

Управление математическими знаниями: онтологические модели и цифровые технологии

© А.М. Елизаров¹ © А.В. Кириллович¹ © Е.К. Липачев¹ © О.А. Невзорова²

¹ Казанский (Приволжский) федеральный университет,

² Институт прикладной семиотики АН Республики Татарстан,

Казань

amelizarov@gmail.com

alikh.kirillovich@gmail.com

elipachev@gmail.com

onevzoro@gmail.com

Аннотация

Представлены основные идеи, подходы и уже полученные результаты, развиваемые в рамках проекта по разработке технологий управления математическими знаниями на основе онтологий. Ключевой идеей является разработка специализированных онтологий в области математики, которые составят основу специализированной цифровой экосистемы OntoMath, состоящей из совокупности онтологий, инструментов текстовой аналитики и приложений для управления математическими знаниями.

Выполненные исследования лежат в русле идеологии проекта создания Всемирной цифровой математической библиотеки (World Digital Mathematical Library – WDML). Основное назначение WDML – объединить в распределенной системе взаимосвязанных хранилищ оцифрованные версии всего корпуса математической научной литературы, включая как современные источники, так и источники, ставшие историческими. Настоящая работа развивает названное направление исследований. В ней, в частности, представлена система сервисов автоматической обработки больших коллекций физико-математических документов.

1 Введение

В настоящее время благодаря широкому внедрению информационно-коммуникационных технологий в научно-исследовательскую деятельность при проведении новых исследований стало возможным использование всего корпуса накопленных научных знаний. Последнее предполагает создание комплекса технологий, обеспечивающих оптимальное управление

имеющимися знаниями, организацию эффективного доступа к ним, а также совместное и многократное использование новых видов структур знаний.

В области профессиональной математики уже накоплен значительный опыт обработки и использования электронного математического контента в рамках различных проектов создания математических электронных библиотек. Например, в настоящее время одной из крупнейших формальных математических библиотек является Mizar [1], которая представляет собой коллекцию Mizar-статей (документов, подготовленных на формальном языке системы Mizar), содержащих определения, теоремы и доказательства [2, 3]. Следует также отметить важные результаты, связанные с формализацией уровня представлений математических статей. Для этих целей разработан широкий спектр языков представления исходных математических текстов – форматы LaTeX, STEX, XML, специализированные формальные языки, а также программные средства конвертации языков (см., например, [4–6]).

Известны также результаты разработки специализированных математических поисковых систем, например, (uni)quation [7], Springer LaTeX Search [8], Wolfram Formula Search [9].

Названные и многие другие реализованные математические проекты подготовили почву для реализации новой идеи – создания Всемирной цифровой математической библиотеки (World Digital Mathematical Library – WDML; термин введен в 2006 году на Генеральной ассамблее Международного математического союза).

2 Проект WDML

Традиционные библиотеки хранят документы, связывая их библиографическими ссылками, и помогают пользователю найти нужный документ на основе его библиографического описания, ключевых слов и выбранных тематических рубрик. По этой причине традиционные библиотеки обладают существенным ограничением – они не обеспечивают

Труды XVIII Международной конференции DAMDID/RCDL'2016 «Аналитика и управление данными в областях с интенсивным использованием данных», Ершово, 11-14 октября 2016

пользователю непосредственного доступа к элементам математического знания, последний вынужден вручную искать в документах интересные его математические понятия и выявлять скрытые связи между ними. Появление электронных библиотек сделало работу с документами более эффективной: они предоставляют мгновенный доступ к документам и обеспечивают полнотекстовый поиск по ним. Однако основной принцип работы электронной библиотеки по-прежнему опирается на хранение и поиск документов. Чтобы преодолеть указанное ограничение, проект WDML предлагает новую систему организации и хранения математического знания. В отличие от традиционных электронных математических библиотек, основными элементом этой системы являются не документы, а математические объекты (определения, аксиомы, теоремы, доказательства, уравнения и т. д.), а также логические связи между ними. Информация в WDML должна храниться в формализованном и понятном компьютеру виде, сформированном на основе технологий Семантического Веба. Такой способ управления математическими знаниями позволит создать инструменты для работы непосредственно с объектами математического знания (средства агрегации, семантического поиска, поиска по формулам и идентификации тождественных объектов) [10–13].

