

UDC 004.8, 004.89

Adaptation of the language model for mathematical texts in the semantic library

Ataeva O.M. (FRC 'Computer Science and Control' RAS),

Tuchkova N.P. (FRC 'Computer Science and Control' RAS)

The paper studies the approach of LLM adaptation for queries in the mathematical subject area. The subject area is presented as an ontology of a semantic library LibMeta, where data navigation is carried out using KG *MathSemanticLib*. The descriptions of the mathematical subject area are based on mathematical encyclopedias of the Soviet and Russian mathematical schools, and the filling of the LibMeta subject area library is carried out by integrating subject areas of specialized mathematical journals. A procedure for integrating LLM and KG *MathSemanticLib* is proposed. It is shown that as a result of this approach, LLM does not go beyond the subject area, which allows us to state a more relevant answer to the query.

Keywords: *library of subject areas, large language model, knowledge graph, automation of access to scientific information, integration of knowledge graph and language model, systems of information support of scientific research, mathematical subject area, industrial engineering, ontological design*

1. Introduction

The fantastic growth of artificial intelligence (AI) methods for communicating with information systems in natural language over the past few years has led to the fact that publications on the topic of large language models (LLM) become obsolete in a year or two. Information technology has probably never experienced such speed and competition. Users are offered various software tools, in particular ChatGPT, which promise answers and predictions on all sorts of topics. A review of LLMs [15] demonstrates the various approaches used in creating these tools, and it can be concluded that the directions laid down at the beginning of the development of AI [19] continue to be implemented on new computing platforms. The authors [15] note such directions as statistical language models, neural language models, pre-trained language models and LLMs and the problems facing developers.

A well-known problem with using LLM is the difficulty of explaining and verifying the conclusions, since the answer does not indicate the sources on the basis of which the answer was formed [13]. As a rule, such sources cannot be indicated even in principle, due to the complexity

(closed nature) of algorithms for processing large amounts of data. This is especially important when it comes to scientific knowledge, especially mathematical knowledge, which is needed in a wide range of classical and applied problems. In mathematical subject areas, it is important to rely on verified sources, to distinguish LLM hallucinations from true search results.

As a rule, for scientific fields (and mathematics, of course), it is necessary to analyze a specific collection of articles, including archived ones, full texts of which are not in the public domain. These articles need to be collected in a digital library, processed, and only then can the result come to the attention of LLM. To extract knowledge from these texts, previously not found in the public domain, it is necessary to provide them for LLM training in a new subject area, i.e. to compile a corpus of articles and describe this set semantically. This endless process is still relevant, since new subject areas appear, and new interdisciplinary studies with new terminology are added to traditional ones, which means that language models need to be adapted.

One of the solutions to improve LLM response on scientific texts is the integration of LLM and the knowledge graph (KG) of the subject area [1]. In this paper, the problem of LLM adaptation is proposed to be solved by using the KG *MathSemanticLib* of the semantic library LibMeta [2], bypassing which, LLM extracts the answer from the subject area of mathematics and its application. The result is achieved due to the fact that the KG represents structured data, relies on the ontology and thesaurus of the subject area.

The structure of the article is as follows: introduction, related works, data model of the KG *MathSemanticLib*, supervised knowledge extraction example LLM answer and conclusion.

2. Related works

The issues of joint consideration of the problems of constructing KG and LLM arose naturally, as a continuation of the ideas of providing access to knowledge as structured data. The Awesome-LLM-KG (<https://github.com/RManLuo/Awesome-LLM-KG>) page presents a collection of links to papers and resources about unifying LLMs and KGs. It graphically displays the advantages and disadvantages of LLMs and KGs in the context of their mutual complementarity.

The main idea is that the KG structure contributes to improving LLM reasoning, and the linguistic capabilities and generalizations of LLMs improve the understanding of the essence of knowledge in the KG. Awesome-LLM-KG also provides generalizations of research directions and applications of the results of unifying LLMs and KGs in searching, building dialog systems and AI assistants, and research methods. The synergetic nature of unifying LLMs and KGs is separately noted, which is based on the mutual enrichment of LLMs and KGs when they are combined.

In the work [17] a cyclic procedure of integration of the domain knowledge and LLM is considered, as a result of which the LLM response and the domain knowledge itself are corrected, which is closest to the idea of our research.

The authors [14] provide an overview of the weaknesses of LLM related to hallucinations. The authors see an improvement in the quality of inference in the use of KG for training LLM, but they note that this process is quite complex, since first a full KG must be constructed, and then LLM reasoning with graph constraints. The work [14] proposes a procedure for transforming KG for further traversal into LLM and for generating correct reasoning paths. The option of creating a semantic description of mathematical concepts from school to university is considered in the work [5]. Here [5] provides an overview of research related to the attempt to reflect the process of cognition of mathematical subject areas and their reflection in digital resources. This idea itself has haunted the scientific community, starting with the GDML project [10]. The research [5] uses the GloVe (<https://nlp.stanford.edu/projects/glove/>) algorithm on a large corpus of mathematical texts to identify the frequency of use of terms and their relationships.

