

ЗАКЛЮЧЕНИЕ ДИССЕРТАЦИОННОГО СОВЕТА Д 002.073.01
на базе Федерального государственного учреждения
«Федеральный исследовательский центр «Информатика и управление»
Российской академии наук»
ПО ДИССЕРТАЦИИ НА СОИСКАНИЕ УЧЕНОЙ СТЕПЕНИ
КАНДИДАТА НАУК

аттестационное дело № _____

решение диссертационного совета от 7 октября 2015 года № 10

О присуждении Шелманову Артему Олеговичу, гражданину Российской Федерации, ученой степени кандидата технических наук.

Диссертация «Исследование методов автоматического анализа текстов и разработка интегрированной системы семантико-синтаксического анализа» по специальности 05.13.17 – Теоретические основы информатики принята к защите «29» июля 2015 года, протокол № 4 диссертационным советом Д 002.073.01 на базе Федерального государственного учреждения «Федеральный исследовательский центр «Информатика и управление» Российской академии наук», 119333, г. Москва, ул. Вавилова, д.44, кор.2, приказ № 714/нк от «2» ноября 2012 г.

Соискатель Шелманов Артем Олегович, 1989 года рождения, в 2011 году окончил Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Национальный исследовательский ядерный университет «МИФИ» по специальности «Прикладная математика и информатика», а в 2014 г. – очную аспирантуру Федерального государственного бюджетного учреждения науки «Институт системного анализа Российской академии наук» по специальности 05.13.17 – Теоретические основы информатики, работает младшим научным сотрудником в Федеральном государственном учреждении «Федеральный исследовательский центр «Информатика и управление» Российской академии наук». Диссертация выполнена в Федеральном государственном

бюджетном учреждении науки «Институт системного анализа Российской академии наук».

Научный руководитель – доктор физико-математических наук, профессор Осипов Геннадий Семенович, заместитель директора Федерального государственного учреждения «Федеральный исследовательский центр «Информатика и управление» Российской академии наук».

Официальные оппоненты:

Лахути Делир Гасемович, доктор технических наук, старший научный сотрудник, руководитель учебно-научного центра программного и лингвистического обеспечения интеллектуальных систем Российского государственного гуманитарного университета (РГГУ),

Лукашевич Наталья Валентиновна, кандидат физико-математических наук, ведущий научный сотрудник Научно-исследовательского вычислительного центра Московского государственного университета имени М.В. Ломоносова (НИВЦ МГУ), – дали положительные отзывы на диссертацию.

Ведущая организация Федеральное государственное бюджетное учреждение науки Институт проблем управления им. В.А. Трапезникова Российской академии наук (ИПУ РАН) (г. Москва) в своем положительном заключении, подписанном Кузнецовым Олегом Петровичем, доктором технических наук, профессором, заведующим лабораторией №11 «Методов интеллектуализации дискретных процессов и систем управления», указала, что диссертация А.О. Шелманова соответствует критериям, установленным Положением о присуждении ученых степеней: в ней содержится решение задачи, имеющей существенное значение для теоретических основ информатики, 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

технических наук по специальности 05.13.17 — Теоретические основы информатики, а ее автор, А.О. Шелманов, заслуживает присуждения ему ученой степени кандидата технических наук по этой специальности.

Соискатель имеет 15 опубликованных работ, в том числе по теме диссертации – 7 работ общим объемом 6,3 печатных листа: 4 из них – статьи в рецензируемых научных изданиях, 2 публикации в трудах российских и международных конференций, 1 зарегистрированная программа для ЭВМ (в соавторстве). Опубликованные в этих работах результаты, относящиеся к методам семантического и семантико-синтаксического анализа, к методам вопросно-ответного поиска и методам извлечения определений и авторских терминов из текстов научных публикаций, получены лично соискателем. Вклад соискателя в указанные работы составляет 5 печатных листов.

Наиболее значительные работы:

1. Смирнов И.В., Шелманов А.О. Семантико-синтаксический анализ естественных языков. Часть I. Обзор методов синтаксического и семантического анализа текстов. // Искусственный интеллект и принятие решений. — М: ИСА РАН. — 2013. — №1. — С. 41–54.
2. Семантико-синтаксический анализ естественных языков Часть II. Метод семантико-синтаксического анализа текстов / И.В. Смирнов, А.О. Шелманов, Е.С. Кузнецова, И.В. Храмоин // Искусственный интеллект и принятие решений. — М: ИСА РАН. — 2014. — № 1. — С. 11–24.
3. Shelmanov A.O., Smirnov I.V. Methods for Semantic Role Labeling of Russian Texts // Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference "Dialogue" (2014). — Issue 13 (20). — 2014. — P. 607–619.
4. Осипов Г.С., Шелманов А.О. Метод повышения качества синтаксического анализа на основе взаимодействия синтаксических и семантических правил // Труды шестой международной конференции

«Системный анализ и информационные технологии» (САИТ). — 2015. — Т. 1. — С. 229–240.

