

УДК 004.65, 005, 001.5

## О МОДЕЛИ ПОИСКА СИНОНИМОВ

О. М. Атаева<sup>1</sup> [0000-0003-0367-5575], В. А. Серебряков<sup>2</sup> [0000-0003-1423-621X],

Н. П. Тучкова<sup>3</sup> [0000-0001-5357-9640]

<sup>1,2,3</sup>Вычислительный центр им. А.А. Дородницына ФИЦ ИУ РАН, г. Москва

<sup>1</sup>oli@ultimeta.ru, <sup>2</sup>serebr@ultimeta.ru, <sup>3</sup>natalia\_tuchkova@mail.ru

### **Аннотация**

Рассмотрена задача нахождения наиболее релевантных документов в результате расширенного и уточненного запроса. Для ее решения предложены модель поиска и механизм предварительной обработки текста, а также совместное использование поисковой системы и модели, построенной на основе индекса с помощью алгоритмов word2vec для генерации расширенного запроса с синонимами и уточнения результатов поиска на основе подбора похожих документов в цифровой семантической библиотеке. В работе исследуется построение векторного представления документов применительно к массиву данных цифровой семантической библиотеки LibMeta. Решалась задача обогащения пользовательских запросов синонимами. При построении модели поиска совместно с алгоритмами word2vec использован подход «сначала индексация, затем обучение», что позволяет получить более точные результаты поиска. Обучение модели проводилось на базе контента библиотеки для предметной области «Математика». Приведены примеры расширенного запроса с использованием синонимов.

**Ключевые слова:** модель поиска, алгоритм word2vec, синонимы, информационный запрос, расширение запроса.

### **ВВЕДЕНИЕ**

Задача поиска синонимов и схожих (похожих, аналогичных, «similar») документов изучается достаточно давно [1, 2]. Известен такой алгоритм, как латентный алгоритм Дирихле (LDA model) [3], основанный на статистической модели Байеса. Наибольшую популярность в свое время обрели алгоритмы векторного представления текстов серии tf-idf [3]. Схема tf-idf сокращает документы произ-

вольной длины до списков фиксированной длины и числа слов, не отражая семантическую структуру внутри документа. LDA-алгоритм использует тематическую привязку слов и тем самым способствует учету семантических связей документов и внутри документов.

Исследования, представленные в [1–3] и других известных работах, позволяют говорить о том, что неверная информация, получаемая по запросу, как правило, является результатом использования в базах данных ошибочных семантических связей, т. е. на этапе предварительной обработки данных не учитываются некоторые семантические связи терминов в тексте [4, 5]. Для научных работ, помещенных в поисковый индекс без учета семантических связей, специфичных для каждой предметной области, это означает, что они могут быть не найдены специалистами и не процитированы. В этом контексте особую роль играют предварительная обработка данных и применение современных подходов к решению задачи поиска достоверной научной информации на основе машинного обучения [6, 7]. Исходные данные, обретая определенную структуру в процессе обработки, могут использоваться уже в качестве источника достоверных знаний [8].

В настоящей работе изучается проблема поиска документов из контента семантической библиотеки, наиболее близких к информационному запросу. Для выбора релевантных документов использована процедура нахождения близких по тематике, схожих документов, которые можно получить в результате расширения запроса синонимами. Целью исследований является построение модели поиска, которая будет удовлетворять условиям наиболее полного удовлетворения *поисковой потребности* пользователя на имеющемся наборе документов семантической библиотеки.

Версию модели, построенную на поисковом индексе LibMeta [9] с помощью алгоритмов word2vec [10–12], далее будем сокращенно называть wsgMath, как принято ранее в работе [13]. Этот подход к совместному использованию индекса поисковой системы и нейросети позволяет получать релевантные модели и функции ранжирования, которые хорошо адаптируются к базовым данным.

В данной работе ставится задача связать модель поиска с предметной областью, границы которой очерчены ее тезаурусом и классификаторами. Таким об-

разом, поиск по контенту библиотеки, поиск новых терминов и новых семантических связей между терминами предметной области становится более осмысленным и точным.

Структура работы следующая: в первом разделе изложены принципы построения модели поиска; во второй части описано построение расширенных запросов на основе векторного представления текстов; далее приведены примеры, заключение и список цитирования.