В глобальной инициативе WDML можно выделить ряд важнейших направлений, связанных как с организационными усилиями мирового математического сообщества, в том числе издателей математической литературы, так и исследовательских и технологических направлений, нацеленных на разработку и внедрение новых (семантических) технологий представления и обработки математического контента. К числу последних можно отнести:

- агрегирование различных онтологий, индексов, других ресурсов, созданных математическим сообществом, и обеспечение широкого доступа к ним для пополнения и редактирования;
- расширение возможностей доступа к математическим публикациям – не только поиск и просмотр, но и аннотирование, навигация, связывание с другими источниками, организация вычислений, визуализация данных и т. п.

Одной из ключевых идей является разработка классов объектов для адекватного описания и исследования математического содержания (контента). Структурированность математического документа позволяет выделить некоторый набор базовых классов математических объектов (последовательности, функции, преобразования, тождества, символы, формулы, теоремы, утверждения и др.). Как отмечается в проекте

WDML, одной из важнейших задач является построение списков математических объектов в разных областях математики.

Переход к представлению внутренней структуры математического знания создает новую парадигму представления, в которой основные акценты смещаются на выделение элементов (классов) и их взаимосвязей, что позволяет создавать различные сетевые концептуальные структуры (например, граф цитирования, граф математических концептов и др.). Выделение классов математических объектов и организация соответствующих репозиториев позволят создать новые вычислительные возможности по обработке данных, такие, как извлечение и обработка формул, поиск близких результатов и т. п.

Семантический поиск позволит по введенному описанию объекта или после выделения термина в статье получить дополнительную информацию (определение, свойства, связи с другими объектами, список документов, в которых встречается объект, указание публикации, где он впервые введен). С помощью такого поиска, например, можно найти все теоремы, в доказательстве которых, прямо или косвенно, используется пятый постулат Евклида. Важно отметить, что в различных документах объект может обозначаться различными терминами.

Инструменты агрегации позволяют автоматически собирать объекты, удовлетворяющие заданному критерию и формировать на их основе автоматически пополняемые списки, например, список объектов из заданной предметной области или список теорем, касающихся заданного математического объекта. Эти списки помогут математикам быть в курсе последних достижений, не тратя время на мониторинг всей литературы и не решая повторно уже решенную задачу.

С помощью инструмента поиска по формулам пользователь сможет выделить формулу и получить о ней дополнительную информацию (название, список литературы и т. д.) или ввести уравнение и получить список статей, в которых оно исследуется. При этом формулы в просматриваемых документах могут иметь различное символическое представление.

Инструменты идентификации предназначены для выявления тождественных объектов, которые упоминаются под разными именами и с использованием различной нотации.

Таким образом, основное назначение WDML – объединить в распределенной системе взаимосвязанных хранилищ оцифрованные версии всего корпуса математической научной литературы, включая как современные источники, так и источники, ставшие историческими.

В базовых документах проекта подчеркнута, что необходимо также обеспечить интеллектуальное извлечение информации для последующей передачи пользователю [12, 13]. В качестве примера назван

сервис, позволяющий пользователю выделить формулу, а затем обратиться к WDML для получения разъяснений и необходимых ссылок.

Для задачи навигации по всему корпусу математических документов ожидаются новые решения, обеспечивающие просмотр и получения дополнительной информации об интересующих объектах: разработка улучшенных механизмов ранжирования документов, в том числе по запросам пользователя; поиск документа, в котором впервые упоминается определенный математический результат, оперативный доступ к справочным ресурсам по теме запроса и т. п.

В целом разработчики проекта WDML полагают, что следующий шаг в развитии и продвижении математики состоит в выходе за пределы традиционных математических публикаций и построении сети информации, основанной на знаниях, содержащихся в этих публикациях.