The comparison is made between Wikipedia terms in French and their English translations, and the use of words from the dictionary by participants (<https://osf.io/dxg2w>) with different mathematical backgrounds. In this way, the GloVe algorithms were tested and a relatively good correspondence between the GloVe [18] vectors and human judgments was established. This study [6] is important for our discussions in terms of the participation of experts in assessing the results of the semantic representation of subject areas and the reflection of these representations in the processes of cognition of mathematical areas. Like the authors of [6] (<https://osf.io/dxg2w>) [6], we use the opinion of experts, but when creating semantic images of subject areas, we rely on classical sources such as encyclopedias and monographs.

The research [7] is devoted to the description of subject areas. The data are given on how individual examples trained on corpora of specific subject areas achieve good results. However, it is noted that this is not enough to create a general approach for different subject areas. The authors propose the KnowledgeDA tool, a unified domain language model development service that can automatically generate a domain language model by performing three steps: (i) localize domain knowledge entities in texts using an embedding-similarity approach; (ii) generate enriched samples by extracting exchangeable pairs of domain entities from two representations of both the knowledge graph and the training data; (iii) select high-quality enriched samples for fine-tuning using confidence-based scoring.

A KnowledgeDA prototype for learning language models for two domains: healthcare and software development. This example of creating text corpora by subject area is quite problematic to

extend to mathematical subject areas, since the original sources may differ radically in structure and presentation features (for example, the presence of formulas changes the process of text preprocessing).

The work [11-12] is devoted to training the SciBERT (<https://github.com/allenai/scibert/>) model on scientific texts, where the possibilities of improving BERT after unsupervised pretraining on a large multi-domain corpus of scientific publications are demonstrated. The BERT model architecture [6] is based on a multilayer bidirectional Transformer [20] is used. Our work presents a technology for constructing KG, starting from arrays of texts of scientific mathematical and interdisciplinary journals, to the integration of the obtained KG *MathSemanticLib* with LLM in the journal recommendation system in the environment of the semantic library LibMeta [4].

3. Data Model of the KG *MathSemanticLib*

A key word is a word or a phrase. The key words are separated by commas. The number of key words is not limited. The key words are used to improve the search quality on the Journal website.

The approach used in this work is that first an ontology and thesaurus of the subject area are built, then a KG based on the ontology, and then LLM is used for communication in the library. The data structure and ontology model of the LibMeta library for the KG *MathSemanticLib* are described in the works [3], [4], here we will note only some of their properties, namely: integration of various sources (encyclopedias, monographs, journals, classifiers, thesauri, dictionaries, formulas) based on the ontology; construction of a KG based on the ontology; use of a KG for organizing a dialogue in the library. Thesauri contain the main terms of LibMeta subject areas, linked by hierarchical and horizontal relationships. The data model in LibMeta is an ontology in OWL (which is represented as an RDF graph). Filling the library is a process of completing the ontology by integrating data in accordance with their descriptions and metadata. The subject area is defined by forming a thematic subspace in the library ontology and establishing semantic links with the basic content of the library [1-3].

The mathematical encyclopedia [8], [21], the encyclopedia of mathematical physics [9], the thesaurus of ordinary differential equations, the dictionary of special functions of mathematical physics and other Russian-language sources and components of the library [3] are used as external basic taxonomies with which publications are linked. The creation and development of the LibMeta library [4] is based on the integration of mathematical knowledge, both in the retrospective and prospective direction, by adding publications from various new subject areas of mathematics, related sciences and applications.

3.1. LibMeta Ontology

The LibMeta digital library ontology defines the data structure. The concepts that make up the LibMeta ontology are conventionally divided into concepts intended for:

- describing the content of a subject area;
- forming a thesaurus of any subject area;
- describing thematic collections;
- describing the task of integrating library content with source data from LOD.

Semantically significant connections are defined between these groups of concepts. The following formal definitions are used to describe the ontology:

Definition 1. Library thesaurus $TH = \{T, R\}$, where T are terms and R are the relationships between them.

Definition 2. Library content $C = \{IR, A, IO\}$, where IR are types of information resources, a set of attributes $A\{a_i\}$, information objects $\{IO\}$.

Definition 3. Semantic labels $M = \{m_i\}$ of an information object are terms that are not included in the thesaurus, but are necessary for thematic division of information objects IO within the subject area.