## **1. ОСОБЕННОСТИ МОДЕЛИ ПОИСКА**

Необходимо отметить, что существуют подходы с использованием моделей, построенных с помощью алгоритмов, обученных на общедоступных наборах данных. Как правило, эти наборы не включают специальные предметные области и не учитывают их терминологическую специфику. Расширение поисковых запросов [14–16] синонимами также требует предварительно составленных словарей синонимов. Можно использовать такие ресурсы, как WordNet<sup>1</sup> или RuWordNet<sup>2</sup>, но основная проблема заключается в том, что синонимы из предварительно составленных словарей не привязаны к индексируемым данным, и их использование не улучшает результаты. Поэтому была выбрана модель использования поискового индекса совместно с векторной моделью индексируемых данных из предметной области «Математика», построенная с помощью алгоритма word2vec и обученная на математической предметной области.

Для реализации такого подхода выбрана последовательная схема работы с данными, а именно:

- определяется предметная область;
- определяется словарь, соответствующий предметной области;
- на основе связей между терминами словаря выявляются связи между документами, статьями, авторами и т. д.

Поиск и отслеживание связей выполняются следующим образом:

- производится предварительная обработка текстов;
- применяются алгоритмы машинного обучения для обработки и анализа текстов;

---

<sup>1</sup> <https://wordnet.princeton.edu/>

<sup>2</sup> <https://ruwordnet.ru/ru>

- применяются векторные представления документов и запросов для *ранжирования* результатов поиска.

Такой подход увеличивает вероятность того, что система будет более *точно реагировать* на информационную потребность пользователя и выдавать более *релевантные* ответы.

В процессе исследований была определена архитектура подсистемы поиска семантической библиотеки, которая состоит из:

- компонента предварительной обработки текста для представления документов в формате, пригодном для поиска, эффективной загрузки и хранения данных и обеспечения быстрого доступа к ним;
- компонента формирования полнотекстового индекса документов;
- компонента построения векторной модели на основе индекса с помощью алгоритмов word2vec;
- компонента обработки запросов и представления их в формате, удобном для выражения информационных потребностей пользователя на естественном языке, обогащенных синонимами из предметной области;
- компонента формирования результатов на основе оценок соответствия документа запросу, с использованием контента библиотеки.

Особенность этого подхода – в гибком сочетании всех инструментов библиотеки, таких как тезаурусы, классификаторы и энциклопедия для поиска синонимов и схожих документов, а также оценки результатов на их основе.

На рис. 1 представлены основные шаги формирования поисковой выдачи в библиотеке LibMeta. Строка запроса, поступающая из интерфейса полнотекстового поиска, проходит через блок *Анализатор*. В нем строка разбивается на слова, затем проводятся их анализ и преобразование. Из модели wsgMath извлекаются и фильтруются синонимы к словам, что позволяет сформировать расширенный запрос, с помощью которого из полнотекстового индекса извлекаются соответствующие документы.

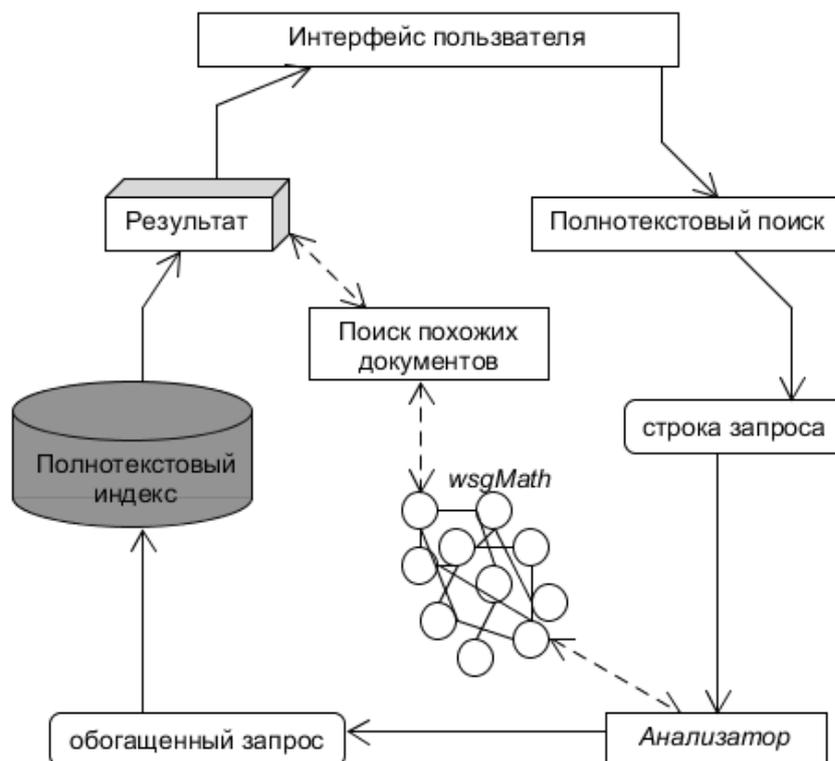


Рис. 1. Схема использования поисковой системы и нейросетевой модели.

