

ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА

на диссертационную работу Андреева Ильи Алексеевича

«Исследование методов и алгоритмов обработки текстовой информации социальных сетей в задачах формирования социального портрета пользователя»,
представленной на соискание ученой степени кандидата технических наук
по специальности 05.13.01 – «Системный анализ, управление и обработка информации (информационные технологии и промышленность)»

1. Актуальность темы диссертационного исследования.

В данной диссертационной работе рассматриваются задачи построения социального портрета пользователя на основе его активности в социальных сетях. Социальные сети являются одной из неотъемлемых частей жизни современного человека и представляют собой крупнейший контент-генератор текстовых данных (новости, посты, комментарии). В психологической науке изучаются психологические особенности, тип характера и темпераменты пользователей социальных сетей на основе анализа предметно-содержательных компонентов публикуемого ими текста (основная тема, употребляемые понятия, эмоциональная модальность и т.п.). При этом текстовая информация в социальных сетях является принципиально неструктурированной и требуется проведение особых исследований для построения различного рода моделей психологических и социальных портретов пользователей. Методы системного анализа отличаются междисциплинарным подходом к решению сложных проблем, что позволяет учесть особенности решаемых в диссертации задач, связанных с наличием множества негативных факторов (многозначность, динамичное развитие, неоднородность естественного языка и т.д.), существенно усложняющих обработки текстовой информации социальных сетей.

Таким образом, применение системного подхода для создания новых методов и алгоритмов психолингвистического и сентимент-анализа русскоязычных текстовых данных социальных сетей на основе системного анализа представляется важной проблемой, а тема диссертации, безусловно, актуальна.

2. Структура и характеристика диссертационной работы.

Объем представленной диссертационной работы – 166 страниц, к работе приложен список из 4 приложений. Количество наименований в списке литературы – 170 единиц. Заявленная цель исследования – снижение трудозатрат на построение социального портрета пользователя социальной сети, объект – открытые русскоязычные текстовые данные, извлекаемые из различного рода сообщений в социальных сетях.

Введение обосновывает актуальность данного исследования, описывает цель и задачи работы, определяет научную новизну и практическую значимость полученных результатов.

Первая глава состоит из нескольких частей:

- первая часть описывает современное состояние методов и существующие системы анализа социальных сетей, в том числе и по идентификации пользователей в различных социальных сетях;
- вторая часть описывает группы методов, связанных с оценкой тональности текста. Отдельно описаны методы формирования обучающих выбор в данной задаче.
- третья часть описывает существующие методы построения психологического портрета человека, в том числе при помощи анализа публичной информации о данном человеке в социальных сетях.

Вторая глава посвящена разработанным в рамках данной диссертационной работы подходам, моделям и алгоритмам – алгоритм формирования обучающей выборки для сентимент-анализа на основе словаря авторских знаков и расширенного словаря WordNetAffect, алгоритм анализа эмоциональной окраски русскоязычных текстовых данных социальных сетей, оригинальный подход к сопоставлению профилей пользователей с использованием анализа структурированных и неструктурированных данных. В главе также представлен классификатор, основанный на психологическом методе «Большой пятёрки», обучение которого происходит при помощи результатов добровольцев, прошедших опросник 5PFQ.

Третья глава состоит из описания реализованного программного обеспечения, приведена общая схема работы системы, общая концепция программного комплекса, отдельно рассмотрена каждая из его частей.

Четвертая глава состоит из экспериментов, доказывающих эффективность разработанных алгоритмов и методов, а также содержит информацию о практическом применении разработанного ПО в различных организациях.

В заключении кратко представлены результаты диссертационного исследования.

Приложения содержат акты внедрения, свидетельства о регистрации программ для ЭВМ и расширенную блок-схему процесса подбора персонала.

3. Теоретическая значимость, научная новизна и практическая ценность результатов диссертации.

