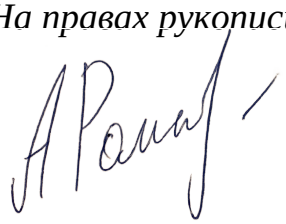
На правах рукописи



Романов Александр Сергеевич

МЕТОДОЛОГИЯ ИДЕНТИФИКАЦИИ АВТОРА ТЕКСТОВОЙ ИНФОРМАЦИИ ДЛЯ РЕШЕНИЯ ЗАДАЧ КИБЕРБЕЗОПАСНОСТИ

Специальность 2.3.6 – Методы и системы защиты информации,
информационная безопасность

Автореферат
диссертации на соискание ученой степени
доктора технических наук

Томск – 2025

Работа выполнена в Федеральном государственном автономном образовательном учреждении высшего образования «Томский государственный университет систем управления и радиоэлектроники» (ТУСУР)

Научный консультант — доктор технических наук профессор **Шелупанов Александр Александрович**

Официальные оппоненты:

Вульфин Алексей Михайлович,
доктор технических наук, профессор кафедры вычислительной техники и защиты информации ФГБОУ ВО «Уфимский университет науки и технологий»

Спицын Владимир Григорьевич,
доктор технических наук, профессор, профессор отделения информационных технологий ФГАОУ ВО «Национальный исследовательский Томский политехнический университет»,

Аникин Игорь Вячеславович,
доктор технических наук, профессор, зав. кафедрой систем информационной безопасности ФГБОУ ВО «Казанский национальный исследовательский технический университет им. А.Н. Туполева»

Ведущая организация — Федеральное государственное автономное образовательное учреждение высшего образования «Новосибирский национальный исследовательский государственный университет»

Защита состоится «09» октября 2025 г. в 15 час. 15 мин. на заседании диссертационного совета 24.2.415.04, созданном на базе ТУСУРа, по адресу: 634050, г. Томск, проспект Ленина, 40, ауд. 201.

С диссертацией можно ознакомиться в библиотеке ТУСУРа по адресу: 634045, г. Томск, ул. Красноармейская, 146 и на официальном сайте ТУСУРа:
https://postgraduate.tusur.ru/urls/51p1ctn6

Автореферат разослан «___» _____ 2025 г.

Ученый секретарь
диссертационного совета  Костюченко Евгений Юрьевич

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы

В эпоху, когда вычислительные технологии стали неотъемлемой частью всех аспектов человеческой деятельности, критически важным является обеспечение защиты данных и систем от киберугроз. Первостепенной задачей становится разработка и внедрение передовых систем и технологий для обеспечения информационной безопасности. Одним из направлений, где требуются такие инструменты, является анализ больших объемов текстовых данных, направленный на решение специфических задач кибербезопасности, в частности, определения авторства текста.

В контексте информационной безопасности идентификация автора текста играет ключевую роль в борьбе с киберпреступностью. С ее помощью становится возможным выявление и пресечение действий распространителей ложной и вредоносной информации, такой как пропаганда экстремизма, терроризма и движений, противоречащих традиционным ценностям России, а также розжига ненависти и розни. В образовательной сфере тема актуальна для обеспечения объективности оценки и подтверждения подлинности работ. В области лингвистики она предоставляет уникальные возможности для изучения стилистических особенностей языка, разработки новых методов анализа и понимания эволюции языка.

Задача идентификации автора текстовой информации не теряет актуальности уже более 100 лет, что подтверждается статистическими данными о публикациях по теме. На рисунке 1 представлены данные о количестве результатов, полученных для запросов к базе цитирования Google Scholar на русском и английском языке на тему идентификации автора текста. Результаты охватывают общее количество публикаций, тезисов и статей по теме за период с 2010 (год защиты автором кандидатской диссертации) по 2024 год. Видно, что интерес к теме российских и зарубежных исследователей только растет.

