



О.М. Атаева, В.А. Серебряков, Н.П.
Тучкова

**Опыт построения онтологии для
предметной области на основе
контента тематического журнала**

Рекомендуемая форма библиографической ссылки

Атаева О.М., Серебряков В.А., Тучкова Н.П. Опыт построения онтологии для предметной области на основе контента тематического журнала // Научный сервис в сети Интернет: труды XXIV Всероссийской научной конференции (19-22 сентября 2022 г., онлайн). — М.: ИПМ им. М.В.Келдыша, 2022. — С. 52-68.

<https://doi.org/10.20948/abrau-2022-35>

<https://keldysh.ru/abrau/2022/theses/35.pdf>

Видеозапись выступления

Опыт построения онтологии для предметной области на основе контента тематического журнала

О.М. Атаева¹, В.А. Серебряков¹, Н.П.Тучкова¹

¹ФИЦ ИУ РАН, Москва, ул. Вавилова, 40

Аннотация. Работа посвящена проблеме построения онтологии прикладной предметной области как части семантической библиотеки. Исходные данные представлены публикациями научного тематического журнала, массивом архивных и текущих версий. Рассматриваются основные этапы формирования онтологии в соответствии с определением и требованиями языков описания онтологий. Приводятся примеры структуры статей тезауруса.

Ключевые слова: прикладная онтология, тезаурус предметной области, источники данных, разработка онтологии

Experience in building an ontology for a subject area based on the content of a thematic journal

O.M.Ataeva¹, V.A.Serebryakov¹, N.P.Tuchkova¹

¹FRS «Computer Sciences and Control», Russian Academy of Sciences,
Vavilov str., 40, Moscow, 119333, Russia

Abstract. The work is devoted to the problem of constructing an ontology of an applied subject area as part of a semantic library. The initial data are presented by publications of a scientific thematic journal, an array of archival and current versions. The main stages of ontology formation are considered in accordance with the definition and requirements of Web Ontology Language. Examples of the structure of thesaurus articles are given.

Keywords: applied ontology, domain thesaurus, data sources, ontology development

Введение

За несколько десятилетий цифровизации термин «онтология»¹ проник в различные области знаний. Многие исследователи, ученые и специалисты из разных областей знаний 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 [1, 2].

По определению² [1-5] обязательными свойствами описания знаний в виде онтологии являются;

(1) конечный и контролируемый словарь понятий и терминов, исключающий их двусмысленную интерпретацию;

(2) строгая иерархия отношений подклассов понятий и терминов, которые описывают знания предметной области

На сегодня разработано огромное количество онтологий для предметных областей, которые охватывают разные аспекты деятельности человека. В этой работе рассматривается 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 имеющихся уже источников данных и контента библиотеки [6]. Поскольку круг этих задач необъятен, то уместно привести цитату академика Василия Сергеевича Владимирова, которому принадлежат фундаментальные труды в этой области:

«Математическая физика – это теория математических моделей физических явлений. Она относится к математическим наукам; критерий истины в ней – математическое доказательство. Однако, в отличие от чисто математических наук, в математической физике исследуются физические задачи на математическом уровне, а результаты представляются в виде теорем, графиков, таблиц и т.д. и получают физическую интерпретацию. При таком широком понимании математической физики к ней следует относить и такие разделы механики, как теоретическая механика, гидродинамика и теория упругости»³.

Построение модели онтологии предметной области позволяет выделить *метаданные для проектирования конкретных структур данных* научных предметных областей и варианты управления этими данными. Для этого необходимо структурировать и связать различные *ресурсы*,

¹ <https://www.w3.org/standards/semanticweb/ontology>

² <https://www.w3.org/TR/owl2-overview/>

³ Владимиров В.С. Что такое математическая физика? Препринт МИАН № НС-06-001, 2006.

извлечь из них и контекстуализировать (определить в контексте) *данные, придавая им свойства знаний* [2]. Для определения предметной области строится модель знаний, в рамках которой могут интегрироваться различные источники данных, использоваться различные таксономии⁴ понятий и терминов, верифицированные признанными экспертами научной области. В качестве базовой таксономии в работе была использована Математическая энциклопедия [7] и онтология научной предметной области, которая лежит в основе конструирования семантической библиотеки LibMeta [8] в совокупности с отраслевыми классификаторами MSC⁵ [9] и УДК⁶.

