

Отзыв

официального оппонента о диссертации Куликова Сергея Юрьевича на тему: «Автоматическое извлечение мнений: лингвистический аспект», представленной на соискание учёной степени кандидата филологических наук по специальности: 10.02.21 – прикладная и математическая лингвистика, М., 2016

Диссертационное исследование Куликова Сергея Юрьевича посвящено разработке лингвистического обеспечения для автоматического сентимент-анализа интернет-текстов на русском языке. Данная работа, бесспорно, является **актуальной** для сегодняшнего общества в условиях гибридных войн, в сложившейся современной политической обстановке, где информационной войне отводится ведущая роль. Также необходимость проведения подобного исследования продиктована отсутствием комплексного лингвистического описания извлечения мнений из большого потока текстовой информации.

Работа логично построена и состоит из введения, трех глав, выводов по главам, заключения, списка использованной научной литературы (178 источников) и источников, четырех приложений. Традиционно Глава 1 посвящена истории и оценке современного состояния автоматического извлечения мнений. Заслугой автора можно считать логичное выделение этапов исторического развития автоматического извлечения мнений. В Главе 2 описывается структура лингвистического обеспечения в системах автоматического извлечения мнений, а также обзор современных систем анализа текста. Результаты исследования автора изложены в Главе 3.

Цели и задачи диссертационного исследования сформулированы автором следующим образом: изучить особенности существующих систем автоматического извлечения мнений; проанализировать типы оценочной информации, моделируемые в существующих системах; определить классы оценочной лексики, необходимые для повышения качества автоматического извлечения мнений; уточнить определение понятия «мнение», принятое в практике автоматического извлечения мнений; разработать методы

филтрации априорно нейтрального контента; определить фрагменты этапов автоматического извлечения мнений, зависящие от аспекта задачи; разработать принципы автоматического и автоматизированного создания оценочных ресурсов для русского языка.

Новизна исследования заключается в том, что автором была расширена классификация типов оценочной лексики, в которую вошли 12 классов. С.Ю. Куликов справедливо отмечает особую важность ограничения количества объектов-референтов у оценочных слов; а также вводит понятие однореферентные оценочные слова. Исследователь также отмечает, что для эффективной работы систем автоматического извлечения мнений необходимо учитывать особенности интернет-коммуникации, информацию об источнике мнения и канале передачи сообщения.

Использованные в работе **методы** исследования соответствуют материалу и общей направленности работы. На основе системного подхода автор в своем исследовании применил методы классификации, корпусного, контекстуального и дискурсивного анализа, метода «черного ящика», статистических методов и метода моделирования, использовались элементы компонентного анализа при исследовании терминологии изучаемой научной области.

Обоснованность научных положений, выводов и рекомендаций, сформулированных автором, их достоверность основана на хорошей научно-теоретической базе трудов отечественных и зарубежных лингвистов, а также на значительной качественной и количественной выборке материала – тексты сети интернет различных жанров (блоги, сообщения информационных агентств и пользовательские отзывы на продукты и события) на русском языке: корпус ruTenTen (15,8 млрд. словоупотреблений); корпус русскоязычных СМИ и блогов (свыше 25,5 млн. словоупотреблений) и ряд других источников (около 150 тыс. словоупотреблений).

Системное представление автором лингвистической информации, необходимой для задач автоматического извлечения мнений; развитие

понятийного аппарата рассматриваемой предметной области; четкая формулировка требований к этапам автоматического анализа текста для автоматического извлечения оценочной информации (переработка, морфологический анализ, фильтрация нерелевантной информации); определяют **теоретическую значимость** рецензируемой диссертационной работы.

Говоря о **достоинствах диссертационного исследования**, мы хотели бы выделить следующие положения диссертации:

- 1) Введено понятие «однореферентных оценочных слов» и предложено их формальное описание при помощи словаря тезаурусного типа, что значительно расширяет понимание имплицитной оценки, которой в последнее время уделяется особое внимание в работах по извлечению мнений. Также предложена методика автоматического формирования словаря-тезауруса для одного из типов однореферентных оценочных слов – ксенофобонимов и ксенофилионимов, что нашло отражение в пп. 3.3. и 3.4. Главы 3.
- 2) Разработана многоступенчатая процедура определения структурных компонентов мнения (начиная с технологии навигации html-разметке веб-документов – Xpath и заканчивая анализом синтаксической структуры предложения).
- 3) В соответствии с моделью передачи сообщения Шеннона-Уивера дополнена новыми элементами структурная схема мнения.
- 4) Предложена классификация оценочной лексики, опирающаяся на ее сочетаемостные свойства (12 классов).
- 5) Описаны различия в структуре лингвистического обеспечения, которые характерны для различных типов систем автоматического извлечения мнений.