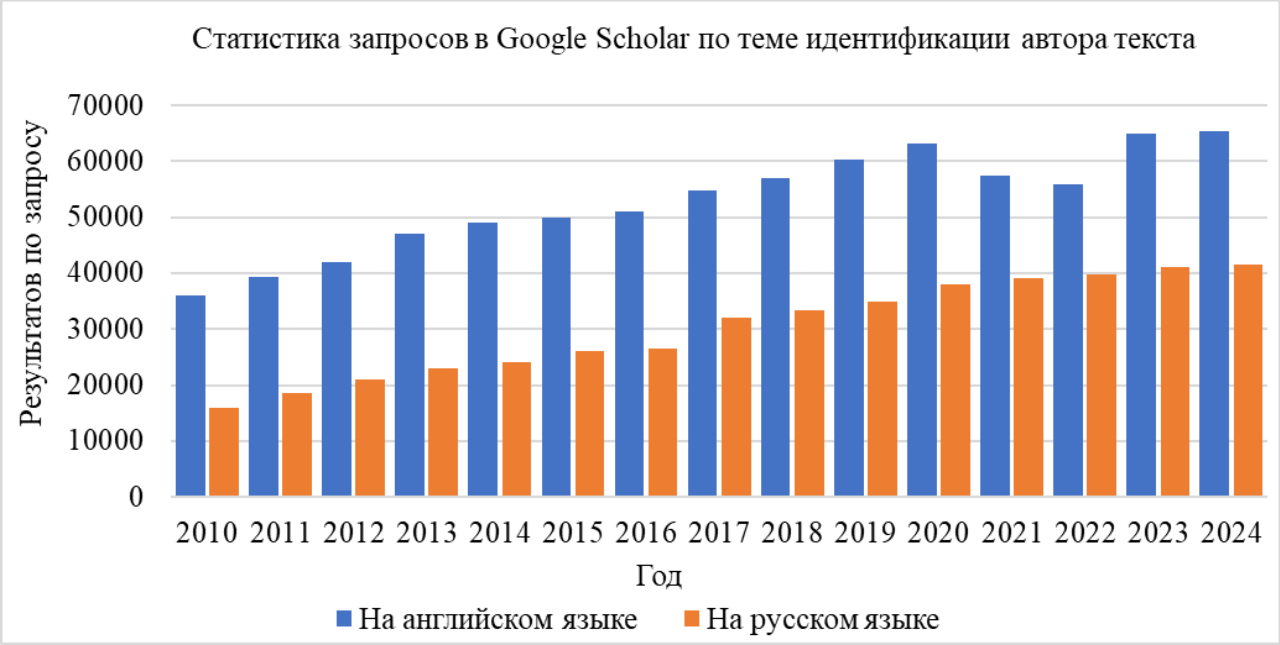


Рисунок 1 – Статистика запросов по теме исследования

Наибольший вклад в решение отдельных задач определения автора текстовой информации и атрибутов автора в России был внесен Н.А. Морозовым, А.А. Марковым, В.П. Фоменко и Т.Г. Фоменко, Д.В. Хмелевым, Ю.В. Сидоровым, А.Ю. Комиссаровым, О.Г. Шевелевым, В.В. Поддубным, М.А. Марусенко, Р.В. Мещеряковым, А.О. Исхаковой, А.А. Роговым, Н.Д. Москиным, А.Г. Сбоевым, М.Е. Сухопаровым, В.Д. Стремоуховым и др.

Наиболее цитируемыми зарубежными авторами, занимавшимися частными вопросами определения авторства и атрибуции, являются A. Abbasi, T.C. Menden hall, A.Q. Morton, E. Stamatatos, J. Weerasinghe, M. Najafi, H. Gómez Adorno, D. Hernández, S. Burrows, B. Alsulami, W. Wisse, F. Ullah, S. Hedegaard, M. Yang, G. Farnadi, D. Nguyen, C.Y. Park, S. Afroz и др.

Перечисленные исследователи успешно решали частные задачи атрибуции, не связанные с обеспечением кибербезопасности напрямую. Однако проблема систематизации существующих подходов к идентификации автора текста, а также разработки комплексной методологии, позволяющей эффективно решать взаимосвязанные задачи авторства в интересах национальной кибербезопасности страны, остается открытой.

В диссертационной работе предлагается комплексное решение **важной научной проблемы** идентификации автора текстовой информации и таких авторских атрибутов, как пол, возраст, идеология и взгляды, основанное на передовых технологиях искусственного интеллекта и машинного обучения, соответствующих приоритету научно-технологического развития России на 2027–2032 г., согласно Указу Президента Российской Федерации № 642 от 01.12.2016 г. (ред. от 15.03.2021) «О Стратегии научно-технологического развития Российской Федерации», п. 20, пп. «а».

Цель и задачи исследования

Целью диссертационной работы является создание методологии идентификации автора текстовой информации, включая естественно-языковые тексты и исходные коды программ, для решения задач кибербезопасности.

Для достижения поставленной цели необходимо решить следующие задачи.

1. Провести анализ современного состояния проблемы идентификации автора текстовой информации и использующихся для этого методов анализа и характеристик текста, выделить актуальные задачи кибербезопасности, связанные с определением авторства текста.
2. Разработать обобщенную методологию идентификации автора текстовой информации для решения задач кибербезопасности, учитывающую специфику естественно- и искусственно-языковых текстов.
3. Разработать прикладные методики идентификации автора текстовой информации, сочетающие отдельные элементы методологии.
4. Разработать информационное, алгоритмическое и программное обеспечение для решения задач кибербезопасности, связанных с идентификацией автора текстовой информации.
5. Исследовать и апробировать разработанные методики для решения задач кибербезопасности на корпусах текстов.

Объектом исследования является печатный текст и его характеристики.

Предметом исследования являются характеристики текста, описывающие авторский стиль, методы и алгоритмы машинного обучения, предназначенные для работы с естественно- и искусственно-языковыми текстами.

Методы исследования

Для решения задач, сформулированных в работе, использовались методы математической статистики, вычислительного эксперимента и искусственного интеллекта. При разработке программной системы использовались методы объектно-ориентированного программирования.

Научная новизна полученных результатов проведенного диссертационного исследования и полученных результатов заключается в следующем:

1. **Впервые** предложена комплексная методология идентификации автора текста, учитывающая особенности естественно- и искусственно-языковых текстов и возможные атаки на методы идентификации.

2. Предложена модель создания текста автором в киберсреде, **впервые** учитывающая семантические особенности и информативные признаки текста на разных уровнях иерархического анализа, специфику среды, атрибуты автора и вид деятельности по созданию текста.