В работе исследуются задачи: как накопленные данные библиотеки можно использовать для описания конкретной области научных исследований в публикациях журнала «ММКМ»⁷ (архив за последние 25 лет) и как встраиваются новые понятия из прикладной области математической физики в контент семантической библиотеки.

О проблемах построения онтологии в частном случае

Несмотря на долгую историю онтологического проектирования, еще нет единого стандарта для построения онтологий, но существуют различные методологии [10, 11, 12] и ГОСТ⁸, содержащий 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.

По сути, после определения целей и задач формирования онтологии выделяются три основных процесса:

- (1) сбор данных из источников, выявления их структуры;
- (2) представление знаний в виде таксономий;

⁴ Таксономия (теория классификации) //Струнино-Тихорецк. М.: Советская энциклопедия, 1976.

⁵ <https://mathscinet.ams.org/msc/msc2010.html>

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

⁷ <https://mkmk.ras.ru>

⁸ https://allgosts.ru/35/060/gost_r_59798-2021

(3) реализация и построение на основе экспертных знаний формальной модели, пригодной для машинного использования в конкретной задаче [11].

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 неоднократно [1, 2, 7], существует определенный разрыв в представлении знаний в цифровых ресурсах и их отражении в библиографических ресурсах. Примерами интеграции математических знаний и публикаций служат такие ресурсы, как Zentralblatt⁹, англоязычная версия Математической энциклопедии¹⁰, LibMeta¹¹ и некоторые другие ресурсы. Многочисленные библиографические ресурсы [6, 13 - 16] служат примерами отдельных баз данных. Это связано, конечно, с тем, что цели этих разработок разные. Тем не менее, нельзя не отметить то обстоятельство, что для научного сообщества важно иметь именно *интеграцию данных*, чтобы, находясь в рамках цифровой предметной области, можно было бы ознакомиться с публикациями на выбранную тему. Такая возможность реализуема при создании семантической библиотеки, где наряду с терминологическими связями предметной области присутствуют связи с публикациями. Особенно это важно в междисциплинарных исследованиях, которые не всегда укладываются в классическую классификацию, что составляет зачастую 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». 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* - это использование цифровой библиотеки книг, журналов, публикаций и различных научных материалов, и ресурсов для научных и образовательных целей. Ресурсы онтологии, с одной стороны, являются *наполнением семантической библиотеки*, а с другой - являются

⁹ <https://zbmath.org>

¹⁰ <https://encyclopediaofmath.org/>

¹¹ <https://libmeta.ru>

источниками знаний в предметной области при определенной их классификации и категоризации.

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

Реализация этой цели в рамках семантической библиотеки позволит перейти к построению, в частности, тезауруса¹² 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, которые демонстрируют во времени развитие предметной области;
- учебники, монографии, словари и классификаторы, которые содержат основную терминологию, на которую опираются в этой предметной области.

В качестве первого типа источника мы использовали публикации тематического журнала¹³, а также тезаурусы, статьи энциклопедии и публикации, накопленные ранее в библиотеке LibMeta¹⁴.

В качестве второго типа источников использовались признанные экспертами предметной области книги, учебники, монографии и терминологические словари, в том числе написанные руководителями классических Российских математических школ:

- Монография академиков А.Н. Тихонова и А.А. Самарского [17]. Из этого источника было извлечено около 400 понятий, которые представляют собой описания основных задач.
- Описание предметной области механики сплошной среды опирается на классификацию академика Л.И. Седова [18]. Классификация Л.И. Седова используется как словарь основных терминов и включает в себя около 1200 терминов.
- Предметная область механики композитов опирается на классификацию академика РАН В.В. Васильева [19, 20, 21] на английском языке и включает около 2500 терминов, которые были извлечены и использованы как словарь рассматриваемой предметной области.

