

Отзыв

официального оппонента 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 работы

В представленной диссертационной работе Атаевой О.М. рассмотрена актуальная задача организации пространства знаний научной предметной области. Актуальность работы связана с тем, что в последние десятилетия объем информации лавинообразно увеличивается и это касается и научных областей. Продолжаются попытки построить формальные модели научных предметных областей. Увеличивается количество источников разнообразных данных и форматов, в которых они представлены. Резко возросло время, необходимое для поиска нужной информации и ее обзора. Решаемая задача требует создания описания обобщенного представления научных знаний. Это необходимо для создания средств помощи экспертам в организации знаний и предоставлении доступа к ним конечных пользователей. При этом средство организации знаний должно быть достаточно универсальным и не требовать глубоких технических познаний. В диссертации такое средство организации знаний представляется открытой цифровой семантической персональной библиотекой, являющейся по существу одновременно средством разработки персональной библиотеки пользователя (или группы пользователей) и средством интеграции информации (научных знаний).

Основной целью работы, обозначенной автором, как раз и является создание такого универсального настраиваемого прототипа библиотеки, в котором учитывалось бы всё разнообразие различных типов ресурсов, которые могут храниться в цифровых библиотеках научных предметных областей, с поддержкой терминологического описания, специфического для предметной области. В контексте создания такого прототипа одной из основных решаемых задач является обеспечение возможности интеграции данных из различных источников, поддерживающих семантическое описание модели данных. Фактически реализуемая система должна представлять собой конструктор для создания цифровой библиотеки любой направленности с адаптируемой моделью контента хранимых данных. Под адаптируемой моделью данных здесь понимается возможность описывать структуры данных контента создаваемой библиотеки в терминах тезауруса предметной области.

Таким образом, направление исследований, обозначенное в диссертации, которое включает вопросы разработки модели информационных ресурсов и объектов науки и связей между ними, а также модели терминологического описания, настраиваемого на специфическую предметную область, следует признать актуальным и важным.

Характеристика содержания диссертационной работы

Диссертация Атаевой О.М. состоит из введения, шести глав, заключения, приложения и библиографического списка, включающего 160 наименований.

Во введении аргументируется актуальность исследований, формируется цель и задачи работы, перечисляются используемые методы исследования, обосновывается научная новизна и практическая значимость полученных результатов, приводятся сведения о результатах внедрения и использования.

В первой главе дается обзор парадигмы Semantic Web и ее практического воплощения Linked Open Data. Рассматриваются онтологии и тезаурусы, которые являются центральными понятиями, исследуемыми в работе, излагается проведенный автором анализ современного состояния стандартов в этой области. Не отклоняясь от основной цели и задач диссертации, автор рассматривает различные решения для семантических библиотек, выполненные различными исследовательскими группами. В заключение главы осуществляется постановка задачи и описывается логическая схема исследования.

Во второй главе проводится анализ семантических цифровых библиотек и модели представления научных знаний в них. Приведены последовательные этапы эволюции библиотек: электронные, цифровые и семантические библиотеки, и определяются основные свойства семантических библиотек. Аргументируется возможность структурированного представления разнообразного контента таких библиотек и возможность связывания данных из разных источников, что в свою очередь, несомненно, отражается на качестве данных контента. Приводится построение обобщенной модели научной предметной области, представляется подход, который делает упор на выделении таких метаданных, которые позволяют проектировать конкретные структуры данных для различных научных предметных областей и выявить общие подходы к управлению этими данными и их обработке. Это позволит построить общую модель научных знаний, в рамках которой можно проводить интеграцию данных из различных источников.

В третьей главе развивается формальный подход к моделированию основных направлений в исследовании построения представления пространства знаний, заключающийся в следующем:

- приведены основные свойства разрабатываемой модели семантической библиотеки научного пространства знаний, описана модель контента библиотеки и введены основные понятия;
- описана модель настраиваемого под специфику предметной области тезауруса;
- описана модель интеграции данных библиотеки с внешними источниками данных;
- введены дополнительные определения формальной модели для выделения дополнительных семантических связей в рамках контента библиотеки;
- приведены выводы и обобщения, которые необходимы для перехода на следующий уровень моделирования онтологии решаемой задачи.

Четвертая глава посвящена построению онтологии научного пространства знаний. Построение такой онтологии рассматривается с точки зрения двух ортогональных подходов: вводятся понятия, позволяющие рассматривать термины, характерные для научной предметной области, соединенные различными связями как иерархическими, так и горизонтальными, а также вводится набор определений, которые на более абстрактном уровне описывают множества объектов научной предметной области, фактически задавая структуру их описания и отношений между ними. Приводятся преимущества использования правил вывода для извлечения новых знаний.

Пятая глава посвящена описанию архитектуры разрабатываемой системы, компонентов системы и их отношений, на основе предложенного в диссертации подхода. Демонстрируется возможность реализации главной задачи семантической библиотеки – семантического/интеллектуального конструирования научного пространства знаний для некоторой предметной области. Для реализации функций открытости научного пространства знаний представлена компонента интеграции других источников данных и возможности связывания с их данными.

В шестой главе представлены сведения о программной реализации семантической библиотеки LibMeta, представляющей собой универсальный прототип, на основе которого и разрабатываются цифровые библиотеки предметных областей. Система реализована согласно требованиям к ее контенту (универсальности, адаптируемости, структурированности), которые сформулированы при постановке задачи, поддерживая при этом основные функции семантической библиотеки такие как: поддержка использования медийных объектов, поддержка семантического описания контента, поддержка интеграции с открытыми источниками, поддержка пользователей и возможности формирования тематической области их интересов.