Андреевым И.А. в научную новизну вынесены следующие результаты исследования:

1) подход к сопоставлению профилей пользователей из различных социальных сетей, который отличается одновременным применением анализа графической информации, данных анкет и социальных графов профилей.

2) алгоритм формирования обучающей выборки, применимой в задачах сентимент-анализа текстовых данных. Особенностью данного алгоритма является использование словарей авторских символов выражения эмоций (так называемых «эмодзи») наряду с использованием ключевых фраз.

3) метод определения психологических характеристик пользователя социальных сетей, отличающийся использованием метода «Большой пятерки» вместе с методами машинного обучения и алгоритмами обработки естественного языка.

4) алгоритм анализа эмоциональной окраски русскоязычных текстовых данных социальных сетей, отличающийся интеграцией семантических подходов и методов машинного обучения.

Теоретическая значимость диссертационной работы заключается в эффективном применении методов системного анализа для построения новых алгоритмов обработки текстовой информации социальных сетей. Практическая значимость диссертационной работы заключается в разработке программного комплекса, который реализует эти методы и алгоритмы, и упрощающего составление социального портрета человека при помощи анализа профилей его социальных сетей.

4. Обоснованность научных положений, выводов, рекомендаций, их достоверность.

Все полученные в ходе исследования положения, выводы и рекомендации являются обоснованными и базируются на следующих методах системного анализа: разработка модели системы, декомпозиция структуры, структурный анализ, сравнительный анализ. Достоверность результатов диссертации обеспечивается корректным использованием методов исследований, подтверждены результатами экспериментов и внедрениями в различные организации, а также апробациями на международных и всероссийских конференциях.

5. Соответствие содержания автореферата положениям диссертации, оценка языка и стиля изложения диссертации и автореферата.

Диссертация и автореферат оформлены достаточно аккуратно. Присутствуют грамматические ошибки, но их количество минимально. Язык изложения научный, но понятный. Автореферат отражает содержание диссертации в полной мере, написан так же,

как и диссертация, понятным языком, и оформлен в соответствии с правилами оформления рефератов диссертаций.

6. Апробация работы, публикации по теме исследования.

Результаты работы были опубликованы в 32х статьях, 4 из которых в журналах из перечня ВАК, 11 статей в изданиях, индексируемых в Scopus и/или Web Of Science, а также 1 монография. Результаты также закреплены в 3 свидетельствах о государственной регистрации программ для ЭВМ. Диссертационная работа была обсуждена не менее, чем на 17 научных конференциях всероссийского и международного масштаба.

7. Замечания официального оппонента по диссертации.

- 1) В диссертации не указано, как формировалась тестовая выборка профилей социальных сетей в задаче психолингвистического анализа этих профилей и насколько эта выборка достоверна. При этом сам автор признаёт, что выборка с большой вероятностью является нерепрезентативной, поскольку добровольцы имеют особый психологический портрет.
- 2) В диссертации описывается метод определения эмоциональной окраски предложения, предлагающий некую среднюю оценку эмоциональной окраски. Нигде не учитывается тот факт, что внутри одного предложения, в частности сложносочиненного, могут встречаться различные и даже противоположные эмоции по отношению к разным объектам.
- 3) Критерий схожести профилей, описанный в пункте 2.3.1 и критерий схожести лиц, описанный в пункте 2.3.2 приведены в работе слишком кратко и поверхностно, без математической формулировки. В представленном объёме работы учёт этих критерием не представляется существенным и целесообразным.
- 4) В пунктах 2.2 и 2.4 написано, в частности в описании алгоритма, что для исправления грамматических ошибок использована библиотека DeepPavlov. Однако в тексте диссертации не встречается описания ее работы и функциональности.
- 5) В работе был применен словарь ключевых фраз, построенный на базе тезауруса WordNet-Affect, который, наряду со словарем авторских символов выражения эмоций был использован при генерации обучающей выборки, однако размер данного словаря нигде не указан.
- 6) В пункте 3.2.1 встречаются участки текста, структурно являющиеся таблицами, которые состоят из одной строки. Эти таблицы не указаны как таблицы, не пронумерованы и не учитываются в общее количество таблиц диссертации.

