

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 № 11

О присуждении Швецу Александру Валерьевичу, гражданину Российской Федерации, ученой степени кандидата технических наук.

**Диссертация** «Взаимодействие информационных и лингвистических методов в задачах анализа качества научных текстов» по специальности 05.13.17 – Теоретические основы информатики принята к защите 29 июля 2015 года, протокол № 5 диссертационным советом Д 002.073.01 на базе Федерального государственного учреждения «Федеральный исследовательский центр «Информатика и управление» Российской академии наук», 119333, Москва, Вавилова, д.44, кор.2, приказ № 714/нк от 02.11.2012.

**Соискатель** Швец Александр Валерьевич, 1989 года рождения. В 2011 году соискатель окончил Федеральное государственное автономное образовательное учреждение высшего профессионального образования «Сибирский федеральный университет», а в 2014 г. – очную аспирантуру Федерального государственного бюджетного учреждения науки Института 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 Институте 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:**

Сулейманов Джавдет Шевкетович, доктор технических наук, академик Академии наук Республики Татарстан, директор Научно-исследовательского института «Прикладная семиотика» Академии наук Республики Татарстан, Князева Анна Анатольевна, кандидат технических наук, младший научный сотрудник лаборатории численного моделирования и высокопроизводительных ресурсов Института вычислительных технологий СО РАН, – дали положительные отзывы на диссертацию.

**Ведущая организация** Федеральное государственное бюджетное образовательное учреждение высшего образования «Национальный исследовательский университет "МЭИ"» (Москва) в своем положительном заключении, подписанном Еремеевым Александром Павловичем, доктором технических наук, профессором, заведующим кафедрой прикладной математики НИУ «МЭИ», лауреатом Премии Президента РФ в области образования, и К.Г. Меньшиковой, кандидатом технических наук, доцентом, ученым секретарем кафедры прикладной математики НИУ «МЭИ», указала, что диссертация А.В. Швеца 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, а ее автор, А.В. Шве́ц, заслуживает присуждения ему ученой степени кандидата технических наук по специальности 05.13.17 – Теоретические основы информатики.

Соискатель имеет 26 опубликованных работ, в том числе по теме диссертации – 9 работ общим объемом 7 печатных листов, в том числе 4 статьи в рецензируемых научных изданиях, по теме работы получено 2 свидетельства о регистрации программ для ЭВМ (в соавторстве). Опубликованные в этих работах результаты, относящиеся к методам и алгоритмам выявления признаков, характеризующих качество текстов научной сферы, получены лично автором.

Наиболее значительные работы:

1. Shvets, A. A Method of Automatic Detection of Pseudoscientific Publications // Proceedings of the 7th IEEE International Conference Intelligent Systems (IS'2014 IEEE). Advances in Intelligent Systems and Computing (AISC). – Warsaw, 2015. – Vol. 2. – P. 533-539.
2. Osipov, G., Smirnov, I., Tikhomirov, I., Sochenkov, I., Shelmanov, A., and Shvets, A. Information Retrieval for R&D Support / Paltoglou, Georgios, Loizides, Fernando, Hansen, Preben (Eds.) Professional Search in the Modern World. Lecture Notes in Computer Science (LNCS). – 2014. – Vol. 8830. – P. 45-69.
3. Шве́ц А.В., Кузнецова Ю.М., Осипов Г.С., Латышев А.В. Метод и алгоритм обнаружения признаков лингвистических дефектов в научно-технических текстах // Информационные технологии и вычислительные системы. – 2013. – № 2. – С. 79-87.
4. Кузнецова Ю.М., Осипов Г.С., Чудова Н.В., Шве́ц А.В. Автоматическое установление соответствия статей требованиям к научным публикациям // Труды ИСА РАН. – 2012. – Т. 62. – Вып. 3. – С. 132-138.

На диссертацию и автореферат поступили положительные отзывы от д.т.н. Марии Анатольевны Михеенковой, д.ф.-м.н., профессора Леонида Антоновича Севастьянова, к.т.н., доцента Павла Викторовича Зеленкова. Авторы отзывов отмечают актуальность темы исследования, высоко

оценивают практическую значимость полученных научных результатов и их новизну. В отзывах ведущей организации, официальных оппонентов и в отзывах на автореферат содержатся следующие критические замечания:

