

УДК 004.82+004
DOI: 10.15827/0236-235X.141.083-096

Дата подачи статьи: 31.10.22, после доработки: 30.11.22
2023. Т. 36. № 1. С. 083–096

Интеграция подпространства предметной области в семантическое пространство «математика»

О.М. Атаева¹, к.т.н., научный сотрудник, oli.ataeva@gmail.com

В.А. Серебряков¹, д.ф.-м.н., профессор, руководитель отдела, serebrvas@gmail.com

Н.П. Тучкова¹, к.ф.-м.н., старший научный сотрудник, natalia_tuchkova@mail.ru

¹ Федеральний исследовательский центр «Информатика и управление» РАН,
г. Москва, 119333, Россия

В работе рассматривается проблема развития семантической библиотеки путем добавления новой прикладной научной области. На примере журнала по прикладным вопросам композиционных материалов строится дополнение к основному контенту библиотеки, расширяется описание исходной предметной области и происходит детализация статей УДК и MSC, соответствующих локальной предметной области. При этом решаются задачи добавления терминов в тезаурус, построения эталонного корпуса прикладной предметной области математики, создания настраиваемого интерфейса. Формулы и уравнения локальной предметной области семантически связываются с основным контентом библиотеки.

Главное преимущество использования семантических библиотек для подобного рода задач состоит в обогащении имеющейся базы знаний библиотеки и выявлении связей в данных. Реализация перечисленных задач ведется средствами семантической библиотеки LibMeta, что позволяет обсуждать такие понятия, как тематическая область/подобласть и их иерархические связи, в частности, автоматическое включение понятий охватывающей области в систему знаний подобласти, возможности персонализации построенных конструкций и использования LibMeta как инструмента построения, в частности, собственно онтологии.

Исследование этих задач требует взаимодействия с экспертами предметной области и использования современных средств и методов обработки естественного языка, машинного обучения подходов к представлению знаний. Интеграция данных в рамках библиотеки позволяет расширять описание предметных областей, связанных с приложениями математики в междисциплинарных исследованиях и технологиях. Для одного из прикладных разделов задач математической физики показана процедура включения специфических словарей, тезаурусов, а также массива публикаций специализированного журнала в семантическую библиотеку.

Предлагаемый подход позволит использовать контент семантической библиотеки «математика» для научных исследований, минимизируя процесс поиска информации в локальной предметной области без потери более общих результатов, содержащихся за пределами этой области.

Ключевые слова: семантическая библиотека, научная предметная область, онтология предметной области, тезаурус предметной области, поиск по локальной предметной области.

За несколько десятилетий цифровизации термин «онтология» проник в различные области знаний. Исследователи, ученые и специалисты используют, разрабатывают или применяют онтологии в качестве механизма представления предметных областей в информационных системах. Термин «онтология», истоки которого можно найти в философии, был принят в исследовательском сообществе искусственного интеллекта для формального описания областей знаний.

По определению, обязательными свойствами описания знаний в виде онтологии являются конечный контролируемый словарь понятий и терминов, исключаяющий их двусмысленную интерпретацию, а также строгая

иерархия отношений подклассов понятий и терминов, которые описывают знания предметной области [1–3].

На сегодня разработано огромное количество онтологий для предметных областей, которые охватывают разные аспекты деятельности человека. В данной работе рассматриваются научная область «математика» и процесс конструирования онтологии для одного из ее прикладных разделов в задачах математической физики (ограниченного тематикой журнала по механике композитов) на основе уже имеющихся источников данных и контента библиотеки [4].

Построение модели онтологии предметной области позволяет выделить метаданные для

проектирования конкретных структур – научных предметных областей – и варианты управления этими данными. Для этого необходимо структурировать и связать различные ресурсы, извлечь из них и контекстуализировать (определить в контексте) данные, придавая им свойства знаний. Для определения предметной области строится модель данных, в рамках которой могут интегрироваться различные источники данных, использоваться различные таксономии понятий и терминов, верифицированные признанными экспертами научной области. В качестве базовой таксономии в работе использованы Математическая энциклопедия и онтология научной предметной области, которая лежит в основе конструирования семантической библиотеки LibMeta в совокупности с отраслевыми классификаторами MSC и УДК [5, 6].

В работе исследуются задачи использования накопленных данных библиотеки для описания конкретной области научных исследований в публикациях журнала «Механика композиционных материалов и конструкций» (МКМК) (<https://mkmk.ras.ru>) (архив за последние 25 лет) и встраивания новых понятий из прикладной области математической физики в контент семантической библиотеки.

Аналогичные исследования ведутся в рамках развития европейских и отечественных баз математических данных. В проекте zbMATH (<https://zbmath.org>) одна из целей развития сформулирована как размещение и индексация в соответствии с MSC2020 печатных изданий с 1868 г. с последующим бесплатным доступом к ним научного сообщества. Новые публикации позволяют расширять статьи классификатора в zbMATH и включать новые области знаний в эту библиотеку. Еще один проект – NIST (<https://dlmf.nist.gov>) использует как базовый справочник формул, где представлены труды с 1947 г. и даже более ранние.

Представлению формул и их распознаванию с применением методов векторного анализа [7] и других алгоритмов искусственного интеллекта [8, 9] посвящено много проектов. В работе [10] рассматривается механизм перевода формул для спецфункций из формата Wolfram Mathematica [11] в формат NIST. В других аналогичных проектах исследуются проблемы семантической идентификации формул [12], представления спецфункций [13, 14] и снабжения формул контекстом [15]. В русскоязычном сегменте Интернета можно отметить проект EqWorld (<http://eqworld.ipmnet.ru/polyanin-ew-ru.htm>), посвященный отдельным

разделам математической физики и механики и опирающийся на справочники авторского коллектива проекта, а также проект OntoMath (<https://ontomathpro.org>), посвященный разработке математической онтологии. Из тематически близких терминов предметной области, которая добавляется в семантическую библиотеку, можно отметить русско-английский словарь нанотерминов (<https://thesaurus.rusnano.com>), который основан на одноименном издании под эгидой проектов «Роснанотех» за 2007–2010 гг. Из перечисленных примеров только в проекте zbMATH, как и в LibMeta, новые статьи используются для извлечения новых терминов, но в этом ресурсе, естественно, не задействован русскоязычный сегмент публикаций.

Особенностью представленного подхода является то, что проектирование выполняется не просто в контексте математической области, а средствами семантической библиотеки LibMeta [16], что позволяет говорить о понятиях предметной области/подобласти, их иерархической связи, в частности, автоматическом включении понятий охватывающей области в систему знаний подобласти, о возможности персонализации построенных конструкций и о том, что LibMeta выступает так же, как инструмент построения, в частности, построения собственно онтологии. Отличие предлагаемого подхода и в развитии методов онтологического проектирования на основе включения новых предметных областей в контент существующей семантической библиотеки путем добавления нового ресурса – тематического журнала. В результате не только создается предметная область журнала в рамках семантической библиотеки, но и углубляются позиции классификаторов за счет новых русскоязычных терминов, формул и связей.

Проблемы построения онтологии

Несмотря на долгую историю онтологического проектирования, нет единого стандарта для построения онтологий, но существуют различные методологии [17] и ГОСТ Р 59798-2021, содержащие определения базисных формальных онтологий.

Методологии определяют различные понятия на разных уровнях концептуализации, но сходятся в основной последовательности шагов, необходимых для построения онтологии:

- определение цели онтологии;
- выделение основных понятий верхнего уровня и их свойств;
- выделение связей между ними.