Одновременно все более востребованными у ученых становятся новые способы обнаружения объектов научного знания непосредственно через Веб, а также инструменты и сервисы, обеспечивающие создание и совместное использование новых видов структур знаний. В контексте концепции связанных данных (Linked Data) и Семантического Веба такие инструменты и сервисы можно использовать для создания «графов сотрудничества» (collaboration graph), которые полезны, например, для вычисления «расстояния сотрудничества» (collaboration distance) между авторами и выделения «близких» документов, что открывает новые возможности тонкой настройки поиска и просмотра (см., например, [14]). Многими авторами (например, [15–17]) подчеркивается важность разработки новых онтологий предметных областей, в частности, математики, поскольку традиционной библиографической каталогизации сегодня уже недостаточно – требуется более глубокая детализация, содержащая описания, созданные с учетом разных точек зрения.

Основные задачи построения WDML и технологии, необходимые для их решения, обсуждены в 2014–2015 гг. широким кругом математиков и закреплены в ряде документов, принятых Всемирным математическим союзом. В частности, одобрено, что следующим шагом в развитии проекта WDML будут выход за пределы традиционных математических публикаций и построение сети информации, основанной на знаниях, содержащихся в этих публикациях. Благодаря сочетанию методов машинного обучения и усилий редакций и редколлегий математических научных журналов, значительная часть информации и знаний (как связанных открытых данных) в глобальном математическом корпусе знаний станет доступной для исследователей через WDML. К реализации названных идей приглашен ряд научно-исследовательских групп по всему миру, в том числе наша группа в Казанском (Приволжском)

федеральном университете (КФУ). Представители многих научных групп, задействованных в реализации проекта WDML, приняли участие в симпозиуме, прошедшем в феврале 2016 года в Филдсовском институте (г. Торонто) [18]. Доклад представителей КФУ на этом симпозиуме был посвящен модельным и программным решениям в области семантического представления математического знания [19]. Эти результаты в большой степени коррелируют с общей идеологией проекта WDML в части семантического представления и обработки математических знаний и являются стратегическим направлением исследований Казанской группы, связанным, в частности, с построением экосистемы OntoMath, которая описана ниже.

Отметим, что уровни представления математического знания, соответствующие форматы представления и применяемые семантические модели могут быть описаны, как показано на рис. 1.

Math Document Mobius Strip

Knowledge	Format	Semantic Model
Text	PDF, LaTeX	-
Metadata	OWL	AKT Portal Ontology
Logical structure	OWL	Mocassin Ontology
Terminology	OWL	OntoMathPro Ontology
Symbolic computation	Coq, Agda	Formalized mathematics



Рисунок 1 Уровни представления математического контента [19]

3 Экосистема OntoMath

OntoMath (<http://ontomathpro.org/>) – это цифровая экосистема онтологий, инструментов текстовой аналитики и приложений, предназначенная для управления математическими знаниями. В ее состав на текущий момент входят:

- онтология логической структуры математических документов Mocassin;
- онтологии профессиональной математики OntoMathPRO;
- программная платформа для подготовки математического набора связанных данных для публикации в облаке Linked Open Data (LOD);
- сервис семантического поиска по математическим формулам.

Кратко опишем указанные основные элементы.

Семантическое аннотирование математических текстов базируется на онтологии, построенной в рамках проекта Mocassin [20], и онтологии профессиональной математики OntoMath^{PRO} [21].

При разработке последней использовались различные терминологические источники: классические книги, интернет-ресурсы (Wikipedia, журнала «Известия вузов. Математика», а также личный опыт работы профессиональных математиков Казанского университета.

Онтология Mocassin – это онтология логической структуры математических документов. Она разработана В.Д. Соловьевым и Н.Г. Жильцовым и предназначена для автоматического анализа математических публикаций в формате LaTeX [22]. Эта онтология описывает семантику структурных элементов математических документов (например, теоремы, леммы, доказательства, определения и т. д.) и связей между ними. Она разработана с использованием языков OWL2/RDFS [23], которые обеспечили ей выразительные возможности, а также теоретические и практические средства вывода, например, с использованием таких современных машин вывода, как Pellet [24] и FaCT++ [25]. Онтология Mocassin содержит типовые концепты и отношения, эффективно извлекаемые из текстов автоматическими методами [26] (см. рис. 2). Она включает следующие концепты: *DocumentSegment* (Сегмент документа), *Axiom* (Аксиома), *Claim* (Утверждение), *Conjecture* (Гипотеза), *Corollary* (Следствие), *Definition* (Определение), *Equation* (Формула), *Example* (Пример) и др. Математический документ в этой модели рассматривается как набор связанных сегментов, которые являются частью документа, имеют начальную и конечную позиции в тексте, а также характеризуются конкретной функциональной ролью. Онтология структуры научных публикаций по математике описывает семантику сегментов и такие возможные отношения между ними, как *dependsOn* (зависит от), *exemplifies* (является примером для), *hasConsequence* (имеет следствие), *hasSegment* (содержит как сегмент),