Рассматривается организация взаимодействия системы с внешними программами, не являющимися web-браузерами. Обеспечение программных клиентов возможностью

оперировать данными, находящимися в системе реализовано на технологии REST (Representational State Transfer – передача представления состояния). Приведены примеры внедрения системы в нескольких предметных областях: микробиология, обыкновенные дифференциальные уравнения, математической физики, приведена реализация пространства знаний в рамках математической энциклопедии. В заключение представлены основные направления дальнейшего развития.

Научная новизна

В диссертационной работе автором предложена онтология, состоящая из минимального количества классов и свойств, что позволяет ее использовать в качестве базовой онтологии для построения более сложных моделей данных различных научных предметных областей. Отличительной особенностью является то, что обобщенный подход к описанию контента позволяет реализовать средства интеграции данных в рамках библиотеки, адаптируемые под условия любой предметной области без оглядки на ее специфику. Это позволяет формализовать процесс интеграции внешних источников в библиотеку и, благодаря гибкости адаптивной модели, позволяет интегрировать любой необходимый источник, удовлетворяющий требованиям в рамках введенных понятий. Оригинальность подхода заключается в том, что появляется возможность смоделировать пространство знаний с точки зрения его отображения в цифровой семантической библиотеке. При этом решаются как задачи сбора и сохранения данных, навигации по ее содержимому, так и задачи по извлечению новых, неявно определенных связей между объектами, составляющими наполнение библиотеки. В работе предложен новый подход к модели тезауруса, что позволяет реализовать гибкий настраиваемый поиск и настраивать модель данных тезауруса под специфику предметной области.

Практическая значимость

Развитый в диссертации подход, реализованный автором как универсальный настраиваемый прототип цифровой библиотеки LibMeta, позволяет применить разработанные в работе модели, подходы и алгоритмы для создания программного обеспечения научных электронных библиотек на основе явного описания модели ресурсов высокоуровневых понятий научной предметной области с использованием технологий Semantic Web. Разработанные прототипы программных систем могут быть использованы для конструирования научных семантических электронных библиотек не на уровне разработки программного обеспечения, а на уровне настройки уже установленной системы, что существенно сокращает время ее построения, что подтверждается рассмотренными ранее примерами практического применения.

Достоверность основных положений и результатов работы

Достоверность результатов обеспечивается прикладными результатами предложенного подхода в различных предметных областях в различных организациях. Результаты диссертации получены автором лично, представлены в 16 публикациях, пять из них опубликованы в научных изданиях, рекомендуемых ВАК РФ. Результаты работы обсуждались на десяти конференциях и научных семинарах. В опубликованных авторах трудах отражены основные положения его диссертации. Диссертационная работа оформлена в соответствии с требованиями, установленными Министерством образования Российской Федерации. Автореферат полностью отражает содержание диссертационной работы. Изложенные в работе материалы обладают внутренним единством и непротиворечивостью.

Необходимо отметить следующие недостатки представленной работы:

1. В четвертой и шестой главе следовало дать более развернутые выводы по результатам изложенного материала в этих главах;
2. В диссертации на рисунке 11 (стр. 105), рисунке 14 (стр. 121) (а также в автореферате на рисунке 3 (стр. 15)) использована не общепринятая структура диаграмм UML при описании модели данных и взаимодействия различных компонент в отличие от остальных диаграмм представленных на других рисунках в работе.
3. В разделе 2.5 (стр. 44) автор указывает на уменьшение размерности разрабатываемой модели на основе адаптивной модели данных, но не приводит никаких примеров и количественных оценок для подтверждения своего утверждения.

Указанные недостатки не снижают значимости полученных автором результатов. Представленная диссертационная работа свидетельствует, несомненно, о высокой научной квалификации автора. Показана возможность практического приложения полученных автором результатов для решения задачи организации пространства знаний в рамках научных предметных областей. Поставленная цель диссертационного исследования достигнута и, несмотря на замечания, заслуживает положительной оценки.

Автореферат соответствует основному содержанию диссертации и отвечает требованиям ВАК РФ.

На основе анализа содержания диссертации и опубликованных автором работ можно сделать заключение, что диссертация Атаевой Ольги Муратовны является законченной научно-квалификационной работой и соответствует требованиям ВАК РФ, предъявляемых к диссертациям на соискание кандидата технических наук. Новые научные результаты, полученные диссертантом, имеют существенное значение для

русской науки и практики в области создания и сопровождения информационных ресурсов различного назначения.

Автор диссертации, Атаева О.М., заслуживает присуждения ученой степени кандидата технических наук по специальности 05.13.11 – математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей.

Официальный оппонент

доктор физико-математических наук (05.13.11 – математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей), профессор РАН, академик РАН, директор Федерального государственного бюджетного учреждения науки Институт системного программирования им. В.П. Иванникова Российской академии наук (ИСП РАН)



Арутюн Ишханович Аветисян

«23» декабря 2019 г.

Сведения об организации

Федеральное государственное бюджетное учреждение науки Институт системного программирования им. В.П. Иванникова Российской академии наук (ИСП РАН)

109004, Россия, г. Москва, ул. Александра Солженицына, дом 25,

Телефон: +7(495) 912-44-25

Факс: +7(495) 912-15-24

E-mail: info-isp@ispras.ru

<https://www.ispras.ru/>