

ОТЗЫВ НАУЧНОГО РУКОВОДИТЕЛЯ

на диссертационную работу Андреева Ильи Алексеевича

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

Актуальность исследования

В настоящее время социальные сети являются огромным источником важных при решении критических задач знаний, которые требуют значительных усилий по фильтрации, предобработки и оригинальных подходов по их анализу вследствие слабой структуризации, а также жанровых особенностей представления.

Социальные сети можно исследовать с разными целями: можно наблюдать общественные мнения по тем или иным событиям, выявлять общественно опасных элементов и проводить иные мероприятия обеспечения безопасности населения. Особенно часто социальные сети в качестве источника знаний используются в маркетинговой деятельности и в подборе персонала организаций.

В связи с этим работа соискателя Андреева И.А., целевым назначением которой является снижение трудозатрат на построение социального портрета пользователей социальных сетей посредством автоматизации и учета дополнительных факторов в процессе анализа открытых русскоязычных текстовых данных, **является, несомненно, актуальной.**

Научная новизна результатов

Проведенный автором анализ особенностей извлечения, хранения, унификации и обработки открытых русскоязычных текстов социальных сетей определил цель и задачи работы.

Результаты исследований, полученные автором путем применения методов онтологического инжиниринга, методов обработки естественного языка, нейросетевых методов, методов машинного обучения (Machine Learning), а также

методов теории анализа социальных сетей (Social Network Analysis, SNA) **обладают научной новизной**. Автором проведены вычислительные эксперименты, подтверждающие достоверность выдвигаемых положений.

Несомненной научной новизной обладает полученные автором результаты:

1. Разработанный алгоритм формирования обучающей выборки для обучения моделей классификации в задачах сентимент-анализа текстовых данных, отличающийся совместным использованием словарей авторских символов выражения эмоций и ключевых фраз.

2. Предложенный подход к сопоставлению профилей пользователей в разных социальных сетях, отличающийся гибридизацией подходов анализа графической информации, структурированных данных анкет, текстовых данных, а также социальных графов профилей.

3. Разработанный метод определения психологических характеристик пользователя социальных сетей, отличающийся гибридизацией алгоритмов обработки естественного языка текстовых данных, машинного обучения и метода «Большой пятерки».

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

Самостоятельность получения результатов

Все перечисленные ниже научные результаты **получены соискателем самостоятельно**, в частности:

1. проведено сравнение современных интеллектуальных методов анализа текстовых данных в рамках психолингвистического и сентимент-анализа данных текстовых сообщений в социальной сети;

2. разработан алгоритм формирования обучающей выборки, состоящей из открытых русскоязычных текстовых ресурсов социальных сетей, классифицированных по 7-ми классам эмоций;

3. разработан метод классификации текстовых постов социальной сети по

классам тональности на основе семантических подходов и машинного обучения;

4. разработан подход к сопоставлению профилей пользователей в разных социальных сетях посредством анализа структурированных и неструктурированных данных анкет, а также социальных графов профилей;

5. разработан подход к определению психологических характеристик пользователя социальных сетей посредством анализа текстовых сообщений в социальных сетях;

6. разработан программный комплекс психолингвистического и сентимент-анализа открытых текстовых русскоязычных данных профилей пользователей социальных сетей;

7. проведены вычислительные эксперименты, основными результатами которых являются:

- наиболее эффективным алгоритмом сентимент анализа русскоязычных текстовых данных социальных сетей стал подход, включающий в качестве классификатора многослойный персептрон, в качестве языковой модели – модель BERT, а также предполагающий в качестве алгоритма формирования обучающей выборки – алгоритм, использующий авторские символы выражения эмоций и расширенный словарь WordNet-Affetct (87% точности);

- наилучшие результаты по классификации пользователей социальных сетей по психологическим характеристикам были получены при использовании в качестве классификатора алгоритм SVM и разбиении обучающей и тестовой выборки соотношением 70/30 (от 0,58 до 0,93 для различных показателей Big5).

- использование подхода к построению психологического портрета пользователя социальных сетей, позволило сократить трудозатраты на поиск волонтеров, обеспечивающих сопровождение лиц с ПОДА, в рамках проекта «Парабадминтон: все силы - для победы» для УОСОО «Федерация бадминтона» на 14 часов.

8. результаты исследований внедрены в практику процесса подбора персонала организаций региона.

Практическая ценность результатов

Практическая ценность работы состоит, прежде всего, в том, что разработан программный комплекс интеллектуального анализа данных открытых русскоязычных текстовых данных социальных сетей, реализующий предложенные модели и алгоритмы и архитектурно состоящий из двух программных модулей:

- модуль sentiment-анализа;
- модуль психолингвистического анализа текстовых данных социальных сетей.

Практическая значимость работы подтверждается использованием ее результатов и рекомендаций в рамках проекта «Интеллектуальная платформа формирования социального портрета соискателя на основании семантико-когнитивного анализа профилей в социальных сетях» компании ООО «Центр программной инженерии и аналитики «ФаззиЛаб», поддержанного Фондом содействия инноваций.

Помимо этого, разработанный программный комплекс был применен общественной организацией «Федерация бадминтона» в рамках проекта «Парабадминтон: все силы - для победы», поддержанного Фондом Президентских грантов, с целью поиска волонтеров для сопровождения спортсменов с ПОДА.

Также результаты диссертационного исследования были использованы в проектной деятельности в рамках проекта «Система интеллектуального поиска и анализа в Интернет-СМИ и социальных сетях», реализуемого совместно с Федеральным научно-производственным центром АО «Научно-производственное объединение «Марс» (ФНПЦ АО НПО «Марс»).

Результаты апробации предложенных моделей, алгоритмов и реализующего их программного комплекса подтверждено соответствующими актами об использовании результатов диссертации.

Направления развития исследований

Проведенные исследования ставят новые исследовательские задачи, которые могут быть положены в развитие настоящей работы, в частности, целесообразно продолжить исследования и разработку моделей и методов сентимент-анализа слабоструктурированных ресурсов с учетом контекстов и особенностей выразительности естественного языка, исследования применимости иных моделей n-арной классификации социальных портретов пользователей по данным социальных сетей по большому числу основных факторов, а также их реализация в экспертной системе, позволяющей оперативно реагировать на технологические изменения, внесенные в API социальных сетей.

Выводы

Содержание результатов исследований позволяет сделать вывод, что диссертационная работа представляет научный интерес и несомненную практическую ценность и полностью удовлетворяет требованиям ВАК РФ, предъявляемым к кандидатским диссертациям, а соискатель Андреев И.А. заслуживает присуждения ученой степени кандидата технических наук по специальности 05.13.01 – «Системный анализ, управление и обработка информации (информационные технологии и промышленность)».

Научный руководитель,
доцент кафедры «Информационные системы»,
ФГБОУ ВО «Ульяновский государственный технический
университет», кандидат технических наук
«9» июня 2022 года

Мошкин В.С.

Подпись Мошкина В.С. заверено
Проректор на научной работе УлГТУ
доктор технических наук, доцент



Наместников А.М.