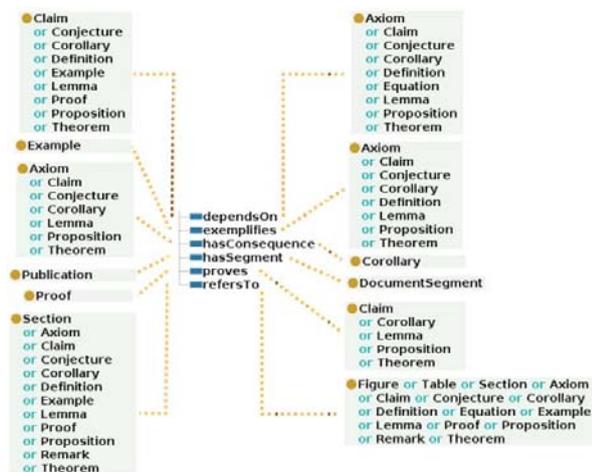


Рисунок 2 Элементы онтологии Mocassin [22]

Онтология профессиональной математики OntoMath^{PRO} – это онтология математического

знания, которая организована в виде двух иерархий (см. рис. 3):

- иерархии областей математики: математическая логика, теория множеств, алгебра, геометрия, топология и т. д.;
- иерархии математических объектов: множество, функция, интеграл, элементарное событие, многочлен Лагранжа и т. д.).

OntoMath^{PRO} разработана на языках OWL-DL/RDFS и содержит 3450 классов, 6 типов свойств объектов, 3630 экземпляров свойства IS-A и 1140 экземпляров остальных свойств. Она содержит пять типов отношений: *Класс* → *Подкласс*, *Определяется с помощью*, *Ассоциативная связь*, *Задача* → *Метод решения* и *Область математики* → *Математический объект*. Концепты онтологии содержат их название на русском и английском языках, определение, ссылки на внешние ресурсы из облака Linked Open Data и связи с другими концептами. Объектами семантического аннотирования также являются формулы, связанные с формулами фрагменты текста, задающие описания переменных формул [17, 27, 28].

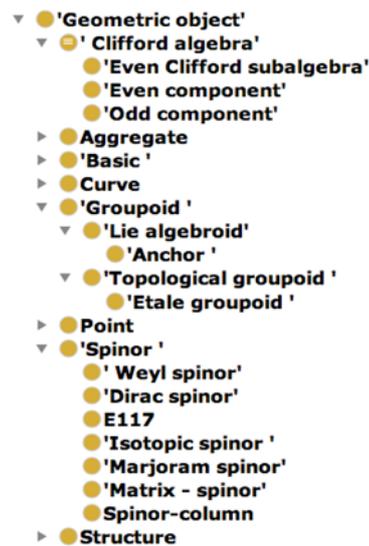


Рисунок 3 Фрагмент онтологии OntoMath^{PRO}

Одним из важных приложений, разработанным на основе указанных выше онтологических моделей, является специальная программная платформа для подготовки математического набора связанных данных для публикации в облаке LOD. Подготовка математического набора связанных данных выполняется на основе разработанных программных инструментов, реализующих комплексный технологический процесс подготовки RDF-набора данных [29]. В качестве экспериментальной коллекции использовались статьи журнала «Известия вузов. Математика» за 1997–2009 гг.