Definition 4. Semantically significant relationships of the library $P = \{P_i\}$ are the following main relationships:

$P1(t, io)$ thesaurus term $\rightarrow$ information object;

$P2(io, t)$ information object $\rightarrow$ thesaurus term;

$P3(r, s)$ information resource $\rightarrow$ class of source objects, where information resource is a general definition for information objects stored in the system; thus, in fact, information objects are instances of information resources;

$P4(a, sa)$ information resource attribute $\rightarrow$ property of source class;

$P5(io, os)$ information object $\rightarrow$ instance of class from data source;

$P6(m, io)$ semantic label $\rightarrow$ information object;

$P7(io, m)$ information object $\rightarrow$ semantic label.

In fact, the concepts are divided into three categories: the first includes definitions of the concepts of the semantic library content, the second category refers to the definition of concepts necessary to support terms in the thesaurus of the subject area, and the third includes definitions necessary to define the processes of integrating the content of these resources. Based on these definitions, the main processes are described, such as, for example, integrating data from different

sources, categorization/classification, mapping different models of source data to a given subject area, constructing equivalence classes, etc. Fig. 1 shows screenshot of LibMeta ODE thesaurus.

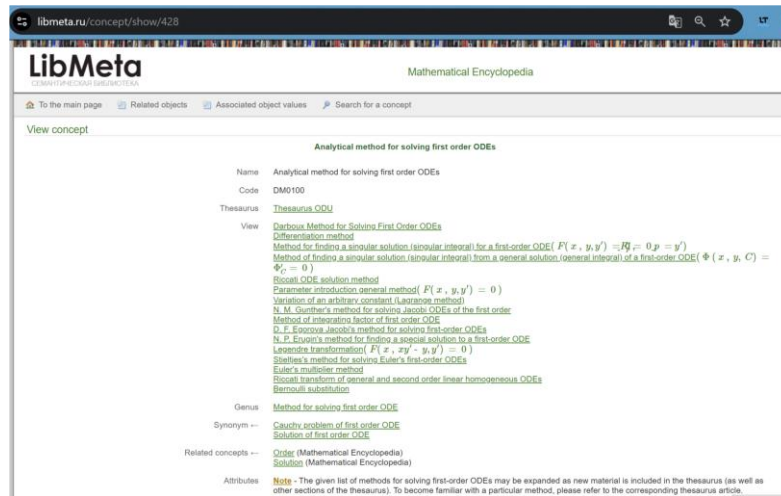


Fig. 1: An example of the ODE thesaurus concept

3.2. Completing the ontology and integrating data

The task of adding new terms (new for this ontology) and links to the ontology arises during integration with new sources (publication arrays). These are, as a rule, terms from new tasks or applications in interdisciplinary research. Integration of new data into LibMeta is implemented by completing the ontology. When integrate publications in a semantic library and KG, they must undergo preliminary processing, diagram Fig. 2.

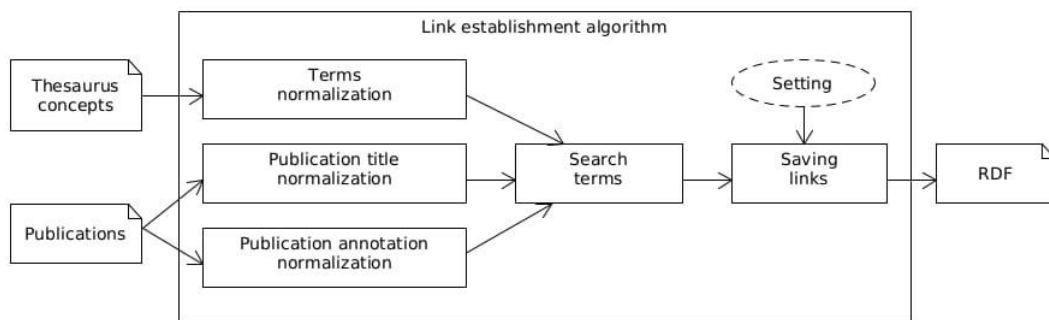


Fig. 2: Stages of publication preprocessing for ontology.

Preprocessing options depend on the source data and may vary depending on the degree of structuring of the articles. Characteristic structures are known for mathematical articles, but it is necessary to identify the main terms and links. At the Link Extraction stage (Fig. 2), semantically significant links of the library $P = \{P_i, i = 1, \dots, 7\}$ are identified. If preprocessing has shown the presence of signs of belonging of the data to a certain subject area, then the publications are placed in the ontology and thesaurus of the subject area.