После определения целей и задач формирования онтологии выделяют три основных процесса:

- сбор данных из источников, выявление их структуры;
- представление знаний в виде таксономий;
- реализация и построение на основе экспертных знаний формальной модели, пригодной для машинного использования в конкретной задаче.

Определение целей разработки онтологии. Цель разработки онтологии приложений предметной области заключается в необходимости интеграции данных для описания приложений математики и научных исследований в области этих приложений, отраженных в научных публикациях. Как уже отмечалось в [5, 17], существует определенный разрыв в представлении знаний в цифровых ресурсах и их отражении в библиографических ресурсах. Примерами интеграции математических знаний и публикаций служат журнал Zentralblatt, англоязычная версия Математической энциклопедии (<https://encyclopediaofmath.org/>), LibMeta [6, 17] (<https://libmeta.ru>) и некоторые другие ресурсы [18, 19]. Многочисленные библиографические ресурсы [20, 21] служат примерами отдельных БД. Это связано, конечно, с тем, что цели этих разработок разные. Тем не менее нельзя не отметить, что для научного сообщества важно иметь именно интеграцию данных, чтобы, находясь в рамках цифровой предметной области, можно было ознакомиться с публикациями на выбранную тему. Такая возможность реализуема при создании семантической библиотеки [12], где наряду с терминологическими связями предметной области присутствуют связи с публикациями. Это особенно важно в междисциплинарных исследованиях, которые не всегда укладываются в классическую классификацию, что зачастую составляет трудность в нахождении таких работ. Междисциплинарными предметными областями являются, например, современные приложения классических уравнений математической физики. Некоторые исследователи предлагают выделить современные приложения в раздел «новая прикладная математика» [22]. Пока этого раздела не существует и даже не определены его рамки, создание онтологии прикладной области математики составляет актуальную задачу онтологического проектирования.

Определяя цели создания онтологии прикладной предметной области математической

физики, прежде всего необходимо определить ее назначение – использование цифровой библиотеки книг, журналов, публикаций, различных научных материалов и ресурсов для научных и образовательных целей. Ресурсы онтологии являются, с одной стороны, наполнением семантической библиотеки, а с другой – источником знаний в предметной области при их определенной классификации и категоризации.

Онтология позволяет обогатить данные горизонтальными и вертикальными семантическими связями, определить неявные связи, например, между задачами и их приложениями, между персонами, связанными с этими ресурсами. Также онтологическое проектирование взаимно обогащает уже связанные ресурсы.

Реализация этой цели в рамках семантической библиотеки позволит перейти к построению, в частности, тезауруса (ГОСТ 7.24-2007) предметной области, который будет расширяться и наполняться с ростом накопленной в библиотеке информации.

Онтологизация: сбор и структурирование данных. Основные источники данных можно разделить на две большие категории:

- данные из журналов и научных публикаций, демонстрирующие во времени развитие предметной области;
- учебники, монографии, словари и классификаторы, содержащие основную терминологию предметной области.

В данном исследовании в качестве источника первого типа использованы публикации тематического журнала, а также тезаурусы, статьи энциклопедии и публикации, накопленные ранее в библиотеке LibMeta.

Вторым типом источников стали признанные экспертами предметной области книги, учебники, монографии и терминологические словари, в том числе написанные руководителями классических российских математических школ.

За терминологическую основу предметных областей взяты классические монографии на русском языке и авторские оригинальные разработки, связанные с терминологическим анализом первоисточников.

Работы российских академиков А.Н. Тихонова, А.А. Самарского, Л.И. Седова и В.В. Васильева [23] посвящены исследованиям теории дифференциальных уравнений, теории упругости, спецфункций математической физики и приложениям к конкретным задачам. Словари терминов на их основе используются для учета

исторических и современных связей в выбранной предметной области.

На основе этих источников были определены главные термины и понятия предметной области. Для выявления структуры связей внутри предметной области и добавления связей с онтологией уже накопленного контента библиотеки также использовались предметные указатели, словари и классификации из перечисленных монографий, классификаторы MSC, УДК, Математическая энциклопедия.

На основе полученных связей формируется тезаурус предметной области, строится таксономия предметной области.

Таксономия предметной области и структура тезауруса. В процессе построения онтологии решается вопрос ее ограничения в рамках конкретной предметной области науки. Для этого вводится базовый набор понятий предметной области, с которыми связывают соответствующие термины предметной области. Как правило, термины предметной области организованы в виде некоторой таксономии с поддержкой связей между ними. Структура этой таксономии может варьироваться по сложности в зависимости от моделируемой области, включая полноценный тезаурус со всем богатством связей. В дальнейшем будем говорить о тезаурусах как о средстве организации понятий (знаний). Представленные в таком виде термины в целом упрощают процесс обработки имеющихся ресурсов. Тезаурус предметной области может быть как результатом работы экспертной группы, так и построенным автоматизированными средствами.

Исходя из особенностей предметной области и входящих в нее понятий, рассмотрим формирование структуры тезауруса предметной области «уравнения математической физики и смежные области».

Из описания уравнений выделим основные типы связей различных лексико-семантических категорий:

- вид уравнения: одномерное, двухмерное, трехмерное;
- тип уравнения: гиперболическое, параболическое, эллиптическое;
- типы коэффициентов: переменные, постоянные.

Таким образом, вариант структуры тезауруса включает в себя:

- тематические разделы: дифференциальные уравнения с частными производными, уравнения смешанного типа;

- категории терминов для описания тематических подразделов: однородное, неоднородное, одномерное, двухмерное, трехмерное, гиперболическое, параболическое, эллиптическое, именное, нарицательное, с переменными коэффициентами, с постоянными коэффициентами;

- список связей между терминами: иерархические – род, вид и горизонтальные – синонимы, ассоциация;

- схему статьи тезауруса для описания понятия: код понятия (идентификатор), коды классификаторов, дескриптор (заглавный термин), недескрипторы (дополнительные термины), символьные представления формул, текстовые дополнения (комментарии, замечания, справки); горизонтальные связи между понятиями, ключевые слова, иерархические связи между понятиями, связи с объектами (при разработке структурной схемы статьи тезауруса нужно также поддерживать возможность разнообразных связей с объектами, не являющимися явно понятиями тезауруса, но необходимыми для полноты описания, в рассматриваемом случае – авторы и публикации; для этого в структуре статьи тезауруса предусмотрен соответствующий набор связей для описания списков литературы, авторов и т.д.).

Описанная структура тезауруса предметной области с возможностью связывания его понятий с различными объектами предметной области хорошо ложится на предложения по моделированию предметной области с терминологической поддержкой в рамках проекта LibMeta по построению цифровой библиотеки.

Разработка онтологии

Онтология прикладной предметной области математической физики конструируется на основе онтологии семантической библиотеки LibMeta. Онтология предметной области «математика» включает терминологическое описание на основе Математической энциклопедии и уже создана ранее [5]. Это позволяет использовать онтологию информационных ресурсов для описания публикаций, персон, задач и связанных с ними формул для описания данных, извлекаемых из новых источников, а также использовать тезаурус для представления таксономий предметной области.

При определении связей и понятий предметной области в рамках онтологии семантической библиотеки возможны три подхода.

- Сверху вниз, когда проектирование понятий и связей начинается с верхнеуровневых понятий.

- Снизу вверх, когда проектирование понятий и связей начинается с низкоуровневых данных, которые группируются и формируют более общие понятия и т.д.

