ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА

на диссертацию Романова Александра Сергеевича «Методология идентификации автора текстовой информации для решения задач кибербезопасности» на соискание ученой степени доктора технических наук по специальности 2.3.6 – Методы и системы защиты информации, информационная безопасность.

Актуальность темы диссертационной работы

Идентификация автора текстовой информации играет ключевую роль в борьбе с киберпреступностью: с ее помощью можно выявлять распространителей ложной и вредоносной информации (пропаганды экстремизма, терроризма, разжигания ненависти и т.п.) и пресекать их деятельность. Актуальность темы также обусловлена потребностями в установлении авторства в образовательной среде (для проверки оригинальности работ) и в сфере защиты интеллектуальной собственности.

Таким образом, исследование, направленное на разработку методологии идентификации автора текстовых данных, полностью соответствует текущим потребностям обеспечения информационной безопасности и паспорту специальности 2.3.6, что подтверждает высокую актуальность выполненной работы.

Анализ содержания диссертационной работы

Диссертационная работа состоит из введения, шести глав, заключения, списка литературы и четырех приложений. Общий объем диссертации составляет 400 страниц. Список литературы содержит 334 источника.

Во **введении** сформулированы общая характеристика и предпосылки исследования, обоснована актуальность темы, определены цель и основные задачи работы, изложены научная новизна, теоретическая и практическая значимость полученных результатов.

Глава 1 представляет обзор современного состояния проблемы и существующих решений в области анализа текстов для задач кибербезопасности. Рассмотрены различные аспекты идентификации автора текста в контексте информационной безопасности, в том числе вопросы защиты авторских прав в цифровой среде и выявления деструктивного контента.

Глава 2 посвящена разработке методологии идентификации автора. В ней предложена модель процесса создания текста в киберсреде, учитывающая атрибуты автора, вид его деятельности при создании текста, специфику среды (например, платформа или формат текста) и особенности решаемых задач кибербезопасности. Предложенная методология интегрирует классические методы анализа с современными подходами глубокого обучения и семантического анализа текста.

Глава 3 содержит описание прикладных методик идентификации автора естественно-языкового текста. Представлены методы решения задач закрытой и открытой атрибуции с использованием комбинации методов машинного обучения и нейросетевых моделей. В частности, автором предложен подход, сочетающий рекуррентно-сверточные нейронные сети (GRU+CNN) с классическими методами машинного обучения – методом опорных векторов, и отбором информативных признаков (генетическим алгоритмом и методом на основе регуляризации). Особое внимание уделено сложным случаям, в которых использовались тексты, созданные генеративными нейросетями, и постановке экспериментов, в которых истинный автор может отсутствовать в обучающей выборке (открытая атрибуция). Отмечу, что точность

предложенной методики превышает результаты аналогичных исследований на 2-14%, требуя меньшего количества символов для анализа.

Глава 4 посвящена идентификации автора исходного кода программы. В этой главе диссертации предложена методика на основе глубокой нейросетевой модели CodeBERT. Методика учитывает ряд факторов, затрудняющих установление авторства кода: обфускацию, стилистические изменения кода, коллективную разработку, использование общепринятых стандартов кодирования, а также случаи, когда код может быть сгенерирован нейросетью. Экспериментально доказано, что использование глубокой модели позволяет эффективно выделять признаки, характерные для стиля программирования конкретных авторов. При апробации данной методики удалось получить результаты, превышающие аналоги до 30%.

Глава 5 рассматривает задачи идентификации отдельных атрибутов автора текста, прежде всего, пола (гендера) и возраста, а также выявлению текстов деструктивной и экстремистской направленности и идентификации их авторов. Приведены результаты экспериментов на корпусах текстов из социальных сетей, демонстрирующие высокую точность предлагаемых подходов. Стоит отметить высокую эффективность методик – применение методики идентификации пола позволяет достигнуть результатов, превышающие точность аналогов до 9%, возраста – до 2%, выявления экстремизма – до 22%.

В **главе 6** представлено исчерпывающее описание алгоритмического и программного обеспечения, реализующего предложенные методики.

В **заключении** подведены итоги проведенного исследования, сформулированы общие выводы и рекомендации. Поставленные задачи решены полностью, достигнута цель работы – разработана комплексная методология идентификации автора текстовой информации, ее создание имеет важное хозяйственное значение для информационной безопасности, защиты авторского права, компьютерной лингвистики.

Каждая глава диссертации завершается краткими выводами, отражающими основные результаты соответствующего этапа работы. Содержание диссертации логично структурировано, материал изложен последовательно – от обзора проблемы и теоретических основ до разработки конкретных методик и их практической апробации. Такое построение свидетельствует о целостности исследования и последовательном решении поставленных задач.