3. Предложена методика идентификации автора естественного текста, **отличающаяся** использованием комбинации GRU+CNN и SVM с отбором признаков генетическим алгоритмом и методом на основе регуляризации, **впервые** учитывающая случаи открытой и закрытой атрибуции и текстов, созданных генеративными нейросетями.

4. Предложена методика идентификации автора исходного кода программ, в которой **впервые** используется классификатор на основе глубокой модели CodeBERT. Методика **отличается** учетом комплекса сложных случаев идентификации (обфускация, командная разработка, использование стандартов кодирования и искусственно-сгенерированных кодов программ).

5. Предложена методика бинарной и мультиклассовой классификации по возрастным группам на основе текста, **отличающаяся** использованием модели fastText в сочетании с методами компьютерного зрения для фильтрации обучающих данных.

6. Предложена методика определения текстов деструктивной и экстремистской направленности и определения автора таких текстов, **отличающаяся** использованием моделей GRU+CNN и BERT, методов семантической кластеризации текста и трансферного обучения, что позволяет выявлять запрещенные законодательством РФ тексты.

7. Разработана **новая** методика определения пола и гендера автора русскоязычного текста, **впервые** учитывающая гендеры ЛГБТ* и отличающаяся использованием ансамбля: SVM, обученного на признаковом пространстве, сформированного на основе методов семантической кластеризации, сверточной сети, обученной на частотных распределениях триграмм, сглаженных методом Катца, и 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

* Запрещенная в Российской Федерации экстремистская организация.

выявления заимствований для случаев открытого множества авторов-кандидатов и использования искусственной генерации текстов.

Достоверность и **обоснованность** научных положений, результатов и основных выводов работы обеспечивается корректностью применяемых методов математической статистики, вычислительного эксперимента и искусственного интеллекта, серией практических экспериментов на представительном и репрезентативном корпусе текстов, а также согласованностью полученных данных с результатами других научных групп и положительным эффектом от внедрения полученных результатов.

Теоретическая значимость работы состоит в проработке и систематизации результатов предыдущих исследований, их дополнении и комплексировании с вновь разработанными методиками, моделями и алгоритмами в составе единой методологии идентификации автора текстовой информации.

Практическая значимость.

1. Разработаны репрезентативные корпуса текстовой информации, включающие художественные, любительские тексты и сообщения из социальных сетей и мессенджеров, исходные коды программ, содержащие атрибуты автора (пол, возраст, идеология и др.), которые можно использовать для обучения моделей при решении задач кибербезопасности.

2. Предложенная методика идентификации автора естественно-языкового текста и её программная реализация позволяют существенно сократить необходимый объем текста для идентификации: до 15000 символов для обычных и до 250 символов для коротких текстов, при этом повышая точность на 2-14% по сравнению с аналогами. Методика учитывает случаи закрытой и открытой атрибуции, устойчива к внедрению текстов, созданных генеративными языковыми моделями на основе образцов текстов реальных авторов.

3. Предложенная методика идентификации автора исходного кода и ее программная реализация позволяют повысить точность идентификации автора на 20-30% по сравнению с существующими подходами для современных интерпретируемых и компилируемых языков программирования в простых и осложненных различными факторами (обфускация, командная разработка, стандарты кодирования, искусственная генерация) случаях.

4. Предложенная методика определения текстов деструктивной и экстремистской направленности и ее программная реализация позволяют получить прирост точности выявления таких текстов в сравнении с аналогами до 32%, а также повысить точность идентификации их автора на 22%, что позволяет выявлять запрещенные законодательством РФ тексты.

5. Предложенная методика определения возраста автора текстовой информации и её программная реализация позволяет определять принадлежность автора к конкретной возрастной группе с точностью на 2% превосходящей аналоги, а также учитывать тексты, написанные несовершеннолетними авторами, что не предусматривается аналогами методики.

6. Предложенная методика идентификации пола и гендера автора текстовой информации и её программная реализация позволяет определять пол с точность на 9% выше, чем аналоги, учитывать гендеры и выявлять тексты, созданные пред-

ставителями ЛГБТ-сообщества* с точностью 93%, что позволяет выявлять пропаганду нетрадиционных ценностей, запрещенную законодательством РФ.

7. Предложенная методика проверки текста на однородность и поиска заимствований позволяет выявлять неоднородные фрагменты с точностью до 94%, определять авторство научных текстов или их фрагментов с точностью более 90% и подтверждать авторство с точностью до 99%. Методика учитывает случаи генерации текста моделями 3-го и 4-го поколений.

8. Методология, модели, алгоритмическое и программное обеспечение могут быть использованы при решении смежных задач кибербезопасности и анализа текстовой информации, не рассмотренных в диссертационном исследовании.

Реализация результатов работы.

Результаты диссертационного исследования использованы при выполнении следующих проектов:

1. «Программное обеспечение для исследования характеристик текста в задачах идентификации автора» программы ФСРМПНТ «У.М.Н.И.К.» (договор № КР 04/07 от 9.06.2007 г.; № 014/08 от 9.09.2009 г), 2007–2009 гг. – Руководитель.

2. «Методология анализа и обеспечения кибербезопасности объектов критической информационной инфраструктуры», проект № FEWM-2020-0037 в рамках базовой части государственного задания ТУСУРа на 2020–2022 гг. – Исполнитель.

