

ЗАКЛЮЧЕНИЕ ДИССЕРТАЦИОННОГО СОВЕТА Д 002.017.02
на базе федерального государственного бюджетного учреждения науки
"Вычислительный центр им. А.А. Дородницына Российской академии наук"
ПО ДИССЕРТАЦИИ НА СОИСКАНИЕ УЧЕНОЙ СТЕПЕНИ
КАНДИДАТА НАУК

аттестационное дело № _____

решение диссертационного совета от "16" октября 2014 г. № 8

О присуждении Соченкову Илье Владимировичу, гражданину Российской Федерации, степени кандидата физико-математических наук.

Диссертация "Реляционно-ситуационные структуры данных, методы и алгоритмы решения поисково-аналитических задач" в виде рукописи по специальности 05.13.17 – "Теоретические основы информатики" принята к защите "19" июня 2014 г., протокол № 4 диссертационным советом Д 002.017.02 на базе федерального государственного бюджетного учреждения науки "Вычислительный центр им. А.А. Дородницына Российской академии наук", 119333, Москва, Вавилова 40, приказ № 192/нк от 9 апреля 2013г.; замена официального оппонента решением совета от "11" сентября 2014 г. – протокол №5.

Соискатель Соченков Илья Владимирович, 1985 г. рождения. В 2009 г. окончил очную магистратуру федерального государственного бюджетного образовательного учреждения высшего профессионального образования «Российский университет дружбы народов» по специальности "Программное обеспечение вычислительных сетей", а в 2012 г. – очную аспирантуру федерального государственного бюджетного образовательного учреждения высшего профессионального образования «Российский университет дружбы народов» по специальности 05.13.17 – теоретические основы информатики. Работает в лаборатории динамических интеллектуальных систем федерального государственного бюджетного учреждения науки "Институт системного анализа Российской академии наук".

Диссертация выполнена в лаборатории динамических интеллектуальных систем федерального государственного бюджетного учреждения науки "Институт системного анализа Российской академии наук".

Научный руководитель: доктор физико-математических наук Осипов Геннадий Семенович, заместитель директора по научной работе федерального государственного бюджетного учреждения науки "Институт системного анализа Российской академии наук".

Официальные оппоненты:

- 1 Аншаков Олег Михайлович, гражданин Российской Федерации, доктор физико-математических наук, профессор, профессор кафедры математики, логики и интеллектуальных систем в гуманитарной сфере отделения интеллектуальных систем в гуманитарной сфере федерального государственного бюджетного учреждения высшего профессионального образования «Российский государственный гуманитарный университет»,
 - 2 Чеповский Андрей Михайлович, гражданин Российской Федерации, кандидат технических наук, доцент, доцент кафедры анализа данных и искусственного интеллекта федерального государственного автономного образовательного учреждения высшего профессионального образования «Национальный исследовательский университет "Высшая школа экономики"»,
- дали положительные отзывы о диссертации.

Ведущая организация: федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Национальный исследовательский университет "МЭИ"», г. Москва, в своём положительном заключении, подписанном доктором технических наук, профессором Еремеевым Александром Павловичем, заведующим кафедрой прикладной математики НИУ "МЭИ", указала, что диссертация И.В. Соченкова соответствует критериям, установленным Положением о присуждении ученых степеней: в ней содержится решение задачи, имеющей существенное значение для развития методов информационного поиска, она написана автором самостоятельно, обладает внутренним единством и содержит новые научные результаты, полученные автором лично, в ней содержатся рекомендации по использованию научных результатов, и она полностью соответствует требованиям ВАК РФ, предъявляемым к диссертациям на соискание ученой степени кандидата физико-математических наук по специальности 05.13.17 — «Теоретические основы информатики», а ее автор, И.В. Соченков, заслуживает

присуждения ему ученой степени кандидата физико-математических наук по этой специальности.

Соискатель имеет 40 опубликованных работ, из них по теме диссертации опубликовано 18 научных работ общим объёмом 14 печатных листов, в том числе 6 статей в научных журналах и изданиях, которые включены в перечень российских рецензируемых научных журналов и изданий для опубликования основных научных результатов диссертаций. Соискателю выданы 2 патента РФ на изобретения, 1 патент РФ на полезную модель и 3 свидетельства о государственной регистрации программ для ЭВМ.