The task of adding new terms (new for this LibMeta ontology) and links to the ontology arises during integration with new sources (publication arrays). These are, as a rule, terms from new tasks or applications in interdisciplinary research. One of such typical examples is applications in equations of mathematical physics. Fig. 3 and Fig. 4 show the scheme and screenshot of adding terms from the journal MKMK [4], thanks to which a new subject area 'elasticity theory', was completed and integrated into the ontology.

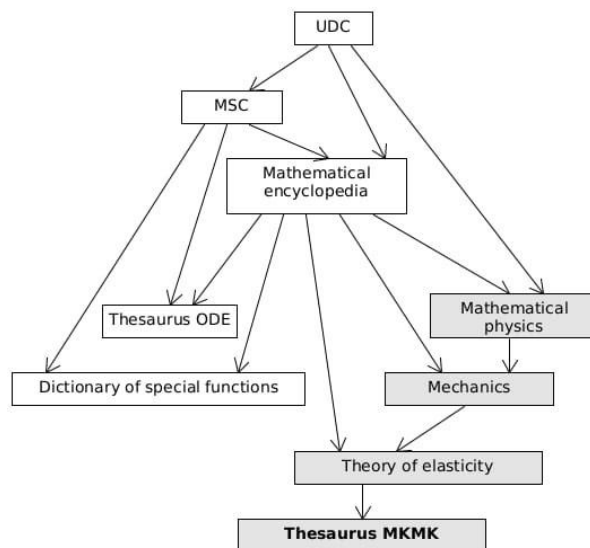


Figure 3: Connection diagram when adding a new subject area to the LibMeta ontology.

Completing the ontology, following the logic of the library data structure, affects the KG *MathSemanticLib*. Since the KG reflects the ontology connections, then when a new subject area appears in the KG of *MathSemanticLib*, a 'subgraph' appears.

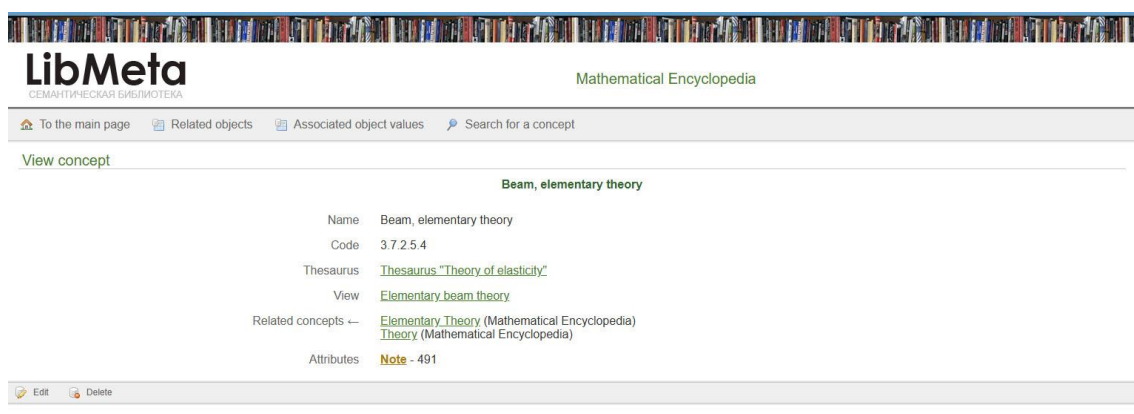


Fig. 4: Example of a thesaurus concept for 'elasticity theory'

3.3. KG *MathSemanticLib*

The LibMeta digital library ontology defines the structure of the library data. Each data element loaded into the library can be associated with an ontology node, which defines the position of the data element in the ontology. Based on the ontology links and the links defined at the design stage, a graph can be constructed. The subject area data can thus be represented as a KG, the structure of which is defined by the ontology, nodes (articles, terms, formulas) are instances of ontology elements, links are links of the subject area thesaurus. This is shown schematically as a three-level ontology in Fig. 5.

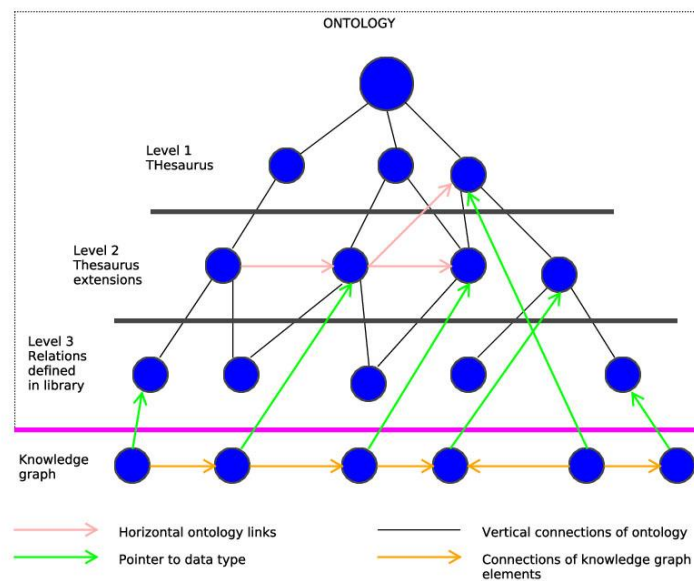


Fig. 5: Scheme of three levels of the LibMeta library ontology.

