

ОТЗЫВ

официального оппонента
кандидата филологических наук, доцента,
доцента кафедры информатики и прикладной лингвистики
учреждения образования «Минский государственный лингвистический
университет» Детскиной Раисы Владимировны
о диссертации **Голяк Юлии Дмитриевны**
**«Предиктивное автодополнение запросов пользователей в
системах информационного поиска (на материале русского языка)»**,
представленной на соискание ученой степени кандидата
филологических наук по специальности
10.02.21 – прикладная и математическая лингвистика

1. Соответствие диссертации специальности и отрасли науки, по которой она представлена к защите

Диссертационная работа соответствует отрасли науки «филология» и специальности 10.02.21 – прикладная и математическая лингвистика, поскольку она посвящена прикладному лингвистическому исследованию и в ней разработаны концепция, метод и широкий спектр масштабных лингвистических ресурсов решения задачи автодополнения пользовательских запросов (ПЗ), которое реализовано и апробировано в составе известной информационно-поисковой платформы автоматизации инженерии знаний и управления знаниями.

2. Актуальность темы диссертации

К настоящему времени широкое распространение получил такой класс информационных систем, как вопросно-ответные системы с ЕЯ-интерфейсом пользователя, ориентированные на корпоративные полнотекстовые базы данных (ПБД) большого объема. Одним из главных компонентов такого интерфейса является автодополнение ПЗ, которое обеспечивает существенное сокращение общего времени ввода ПЗ и, что очень важно, потенциально обладает возможностью повышения качества формирования ПЗ с точки зрения информационной потребности пользователя.

Существующие решения задачи автодополнения ПЗ (задачи QTA) имеют ряд ограничений: как правило, они ориентированы только на историю информационного поиска и на последовательное завершение строки вводимого запроса. Они не учитывают типы пользовательских запросов и их синтаксических структур, а также само наполнение поискового пространства, что вполне возможно в условиях корпоративных информационных систем. Как результат, создаваемые системы слабо ориентированы на использование лингвистических ресурсов решения задачи

QTA. Поэтому тематика диссертационной работы, посвященной решению задачи QTA, свободному от указанных выше недостатков, безусловно, актуальна.

3. Степень новизны результатов, полученных в диссертации, и научных положений, выносимых на защиту

В диссертации автором получены следующие **новые результаты**, важные как с научной точки зрения, так и для практического использования:

- с учетом недостатков существующих решений задачи QTA сформулирована новая концепция ее решения, ориентированная на предполагаемое поисковое пространство в виде корпоративной полнотекстовой базы данных и ее развитый лингвистический анализ, а также на погружение строки вводимого запроса в контекст;

- на основе анализа внушительного по объему корпуса ПЗ впервые дана классификация основных типов ПЗ и их базовых синтаксических структур, что позволило соискателю разработать метод решения задачи QTA, который в отличие от существующих сводится к автоматическому распознаванию этих структур в ПБД и выбору их лексических наполнений, составляющих базу подсказок для автодополнения пользовательских запросов;

- ориентируясь на сформулированную концепцию и предложенный метод решения задачи QTA, соискателем разработаны структурно-функциональная схема, лингвистическое и алгоритмическое обеспечение системы автодополнения русскоязычных ПЗ, которые отличаются от существующих как по своей функциональности, так и по используемым лингвистическим ресурсам;

- на основе разработанных концептуальных и алгоритмических решений, а также лингвистического обеспечения создан и внедрен в эксплуатацию прототип оригинальной системы автодополнения русскоязычных ПЗ для корпоративных информационных систем, отличительной особенностью которого является предоставление пользователю возможности более точно сформулировать свою информационную потребность, не ориентироваться на историю информационного поиска, получать гарантированно релевантную информацию на поисковой запрос.

4. Обоснованность и достоверность выводов и рекомендаций, сформулированных в диссертации

подтверждается тщательным анализом релевантных тематике диссертационной работы литературных источников, впечатляющей репрезентативностью использованного для достижения поставленной цели

лингвистического материала (корпуса пользовательских запросов объемом порядка 13 тыс. ПЗ и корпуса текстовых документов объемом более 7,5 млн. предложений, являющегося информационной моделью корпоративной ПБД), корректным использованием широкого спектра современных методов исследования в области прикладной и математической лингвистики (контекстного и сопоставительного анализа, компонентного и статистического анализа, алгоритмизации лингвистического анализа текста, экспертного тестирования), апробацией полученных результатов на научных конференциях и на практике в составе известной многоязычной информационно-поисковой платформы IHS Goldfire.

5. Научная, практическая, экономическая и социальная значимость результатов диссертации с указанием рекомендаций по их использованию

Научная значимость результатов диссертации состоит в том, что они вносят существенный вклад в исследование и решение актуальных задач автоматической обработки текста, поскольку затрагивают такие важные аспекты, как эффективная организация требуемого лингвистического материала, анализ возможностей использования уже существующих, хорошо зарекомендовавших себя, и разработка собственных средств автоматического лингвистического анализа текста, экспертное тестирование получаемых решений, их ориентация на применение для других ЕЯ и на серьезное практическое внедрение. В этом смысле данную диссертационную работу можно считать эталонной.

Практическая значимость полученных соискателем результатов высока. Разработанный в ходе выполнения диссертационной работы прототип оригинальной системы автодополнения русскоязычных пользовательских запросов внедрен в составе известной многоязычной промышленной системы автоматизации инженерии знаний и управления знаниями IHS Goldfire, пользователями которой являются более 1500 крупнейших компаний мира. Полученные результаты также внедрены в учебный процесс в МГЛУ. Представлены соответствующие акты о внедрении.