3. «Методология использования методов машинного обучения для обеспечения кибербезопасности», проект № FEWM-2023-0015 в рамках базовой части государственного задания ТУСУРа на 2023–2025 гг. – Исполнитель.

4. «Технология семантической контент-фильтрации сетевого трафика (нежелательного контента)». Грант на государственную поддержку центров Национальной технологической инициативы на базе образовательных организаций высшего образования и научных организаций. Соглашение о предоставлении субсидии от 14.12.2021 г. № 70-2021-00246, 2022 г. – Руководитель.

5. «Прикладные методики семантического анализа текста в сети Интернет», хоз. договор с ООО «СИБ» (Договор № 10-ТДВ от 28.03.2023), 2023 г. – Исполнитель.

6. «Интерфейсы взаимодействия в киберфизических системах», подпроект № 18 в рамках стратегического проекта № 2 «ИТ, безопасная цифровая среда и киберфизические системы» госпрограммы «Приоритет 2030», 2022–2023 гг. – Исполнитель.

7. «Кибербезопасность объектов критической информационной инфраструктуры. Безопасный искусственный интеллект», подпроект № 53 в рамках стратегического проекта № 2 «ИТ, безопасная цифровая среда и киберфизические системы» госпрограммы «Приоритет 2030», 2024 г. – Исполнитель.

Положения, выносимые на защиту

1. Новая комплексная методология идентификации автора текста, учитывающая особенности естественно- и искусственно-языковых текстов, атаки на методы идентификации автора, основанная на классических и современных методах глубокого и семантического анализа, позволяет создавать прикладные инструменты и решать широкий класс задач кибербезопасности, связанных с обработкой текстовой информации.

---

* Запрещенная в Российской Федерации экстремистская организация.

Соответствует п. 1 паспорта специальности 2.3.6: Теория и методология обеспечения информационной безопасности и защиты информации.

2. Модель создания текста автором в киберсреде позволяет учитывать влияние атрибутов авторов, вида деятельности по созданию текста и специфику среды на семантику текста и авторский стиль, а также особенности решаемой задачи кибербезопасности.

Соответствует п. 1 паспорта специальности 2.3.6: Теория и методология обеспечения информационной безопасности и защиты информации.

3. Методика идентификации автора естественного текста, отличающаяся использованием комбинации GRU+CNN и SVM, обученных на информативных признаках, полученных генетическим алгоритмом и методом на основе регуляризации, позволяет определять автора в случае открытого и закрытого множества кандидатов с точностью до 99% (на 2-14% выше, чем аналоги), устойчива к внедрению имитирующих авторский стиль образцов, созданных генеративными моделями, применима для коротких (до 250 символов) и длинных текстов (от 15000 символов).

Соответствует п. 13 паспорта специальности 2.3.6: Методы и модели выявления и противодействия распространению ложной и вредоносной информации и п. 12 Технологии идентификации и аутентификации пользователей и субъектов информационных процессов. Системы разграничения доступа.

4. Методика идентификации автора исходного кода программы, отличающаяся использованием нового классификатора на основе CodeBERT, позволяет определять автора исходного кода, написанного на интерпретируемом или компилируемом языке, с точностью до 93% в простых случаях и с точностью до 90% (на 20-30% выше, чем у аналогов) в случаях, осложненных обфускацией, стандартами кодирования, искусственной генерацией или командной разработкой.

Соответствует п. 13 паспорта специальности 2.3.6: Методы и модели выявления и противодействия распространению ложной и вредоносной информации.

5. Методика бинарной и мультиклассовой классификации текстов по возрастным группам автора, отличающаяся использованием метода компьютерного зрения VGG-Face для фильтрации недостоверных данных и fastText для классификации, позволяет повысить точность определения возрастной группы, включая группу до 18 лет, до 82% (на 2% выше аналогов).

Соответствует п. 13 паспорта специальности 2.3.6: Методы и модели выявления и противодействия распространению ложной и вредоносной информации.

6. Методика определения текстов деструктивной и экстремистской направленности, отличающаяся использованием моделей GRU+CNN и BERT, методов семантической кластеризации текста и трансферного обучения, позволяет снизить требуемый объем текстового образца до 250 символов и повысить точность выявления текстов деструктивной и экстремистской направленности до 94% и их распространителей до 95% (на 1–22% выше в сравнении с аналогами).

Соответствует п. 13 паспорта специальности 2.3.6: Методы и модели выявления и противодействия распространению ложной и вредоносной информации.

7. Методика определения пола и гендера автора текста на основе ансамбля SVM, BERT и CNN, отличающаяся определением гендеров ЛГБТ-сообщества*,

* Запрещенная в Российской Федерации экстремистская организация.

признанного экстремистским на территории Российской Федерации, позволяет определять пол с точностью 88–92% (на 9% выше аналогов) и признаки ЛГБТ* в тексте с точностью 93%.

Соответствует п. 13 паспорта специальности 2.3.6: Методы и модели выявления и противодействия распространению ложной и вредоносной информации.

8. Методика проверки однородности текста и поиска заимствований, отличающаяся использованием сиамских сетей SimNN, обученных с контрастивной и тройной функциями потерь, позволяет выявлять неоднородные фрагменты научного текста с точностью 94%, определять их авторство на открытом множестве кандидатов, в том числе генеративные модели, с точностью более 90% и подтверждать автора двух и более текстов с точностью до 99%.