- Комбинированный, когда сформулированы основные понятия конкретной предметной области, ее данные очищены и частично структурированы, например, представлены в виде отдельных таксономий. Необходимо связать разрозненные данные в рамках заданной предметной области, одновременно обогащая ее и уточняя включаемые ресурсы. В данном случае использован последний (комбинированный) подход. Это означает, что предварительная работа по выделению верхнеуровневых понятий предметной области и понятий, необходимых для описания структуры и связей ее тезауруса, была выполнена ранее при проектировании библиотеки, а препроцессинг (подготовка данных к загрузке в библиотеку) в рамках подготовки данных к загрузке.

Приведем список некоторых понятий, использованных для формирования онтологии тезауруса рассматриваемой предметной области. Условно их можно разделить на две группы. Первая группа включает понятия, необходимые для описания модели тезауруса, и соответствует по составу набору понятий стандарта для многоязычных тезаурусов, поэтому их описание опускается. Вторая группа составлена с учетом специфики тезауруса предметной области и содержит следующие понятия.

1. Группа: методы, задачи, уравнения.
2. Виды: эллиптический, гиперболический, параболический.
3. Формула.
4. Персона (для поддержки именных понятий (например, уравнений) предметной области);
5. Публикация и т.д.

Все понятия второй группы связаны с понятием «концепт» следующими связями: относится к группе, имеет вид, включает формулу, связан с персоной, связан с публикацией, имеет код.

Эти дополнительные понятия и связи позволяют ввести также дополнительные уровни классификации в предметной области. При включении тезауруса в контент предметной области (используя цепочки связей) данные дополнительно обогащаются ключевыми словами публикаций, авторскими ключевыми

словами, дополнительными кодами классификаций. Благодаря этим связям эксперт может анализировать развитие некоторой области знаний и использовать их для расширения тезауруса.

Интеграция математических знаний в цифровой среде

Основу рассматриваемого подхода формирования тематического подпространства составляет установление семантических связей с уже накопленным контентом библиотеки.

Базовая идея метода включения состоит в использовании словарей, тезаурусов и связей с Математической энциклопедией [21, 22]. Это множество данных применяется для присоединения нового массива метаданных научных изданий к существующему множеству метаданных библиотеки.

В процессе предварительной обработки текстов научных изданий производится их семантический анализ и формируется локальное описание предметной области этого массива в рамках контента библиотеки.

Рассматриваются такие предметные области, как *обыкновенные дифференциальные уравнения* (ОДУ), уравнения в частных производных, уравнения механики сплошной среды, уравнения механики композитов и их решения, выраженные через специальные функции математической физики. Центральным ресурсом выступает описание Математической энциклопедии в ее классическом виде. Таким образом, через связи уравнений прикладных задач и их решений можно выявить соответствующие классификаторы и использовать их также для формирования описания предметной области. Предметная область журнала МКМК является новой и ранее в цифровой библиотеке не описывалась.

Такая процедура становится возможной благодаря интеграции данных в библиотеку LibMeta [6, 24], что позволяет расширять описание предметных областей, связанных с приложениями математики в междисциплинарных исследованиях и технологиях. Реализация этого подхода составляет одно из новых направлений наполнения семантической библиотеки. Здесь используется метод интеграции данных и насыщения связями, которые выявляются только в процессе анализа новых данных. В результате выявляются пробелы в описании смежных областей математики, связанных с приложениями решений классических задач

математической физики. Это происходит, если в процессе добавления нового описания в составе библиотеки обнаружатся новые связи.

На рисунке 1 приведен фрагмент схематичного представления семантического пространства предметной области «математика», связанный с Математической энциклопедией. Представлены также в виде ресурсов тезаурус ОДУ, словарь спецфункций, классификаторы MSC и УДК, публикации с некоторыми метаданными (библиография, автор). Источник данных представлен публикациями журнала МКМК. В качестве внешних источников рассматриваются DBPedia, английская версия Математической энциклопедии. Все эти ресурсы связаны между собой, также возможны извлечения новых связей на базе имеющихся в семантической библиотеке, которые обозначены пунктирными линиями.

Ресурсы предметной области «математика». При описании предметной области нужно, с одной стороны, учитывать разнообра-

зие типов ресурсов в этой области и, с другой, обеспечить ее описание в терминах, верифицированных экспертами данной области. Использование верифицированных терминов обеспечивает высокое качество описания ресурсов и интеграции данных из разных источников.

Применение подхода на основе онтологий как для создания модели данных предметной области, так и для представления данных в терминах этой онтологии [25, 26] позволяет быстрее и легче интегрировать разные источники данных, а также выполнять более осмысленный поиск информации, связывать данные из различных источников, обогащать и дополнять имеющуюся информацию. Благодаря онтологии можно перейти от простого атрибутивного и полнотекстового вариантов поиска, основанных на точных совпадениях слов и словоформ, к семантическому поиску, когда дополнительно используются терминологические связи предметной области для формирования результата поискового запроса [26]. Более подробно с опи-