Основными функциями разработанного программного прототипа для публикации данных в облаке LOD являются:

- индексирование математических статей в формате LaTeX в виде LOD-совместимых RDF-данных;
- извлечение метаданных статьи в виде концептов онтологии AKT Ontology [30];
- извлечение логической структуры документа с использованием онтологии Mocassin;
- извлечение экземпляров математических сущностей в виде концептов онтологии OntoMathPRO и связывание с ресурсами DBpedia;
- распознавание семантики формул через связывание полученных экземпляров математических сущностей с математическими выражениями и формулами в тексте;
- установление взаимосвязи между опубликованными RDF-данными и существующими наборами данных LOD.

Разработанная технология имеет следующие отличительные особенности:

- математический RDF-набор строится на основе коллекции математических статей на русском языке;
- построенный RDF-набор помимо метаданных статей включает специальные семантические знания: знания, формируемые в результате специальной обработки математических формул – семантического связывания текстовых определений переменных формул с их символьными обозначениями; знания, связанные с идентификацией в тексте экземпляров онтологии OntoMathPRO; также знания о структурных элементах математической статьи.

Архитектура прототипа программной системы включает 8 модулей, которые могут быть сгруппированы в следующие подсистемы:

- преобразование формата;
- аннотирование текста;
- семантическое аннотирование;
- аннотирование метаданных;
- генерация RDF;
- связывание.

Подробное описание названных модулей дано в [29].

Другим важным инструментом является разработанный нами семантический поисковик по математическим формулам [27]. Отметим, что известные сегодня сервисы поиска по формулам (такие, как (uni)quation, Springer LaTeX Search, Wikipedia Formula Search, Wolfram Formula Search) являются синтаксическими и позволяют лишь найти формулы, содержащие заданные фрагменты (например, $(a+b)^2$). В отличие от них семантический поисковый сервис системы OntoMath решает задачу поиска формул по именам ее символьных переменных, в частности, выполняет поиск формулы, содержащей переменную, обозначающую

заданное математическое понятие (например, формулы, содержащие обозначение угла или связь давления и массы).

Этот сервис использует семантическое представление документа, построенное с помощью платформы семантической публикации, описанной выше. Названное представление отражает связи между элементами логической структуры документа, используемой терминологией, переменными и формулами. При этом пользователь может ограничивать контекст поиска, например, искать только в текстах определений и формулировках теорем.

Еще одним приложением экосистемы OntoMath является рекомендательная система для коллекций физико-математических документов, которая для каждого из них строит список «близких» документов (см. [31]). Традиционно список «близких» документов формируется на основе выбранной меры близости ключевых слов, приведенных авторами, а также библиографических ссылок, имеющихся в документах. Этот подход имеет ряд недостатков:

- список ключевых слов может быть неполным или отсутствовать;
- проблема омонимии: одно и то же понятие может обозначаться разными ключевыми словами, например, «полином» и «многочлен»;
- не учитываются родовидовые отношения между понятиями, например, статья с ключевым словом «пятиугольник» не будет определена как «близкая» к статье с ключевым словом «многоугольник»;
- существенна привязка к языку, например, статья с ключевым словом «матрица» не будет «близка» к статье с ключевым словом «matrix».

Таким образом, авторского списка ключевых слов недостаточно, и необходим более глубокий анализ содержания документа. Одним из методов такого анализа является терминологическое аннотирование, основанное на онтологиях предметных областей (например, [32]).

В разработанной рекомендательной системе реализуется следующие основные этапы:

- на основе онтологии OntoMath^{PRO} производится извлечение ключевых слов из документов коллекции;
- каждая публикации представляются в виде вектора, компоненты которого соответствуют концептам онтологии;
- значение компоненты – это вес соответствующего понятия в данной статье (вычисляется с использованием количества его упоминаний в тексте статьи и количества упоминаний связанных понятий) (см. [31]);
- в качестве меры близости между публикациями используется косинусная мера близости между их векторами.

Заключение

Описаны основные идеи, подходы и полученные результаты по разработке технологий управления математическими знаниями на основе специализированных онтологий в области математики. Эти решения составляют основу специализированной цифровой экосистемы OntoMath, которая состоит из совокупности онтологий, инструментов текстовой аналитики и приложений для управления математическими знаниями. Выполненные исследования лежат в русле проекта создания Всемирной цифровой математической библиотеки World Digital Mathematical Library.