¹² ГОСТ 7.24-2007 Тезаурус информационно-поисковый многоязычный.
<https://ifap.ru/library/gost/7242007.pdf>

¹³ <https://mkmk.ras.ru>

¹⁴ <https://libmeta.ru>

- Словарь терминов из раздела «Фуллерены» и смежных областей [23] (соавтор и редактор академик Е.И. Моисеев). Основной акцент словаря делается на использование терминов в русскоязычных монографиях, учебной литературе, научных статьях и электронных источниках. Содержит свыше 850 терминов на русском языке.
- Словарь терминов из раздела «Нанотрубки» и смежных областей [22] (соавтор и редактор академик Е.И. Моисеев). Основной акцент словаря также делается на использование терминов в русскоязычных монографиях, учебной литературе, научных статьях и электронных источниках. Содержит свыше 1000 терминов на русском языке.
- Словарь полимерных композитов на основе ГОСТ 32794-2014¹⁵. Содержит около 500 терминов, которые представлены на нескольких языках,

Заметим, что работы Российских академиков Л.И. Седова и В.В. Васильева посвящены исследованиям 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 [17-22], классификаторы MSC, УДК, математическая энциклопедия.

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

О таксономии предметной области и структуре тезауруса

В процессе построения онтологии решается вопрос ее ограничения в рамках конкретной предметной области науки. Для этого вводится базовый набор понятий предметной области. Соответствующие термины предметной области связывают с этими понятиями. 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. Отдельно стоит упомянуть о том, что

¹⁵<https://docs.cntd.ru/document/1200113813>

тезаурус предметной области может быть, как результатом работы экспертной группы, так и построен автоматизированными средствами.

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

Из описания уравнений [17] выделяем основные типы связей различных лексико-семантических категорий, таких как:

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

Итак, *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,
 - именованное, нарицательное,
 - с переменными коэффициентами, с постоянными коэффициентами;
- список связей между терминами
 - иерархические: род, вид;
 - горизонтальные: синонимы, ассоциация;
- схема статьи тезауруса для описания понятия:
 - код понятия (идентификатор);
 - коды классификаторов;
 - дескриптор (заглавный термин);
 - недескрипторы (дополнительные термины);
 - символные 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. К таким объектам в рассматриваемом случае относятся *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», уже создана ранее [7]. Это позволяет использовать онтологию информационных ресурсов для описания публикаций, персон, задач и связанных с ними формул для описания данных, извлекаемых из новых источников данных, а также использовать тезаурус для представления таксономий предметной области.

При определении связей и понятий предметной области в рамках онтологии семантической библиотеки возможно три подхода:

1. Подход сверху-вниз, когда проектирование понятий и связей начинается с наиболее верхнеуровневых понятий.
2. Подход снизу-вверх, когда проектирование понятий и связей начинается с наиболее низкоуровневых данных, которые группируются и формируют более общие понятия и т.д.
3. И комбинированный подход, когда сформулированы основные понятия конкретной предметной области, ее данные очищены и частично-структурированы (например, представлены в виде отдельных таксономий) и необходимо связать разрозненные данные в рамках заданной предметной области, одновременно обогащая ее и специализируя и уточняя включаемые ресурсы.

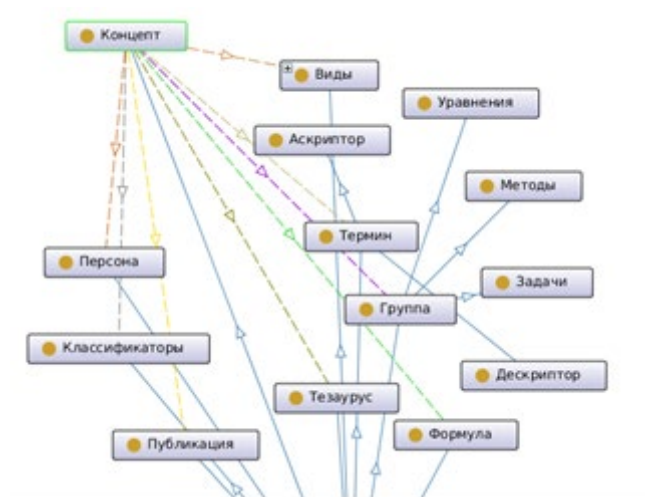


Рис. 1. Понятие "Концепт" и его связи.

В нашем случае использовался 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 был выполнен в рамках подготовки данных к загрузке.