The construction of KG *MathSemanticLib* can be described in two global stages. At the first stage, a 'zero' version of KG is constructed from some source, and at the 'second' stage, the integration of the graph of incoming data with the general graph of the library occurs by establishing links with the thesaurus [4]. The 'zero' version of the graph KG *MathSemanticLib* is the KG of the mathematical encyclopedia [8], [21], and the 'second' stage is the integration of an array of scientific articles. When completing the ontology, KG is also completed, that is, the 'second' stage is each subsequent stage.

The main stages of data processing for KG are closely related to the sources from which the data comes. Often the data is presented in an unstructured or semi-structured form. In our case, we consider, among other things, unstructured texts of Russian-language scientific articles. Nodes can be larger ontology objects Fig. 5, or objects - publication, term, person, formula.

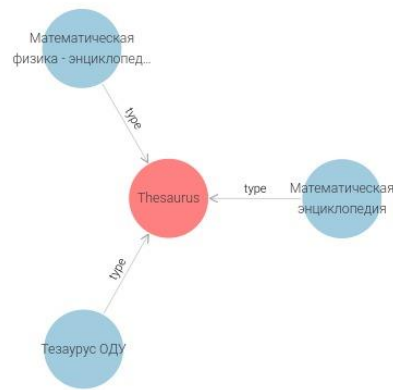


Fig. 6: The central node of the *MathSemanticLib* thesaurus.

3.4. Formulas in KG *MathSemanticLib*

The use of formula language in mathematical subject areas is a natural stage in modern dialog programs. One of the most authoritative databases of scientific publications zbMATH (<https://zbmath.org>) has long allowed formula entry in the search line. Nevertheless, search by formulas remains one of the tasks in information technology, since it is important to consider the meaning of their use in a scientific publication.

The LibMeta library accumulates and integrates formulas from the mathematical encyclopedia [8], [21], encyclopedia of mathematical physics [9], the thesaurus of ordinary differential equations [4], [16] dictionary of special functions and others into the ontology. The peculiarity of this integration is that the formula is saved with the context, and thus, a dictionary of formulas with links is formed, that is, the semantic image of the formula is saved. This approach allows you to include symbolic expressions in search queries. Since the formula is a semantic object in LibMeta, it can also be a KG node, for example, Fig. 6, which shows the links of the formula with other KG *MathSemanticLib* objects of the LibMeta library.

There are several types of work with formulas in the LibMeta library:

- selection of basic formulas related to concepts;
- selection of formulas related to the concept;
- selection of all formula links.

4. Supervised Knowledge Extraction Example LLM Answer

Diving into the terminology of the subject area means accepting some non-standard terms established by specialists. For mathematical subject areas, it is important to reflect the meaning of

formulas and terms when describing subject areas in digital libraries. Fig. 7–8 shows fragments of the KG *MathSemanticLib* for a concept, formula, classifiers.



Fig. 7: Fragment of the subgraph of the concept 'Lame Equation', KG *MathSemanticLib*.

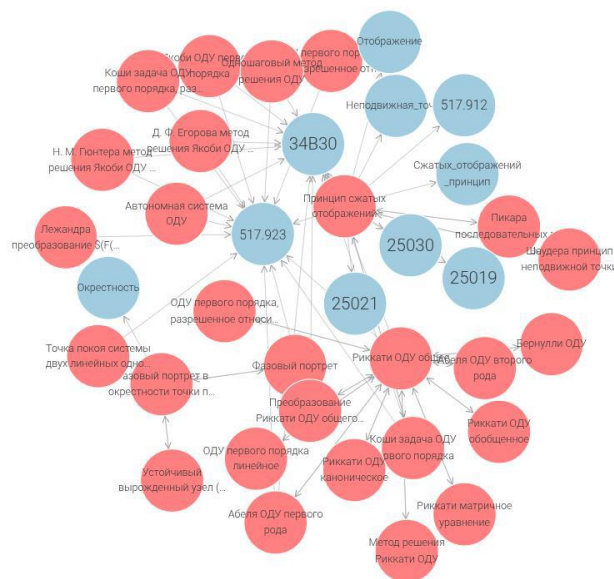


Fig. 8: Subgraph fragment with MSC codes –nodes, KG *MathSemanticLib*.