1. Д.Ш. Сулейманов отмечает: 1) сформированный перечень признаков с их значениями, приведенный в таблице в разделе 3.3, логичнее было бы поместить в Главе 2 диссертации, где и устанавливаются возможные значения признаков; 2) в описаниях предлагаемых алгоритмов отсутствуют такие формальные элементы, как «начало» и «конец», однако на соответствующих блок-схемах они есть и обозначены как «вход» и «выход»; 3) не совсем удачно названа Глава 3 «Применение методов машинного обучения для выявления признаков псевдонаучных текстов», поскольку кроме методов машинного обучения используется индуктивный метод порождения гипотез; 4) в обзорной части работы автор приводит исследования, в которых структура научных публикаций рассматривается шире, чем IMRAD. Например, раздел «Методы» разбивают на подразделы «Метод», «Эксперимент», «Модель», «Наблюдение». Автору следует учесть это в дальнейшей работе, возможно, более детальный анализ структуры даст лучшие результаты при анализе качества научных текстов.
2. А.А. Князева, отмечая высокое качество проработки решаемых задач, в то же время обращает внимание автора на ряд критических замечаний по существу отдельных положений диссертационной работы: 1) во введении отсутствуют определения некоторых понятий, которые раскрываются лишь по мере появления в основной части работы (например, «псевдонаучный текст», «структура текста»); 2) в работе отсутствует пояснение, почему значения извлекаемых признаков, характеризующих качество научных текстов, являются дискретными. Лишь в третьей главе показано, что они используются для получения интерпретируемых экспертом правил. Стоило

уточнить причину дискретности раньше, в первой или второй главе, где устанавливаются значения признаков; 3) наряду с термином «снижение размерности признакового пространства» равнозначно используется термин «сокращение пространства признаков», который не является общепринятым. Следовало избежать употребления одного из терминов или, по крайней мере, пояснить, что они несут одинаковое значение; 4) в работе отсутствует анализ применимости предлагаемых инструментов в зависимости от области науки, которой принадлежат рассматриваемые публикации. В частности, неясно, подходят ли предложенные методы для гуманитарных наук; 5) в проведенном исследовании используется допущение, согласно которому научные публикации однородны по своему составу (научные состоят только из научных фрагментов, а псевдонаучные из псевдонаучных). На практике такие условия вряд ли выполняются для всех научных публикаций.

3. Ведущая организация отмечает, что полученные в работе результаты обладают не только теоретической значимостью, но и практической ценностью. Замечания по работе: 1) в работе фактически отсутствуют сравнения с ранее известными методами решения аналогичных задач. Следовало бы включить в сравнение методов классификации также методы, основанные на известных классических методах признакового описания текстов; 2) автор в недостаточной степени исследует детали предлагаемых им алгоритмов и методов. Например, автор предлагает способ вычисления численной оценки соответствия текстового фрагмента структурному разделу, но не обосновывает конкретный вид формулы, не исследует другие возможные способы вычисления указанной оценки и не проводит эмпирического сравнения этих способов; 3) не обоснован выбор типа нейронной сети при сравнении различных методов классификации; 4) из рисунка 10 следует, что статьи разных отраслей науки отличаются уровнем

использования общенаучной лексики. Однако, статьи также могут иметь отличия в структуре и степени выраженности различных нарушений. Целесообразно было бы проанализировать особенности статей разных отраслей науки и установить для них разные (наиболее соответствующие данной отрасли) критерии оценки качества текстов; 5) имеются ошибки и неточности при описании алгоритмов. Так, во всех описаниях алгоритмов формально отсутствуют критерии окончания алгоритмов. Алгоритмы 2.4 и 2.5 содержат недетерминированные шаги (шаги с недетерминированным выбором), а именно: «Шаг 1. Разделить текст  $T$  на фрагменты  $F_i$  равной длины. Шаг 2. Выбрать один из фрагментов  $F...$ » (алгоритм 2.4); «Шаг 1. Выбрать одно из правил русского языка  $r'$ .» (алгоритм 2.5). Кроме того, для алгоритма 2.4 указана сложность  $O(N)$ , где  $N$  – число семантических и синтаксических конструкций в тексте, но какова сложность выявления этих конструкций не указывается.

4. М.А. Михеенкова указывает на то, что в автореферате не уточняется какая стратегия ДСМ-метода (с запретом или без запрета на контрпримеры) используется при формировании гипотез для определения принадлежности текста множеству псевдонаучных текстов и что отсутствует описание того, какие из правил для обнаружения псевдонаучных текстов предоставляются эксперту для дальнейшего анализа. Поэтому остается неясно, работает ли эксперт со всем множеством правил, условия которых выполнены при проверке некоторого текста, или только с наиболее значимыми из них.
5. Л.А. Севастьянов отмечает, что в автореферате на странице 10 не указано, какая коллекция текстов использовалась в качестве обучающей выборки и что в описании алгоритма 2.4 не раскрыто, каким образом выполняется разделение текста на фрагменты.