Научная новизна

Диссертационная работа отличается высокой научной новизной.

1. Автором впервые предложена комплексная методология идентификации автора текстовой информации в интересах кибербезопасности, которая учитывает специфику как естественно-языковых текстов, так и исходных кодов программ, а также возможные атаки и противодействие методам атрибуции.
2. В рамках методологии разработан новый подход к моделированию процесса создания текста в киберсреде, позволяющий учитывать влияние разнообразных факторов (семантические особенности текста на разных уровнях, характеристики автора, условия создания текста и др.) на формирование авторского стиля.
3. Кроме того, в диссертации предложен ряд новых методик для решения отдельных задач идентификации и профилирования автора текста, которые ранее комплексно не рассматривались в контексте информационной безопасности. В их числе методика идентификации автора естественного языка с использованием комбинации современных нейросетевых моделей (GRU+CNN) и методов классического машинного обучения (SVM с генетическим отбором признаков и методами регуляризации),

учитывающая случаи закрытой и открытой атрибуции, в том числе для текстов, сгенерированных нейросетями.

4. Методика идентификации автора программного кода на базе глубокой модели CodeBERT, адаптированной к задачам авторства в условиях обфускации и командной разработки.

5. Методика определения возраста автора текста, основанная на модели fastText. Особенностью является автоматизация процесса фильтрации данных на основе методов компьютерного зрения. Методика уникальна предоставлением возможности анализа текстов, созданных несовершеннолетними.

6. Методика обнаружения экстремистских текстов с применением BERT и семантического анализа, основанная на комбинации глубоких нейросетевых моделей GRU+CNN и BERT.

7. Методика классификации автора по гендеру и биологическому полу впервые учитывает основные ЛГБТ-гендеры при анализе текстов на русском языке. Методика основана на комбинации методов семантического анализа, частотных распределений триграмм и применении глубокой нейросетевой модели BERT.

8. Предложена методика проверки однородности текста и поиска заимствований, отличающаяся применением сиамских НС SimNN, обученных с контрастивной и тройной функциями потерь. Методика впервые решает задачу выявления заимствований для случаев открытого множества авторов-кандидатов и использования искусственной генерации текстов.

Все перечисленные решения являются новыми и оригинальными, получены лично соискателем.

Теоретическая и практическая значимость

Теоретическая значимость работы состоит в развитии научных представлений о методах и системах защиты информации применительно к анализу текстовых данных. Предложенная в диссертации методология и связанные с ней модели расширяют теоретическую базу в области идентификации автора текста и компьютерной лингвистики. Работа обобщает и систематизирует ранее разрозненные подходы, связывая классические лингвистические методы с современными алгоритмами искусственного интеллекта. Таким образом, результаты исследования обогащают научно-методический аппарат информационной безопасности новыми знаниями о характеристиках авторского стиля и о том, как эти характеристики могут быть использованы для решения различных задач (идентификация автора, определение его социальных атрибутов, выявление аномалий в тексте и т.д.).

Практическая значимость результатов диссертации также является весьма высокой. Разработанные методики и алгоритмы могут найти прямое применение в системах обеспечения кибербезопасности и в правоохранительной практике. Например, предложенные автором подходы позволяют создавать инструменты для автоматического выявления авторов деструктивных сообщений в Интернете, что актуально для предотвращения распространения экстремистских материалов и угроз национальной безопасности. Методы идентификации автора и профилирования (пол, возраст, идеологические взгляды) могут использоваться при расследовании киберпреступлений, в экспертизах по делам, связанным с распространением запрещенного контента, а также для защиты авторских прав (для подтверждения или оспаривания авторства документа). Практическая ценность подтверждается и фактом создания программного обеспечения (получены свидетельства о регистрации программ для ЭВМ и баз данных), внедренного для решения практических задач информационной безопасности в организации. Также о

высокой практической значимости свидетельствуют и полученные автором результаты, превосходящие аналоги до 30%.

Обоснованность и достоверность результатов

Полученные в диссертации научные результаты и выводы являются обоснованными и достоверными. Это обеспечено корректным применением современных методов исследований – математической статистики, машинного обучения и вычислительных экспериментов при решении поставленных задач. Автор провел обширные экспериментальные исследования на разнообразных корпусах текстовых данных (включая как художественные тексты и пользовательские комментарии, так и исходные коды программ), что позволило всесторонне проверить работоспособность предложенных методов. Использование стандартных метрик оценки качества классификации и сравнение с базовыми подходами подтверждают правильность сделанных выводов. В диссертации прослеживается согласованность полученных результатов с выдвинутыми гипотезами и целями исследования.