Below is a comparison of responses from some language models.

4.1. Examples with ChatGpt, YandexGPT, LLAMA

ChatGpt 4 and YandexGPT 4 Pro (<https://console.yandex.cloud/>) coped with the query 'Lame equation', but gave only general information without links and formulas with an error, Perplexity and LLAMA (Large Language Model Meta AI <https://app.hyperbolic.xyz/models/llama-3-3-70b-instruct>) did not cope at all.

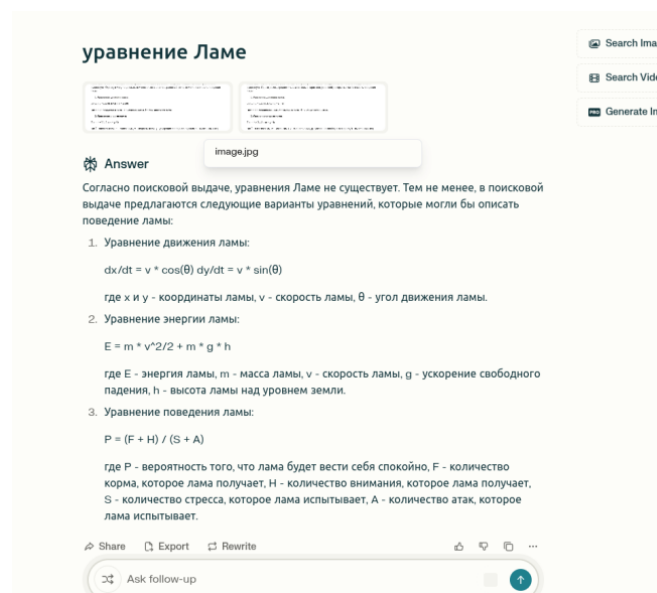


Fig. 9: Perplexity answers to the query 'Lame equation'.

The language model + knowledge graph allow using the generalizing capabilities of the language model to explore the domain graph in natural language. Questions formulated in natural language are transformed into SPARQL queries based on the ontological model of the graph. The model's answers are given based on the mathematical encyclopedia articles and their connections Fig. 10.

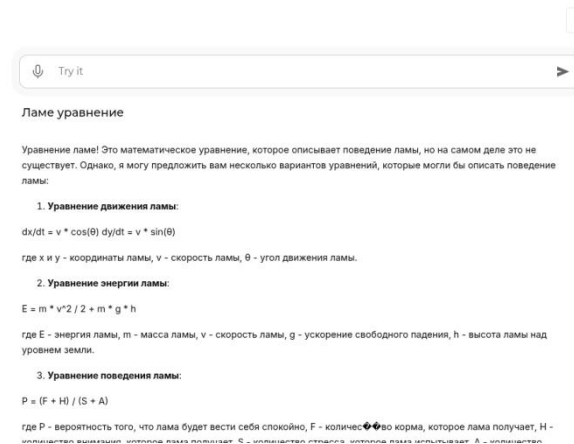


Fig. 10: LLAMA Answers to the Query 'Lame Equation'.

The model Gemma2-9b-It was used with following 3 questions. Gemma is a family of lightweight, state-of-the-art open models from Google, built from the same research and technology used to create the Gemini models. They are text-to-text, decoder-only large language models, available in English, with open weights for both pre-trained variants and instructiontuned variants. Gemma models are well-suited for a variety of text generation tasks, including question answering,

summarization, and reasoning. Their relatively small size makes it possible to deploy them in environments with limited resources such as a laptop, desktop or your own cloud infrastructure, democratizing access to state of the art AI models and helping foster innovation for everyone (<https://huggingface.co/google/gemma-2-9b-it>).

Question 1: How many publications?

- Generated SPARQL:

SE

LECT (COUNT(*) AS ?publications)

Question 2: How many concepts?

- SELECT (COUNT(DISTINCT ?concept) AS ?count)

WHERE

?concept a :Concept .

Question 3: How many concepts are connected with Lame formula?

- SELECT (COUNT(DISTINCT ?concept) AS ?count)

WHERE

?formula a :LameFormula .

?concept a :Concept .

?formula :connectedTo ?concept

4.2. Stages of generating a SPARQL query and response to it using a GPT like model

Stage 1. Translation of text into SPARQL query.

General format of the prompt for a model to generate a SPARQL query includes:

Question (prompt): in natural language

Scheme: The graph data model is specified as an rdfs scheme, which is annotated with the rdfs:label and rdfs:comment properties, which contain a natural language description of the classes and properties of the ontology

Instructions:

- Generate a SPARQL SELECT query to query a graph database using the following ontology schema in Turtle format: {schema}
- Use only the classes and properties specified in the schema.
- Do not include any that are not explicitly provided.
- Ensure that all required prefixes are included.