Применение расширенной версии word2vec (doc2vec или paragraph2vec, в разных источниках) позволяет ввести дополнительный элемент, такой как *Метка фрагмента текста* или всего документа, и, основываясь на векторах этих меток, подбирать похожие документы не только по точному совпадению ключевых слов или терминов, но и, основываясь на контексте отдельных фрагментов или всего документа.

*Замечание 1.* Метка фрагмента текста используется для выдачи близких по смыслу документов, которые не попадают в поисковую выдачу, но могут представлять интерес для пользователя.

## 2. ПОСТРОЕНИЕ И ОБУЧЕНИЕ МОДЕЛИ ВЕКТОРНОГО ПРОСТРАНСТВА ПОИСКОВОЙ СИСТЕМЫ

### 2.1. Предобработка статей

Одним из необходимых этапов подготовки данных к их загрузке в определенных текстовых форматах в уже подготовленную инфраструктуру данных являются *предобработка* и *очистка* этих данных.

В нашем случае данные предоставлялись файлами в формате TeX, оформленными с разными стилями и метакомандами, т. е. сначала было необходимо заменить все авторские тэги на стандартные, очистить документы от специальных символов и неизвестных тэгов. При этом совсем избежать ручной обработки не удалось, но удалось свести ее к минимуму.

На рис. 2 приведен пример просмотра терминов в системе LiMeta, связи которых сформировались на этапах *предобработки* и *очистки* данных.

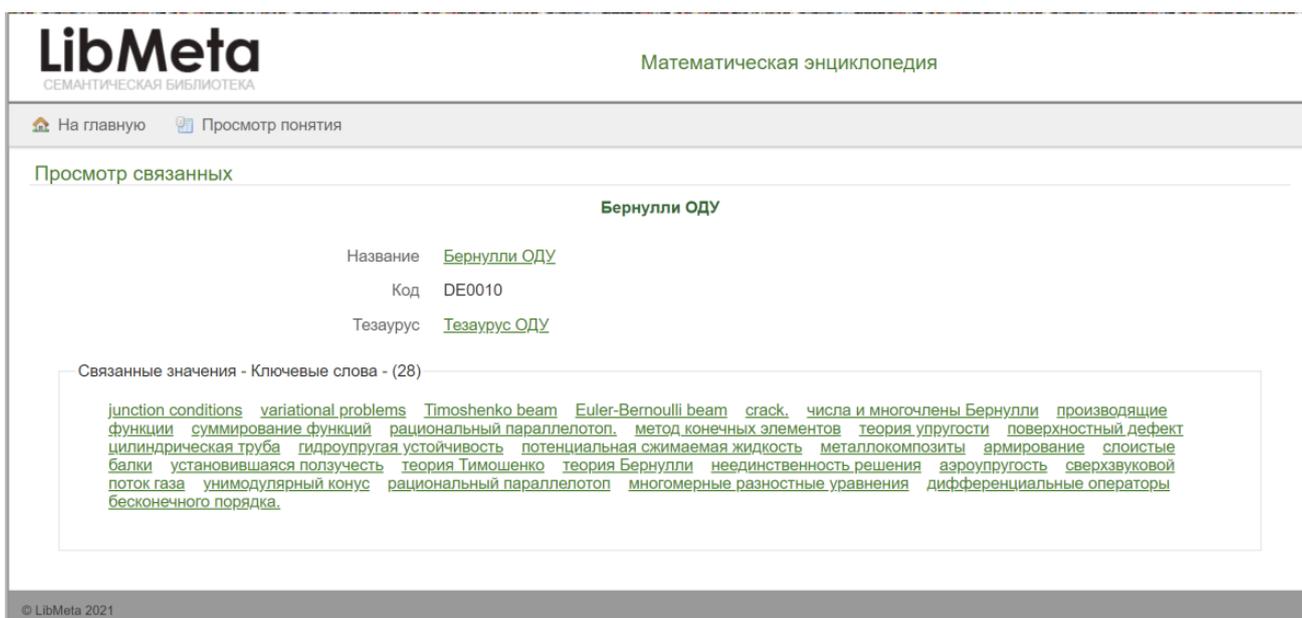


Рис. 2. Пример реализации модуля предобработки в LibMeta.

Модуль предварительной обработки выполнен на языке программирования Python вместе с интеграцией open-source библиотеки TexSoup версии 2015 года и разбит на следующие блоки:

- очистка документа;
- преобразование статьи в древовидное представление;

- обработка всех узлов дерева, запись исправленного документа.