Экономическая значимость заключается в том, что разработанные в диссертации концепция, метод, лингвистическое обеспечение и алгоритмы могут применяться при создании новых эффективных информационных систем и технологий. Разработанный в ходе исследования прототип системы автодополнения русскоязычных ПЗ в результате своего последующего промышленного внедрения позволил существенно повысить эффективность работы пользователей при решении ими задач информационного поиска и инновационных задач.

Социальная значимость полученных в диссертационной работе результатов заключается в возможном их применении в самых

разнообразных сферах человеческой деятельности, требующей оперативной информационно-интеллектуальной поддержки, а также в учебном процессе при подготовке специалистов по информатике и прикладной лингвистике.

Рекомендуется использование результатов диссертационного исследования в научных организациях и учреждениях, занимающихся проектированием и реализацией информационных систем различного назначения, а также в учебном процессе в высших учебных заведениях, осуществляющих подготовку специалистов соответствующего профиля.

6. Опубликованность результатов диссертации в научной печати

По основным теоретическим и практическим результатам диссертации опубликовано 8 научных работ, в том числе 5 статей (две из них в соавторстве) в рецензируемых периодических изданиях, рекомендуемых ВАК РБ (из них четыре – в изданиях по филологическим наукам и одна – по техническим), 3 статьи в сборниках материалов научных конференций. Общий их объем составляет 4,21 авторских листа.

Публикации и автореферат диссертации соответствуют ее содержанию, результатам и положениям, которые выносятся на защиту.

7. Соответствие оформления диссертации требованиям ВАК

Диссертация состоит из введения, общей характеристики работы, трех глав, заключения, библиографического списка и пяти приложений. Разделение диссертации на главы и их содержание соответствует логике изложения и максимальной информативности представляемого материала. Выводы отражают основные результаты работы.

Диссертация оформлена в соответствии с требованиями Инструкции Высшей аттестационной комиссии Республики Беларусь о порядке оформления квалификационной работы (диссертации) и автореферата.

8. Недостатки диссертации

При общей положительной оценке работы хотелось бы обратить внимание на некоторые дискуссионные положения и неточности:

1. Вносились ли соискателем в интересах достижения цели диссертационной работы изменения в русскоязычный компонент многоязычного базового лингвистического процессора (в частности, в его лингвистическую базу знаний)?

2. Обосновывая необходимость включения в используемый текстовый материал англоязычных данных, автор подчеркивает (п. 2.1, с. 32), что она обусловлена, в том числе, потребностью оценить степень

универсальности полученных в работе результатов по отношению к различным ЕЯ. Однако, в тексте диссертации информация по этому поводу не приведена.

3. Очевидно, что построение работающего прототипа системы QTA/R не сводится только к построению его лингвистического обеспечения. Сюда входят также задачи алгоритмизации, программирования, тестирования. Хотелось бы уточнить, что в этом сложном многозадачном процессе реализовано самим соискателем?

4. С предыдущим связан и такой вопрос. Из приведенного в работе сквозного примера видно, что объем формируемой системой базы подсказок очень велик. Насколько это оправдано? Ведь время поиска подходящих подсказок тоже существенно возрастает.

Указанные замечания и неточности не носят принципиального характера, затрагивающего концепцию, основные результаты и выводы диссертационного исследования, которые обладают несомненной новизной, имеют теоретическую и практическую значимость.

9. Соответствие научной квалификации соискателя ученой степени, на которую он претендует

Содержание диссертационной работы свидетельствует об основательной теоретической подготовке соискателя в области прикладной лингвистики, владении современными методами научного исследования и инструментально-программными средствами его эффективной организации и практической реализации приложений. Наличие безусловных аналитических способностей, критического и профессионального подхода к анализу существующих решений проблемы, владение теорией воспроизводящего моделирования лингвистических объектов и процессов, убедительная аргументация принимаемых решений – всё это характеризует соискателя как состоявшегося исследователя, который обладает научной квалификацией, соответствующей ученой степени кандидата филологических наук.

10. Заключение

Диссертационная работа Голяк Юлии Дмитриевны, выполненная под руководством доктора технических наук, профессора Совпеля Игоря Васильевича, является самостоятельно выполненной законченной квалификационной работой, содержащей новые научно обоснованные результаты, совокупность которых является существенным вкладом в развитие теории и практики автоматической обработки естественного языка, соответствует специальности 10.02.21 – прикладная и математическая

лингвистика, а также требованиям пп. 19 и 20 Положения о присуждении ученых степеней и присвоении ученых званий.

Автор диссертационной работы заслуживает присуждения ученой степени кандидата филологических наук за:

- разработку актуальной концепции решения задачи QTA, ориентированной на корпоративные информационные системы, развитый лингвистический анализ поискового пространства и гарантированно релевантную реакцию поисковой системы;
- построение классификации основных типов пользовательских запросов, их базовых синтаксических конструкций и основанного на ней метода решения задачи QTA;
- разработку структурно-функциональной схемы системы автодополнения пользовательских запросов, лингвистического обеспечения и алгоритмов автоматического построения базы подсказок, являющейся основным ресурсом решения задачи QTA;
- разработку, экспертное тестирование и внедрение в промышленную эксплуатацию прототипа оригинальной системы автодополнения русскоязычных ПЗ в корпоративных системах информационного поиска, обладающей высокими качественными характеристиками.

Выражаю свое согласие на размещение моего отзыва о диссертации Голяк Ю. Д. «Предиктивное автодополнение запросов пользователей в системах информационного поиска (на материале русского языка)» на сайте учреждения образования «Минский государственный лингвистический университет».

Официальный оппонент
доцент кафедры информатики
и прикладной лингвистики
УО «Минский государственный
лингвистический университет»,
кандидат филологических наук, доцент


Р. В. Деткина

03.03.2023 г.

