

ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА

на диссертационную работу Ольги Муратовны Атаевой «Разработка и реализация семантической цифровой библиотеки как основы для построения пространства научных знаний», представленную к защите на соискание ученой степени кандидата технических наук по специальности 05.13.11 – Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей

В диссертационной работе О.М. Атаевой исследованы актуальные и новые задачи организации пространства знаний для произвольной предметной научной области на основе открытой семантической цифровой библиотеки. В работе дано определение последней, а задача представления цифрового пространства научного знания рассмотрена, исходя из признания базовым того положения, что знание целостно, а деление его на отдельные дисциплины лишь условно. Это положение выводит на передний план задачи создания модели обобщенного представления научных знаний для произвольной предметной научной области и построения на основе этой модели инструмента организации знаний с целями помощи экспертам в навигации и извлечении новых знаний и облегчения процессов интеграции данных из различных источников, обладающих семантическим содержанием. Для решения названных задач использована адаптивная модель данных, позволяющая не ограничиваться при разработке строго очерченным набором ресурсов. Использование такой модели приводит к построению онтологии, близкой к высокоуровневым онтологиям предметных областей науки и описывающей наиболее общие понятия, не зависящие от конкретных проблемы или области.

Актуальность решения задачи построения цифрового пространства знаний, исследованной в диссертации, обусловлена стремительным увеличением объема информации в научных областях, наблюдаемым в настоящее время и требующим создания новых инструментов для ее хранения и обработки. При этом эффективная навигация, извлечение полезной информации и ее обзор приводят к значительным временным затратам. Одновременно увеличивается количество источников разнообразных данных и форматов, в которых информация представлена, что дополнительно усложняет доступ к ней. Онтологический подход к построению модели предметной научной области является одним из основных и ак-

туальных способов концептуализации пространства знаний. Он позволяет отделить понятие контента семантической цифровой библиотеки от понятия информационной системы, связанной с ней, и значительно наращивать функциональность этой системы, а также добавлять новые подсистемы или изменять уже имеющиеся без изменения остальных.

В диссертации О.М. Атаевой при построении семантической цифровой библиотеки для формирования пространства знаний рассматриваемой предметной области предложено специальное внимание уделять возможности терминологически очертить границу пространства знаний этой области, используя ее тезаурус. Опираясь на общепринятые стандарты в этой области, в работе рассмотрена расширяемая модель тезауруса, прежде всего, с возможностью наращивания базовой структуры концептов тезауруса. Для этого предложено поддерживать иерархию классов для дополнительных атрибутов концептов, которая позволяет включать в описание концептов тезауруса объекты, являющиеся непосредственно элементами пространства знаний.

Кратко опишем структуру и содержание диссертации.

Диссертация состоит из введения, шести глав, заключения, общий ее объем – 155 страниц. Библиография содержит 160 наименований.

Во введении достаточно подробно описана идея использования стека технологий Семантического веба для решения поставленной задачи. Дано определение семантических цифровых библиотек и охарактеризована их роль при построении цифрового пространства знаний. Сформулирована цель работы – применить онтологический подхода к построению цифрового пространства знаний и средств навигации в нем. Обоснована актуальность темы, охарактеризовано содержание работы, приведены основные результаты.

Первая глава посвящена обзору основных концепций и инструментов, которые легли в основу работы. Проведен анализ существующих решений по семантическому представлению знаний и интеграции рассматриваемых данных. Представлены парадигма Семантического веба и онтологии, которые стали активно использоваться с ее появлением; охарактеризованы использование онтологий для формализации знаний в различных предметных областях, их отличие от тезаурусов и связь между ними. Описана концептуальная модель электронных библиотек DELOS с определениями важнейших представлений об их архитектуре, ресурсах и

функциональности. Дан обзор семантических библиотек, созданных разными исследовательскими группами. Дана постановка решаемой задачи и приведена логическая схема проведенного исследования.

Вторая глава характеризует концепции, на которых базируется определение семантической цифровой библиотеки для предметной научной области. Рассмотрена роль информационной системы в процессе конструирования такой библиотеки. Аргументировано использование адаптивной модели данных.

