

ОТЗЫВ

члена диссертационного совета Городецкого Владимира Ивановича 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

Актуальность темы диссертационной работы. Концепция онтологий как средства формализации семантики предметной области и ее представления в базах знаний, которая была предложена в начале 1990-х годов, в настоящее время приобретает все более важное значение в технологиях искусственного интеллекта, главной движущей силой современных информационных технологий (ИТ). Онтологии – это основа современных семантических моделей и семантических технологий, которые в настоящее время формируют базовые тренды развития искусственного интеллекта, в частности, инструментов интеграции знаний и данных сложных приложений в едином семантическом пространстве. Однако несмотря на длительное время исследований и разработок в области методологий, методов, алгоритмов и инструментов создания онтологий, существующий уровень технологий их создания на индустриальном уровне нельзя признать зрелым – они по-прежнему остаются узким местом в разработке ИТ-приложений. Они остаются трудоемкими и во многом полагаются на ручной труд специалистов в области инжиниринга онтологий. Кроме того, поменялось существо самих онтологий. Если раньше онтологии рассматривались только как средство формализации семантики предметной области, то сейчас они стали неотъемлемой частью технологий машинного обучения, что привело к усложнению проблем интеграции разнородных данных из распределенных источников в онтологию. В этой ситуации онтологии стали рассматриваться также и как средство интеграции гетерогенных х данных и обеспечения интероперабельности подсистем приложений.

Уже этих аргументов достаточно для того, чтобы оценить тематику работы И.А. Лещевой, посвященной теоретическому обоснованию и практическим проблемам автоматизации процессов наполнения онтологий, как актуальную для широкого класса современных бизнес-приложений и приложений за пределами бизнес-систем.

Основные результаты исследования и их научная новизна. В Главе 1 диссертационной работы приводится детальный анализ современного состояния исследований в области онтологий. В частности, анализируются различные определения онтологий и варианты ее формального описания на теоретико-множественном уровне, приводится сравнительный анализ формальных языков представления онтологий в историческом контексте и программных инструментов поддержки онтологического инжиниринга. Эта часть работы демонстрирует высокую квалификацию автора в выбранной области исследований.

Глава 2 представляет научные результаты работы. В ней дается описание разработанного метода и алгоритма наполнения онтологии. Под этой задачей понимается компонента онтологического инжиниринга баз знаний, которая отвечает преобразованию данных из различных хранилищ, баз и отдельных источников в формат примеров базы знаний. Приводится также описание разработанной архитектуры инструмента программной поддержки процессов наполнения онтологий.

Основная научная и прикладная проблема в этой части онтологического инжиниринга состоит в интеграции гетерогенных источников данных организации, возможно, созданных в разное время и хранящихся в разных структурах, а также данных внешних источников, которые используются для управления бизнесом и совершенствования механизмов принятия решений, например, на основе методов машинного обучения.

Глава 3 посвящена демонстрации разработанной компоненты технологии онтологического инжиниринга и доказательству ее практической пригодности на нескольких практических примерах ее использования в конкретных приложениях.

Выводы по главам корректно резюмируют основное содержание и/или основные результаты, полученные в них и их место в диссертационной работе.

Основные новые результаты исследования:

1. Детальный сравнительный критический анализ методов и средств спецификации онтологий и инструментов онтологического инжиниринга, который дает возможность выявить ключевые проблемы в этой области и корректно обосновать цели и направления исследований и разработок диссертационной работы. Представленный анализ демонстрирует высокую квалификацию автора.

2. Метод (авторское название метода – «МЕТЕОР») и архитектура (в виде потока данных и их преобразования) наполнения единой базы знаний из гетерогенных структурированных источников данных. Основная идея метода состоит том, что для источников данных сначала строится онтология типов данных (мета-шаблоны описания типов данных) и процесс извлечения, преобразования и загрузки данных в базу знаний выполняется по типу технологии ETL. В качестве меташаблонов (им отвечают определенные классы в стиле ООП) рассматриваются электронные таблицы, реляционные базы данных и документы XML. Для каждого типа источников извлеченных данных механизм преобразования описывается соответствующим множеством правил их отображения в данные, загружаемые в базу знаний. Эта идея реализована конкретным потоком преобразования данных (автор называет его архитектурой наполнения онтологической базы знаний) для различных вариантов их исходных структур при заданной онтологии предметной области. Предложенный подход, детально описанный множеством классов шаблонов, позволяет динамически расширять множество типов источников. Представлено также детальное описание технологии наполнения баз знаний. Данный результат является основным в работе. Его научная новизна состоит в том, что он базируется на использовании расширяемой онтологии источников данных, что является новой и конструктивной идеей. Реализация разработанного метода в форме программного прототипа и его экспериментальное применение (об этом идет речь в Главе 3) следует рассматривать как аргумент в пользу практической важности результатов работы.