На диссертацию и автореферат поступили положительные отзывы от д.т.н., профессора Стефанюка Вадима Львовича, к.ф.-м.н., профессора Толмачева Игоря Леонидовича, к.т.н., доцента Зеленкова Павла Викторовича. Авторы отзывов отмечают актуальность темы исследования, высоко оценивают важность полученных научных результатов и их новизну. В отзывах ведущей организации, официальных оппонентов и в отзывах об автореферате содержатся следующие критические замечания:

1. Д.Г. Лахути отмечает, что к числу недостатков рецензируемой работы можно отнести разве что нечеткость или спорность двух формулировок. 1) Так, на с. 5 диссертации говорится: «Подчинение обуславливается набором общих принципов, которые в целом сводятся к тому, что зависимое слово в предложении является уточняющим, необязательным, менее важным для передачи смысла высказывания, чем главное». С этим вряд ли можно безусловно согласиться. Достаточно типичным является случай, когда главное слово является достаточно общим, в какой-то мере «пустым»; например, в словосочетаниях «синтаксический анализ» или «семантический анализ» главное слово «анализ» является чрезвычайно общим (особенно в случае научного текста), и только зависимые слова «синтаксический» и «семантический» придают ему определенное значение, в отличие, например, от «химического» или «ситуационного» анализа. 2) Неудачным представляется определение синтаксического анализа на с. 15 как «сопоставление лексем некоторого языка (естественного или формального) с его формальной грамматикой»: лексемы и грамматика – объекты разной природы, они несопоставимы.
2. Н.В. Лукашевич отмечает, что: 1) тестирование подхода выполнено на неопубликованных данных (семантическая разметка), что

затрудняет проведение сравнительного анализа в будущих работах других исследователей; 2) на странице 87 делается ссылка на «сильные» правила, описанные в разделе 2.2.3, но в самом разделе 2.2.3 перечисляемые правила не названы «сильными», что усложняет понимание текста; 3) обнаружен ряд опечаток (стр. 88, 89).

3. Ведущая организация отмечает, что: 1) при описании алгоритмов семантического и семантико-синтаксического анализа следовало воспользоваться теоретико-множественными методами, что позволило бы представить алгоритмы более формально и упростить их восприятие. Следовало внести отдельную нумерацию для алгоритмов; 2) в формуле (6) на странице 83 допущена опечатка, изменяющая смысл формулы: вместо символа \notin используется символ \in ; 3) в задаче распределения ролей (гл.2, п.2.2.5) вместо термина «функция стоимости» при постановке задачи оптимизации следовало использовать другой термин, например, «функция веса», поскольку в диссертации ставится задача не минимизации, а максимизации совокупного веса варианта распределения ролей; 4) в диссертации недостаточно описаны разработанные программные средства. Следовало более подробно описать архитектуру, входные-выходные данные, модули и компоненты разработанных систем.
4. П.В. Зеленков делает следующие замечания по автореферату: 1) на страницах 11–12 недостаточно ясно сформулировано различие между «центральными» и «периферийными» семантическими ролями; 2) на странице 15 используется термин «проективность связей», однако определение этого термина в тексте автореферата не приведено; 3) на странице 15 упоминается статистико-эвристический критерий, но в тексте автореферата он не сформулирован.
5. И.Л. Толмачев указывает на то, что в автореферате недостаточно подробно описан разработанный автором метод ранжирования для

вопросно-ответного поиска, учитывающий лексику, семантические роли и отношения семантической сети.

6. В.Л. Стефанюк делает следующие замечания по автореферату: 1) в постановке оптимизационной задачи о распределении ролей по аргументам приведена только целевая функция и сказано, что решается задача о назначениях. При этом отсутствует полная математическая постановка задачи, что затрудняет восприятие; 2) в тексте встречается термин «проективность связей», однако отсутствует его пояснение; 3) в автореферате следовало бы привести формулы метрик оценки качества для вопросно-ответного поиска.

Авторы отзывов указывают на то, что замечания не носят принципиального характера и не снижают высокой оценки проделанной соискателем работы. Авторы отзывов отмечают, что диссертация А.О Шелманова выполнена на высоком научном уровне, является завершённой научно-квалификационной работой и удовлетворяет требованиям ВАК РФ, а А.О. Шелманов заслуживает присуждения ему учёной степени кандидата технических наук по специальности 05.13.17 – Теоретические основы информатики.

Выбор официальных оппонентов и выбор ведущей организации обосновывается их компетенцией и достижениями в соответствующей отрасли науки, наличием публикаций по теме диссертации А.О. Шелманова.

Диссертационный совет отмечает, что на основании выполненных соискателем исследований:

- **предложена** модификация метода автоматического определения ролевых структур высказываний, основанного на коммуникативной грамматике русского языка, которая существенно повышает точность и полноту семантического анализа;
- **разработан** новый метод компьютерного семантико-синтаксического анализа текстов, в котором интегрированы

методы построения синтаксических деревьев зависимостей и определения ролевых структур высказываний, позволяющий повысить полноту синтаксического и семантического видов анализа по сравнению с реализацией, в которой эти виды анализа выполняются отдельно;

- **разработана и реализована** интегрированная система семантико-синтаксического анализа, которая была применена для решения задач вопросно-ответного поиска, извлечения определений и авторских терминов из текстов научных публикаций.