Наиболее значимые научные работы по теме диссертации:

1. И.В. Соченков. Метод сравнения текстов для решения поисково-аналитических задач // Искусственный интеллект и принятие решений. №2, 2013. С.95-106.
2. И. В. Соченков, Р. Е. Суворов. Сервисы полнотекстового поиска в информационно-аналитической системе (Часть 1) // Информационные технологии и вычислительные системы. №2, 2013. С. 69-78.
3. И. В. Соченков, Р. Е. Суворов. Сервисы полнотекстового поиска в информационно-аналитической системе (Часть 2) // Информационные технологии и вычислительные системы. №3, 2013. С. 71-87.
4. Р.Е. Суворов, И.В. Соченков. Определение связанности научно-технических документов на основе характеристики тематической значимости. // Искусственный интеллект и принятие решений. №1, 2013. С.33-40.
5. Осипов Г.С., Тихомиров И.А., Соченков И.В., Смирнов И.В. Способ и система семантического поиска электронных документов. / Патент РФ на изобретение №2473119, дата отсчета срока действия патента: 05.08.2011

На диссертацию и автореферат поступили отзывы: д.т.н. профессора Кузнецова Олега Петровича, д.ф.-м.н., профессора Соловьёва Сергея Юрьевича, к.ф.-м.н., доцента Юмагужиной Валерии Николаевны. Авторы отзывов отмечают актуальность темы исследования, высоко оценивают важность полученных научно-технических результатов и их новизну. В отзывах ведущей организации, отзывах официальных оппонентов и отзывах на автореферат содержатся следующие основные замечания.

1. А.М. Чеповский отмечает, что из текста диссертации неясно, что происходит с последовательностями, которые не являются словами текста естественного языка, производится ли их отбрасывание или информация об их вхождении также сохраняется в индексе.
2. А.М. Чеповский обращает внимание, что в диссертации не описано, как организовано хранение списков, если оперативной памяти недостаточно.
3. Ведущая организация отмечает, что в обзоре литературы не упоминаются базы данных на основе графовых моделей как один из возможных способов хранения и обработки информации в виде сетевых структур и деревьев.
4. По замечанию ведущей организации, в диссертации не обоснован выбор именно линейной комбинации значений критериев в качестве результирующей оценки сходства текстов. Кроме того, в сериях экспериментов при оценке метрик качества результатов поиска использованы только русскоязычные коллекции текстов.
5. С.Ю. Соловьёв отмечает, что в работе И.В. Соченкова отсутствует сравнение эффективности предложенных автором структур данных инвертированного индекса с реляционными и графовыми базами данных по параметрам временной производительности и объёмов памяти, требуемой для хранения данных.
6. В.Н. Юмагузина указывает, что в автореферате не сформулированы явно ограничения, которые разработанные структуры данных накладывают на хранимые результаты лингвистического разбора.

Авторы отзывов отмечают, что замечания не носят принципиального характера, являются рекомендательными и не снижают высокой оценки проделанной соискателем работы. Авторы отзывов указывают, что диссертация И.В. Соченкова выполнена на высоком уровне, является завершённой научно-квалификационной работой и удовлетворяет требованиям ВАК РФ, а И.В. Соченков заслуживает присуждения ему учёной степени кандидата физико-математических наук по специальности 05.13.17 – «Теоретические основы информатики».

Выбор официальных оппонентов обоснован тем, что они являются известными специалистами и имеют работы, относящиеся к теме диссертации и

опубликованные в ведущих научных изданиях. Выбор ведущей организации обоснован тем, что в НИУ "МЭИ" ведутся работы по теме диссертации И.В. Соченкова, результаты этих работ известны научному сообществу и опубликованы в ведущих научных изданиях.

Диссертационный совет отмечает, что на основании выполненных соискателем исследований:

разработано новое решение задачи оценки сходства текстов, имеющее важное значение для развития методов информационного поиска с использованием результатов лексико-морфологического, синтаксического и семантического анализа текстов, направленное на повышение полноты и точности работы этих методов,

предложен новый подход к решению задачи хранения и обработки результатов лексико-морфологического, синтаксического и семантического анализа текстов при решении задач информационного поиска, в частности, предложена модель данных и исследованы свойства разработанных соискателем новых структур данных и алгоритмов решения поисково-аналитических задач,

доказана практическая применимость разработанных соискателем структур данных инвертированного поискового индекса, а также новых алгоритмов индексирования и поиска информации, обеспечивающих повышение полноты и точности решения задач информационного поиска,

введены новые понятия для описания и исследования свойств структур данных и алгоритмов информационного поиска в рамках предложенной соискателем модели данных.

Теоретическая значимость исследования обоснована тем, что:

доказаны утверждения о свойствах разработанных алгоритмов и оценках их вычислительной сложности, а также утверждения о свойствах операций, определенных в предложенной соискателем модели данных,

применительно к проблематике диссертации результативно (эффективно, то есть с получением обладающих новизной результатов) использован комплекс методов, включающий методы теории множеств и алгебры логики, методы объектно-ориентированного проектирования программного обеспечения, методы исследования качества результатов информационного поиска.

