

ОТЗЫВ НА АВТОРЕФЕРАТ ДИССЕРТАЦИИ

Ильовского Дмитрия Алексеевича

«Методы и алгоритмы обработки текстовых данных
на основе графовых дискурсивных моделей»,

представленной на соискание ученой степени кандидата технических наук по специальности
05.13.18 – Математическое моделирование, численные методы и комплексы программ

В диссертационной работе описано применение математических методов для решения задач по обработке текстовых данных на английском языке. Автор предлагает модель представления («чашу разбора») текстового абзаца, дополненную операцией сходства между абзацами текста. Предложенная модель позволяет решать целый ряд задач, связанных с повышением релевантности информационного поиска и интеллектуальной обработкой результатов этого поиска.

Исследуемую Д.А. Ильовским тему следует считать актуальной, поскольку в настоящий момент на решение задач, связанных с интеллектуальной обработкой текста за счет привлечения дополнительных лингвистических сведений, направлены усилия как ведущих международных и российских компаний, так и академических исследовательских групп в различных странах мира.

Главным преимуществом работы Дмитрия является создание им математического и программного обеспечения, которое может быть использовано программистами, не обладающими глубокими познаниями в математике и лингвистике, которое позволит им улучшить качество работы существующих алгоритмов и программ.

В числе основных достоинств работы можно перечислить следующее:

- Автор проделал существенный объем работы, изучив значительное количество научных источников, причем не только математических, но и лингвистических.
- Разработанные методы и алгоритмы используются в конкретных прикладных проектах для решения реальных задач.
- Все основные результаты диссертации были опубликованы (в том числе, в изданиях, рекомендованных ВАК), автор сделал доклады по результатам проделанной работы на крупнейших конференциях в своей области, включая конференции CICLING-2014, ACL-2014, RaNLP 2015, ACL-2015.

Конечно, необходимо отметить и несколько недостатков работы:

- Хотя работа посвящена в первую очередь анализу текстов на английском языке, автору стоило пояснить, почему рассматриваемые в работе модели не могут быть (полностью или частично) применены для других языков.
- Из реферата не вполне ясно, все ли перечисленные типы связей внутри абзаца использовались автором в практических приложениях. Этот момент стоило отметить явно.
- При сравнении предложенных автором моделей представления текстовых данных с существующими не рассмотрены одни из наиболее популярных современных моделей типа word2vec и doc2vec.
- В автореферате не приведен алгоритм или другое формальное описание расчета предложенной автором ядерной (kernel) функции для оценки близости документов, хотя это один из важных результатов.

Вместе с тем, указанные недостатки ни в коей мере не влияют на общий положительный вывод. Диссертация Д.А. Ильовского отвечает всем требованиям, предъявляемым ВАК к специальности 05.13.18, а ее автор заслуживает присуждения степени кандидата технических наук.

Доцент ВМиК МГУ
кандидат физико-математических наук
Петровский Михаил Игоревич

Адрес: 119991, Москва, ГСП-1, Ленинские горы, д. 1, стр. 52, факультет ВМК.
E-mail: michael@cs.msu.su, телефон: +7 495 9391789