- 6) Предложена процедура автоматизированного определения противоправного контента, опирающаяся на сложную систему фильтров для входных данных.

Завершая констатирующую часть отзыва и переходя к дискуссионной, хотелось бы еще раз отметить, что теоретические положения, вынесенные соискателем на защиту, не вызывают сомнений в их достоверности. Все задачи, поставленные соискателем в диссертационном исследовании, успешно решены.

При чтении работы С.Ю. Куликова возникает ряд вопросов дискуссионного характера, а также имеются некоторые технические замечания:

1. Говоря об интернете и текстах в нем бытующих, нельзя не принимать во внимание особенности креолизации, которые имеют важное структурное, оценочное и эмоциональное значение в текстах различных жанров. Однако автор не уделяет должного внимания всем невербальным проявлениям эмоций (графике, шрифту и т.д.). Учитывая важную роль, отводимую поликодовым элементам в текстах блогов, форумов, чатов и т.д., хотелось бы уточнить, как автор планирует обрабатывать тексты, содержащие подобные элементы?
2. Традиционно пунктуации отводится одна из центральных функций в передаче эмоционально-оценочных характеристик письменного текста. Данная проблема не находит достаточного освещения в тексте диссертации. Нам представляется важным уточнить, как автор предполагает учитывать пунктуационные особенности текста?
3. Работа содержит ряд опечаток (например, сс. 16, 27 и т.д.)
4. Представляется избыточной информация представленная в Приложениях 1, 2, 3, которые содержат результаты работы

других авторов, что представляется неуместным в работах подобного типа.

Заданные вопросы и замечания не затрагивают концептуальной сущности данной научной работы, не умоляют вышеперечисленных достоинств, но продиктованы нашим интересом к проведенному С.Ю. Крючковым исследованию.

Значительное количество анализируемых теоретических работ по проблематике диссертационного исследования свидетельствуют о широте охвата исследования, а большой объем эмпирического материала делает выводы взвешенными, обоснованными и, соответственно, достоверными. Для достижения поставленных в работе целей и решения задач выбраны адекватные методы исследования, что приводит автора работы к непротиворечивым и убедительным выводам.

Диссертация С.Ю. Куликова характеризуется продуманностью композиции, логичным и стройным изложением своих мыслей. Содержание диссертации соответствует заявленной теме, а тема, в свою очередь, соответствует заявленной специальности.

Рецензируемая диссертация является самостоятельной научно-квалификационной работой. Текст диссертации написан хорошим и ясным русским языком, логично выстроен, обладает внутренним единством.

Содержание работы и полученные результаты отражены в автореферате и 24 публикациях, в том числе 3 публикациях в ведущих рецензируемых научных журналах, входящих в перечень ВАК Министерства образования и науки РФ.

Работа прошла надежную апробацию в секторе прикладного языкознания ФГБУН «Институт языкознания Российской академии наук», в докладах автора на международных конференциях и конгрессах, а также в 24 публикациях в сборниках научных трудов, включая три статьи в рецензируемых журналах, рекомендованных ВАК РФ.

Диссертационное исследование «Автоматическое извлечение мнений:

лингвистический аспект» выполнено на высоком научно-теоретическом и методологическом уровне и соответствует требованиям п. 9 Положения о присуждении учёных степеней, а его автор, Куликов Сергей Юрьевич, заслуживает присуждения ему искомой учёной степени кандидата филологических наук по специальности 10.02.21 – прикладная и математическая лингвистика.

Официальный оппонент
кандидат филологических наук,
доцент,
доцент кафедры английского языка
ГОУ ВО МО «Государственный социально-
гуманитарный университет»

Н.А. Ахренова
23 августа 2016 г.

Проректор по научной работе
ГОУ ВО МО «ГСГУ»

С.П. Хэкало



Личная подпись Ахреновой Н.А.
Заведующий Жако В.П.
Нач. отдела кадров ГОУ ВО МО «ГСГУ»
«13» 08 2016г.

ГОУ ВО МО «Государственный
социально-гуманитарный университет»

140410, Московская область,
г. Коломна, ул. Зеленая, д.30

kafedra@ech-mgosgi.ru

+7 (496) 610 15 76