- Output only the SPARQL query without any backticks or additional text.
- The question is: {prompt}

Example 1.

Question (prompt): in natural language – How many concepts are there?

Scheme (fragment):

....

<http://libmeta.ru/thesaurus/concept/DE0002>

<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>

<http://libmeta.ru/Concept> .

<http://libmeta.ru/thesaurus/ODU>

<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>

<http://libmeta.ru/Thesaurus>

Answer:

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

SELECT (COUNT(?concept) AS ?conceptCount)

WHERE

?concept rdf:type <http://libmeta.ru/Concept> .

Stage 2. Graph query

Next comes the step of sending the generated query through the SPARQL access point and receiving the response. For the query specified in the example, the response comes in the form 'conceptCount': 'datatype': 'http://www.w3.org/2001/XMLSchema#integer', 'type': 'literal', 'value': '1221'

Stage 3. Translate the answer into human readable format

General format of the prompt to translate the response into humanreadable format includes:

Question (prompt): in natural language

Scheme: The graph data model is specified as an rdfs scheme, which is annotated with the rdfs:label and rdfs:comment properties, which contain a natural language description of the classes and properties of the ontology

Instructions:

- Create a natural language response based solely on the results of a SPARQL query.
- You are an assistant who produces clear, human-friendly answers.

- Use only the provided information to build your response.
- This information is authoritative
- Do not question or alter it with your own knowledge.
- Ensure your answer reflects an AI assistant's tone without adding any extra details.
- If no information is available, simply state that you don't know.
- Schema: {schema}
- Context: {context}
- Question: {prompt}

Example 2.

Question (prompt): in natural language – How many concepts are there?

Scheme (fragment):

<http://libmeta.ru/thesaurus/concept/DE0002>

<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>

<http://libmeta.ru/Concept> .

<http://libmeta.ru/thesaurus/ODU>

<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>

<http://libmeta.ru/Thesaurus>. . . . Context:

{'conceptCount': {'datatype': 'http://www.w3.org/2001/XMLSchema#integer', 'type':

'literal',

'value': '1221'}}}

Answer:

There are 1,221 concepts in total.

Stage 4. Correction of the query

If an invalid request was generated in response to the first step, an attempt is made to correct it using the following instructions:

Question (prompt): in natural language

Scheme (fragment):

<http://libmeta.ru/thesaurus/concept/DE0002>

<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>

<http://libmeta.ru/Concept> .

<http://libmeta.ru/thesaurus/ODU>

<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>

<http://libmeta.ru/Thesaurus> **Error:**

Error text

SPARQL

wrong query

Instruction

- The SPARQL query below
{generated sparql is invalid.
- The error is: {error message}
- Provide a corrected version of the SPARQL query without altering its logic.
- Include only the corrected query.
- No explanations or additional text.
- Do not enclose the query in backticks.
- The ontology schema in Turtle format is provided below: {schema}

libmeta.ru/Concept .

<http://libmeta.ru/thesaurus/ODU>

<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>

<http://libmeta.ru/Thesaurus> .

Error:

400: Error MALFORMED QUERY: QName 'lbm:ThesaurusConcept' uses an undefined

prefix

SPARQL

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

SELECT (COUNT(?concept) AS ?conceptCount)

WHERE

?concept rdf:type lbm:ThesaurusConcept .

Answer

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

PREFIX lbm: <http://libmeta.ru/>

SELECT (COUNT(?concept) AS ?conceptCount)

WHERE

?concept rdf:type lbm:Concept .

5. Conclusion

The proposed approach to adapting LLM to a specific mathematical direction of scientific Russian-language journals has been tested in the semantic electronic library LibMeta. The experience of integrating LLM and KG *MathSemanticLib* allows us to conclude that it is possible to organize a user dialogue with the library and create a digital assistant with the functions of a reader, author, editor, and reviewer of the journal.

It should be noted that when completing the LibMeta digital library ontology, it will be possible to ensure further integration of LLM and the KG *MathSemanticLib*, which means that the capabilities of a controlled Russian-language dialogue on mathematical topics will develop.

At the same time, the ability to check the LLM output within the LibMeta content and external sources, which will continue to be integrated, will be preserved. Further research will continue in the direction of developing dialogue and recommender systems.

