

ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА
на диссертацию **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»

Задачами диссертационной работы И.В. Соченкова является разработка и исследование метода многокритериальной оценки сходства текстов с учётом информации, получаемой в результате морфологического, синтаксического и семантического анализа, а также разработки и исследование структур данных и алгоритмов для реализации этого метода в информационно-аналитических системах. Целью исследования И.В. Соченкова является повышение вычислительной эффективности, а также увеличение полноты и точности решения поисково-аналитических задач.

Актуальность. Поставленные задачи актуальны, поскольку повышение качества решения известных задач информационного поиска, а также решение новых задач, возникающих в области анализа больших текстовых коллекций (например, в Интернете), требуют применения методов полного лингвистического анализа текстов. Однако в настоящее время недостаточно проработаны эффективные способы хранения и обработки результатов полного лингвистического анализа больших коллекций текстов. Распространенные технологии информационного поиска опираются на векторную модель текста и используют инвертированные индексы для хранения и эффективной выборки векторов признаков, характеризующих текстовые документы. Результаты синтаксического и семантического анализа представляют собой древовидные и сетевые структуры, что ограничивает возможности их хранения и обработка с помощью инвертированных индексов. Альтернативные решения этой задачи (реляционные постреляционные базы данных) не обладают вычислительной эффективностью при работе с количеством объектов и связей необходимым для обработки больших коллекций текстов. Таким образом, И.В. Соченковым в диссертации решены актуальные научные задачи: показано, как результаты морфологического, синтаксического и семантического анализа могут быть совместно использованы для оценки структурного сходства текстов по предложениям, а также предложен способ хранения результатов лингвистического анализа в инвертированных индексах специальной структуры.

Диссертация И.В. Соченкова состоит из введения, четырёх глав, заключения, списка сокращений и условных обозначений, а также списка использованной литературы, включающего 152 источника. Диссертация содержит 148 страниц, 25 рисунков и 2

таблицы. Во введении обоснована актуальность темы, определен предмет исследования, сформулированы цель и задачи исследования, научная новизна, а также охарактеризована практическая и теоретическая значимость результатов исследования.

В первой главе приведен обзор методов компьютерного анализа и информационного поиска текстовой информации, а также рассмотрены структуры данных для обработки текстовой информации в задачах информационного поиска. Вторая глава содержит описание предложенного И.В. Соченковым метода оценки сходства текстов, определены критерии сравнения текстов по предложениям, сформулированы и доказаны утверждения об их свойствах. Третья глава посвящена модели данных, структурам данных и алгоритмам решения поисково-аналитических задач с применением метода оценки сходства текстов. В частности, приведено описание разработанных автором алгоритмов индексации текстовой информации, а также алгоритмов оценки сходства текстов с помощью предложенного инвертированного индекса, предназначенного для эффективного хранения результатов полного лингвистического анализа. Кроме того, сформулированы и доказаны утверждения об оценках вычислительной сложности указанных алгоритмов. В этой же главе доказаны утверждения о свойствах операций над данными, введенных И.В. Соченковым в рамках модели данных, применяемой для исследования индексных структур. Четвёртая, заключительная, глава посвящена экспериментальному исследованию метода оценки сходства текстов, структур данных и алгоритмов информационного поиска, реализованных автором диссертации. В заключении приведены основные результаты и выводы диссертационной работы, а также рассмотрены направления дальнейших исследований.

Научную новизну диссертации составляют разработанные структуры данных инвертированного поискового индекса, предложенная и исследованная модель данных, предназначенная для поисково-аналитической обработки текстовой информации, метод многокритериальной оценки сходства текстов.

Практическая значимость диссертации определяется разработанными для реализации информационных систем алгоритмов и структур данных для решения поисково-аналитических задач, реализацией метода многокритериальной оценки сходства текстов, в основе которого лежит сопоставление лексико-морфологической, синтаксической, семантической информации.

Достоверность и обоснованность результатов диссертации подтверждена строгими математическими доказательствами сформулированных автором утверждений. О достижении целей работы свидетельствуют данные экспериментального исследования.

Результаты работы обоснованы и достаточно полно отражены в публикациях автора по теме исследования; по теме диссертации И.В. Соченковым опубликовано 14 работ: 6 из них в рецензируемых журналах, входящих в список ВАК РФ, 8 работ опубликованы в материалах российских и международных конференций.

По работе следует сделать ряд замечаний:

1. Из текста диссертации неясно, что происходит с последовательностями, которые не являются словами текста естественного языка, например алфавитно-цифровыми последовательностями. Производится ли их отбрасывание или информация об их вхождении также сохраняется в индексе.
2. В диссертации не описано как организовано хранение списков в случае когда оперативной памяти недостаточно. При этом в экспериментах задействовано не большие объемы данных (до 10 ГБ).
3. Объемы полученных индексов автор сравнивает только между собой. При этом не объясняется, как объем индекса вообще соотносится с объемом исходных данных.

Приведенные замечания не снижают общей высокой оценки диссертационной работы.

Заключение.

Новизна конструкции инвертированного индекса предложенного Саченковым И.В. состоит в разработанном им способе хранения морфологической и семантической информации о конкретном словоупотреблении. Такое решение может быть использовано для эффективного фразового и семантического видов поиска (в том числе вопросно-ответного), а также для оценки релевантности найденных документов.

Результаты диссертационного исследования И.В. Соченкова могут быть использованы в университетах и научных учреждениях, работающих по проблематике компьютерного анализа текстов. Результаты диссертационной работы будут применимы при разработке информационных систем для правоохранительных органов и специальных служб, государственных органов Российской Федерации.

Все результаты диссертации, выносимые на защиту, являются новыми, получены лично И.В. Соченковым. Диссертация И.В. Соченкова является завершённой и целостной научно-исследовательской работой, выполнена на высоком научном уровне. Все утверждения подкрепляются строгими математическими доказательствами. Материал диссертации хорошо структурирован.

Автореферат в полной мере отражает содержание диссертационной работы.

Диссертационная работа полностью соответствует критериям, установленным Положением о присуждении ученых степеней: в ней содержится решение задачи, имеющей существенное значение для развития методов поиска и анализа текстовой информации. Диссертация написана автором самостоятельно, обладает внутренним единством и содержит новые научные результаты, полученные автором лично. В диссертации приведены рекомендации по использованию научных результатов. Считаю, что работа полностью соответствует требованиям ВАК РФ, предъявляемым к кандидатским диссертациям, а ее автор, Соченков И.В., заслуживает присуждения ученой степени кандидата физико-математических наук по специальности 05.13.17 — Теоретические основы информатики.

Официальный оппонент,

Доцент Национального исследовательского университета

"Высшая школа экономики",

к.т.н., доцент

«29» сентября 2014 г.



А.М. Чеповский

ПОДПИСЬ ЗАВЕРЯЮ

УПРАВЛЕНИЕ ПЕРСОНАЛА
ЗАМ. НАЧ. ОТДЕЛА
РАБОТЫ С ННП
ТИХОНОВА Е.Р.

29.09.2014