изложены новые синтаксические и семантические критерии оценки сходства текстов, а также новый метод многокритериальной оценки сходства текстов, **раскрыты** особенности применения разработанных соискателем новых алгоритмов и структур данных для решения задач информационного поиска, в частности, фразового, семантического, вопросно-ответного поиска информации в масштабных коллекциях текстовых документов,

изучены свойства операций, введённых соискателем в новой модели данных для описания и исследования свойств структур данных и алгоритмов, предназначенных для хранения и обработки результатов лексико-морфологического, синтаксического и семантического анализа текстов при решении задач поиска и анализа информации,

проведена модернизация существующих методов информационного поиска с применением предложенного соискателем метода оценки сходства текстов, направленная на повышение полноты и точности результатов поиска, а также расширение класса решаемых информационно-аналитических задач, в частности, семантического аннотирования и сопоставления текстов.

Значение полученных соискателем результатов исследования для практики подтверждается тем, что:

разработаны и внедрены в существующие информационно-поисковые и информационно аналитические системы (электронная библиотека международных клинических руководств в Медицинском центре Банка России, информационно-поисковые сервисы портала «Руконт» – национальный цифровой ресурс, информационно-аналитическая система Exactus Expert и поисковая машина Exactus) методы поиска и анализа текстовой информации, базирующиеся на предложенном соискателем методе многокритериальной оценки сходства текстов, алгоритмы и структуры данных информационного поиска;

определен эффект от использования результатов лингвистического анализа текстов, заключающийся в повышении качества решения задач информационного поиска,

создан комплекс программных средств для решения информационно-аналитических задач на основе разработанных соискателем алгоритмов и структур данных,

представлены предложения по дальнейшему использованию результатов исследования и перспективным направлениям будущих разработок по теме диссертации.

Результаты исследований по теме диссертационной работы использованы при выполнении научно-исследовательских работ по 10 проектам Минобрнауки РФ, программам ОНИТ РАН и грантам РФФИ.

Оценка достоверности результатов исследования выявила:

результаты экспериментальной проверки разработанных соискателем методов текстового поиска информации достоверны, они получены с применением стандартных методик оценки качества результатов информационного поиска и стандартных тестовых массивов,

теория построена с привлечением принципов логико-алгебраического моделирования и теории множеств, опирается на известные, проверяемые данные и факты и согласуется с опубликованными экспериментальными данными по теме диссертации,

идея базируется на анализе практики и современных мировых тенденций в области методов информационного поиска и компьютерного анализа текстов,

использованы актуальные данные о развитии и применении методов компьютерного анализа текстов при решении задач информационного поиска,

установлено качественное совпадение, выраженное в повышении полноты и точности решения задач информационного поиска с применением результатов лингвистического анализа текстов, с аналогичными тенденциями, представленными в работах ведущих российских и зарубежных исследователей в области исследования,

использованы современные методики сбора и обработки исходной информации.

Личный вклад соискателя состоит в следующем. Соискателем лично разработаны: новые **синтаксические и семантические критерии оценки сходства текстов**, **новый метод многокритериальной оценки сходства текстов**, новая **модель данных** для исследования свойств структур данных и алгоритмов, предназначенных для хранения и обработки лексико-морфологической, синтаксической и семантической информации текстов при решении задач поиска и анализа информации, новые **структуры данных инвертированного**

поискового индекса, обеспечивающие эффективное хранение и обработку лексико-морфологической, синтаксической и семантической информации текстов, а также новые **алгоритмы индексирования и поиска информации** с учётом морфологической, синтаксической и семантической информации текстов. Соискателем выполнена **программная реализация** разработанных структур данных и алгоритмов, а также получены **результаты экспериментального исследования**, демонстрирующие, что разработанные алгоритмы и структуры данных обладают большей вычислительной эффективностью и обеспечивают более качественное решение задач информационного поиска, нежели известные методы. Результаты, относящиеся к теме диссертации, получены автором лично и подготовлены им к публикации самостоятельно.

Диссертация охватывает основные вопросы поставленной научной задачи (проблемы) и соответствует критерию внутреннего единства, что подтверждается наличием последовательного плана исследования, непротиворечивой методологической платформы, основной идейной линии, концептуальности и взаимосвязи выводов, полученных с опорой на доказанные положения.

На заседании "16" октября 2014 г. диссертационный совет принял решение присудить Соченкову Илье Владимировичу ученую степень кандидата физико-математических наук. При проведении тайного голосования диссертационный совет в количестве 27 человек, из них 7 докторов наук, участвовавших в заседании, из 35 человек, входящих в состав совета, проголосовали: за присуждение учёной степени 27, против присуждения учёной степени 0, недействительных бюллетеней 0.

Председатель диссертационного совета

академик РАН

Ученый секретарь диссертационного совета

д.ф.-м.н., профессор

"16" октября 2014 г.



Журавлёв Ю.И.

Рязанов В.В.