Соответствует п. 13 паспорта специальности 2.3.6: Методы и модели выявления и противодействия распространению ложной и вредоносной информации.

9. Разработанное алгоритмическое и программное обеспечение для анализа естественно- и искусственно-языковых текстов, созданные репрезентативные корпуса текстов позволяют идентифицировать автора и его атрибуты (пол, гендер, возраст, настроение, убеждения), выявлять тексты, нарушающие законодательство РФ, и решать прикладные задачи кибербезопасности.

Соответствует п. 5 паспорта специальности 2.3.6: Методы, модели и средства (комплексы средств) противодействия угрозам нарушения информационной безопасности в открытых компьютерных сетях, включая Интернет.

Внедрение результатов работы

Элементы методологии, разработанное программное обеспечение внедрены в АО «Национальный Инновационный Центр», ООО «СИБ», ООО «НТР», ООО «Сибэдж», войсковую часть 51952, ИШИТР ТПУ, ОБК УМВД России по Томской области, на экономическом факультете МГУ имени М.В. Ломоносова, для решения задач кибербезопасности и обработки текста, включая методики определения автора естественного текста, его возраста, пола и гендера, а также обнаружения текстов деструктивной и экстремистской тематики, определения авторства исходных кодов программ. Внедрение показало положительный эффект, состоящий в сокращении временных затрат, повышении точности и эффективности работы отдельных систем и процессов, минимизации потенциальных финансовых и репутационных потерь за счет применения разработанных в диссертации подходов.

Результаты диссертационной работы используются в учебном процессе Томского государственного университета систем управления и радиоэлектроники при изучении дисциплин «Языки программирования» и «Основы построения систем искусственного интеллекта и машинного обучения». Методика проверки однородности текста и поиска заимствований используется для оценки работ студентов.

Личный вклад автора

Все результаты и положения, составляющие научную основу диссертационной работы и выносимые на защиту, получены автором лично. Соавторы, принимавшие участие в отдельных направлениях исследований, указаны в списке основных публикаций по теме диссертации.

Автором самостоятельно разработаны методология идентификации автора текстовой информации, модель создания текста автором в киберсреде, методики

идентификации автора естественно-языкового текста, его возраста, пола и гендера, автора исходных кодов программ, деструктивной и экстремистской направленности в текстах и определения автора таких текстов, проверки однородности текста и поиска заимствований.

Эксперименты проведены с помощью программного комплекса, разработанного по инициативе, под руководством и при непосредственном участии автора, на основе разработанной им методологии, методик, моделей и алгоритмов совместно с аспирантами и студентами ТУСУР научной группы автора.

Апробация работы

Материалы работы докладывались и обсуждались:

- на Всероссийской научно-технической конференции студентов, аспирантов и молодых ученых «Научная сессия ТУСУР», 2006–2024 гг., Томск;
- Международной научно-методической конференции, посвященной 90-летию высшего математического образования на Урале «Актуальные проблемы математики, механики, информатики», 2006 г., Пермь;
- Международной конференции «Interactive Systems and Technologies: The Problems of Human-Computer Interaction», 2007 г., Ульяновск;
- Международной научно-практической конференции «Электронные средства и системы управления», 2007 г., 2009 г., 2013 г., 2015 г., 2017 г., 2018 г., 2019 г., 2020 г. Томск;
- Всероссийской научной конференции «Техническая кибернетика, радиоэлектроника и системы управления», 2008 г., Таганрог;
- Всероссийской научно-практической конференции студентов, аспирантов и молодых ученых с международным участием «Молодежь и современные информационные технологии», 2008 г., 2009 г., Томск;
- Всероссийской научно-практической конференции «Проблемы информационной безопасности государства, общества и личности, безопасность нанотехнологий», 2009 г., Томск;
- Международной конференции по компьютерной лингвистике «Диалог», 2009 г., 2010 г., 2010 г., 2021 г., Москва;
- Всероссийской научно-практической конференции студентов, аспирантов и молодых ученых «Технологии Microsoft в теории и практике программирования», 2010 г., Томск;
- Международной конференции «The II International conference R. Piotrowski's Readings in Language Engineering and Applied Linguistics», 2017 г., 2019 г., Санкт-Петербург;
- XVI Международной научно-практической конференции «Татищевские чтения: актуальные проблемы науки и практики», 2019 г., Тольятти;
- Международной научно-методической конференции «Современное образование: качество образования и актуальные проблемы современной высшей школы», 2019 г., Томск;
- XVI Международной конференции студентов, аспирантов и молодых ученых Перспективы развития фундаментальных наук, 2019 г., Томск;
- Международной конференции «2019 International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON)», 2019 г., Томск;

- Международной конференции «YRID-2020: International Workshop on Data Mining and Knowledge engineering», 2020 г., Ставрополь;
- XVII Международной научно-практической конференции «Перспективы развития фундаментальных наук». Город Томск, 2020 г.;
- на XVIII Международной школе-конференции студентов, аспирантов и молодых ученых Инноватика-2022. Город Томск, 2022 г.