Приведем список некоторых понятий, которые использовались для формирования онтологии тезауруса и отображены на рис. 1. Условно их можно разделить на 2 группы. Первая группа включает в себя понятия, необходимые для описания модели тезауруса, и соответствует по составу набору понятий стандарта для многоязычных тезаурусов¹⁶. Перечислим основные из них:

1. *Тезаурус,*
 2. *Концепт,*
 3. *Термин,*
- *Первостепенный термин,*
 - *Второстепенный термин.*

Также для описания тезауруса включены свойства для поддержки связей между концептами:

4. *Иерархические связи:*
- *Выше,*
 - *Ниже.*
5. *Горизонтальные:*
- *Ассоциация,*

¹⁶ ГОСТ 7.24-2007 <https://docs.cntd.ru/document/1200057506>

- *Синоним,*
- *Смотри также.*

И вторая группа, которая учитывая специфику тезауруса предметной области содержит следующие понятия:

6. *Группа:*

- *Методы,*
- *Задачи,*
- *Уравнения.*

7. *Виды:*

- *Эллиптический,*
- *Гиперболический,*
- *Параболический.*

8. *Формула*

9. *Персона* (для поддержки именных понятий (например, уравнений) предметной области).

10. *Публикация, и т.д.*

Все понятия второй группы связаны с понятием *Концепт* следующими связями:

1. *относится к группе,*
2. *имеет вид,*
3. *включает формулу,*
4. *связан с персоной,*
5. *связан с публикацией,*
6. *имеет код.*

Эти дополнительные связи и понятия позволяют ввести 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.

В качестве примера приведем небольшую часть тезауруса, посвященную задаче Ламе [24, 25, 26]. Это статья лексико-семантического указателя для предметной области «механика сплошной среды», раздела композиционные материалы соответствующая понятию «Ламе обобщенная краевая задача для градиентной теории упругости изотропных тел».



Рис. 2. Понятие "Уравнение Ламе" и его связи в семантической библиотеке

На рис. 2 представлен фрагмент связей понятия «Уравнение Ламе». И далее в двух 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, ссылки на синонимы, ссылки на связанные понятия, ссылки на литературу на которую опирались при составлении тезауруса. Статья тезауруса также включает в себя набор математических формул связанных с этим понятием.

Набор ключевых слов из статьи тезауруса составлен на основе связанных справочников и словарей, которые были перечислены ранее в разделе «Онтологизация». Эти ключевые слова наряду с дескриптором и аскрипотрами понятия используются в первую очередь для выявления связанных материалов из контента библиотеки для выявления связей при анализе текста.

Пример 1: Дескриптор «Уравнение Ламе»

:	Уравнение Ламе
YN:	Lame equation
	Уравнение равновесия в перемещениях

YN:	
YN:	Balance equation in displacements
T:	Уравнения теории упругости
T:	Ламе обобщенная краевая задача для градиентной теории упругости изотропных тел
ath:	$(\lambda + \mu) \frac{\partial \theta}{\partial x} + \mu \Delta u_x = 0$ $(\lambda + \mu) \frac{\partial \theta}{\partial y} + \mu \Delta u_y = 0$ $(\lambda + \mu) \frac{\partial \theta}{\partial z} + \mu \Delta u_z = 0$ $\lambda = \frac{E\nu}{(1+\nu)(1-2\nu)}, \mu = G, \Delta - \text{Laplace operator}$
ath:	$\left\{ \begin{array}{l} (\lambda + \mu) \frac{\partial \theta}{\partial x} + \mu \Delta u_x = 0 \\ (\lambda + \mu) \frac{\partial \theta}{\partial y} + \mu \Delta u_y = 0 \\ (\lambda + \mu) \frac{\partial \theta}{\partial z} + \mu \Delta u_z = 0 \end{array} \right. , \lambda = \frac{E\nu}{(1+\nu)(1-2\nu)}, \mu = G, \Delta - \text{Laplace operator}$

В первую очередь были выявлены подмножества понятий из математической энциклопедии и словаря специальных функций для понятия «уравнение Ламе». Эти понятия в свою очередь связаны с элементами отраслевых классификаторов, которые могут использоваться в качестве рекомендаций для использования при классификации публикаций, которые не содержат такой информации.

