

## ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА

о диссертации Швеца Александра Валерьевича  
«Взаимодействие информационных и лингвистических методов в задачах  
анализа качества научных текстов», представленной на соискание ученой  
степени кандидата технических наук  
по специальности 05.13.17 – «Теоретические основы информатики»

Диссертационная работа А.В. Швеца посвящена разработке методов автоматического определения качества научных текстов. Потребность в оценке текстов вызвана регулярным появлением в открытой печати статей, содержащих различные нарушения (использование ненаучной лексики, несоблюдение структуры, отсутствие семантической связности, нарушения последовательности изложения и другие). Указанные нарушения, как правило, являются причиной низкой ясности изложения текста и отрицательно влияют на возможность корректного восприятия информации. Ввиду отсутствия эффективных методов определения признаков, характеризующих качество научных текстов, актуальность работы не вызывает сомнений.

Диссертация состоит из введения, трех глав, заключения, списка использованных источников и приложения.

Во введении автором обоснована актуальность темы, определен предмет исследования, сформулированы цель и задачи исследования, научная новизна, теоретическая и практическая значимость полученных результатов.

В первой главе А.В. Швец рассматривает основные нарушения, приводит соответствующие им примеры из реальных публикаций и исследует возможность выявления этих нарушений с использованием лексики, синтаксических и семантических структур. Автором выполнен обзор методов анализа научных текстов, отмечены ограничения к применению этих методов для определения качества текстов. Предложена схема выявления признаков текста, характеризующих его качество, и поставлены задачи исследования.

Во второй главе последовательно, для каждого типа нарушений, представленных в первой главе, А.В. Швецом рассматриваются данные, необходимые для их обнаружения, предлагаются новые методы формирования этих данных (метод формирования общенаучного словаря устойчивых словосочетаний, метод формирования правил, характеризующих лингвистические ошибки, метод извлечения маркеров структурных разделов) и приводятся разработанные методы выявления нарушений. Выполнено экспериментальное исследование предложенных методов, в ходе которого установлены возможные значения признаков качества текстов.

В третьей главе исследуется применимость разработанных методов для выявления псевдонаучных текстов. В первом параграфе автор приводит используемое в работе определение псевдонауки. Второй параграф посвящен разработке метода автоматического определения псевдонаучных фрагментов. В третьем параграфе автор формирует признаковое пространство и в четвертом параграфе выполняет построение множества правил для обнаружения псевдонаучных текстов с помощью индуктивного метода порождения гипотез. В заключительной части А.В. Швецом показано, что использование полученных признаков позволяет выполнять классификацию научных текстов с высокой точностью и полнотой. Автор делает вывод о применимости сформированного пространства признаков к определению качества текстов и целесообразности использования разработанных методов.

В заключении приводятся следующие основные результаты, полученные в диссертационном исследовании:

1. Разработан метод автоматического формирования общенаучного словаря устойчивых словосочетаний.
2. Разработан метод автоматического выявления структуры научной публикации.
3. Разработан метод обнаружения нарушений правил согласования, нарушений синтаксической и семантической связности, лексической избыточности, нарушений последовательности изложения.
4. Разработан метод автоматического выявления псевдонаучных фрагментов текстов научной сферы.
5. Сформировано множество признаков, характеризующих качество текстов научной сферы.
6. Построено множество правил для обнаружения псевдонаучных текстов.

По работе можно сделать следующие замечания:

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.

Приведенные замечания не влияют на положительную оценку работы.

Полученные в работе результаты внедрены в информационно-аналитические и электронно-библиотечные системы, и также могут быть использованы в поисково-аналитических системах для решения задач интеллектуального анализа научной информации, в научных фондах при отборе проектов и в издательствах при приеме статей к публикации.

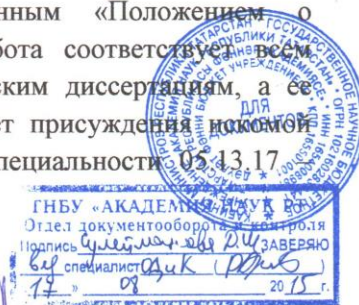
Основные научные результаты отражены в 9 работах, опубликованных автором: в 4 публикациях в рецензируемых изданиях из Перечня ВАК и приравненных к ним, в 3 публикациях в трудах международных и российских конференций и в 2 зарегистрированных программных системах.

Все результаты диссертации, выносимые на защиту, являются новыми, получены А.В. Швецом лично. Диссертация А.В. Швеца является завершённой научно-исследовательской работой, выполнена на высоком научном уровне. Все выводы подкрепляются результатами экспериментов. Материал диссертации изложен ясным языком и хорошо структурирован.

Автореферат полностью отражает содержание диссертации.

Диссертационная работа А.В. Швеца «Взаимодействие информационных и лингвистических методов в задачах анализа качества научных текстов» полностью удовлетворяет критериям, установленным «Положением о присуждении ученых степеней». Считаю, что работа соответствует всем требованиям ВАК РФ, предъявляемым к кандидатским диссертациям, а ее автор – Швец Александр Валерьевич – заслуживает присуждения ученой степени кандидата технических наук по специальности «Теоретические основы информатики».

Директор Научно-исследовательского  
института «Прикладная семиотика» АН РТ,  
д.т.н., академик АН РТ



 Д.Ш. Сулейманов  
«15» августа 2015 г.

Сулейманов Джавдет Шевкетович, НИИ «Прикладная семиотика» Академии наук Республики Татарстан, 420111, Республика Татарстан, г. Казань, ул.Левобулачная, 36а, +7 (843) 292-49-14, ips.antat@gmail.com