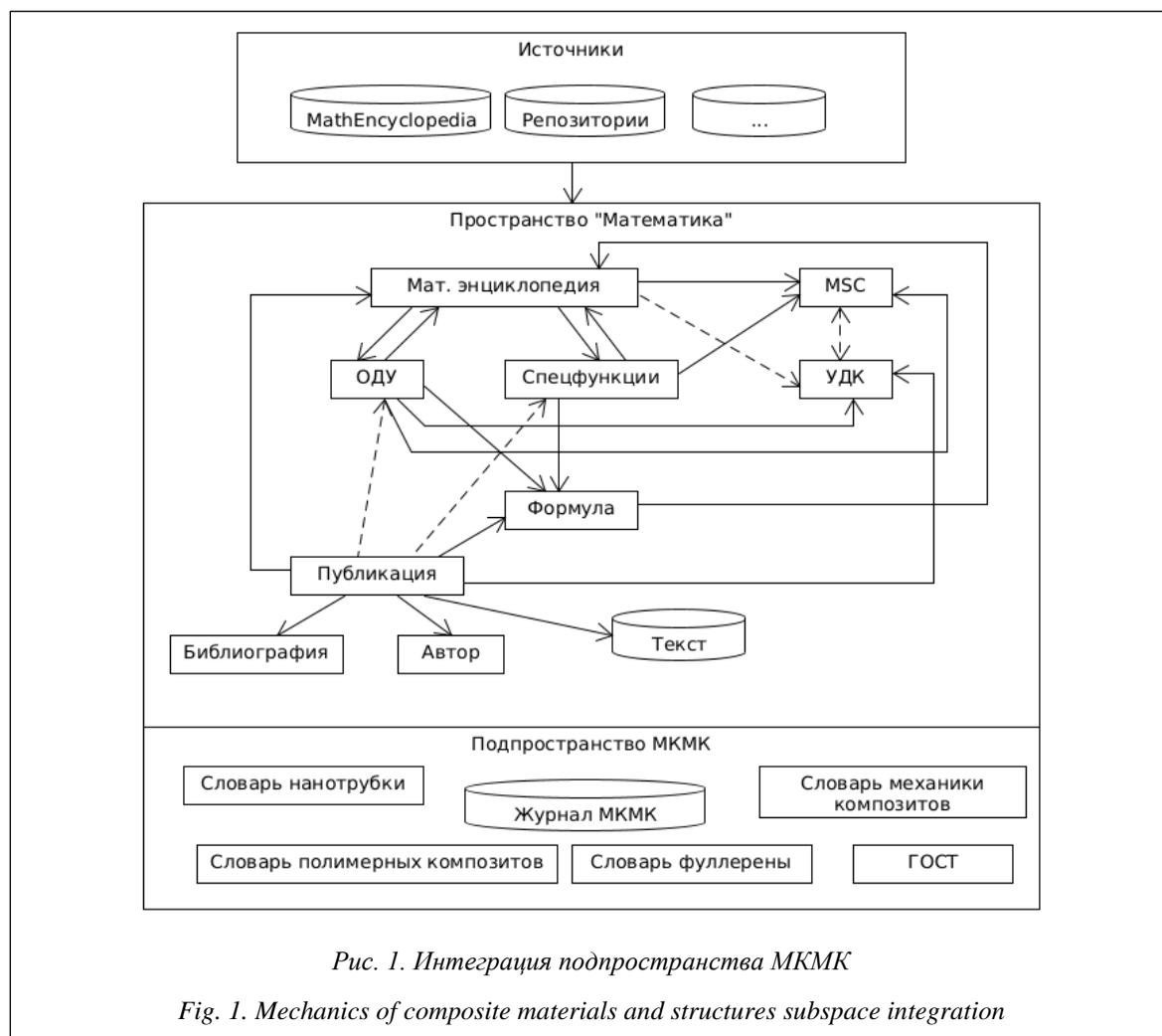


Рис. 1. Интеграция подпространства МКМК

Fig. 1. Mechanics of composite materials and structures subspace integration

санием используемых ресурсов можно ознакомиться в работах [27, 28].

Представление формул. Для поддержки формул в онтологию LibMeta было введено понятие «формула», которое позволяет хранить оригинальную строку формулы из источника, откуда она получена. Строка может быть в форматах Content MathML, Presentation MathML, LaTeX [29, 30]. При необходимости количество типов представления формулы в различных нотациях легко расширяется. Понятие формулы связано отношениями с понятиями тезауруса и с объектами из источников данных [31]. Таким образом, можно построить сеть связей формулы.

На рисунке 2 представлена сеть, каждый узел которой доступен из узла «Формула». При этом к каждой формуле могут быть привязаны ключевые слова. Они могут как проставляться экспертом, так и добавляться автоматически, поступая вместе с формулой из ее источника, а также пополняясь ключевыми словами связанных объектов. Формулы могут использоваться в связях «смотри также» и как обозначения для описания тематик. Следует подчеркнуть, что на текущем этапе используются только формулы, связанные с понятиями предметной области.

Примеры. Приведем небольшую часть тезауруса, посвященную задаче Ламе. Это статья лексико-семантического указателя для предметной области «механика сплошной среды»,

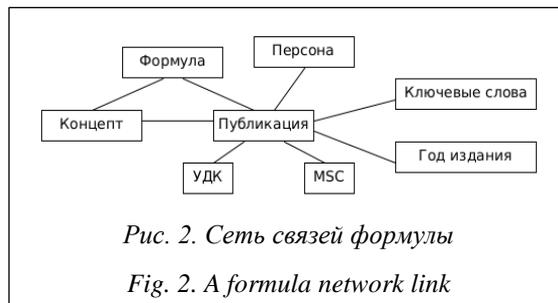


Рис. 2. Сеть связей формулы

Fig. 2. A formula network link

раздела «композиционные материалы», соответствующая понятию «Ламе обобщенная краевая задача для градиентной теории упругости изотропных тел».

На рисунке 3 представлен фрагмент связей понятия «уравнение Ламе». В Примерах 1 и 2 приведены статьи тезауруса для дескрипторов «уравнение Ламе» и «Ламе обобщенная краевая задача для градиентной теории упругости изотропных тел».

Структура статьи тезауруса включает в себя название (D – descriptor), ссылки на синонимы (SYN – synonym), ссылки на связанные понятия (связь с более общим понятием обозначается как BT – broader term, связь с более узким понятием как NT – narrower term), ссылки на литературу (REF – references), на которую опирались при составлении тезауруса. Статья тезауруса также включает в себя набор математических формул (Math), связанных с этим понятием, ключевые слова (KW), дополнительную

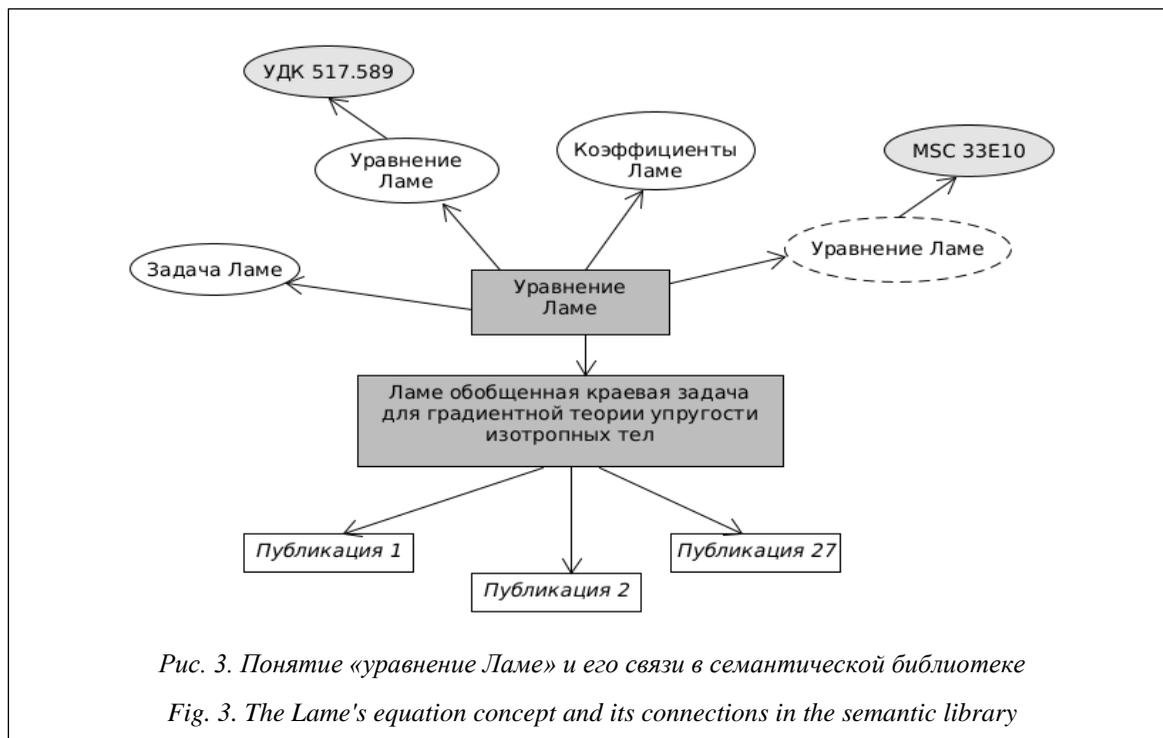


Рис. 3. Понятие «уравнение Ламе» и его связи в семантической библиотеке

Fig. 3. The Lamé's equation concept and its connections in the semantic library

Пример 1. Дескриптор «Уравнение Ламе»

Example 1. The Lamé Equation descriptor

D:	Уравнение Ламе
SYN:	Lame equation
SYN:	Уравнение равновесия в перемещениях
SYN:	Balance equation in displacements
BT:	Уравнения теории упругости
NT:	Ламе обобщенная краевая задача для градиентной теории упругости изотропных тел
Math:	$\left[\begin{array}{l} (\lambda + \mu) \frac{\partial \Theta}{\partial x} + \mu \Delta u_x = 0 \\ (\lambda + \mu) \frac{\partial \Theta}{\partial y} + \mu \Delta u_y = 0 \\ (\lambda + \mu) \frac{\partial \Theta}{\partial z} + \mu \Delta u_z = 0 \end{array} \right]$ $\lambda = \frac{E \nu}{(1+\nu)(1-2\nu)}, \mu = G, \Delta - \text{Laplace operator}$