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

:	D	Ламе обобщенная краевая задача для градиентной теории упругости изотропных тел
YN:	S	Ламе обобщенная краевая задача для градиентной теории упругости изотропных тел с микроструктурой
T:	B	Уравнение Ламе
ATH:	M	Ламе уравнение обобщенное (индексная форма) $L_{ij} (\delta_{jk} R_k - \frac{1}{C} L_{jk} (R_k)) + \rho F_i = 0, \text{ где } \delta_{jk} - \text{дельта Кронекера}$

ATH:	M	<p>Ламе уравнение обобщенное (индексная форма)</p> $\left(H_{jk}(R_k) \right) + \rho F_i = 0, \quad H_{jk}(\dots) = -\frac{1}{C} L_{jk}(\dots) + \delta_{jk}(\dots)$ <p>где $H_{jk}(\dots)$ - оператор Гельмгольца для оператора Ламе</p>
ATH:	M	$\Delta_{ij}(\delta_{ik} R_k - \frac{1}{C} L_{jk}(R_k)) + \rho F_i = 0,$ <p>где Δ_{ij} - дельта Кронекера</p> <p>или $\Delta_{ij}(H_{jk}(R_k)) + \rho F_i = 0, \quad \Delta_{jk}(\dots) = \frac{1}{C} L_{jk}(\dots) + \delta_{jk}(\dots),$</p> <p>где $\Delta_{jk}(\dots)$ - оператор Гельмгольца для оператора Ламе (обобщенный оператор Гельмгольца)</p>
ATH:	M	<p>Ламе уравнение обобщенное (векторная форма):</p> $-\frac{1}{C} LL_c(\vec{R}) + \rho \vec{F} = 0, \quad L(\vec{R}) = \mu \nabla^2 \vec{R} + (\mu + \lambda) \nabla \text{div} \vec{R}$ $\vec{R}, \quad L_c(\vec{R}) = L(\vec{R}) - C \vec{R}$
ATH:	M	$-\frac{1}{C} LL_c(\vec{R}) + \rho \vec{F} = 0,$ $L(\vec{R}) = \mu \nabla^2 \vec{R} + (\mu + \lambda) \nabla \text{div} \vec{R},$ $L_c(\vec{R}) = L(\vec{R}) - C \vec{R}.$
EF:	R	<p>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.</p>
EF:	R	<p>Ильюшин А.А. Механика сплошной среды. М.: Изд-во Моск. ун-та, 1990. 310 с.</p>
W:	K	<p>Ламе уравнение обобщенное, краевая задача для Ламе уравнения обобщенного, математическая модель межфазного слоя в механике материалов, модель композитных структур с микро- и нано-включениями, 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.</p>
W:	K	<p>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</p>
OTE:	N	<p>Ламе уравнение обобщенное и краевая задача для Ламе уравнения обобщенного определяет математическую модель межфазного слоя в механике материалов или модель композитных структур с микро- и нано-включениями и модель тонкопленочных наноструктур.</p>
OTE:	N	<p>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-</p>

Заключение

Работа посвящена проблеме обработки плохо (или недостаточно) структурированной информации, к которой относятся архивные статьи национальных специализированных журналов. Довольно часто в них не содержатся выделенные разделы аннотаций, ключевых слов, классификаторов и других, стандартных на сегодняшний день, структурных частей научной публикации. Это затрудняет их поиск. Тем не менее, эти работы составляют определенную научную ценность как часть прикладной научной области. Современные средства семантической обработки позволяют проанализировать эти тексты и дополнить ими цифровую библиотеку в соответствующих предметных областях. В представленной работе показано, как для одного из прикладных разделов задач математической физики реализована процедура включения массивов публикаций журнала «ММКМ» в онтологию семантической библиотеки на основе имеющихся источников данных и контента библиотеки. Получены связи нового массива публикаций с математической энциклопедией, классификаторами, выделены основные понятия локальной предметной области журнала, ключевые слова, что позволит в дальнейшем составить локальный тезаурус журнала. 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.) и могут быть представлены в поисковых запросах и найдены благодаря полученным семантическим связям.

Работа выполнена при поддержке Российского фонда фундаментальных исследований, проект #22-21-00449.

Литература

1. Gruber T. Ontology of folksonomy: A mash-up of apples and oranges // International Journal on Semantic Web and Information Systems (IJSWIS). – 2007. – Т. 3. – №. 1. – С. 1-11.
2. Гаврилова Т. А., Хорошевский В. Ф. Базы знаний интеллектуальных систем. – Питер, 2000.
3. Hlava M. M. K. The Taxobook: History, Theories, and Concepts of Knowledge Organization, Part 1 of a 3-Part Series // Synthesis Lectures on Information Concepts, Retrieval, and Services. – 2014. – Т. 6. – №. 3. – С. 1-80.

