

## ОТЗЫВ

на автореферат диссертации 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)

Развитие вычислительной техники и глобальная цифровизация всех областей жизни человека вызвало в том числе и модификацию социальных отношений, выражающуюся в переносе общения в новую среду: чаты, блоги, и различные социальные сети ориентированные на текстовый, аудио, графический или видео-контент. Для поддержки этих процессов современный мир наполнен огромным количеством электронных устройств, web и мобильных приложений, одной из основных задач которых является коммуникация между людьми.

Каждый год аудитория гаджетов и социальных сетей растет, в связи с этим происходит и рост объема генерируемого контента и данных пользователей. Примером платформы, где каждый человек может создать своего цифрового двойника является «Мир социальных сетей». Каждый подобный цифровой портрет частично отражает некоторые характеристики своего создателя и анализ данной неструктурированной информации, требует разработки новых подходов к переработке данных виртуального мира.

Поэтому можно сделать вывод, что исследования в области обработки текстовой информации социальных сетей в задачах формирования социального портрета пользователя являются актуальными и решают важную научно-техническую задачу.

Представленная в автореферате работа обладает научной новизной. Кратко её можно описать следующим образом:

1. Алгоритм формирования обучающей выборки для обучения моделей классификации в задачах сентимент-анализа текстовых данных, обладающий новизной, выражаемой совместным использованием словарей авторских символов выражения эмоций и ключевых фраз.

2. 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. Метод определения психологических характеристик пользователя социальных сетей, обладающий новизной, выраженной в гибридизации алгоритмов обработки естественного языка текстовых данных, машинного обучения и метода «Большой пятерки».

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

Работа обладает как теоретической значимостью, которая включает в себя разработанные модели и алгоритмы sentiment-анализа и психолингвистического анализа, так и практической значимостью, которая заключается в разработке программного продукта, позволяющего автоматизировать построение портрета пользователя социальных сетей.

В автореферате можно заметить следующий недостаток: приведены эксперименты по оценке эффективности алгоритмов классификации тональности с использованием только двух языковых моделей – Word2Vec и BERT, при этом не встречается обоснования их применения. В настоящее время существует большой набор модификаций языковых моделей (ELMo, GloVe, TinyBERT), которые могли бы показать высокую эффективность в данной задаче.

Упомянутый недостаток не снижает научной и практической ценности работы, которая удовлетворяет требованиям, предъявляемым ВАК РФ к кандидатским диссертациям, а автор диссертационной работы Андреев Илья Алексеевич заслуживает присуждения ученой степени кандидата технических наук по специальности 05.13.01 – Системный анализ, управление и обработка информации (информационные технологии и промышленность).

«12» июля 2022 г

Чоркаев Олег Эдуардович

Подпись Чоркаев О. Э.  
Начальник отдела управления персоналом  
С.А. Зинченко



М.П.

Чоркаев Олег Эдуардович, к.т.н., специальность 05.13.12, доцент кафедры «Организации аэропортовой деятельности и информационных технологий».

Федеральное государственное бюджетное образовательное учреждение высшего образования «Ульяновский институт гражданской авиации имени главного маршала авиации Б.П. Бугаева»

432071, Ульяновская обл., г. Ульяновск, ул. Можайского д. 8/8.

телефон: + 7 (8422) 39-81-81

e-mail:uvau@list.ru

## ОТЗЫВ

на автореферат диссертации 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)

Неотъемлемой частью современного социума является общение в «онлайне». Большая часть информации, находящейся в сети Интернет, представлена в текстовом виде на естественном языке. Необходимость обработки подобной информации породило новую прикладную область - компьютерную лингвистику, которая является направлением прикладной лингвистики, ориентированное на использование компьютерных программ обработки языка в тех или иных условиях. Данная дисциплина появилась в 1960-е годы и вначале представляла собой разработку методов для общения человека с ЭВМ на естественном или ограниченно естественном языке.

Стремительный рост массивов информации, состоящих из наборов коротких текстов, способствует интенсификации исследований в области развития методов обработки текстов с применением машинного обучения. Недостаточное количество исследований для русского языка объясняется двумя основными причинами: сложность обработки русского языка для автоматизированной обработки и меньшее число задействованных исследователей.

Из вышесказанного можно сделать вывод, что исследования в области обработки текстовой информации социальных сетей в задачах формирования социального портрета пользователя являются актуальными и решают важную научно-техническую задачу.

Проведенный автором автореферата анализ современных моделей и методов обработки коротких текстов определил следующие основные задачи диссертационной работы: разработать алгоритм формирования обучающей выборки для последующего использования при обучении моделей классификации в задачах sentiment-анализа текстовых данных, разработать подход к сопоставлению профилей пользователей в разных социальных сетях, разработать метод определения психологических характеристик пользователя социальных сетей и разработать алгоритм анализа эмоциональной окраски русскоязычных текстовых данных социальных сетей. Кроме вышечисленного, автором была поставлена задача провести апробацию разработанных алгоритмов и методов и внедрить полученные результаты в практическую задачу.