Пример 2. Дескриптор «Ламе обобщенная краевая задача для градиентной теории упругости изотропных тел»

Example 2. The descriptor Lamé generalized boundary value problem for the gradient theory of elasticity of isotropic bodies

D:	Ламе обобщенная краевая задача для градиентной теории упругости изотропных тел
SYN:	Ламе обобщенная краевая задача для градиентной теории упругости изотропных тел с микроструктурой
BT:	Уравнение Ламе
MATH:	$L_{ij}(\delta_{ik} R_k - \frac{1}{C} L_{jk}(R_k)) + \rho F_i = 0$, где δ_{ik} – дельта Кронекера или $L_{ij}(H_{jk}(R_k)) + \rho F_i = 0$, $H_{jk}(\cdot) = \frac{1}{C} L_{jk}(\cdot) + \delta_{ik}(\cdot)$, где $H_{jk}(\cdot)$ – оператор Гельмгольца для оператора Ламе (обобщенный оператор Гельмгольца)
MATH:	$L(\vec{R}) = \mu \nabla^2 \vec{R} + (\mu + \lambda) \nabla \text{div} \vec{R}$, $L_c(\vec{R}) = L(\vec{R}) - C \vec{R}$
REF:	Волков-Богородский Д.Б., Евтушенко Ю.Г., Зубов В.И., Лурье С.А. Численно-аналитический учет масштабных эффектов при расчете деформаций нанокompозитов с использованием блочного метода мультиполей // Вычислительная математика и математическая физика. 2006. Т. 46. № 7. С. 1318–1337
REF:	Ильюшин А.А. Механика сплошной среды. М.: Изд-во Моск. ун-та, 1990. 310 с.
KW:	Ламе уравнение обобщенное, краевая задача для Ламе уравнения обобщенного, математическая модель межфазного слоя в механике материалов, модель композитных структур с микро- и нановключениями, модель тонкопленочных наноструктур, представление Нейбера–Папковича обобщенное, Ламе неоднородное уравнение динамики, Ламе однородное уравнение, Ламе оператор, векторная форма Ламе уравнения, индексная форма Ламе уравнения, решение уравнения Ламе представление Галеркина, Ламе волновое уравнение, Ламе полином, Ламе функция
KW:	Lame equation generalized, boundary value problem for the Lamé equation generalized, mathematical model of the interfacial layer in mechanics of materials, model of composite structures with micro- and nano-inclusions, model of thin-film nanostructures, generalized Neuber–Papkovich representation, Lamé inhomogeneous equation of dynamics, Lamé homogeneous equation, Lamé operator, vector form of the Lamé equation, index form of the Lamé equation, solution of the Lamé equation Galerkin representation, Lamé wave equation, Lamé polynomial, Lamé function
NOTE:	Ламе уравнение обобщенное и краевая задача для Ламе уравнения обобщенного определяют математическую модель межфазного слоя в механике материалов или модель композитных структур с микро- и нановключениями и модель тонкопленочных наноструктур
NOTE:	The Lamé equation generalized and the boundary value problem for the Lamé equation generalized determines the mathematical model of the interfacial layer in the mechanics of materials or the model of composite structures with micro- and nano-inclusions and the model of thin-film nanostructures

информацию, относящуюся к дескриптору, но не входящую в стандартную статью тезауруса (NOTE).

Набор ключевых слов из статьи тезауруса составлен на основе связанных справочников и словарей, которые были перечислены ранее. Эти ключевые слова наряду с терминами, опре-

деленными при описании понятия тезауруса, используются для выявления связанных материалов из контента библиотеки для выявления связей при анализе текста.

В первую очередь были выявлены подмножества понятий из Математической энциклопедии и словаря специальных функций для поня-

тия «уравнение Ламе». Эти понятия, в свою очередь, связаны с элементами отраслевых классификаторов, которые могут использоваться в качестве рекомендаций при классификации публикаций, не содержащих такой информации.

Создание подпространства композиционных материалов на базе журнала МКМК

Рассмотрим этапы построения в LibMeta подпространства предметной области «математика» и выделение связей в ней. На базе Математической энциклопедии с помощью классификаторов выделяются термины, которые относятся к тематике подпространства и используются в качестве каркаса для построения тезауруса этой подобласти. Этапы построения подпространства предметной области представлены на рисунке 4.

Этап 1. На первом этапе использование математической энциклопедии позволило охватить все разнообразие основных терминов предметной области «математика» и предварительно очертить ее границы (в энциклопедии). Статьи электронной версии энциклопедии представляют собой неструктурированный текст и формулы в виде картинок, не содержат каких-либо ссылок на связанные статьи энциклопедии или других ресурсов, не имеют указаний на раздел математики. Для описания энциклопедии была проведена работа по предварительной обработке статей для структуризации ее контента и выделению семантических

связей в автоматическом режиме. Необходимость подобной разметки продиктована тем, что для пользователя электронной версии ценность представляют не только сами статьи, но и возможность навигации по библиотеке, поиска интересующих материалов и связанных с ними данных. В таком виде пользователь имеет возможность строить запросы к массиву данных, наблюдать связи между математическими понятиями, публикациями и авторами.

Этап 2. Для определения семантики ресурсов разработаны различные виды классификации, которые отличаются друг от друга охватом предметных областей и степенью гранулярности при их классификации. Для тематического описания предметной области используются широко распространенные классификаторы, такие как УДК (универсальная десятичная классификация) и ГРНТИ (Государственный рубрикатор научно-технической информации). Эти классификаторы охватывают почти все области научного знания и перечень понятий, характерных для этих областей. Обычно понятия носят довольно общий характер и не отражают разнообразия направлений в каждой отдельной области научного знания.

Классификатор MSC обеспечивает более детальную классификацию ресурсов и более точное соотношение смысловых понятий ресурсов с определенным направлением области знания для предметной области «математика».

На текущий момент электронная версия английской энциклопедии поддерживается международным издательством Springer и доступна в сети. В Encyclopedia of Mathematics статьи

Этап 1 Без классификации по подразделам	Предметная область «математика»				
	Тезаурус «Математическая энциклопедия»				
Этап 2 Классификация и выделение подразделов	Классификаторы				
	MSC		УДК		ГРНТИ
	Подпространства предметной области «математика»				
	Дифференциальные уравнения	Математическая физика	Вычислительная математика	Математическое моделирование	...
Этап 3 Выделение терминов для подразделов	Тезаурусы и словари подпространств предметной области «математика»				
	Тезаурус ОДУ	Словарь спец. функций	Словарь уравнений смешанного типа	Словарь механики сплошной среды	Словарь механики композитов
	Словарь фуллерены	Словарь нанотрубки	Словарь полимерных композитов	...	
Этап 4 Классификация источников данных и обогащение описания предметной области	Источники данных				
	Неструктурированные			Структурированные	
	Журналы	Публикации	...	Репозитории	Mathnet

Рис. 4. Этапы построения подпространства предметной области

Fig. 4. The stages of building a subject area subspace

снабжены формулами в формате TeX, пригодном для машинной обработки, ссылками на связанные статьи энциклопедии. Каждой статье сопоставлен индекс MSC, который используется для классификации по разделам математики. В совокупности эти метаданные открывают широкие возможности по поиску интересующих статей и изучению смежных тем. Таким образом, сопоставление классификаторов и понятий Математической энциклопедии позволяет выделить для некоторого раздела математики множество понятий, которые составят каркас его тезауруса. Также за счет назначения понятиям энциклопедии элементов различных классификаторов появляются новые связи между классификаторами.