Полнота опубликования результатов работы

Основные результаты диссертационного исследования опубликованы в достаточном объеме, что соответствует установленным требованиям ВАК. По теме диссертации автором опубликован ряд статей в рецензируемых научных изданиях, рекомендованных ВАК (18 статей), а также в зарубежных журналах, индексируемых в Web of Science и Scopus (13 статей). Также результаты отражены в монографии, а практическая ценность подтверждена получением 8 свидетельств о регистрации программ и 3 для ЭВМ и баз данных.

Соответствие автореферата содержанию диссертации

Автореферат диссертации полностью соответствует ее содержанию. В автореферате отражены все основные положения, выводы и результаты исследования, представленные в полном тексте диссертации. Материалы автореферата структурированы аналогично диссертации, что позволяет получить корректное представление о цели, задачах, методах и результатах работы. Существенных расхождений между авторефератом и диссертацией не выявлено – ключевые идеи и результаты совпадают, а форма изложения в автореферате адекватно передает суть проделанной работы.

Замечания по диссертационной работе

Вместе с тем по диссертационной работе имеются следующие замечания:

1. В главе 3 представлена методика идентификации автора естественно-языкового текста для случая открытой атрибуции, однако критерии выбора порогов для евклидова расстояния не представлены. Следовало бы пояснить, как именно выбирались эти пороговые расстояния.
2. В таблицах 3.2. и 3.9 представлены результаты экспериментов по отбору информативных признаков для художественных текстов. Приведена достигнутая точность и количество подобранных признаков в лучшем подмножестве. При этом, полученное количество признаков – достаточно велико (от 2100 до 5821). В то же время в своих работах автор указывает на эффективность совместного применения ГА с SVM на наборе 1168 признаков, получая более точные результаты для меньшего количества признаков (100, 400). Возникает вопрос – не следовало ли автору отталкиваться при проведении экспериментов от ранее полученного в работе [289] набора из 1168 элементов?
3. В подразделах 3.5.1–3.5.4 описаны методы открытой атрибуции, но не указано, каким образом производилась валидация построенных моделей.

4. При формировании набора данных из мессенджера Telegram для проведения экспериментов по определению пола и гендера автора текста, реализован механизм автоматической разметки данных. В работе следовало представить данный механизм более подробно.

5. В главе 5 используется модель с четырьмя гендерами, в то время как современные источники указывают на существование более широкого спектра (от 46 до 78). Следовало бы пояснить ограничение до четырех.

Отмеченные недостатки носят частный характер и не снижают общей положительной оценки работы.

Заключение

Диссертационная работа А.С. Романова «Методология идентификации автора текстовой информации для решения задач кибербезопасности» является самостоятельной и завершенной научно-квалификационной работой. Полученные результаты представляют собой решение актуальной научной проблемы разработки методологических основ и практических подходов идентификации автора текстовых данных в интересах обеспечения информационной безопасности страны, имеющей важное хозяйственное значение. Формулировки выводов и предложенные методики свидетельствуют о высоком уровне проведенного исследования, а внедрение полученных решений подтверждает их практическую ценность. Диссертация полностью соответствует требованиям п. 9-14 «Положения о присуждении ученых степеней», утвержденного Постановлением Правительства РФ от 24.09.2013 г. № 842. Ее автор, Александр Сергеевич Романов, безусловно заслуживает присуждения ученой степени **доктора технических наук** по специальности 2.3.6 – Методы и системы защиты информации, информационная безопасность.

Я, Аникин Игорь Вячеславович, даю свое согласие на включение своих персональных данных в документы, связанные с работой диссертационного совета, и их дальнейшую обработку.

Официальный оппонент,
доктор технических наук, профессор, заведующий кафедрой систем информационной безопасности Казанского национального исследовательского технического университета им. А.Н. Туполева-КАИ

«2» 09 2025 _____ Аникин Игорь Вячеславович

Диссертация на соискание ученой степени доктора технических наук защищена по специальности 05.13.19 – Методы и системы защиты информации, информационная безопасность (2018 г).

ФГБОУ ВО Казанский национальный исследовательский
технический университет им. А.Н. Туполева-КАИ (КНИТУ-КАИ)
420111, Республика Татарстан, город Казань, ул. Карла Маркса, д. 10
Тел: +7 (843) 231-97-34
Email: anikinigor777@mail.ru

Подпись Аникина И. В.
заверяю. Начальник управления
делопроизводства и контроля

Подпись Аникина Игоря Вячеславовича заверяю