

УТВЕРЖДАЮ

Зам. директора

ИПМ им.М.В.Келдыша РАН

член-корреспондент РАН

М.В.Якобовский



24 декабря 2019 г.

Отзыв

ведущей организации Федерального государственного учреждения
«Федеральный исследовательский центр Институт прикладной математики
им. М.В. Келдыша Российской академии наук»
на диссертационную работу Атаевой Ольги Муратовны
«Разработка и реализация семантической цифровой библиотеки как основы
для построения пространства научных знаний», представленной на соискание
ученой степени кандидата технических наук по специальности 05.13.11 –
математическое и программное обеспечение вычислительных машин,
комплексов и компьютерных сетей

Актуальность темы исследования

Семантическая библиотека — современный эффективный механизм организации знаний и извлечения из них нового знания. Источников и форм представления цифровой научной информации с каждым годом становится все больше и больше, объемы нуждающейся в обработке информации стремительно растут. Семантические библиотеки являются сейчас одной из наиболее перспективных форм представления данных. Они не только вовлекают в обработку существующие и вновь появляющиеся структуры представления данных, но и успешно взаимодействуют со сложившимися формами организации хранения знаний в интернете.

Структура и содержание диссертации

Во **введении** представлено описание развития парадигмы Semantic Web и связанных с ней концепций. Аргументируется актуальность исследований, формулируются цель и задачи работы, перечисляются

используемые методы исследования, обосновывается научная новизна и практическая значимость полученных результатов, приводятся сведения о результатах внедрения и использования.

В главе 1 представлен обзор основных концепций и инструментов, положенных в основу работы. В частности, представлена парадигма Semantic Web и ее современное воплощение Linked Open Data, рассматривается место онтологий в этой парадигме. Анализируются современные решения, используемые для построения семантической библиотеки.

В главе 2 прослеживается эволюция семантической библиотеки, выделяется понятие ее контента согласно концептуальной модели DELOS. Представлены этапы процесса развития семантических библиотек и их основные свойства. Освещается роль информационных систем в процессе построения семантических библиотек и выделяется понятие онтологии информационной системы. Определяются центральные понятия работы — контент библиотеки и представление научной информации в рамках этого контента.

Глава 3 посвящена анализу и описанию обобщенной модели научной предметной области, приводятся основные свойства этой модели. Вводятся основные понятия: информационный объект, информационный ресурс, атрибут, набор атрибутов, свойства и связи между ними. Проводится классификация атрибутов

Строится модель тезауруса предметной области согласно стандарта ISO 25964. Тезаурус является каркасом для построения форматизированного представления терминологической модели знаний. Интересной особенностью предложенной автором модели является возможность настройки структуры тезаурусных статей для понятий и их связей, отвечающей специфике рассматриваемой предметной области. Предлагаемый подход избавляет пользователя библиотеки от необходимости обладать специальными техническими знаниями.

Глава 4 посвящена построению онтологии семантической библиотеки для научного пространства знаний на основе информационной модели из предыдущей главы. Вводятся понятия, необходимые для представления терминологии рассматриваемой научной предметной области, соединенные различными связями как иерархическими, так и горизонтальными. Вводится

набор определений, на абстрактном уровне описывающий множество объектов научной предметной области. Построенная онтология характеризуется открытой иерархической и динамичной структурой и служит как для хранения знаний и их структуризации, так и для извлечения новых знаний с использованием правил вывода.

В главе 5 представлены сведения об архитектуре семантической библиотеки LibMeta, разработанной на основе предлагаемого в диссертации подхода. Подробно описаны основные компоненты системы, приведены диаграммы, описывающие взаимодействие компонент. Важной особенностью предложенной автором реализации является то, что при использовании LibMeta как средства представления цифрового пространства знаний нет необходимости опускаться на уровень кода для уточнения специфики области.

В главе 6 приведены сведения о программной реализации LibMeta и о ее применении в следующих предметных областях:

- обыкновенные дифференциальные уравнения;
- задачи математической физики;
- микробиология и физиология растений;
- математическая энциклопедия.

Затем рассматривается организация взаимодействия системы с внешними программами, не являющимися web-браузерами, для чего был реализован специальный интерфейс. Предлагаются направления дальнейшего развития системы.

В заключении формулируются основные результаты и выводы диссертационной работы. В приложении приведены свидетельства о регистрации программ, реализованных во время проведения исследования.

Научно-техническая новизна диссертационной работы состоит в следующем.

