

ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА

на диссертацию И.В.Соченкова "Реляционно-ситуационные структуры данных, методы и алгоритмы решения поисково-аналитических задач", представленную на соискание ученой степени кандидата физико-математических наук по специальности 05.13.17 – теоретические основы информатики

Диссертационная работа И.В.Соченкова посвящена проблеме поисково-аналитической обработки больших коллекций полнотекстовых документов, представленных в сети Интернет. Актуальность исследования не вызывает сомнений, в связи с нарастающим противоречием между потребностью в «глубокой» обработке доступных массивов документов и резким ростом количества документов.

Диссертация состоит из введения, четырех глав, заключения и списка литературы.

В первой главе диссертации приводятся обзор существующих методов компьютерного анализа текстов для решения задач информационного поиска, а также обзор индексных структур данных и методов ранжирования результатов поиска в информационно-аналитических и поисковых системах. Обзоры позволили автору вскрыть современные проблемы информационного поиска и наметить пути их решения, развитые в следующих главах диссертации. В качестве основной задачи научного исследования И.В.Соченков вполне обоснованно выбирает задачу сравнения двух текстов. В части представления данных выделяется проблема эффективной реализации большого объема разнообразной справочной информации.

Во второй главе диссертации излагается новый метод многокритериальной оценки сходства текстов на основе лексико-морфологической, синтаксической и семантической информации. Автор предлагает рассматривать представление текстовой информации посредством достаточно сложной формальной модели, опирающейся на универсальные словари лемм и лексем и включающей ряд отношений и функций предпочтения. В предложенной модели обладающий внутренним единством текст описывается признаками пяти классов. На базе разработанного формализма вводятся частные критерии оценки сходства текстов, а для нахождения общей оценки предлагается использовать взвешенную сумму частных критериев. Особый интерес во второй главе представляют доказанные утверждения-свойства критериев оценки сходства текстов; эти утверждения существенно используются алгоритмами оценки сходства текстов и алгоритмом ранжирования результатов поиска.

В третьей главе диссертации на основании разработанной модели приводятся и обосновываются структуры данных и алгоритмы формирования поисковых индексов, способ представления поискового запроса, а также алгоритмы оценки релевантности и ранжирования информационного поиска. В интересах точного описания разработанных структур данных и алгоритмов предлагается формальная модель данных, базирующаяся на строгих определениях ряда понятий и позволяющая формализовать операции над ними. Для всех алгоритмов приводятся и доказываются оценки сложности, позволяющие предположить высокую практическую эффективность предлагаемой технологии сравнения текстов.

Четвертая глава диссертации посвящена вопросам программной реализации описанных структур данных и алгоритмов в информационно-аналитических системах. Здесь приводятся и анализируются экспериментально полученные оценки теоретически разработанных структур данных и

алгоритмов. Проверка проводилась на весьма представительной коллекции нормативных документов. Экспериментально установлено значительное повышение качества информационного поиска, что выгодно отличает предложенный подход от иных решений сходных задач.

Основными результатами, полученными в рамках диссертационного исследования, являются

- ▲ новые критерии оценки сходства текстов;
- ▲ новый метод многокритериальной оценки сходства текстов;
- ▲ новая модель данных, ориентированная на описание алгоритмов поисково-аналитической обработки текстов;
- ▲ программная реализация и экспериментальная проверка разработанных подходов.

Совокупность результатов полученных в диссертации позволяет их рассматривать как решение задачи сравнения текстов, имеющей значение для развития информационного поиска.

Достоинством диссертации является формальный подход к задаче оценки сходства текстов, позволивший выявить значительное количество строго доказанных свойств, определяющих содержание и алгоритмов, и структур данных. Достоверность полученных результатов основывается на том, что сформулированные в диссертации утверждения строго доказаны, а алгоритмы дополнительно проверены на практических задачах. Также считаю необходимым отметить высокое качество диссертации-как-текста, что – в совокупности – свидетельствует о высокой научной квалификации диссертанта.

К недостаткам диссертационной работы следует отнести:

1. Отсутствие подробного исследования особенностей поисково-аналитических задач.

2. Перегруженность изложения § 2.2, в котором без разбиения на подразделы излагаются и частные критерии сходства текстов, и общий критерий, и формальные исследования свойств.

3. Конспективный стиль описания программной реализации алгоритмов и структур данных.

Перечисленные недостатки, тем не менее, не снижают общую положительную оценку диссертационной работы И.В.Соченкова.

Автореферат диссертации полно раскрывает ее содержание и удовлетворяет всем требованиям Положения о присуждении ученых степеней, утвержденном Правительством Российской Федерации (Положения). В целом диссертационная работа является законченным научным исследованием, удовлетворяет всем требованиям Положения, предъявляемым к диссертациям на соискание ученой степени кандидата физико-математических наук, а ее автор – Соченков Илья Владимирович – заслуживает присуждения ученой степени кандидата физико-математических наук по специальности 05.13.17 – теоретические основы информатики.

Профессор кафедры математики, логики и интеллектуальных систем в гуманитарной сфере отделения интеллектуальных систем в гуманитарной сфере Российского государственного гуманитарного университета, д.ф.-м.н, проф.


29.09.14

О.М.Аншаков

Аншаков Олег Михайлович, РГГУ,
125993, г. Москва, Миусская площадь, д.6, +7(499)250-63-29, oansh@yandex.ru

Подпись О.М. Аншакова

УДОСТОВЕРЯЮ

Ученый секретарь Трант



29.09.2014