Публикации по теме диссертационной работы

Результаты диссертационной работы отражены в 94 публикациях, в том числе в 31 публикации в рецензируемых научных изданиях, рекомендуемых для опубликования основных научных результатов диссертаций на соискание ученых степеней кандидата и доктора наук, из них 18 публикаций в рецензируемых журналах из перечня ВАК и 13 публикаций в изданиях, индексируемых Web of Science и/или Scopus (из них 7 – в журналах, входящих в первый и второй квартили), а также в 1 монографии, 8 свидетельствах о регистрации программы для ЭВМ и 3 свидетельствах о регистрации баз данных.

Структура и объем работы

Диссертация состоит из введения, шести глав, заключения, списка литературы и 4 приложений. Объем диссертации с приложениями составляет 400 страниц, в том числе 90 рисунков и 59 таблиц. Список литературы включает 334 наименования.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **введении** обоснована актуальность темы диссертационной работы, определены цель и задачи исследования, показаны научная новизна и практическая ценность работы, изложены научные положения, выносимые на защиту.

В **первой главе** приведен обзор проблем и решений в области анализа текстов для задач кибербезопасности.

Один из видов нарушений в киберпространстве – нарушение авторских и смежных прав на текстовые произведения, которое может выражаться, например, в присвоении текста другого человека с целью получения материальной выгоды или попытке выдать авторство созданного текста за авторство другого лица. Методы определения авторства позволяют выявить подобные нарушения и установить личность создателя текста. Атрибуция текстов является важной проблемой в компьютерной лингвистике, журналистике, а также в криминалистике, где знание истинного автора анонимного текста (например, предсмертной записки) может облегчить и ускорить работу правоохранительных органов. Таким образом проблема идентификации автора текста для защиты интеллектуальной собственности является важной задачей кибербезопасности.

В числе атрибутов автора выделяют пол, возраст, образование, профессию, личностные качества и др. Первичными считают гендерный признак и возраст, потому что их конкретизация позволяет установить остальные атрибуты (вторичные) и сузить круг кандидатов при определении автора. Совокупность атрибутов формирует уникальную языковую личность. Одним из направлений практического применения методик определения пола и возраста является криминалистика, где важно определение психологического портрета преступника и профилирование автора. В контексте проблем информационной безопасности подобные

методики являются основой для мониторинга социальных сетей для выявления информации, пропагандирующей нетрадиционные сексуальные отношения, педофилию, смену пола. Важной проблемой также является недопущение детей и подростков до запрещенного или шокирующего контента, который имеет возрастное ограничение «18+», либо ограничение общения с целью предотвращения педофилии. Определение гендерной принадлежности автора текста актуально, поскольку решением Верховного Суда Российской Федерации от 30.11.2023 по делу № АКПИ23-990С «движение ЛГБТ» признано экстремистской организацией и его деятельность запрещена на территории России. Таким образом определение пола и возраста по тексту для дальнейшего выявления признаков пропаганды ЛГБТ* и педофилии является важной задачей кибербезопасности. Актуальным является также вопрос создания программного обеспечения для мониторинга социальных сетей.

Методы определения однородности авторского стиля сообщений можно использовать для продленной аутентификации в социальной сетях и мессенджерах, обнаружения аномалий и необычных паттернов в потоке текстовых данных пользователей сети Интернет. Таким образом задача идентификации автора коротких сообщений в сети Интернет и продленная аутентификация пользователя социальных сетей на основе текста являются важными задачами кибербезопасности.

Анализ и определение авторства текста с учетом эмоциональной составляющей имеют особую важность в контексте борьбы с экстремизмом и терроризмом. В современном информационном обществе социальные сети и онлайн-платформы стали пространством для распространения экстремистских и радикальных идей. Опасность заключается в том, что Интернет-платформы предоставляют экстремистам доступ к широкой аудитории и возможность быстрой и масштабной пропаганды своих идей. Это может воздействовать на уязвимых или подверженных влиянию лиц, включая молодежь, подстрекая их к насилию, террористическим действиям или участию в экстремистских организациях. Мониторинг этой информации в определенные моменты времени и выявление лиц, имеющих целью совершение злонамеренных действий, становится актуальной практической задачей противостояния террористической угрозе и защиты государства. Таким образом анализ настроения автора, определения эмоциональной окраски текста, деструктивных текстов, а также текстов экстремистской направленности, запрещенных законодательством Российской Федерации, является важной задачей кибербезопасности.

Решение задачи определения автора программного кода является критически значимыми для обеспечения безопасности в цифровой среде. Это связано с тем, что подавляющее большинство технологий, а также программных систем и комплексов, упрощающих профессиональную и повседневную деятельность человека, подвержены сбоям. Подобные проблемы могут возникать по ряду причин, например, в результате ошибок разработчиков (при проектировании, реализации и/или внедрении), неправильной эксплуатации пользователями, неполадок смежных систем. Однако наибольшую угрозу несут сбои, происходящие ввиду преднамеренного и/или злоумышленного вмешательства. Несмотря на то, что дея-

* Запрещенная в Российской Федерации экстремистская организация.