Теоретическая значимость исследования обоснована тем, что **применительно к проблематике диссертации результативно** использован комплекс методов, включающий методы оптимизации, методы машинного обучения, методы компьютерной лингвистики, методы оценки качества алгоритмов машинного обучения, методы проверки статистической значимости полученных результатов, методы исследования качества синтаксического и семантического анализа, методы объектно-ориентированного проектирования программного обеспечения; разработан новый метод интеграции и взаимодействия синтаксического и семантического видов анализа текстов на естественном языке; **проведена модернизация** метода семантического анализа.

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

- **разработаны** методы семантического и семантико-синтаксического анализа, вопросно-ответного поиска и извлечения определений и авторских терминов из текстов научных публикаций; показаны преимущества разработанных методов в задачах вопросно-ответного поиска и построения правил для извлечения определений и авторских терминов; показаны преимущества семантико-синтаксического анализа по сравнению с последовательным выполнением синтаксического и

семантического видов анализа. Реализованные на основе разработанных методов программные средства **внедрены** в информационно-поисковые сервисы портала «Руконт», электронно-библиотечную систему «Znanium.com», информационно-аналитическую систему «Ехactus Expert» и метапоисковую машину «Ехactus»;

- **определены** области применения разработанных методов семантического и семантико-синтаксического анализа;
- на основе разработанных методов **создан** комплекс программных средств для семантического и семантико-синтаксического видов анализа текстов на естественном языке;
- **представлены рекомендации** по применению разработанных методов и алгоритмов в задачах информационного поиска и анализа текстов на естественном языке;
- результаты диссертационной работы использованы при выполнении ряда научно-исследовательских работ по проектам Минобрнауки РФ, программам ОНИТ РАН и грантам РФФИ.

Оценка достоверности результатов исследования выявила:

- **результаты экспериментальной проверки** разработанных соискателем методов семантического и семантико-синтаксического анализа, а также методов вопросно-ответного поиска и извлечения определений и авторских терминов, достоверны, они получены с применением стандартных методик оценки качества решения соответствующих задач; методы оценивались на размеченных корпусах; использовались стандартные метрики: точность, полнота, F_1 -мера, средний обратный ранг ответа; при анализе качества синтаксического анализа использовался стандартный тестовый массив – синтаксически размеченный корпус русского языка СинТагРус;

проводилась оценка статистической значимости полученных результатов;

- **теоретические положения**, относящиеся к методам и алгоритмам синтаксического, семантического и семантико-синтаксического анализа, опираются на известные, проверяемые данные и факты и согласуются с опубликованными экспериментальными данными по теме диссертации;
- **идея базируется** на анализе практики и современных мировых тенденций в области методов семантического и семантико-синтаксического анализа текстов на естественном языке;
- **использованы** актуальные данные о развитии методов синтаксического, семантического и семантико-синтаксического анализа текстов на естественном языке;
- **использованы** современные методики сбора и обработки исходной информации.

Личный вклад соискателя состоит в следующем. Соискателем лично модифицирован **метод автоматического определения ролевых структур высказываний, основанный на коммуникативной грамматике русского языка**, разработан **новый метод компьютерного семантико-синтаксического анализа текстов**, разработана и реализована **интегрированная система семантико-синтаксического анализа**, экспериментально показана эффективность разработанных методов в задачах синтаксического и семантического анализа при их тестировании на размеченных корпусах русскоязычных текстов, разработаны методы решения практических задач, использующие результаты семантического и семантико-синтаксического анализа, проведена экспериментальная оценка этих методов и показана значимость применения методов семантического и семантико-синтаксического анализа для решения практических задач обработки текстов на естественном языке.

Соискателем решена **научная задача**: разработан метод семантико-синтаксического анализа, интегрирующий построение синтаксического дерева зависимостей с определением ролевых структур высказываний.

Диссертация охватывает основные вопросы поставленной научной задачи 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 экспериментально обоснованные положения. Результаты, относящиеся к теме диссертации, получены автором лично и подготовлены им к публикации самостоятельно.

На заседании «7» октября 2015 г. диссертационный совет принял решение присудить А.О. Шелманову ученую степень кандидата технических наук. При проведении тайного голосования диссертационный совет в количестве 20 человек, из них 5 докторов наук по специальности 05.13.17 – Теоретические основы информатики, участвовавших в заседании, из 24 человек, входящих в состав совета, дополнительно введены на разовую защиту 0 человек, проголосовали: за присуждение ученой степени – 20, против присуждения ученой степени – нет, недействительных бюллетеней – нет.

Председатель

диссертационного совета

Соколов И.А.

Ученый секретарь

диссертационного совета

Гринченко С.Н.



Дата оформления Заключения « 9 » октября 2015 г.