На рис. 3 представлены основные этапы предварительной обработки текстов.

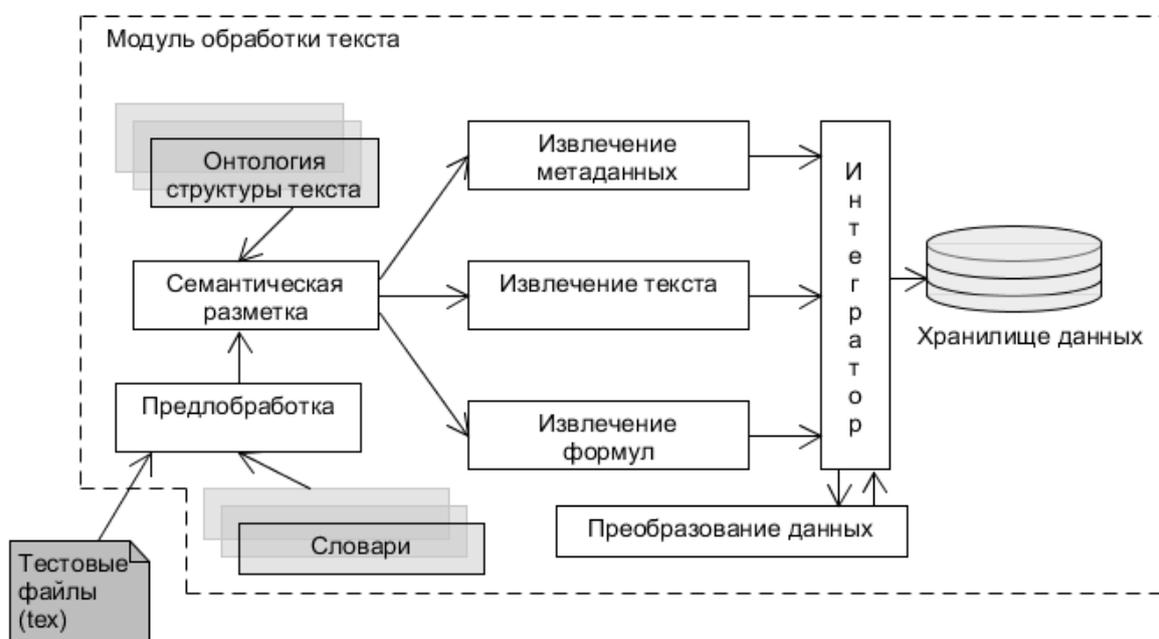


Рис. 3. Схема предварительной обработки текстов.

## 2.2. Построение индекса и обучение word2vec

Целью обучения модели wsgMath было получение синонимов, которыми можно было бы расширить поисковый запрос и получить неучтенные ранее семантические связи для дальнейшего извлечения информации, релевантной запросу.

Используемая в работе модель поиска реализует интеграцию модели, построенной на основе нейронной сети word2vec, и полнотекстового индекса. Интеграция нейронной сети и индекса может выполняться следующими способами:

- сначала обучение на корпусе текстов, затем индексация текстов и совместное использование обученной модели и индекса при поиске;
- сначала индексация, затем обучение на индексированных данных и совместное использование при поиске;
- сначала обучение, затем извлечение/создание полезных ресурсов обученной сетью, а потом индексация всех ресурсов, и новых, и исходных.

В библиотеке LibMeta использован подход «сначала индексация, затем обучение». Также изучены проблемы, как предоставить больше точных результатов на основе расширенных запросов [14–16] и как дать пользователям более «умные рекомендации» для дальнейшего поиска на основе найденных документов из предметной области в LibMeta.

На основе массива предварительно обработанных статей был построен *индекс полных текстов* на базе библиотеки поиска с открытым исходным кодом Apache Lucene<sup>3</sup>, написанный на Java<sup>4</sup>. Этот индекс используется подсистемой полнотекстового поиска библиотеки, и он же использовался для обучения алгоритма и извлечения контекстов.

Слова в контексте, близкие к рассматриваемому, трактуются как синонимы (контекстно-зависимые синонимы, в данном случае) и анализируются. Осуществляется их лексико-семантический анализ, т. е. определяются части речи, словоформы и собственные связи, в том числе со словарями и тезаурусами предметной области. На основе модели *wsgMath* численно оценивается близость контекстно-зависимых синонимов. С помощью этих оценок выбираются кандидаты, а затем из них – наилучшие с наибольшими оценками. Для дальнейшего сравнения могут быть использованы коды классификаторов, если с ними связаны выбранные слова.

