

## ОТЗЫВ НАУЧНОГО РУКОВОДИТЕЛЯ

о диссертационной работе Голяк Юлии Дмитриевны «Предиктивное автодополнение запросов пользователей в системах информационного поиска (на материале русского языка)», представленной на соискание учёной степени кандидата филологических наук по специальности 10.02.21 – прикладная и математическая лингвистика

Диссертационная работа Ю. Д. Голяк посвящена разработке метода, алгоритмов и лингвистического обеспечения решения задачи автодополнения русскоязычных пользовательских запросов (ПЗ) в системах информационного поиска (СИП) и их реализации в виде промышленного прототипа.

Разработка и использование ЕЯ-интерфейса пользователя стала устойчивой тенденцией в построении современных СИП, а одним из главных компонентов в его составе является автодополнение (автоматическое завершение, предиктивный ввод) ПЗ, в дальнейшем QTA (query type-ahead).

Актуальность задачи QTA вызвана, прежде всего, тем, что ее решение обеспечивает существенное сокращение общего времени, затрачиваемого пользователем на ввод поискового запроса. При этом, что очень важно, повышается качество формируемого запроса с точки зрения информационной потребности пользователя. Это особенно проявляется в получивших широкое распространение вопросно-ответных СИП, работающих с полнотекстовыми БД большого объема.

Существующие решения задачи QTA наряду с достоинствами имеют ряд характерных недостатков: ориентация только на историю уже проведенного поиска; отсутствие возможности погружения вводимой строки ПЗ в широкий контекст; слабое использование лингвистической составляющей, в частности, средств автоматического лингвистического анализа текста, классификации основных типов ПЗ и, например, их синтаксических структур; неиспользование самого контента поискового пространства при построении подсказок для автодополнения ПЗ. На

преодоление указанных недостатков и были направлены научные исследования Ю.Д. Голяк.

В диссертационной работе получены следующие основные результаты:

- предложена и обоснована концепция решения задачи QTA, свободная от указанных выше недостатков существующих ее решений;

- получена классификация основных типов ПЗ и их базовых синтаксических структур и предложен основанный на них метод решения задачи QTA, предполагающий автоматическое распознавание этих структур в ПБД и выбор их лексических наполнений, которые в совокупности составляют базу подсказок для автодополнения ПЗ;

- в соответствии с предложенными концепцией и методом разработана структурно-функциональная схема системы QTA/R автодополнения русскоязычных пользовательских запросов в СИП, а также в виде расширения функциональности используемого известного базового лингвистического процессора – собственное лингвистическое обеспечение и алгоритмы автоматического распознавания в ПБД подсказок для автодополнения ПЗ.

- разработан прототип оригинальной системы автодополнения русскоязычных ПЗ, проведено ее тестирование, которое показало, что качество получаемых решений по показателям точности и полноты близко к их абсолютным значениям по отношению к БД САО-отношений и составляет 84-85% по отношению к ПБД, осуществлено внедрение системы в промышленную эксплуатацию в составе известной многоязычной информационно-поисковой платформы IHS Goldfire, используемой для решения инновационных задач многими крупнейшими компаниями мира.

Разработанные методы, алгоритмы и лингвистические ресурсы рекомендуются к использованию при решении различных прикладных задач автоматической обработки текста, информационного поиска и т.п., а также в учебном процессе в

высших учебных заведениях, осуществляющих подготовку специалистов в области интеллектуальных информационных систем и компьютерной лингвистики.

Голяк Ю. Д. закончила в 2014 г. филологический факультет Белорусского государственного университета. В 2016 г. поступила в аспирантуру (форма обучения – соискательство) на кафедру прикладной лингвистики БГУ и окончила ее в 2021 г. За время учёбы в аспирантуре Голяк Ю. Д. проявила себя способным исследователем, владеющим современными методами и технологиями решения актуальных научных и прикладных задач в области прикладной лингвистики, умеющим самостоятельно ставить и решать эти задачи.

В диссертационной работе Голяк Ю. Д. содержатся новые научные результаты по проблеме автодополнения русскоязычных пользовательских запросов, которые широко апробированы и имеют большое практическое значение. Представленная диссертация является законченной научной работой, её основные результаты чётко сформулированы и обоснованы, работа удовлетворяет требованиям, предъявляемым к диссертациям на соискание учёной степени кандидата филологических наук. Все научные публикации Голяк Ю. Д. по теме диссертационной работы выполнены на достаточно высоком уровне.

Учитывая вышеизложенное, считаю, что Голяк Ю. Д. заслуживает присуждения ей учёной степени кандидата филологических наук по специальности 10.02.21 – прикладная и математическая лингвистика – за разработку и внедрение метода, алгоритмов, лингвистического обеспечения и прототипа системы для решения задачи автодополнения русскоязычных пользовательских запросов.

Научный руководитель

доктор технических наук, профессор

И.В. Совпель