В третьей главе описана информационная модель семантической цифровой библиотеки, используемая как средство построения цифрового пространства знаний. В этой модели выделен явным образом набор понятий, с помощью которых можно описать содержимое библиотеки, выделяемое в стандарте DELOS. Вводимые понятия, такие, как информационный объект, информационный ресурс, атрибут, в дальнейшем использованы как понятийная основа конкретной предметной научной области для описания ее пространства знаний. Фактически понятия информационной модели семантической цифровой библиотеки разделены на три категории: первая включает определения понятий контента этой библиотеки, вторая относится к определению понятий, необходимых для поддержки терминов в тезаурусе предметной области, а третья включает определения, необходимые для определения процессов интеграции контента этих ресурсов. На основе названных определений описаны такие основные процессы, как, например, интеграция данных из разных источников, категоризация/классификация, отображение разных моделей источников данных на заданную предметную область, построение классов эквивалентности. Определен набор правил для явных и неявных связей с целью извлечения дополнительных знаний.

В четвертой главе дано описание онтологии контента семантической цифровой библиотеки для представления цифрового пространства знаний, которое представлено с точки зрения двух подходов: определен необходимый набор понятий и их связей для конструирования тезауруса предметной области и введен набор определений, который на более абстрактном уровне описывает множества объектов предметной научной области, фактически задавая структуру их описания и отношений между ними. Аргументирована идея конструирования трехуровневой онтологии конкретной предметной области, в которой экземпляры первого уровня

– это высокоуровневые понятия разработанной онтологии, на втором уровне описаны понятия конкретной предметной области как экземпляры в терминах первого уровня; на третьем уровне они используются как определения классов при заполнении онтологии данными. Приведены примеры использования правил вывода новых знаний.

Пятая глава посвящена описанию основной функциональности системы, реализованной на основе разработанной онтологии. Функциональность системы составляют множество функций, доступных для всех публичных пользователей, и множество функций авторизованного пользователя, зависящее от его роли в системе. Подробно описаны основные компоненты системы, каждая из которых отвечает за определенную функциональность и использует определенное подмножество понятий информационной модели, рассмотрены способы их взаимодействия.

Шестая глава посвящена программной реализации семантической библиотеки LibMeta. Проведен анализ характеристик прототипа этой системы, созданного автором на основе архитектуры, разработанной в диссертации. Описаны возможности настройки прототипа под специфическую предметную область и приведены результаты его исследования на соответствие предъявляемым к системе требованиям, которые сформулированы при постановке задачи: универсальность, адаптируемость, структурированность. Представлены результаты использования системы применительно к следующим предметным областям: обыкновенные дифференциальные уравнения, задачи математической физики, микробиология и физиология растений, математическая энциклопедия. Рассмотрено взаимодействие системы с внешними программами, реализованное с использованием технологии REST (Representational State Transfer – передача представления состояния).

В заключении сформулированы основные результаты диссертационной работы.

Основными результатами диссертации, по моему мнению, являются следующие:

- Разработан подход к построению обобщенной модели предметной научной области: он нацелен на выделение таких метаданных, которые позволяют спроектировать конкретные структуры данных для различных предметных научных областей и выявить общие подходы к управлению этими данными и их обработке;

- Построена модель интеграции знаний в рамках предметной научной области на основе обобщенной модели описания информационных ресурсов, которая определяет возможности интеграции с различными источниками данных;
- Предложен подход к построению семантических информационных систем, способных гибко настраиваться под запросы конкретных предметных научных областей и формировать их онтологии на основе высокоуровневых понятий этих областей;
- На основе тезауруса, учитывающего связи между объектами цифровой семантической библиотеки, усовершенствованы доступ и восприятие пользователем больших и сложно структурированных объемов информации, что дает возможность более полно формировать картину научного знания в рассматриваемой области;
- Разработана гибкая и настраиваемая модель поддержки тезауруса в семантической цифровой библиотеке, которая позволяет выявлять и фиксировать новые связи между элементами тезауруса и контентом библиотеки и представлять научные знания в структурированном виде;
- На базе прототипа системы LibMeta, разработанной автором на основе подхода к построению семантических цифровых библиотек, предложенного в диссертации, созданы конкретные семантические цифровые библиотеки для четырех предметных областей.