References

1. Ataeva O., Serebryakov V., Tuchkova N. Development of the semantic space 'Mathematics' by integrating a subspace of its applied area // *Lobachevskii J. of Mathematics*. 2022. Vol. 43, No. 12. P. 29-40.
2. Ataeva O., Serebryakov V., Tuchkova N., Ontological approach to a knowledge graph construction in a semantic library // *Lobachevskii J. of Mathematics*. 2023. Vol. 44, No. 6. P. 2229–2239. <https://doi.org/10.1134/S1995080223060471>
3. Ataeva O., Serebryakov V., and Tuchkova N. From Texts to Knowledge Graph in the Semantic Library LibMeta. // *Lobachevskii J. of Mathematics*. 2024. Vol. 45, P. 2211–2219. <https://doi.org/10.1134/S1995080224602625>
4. Ataeva O., Serebryakov V., and Tuchkova N. Ontology-Driven Knowledge Graph Construction in the Mathematics Semantic Library. *Pattern Recognition and Image Analysis*, 2024, Vol. 34, No. 3, P. 451–458. <https://doi.org/10.1134/S1054661824700196>
5. Dehaene S., Debray S. Mapping and modeling the semantic space of math concepts. *Cognition*, 254:1–8, 2025.
6. Devlin J., Chang M.-W., Lee K., and Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. In *NAACL-HLT*. 2019.
7. Ding R., Han X., and Wang L. A Unified Knowledge Graph Augmentation Service for Boosting Domain-specific NLP Tasks. In *Findings of the Association for Computational Linguistics: ACL 2023*. P. 353–369, Toronto, Canada. Association for Computational Linguistics. <https://aclanthology.org/2023.findings-acl.24/>

8. Encyclopedia of Mathematics. https://www.encyclopediaofmath.org/index.php/Main_Page [Электронный ресурс] (дата обращения: 30.06.2025)
9. Faddeev L. Matematicheskaya fizika. Enciklopediya, Vol. 1. Bol'shaya Rossiyskaya enciklopediya, 1998.
10. Ion P. d. F., Bouche T., Misra G., Onshuus A. A., Watt S. M., and Zheng L. International Mathematical Knowledge Trust IMKT: An Update On The Global Digital Mathematics Library. Proceedings of the International Congress of Mathematicians (ICM 2018), 2019, pp. 1157-1175 https://doi.org/10.1142/9789813272880_0041
11. Iz B., Lo K. and Cohan A. SciBERT: A Pretrained Language Model for Scientific Text. Conference on Empirical Methods in Natural Language Processing (2019). <https://arxiv.org/abs/1903.10676>
12. <https://github.com/allenai/scibert/> [Электронный ресурс] (дата обращения: 30.06.2025)
13. Kaddour J., at all. Challenges and Applications of Large Language Models / Kaddour J., Harris J., Mozes M., Bradley H., Raileanu R., McHardy R. 2023. <https://arxiv.org/abs/2307.10169> <https://doi.org/10.48550/arXiv.2307.10169> [Электронный ресурс] (дата обращения: 30.06.2025)
14. Luo L., Zhao Z., Gong C., Haffari G., Pan S. Graph-constrained reasoning: Faithful reasoning on knowledge graphs with large language models. arXiv preprint arXiv:2410.13080, 2024 <https://doi.org/10.48550/arXiv.2410.13080>
15. Malinka K., Perešini M., Firc A., Hujník O. and Januš F. On the Educational Impact of ChatGPT: Is Artificial Intelligence Ready to Obtain a University Degree? In Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 1 (ITiCSE 2023), July 8–12, 2023, Turku, Finland. ACM, New York, NY, USA, 7 pages. <https://doi.org/10.1145/3587102.3588827>
16. Moiseev E.I., Muromskij A.A., Tuchkova N.P. Tezaurus informatsionno – poiskovyy po predmetnoy oblasti «obyknovennyye differentsial'nye uravneniya» [Information search thesaurus of subject area "Ordinary Differential Equations"]. M.: MAKSS Press, 2005. 116p.
17. Pan S., at all. Unifying Large Language Models and Knowledge Graphs: A Roadmap // in *IEEE Transactions on Knowledge and Data Engineering*. Pan S., Luo L., Wang Y., Chen C., Wang J., Wu X., Vol. 36. №. 7. P. 3580-3599, July 2024, <https://doi.org/10.1109/TKDE.2024.3352100>
18. Pennington J., Socher R., Manning C. D. GloVe: Global Vectors for Word Representation. <https://nlp.stanford.edu/projects/glove/> [Электронный ресурс] (дата обращения: 30.06.2025)
19. Russell S., Norvig P. Artificial Intelligence, Global Edition. A Modern Approach. 4th edition. Munich, Pearson. 2021. 1168 p.
20. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A. N., Kaiser L., and Polosukhin I. 2017. Attention is all you need. In NIPS
21. Vinogradov, I.M. (red.), Matematicheskaya enciklopediya (v 5 tomah) M.: Sovetskaya enciklopediya (1977—1985)

