

## ОТЗЫВ

члена диссертационного совета Кознов Дмитрий Владимирович на диссертацию Лещевой Ирины Анатольевны на тему «Метод автоматизированного наполнения баз знаний онтологического типа», представленную на соискание ученой степени кандидата технических наук по специальности 2.3.5. — Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей

Управление знаниями в различных бизнес-контекстах является очень востребованной, и в тоже время наукоёмкой задачей. Различные компании – телекоммуникационные, добывающие, торговые, энергетические и т.д. – производят, хранят и используют огромное количество знаний. В связи с этим консолидация знаний и поддержка их консистентности является очень актуальной задачей, несмотря на обилие существующих методов, приёмов и инструментов. Так, эта задача особенно остро стоит в онтологическом инжиниринге, где исследователи в большей степени обращены к так называемым принципиальным вопросам, оставляя в стороне детали реализации предлагаемых ими подходов.

Представленная диссертация посвящена разработке метода автоматизированной консолидации онтологии из гетерогенных источников наполнения баз знаний онтологического типа. Существующие в этой области методы и программные средства обладают значительными недостатками, в частности, им свойственна узкая направленность на решение задач конкретного проекта, сложность описания моделей источников данных и их трансформации, высокие требования к квалификации пользователей, обеспечение возможности интеграции данных из источников только одного типа, отсутствие модульности и расширяемости и пр. Таким образом, данное диссертационное исследование является актуальным и нужным.

Основными результатами данной диссертационной работы являются следующие.

1. Предложена новая генеалогия языков представления знаний, которая существенно расширяет и дополняет существующие.
2. Полученные новые статистически значимые зависимости между когнитивным типом человека и характеристиками построенной им онтологии.
3. Предложен метод консолидации онтологии из гетерогенных источников, решена проблема разрешения конфликтов (merge conflicts). Разработанный метод не использует дополнительных формальных языков, а построен на основе средств онтологического

моделирования. Также разработан ряд шаблонов для построения правил сопоставления (mapping) исходной онтологии в базу знаний предметной области.

4. Разработана архитектура и создан прототип программного инструмента, реализующий предложенный метод консолидации данных. Данный инструмент был апробирован на четырех реальных проектах в разных предметных областях: автоматизация научных исследований, медицина, автомобильная промышленность, администрирование учебной деятельности ВУЗа.

**Во введении** диссертации обоснована актуальность тематики, сформулированы цель и задачи работы, а также объяснена научная новизна полученных результатов.

**Первая глава** представляет обзор литературы и затрагивает круг вопросов, необходимых для понимания сути проблемы и предложенного решения.

**Вторая глава** начинается с описания проблемы наполнения онтологий и подводит читателя к предложенному методу консолидации данных. Идея метода такова: использовать дополнительную онтологию для описания структуры источников данных с тем, чтобы сформулировать правила отображения между этой дополнительной онтологией и наполняемой базой знаний и воспользоваться автоматическими средствами онтологического инжиниринга. Также в главе описаны архитектура системы и процесс наполнения онтологии.

**Третья глава** представляет собой описание апробации предложенного метода в различных предметных областях: при эмпирических исследованиях, при решении задачи определения диагностики орфанных заболеваний, при классификации сборочных единиц в процессе автоматизированного производства автомобилей, для администрирования учебной деятельности ВУЗа.

**В заключении** приводятся основные результаты работы, сопровождаемые дополнительным обсуждением, а также приводятся данные по научной апробации результатов работы.

В целом диссертация Лещевой Ирины Анатольевны представляет собой интересную и добротную научную работу с хорошей постановкой задачи, интересными обзорами и сопоставлениями. Ясно видна научная школа, в рамках которой выполнено исследование. Работа органично вписана в научный и индустриальный контексты (всё же в первый лучше, чем во второй). Диссертация написана грамотным научным языком, и, хотя местами

повествование резковато, но в целом текст свободно читается и производит позитивное впечатление.

Особо хочется отметить апробацию работы в виде реализации реальных проектов из разнообразных предметных областей. Эти проекты описаны достаточно детально и демонстрируют большой объем работы, проделанный автором. Также важной характеристикой диссертационной работы является большое количество разнообразных публикаций — 8 работ в изданиях, индексируемых Scopus, включая журналы (что является редкостью для работ по данной специальности), 7 – в журналах из перечня ВАК.

Однако данная диссертационная работа не свободна от некоторых недостатков.

1. Мне кажется, что название работы не вполне удачно отражает суть – название сообщает о задаче наполнения онтологических баз данных, и на ум приходит рутинный труд по вводу данных с клавиатуру и автоматизация этого труда. Но на самом деле в работе, во многом, речь идёт об интеграции данных из различных гетерогенных источников, что гораздо содержательнее первого впечатления от названия.
2. Работа, несомненно, выиграла бы, если бы содержала связи с реалиями разработки информационных систем. Хоть автор и ссылается на известную в области управления данных концепцию ETL, но ограничивается лишь верхнеуровневым позиционированием, не используя ни соответствующих методов, ни программных инструментов. Также очень близкой задачей является управление мастер-данными (Master Data Management, MDM), которая как раз и нацелена на извлечение и интеграцию данных из гетерогенных источников. Кроме того, в рамках MDM решается также задача доставки извлеченных данных конечным потребителям – а данный вопрос вовсе опущен в данном исследовании. В целом, создаётся впечатление о том, что автор воспринимает онтологический инжиниринг как нечто отдельное, самоценное, в то время как последний в неявном виде активно используется в различных областях разработки информационных систем (тут часто бывает, как у Мольера – месть Журден не знал, что всю жизнь говорил прозой).
3. На странице 6, прямо в формулировке одного из результатов работы, содержится досадная опечатка: «Предложен новый универсальный алгоритм наполнения онтологий, лишенный большинства недостатков существующих решений интеграции данных распределенной природы в онтологии (см. выше) **в силу предложенного алгоритма**».

4. В тексте наблюдается неправильное, с точки зрения русского языка, оформление вынесенных перечислений. Так, на стр. 11 формально говоря, имеем такое вот предложение «Основные принципы, которым должны удовлетворять онтологии [Gruber, 1995]: 1.». Или в конце стр. 14 имеем после двоеточия набор предложений с большой буквы, ни одно из которых не заканчивается точкой.
5. При описании апробации предложенных результатов в реальных проектах не понятно, насколько эти проекты действительно реальны: были ли они внедрены, имеются ли у разработанных инструментов реальные пользователи и т.д.

В заключении хочу сказать, что несмотря на указанные недостатки, диссертация Лещевой Ирины Анатольевны на тему «Метод автоматизированного наполнения баз знаний онтологического типа» полностью соответствует основным требованиям, установленным приказом № 11181/1 от 19.11.2021 «О порядке присуждения ученых степеней в Санкт-Петербургском государственном университете», а сама соискатель Лещева Ирина Анатольевна заслуживает присуждения ей ученой степени кандидата технических наук по научной специальности 2.3.5. «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей». Нарушения пунктов 9 и 11 указанного Порядка в диссертации не выявлены.

Член диссертационного совета, доктор технических наук, профессор Санкт-Петербургского государственного университета, г. Санкт-Петербург, РФ

/подпись/

Кознов Дмитрий Владимирович

Дата

29 сент 2022 года