тельность по созданию, использованию и распространению вредоносного программного обеспечения запрещена на законодательном уровне и закреплена в ст. 272, 273, 274 Уголовного Кодекса Российской Федерации, технические средства, обеспечивающие эффективное и своевременное установления авторства исходного кода в рамках компьютерных экспертиз, на 2024 год отсутствуют. Анализ исходных кодов программ на предмет авторства осуществляется специалистами в области компьютерной криминалистики вручную или с использованием малоэффективных для данной задачи средств текстового анализа. Таким образом, задача определения автора-вирусописателя представляют особую важность для кибербезопасности и компьютерной криминалистики.

Актуальной задачей кибербезопасности становится проблема создания и усовершенствования методик, учитывающих способы сокрытия авторского стиля (обфускация) и имитации авторского стиля, а также генеративных моделей, позволяющих автоматически генерировать тексты на основе глубоких моделей, обученных на больших текстовых корпусах (GPT). В связи с этим возникает необходимость в проведении дополнительных исследований, направленных на оценку устойчивости методов определения авторства текста к такого рода атакам.

Анализ основных подходов к идентификации автора текста и смежных задач выявил следующие проблемы.

1. Результаты анализа существующих работ в области идентификации автора текстовой информации для целей кибербезопасности подтверждают сложность и неоднозначность решаемых задач. Для эффективного решения задач определения автора естественно- и искусственно-языкового текста, его пола и возраста, эмоционального состояния и идеологии необходим анализ творческой, повседневной и профессиональной деятельности и порождаемой при этом текстовой информации.

2. Методологическая база для комплексного и эффективного интеллектуального анализа текста для кибербезопасности отсутствует. Большинство применяемых исследователями подходов являются морально устаревшими. В то время, как в исследованиях фигурируют популярные алгоритмы – метод опорных векторов (SVM), абстрактные синтаксические деревья (АСД), рекуррентные нейронные сети (RNN) и сверточные нейронные сети (CNN), число глубоких архитектур на базе трансформных моделей (BERT, GPT и др.) и число их гиперпараметров активно растут. На базе таких трансформерных моделей создаются атаки на методы, способные запутать классификатор, снизить точность и, более того, свести ее к нулю. Актуальные подходы должны быть устойчивы к таким атакам за счет своей способности к глубокому анализу неподконтрольных автору и неявных признаков, не поддающихся целенаправленному искажению. Процесс разработки методов определения авторства должен охватывать апробацию зарекомендовавших себя state-of-the-art алгоритмов и современных глубоких нейронных сетей (НС), включая их двунаправленные вариации, комбинации архитектур и ансамбли.

3. Каждая задача, направленная на установление авторства, требует индивидуального подхода. Набор авторских признаков, на который опирается метод принятия решения, может иметь разную эффективность в зависимости от типа задачи, анализируемых текстовых данных, типа языка (аналитический или синтетический). Методы, работающие с высокой эффективностью для англоязычных текстов, необходимо существенно модернизировать для других языков с учетом их особенностей. Различия в типах и жанрах текстов, объеме и числе доступных для

анализа образцов накладывают дополнительные ограничения на возможность применения методов определения авторства.

4. Методы интеллектуального анализа текста вне зависимости от решаемой задачи кибербезопасности должны учитывать семантику и контекст. Семантические признаки являются наиболее информативными. Они отражают краткосрочные и долгосрочные зависимости между языковыми сущностями, приверженность к определенным синтаксическим конструкциям и речевым оборотам, авторские оценочные суждения и идеологию. Контекст может быть особенно полезен при решении задач, имеющих сопутствующий контент – например, при анализе комментариев к посту в тематическом сообществе.

5. Исходные коды программ требуют применения принципиально иных техник. В отличие от обычных текстов исходные коды создаются в соответствии со строгими синтаксическими правилами и парадигмами программирования. В них зачастую используются известные реализации алгоритмов и шаблоны.

6. Количество образцов авторского текста и их длина могут оказывать существенное влияние на эффективность алгоритмов. Важно учитывать особенности, возникающие при работе с небольшим числом образцов или с образцами, имеющими ограниченную длину, и выбирать соответствующие методы.

7. Множество признаков, используемых для обучения алгоритмов, должно быть информативным. Для получения такого множества следует использовать методы фильтрации, направленные на отсев запутывающих или мало информативных признаков. Без использования фильтрации уменьшается статистическая надежность работы алгоритма, возрастает время его работы, а также увеличивается вычислительная сложность обучения.

Во **второй главе** предложены методология идентификации автора текстовой информации для решения задач кибербезопасности, и рассмотрены ее составляющие: модель создания автором текста в киберсреде, модели представления текста, алгоритмы разбора текста, методы принятия решений, методы оптимизации гиперпараметров, методы снижения размерности и отбора признаков, атаки на методы идентификации, алгоритмы сглаживания.

На рисунке 2 представлены этапы создания методологии, сама методология представлена на рисунке 3.

Под **методологией** будем понимать совокупность моделей, алгоритмов, методов и методик, предназначенную для достижения цели работы – идентификации автора текстовой информации, включая естественно-языковые тексты и исходные коды программ, для решения задач кибербезопасности.

Под **методикой** будем понимать систему, включающую определенные элементы методологии с указанием значений их параметров и последовательности применения этих элементов для решения практической задачи.