Этап 3. В отличие от второго этапа, на котором благодаря классификаторам на основе терминологической базы Математической энциклопедии формируется каркас тезауруса некоторой подобласти, на третьем этапе описание предметной области углубляется и уточняется за счет подключения специализированных тезаурусов и словарей для некоторой подобласти. Естественно, что эти тезаурусы и словари связываются с каркасными понятиями из Математической энциклопедии. Несмотря на то, что включаемые тезаурусы разрабатываются или разрабатывались отдельно от энциклопедии, терминологическое пересечение в ключевых понятиях всегда присутствует. Эти ключевые понятия становятся вершинами новых веток тезауруса предметной области за счет подключения новых тезаурусов/словарей. Классификаторы, используемые на втором этапе для распределения понятий энциклопедии по областям, также используются и для подключаемых понятий и тезаурусов, создавая новые тематические связи в предметной области.

При построении подпространства «позиционные материалы» в качестве источника данных рассматривались публикации журнала МКМК, тесно связанного с такими словарями и тезаурусами, как словарь спецфункций, тезаурус ОДУ, тезаурус матфизики и т.д., поэтому на третьем этапе появляются соответствующие словари и тезаурусы.

Этап 4. Используются структурированные источники данных, такие как БД, и неструктурированные в виде полных текстов статей журналов. Также встречаются источники, данные которых частично структурированы. Для каждого источника определяется модель его данных, устанавливаются связи с тезаурусом предметной области, который формировался на

предыдущих этапах. Основным источником данных – это метаданные математических статей и сами полные тексты. Модель данных таких источников может быть углублена за счет учета модели самих текстов. Это позволяет проводить более точный анализ и обработку извлекаемых из текстов данных/знаний.

Помимо подключения и загрузки данных, основная задача этого этапа – классификация и выявление связей с расширенным на предыдущих этапах тезаурусом предметной области. Таким образом, появляются новые междисциплинарные связи в данных, которые, в свою очередь, обогащают описание предметной области. Полные тексты также используются для извлечения новых понятий предметной области.

Заключение

Работа посвящена проблеме семантического описания подраздела предметной области в рамках цифровой библиотеки, ресурсы которой представлены в виде плохо (или недостаточно) структурированной информации, в виде текстов архивных статей национальных специализированных журналов и различных словарей и тезаурусов. Довольно часто отсутствие такого описания приводит к тому, что специфические знания, содержащиеся в подразделе предметной области, бывают невидимыми, хотя и имеют научную ценность как часть прикладной области. В представленной работе показано, как для одного из прикладных разделов задач математической физики реализована процедура включения массивов публикаций журнала МКМК в онтологию семантической библиотеки на основе имеющихся источников данных, контента библиотеки и специфических словарей и тезаурусов. Получены связи нового массива публикаций с Математической энциклопедией и классификаторами, выделены основные понятия локальной предметной области журнала, ключевые слова, что позволит составить локальный тезаурус журнала и представить семантическое описание соответствующего подраздела предметной области.

Таким образом, была решена задача включения предметной области журнала в контент семантической библиотеки.

Важно отметить, что архивные статьи обретают дополнительные свойства (при индексации в библиотеке дополнены ключевыми словами, кодами классификаторов и т.д.), могут быть представлены в поисковых запросах и

найлены благодаря полученным семантическим связям.

На базе нескольких математических журналов сейчас формируется корпус размеченных текстов из разных разделов математики. Такой корпус текстов необходим для решения возникающих задач, в том числе методами машинного обучения, оценки качества построенных решений. Использовать этот корпус планируется в следующих задачах:

- создание контента семантически связанной экспертной информации о математических предметных областях (энциклопедий, тезаурусов, классификаторов, предметных и авторских указателей из рецензируемых изданий);
- создание алгоритмов и технологии систематизации данных в соответствии с предметными математическими областями на основе онтологического подхода;
- использование алгоритмов машинного обучения для определения предметной и тема-

тической принадлежности публикаций, их схожести и авторства;

- применение алгоритмов искусственного интеллекта для выявления информационной потребности пользователя для динамического формирования и настройки интерфейса на определенную предметную область контента математической библиотеки;
- использование алгоритмов расширенного поискового запроса для поиска приложений в предметной области;
- применение методов векторного анализа для выявления дополнительных связей терминов и актуализации описания предметных областей (насыщения) в контенте семантической библиотеки;
- создание архивов общего доступа для размещения достоверных знаний по предметным областям.

Дальнейшие исследования предполагается проводить в направлении этих задач.

Работа представлена в рамках выполнения темы НИР 0063-2019-0003 ФИЦ ИУ РАН.

Литература

1. Hlava M.M.K. The taxobook: History, theories, and concepts of knowledge organization. Pt. 1. In: SLICRS, 2014, vol. 6, no. 3, pp. 1–80. DOI: 10.2200/S00602ED1V01Y201410ICR035.
2. Hlava M.M.K. The taxobook: Principles and practices of building taxonomies. Pt. 2. In: SLICRS, 2014, vol. 6, no. 4, pp. 1–164. DOI: 10.2200/S00603ED1V02Y201410ICR036.
3. Hlava M.M.K. The taxobook: Applications, implementation, and integration in search. Pt. 3. In: SLICRS, 2014, vol. 6, no. 4, pp. 1–156. DOI: 10.2200/S00604ED1V03Y201410ICR037.
4. Ataeva O.M., Serebryakov V.A., Tuchkova N.P. Mathematical physics branches: Identifying mixed type equations. *Lobachevskii J. of Math.*, 2019, vol. 40, no. 7, pp. 876–886. DOI: 10.1134/S1995080219070047.
5. Ataeva O., Serebryakov V.A., Sinelnikova E. Thesaurus and ontology building for semantic library based on mathematical encyclopedia. *Proc. DAMDID/RCDL*, 2019, pp. 148–157.
6. Атаева О.М., Серебряков В.А. Онтология цифровой семантической библиотеки LibMeta // Информатика и ее применения. 2018. Т. 12. № 1. С. 2–10. DOI: 10.14357/19922264180101.
7. Wojanowski P., Grave E., Joulin A., Mikolov T. Enriching word vectors with subword information. *Trans. Assoc. Comput. Linguist.*, 2017, vol. 5, pp. 135–146. DOI: 10.1162/tacl_a_00051.
8. Goodfellow I., Bengio Y., Courville A. *Deep Learning*. Cambridge, MIT Press Publ., 2016, 787 p.
9. Bengio Y., LeCun Y., Hinton G. Deep learning. *Nature*, 2015, vol. 521, no. 7553, pp. 436–444. DOI: 10.1038/nature14539.
10. Cohl H., Schubotz M., Youssef A., Greiner-Petter A., Gerhard J. et al. Semantic preserving bijective mappings of mathematical formulae between word processors and computer algebra systems. In: LNCS. *Proc. CICM*, 2017, pp. 115–131.
11. Trott M., Weisstein E.W. Computational knowledge of continued fractions. *WolframAlpha*, 2013. URL: <http://blog.wolframalpha.com/2013/05/16/computational-knowledge-of-continued-fractions> (дата обращения: 29.11.2022).
12. Schubotz M., Grigorev A., Cohl H.S., Meuschke N. et al. Semantification of identifiers in mathematics for better math information retrieval. *Proc. 39th Int. ACM SIGIR Conf. Research and Development in Information Retrieval*, 2016, pp. 135–144. DOI: 10.1145/2911451.2911503.
13. Cohl H.S., Schubotz M., McClain M.A., Saunders B.V., Zou C.Y., Mohammed A.S., Danoff A.A. Growing the digital repository of mathematical formulae with generic LaTeX sources. In: LNCS. *Proc. CICM*, 2015, pp. 280–287. DOI: 10.1007/978-3-319-20615-8_18.

