

ОТЗЫВ НА АВТОРЕФЕРАТ ДИССЕРТАЦИИ

Лукашевич Натальи Валентиновны

“Модели и методы автоматического анализа текстов на основе лингвистических онтологий в информационно-аналитических системах”,

представленной на соискание ученой степени доктора технических наук по специальности 05.13.17 – Теоретические основы информатики

В настоящее время актуальной является обработка больших текстовых коллекций и потоков неструктурированной текстовой информации. Для обработки больших объемов поступающей информации часто используются статистические методы, без учета знаний о понятиях и терминах предметной области.

В диссертации Лукашевич Н.В. рассматривается специальный вид компьютерных ресурсов для автоматической обработки текстов – так называемые лингвистические онтологии, которые содержат как описание предметной области в виде понятий и отношений между ними, так и способы выражения понятий в тексте посредством разного рода языковых выражений.

Особое внимание в работе уделено выбору и обоснованию набора отношений между понятиями в предлагаемой модели лингвистической онтологии, которые должны быть применимы в широких предметных областях, которые включают большое количество различных типов объектов, действий, процессов; переносимы между предметными областями.

По описанной в работе модели лингвистической онтологии был создан ряд сверхбольших компьютерных ресурсов для автоматической обработки текстов, включая Общественно-политический тезаурус, Тезаурус русского языка, Авиа-Онтология и др.

В работе представлен ряд методов использования созданных лингвистических онтологий для различных приложений автоматической обработки текстов в рамках текстовых информационных систем: анализ основных тем, обсуждаемых в документе (т.н. тематическое представление); расширение поискового запроса (отдельно исследуются особенности расширения длинных поисковых запросов); автоматическая классификация документов; автоматическое аннотирование одного или многих документов.

Отдельная глава посвящена методам извлечения терминологии из текстов предметной области, для чего используется модель, основанная на учете различных лингвистических и статистических характеристик терминов.

По работе можно сделать следующее замечание:

- в автореферате не хватает примеров лексических цепочек, тематических узлов, обсуждаемых в третьей и четвертой главе диссертации.

Указанное замечание не снижает значимость представленной работы.

Диссертация Лукашевич Натальи Валентиновны является законченной самостоятельной научно-исследовательской работой, совокупность результатов которой можно квалифицировать как существенное продвижение в решении актуальной научной проблемы создания сверхбольших лингвистических ресурсов для автоматической обработки текстов, автоматического анализа структуры связного текста.

Принимая во внимание актуальность темы диссертации, научную новизну и практическую значимость ее результатов, считаю, что диссертационная работа «Модели и методы автоматического анализа текстов на основе лингвистических онтологий в информационно-аналитических системах» удовлетворяет всем требованиям ВАК РФ, предъявляемым к докторским диссертациям, а ее автор, Лукашевич Наталья Валентиновна, безусловно, заслуживает присуждения ученой степени доктора технических наук по специальности 05.13.17 – "Теоретические основы информатики".

Зав. лабораторией искусственного интеллекта
Федерального государственного бюджетного
учреждения науки Институт систем информатики
им. А.П. Ершова Сибирского отделения
Российской академии наук,
к. т. н.

Ю.А. Загорулько

21 сентября 2016 г.

Адрес: 630090, Новосибирск, проспект Лаврентьева, 6

Телефон: (383) 3308652, +7 9137852655

E-mail: zagor@iis.nsk.su

