

ОТЗЫВ НА АВТОРЕФЕРАТ 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»

Настоящее время характеризуется большими объемами электронной текстовой информации, которую необходимо анализировать, осуществлять в ней поиск, подвергать разного рода обработке типа классификации, кластеризации, извлекать различную информацию. Тексты на естественном языке представляют собой чрезвычайно сложные структурные объекты, однако преобладающей технологией представления содержания связного текста является подход на основе мешка слов, не учитывающий синтаксические и семантические взаимосвязи между словами. Это связано с тем, что текст содержит большое количество такого рода связей, которые трудно эффективно имплементировать в алгоритмы. В то же время ясно, что для дальнейшего улучшения качества обработки текстов необходимо учитывать дополнительные знания о языке и об окружающем мире.

Диссертация Ильвовского Д.А. посвящена использованию в задачах информационного поиска, классификации и кластеризации текстов нескольких структурных типов отношений между словами, а именно используются синтаксические деревья разбора предложений, отношения кореференции, структура риторических отношений между предложениями в тексте.

В первой главе приводятся основные определения, связанные с частично упорядоченными множествами и решетками, решетками замкнутых описаний, синтаксическими и семантическими моделями представления текста. Также вводится модель структурного представления текстовых абзацев – чаща разбора.

Во второй главе описывается графовая модель представления структуры текстовых абзацев и её применение в задаче информационного поиска для английского языка. Рассматриваются методы вычисления полного и приближенного структурного сходства текстовых абзацев, определяется проекция структурного представления текстового абзаца в виде расширенных синтаксических групп.

В третьей главе описывается применение построенной модели для задачи обучения с учителем на текстовых абзацах (для английского языка), основанное на использовании ядерных функций (kernels) в методе опорных векторов (SVM). Демонстрируется

преимущество применения новой модели в задаче классификации поисковых результатов.

В четвертой главе рассматривается задача выявления тождественных денотатов для случая формальных описаний, построенных на основе предварительно обработанных текстовых данных. В пятой главе приводится описание построенного в рамках исследования программного комплекса, реализующего разработанные в исследовании модели и методы.

В целом, работа представляет интересный подход к представлению текстовых абзацев для улучшения качества приложений автоматической обработки текстов.

Замечания к автореферату:

1. В автореферате неоднократно цитируется работа Alessandro Moschitti, однако нигде не дается ссылка на его конкретную работу,
2. Десятичные числа записываются то через запятую, то через точку.

Указанные недостатки являются несущественными и не влияют на понимание сути полученных результатов.

Исходя из анализа текста автореферата, можно заключить, что диссертация Ильвовского Дмитрия Алексеевича является законченной самостоятельной научно-исследовательской работой. Принимая во внимание актуальность темы диссертации, научную новизну и практическую значимость ее результатов, считаю, что диссертационная работа «Методы и алгоритмы обработки текстовых данных на основе графовых дискурсивных моделей» удовлетворяет всем требованиям ВАК при Минобрнауки России, предъявляемым к кандидатским диссертациям, а ее автор, Ильвовский Дмитрий Алексеевич, безусловно, заслуживает присуждения ученой степени кандидата технических наук по специальности 05.13.18 – "Математическое моделирование, численные методы и комплексы программ".

Лукашевич Наталья Валентиновна

д.т.н.

ведущий научный сотрудник НИВЦ МГУ им. М.В. Ломоносова

8-926-1446163

louk_nat@mail.ru

119234, Москва, Ленинские горы 1/4

22.05.2017

НАЧАЛЬНИК ОТДЕЛА
КАДРОВ

НИИ ПАУ