В Таблице 1 приведены примеры слов со связями между словами (в первой строке стоит главное слово, в столбцах ниже – выявленные).

---

<sup>3</sup> <https://lucene.apache.org/>

<sup>4</sup> <https://www.java.com/ru/>

**Таблица 1.** Все связи слова

<i>пространство</i>	<i>краевой</i>	<i>задача</i>	<i>краевой</i>	<i>напряжение</i>	<i>остаточный</i>
оператор	граничный	решение	интегральный	деформация	концентратор
множество	интегральный	уравнение	дифференциальный	упрочнение	упрочнение
функция		условие	уравнение	пластический	усталость
		система			
		функция			

Таблица 1 содержит примеры синонимов, определенных по частям речи и коэффициенту близости слов контента библиотеки на основе модели wsgMath.

### **3. ПРИМЕРЫ (ЗАПРОС С СИНОНИМАМИ)**

#### **3.1. Расширение синонимами**

Рассмотрим термин «*краевая задача*», состоящий из двух слов – «*задача*» и «*краевая*», каждое из которых имеет собственные синонимы, представленные в Таблице 1.

В контекст термина как одной единицы попадают такие синонимы, как [*решение, уравнение, условие, система, тип, функция, область, работа*]. При этом сам термин «*краевая задача*» имеет следующие синонимы: «*граничное уравнение*», «*граничное условие*», «*граничная функция*», «*интегральная функция*», которые были определены на основе высоких оценок близости следующих пар синонимов и в соответствии с паттерном «*прилагательное + термин(сущ)*» на модели wsgMath:

$$sim ( задача, решение ) = 0.91$$

$$sim ( задача, уравнение ) = 0.86$$

$$sim ( задача, условие ) = 0.82$$

$$sim ( задача, система ) = 0.79$$

$$sim ( задача, функция ) = 0.73$$

*Замечание 2.* При построении синонимичных терминов не используются синонимы слов, определенных как именованная сущность на основе словаря, который включает в себя список персон, встречающихся в математической энциклопедии. Но при этом отметим, что в множество синонимов {коши (Коши)} попало слово риман (Риман), а в множество синонимов {лаплас (Лаплас)} попало слово фурье (Фурье).

Таким образом, были выбраны работы с высокими оценками близости синонимов.

1. *О положительном радиально-симметрическом решении задачи Дирихле для одного нелинейного уравнения и численном методе его получения*

*score = 0.90484273*

2. *О корректности краевой задачи на прямой для трех аналитических функций*

*score = 0.902505*

3. *Проекционные процедуры нелокального улучшения линейно управляемых процессов*

*score = 0.8816618*

4. *Краевая задача для частного вида уравнения Эйлера-Дарбу с интегральными условиями и специальными условиями сопряжения на характеристике*

*score = 0.846388*

5. *Теорема Валле-Пуссена для одного класса функционально-дифференциальных уравнений*

*score = 0.84127665*

### **3.2. Поиск похожих документов**

Рассмотрим пример использования элемента *Метка фрагмента текста* для процесса ранжирования документов на основе модели wsgMath при поиске схожих документов.

Когда документ поступает в систему, то извлекается его текущее векторное представление, выполняется поиск и возвращаются метки ближайших документов, косинусное расстояние которых превышает некоторый порог, определенный экспериментально как 0,6.

Далее можно также использовать для сравнения коды классификаторов как один из вариантов оценки похожих документов. При этом возможны различные варианты, связанные с наличием или отсутствием кодов классификации MSC<sup>5</sup> и УДК<sup>6</sup> у исходных документов:

- Документы, поступающие в систему, размечены кодами классификаторов MSC и УДК. В этом случае при выявлении документов, косинусное расстояние у которых превысило заданный порог, можно сравнивать коды классификаторов и устанавливать соответствие MSC и УДК. Если коды УДК отличаются у схожих документов, то можно их указать как смежные предметные области (приложения результатов, междисциплинарные исследования и пр.).
- Документы не снабжены кодами, но ключевые слова соответствуют предметной области, и в словаре (тезаурусе, энциклопедии) есть коды классификаторов. В этом случае сравниваются коды ключевых слов, и документам приписываются соответствующие коды.

На рис. 4 приведен пример соответствия кодов классификаторов, полученных на основе контента LibMeta и процедуры выявления синонимов. В данном случае было выявлено, что коду УДК 515.128 соответствуют такие коды MSC, как 54E20, 54E40, 54D65 и т. д.