4. Hlava M. M. K. The Taxobook: Principles and practices of building taxonomies, part 2 of a 3-part series // Synthesis Lectures on Information Concepts, Retrieval, and Services. – 2014. – Т. 6. – №. 4. – С. 1-164.
5. Hlava M. M. K. The Taxobook: Applications, Implementation, and Integration in Search: Part 3 of a 3-Part Series // Synthesis Lectures on Information Concepts, Retrieval, and Services. – 2014. – Т. 6. – №. 4. – С. 1-156.
6. Ataeva O. M., Serebryakov V. A., Tuchkova N. P. Mathematical physics branches: Identifying mixed type equations // Lobachevskii Journal of Mathematics. – 2019. – Т. 40. – №. 7. – С. 876-886.
7. Ataeva O., Serebryakov V. A., Sinelnikova E. Thesaurus and Ontology Building for Semantic Library Based on Mathematical Encyclopedia // DAMDID/RCDL. – 2019. – С. 148-157.
8. Serebryakov V. A., Ataeva O. M. Ontology Based Approach to Modeling of the Subject Domain “Mathematics” in the Digital Library // Lobachevskii Journal of Mathematics. – 2021. – Т. 42. – №. 8. – С. 1920-1934. <https://mathscinet.ams.org/msc/msc2010.html>.
9. Elizarov A. M. et al. Mathematical knowledge representation: semantic models and formalisms // Lobachevskii Journal of Mathematics. – 2014. – Т. 35. – №. 4. – С. 348-354.
10. Bravo M., Hoyos Reyes L. F., Reyes Ortiz J. A. Methodology for ontology design and construction // Contaduría y administración. – 2019. – Т. 64. – №. 4.
11. Allemang D., Hendler J., Gandon F. Semantic web for the working ontologist. – Association for Computing Machinery, 2020.
12. Lassila, O., & McGuinness, D. The role of frame-based representation on the semantic web. Linköping Electronic Articles in Computer and Information Science, 6(5), 2001.
13. Елизаров А. М. и др. Онтологии математического знания и рекомендательная система для коллекций физико-математических документов // Докл. РАН. – 2016. – Т. 467. – №. 4. – С. 392-395.
14. Когаловский М. Р., Паринов С. И. Семантическое структурирование контента научных электронных библиотек на основе онтологий. – 2015.
15. Doerr M. et al. The europeana data model (edm) // World Library and Information Congress: 76th IFLA general conference and assembly. – 2010. – С. 10-15.
16. Серебряков В. А. Что такое семантическая цифровая библиотека // Труды 16-й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» — RCDL. - Дубна, Объединенный институт ядерных исследований. - 2014. - С. 21- 25.

17. Самарский А. А., Тихонов А. Н. Уравнения математической физики — М., изд-во МГУ — Наука, 2004, 7-е изд.
18. Седов Л.И. Механика сплошной среды. Т. 1, 2. М.: Наука, 1970.
19. Композиционные материалы. Справочник. Под общ. ред. В. В. Васильева, Ю. М. Тарнопольского. М.: Машиностроение, 1990.
20. Васильев В.В. Механика конструкций из композиционных материалов. М.: Машиностроение, 1988.
21. Vasiliev V.V., Morozov E.V. Advanced Mechanics of Composite Materials and Structures. Fourth Edition, Elsevier, 2018, 856 p.
22. Моисеев Е.И., Муромский А.А., Тучкова Н.П. Нанотрубки в терминах макс пресс москва, isbn 978-5-317-04236-3, 231 с., 2012
23. Моисеев Е.И., Муромский А.А., Тучкова Н.П. Фуллерены в терминах МАКС ПРЕСС Москва, ISBN 978-5-317-03548-8, 150 с., 2010.

