

Отзыв на автореферат кандидатской диссертации Ильвовского 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.

Диссертационная работа Ильвовского Д.А. посвящена актуальной области информационного поиска, связанной со структурным представлением текстов. В работе рассматривается решение следующих основных задач:

- создать формальные модели текстов на естественном языке, которые отражают не только их теоретико-множественную, а также и синтаксическую, и дискурсивную структуры;
- разработать на основе предложенных моделей комплекс алгоритмов и программ, предназначенных для поиска, классификации и кластеризации текстовых данных, а также опережающих по эффективности частотные модели, активно применяемых в настоящее время в индустрии, в частности, в современных поисковых системах.

Результат поиска текстов (документов) по запросу в таких системах представляется в виде последовательности документов, расположенных в порядке убывания их релевантности запросу. Эффективность, или качество, поиска определяется тем, насколько этот порядок близок к содержательным представлениям о релевантности документов. В абсолютном большинстве случаев массовый пользователь просматривает ограниченное число (от 10 до 100) выданных поисковой системой документов. Поисковая система тем лучше (качество обеспечиваемого ею поиска тем выше), чем скорее пользователь найдет текст, содержащий интересующую его информацию, то есть чем меньше бесполезных для него документов ему придется просмотреть. Следует отметить, что структуры, используемые современными поисковыми сервисами, недостаточно хорошо отражают смысловую структуру текста, что в ряде случаев проявляется в низком качестве их работы.

В диссертации показано, что, с одной стороны, возможность автоматически учитывать синтактико-дискурсивную структуру текста позволяет существенно повысить качество поиска. С другой стороны, сложность синтаксического и дискурсивного анализа текста не позволяет эффективно реализовать его на очень больших поисковых массивах. Тем не менее, данный метод позволяет сделать это при обработке, например, первых 50 или 100 выданных первичным поиском документов, что позволяет применять предлагаемый анализ на этапе «переранжирования» современных глобальных поисковых систем в Интернете (таких как, Яндекс и Google). В качестве недостатка работы стоит отметить, что в автореферате (в отличие от диссертации) это соображение не высказано в явном виде. Также в автореферате некоторые представленные результаты не снабжены информацией об уровне статистической значимости превосходства качества одних алгоритмов над другими (см. таблицы 3.1-3.4 и 4.1-4.3). Эти недостатки не снижают теоретической и практической ценности работы.

Результаты работы являются новыми, реализованы в виде программного обеспечения, апробированы на реальных задачах, а также опубликованы в рецензируемых трудах ведущих международных конференций и в 3 статьях в журналах из списка ВАК. Считаю, что диссертация «Методы и алгоритмы обработки текстовых данных на основе графовых дискурсивных моделей» соответствует заявленной научной специальности 05.13.18, обладает научной новизной и практической ценностью, соответствует квалификационным требованиям ВАК, предъявляемым к кандидатским диссертациям, а ее автор, Ильвовский Дмитрий Алексеевич, заслуживает присвоения ему ученой степени кандидата технических наук.

ПОДПИСЬ  
ЗАВЕРЯЮ

БЕЛОВА Л. Г.

РУКОВОДИТЕЛЬ СЛУЖБЫ ПО РАБОТЕ  
СО ШТАТНЫМИ СОТРУДНИКАМИ  
НА ОСНОВАНИИ ДОВЕРЕННОСТИ № 51  
ОТ 10 МАРТА 2016 Г.



11.05.2017  
Друца Алексей Валерьевич,  
кандидат физико-математических наук

Разработчик-исследователь программного обеспечения  
ООО «Яндекс»

119021, Москва, ул. Льва Толстого, 16