---

<sup>5</sup> <https://cran.r-project.org/web/classifications/MSC.html>

<sup>6</sup> <https://teacode.com/online/udc/>

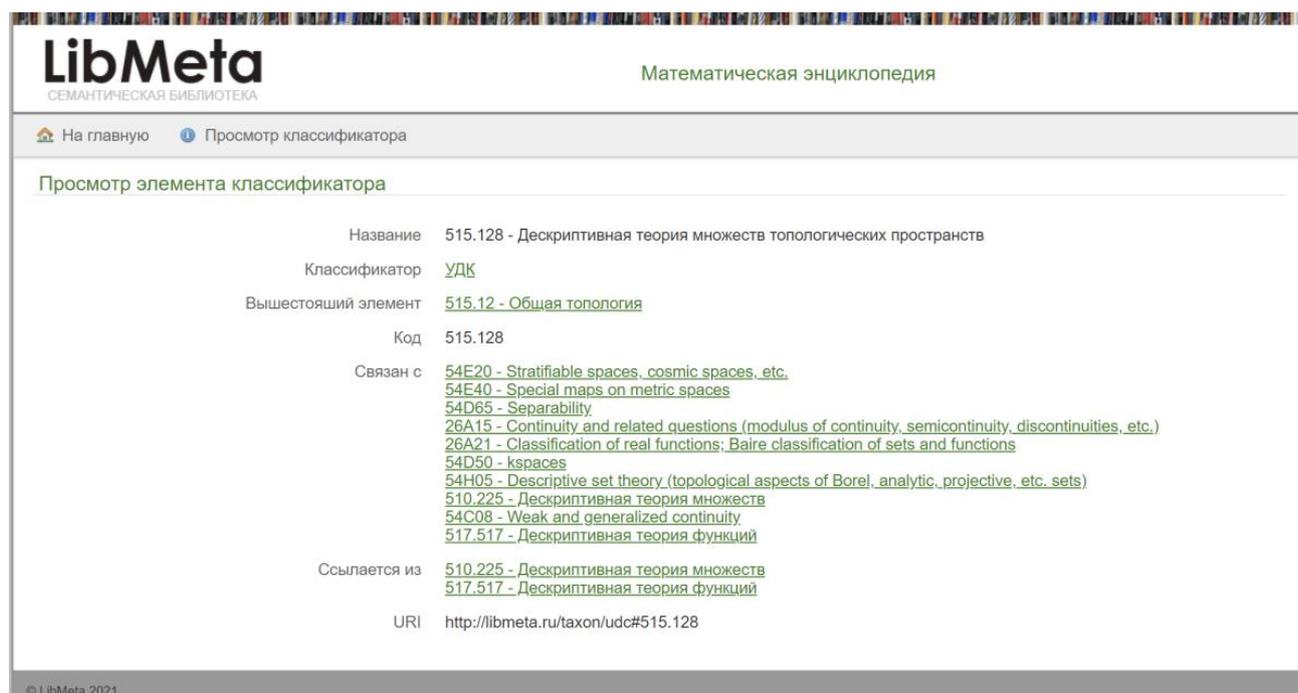


Рис. 4. Пример соответствия кодов классификаторов MSC и УДК.

## ЗАКЛЮЧЕНИЕ

В представленном исследовании получены следующие основные результаты.

Показано, что предварительная обработка входных массивов данных (текстов научных статей) позволяет учесть в дальнейшем дополнительные семантические связи и улучшить качество поиска.

Использование механизма интеграции нейронной сети и индекса позволяет реализовать варианты поисковой модели для получения релевантных документов с заданной точностью.

Совместное использование индекса поисковой системы и нейросети позволяет получать релевантные модели и функции ранжирования, которые хорошо адаптируются к базовым данным.

Предложенная модель поиска позволяет также устанавливать соответствие кодов классификаторов для близких документов, находить синонимы при контекстном сравнении и ранжировать документы на основе метки фрагмента.

Выявлены проблемы для дальнейшего изучения – развитие механизма оценки качества поиска с использованием различных метрик, использование английских и русских синонимов для обогащения запроса и улучшения качества поиска, оценки скорости обучения модели. Решение этих проблем вытекает из проделанных исследований, которые позволяют сформулировать конкретные задачи для улучшения качества поиска. Это составление словарей синонимов предметной области, связанных с классификаторами и эталонных документов, связанных с терминами тезауруса предметной области. Такие ресурсы могут позволить в дальнейшем улучшить качество поиска на основе алгоритмов машинного обучения.

Работа выполнена при частичной поддержке Российского фонда фундаментальных исследований, проект № 20-07-00324, и в рамках темы Министерства науки и высшего образования РФ «Математические методы анализа данных и прогнозирования».