14. Miller B.R. Drafting DLMF content dictionaries. Proc. OpenMath Workshop of the 9th CICM, 2016.
15. Pathak A., Das R., Pakray P., Gelbukh A. Extracting context of math formulae contained inside scientific documents. *Computación y Sistemas*, 2019, vol. 23, no. 3, pp. 803–818. DOI: 10.13053/cys-23-3-3246.
16. Serebryakov V.A., Ataeva O.M. Ontology based approach to modeling of the subject domain "Mathematics" in the digital library. *Lobachevskii J. of Math.*, 2021, vol. 42, no. 8, pp. 1920–1934. DOI: 10.1134/S199508022108028X.
17. Allemang D., Hendler J., Gandon F. *Semantic Web for the Working Ontologist*. ACM Books Publ., 2020, 512 p. DOI: 10.1016/b978-0-12-373556-0.x0001-9.
18. Elizarov A.M., Kirillovich A.V., Lipachev E.K., Nevzorova O.A., Solovyev V.D., Zhiltsov N.G. Mathematical knowledge representation: semantic models and formalisms. *Lobachevskii J. of Math.*, 2014, vol. 35, no. 4, pp. 348–354. DOI: 10.1134/S1995080214040143.
19. Елизаров А.М., Жижченко А.Б., Жильцов Н.Г., Кириллович А.В., Липачев Е.К. Онтологии математического знания и рекомендательная система для коллекций физико-математических документов // Докл. Академии наук. 2016. Т. 467. № 4. С. 392–395. DOI: 10.7868/S0869565216100042.
20. Паринов С.И., Коголовский М.Р. Семантическое структурирование контента научных электронных библиотек на основе онтологий // RCDL: тр. XIII Всерос. научн. конф. 2011. С. 94–103.
21. Серебряков В.А. Что такое семантическая цифровая библиотека // RCDL: тр. XVI Всерос. научн. конф. 2014. С. 21–25.
22. Weinan E. The Dawning of a new era in applied mathematics. *Notices of the American Mathematical Society*, 2021, vol. 68, no. 4, pp. 565–571. DOI: 10.1090/NOTI2259.
23. Vasiliev V.V., Morozov E.V. *Advanced Mechanics of Composite Materials and Structures*. Elsevier Publ., 2018, 856 p.
24. Bravo M., Hoyos Reyes L.F., Reyes Ortiz J.A. Methodology for ontology design and construction. *Contaduría y Administración*, 2019, vol. 64, no. 4, pp. 1–24. DOI: 10.22201/fca.24488410e.2020.2368.
25. Ataeva O.M., Serebryakov V.A., Tuchkova N.P. Search query extension semantics. *Ceur Workshop Proceedings. Proc. XXIII Int. Conf. DAMDID/RCDL*, 2021, vol. 036, pp. 325–339.
26. Ataeva O.M., Serebryakov V.A., Tuchkova N.P. Creating the applied subject area ontology by means of the content of the digital semantic library. *Lobachevskii J. of Math.*, 2022, vol. 43, no. 7, pp. 1557–1566. DOI: 10.1134/S1995080222100043.
27. Scharpf P., Schubotz M., Youssef A., Hamborg F., Meuschke N., Gipp B. Classification and clustering of arxiv documents, sections, and abstracts, comparing encodings of natural and mathematical language. *Proc. JCDL*, 2020, pp. 137–146. DOI: 10.1145/3383583.3398529.
28. Youssef A., Miller B. Deep learning for math knowledge processing. *Proc. XI Int. Conf. CICM*, 2018, pp. 271–286. DOI: 10.1007/978-3-319-96812-4_23.
29. Mihaljević-Brandt H., Kohlhase M., Teschke O., Sperber W. Mathematical formula search. *European Math. Soc. Newsletter*, 2013, vol. 89, pp. 56–58.
30. Pathak A., Pakray P., Gelbukh A. A formula embedding approach to math information retrieval. *Computación y Sistemas*, 2018, vol. 22, no. 3, pp. 819–833. DOI: 10.13053/cys-22-3-3015.
31. Hong S., Su W., Lin H. and Lv X. Functional classification study for mathematical formulas retrieval. *Proc. XVII IEEE/ACIS Int. Conf. SNPD*, 2016, pp. 99–104. DOI: 10.1109/SNPD.2016.7515885.

Integrating the subject area subspace into the “mathematics” semantic space

O.M. Ataeva¹, Ph.D. (Engineering), Research Associate, oli.ataeva@gmail.com

V.A. Serebryakov¹, Dr.Sc. (Physics and Mathematics), Professor, Head of Department, serebrvas@gmail.com

N.P. Tuchkova¹, Ph.D. (Physics and Mathematics), Senior Researcher, natalia_tuchkova@mail.ru

¹ Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences (FRC CSC RAS), Moscow, 119333, Russian Federation

Abstract. The paper studies the problem of developing a semantic library by adding a new applied scientific area. The authors use the example of a journal on applied issues of composite materials in order to build an

addition to the main content of the library, to expand the description of the original subject area, and to detail UDC and MSC articles corresponding to the local subject area. At the same time, they solve the tasks of adding terms to the thesaurus, building a reference corpus of the applied subject area of mathematics, and creating a custom interface. Formulas and equations of the local subject area are semantically linked to the main content of the library.

The main advantage of using semantic libraries for this kind of tasks is to enrich the existing knowledge base of the library and identify relationships in data. To study these problems, it is necessary to interact with subject matter experts and use modern tools and methods for natural language processing, machine learning approaches to knowledge representation. The above tasks are implemented by means of the LibMeta semantic library, which allows considering such concepts as the thematic area/subdomain and their hierarchical relationships, in particular, the automatic inclusion of the enclosing area concepts in the subdomain knowledge system; the possibility of personalizing the built structures and using LibMeta as a construction tool, in particular, building the ontology itself. Data integration within the library allows expanding the description of subject areas related to the applications of mathematics in interdisciplinary research and technology. As an example of one of the applied sections of mathematical physics problems, the procedure for including specific dictionaries, thesauri, as well as an array of publications of a specialized journal in a semantic library is shown. The proposed approach will allow using the content of the “mathematics” semantic library for scientific research and minimizing the process of searching for information in the local subject area, without losing more general results contained outside this area.

Keywords: semantic library, applied ontology, local thesaurus, local subject domain search.