Научная новизна результатов диссертации состоит в следующем:

- впервые разработана онтология, состоящая из минимального количества классов и свойств и служащая базой для построения моделей данных для различных предметных научных областей;
- впервые разработан обобщенный подход к формированию контента семантической цифровой библиотеки, реализующий интеграцию данных в условиях любой предметной области из любых внешних источников с выполнением требований, установленных в рамках введенных понятий;
- в отличие от разработанных ранее решений для семантических цифровых библиотек обеспечены конструирование библиотеки не на уровне разработки программного обеспечения, а на уровне настройки уже установленной системы, а также существенное сокращение временных затрат;

- обеспечена возможность моделирования пространства знаний с помощью его отображения в контексте семантической цифровой библиотеки, которая реализует как классические возможности сбора, хранения и навигации по ее содержимому, так и новые возможности по извлечению новых, неявно определенных связей между объектами, составляющими наполнение библиотеки;
- предложен новый подход к построению модели тезауруса, который позволяет реализовать гибкий настраиваемый поиск объектов предметной области, а также настраивать модель данных тезауруса под специфику предметной области.

Теоретическая значимость исследования, проведенного в диссертации, заключается в том, что полученные результаты вносят вклад в развитие теории информационных систем для предметных научных областей, основанных на технологиях Семантического веба в условиях непрерывно расширяющегося поступающего потока информации. Тем самым внесен значимый вклад в теорию цифровых библиотек.

Практическая значимость результатов диссертации: разработанные подходы, модели и алгоритмы применены для создания программного обеспечения цифровых библиотек в нескольких конкретных научных предметных областях и могут быть использованы для конструирования различных научных семантических цифровых библиотек с использованием технологий Семантического веба.

Можно отметить следующие **недостатки и замечания по диссертации**.

Замечания по содержанию:

1. В главе 2 предложено рассмотреть последовательные этапы эволюции библиотек: электронные, цифровые и семантические библиотеки. Далее даны определения этих понятий. Во введении и первой главе разные упоминаемые библиотеки охарактеризованы одним из указанных терминов (например, Europeana названа глобальной семантической цифровой библиотекой; Jerome DL названа приложением для ведения цифровой библиотеки, но при этом отмечено, что последняя сейчас недоступна (возникает вопрос, зачем тогда указаны ее недостатки и преимущества)).

Во-первых, следовало сначала определить используемые понятия, а уж затем их применять. Так, например, на с. 36 утверждается, что контент электронных библиотек представляет собой множество библиографических записей объектов реальной классической библиотеки. Такое определение электронной библиотеки нигде ниже не используется, но вызывает возражения: в исследованиях многих авторов под контентом электронных библиотек подразумеваются не просто оцифрованные библиографические записи, а разнообразные цифровые документы. В этом смысле в определении цифровых библиотек, приведенном в диссертации, как раз говорится о таком контенте. Отмечу также, что термин «электронные библиотеки» в русскоязычных изданиях является аналогом термина «digital libraries» из англоязычных публикаций.

Во-вторых, по определениям диссертанта, цифровые библиотеки «решают те же задачи поиска и хранения контента, что и электронные библиотеки, но существенно расширяют свою функциональность и определение своего контента» (с. 36), а семантическая цифровая библиотека – это уже расширение цифровой библиотеки за счет применения семантических технологий для обработки контента (вот это определение и является базовым для диссертации). Отмечу, что приведенные определения не содержат никаких ссылок на определения, которые были даны другими авторами, а это следовало сделать.