- 7) Оценка качества разработанных алгоритмов проводится на основе сравнения точности классификации, при этом в тексте диссертации отсутствует формальное математическое определение используемого понятия. При этом в большей части диссертации точность приводится в абсолютных значениях, а в таблице 4.6 и в заключении точность приводится в процентах.
- 8) Основные положения, выносимые на защиту, содержат абстрактные формулировки полученных результатов («позволяет классифицировать», «повышает точность» и т.п.) и только в выводах к главам автор приводит конкретные значения улучшаемых показателей.
- 9) Создаётся ощущение, что задача объединения профилей пользователей в различных социальных сетях, задача сентимент-анализа текстовых данных и задача психолингвистического анализа текста профилей социальных сетей являются тремя различными задачами с различным математическим аппаратом и различными исходными данными.
- 10) В диссертации на отдельных страницах встречаются некоторые ошибки в оформлении:
 - частично нумерация разделов, таблиц и рисунков оформлена не по ГОСТу.
 - небрежность форматирования приводит к наличию малозаполненных страниц (например, на пустом месте страницы 117 следовало разместить рисунок 4.1, расположенный на следующей странице).
 - либо таблицы начинаются внизу страницы (например, таблица 1.1. на стр. 42), вследствие чего появляются разрывы и переносы таблицы на следующую страницу, что снижает удобочитаемость работы.

8. Формулировка научной задачи, решенной в диссертации и вывод о ее соответствии Положению о присуждении ученых степеней.

Содержание диссертационной работы соответствует паспорту специальности 05.13.01 – «Системный анализ, управление и обработка информации (информационные технологии и промышленность)», а именно пунктам п. 4 – разработка методов и алгоритмов решения задач системного анализа, оптимизации, управления, принятия решений и обработки информации и п. 10 – методы и алгоритмы интеллектуальной поддержки при принятии управленческих решений в технических, экономических, биологических, медицинских и социальных системах.

Диссертация Андреева Ильи Алексеевича на тему «Исследование методов и алгоритмов обработки текстовой информации социальных сетей в задачах формирования социального портрета пользователя» является самостоятельно выполненной, завершенной

научно-квалификационной работой. Цель диссертация была достигнута, новые научные результаты имеют существенное научное и практическое значение.

Учитывая глубину проработки материала, научную и практическую ценность работы считаю, что диссертационная работа удовлетворяет критериям Положения о присуждении ученых степеней, предъявляемым к диссертациям на соискание ученой степени кандидата технических наук, содержит решение задачи построения модели и алгоритмов психолингвистического и сентимент-анализа русскоязычных текстовых данных социальных сетей, имеющей существенное значение для развития современных технологий обработки текстовой информации.

Автор, Андреев Илья Алексеевич, заслуживает присуждения степени кандидата технических наук по специальности 05.13.01 – «Системный анализ, управление и обработка информации (информационные технологии и промышленность)».

4 августа 2022 г.



А.В. Куприянов

Сведения об официальном оппоненте:

Доктор технических наук по специальности 05.13.18 «Математическое моделирование, численные методы и комплексы программ». Заведующий кафедрой технической кибернетики, исполнительный директор института информатики и кибернетики «Самарский национальный исследовательский университет имени академика С.П. Королева».

Федеральное государственное автономное образовательное учреждение высшего образования «Самарский национальный исследовательский университет имени академика С.П. Королева, 443086, Приволжский федеральный округ, Самарская область, г. Самара, Московское шоссе, д. 34, тел.: + 7 (846) 335-18-26, факс: + 7 (846) 335-18-36, web-сайт: <https://ssau.ru/>, e-mail: akupr@ssau.ru

данные верны, подпись Куприянова Александра Викторовича заверяю

