

## ОТЗЫВ

на автореферат диссертации Соченкова Ильи Владимировича  
«Реляционно-ситуационные структуры данных, методы и  
алгоритмы решения поисково-аналитических задач»,  
представленной на соискание ученой степени кандидата физико-математических  
наук по специальности 05.13.17 – «Теоретические основы информатики»

Одной из преобладающих тенденций в области информационного поиска в настоящее время является исследование и развитие методов обработки текстов, основанных на использовании результатов лексико-морфологического, синтаксического и семантического анализа. Переход от статистических критериев, основанных на анализе частотности отдельных слов, к критериям, учитывающим синтаксические и семантические связи между словами в предложениях, позволяет повысить качество результатов при решении таких задач, как поиск информации по запросу, поиск текстовых заимствований, выявление значимых фрагментов текстов для автоматического аннотирования и многих других. Однако в промышленных поисково-аналитических системах, предназначенных для обработки масштабных коллекций текстов, использование результатов синтаксического и семантического анализа ограничено в виду отсутствия эффективных методов хранения и обработки этой информации. Получаемые в ходе полного лингвистического анализа древовидные и сетевые структуры требуют применения реляционных или графовых баз данных, которые, однако, не решают проблемы, связанные с большими объёмами хранимых данных, скоростью наполнения баз и скоростью выборки информации. В диссертации И.В. Соченкова предлагаются средства и методы решения этих проблем, что свидетельствует об актуальности ее тематики.

Диссертация И.В. Соченкова посвящена исследованию методов поиска текстовой информации, основанных на оценке сходства текстов с применением лексико-морфологических признаков словоупотреблений, а также синтаксических и семантических структур предложений. Предложен метод оценки сходства текстов для решения задач полнотекстового поиска. Проблема индексации древовидных и сетевых структур, являющихся результатом семантико-синтаксического анализа масштабных коллекций текстов, решена И.В.Соченковым путём разработки новых структур данных и алгоритмов обработки инвертированного индекса. Предложенные и исследованные в работе реляционно-ситуационные структуры данных позволяют хранить и эффективно использовать при решении задач информационного поиска не только частотные характеристики отдельных слов, повсеместно применяемые в поисковых системах, но также синтаксические и семантические связи между словоупотреблениями в предложениях текстов. Разработанная автором модель данных позволила формально описать предложенные структуры данных и операции над ними. С её помощью доказан ряд утверждений о свойствах этих операций и реализующих их

алгоритмов, которые используются для обработки реляционно-ситуационных структур данных. Эксперименты, выполненные автором в ходе исследований, воспроизводимы и подтверждают работоспособность и эффективность предложенных структур данных и алгоритмов.

В целом кандидатская диссертация И.В. Соченкова представляет собой законченное научное исследование, выполненное на высоком уровне и отвечающее требованиям ВАК РФ, а соискатель ученой степени заслуживает присуждения ему учёной степени кандидата физико-математических наук по специальности 05.13.17 – «Теоретические основы информатики».

Зав. лабораторией Института проблем управления им. В.А. Трапезникова РАН

доктор технических наук, профессор

О.П.Кузнецов

02.10.2014