References

1. Gruber T. Ontology of folksonomy: A mash-up of apples and oranges // International Journal on Semantic Web and Information Systems (IJSWIS). – 2007. – V. 3. – №. 1. – S. 1-11.
2. Т. А. Gavrilova and V. F. Horoshevskij, Bazy znaniy intellektual'nyh sistem // Piter, St.-Peterburg, 2000.
3. Hlava M. M. K. The Taxobook: History, Theories, and Concepts of Knowledge Organization, Part 1 of a 3-Part Series // Synthesis Lectures on Information Concepts, Retrieval, and Services. – 2014. – V. 6. – №. 3. – S. 1-80.
4. Hlava M. M. K. The Taxobook: Principles and practices of building taxonomies, part 2 of a 3-part series // Synthesis Lectures on Information Concepts, Retrieval, and Services. – 2014. – V. 6. – №. 4. – S. 1-164.
5. Hlava M. M. K. The Taxobook: Applications, Implementation, and Integration in Search: Part 3 of a 3-Part Series // Synthesis Lectures on Information Concepts, Retrieval, and Services. – 2014. – V. 6. – №. 4. – S. 1-156.
6. Ataeva O. M., Serebryakov V. A., Tuchkova N. P. Mathematical physics branches: Identifying mixed type equations // Lobachevskii Journal of Mathematics. – 2019. – V. 40. – №. 7. – S. 876-886.
7. Ataeva O., Serebryakov V. A., Sinelnikova E. Thesaurus and Ontology Building for Semantic Library Based on Mathematical Encyclopedia // DAMDID/RCDL. – 2019. – S. 148-157.
8. Serebryakov V. A., Ataeva O. M. Ontology Based Approach to Modeling of the Subject Domain “Mathematics” in the Digital Library // Lobachevskii Journal of Mathematics. – 2021. – V. 42. – №. 8. – S. 1920-1934. <https://mathscinet.ams.org/msc/msc2010.html>.

9. Elizarov A. M. et al. Mathematical knowledge representation: semantic models and formalisms // Lobachevskii Journal of Mathematics. – 2014. – T. 35. – №. 4. – С. 348-354.
10. Bravo M., Hoyos Reyes L. F., Reyes Ortiz J. A. Methodology for ontology design and construction // Contaduría y administración. – 2019. – V. 64. – №. 4.
11. Allemang D., Hendler J., Gandon F. Semantic web for the working ontologist. – Association for Computing Machinery, 2020.
12. Lassila, O., & McGuinness, D. The role of frame-based representation on the semantic web. Linköping Electronic Articles in Computer and Information Science, 6(5), 2001.
13. Elizarov A. M. at all. Ontologii matematicheskogo znaniya i rekomendatel'naya sistema dlya kollekcij fiziko-matematicheskikh dokumentov // Dokl. RAN. – 2016. – T. 467. – №. 4. – S. 392-395.
14. Kogalovskij M. R., Parinov S. I. Semanticheskoe strukturirovanie kontenta nauchnyh elektronnyh bibliotek na osnove ontologij. – 2015.
15. Doerr M. et al. The europeana data model (edm) // World Library and Information Congress: 76th IFLA general conference and assembly. – 2010. – S. 10-15.
16. Serebryakov V. A. Chto takoe semanticheskaya cifrovaya biblioteka // Trudy 16-j Vserossijskoj nauchnoj konferencii «Elektronnye biblioteki: perspektivnye metody i tekhnologii, elektronnye kollekcii» — RCDL. - Dubna, Ob"edinennyj institut yadernyh issledovanij. - 2014. - S. 21- 25.
17. Samarskij A. A., Tihonov A. N. Uravneniya matematicheskoy fiziki — M., izd-vo MGU — Nauka, 2004, 7-e izd.
18. Sedov L.I. Mekhanika sploshnoj sredy. T. 1, 2. M.: Nauka, 1970.
19. Kompozicionnye materialy. Spravochnik. Pod obshch. red. V. V. Vasil'eva, YU. M. Tarnopol'skogo. M.: Mashinostroenie, 1990.
20. Vasil'ev V.V. Mekhanika konstrukcij iz kompozicionnyh materialov. M.: Mashinostroenie, 1988.
21. Vasiliev V.V., Morozov E.V. Advanced Mechanics of Composite Materials and Structures. Fourth Edition, Elsevier, 2018, 856 p.
22. Moiseev E.I., Muromskij A.A., Tuchkova N.P. Nanotrubki v terminah maks press moskva, isbn 978-5-317-04236-3, 231 s., 2012.
23. Moiseev E.I., Muromskij A.A., Tuchkova N.P. Fullereny v terminah MAKS PRESS. Moskva , isbn 978-5-317-03548-8, 150 c., 2010.