

ОТЗЫВ

на автореферат диссертации Лунева Кирилла Владимировича
«Теоретико-графовые алгоритмы выявления семантической близости между
понятиями на основе анализа наборов ключевых слов взаимосвязанных
объектов»

на соискание учёной степени кандидата физико-математических наук по
специальности 05.13.17 – теоретические основы информатики
(физико-математические науки)

Судя по автореферату, диссертационная работа посвящена моделям семантического анализа объектов научных информационных систем, включая наукометрические системы. Объектами таких систем могут выступать персоналии, научные события (например, конференции), организационные сущности (например, лаборатории), публикации и др. Объекты могут быть связаны отношениями различных типов. Например, сотрудник является членом организации, участником научно-технической конференции, автором публикаций и т.п.

В диссертации исследуется особый случай или вид научных информационных систем, в которых объекты характеризуются небольшим доступным объемом информации и узкой направленностью: объекты такой системы описываются конечным множеством ключевых слов, а общее число объектов системы сравнительно невелико. Подобные случаи встречаются на практике и поэтому поставленная задача является актуальной.

Основными научными результатами выполненного исследования являются следующие:

1. Разработан ряд моделей вычисления уровня семантической близости между ключевыми словами интеллектуальной системы. Программные реализации моделей позволяют вычислять такую близость, автоматически учитывая специфику системы, в которую модели внедряются. Проведены многочисленные тестовые испытания, подтверждающие высокий уровень полученных результатов. Получены аналитические оценки, характеризующие вычислительную сложность программных реализаций этих моделей.
2. Созданы уникальные методы автоматической генерации обучающей выборки для определения семантически похожих ключевых слов. Разработанные методы избавляют от больших трудозатрат для выставления экспертных оценок. Кроме того, алгоритмы генерации

позволяют использовать методы машинного обучения с учителем для решения задачи семантической близости пары слов. Это обстоятельство в значительной мере упрощает подбор параметров модели и улучшает качество определения семантической близости.

3. Разработана модель и ее программная реализация для вычисления семантической близости между парой объектов. Для решения задачи были использованы дополнительные связи между сущностями системы. Модель протестирована на данных из информационно-аналитической системы (ИАС) «ИСТИНА». Показано, что модель удовлетворяет предъявляемым к ней требованиям.
4. На основе разработанных моделей решены важные, востребованные практикой задачи поиска экспертов в различных областях научных знаний, кластеризации ключевых слов, определения тематической направленности объекта информационно-аналитической системы. Программные решения опробированы на данных из ИАС «ИСТИНА», включены в состав ее программного обеспечения, а также получили высокие оценки качества квалифицированных экспертов.

В результате создан программный комплекс, который является более гибким решением для научных информационных систем определенного вида, не обладающих большим объемом данных. Данный результат диссертации имеет научную новизну, а проведенная работа имеет теоретическую и практическую научную значимость.

Достоинством работы является хорошая логическая организация и структурирование материала. Также, достоинством работы является тщательно продуманные и удачно реализованные тестовые испытания для всех полученных результатов, включая привлечение экспертов в некоторых случаях. Представляется ценным проведенные автором теоретическое обоснование и практическая классификация трех типов анализируемых ключевых слов: абстрактные, тематические и ключевые слова-термины.

Содержание автореферата меня заинтересовало, и я прочитал текст диссертации. К сожалению, в тексте диссертации встречаются опечатки, стилистические погрешности (например, на стр. 27, 2-й абзац), пропущенные слова в предложениях (например, на стр. 29, 2-й абзац, первое предложение) и т.п. Данные недостатки главным образом встречаются в первой главе диссертации.

Отмеченные недостатки не уменьшают ценность проделанной работы и полученных результатов. Хотя разработанный подход относится к анализу семантической близости объектов научных информационных систем на основе небольшого количества ключевых слов и связей между объектами, а такие ситуации встречаются все реже, но появляются новые области, где данный подход может применяться. Например, для анализа тематической структуры содержания цитирований в полных текстах публикаций. Контексты цитирований, как правило, представлены небольшими текстовыми фрагментами и представленный подход мог бы дать хорошие результаты применительно к их тематическому анализу.

Судя по проведенной в автореферате информации, результаты диссертационного исследования прошли достаточную апробацию. Это показало, что ценность выполненных автором научных работ в том числе заключается в востребованности решаемых задач на практике, в частности, в информационно-аналитической системе ИСТИНА.

Автореферат, подготовленный по диссертации, удовлетворяет всем требованиям Положения о присуждении ученых степеней в Московском государственном университете имени М.В. Ломоносова, предъявляемым к кандидатским диссертациям, а соискатель Лунев К.В. заслуживает присуждения ему ученой степени кандидата физико-математических наук по специальности 05.13.17 – теоретические основы информатики.

Главный научный сотрудник
ЦЭМИ РАН, д.т.н.

С.И. Паринов

Подпись		ЗАВЕРЯЮ
Зав. ОК		