6. П.В. Зеленков отмечает, что по автореферату имеются небольшие замечания и рекомендации: из описаний предложенных методов не следует, как выполняется анализ качества текстов научных работ, содержащих большое количество математических формул; в перечень основных результатов работы, приведенных в заключении автореферата, следовало включить пункт с указанием факта разработки программных средств.

Авторы отзывов отмечают, что замечания являются рекомендательными и не снижают высокой оценки проделанной соискателем работы. Авторы отзывов указывают, что диссертация А.В. Швеца является законченной научно-квалификационной работой, выполнена на высоком научном уровне и удовлетворяет требованиям ВАК РФ, а А.В. Швец заслуживает присуждения искомой учёной степени кандидата технических наук по специальности 05.13.17 – Теоретические основы информатики.

**Выбор официальных оппонентов** обосновывается 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 А.В. Швеца, результаты этих работ известны научному сообществу и опубликованы в ведущих научных изданиях.

Диссертационный совет отмечает, что на основании выполненных соискателем исследований:

- **разработан** новый метод автоматического формирования общенаучного словаря устойчивых словосочетаний;
- **разработан** новый метод автоматического выявления структуры научной публикации;
- **разработан** новый метод обнаружения нарушений ряда правил согласования, нарушений синтаксической и семантической связности, лексической избыточности, нарушений последовательности изложения;

- **разработан** метод автоматического выявления псевдонаучных фрагментов текстов научной сферы.

**Теоретическая значимость исследования** обоснована тем, что были созданы новые методы автоматического выявления признаков, характеризующих качество текстов научной сферы, на основе взаимодействия информационных и лингвистических методов; применительно к проблематике диссертации результативно (эффективно, то есть с получением обладающих новизной результатов) **использован** комплекс методов, включающий методы интеллектуального анализа текстов, статистические методы, методы машинного обучения, методы снижения размерности признакового пространства, индуктивные методы порождения гипотез, метод реляционно-ситуационного анализа текстов.

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

- результаты исследований по теме диссертационной работы **использованы** при выполнении ряда научно-исследовательских работ по проектам Министерства образования и науки РФ, программам РАН и грантам РФФИ;
- разработанные методы определения качества текстов научной сферы **реализованы** в виде программных средств и **внедрены** в информационную систему «ЭКБСОН» Государственной публичной научно-технической библиотеки, электронно-библиотечную систему «Рукопт» ООО «Национальный цифровой ресурс «Рукопт», электронно-библиотечную систему «Znanium.com» ООО «Научно-издательский центр ИНФРА-М», систему интеллектуального поиска и анализа научных публикаций «Exactus Expert» ЗАО «РосИнтернет технологии»;



- **определена** возможность использования разработанных методов и алгоритмов в системах поддержки принятия решений и в системах поиска и анализа научной информации.

**Оценка достоверности результатов исследования выявила:**

- **результаты** экспериментальной **проверки** разработанных соискателем методов достоверны, они получены с применением стандартных методик оценки качества решения соответствующих задач: использован масштабный тестовый набор данных, в том числе использована большая коллекция псевдонаучных текстов, рекомендованная для тестирования комиссией РАН по борьбе с лженаукой и фальсификацией научных исследований, при проведении экспериментов применялся метод перепроверки с многократным запуском на независимых выборках текстов для усреднения результатов, результаты продемонстрировали эффективность разработанных методов (получены высокие значения точности и полноты классификации), а также показана применимость сформированного пространства признаков к автоматическому обнаружению псевдонаучных текстов с помощью ряда различных методов классификации и целесообразность применения предложенных в работе методов.
- **теоретические положения** 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 интеллектуального анализа текстов, методов машинного обучения и методов индуктивного порождения гипотез;
- **использованы** современные методики сбора и обработки исходной информации.

**Личный вклад** соискателя состоит в следующем. Соискателем лично разработаны новые методы автоматического формирования общенаучного словаря устойчивых словосочетаний, автоматического выявления структуры научной публикации, обнаружения нарушений ряда правил согласования, нарушений синтаксической и семантической связности, лексической избыточности, нарушений последовательности изложения, автоматического выявления псевдонаучных фрагментов текстов научной сферы, выполнена программная реализация разработанных методов, экспериментально показана эффективность и значимость разработанных методов при решении практической задачи автоматического обнаружения псевдонаучных текстов. 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.В. Швецу 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



И.А. Соколов

Ученый секретарь диссертационного совета

С.Н. Гринченко

« 9 » октября 2015 года

М.П.