В качестве еще одного научного результата следует отметить конкретные шаблоны построения правил отображения источников данных различных типов в данные онтологии. Результаты, приведенные в главе 3, представляют собой важную методологическую часть исследования, которая демонстрирует технологию использования результатов работы и их программной поддержки для решения задач наполнения онтологий при создании различных приложений. Бесспорно, каждую из выполненных разработок следует рассматривать как фрагмент общего научного исследования, представленного в работе.

Достоверность полученных результатов. Результаты исследования, представленные в работе, являются достоверными. Это следует из детального сравнительного и критического анализа современного состояния исследований по тематике работы в контексте других проблем, связанных с исследуемой. Доказательству достоверности полученных результатов, их теоретической и практической достоверности, по существу, посвящена Глава 3 работы.

Теоретическая и практическая значимость диссертационной работы состоит в том, что ее основные научные результаты фактически заполняют существенный пробел в современной технологии создания онтологий, который касается автоматизации процессов наполнения онтологий примерами понятий. Теоретическая значимость подтверждается также публикацией основных результатов работы в высокорейтинговых российских и зарубежных изданиях. Практическая значимость работы заключается в том, что в ней получены и практически апробированы результаты, которые могут быть положены в основу разработки новых инструментариев создания онтологий индустриального уровня.

Замечания к диссертационной работе:

1. Было бы полезно представить позицию автора по вопросу о наполнении онтологий, представленных в структурах графовых баз данных и знаний.

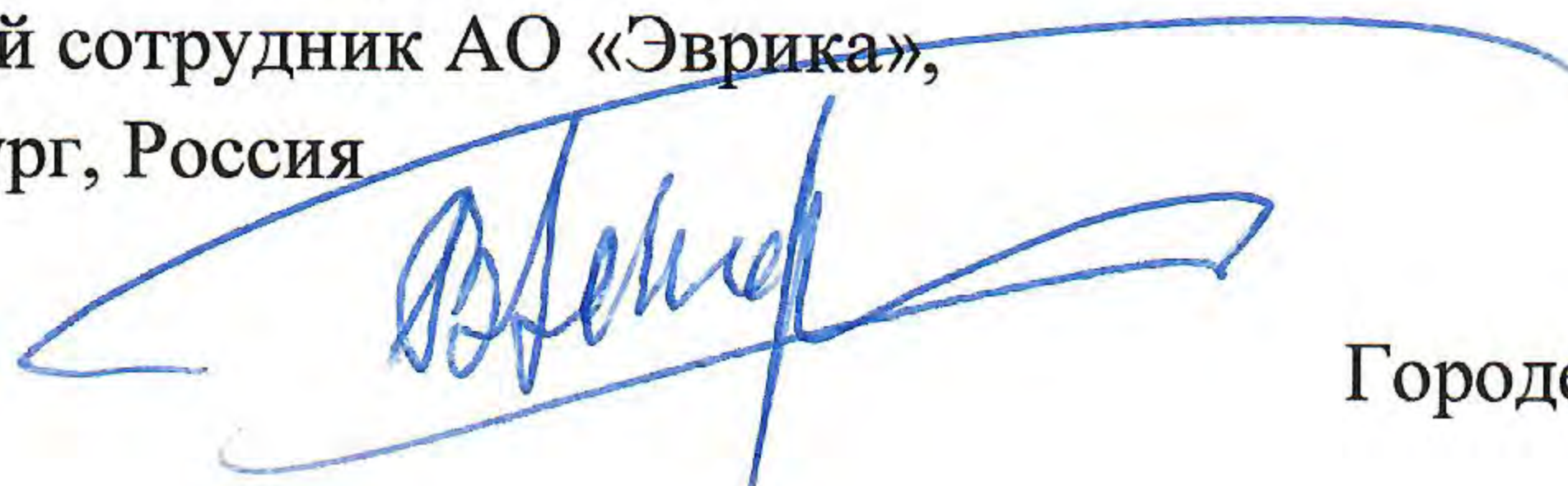
2. В настоящее время большая часть данных является неструктурированными (по мнению многих экспертов, доля таких данных составляет перспективы автоматизации наполнения баз знаний для данных такого типа).

3. В настоящее время существуют более прогрессивные технологии интеграции данных по сравнению с технологией ETL, например, технология виртуализации данных, поддерживаемая многочисленными инструментами индустриального уровня, которые имеют мощные средства трансформации данных в требуемые структуры. Представляется, что эти технологии могли бы решить проблему, исследуемую в диссертационной работе, для более широкого множества структур источников данных.

Выводы. Диссертация Лещевой Ирины Анатольевны на тему «Метод автоматизированного наполнения баз знаний онтологического типа» соответствует основным требованиям, установленным приказом № 11181/1 от 19.11.2021 «О порядке присуждения ученых степеней в Санкт-Петербургском государственном университете», а соискатель Лещева Ирина Анатольевна заслуживает присуждения 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 установлены.

Член диссертационного совета

Доктор технических наук профессор
ведущий научный сотрудник АО «Эврика»,
г. Санкт-Петербург, Россия



Городецкий Владимир Иванович

Дата: 30 сентября 2022 года.