Acknowledgements. *The work is presented as a part of implementing the research topic 0063-2019-0003 FRC CSC RAS.*

References

1. Hlava M.M.K. The taxobook: History, theories, and concepts of knowledge organization. Pt. 1. In: *SLICRS*, 2014, vol. 6, no. 3, pp. 1–80. DOI: 10.2200/S00602ED1V01Y201410ICR035.
2. Hlava M.M.K. The taxobook: Principles and practices of building taxonomies. Pt. 2. In: *SLICRS*, 2014, vol. 6, no. 4, pp. 1–164. DOI: 10.2200/S00603ED1V02Y201410ICR036.
3. Hlava M.M.K. The taxobook: Applications, implementation, and integration in search. Pt. 3. In: *SLICRS*, 2014, vol. 6, no. 4, pp. 1–156. DOI: 10.2200/S00604ED1V03Y201410ICR037.
4. Ataeva O.M., Serebryakov V.A., Tuchkova N.P. Mathematical physics branches: Identifying mixed type equations. *Lobachevskii J. of Math.*, 2019, vol. 40, no. 7, pp. 876–886. DOI: 10.1134/S1995080219070047.
5. Ataeva O., Serebryakov V.A., Sinelnikova E. Thesaurus and ontology building for semantic library based on mathematical encyclopedia. *Proc. DAMDID/RCDL*, 2019, pp. 148–157.
6. Ataeva O.M., Serebryakov V.A. Ontology of the digital semantic library LibMeta. *Informatics and its Applications*, 2018, vol. 12, no. 1, pp. 2–10. DOI: 10.14357/19922264180101 (in Russ.).
7. Bojanowski P., Grave E., Joulin A., Mikolov T. Enriching word vectors with subword information. *Trans. Assoc. Comput. Linguist.*, 2017, vol. 5, pp. 135–146. DOI: 10.1162/tacl_a_00051.
8. Goodfellow I., Bengio Y., Courville A. *Deep Learning*. Cambridge, MIT Press Publ., 2016, 787 p.
9. Bengio Y., LeCun Y., Hinton G. Deep learning. *Nature*, 2015, vol. 521, no. 7553, pp. 436–444. DOI: 10.1038/nature14539.
10. Cohl H., Schubotz M., Youssef A., Greiner-Petter A., Gerhard J. et al. Semantic preserving bijective mappings of mathematical formulae between word processors and computer algebra systems. In: *LNCS. Proc. CICM*, 2017, pp. 115–131.
11. Trott M., Weisstein E.W. Computational knowledge of continued fractions. *WolframAlpha*, 2013. Available at: <http://blog.wolframalpha.com/2013/05/16/computational-knowledge-of-continued-fractions> (accessed November 29, 2022).
12. Schubotz M., Grigorev A., Cohl H.S., Meuschke N. et al. Semantification of identifiers in mathematics for better math information retrieval. *Proc. 39th Int. ACM SIGIR Conf. Research and Development in Information Retrieval*, 2016, pp. 135–144. DOI: 10.1145/2911451.2911503.
13. Cohl H.S., Schubotz M., McClain M.A., Saunders B.V., Zou C.Y., Mohammed A.S., Danoff A.A. Growing the digital repository of mathematical formulae with generic LaTeX sources. In: *LNCS. Proc. CICM*, 2015, pp. 280–287. DOI: 10.1007/978-3-319-20615-8_18.
14. Miller B.R. Drafting DLMF content dictionaries. *Proc. OpenMath Workshop of the 9th CICM*, 2016.

15. Pathak A., Das R., Pakray P., Gelbukh A. Extracting context of math formulae contained inside scientific documents. *Computación y Sistemas*, 2019, vol. 23, no. 3, pp. 803–818. DOI: 10.13053/cys-23-3-3246.
16. Serebryakov V.A., Ataeva O.M. Ontology based approach to modeling of the subject domain "Mathematics" in the digital library. *Lobachevskii J. of Math.*, 2021, vol. 42, no. 8, pp. 1920–1934. DOI: 10.1134/S199508022108028X.
17. Allemang D., Hendler J., Gandon F. *Semantic Web for the Working Ontologist*. ACM Books Publ., 2020, 512 p. DOI: 10.1016/b978-0-12-373556-0.x0001-9.
18. Elizarov A.M., Kirillovich A.V., Lipachev E.K., Nevzorova O.A., Solovyev V.D., Zhiltsov N.G. Mathematical knowledge representation: semantic models and formalisms. *Lobachevskii J. of Math.*, 2014, vol. 35, no. 4, pp. 348–354. DOI: 10.1134/S1995080214040143.
19. Elizarov A.M., Zhizhchenko A.B., Zhiltsov N.G., Kirillovich A.V., Lipachev E.K. Ontologies of mathematical knowledge and a recommender system for collections of physical and mathematical documents. *Doklady Akademii Nauk*, 2016, vol. 467, no. 4, pp. 392–395 (in Russ.).
20. Parinov S.I., Kogalovsky M.R. A technology for semantic structuring of scientific digital library content. *Proc. XIII Sci.Conf. RCDL*, 2011, pp. 94–103 (in Russ.).
21. Serebryakov V.A. Semantic digital libraries. What is it? *Proc. XVI Sci.Conf. RCDL*, 2014, pp. 21–25 (in Russ.).
22. Weinan E. The Dawning of a new era in applied mathematics. *Notices of the American Mathematical Society*, 2021, vol. 68, no. 4, pp. 565–571. DOI: 10.1090/NOTI2259.
23. Vasiliev V.V., Morozov E.V. *Advanced Mechanics of Composite Materials and Structures*. Elsevier Publ., 2018, 856 p.
24. Bravo M., Hoyos Reyes L.F., Reyes Ortiz J.A. Methodology for ontology design and construction. *Contaduría y Administración*, 2019, vol. 64, no. 4, pp. 1–24. DOI: 10.22201/fca.24488410e.2020.2368.
25. Ataeva O.M., Serebryakov V.A., Tuchkova N.P. Search query extension semantics. Ceur Workshop Proceedings. *Proc. XXIII Int. Conf. DAMDID/RCDL*, 2021, vol. 036, pp. 325–3399.
26. Ataeva O.M., Serebryakov V.A., Tuchkova N.P. Creating the applied subject area ontology by means of the content of the digital semantic library. *Lobachevskii J. of Math.*, 2022, vol. 43, no. 7, pp. 1557–1566. DOI: 10.1134/S1995080222100043.
27. Scharpf P., Schubotz M., Youssef A., Hamborg F., Meuschke N., Gipp B. Classification and clustering of arxiv documents, sections, and abstracts, comparing encodings of natural and mathematical language. *Proc. JCDL*, 2020, pp. 137–146. DOI: 10.1145/3383583.3398529.
28. Youssef A., Miller B. Deep learning for math knowledge processing. *Proc. XI Int. Conf. CICM*, 2018, pp. 271–286. DOI: 10.1007/978-3-319-96812-4_23.
29. Mihaljević-Brandt H., Kohlhase M., Teschke O., Sperber W. Mathematical formula search. *European Math. Society Newsletter*, 2013, vol. 89, pp. 56–58.
30. Pathak A., Pakray P., Gelbukh A. A formula embedding approach to math information retrieval. *Computación y Sistemas*, 2018, vol. 22, no. 3, pp. 819–833. DOI: 10.13053/cys-22-3-3015.
31. Hong S., Su W., Lin H. and Lv X. Functional classification study for mathematical formulas retrieval. *Proc. XVII IEEE/ACIS Int. Conf. SNPD*, 2016, pp. 99–104. DOI: 10.1109/SNPD.2016.7515885.

Для цитирования

Атаева О.М., Серебряков В.А., Тучкова Н.П. Интеграция подпространства предметной области в семантическое пространство «математика» // Программные продукты и системы. 2023. Т. 36. № 1. С. 083–096. DOI: 10.15827/0236-235X.141.083-096.

For citation

Ataeva O.M., Serebryakov V.A., Tuchkova N.P. Integrating the subject area subspace into the "mathematics" semantic space. *Software & Systems*, 2023, vol. 36, no. 1, pp. 083–096 (in Russ.). DOI: 10.15827/0236-235X.141.083-096.