2. В п. 1.6 сформулирована, как мне представляется, основная задача диссертации – «задача разработки семантической библиотеки для некоторой научной предметной области». Далее в тексте упоминается еще ряд задач. На с. 33 сформулированы основные решаемые задачи (их семь), а во введении они названы подзадачами. Далее, во введении к этим основным подзадачам добавлены еще три (из них две не пересекаются с названными семи, а одна – полностью совпадает с пятой из этого перечня; все то же самое – в автореферате), а на с. 33 эти последние названы целями работы. Желательно на защите расставить все по своим местам.

3. Диссертация имеет существенные приложения и внедрения. Как отмечено в тексте работы, внедрение подтверждено двумя справками (в автореферате написано – одной). Ни одна из упомянутых справок не приложена к диссертации. В то же время Приложение к диссертации содержит семь свидетельств о регистрации программ для ЭВМ, которые нигде

более не упомянуты и не включены в список публикаций автора (а ведь такие свидетельства по действующим правилам приравниваются к публикациям в журналах из списка ВАК).

4. В главе 2 автор указывает на уменьшение размерности модели, разрабатываемой на основе адаптивной модели данных, но не приводит никаких оценок для подтверждения своего утверждения.

5. В главе 4 следовало привести более формальное определение правил, которые используются для вывода знаний на основе введенной онтологии. В четвертой и шестой главах можно было дать более развернутые выводы по результатам изложенного материала.

Отмеченные недостатки по содержанию работы не влияют на общую положительную оценку диссертации, которая выполнена на хорошем научном уровне.

Замечания по оформлению:

Диссертация и особенно автореферат содержат много стилистических погрешностей (расстановка запятых, кавычек и др.):

- общее число запятых близко к нужному, но не везде выполнены правила их расстановки;
- не всегда пояснены впервые введенные аббревиатуры, хотя смысл введенных сокращений в принципе понятен (ЭБ, ПО и др.);
- в названиях конференций и организаций не везде использованы кавычки, что искажает смысл (например, в название Всероссийской научной конференции «Научный сервис в сети Интернет» включено слово «труды», а сама конференция названа конференцией ИПМ им. М.В. Келдыша; названия БЕН и МСЦ РАН в полном написании прописаны строчными буквами);
- библиография оформлена не по единому стандарту, иногда с пропуском годов издания соответствующих цитируемых источников; некоторые записи не содержат необходимый полный объем информации;
- в автореферате на с. 20 без пояснений жирным шрифтом выделены библиографические описания пяти публикаций (видимо, в журналах из Перечня ВАК); рис. 1 и 2 в автореферате не содержат пояснений и трудно воспринимаются; в диссертации на рис. 2 (с. 88) вообще нет ссылок.

Отмечу, что разделы диссертации, содержащие описание технических и программных результатов проведенного исследования, содержат

существенно меньшее количество погрешностей, чем остальные разделы.

В целом считаю, что полученные автором результаты обладают новизной, достоверны, обоснованы, практически значимы и полностью отражены в опубликованных работах. Они доложены и обсуждены на ряде международных и всероссийских конференций и семинаров и получили одобрение. Автореферат соответствует содержанию диссертации.

Диссертационная работа О.М. Атаевой отвечает требованиям ВАК, предъявляемым к кандидатским диссертациям по специальности 05.13.11 – Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей, а ее автор, Ольга Муратовна Атаева, заслуживает присуждения ей ученой степени кандидата технических наук по специальности 05.13.11.

Официальный оппонент, профессор Подпись А.М. Елизаров удостоверяю кафедры «Программная инженерия»

Высшей школы информационных технологий и интеллектуальных систем Казанского (Приволжского) федерального университета, доктор физико-математических наук (01.02.05 – Механика жидкости и газа), профессор, Заслуженный деятель науки Республики Татарстан

Проректор по научной деятельности Казанского (Приволжского) федерального университета, профессор

Александр Михайлович Елизаров

«7» января 2020 г.

Казанский (Приволжский) федеральный университет, 420008, г. Казань, ул. Кремлевская, 35.

Тел. 8-9376151553,

e-mail: amelizarov@gmail.com



Данис Карлович Нургалиев

«7» января 2020 г.