Благодарности

Работа выполнена при финансовой поддержке РФФИ (проекты №№ 15-07-08522, 15-47-02472).

Литература

- [1] <http://www.mizar.org/>.
- [2] A. Naumowicz, A. Kornilowicz. A brief overview of Mizar. In: S. Berghofer et al. (Eds.), TPHOLS 2009, Lecture Notes in Computer Science 5674, Springer-Verlag Berlin Heidelberg, p. 67–72, 2009.
- [3] G. Bancerek, C. Bylinski, A. Grabowski, A. Kornilowicz, R. Matuszewski, A. Naumowicz, K. Pak, J. Urban. Mizar: State-of-the-Art and Beyond. In M. Kerber et al. (Eds.), Intelligent Computer Mathematics, CICM 2015, Lecture Notes in Artificial Intelligence 9150, p. 261–279, 2015.
- [4] M. Kohlhase. Using LaTeX as a Semantic markup format. Mathematics in Computer Science (2:2); p. 279–304, Birkhäuser 2008.
- [5] H. Stamerjohanns, D. Ginev, C. David, D. Misev, V. Zamdzhiev, M. Kohlhase. Conversion d'articles en LaTeX vers XML avec MathML: une étude comparative. Cahiers GUTenberg, 51; p. 7–28, 2010, http://cahiers.gutenberg.eu.org/cg-bin/article/CG_2008_51_7_0.pdf.
- [6] M. Iancu, M. Kohlhase, F. Rabe, J. Urban. The Mizar Mathematical Library in OMDoc: Translation and Applications; Journal of Automated Reasoning, 50:2; p. 191–202, Springer Verlag 2013.
- [7] Сайт (uni)quationalpha math expression search engine. <http://uniquation.com/en/>.
- [8] Сайт LaTeX Search Beta. <http://latexsearch.com/>.
- [9] The Wolfram Functions Site. <http://functions.wolfram.com/>.
- [10] Digital Mathematics Library: a vision for the future. International Mathematical Union, 2006. http://www.mathunion.org/fileadmin/IMU/Report/dml_vision.pdf.
- [11] P. J. Olver. What's happening with the World Digital Mathematics Library? http://www.math.umn.edu/~olver/t_wdmlb.pdf.
- [12] Developing a 21st century global library for mathematics research. Washington, D.C.: The National Academies Press, 2014. 131 p. arxiv.org/pdf/1404.1905;
- <http://www.nap.edu/catalog/18619/developing-a-21st-century-global-library-for-mathematics-research>.
- [13] P. J. Olver. The World Digital Mathematics Library: report of a panel discussion. Proceedings of the International Congress of Mathematicians, August 13–21, 2014, Seoul, Korea. Kyung Moon SA, V. 1, p. 773–785, 2014.
- [14] R. Todeschini, A. Baccini. Handbook of bibliometric indicators: quantitative tools for studying and evaluating research. Wiley-VCH Verlag, 2016.
- [15] S. Staab, R. Studer (Eds.). Handbook on ontologies. Berlin, Heidelberg: Springer Verlag, 2003, 2009. 811 p.
- [16] C. Lange. Ontologies and languages for representing mathematical knowledge on the Semantic Web. Semantic Web Journal, 2010. <http://www.semantic-web-journal.net/content/ontologies-and-languages-representing-mathematical-knowledge-semantic-web>.
- [17] A. Elizarov, A. Kirillovich, E. Lipachev, O. Nevzorova, V. Solovyev, N. Zhiltsov. Mathematical knowledge representation: semantic models and formalisms. Lobachevskii Journal of Mathematics, V. 35, No 4, p. 347–353, 2014.
- [18] Semantic representation of mathematical knowledge workshop, 5 February 2016: <https://www.fields.utoronto.ca/programs/scientific/15-16/semantic/>.
- [19] A. M. Elizarov, N. G. Zhiltsov, A. V. Kirilovich, E. K. Lipachev, O. A. Nevzorova, V. D. Solovyev. The OntoMath ecosystem: ontologies and applications for math knowledge management. Semantic Representation of Mathematical Knowledge Workshop 5 February 2016. <http://www.fields.utoronto.ca/video-archive/2016/02/2053-14698>.
- [20] Mathematical Semantic Search engINe (MocaSSIN). <https://code.google.com/archive/p/mocassin/>.
- [21] OntoMathPRO: A Hub for Math Linking Open Data (LOD). <https://github.com/CLLKazan/OntoMathPro>.
- [22] V. Solovyev, N. Zhiltsov. Logical structure analysis of scientific publications in mathematics. In: Proceedings of the Int. Conf. on Web Intelligence, Mining and Semantics (WIMS'11). ACM, p. 21:1–21:9, 2011. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.473.9071&rep=rep1&type=pdf>.
- [23] OWL 2 Web Ontology Language. RDF-Based Semantics (Second Edition). W3C Recommendation 11 December 2012. <https://www.w3.org/2012/pdf/REC-owl2-rdf-based-semantics-20121211.pdf>.
- [24] Pellet: An Open Source OWL DL reasoner for Java. <http://clarkparsia.com/pellet>.
- [25] FaCT++. <http://owl.man.ac.uk/factplusplus/>.
- [26] О. А. Невзорова, Е. В. Биряльцев, Н. Г. Жильцов. Коллекции математических