#### **СПИСОК ЛИТЕРАТУРЫ**

1. *Baeza-Yates R., Ribeiro-Neto B.* Modern Information Retrieval. ACM Press, New York, 1999. 518 p.
2. *Salton G.* Introduction to Modern Information Retrieval. McGraw-Hill, 1983, 513 p.
3. *Blei D.M., Ng A.Y., Jordan M.I.* Latent Dirichlet Allocation // Journal of Machine Learning Research. 2003. V. 3. P. 993–1022.
4. *Furnas G.W., Landauer T.K., Gomez L.M., Dumais S.T.* The vocabulary problem in human-system communication // Commun. ACM. 1987. V. 30, No. 11 P. 964–971.
5. *Biswas G., Bezdek J., Oakman R.L.* A knowledge-based approach to online document retrieval system design. In Proc. ACM SIGART Int. Symp. Methodol. Intell. Syst. 1986. P. 112–120.
6. *Мак-Каллок У.С., Питтс В.* Логическое исчисление идей, относящихся к нервной активности // Автоматы. Под ред. К. Э. Шеннона и Дж. Маккарти. М.: Изд-во иностр. лит., 1956. С. 363–384 (Перевод англоязычной статьи 1943 г.).
7. Профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных. URL: <http://www.machinelearning.ru/> (доступно 26.10.2021)

8. Гаврилова Т.А., Хорошевский В.Ф. Базы знаний интеллектуальных систем. СПб.: Питер, 2000. 384 с.

9. Атаева О.М., Серебряков В.А. Онтология цифровой семантической библиотеки LibMeta // Информатика и её применения. 2018. Т. 12. № 1. С. 2–10.

10. Mikolov T., Chen K., Corrado G., Dean J. Efficient Estimation of Word Representations in Vector Space // Proceedings of Workshop at ICLR, 2013.

11. Mikolov T., Yih W.T., Zweig C. Linguistic Regularities in Continuous Space Word Representations // Proceedings of NAACL HLT, 2013.

12. Le Q., Mikolov T. Distributed Representations of Sentences and Document // International Conference on Machine Learning. 2014. P. 1188–1196.

13. Атаева О.М., Серебряков В.А., Тучкова Н.П. Using Applied Ontology to Saturate Semantic Relations // Lobachevskii Journal of Mathematics. 2021. V. 42. No. 8. P. 1776–1785.

14. Voorhees E.M. Query expansion using lexical-semantic relations. 17th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr., Dublin, Ireland, 1994.

15. Buckley C., Salton G., Allan J., Singhal A. Automatic query expansion using SMART: TREC 3, presented at the 3rd Text Retr. Conf. (TREC), 1995.

16. Efthimiadis E.N. Query expansion // Annu. Rev. Inf. Sci. Technol. 1996. V. 31. No. 5. P. 121–187.

---

## ON THE SYNONYM SEARCH MODEL

O. M. Ataeva<sup>1</sup> [0000-0003-0367-5575], V. A. Serebriakov<sup>2</sup> [0000-0003-1423-621X],

N. P. Tuchkova<sup>3</sup> [0000-0001-5357-9640]

<sup>1,2,3</sup>Dorodnicyn Computing Centre FRC CSC RAS, Moscow

<sup>1</sup>oli@ultimeta.ru, <sup>2</sup>serebr@ultimeta.ru, <sup>3</sup>natalia\_tuchkova@mail.ru

### Abstract

The problem of finding the most relevant documents as a result of an extended and refined query is considered. For this, a search model and a text preprocessing mechanism are proposed, as well as the joint use of a search engine and a neural network model built on the basis of an index using word2vec algorithms to generate an

---

extended query with synonyms and refine search results based on a selection of similar documents in a digital semantic library. The paper investigates the construction of a vector representation of documents based on paragraphs in relation to the data array of the digital semantic library LibMeta. Each piece of text is labeled. Both the whole document and its separate parts can be marked. The problem of enriching user queries with synonyms was solved, then when building a search model together with word2vec algorithms, an approach of "indexing first, then training" was used to cover more information and give more accurate search results. The model was trained on the basis of the library's mathematical content. Examples of training, extended query and search quality assessment using training and synonyms are given.