Результаты исследования, полученные автором диссертации, обладают научной новизной. Вычислительные эксперименты, описанные в

автореферате, подтверждают достоверность выдвинутых положений. Практическая ценность также присутствует и состоит в разработанном программном комплексе.

В тексте автореферата можно выделить следующие недостатки:

1. Одним из пунктов, вынесенных автором в качестве научной новизны, является методика объединения профилей пользователей из различных социальных сетей. В экспериментах показано, что текущего набора алгоритмов сравнения данных профилей, примененных последовательно, достаточно для объединения профилей 100 человек (всего 204 профиля) разных социальных сетей. Однако, данная методика может показать иной уровень эффективности при обработке данных более крупной выборки.

2. В списке публикаций встречаются опечатки и ошибки форматирования.

Указанные недостатки не снижают общую научную и практическую ценность работы. Работа полностью удовлетворяет требованиям, предъявляемым ВАК РФ к кандидатским диссертациям, а соискатель Андреев Илья Алексеевич заслуживает присуждения ученой степени кандидата технических наук по специальности 05.13.01 – Системный анализ, управление и обработка информации (информационные технологии и промышленность).

«30» августа 2022 г

Виктор Сергеевич Луфаров

Луфаров Виктор Сергеевич, к.т.н., специальность **05.13.01** (Системный анализ, управление и обработка информации (по отраслям)), ведущий программист лаборатории.

Филиал федеральное государственное бюджетное образовательное учреждение высшего образования «Национальный исследовательский университет «МЭИ» в г. Смоленске

214013, РФ, г. Смоленск, Энергетический проезд, дом 1.

телефон: +7 (4812) 65-14-61

e-mail: mail@sbmpei.ru



Синдром перин  
Луфаров В.С. уполномочен  
Кадровый отдел кадров  
Росиф Русенков В.А.

### ОТЗЫВ

на автореферат диссертации 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)

Последние годы характеризуются бурным всплеском числа пользователей различных мессенджеров и социальных сетей. При этом информация, оставляемая пользователями в виде фотографий, видеоматериала, аудиозаписей, собственных постов, комментариев или даже эмоджи, потенциально может быть полезной при составлении некоторого портрета пользователя социальной сети. Дальнейшее углубление в ходе анализа в перспективе позволяет получить цифровой портрет пользователя. Одним из крупнейших источников данных в соцсетях наряду с изображениями и видео, является текст. В силу некоторых особенностей, свойственных текстовым данным, они требуют специализированных алгоритмов обработки. Однако классические алгоритмы не всегда эффективно справляются с задачей анализа текстов, и сегодня во многих сферах часто такая аналитика выполняется человеком. В связи с этим, тематика, цель и задачи диссертационного исследования Андреева И.А. непременно являются актуальными, а полученные результаты имеют прикладной характер.

Научной новизной обладают решения, связанные с комплексированием информации профилей пользователя из различных социальных сетей, обученные модели искусственного интеллекта, гибридные методы обработки на базе семантического анализа и машинного обучения. Теоретическая значимость заключается в создании новых моделей и алгоритмов психолингвистического и семантического анализа текстов.

Вместе с тем к тексту автореферата имеется ряд замечаний:

1) В описании алгоритма формирования обучающей выборки Шаг №4 включает в себя предобработку данных. Однако предобработка – это достаточно важный шаг, детальное описание которого в автореферате опущено.

2) В таблице 1 и далее приводятся метрики точности разных моделей. Однако отсутствует информация о распределении классов в обучающей и тестовой выборках, без которой сложно судить о качестве работы модели, даже если точность составляет 87%.

3) В таблице 2 в качестве алгоритма классификации указана «Линейная регрессия». Вероятно, автор имел в виду логистическую регрессию?

Несмотря на указанные замечания, диссертационная работа Андреева И.А. заслуживает положительной оценки, поскольку в ней решены актуальные задачи, связанные с разработкой методического, алгоритмического и программного обеспечения для сокращения временных затрат при составлении цифровых

портретов пользователей соцсетей. По результатам исследования опубликовано 4 работы в изданиях, входящих в перечень ВАК, а также 11 работ в изданиях, рецензируемых международными системами цитирования – Web of Science и Scopus. Автором проведена достаточная апробация результатов диссертационного исследования в рамках участия на конференциях различного уровня.

Учитывая вышесказанное, считаю, что работа «Исследование методов и алгоритмов обработки текстовой информации социальных сетей в задачах формирования социального портрета пользователя» соответствует требованиям п.9 Положения о присуждении ученых степеней, утвержденного постановлением Правительства Российской Федерации от 24.09.2013 №842, содержание автореферата диссертации соответствует специальности 05.13.01 – Системный анализ, управление и обработка информации (информационные технологии и промышленность), а ее автор, Андреев Илья Алексеевич, заслуживает присуждения ученой степени кандидата технических наук.

