

ОТЗЫВ

научного консультанта на диссертационную работу Мухсиной Куралай Женисбековны «Разработка системы анализа многоязычной текстовой информации на основе машинного обучения», представленную на соискание ученой степени доктора философии по специальности «6D070400 – Вычислительная техника и программное обеспечение».

Диссертационная работа Мухсиной К. Ж. посвящена комплексному исследованию и решению проблемы семантического анализа многоязычной текстовой информации на основе машинного обучения.

Несмотря на большое количество существующих сегодня математических и лингвистических моделей NLP, большинство этих моделей направленно и успешно решает задачи на уровне грамматики, в то время как основные сложности естественного языка, влияющие на качества его автоматического анализа, заложены в семантическом представлении.

Кроме того, существующая сегодня парадигма наращивания лингвистического описания от данных к определяющей структуре, лежащая в основе построения математических моделей от поиска закономерностей в текстовых корпусах к их обобщению, требует отдельного моделирования каждого конкретного языка.

Исходя из вышеперечисленного, диссертационное исследование, посвященное построению моделей, методов и алгоритмов семантического анализа текстов казахского, русского и английского языков, является, безусловно, актуальным.

В качестве основного математического аппарата построения логико-лингвистической модели Open IE для казахского, русского и английского языков выбран и обоснован инструментарий алгебры конечных предикатов; разработанные в диссертации методы морфологического и семантического анализа многоязычных текстов базируются на использовании Hidden Markov Model, использование Vector Space Model позволило построить метод определения семантической близости многоязычных текстовых документов.

Научные положения, выводы и рекомендации, содержащиеся в диссертации, являются достоверными и обоснованными, поскольку они соответствуют фундаментальным теоретическим принципам, согласуются с данными научно-технической литературы, а также подтверждаются результатами математического моделирования и практических экспериментов.

Практическое значение полученных автором научных результатов заключается в программной реализации разработанных моделей и методов, ис-

пользование которых позволяет повысить эффективность работы систем автоматического обработки многоязычной текстовой информации, а именно: (1) созданного приложения для определения семантической близости текстов к заданной узкоспециализированной тематике; и (2) созданного параллельного аннотированного корпуса текстов криминально-окрашенной информации казахского и русского языков.

Таким образом, результаты диссертационного исследования Мухсиной К. Ж. имеют научную и практическую значимость, а использование результатов обеспечивает решение важной прикладной задачи автоматизации обработки и анализа многоязычной текстовой информации.

Учитывая вышеизложенное, считаю, что диссертационная работа Мухсиной Куралай Женисбековны на тему «Разработка системы анализа многоязычной текстовой информации на основе машинного обучения» может быть рекомендована к защите по специальности «6D070400 – Вычислительная техника и программное обеспечение».

Научный консультант, д.т.н., профессор,
профессор кафедры интеллектуальных
компьютерных систем Национального
технического университета «Харьковский
политехнический институт»



Н. Ф. Хайрова