- текстов: аннотирование и применение в поисковых задачах. Искусственный интеллект и принятие решений, № 3, с. 51–62, 2012.
- [27] O. Nevzorova, N. Zhiltsov, A. Kirillovich, E. Lipachev. OntoMathPro ontology: a linked data hub for mathematics. *Communications in Computer and Information Science*, V. 468, p. 105–119, 2014.
- [28] А. М. Елизаров, Е. К. Липачёв, О. А. Невзо-рова, В. Д. Соловьев. Методы и средства семантического структурирования электронных математических документов. Докл. РАН, Т. 457 (6), с. 642–645, 2014.
- [29] O. Nevzorova, N. Zhiltsov, D. Zaikin, O. Zhib-rik, A. Kirillovich, V. Nevzorov, E. Birialtsev. Bringing Math to LOD: a semantic publishing platform prototype for scientific collections in mathematics. 12th Int. Semantic Web Conference, Sydney, NSW, Australia, October 21–25, 2013, Proceedings, Part I. *Lecture Notes in Computer Science*, Vol. 8218. Springer Berlin Heidelberg, p. 379–394, 2013.
- [30] АКТ Ontology. <http://dream.inf.ed.ac.uk/projects/dor/akt/akt.html>.
- [31] А. М. Елизаров, А. Б. Жижченко, Н. Г. Жильцов, А. В. Кириллович, Е. К. Липачёв. Онтологии математического знания и рекомендательная система для коллекций физико-математических документов. Докл. РАН, Т. 467, № 4, с. 392–395, 2016.
- [32] А. М. Елизаров, Н. Г. Жильцов, А. В. Кириллович, Е. К. Липачёв.

Терминологическое аннотирование и рекомендательный сервис в системе управления физико-математическим контентом. Труды XVII Межд. конф. DAMDID/RCDL'2015 «Аналитика и управление данными в областях с интенсивным использованием данных». Обнинск: ИАТЭ НИЯУ МИФИ, с. 347–350, 2015.

Mathematical knowledge management: ontological models and digital technology

Alexander M. Elizarov, Alexander V. Kirilovich, Evgeny K. Lipachev, Olga A. Nevzorova

This paper is discussed basic ideas, approaches and the results obtained in the research project the objective of which is to develop mathematical knowledge management technologies based on ontologies. We are developing the digital ecosystem OntoMath for mathematical knowledge management, which includes a set of specialized ontologies, text analytics tools and applications for managing mathematical knowledge. The results obtained are close to main problems declared in the World Digital Math Library (WDML) project. The main purpose of WDML is to build a global system of linked repositories for saving all digital mathematical documents, including contemporary and historic sources. This paper is devoted to decisions of some problems in this global initiative. In particular, we developed the program services for processing large collections of mathematical papers.