Доцент Департамента анализа данных и  
машинного обучения Факультета информационных  
технологий и анализа больших данных  
федерального государственного образовательного  
бюджетного учреждения высшего образования  
"Финансовый университет при  
Правительстве Российской Федерации",  
кандидат технических наук



Андрянов Никита Андреевич

Контактная информация организации:

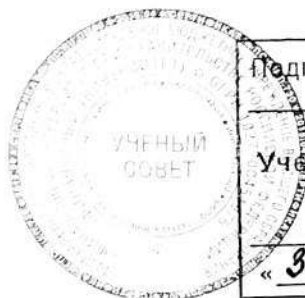
федеральное государственное образовательное бюджетное учреждение высшего образования  
«Финансовый университет при Правительстве Российской Федерации»

Адрес: Российская Федерация, 125167, Москва, пр-кт Ленинградский, д. 49/2

Сайт: <http://www.fa.ru/>

Тел.: +7(499)943-98-29

E-mail: [naandriyanov@fa.ru](mailto:naandriyanov@fa.ru)



Подпись

*Андрянова Н.А.*

ЗАВЕРЯЮ

Ученый секретарь Ученого совета  
Финансового университета

*Звильс* В.В. Звягинцева

«30» августа 2022 г.

## ОТЗЫВ

на автореферат диссертации 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)

Большая часть информации, находящейся в сети Интернет, представлена в текстовом виде на естественном языке. Обработка такой информации – сложный процесс, требующий вмешательства эксперта или использования методов компьютерной лингвистики для трансформации данной информации в структурированную, в связи с чем в настоящее время возрастает актуальность лингвистических исследований извлечения структурированных данных из неструктурированных массивов текстовой информации, классификации и кластеризации информации, нацеленных на анализ самих сообщений в сети.

Технологии обработки естественного языка сделали большой прорыв в области вычислений и искусственного интеллекта. Сейчас результаты обработки естественного языка применяются во многих задачах построения чат-ботов, определении тональности отзывов о компании или продукте, разработке голосовых помощников (в этом случае звуковая информация предварительно превращается в текстовую), автоматическом переводе текста и фильтрации нежелательных сообщений. Традиционными задачами в области анализа текста в настоящее время являются анализ тональности текста и автоматическая тематическая классификация.

Из вышеперечисленного можно сделать вывод, что исследования в области обработки текстовой информации социальных сетей в задачах формирования социального портрета пользователя являются актуальными и решают важную научно-техническую задачу.

Заявленная цель диссертационного исследования Андреева Ильи Алексеевича - это снижение трудозатрат на построение портрета пользователей социальных сетей, который достигается автоматизацией и анализа открытых русскоязычных текстовых данных.

Полученные автором диссертационного исследования результаты обладают научной новизной, которая заключается в следующем:

- Новый алгоритм формирования обучающей выборки, который можно применить при обучении моделей классификации в задачах сентимент-анализа текстовых данных, отличающийся совместным использованием словарей авторских символов выражения эмоций и ключевых фраз;

- Новый подход к сопоставлению профилей пользователей в разных социальных сетях, который отличается совместным использованием

подходов анализа графической информации, структурированных данных анкет, текстовых данных и социальных графов профилей;

- Разработанный метод определения психологических характеристик пользователя социальных сетей по методу «Большой пятерки», отличающийся гибридизацией алгоритмов обработки естественного языка текстовых данных и машинного обучения;

- Предложенный алгоритм анализа эмоциональной окраски русскоязычных текстовых данных социальных сетей, отличающийся совместным использованием семантических подходов и методов машинного обучения.

Практическая ценность работы заключается в разработанном программном комплексе, который автоматизирует построение портрета пользователя социальных сетей на основе его публичных данных.

Недостатком автореферата является тот факт, что автором не указан объем применяемых им словарей: ключевых фраз на базе тезауруса WordNet-Affect и авторских символов выражения эмоций, которые используются при генерации обучающей выборки.

Стоит заметить, что данный недостаток не снижает научной и практической ценности работы, результаты которой представляют несомненный научный интерес и которая удовлетворяет требованиям, предъявляемым ВАК РФ к кандидатским диссертациям, а автор диссертационной работы 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).

« 02 » \_\_ сентября \_\_\_\_\_ 2022 г



Невзорова Ольга Авенировна, к.т.н., специальность 05.13.17, доцент кафедры информационных систем

Федеральное государственное автономное образовательное учреждение высшего образования «Казанский (Приволжский) федеральный университет»

420008, Россия, Республика Татарстан, г. Казань, ул. Кремлевская, д. 18, корп. 1.

телефон: +7(843)292-69-77

e-mail: public.mail@kpfu.ru