**Keywords:** *search model, word2vec algorithm, synonyms, information query, query extension*

## REFERENCES

1. *Baeza-Yates R., Ribeiro-Neto B.* Modern Information Retrieval. ACM Press, New York, 1999. 518 p.
2. *Salton G.* Introduction to Modern Information Retrieval. McGraw-Hill, 1983, 513 p.
3. *Blei D.M., Ng A.Y., Jordan M.I.* Latent Dirichlet Allocation // Journal of Machine Learning Research. 2003. V. 3. P. 993–1022.
4. *Furnas G.W., Landauer T.K., Gomez L.M., Dumais S.T.* The vocabulary problem in human-system communication // Commun. ACM. 1987. V. 30. No. 11. P. 964–971.
5. *Biswas G., Bezdek J., Oakman R.L.* A knowledge-based approach to online document retrieval system design. In Proc. ACM SIGART Int. Symp. Methodol. Intell. Syst. 1986. P. 112–120.
6. *Mak Kallok U.S., Pitts V.* Logicheskoe ischislenie idej odnosyashchihsy k nervnoj aktivnosti Avtomaty Ed. Shennon i Dzh Makkarti M: Izd-vo inostr. Lit. 1956. S. 363–384. Pervod anglijskoj stati 1943 g.
7. Professionalnyj informacionno analiticheskij resurs posvyashchennyj mashin nomu obucheniyu raspoznavaniyu obrazov i intellektualnomu analizu dannyh URL: <http://www.machinelearning.ru/> (access 26.10.2021)

8. *Gavrilova T.A., Horoshevskij V.F.* Bazy znaniy intellektualnyh sistem SPb: Piter. 2000, 384 s.
9. *Ataeva O.M., Serebryakov V.A.* Ontologiya cifrovoj semanticheskoy biblioteki LibMeta // Informatika i eyo primeneniya. 2018. V. 12. No. 1. S. 2–10.
10. *Mikolov T., Chen K., Corrado G., Dean J.* Efficient Estimation of Word Representations in Vector Space // Proceedings of Workshop at ICLR, 2013.
11. *Mikolov T., Yih W.T., Zweig C.* Linguistic Regularities in Continuous Space Word Representations // Proceedings of NAACL HLT, 2013.
12. *Le Q., Mikolov T.* Distributed Representations of Sentences and Document // International Conference on Machine Learning. 2014. P. 1188–1196.
13. *Ataeva O.M., Sererbryakov V.A., Tuchkova N.P.* Using Applied Ontology to Saturate Semantic Relations // Lobachevskij Journal of Mathematics. 2021. V. 42. No. 8. P. 1776–1785.
14. *Voorhees E.M.* Query expansion using lexical-semantic relations. 17th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr., Dublin, Ireland, 1994.
15. *Buckley C., Salton G., Allan J., Singhal A.* Automatic query expansion using SMART: TREC 3, presented at the 3rd Text Retr. Conf. (TREC), 1995.
16. *Efthimiadis E.N.* Query expansion // Annu. Rev. Inf. Sci. Technol. 1996. V. 31. No. 5. P. 121–187.

## СВЕДЕНИЯ ОБ АВТОРАХ



**АТАЕВА Ольга Муратовна** – научный сотрудник Вычислительного центра им. А.А. Дородницына ФИЦ ИУ РАН, кандидат техн. наук, специалист в области системного программирования и баз данных.

**Olga Muratovna ATAeva** – researcher of the of Dorodnicyn computing center FRC SCS RAS, PhD, expert in the field of system programming and data-bases.

email: oli@ultimeta.ru

ORCID: 0000-0003-0367-5575



**СЕРЕБРЯКОВ Владимир Алексеевич** – специалист в области теории формальных языков и её приложений, доктор физ.-мат. наук, профессор, зав. отделом Вычислительного центра им. А.А. Дородницына ФИЦ ИУ РАН. Руководитель и участник разработки ряда известных программных проектов, в частности, ИСИР и ИСИР РАН, «Научный портал РАН».

**Vladimir Alekseevich SEREBRIAKOV** – expert in the field of theory of formal languages and its applications, doctor of sciences, professor, head of Dorodnicyn computing center FRC SCS RAS department. Head and participant in the development of a number of well-known program projects, in particular, ISIR and ISIR RAS, Scientific portal RAS.

email: serebr@ultimeta.ru

ORCID: 0000-0003-1423-621X



**ТУЧКОВА Наталия Павловна** – старший научный сотрудник Вычислительного центра им. А.А. Дородницына ФИЦ ИУ РАН, кандидат физ.-мат. наук, окончила ВМиК МГУ им. М.В. Ломоносова. Специалист в области алгоритмических языков и информационных технологий.

**Natalia Pavlovna TUCHKOVA** – senior researcher of Dorodnicyn computing center FRC SCS RAS, PhD in physics with a math degree, graduated from CS Faculty of Lomonosov MSU. The expert in the field of algorithmic languages and information technologies.

email: natalia\_tuchkova@mail.ru

ORCID: 0000-0001-5357-9640

*Материал поступил в редакцию 05 ноября 2021 года*