- предложена базовая онтология, состоящая из относительно небольшого числа классов и свойств, позволяющая строить на ее основе более сложных модели данных различных научных предметных областей;

- предложен обобщенный подход к описанию контента, позволяющий реализовать средства интеграции данных в рамках библиотеки, адаптируемые под условия предметной области;
- поддерживается конструирование семантической библиотеки не на уровне разработки программного обеспечения, а на уровне настройки уже установленной системы, что существенно сокращает время ее построения;
- предложен новый подход к модели тезауруса, позволяющий реализовать гибкий настраиваемый поиск.

Достоверность, практическая значимость результатов и выводов, представленных в диссертации, подтверждаются их успешной апробацией в Библиотеке по естественным наукам РАН и в межведомственном суперкомпьютерном центре РАН. Результаты не раз докладывались на российских и международных научных форумах, они опубликованы в рецензируемых российских изданиях, в числе которых 5 из Перечня ВАК РФ, и получили положительный отклик.

Соответствие диссертации и автореферата требованиям «Положение о порядке присуждения научных степеней». Автореферат диссертации адекватно отражает ее содержание. Диссертация и автореферат удовлетворяют требованиям «Положения о присуждении ученых степеней».

Замечания

Изложение первых глав диссертации несколько суховато, и читатель с нетерпением ждет заключительную, шестую главу, где автор сможет показать "товар лицом", реализовав свои теоретические построения. Однако тут читателя возникает определенное разочарование: и эта глава представлена достаточно сухо, без наиболее востребованных читателем подробностей.

По-видимому, основная апробация теоретических положений диссертации состоялась при реализации семантической библиотеки «Обыкновенные дифференциальные уравнения». Но этот раздел подчеркнуто лаконичен, в нем не приводятся примеры того, как выглядят характерные запросы к построенной библиотеке, включенные в текст скриншоты практически не иллюстрируют сформированную для читателя среду. Прохладное отношение автора к представлению внешнего вида программного продукта заметно, в частности, и из таблицы объемов

созданного программного кода, приведенной на стр. 113: в таблице учтены все строки кода, кроме кодов CSS, отвечающих за внешнее представление результатов.

Один из наиболее интересных разделов библиотеки «Обыкновенные дифференциальные уравнения» — организация поиска математических формул. И здесь не приведено примеров конкретных, живых поисковых запросов и полученных результатов поиска. Отсутствует сопоставление построенных поисковых средств с наиболее известным в нашей стране аппаратом поиска формул, предложенным А.М. Елизаровым.

Самый многообещающий раздел шестой главы связан с "Математической энциклопедией" — классическим изданием, весьма популярным среди профессионалов. Казалось бы, вот великолепный шанс продемонстрировать возможности предлагаемого в диссертации механизма семантической библиотеки: систематически препарировать тексты "Математической энциклопедии" и разместить получившуюся семантическую библиотеку в открытом доступе. Все российские математики несомненно были бы благодарны диссертанту за такой подарок. Однако "Математическая энциклопедия" используется в диссертации лишь в качестве второстепенного вспомогательного материала, а очевидная амбициозная задача превращения этой энциклопедии в полноценную общедоступную семантическую библиотеку не ставится диссертантом даже в разделе "Дальнейшее развитие системы LibMeta".

В текстах диссертации и автореферата встречаются грамматические ошибки, несогласованные предложения, что несколько затрудняет их чтение.

Заключение

Отмеченные выше замечания не меняют общего положительного впечатления о диссертационной работе в целом. В ней излагаются результаты востребованной практикой стратегически важной задачи разработки и программной реализации средств организации научных знаний и интеграции их в мировое информационное пространство.

Диссертация О.М. Атаевой обладает внутренним единством, представляет завершенные результаты на актуальную тему и является научно-квалификационной работой, в которой на основании выполненных автором исследований разработаны теоретические положения, вносящие значительный вклад в разработку математического и программного обеспечения вычислительных машин, комплексов и сетей. Автореферат адекватно отображает ее содержание.

Диссертационная работа О.М. Атаевой удовлетворяет требованиям, предъявляемым к диссертациям по специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей», а ее автор заслуживает присуждения ей ученой степени кандидата технических наук по этой специальности.

Диссертация была заслушана 21 ноября 2019 г. на расширенном заседании семинара информационно-издательского отдела ИПМ им.М.В.Келдыша РАН (протокол № 11) и получила положительную оценку.

Зав. отделом ИПМ им.М.В.Келдыша РАН
д.ф.-м.н.



М.М.Горбунов-Посадов

Сведения об организации

Федеральное государственное учреждение «Федеральный исследовательский центр Институт прикладной математики им. М.В. Келдыша Российской академии наук» (ИПМ им. М.В. Келдыша РАН)

125047, г. Москва, Миусская пл. 4; 8 (499) 250-79-86
<https://keldysh.ru>; office@keldysh.ru