be complete in the sense of pertinence. This is one of the necessary properties of the usefulness of the information system and the success of utilizing it, which was noted by the founder of the term "information retrieval" Calvin Mooers [11].

4 Conclusion

Specific developments in the mathematical subject areas representation still in the focus modern interdisciplinary research papers. The proposed information resource implements the combining specified semantic relations, symbolic language of formulas and systematic representation of mathematical resources (publications with secondary information, indexed in accordance with the thesaurus) in the digital library. This approach corresponds to the current trends in the development of information technology and allows for expansion of the mathematical ontological domain in the digital space. The choice of the subject area with the lack of representation in literature and descriptor dictionaries, and the digital library approach renders this study relevant. Also a tool is being created for the formation of Russian-language content on this topic. The discovered solution for mixed-type problems does not limit the other MPh problems ontology generality. The functions of the digital library LibMeta allows for updating this subject area as the development of relevant resources (thesaurus, dictionaries, term lists) goes on.

It takes into account all the properties of a mathematical text and a combination of symbolic (formula) and textual information. And this is the main feature of the proposed ontological model of a thesaurus for equations of mathematical physics. As a result, all properties of a mathematical text are taken into account when searching and indexing publications. Now the thesaurus of mathematical physics is presented in terms of the ontology LibMeta includes about 100 concepts and more than 200 terms. The ontology language is OWL, the syntax is RDF/XML.

Acknowledgements

The study was supported by Russian Foundation for Basic Research, projects Nos 17-07-00217, 18-00-00297, 17-07-00214.

References

1. Bond, Ph.: The Era of Mathematics. An Independent Review of Knowledge Exchange in the Mathematical Sciences. Professor Philip Bond. https://epsrc.ukri.org/newsevents/pubs/era-of-maths/, 02.12.2018
2. Mayans, R.: The Future of Mathematical Text: A Proposal for a New Internet Hypertext for Mathematics. Journal of Digital Information [S.l.] **5** (1) (2006). https://journals.tdl.org/jodi/index.php/jodi/article/view/128/126, 02.12.2018
3. Mathematical encyclopedia. Chief Editor. I.M. Vinogradov. Moscow: Soviet encyclopedia. 1979. 1104 p.
4. Tihonov, A.N. and Samarskij, A.A. Equations of mathematical physics. M.: Publisher Nauka, 1972. 736 p. (ru)
5. Sveshnikov, A.G., Bogolyubov, A.N., and Kravcov, V.V.: Lectures on mathematical physics: Textbook. Moscow: Publisher MSU, 1993. 352 p. (ru).
6. Moiseev, E.I., Muromskij, A.A., and Tuchkova, N.P.: Internet and mathematical knowledge: representation of equations of mathematical physics in the information retrieval environment. M: Publisher MAKS Press, 2008. 80 p. (ru).
7. Smirnov, M.M.: Equations of mixed type. M: Publisher Nauka, 1970. 296 p. (ru).
8. Serebryakov, V.A. and Ataeva, O.M.: Information model of the open personal semantic library LibMeta. Proceedings of the XVIII Russian Scientific Conference "Scientific Service on the Internet". Novorossijsk, 19–24 September 2016. Keldysh Institute of Applied Mathematics (Russian Academy of Sciences). P. 304–313 (ru).
9. ISO 25964 thesaurus schemas. http://www.niso.org/schemas/iso25964
10. Ataeva, O.M. and Serebryakov, V.A.: Ontology of the digital semantic library LibMeta. Informatics and its applications **12**, 2–10 (2018) (ru).
11. Muromskij, A.A. and Tuchkova, N.P.: On the ontology of the addressee in the mathematical subject field. Electronic Libraries **21** (6), 506–533 (2018) (ru).

12. Moiseev E.I., Muromskij A.A., Tuchkova N.P. Thesaurus information retrieval in the subject area: ordinary differential equations. M.: Publisher MAKS Press, 2005. 116 p. (ru)
13. Steklov, V.A.: The main tasks of mathematical physics. Moscow: Publisher Nauka, 1983. 433 p. (ru).
14. Vladimirov, V.S.: Equations of mathematical physics. Ed 4. Moscow: Publisher Nauka, Main editorial office of the physical and mathematical literature, 1981. 512 p. (ru).
15. Kurant, R. and Gil'bert, D.: Methods of mathematical physics, v. I, II, M.: Publisher Gostekhizdat, 1951 (ru).
16. Smirnov, M.M. Equations of mixed type. Moscow: Publisher Nauka, 1970. 296 p. (ru).
17. Bicadze, A.V.: Equations of mathematical physics. Moscow: Publisher Nauka, 1982. 336 p. (ru).
18. Il'in, V.A. and Moiseev, E.I.: Nonlocal boundary-value problem of the second kind for the Sturm–Liouville operator. Differ. Equations **23** (8), 1422–1431 (1987) (ru).
19. Tihonov, A.N. and Samarskij, A.A.: Equations of mathematical physics. Moscow: Publisher Nauka, 1972. 736 p. (ru).
20. Sveshnikov, A.G., Bogolyubov, A.N., and Kravcov, V.V.: Lectures on mathematical physics: Textbook. Moscow: Publisher MSU, 1993. 352 p. (ru).
21. Bizer, C., Heath, T., and Berners-Lee, T.: Linked data: The story so far. Semantic services, interoperability and web applications: emerging concepts. IGI Global, 205–227 (2011).