Методология предполагает использование методов статистического анализа, машинного и глубокого обучения (МО). Исходя из специфики данных, решаемой задачи кибербезопасности и потенциальных атак на метод, на каждом из этапов могут быть задействованы разные по принципу действия подходы. Традиционные методы машинного обучения обеспечивают высокую степень интерпретируемости и скорости, поэтому могут применяться как самостоятельно, так и в составе ансамблей методов принятия решений, однако являются менее эффективными в сравнении с НС при наличии осложняющих факторов. НС являются более устой-

чивыми к атакам на метод и более эффективными для поиска явных и неявных признаков авторского стиля.

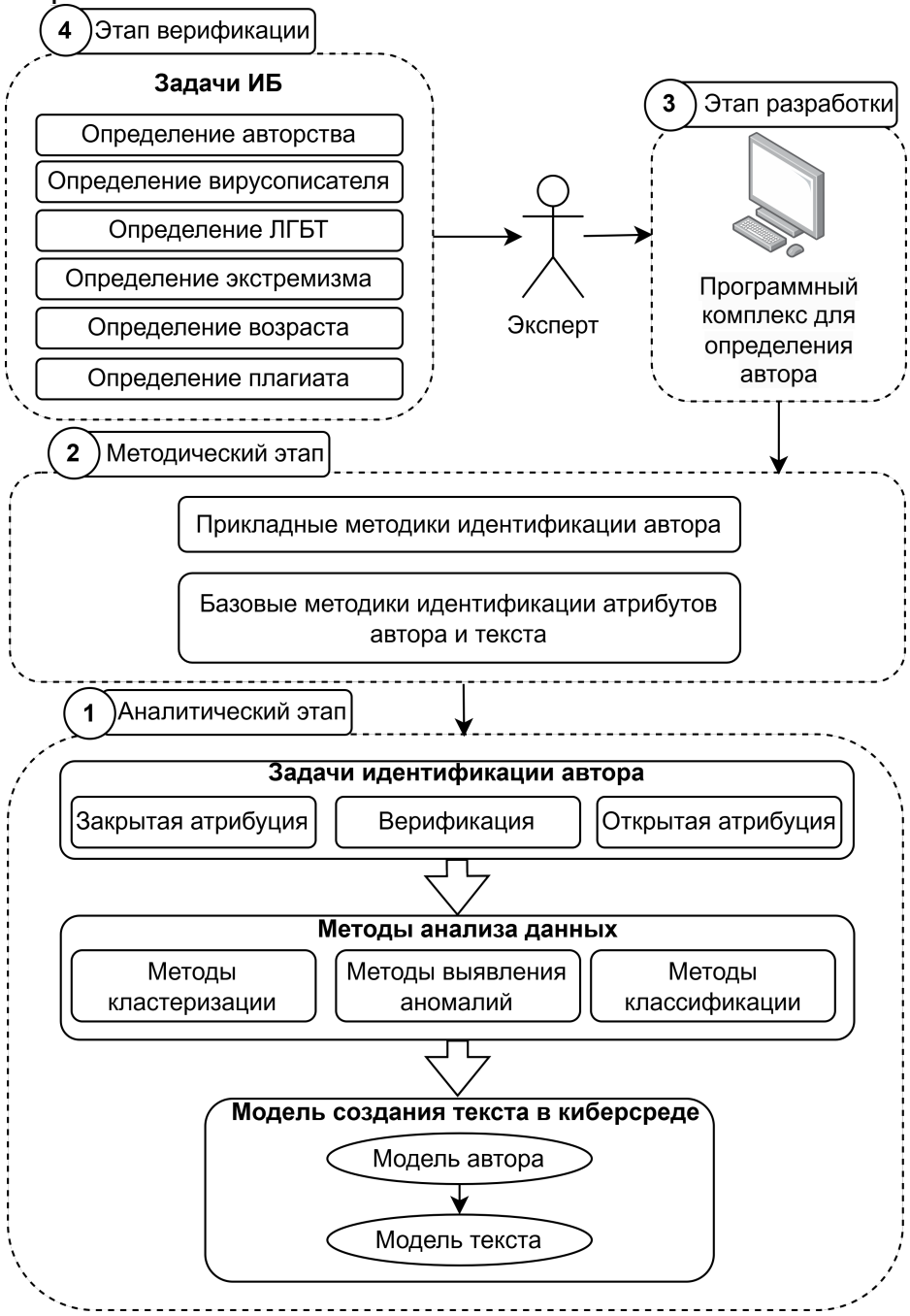


Рисунок 2 – Этапы создания методологии

Модель создания текста автором в киберсреде с учетом семантики представим тройкой:

$$M = (\mathbf{A}, \mathbf{T}, \mathbf{E}),$$

где **A** – множество авторов, **T** – множество текстов, **E** – множество сред.

Пусть имеется множество текстов $\mathbf{T} = \{t_1, ..., t_{|\mathbf{T}|}\}$ и множество авторов Ы. Введем бинарное отношение «текст написан автором» $R \subset \mathbf{T} \times \mathbf{A}$ на декартовом произведении множеств **T** и **A** такое, что выполняется *tRa* если текст $t \in \mathbf{T}$ написан автором $a \in \mathbf{A}$:

$$\exists t \in \mathbf{T}, \exists a \in \mathbf{A} : (tRa).$$

Автор может взаимодействовать или не взаимодействовать с другими людьми, оказывать на них влияние посредством написанных им текстов.

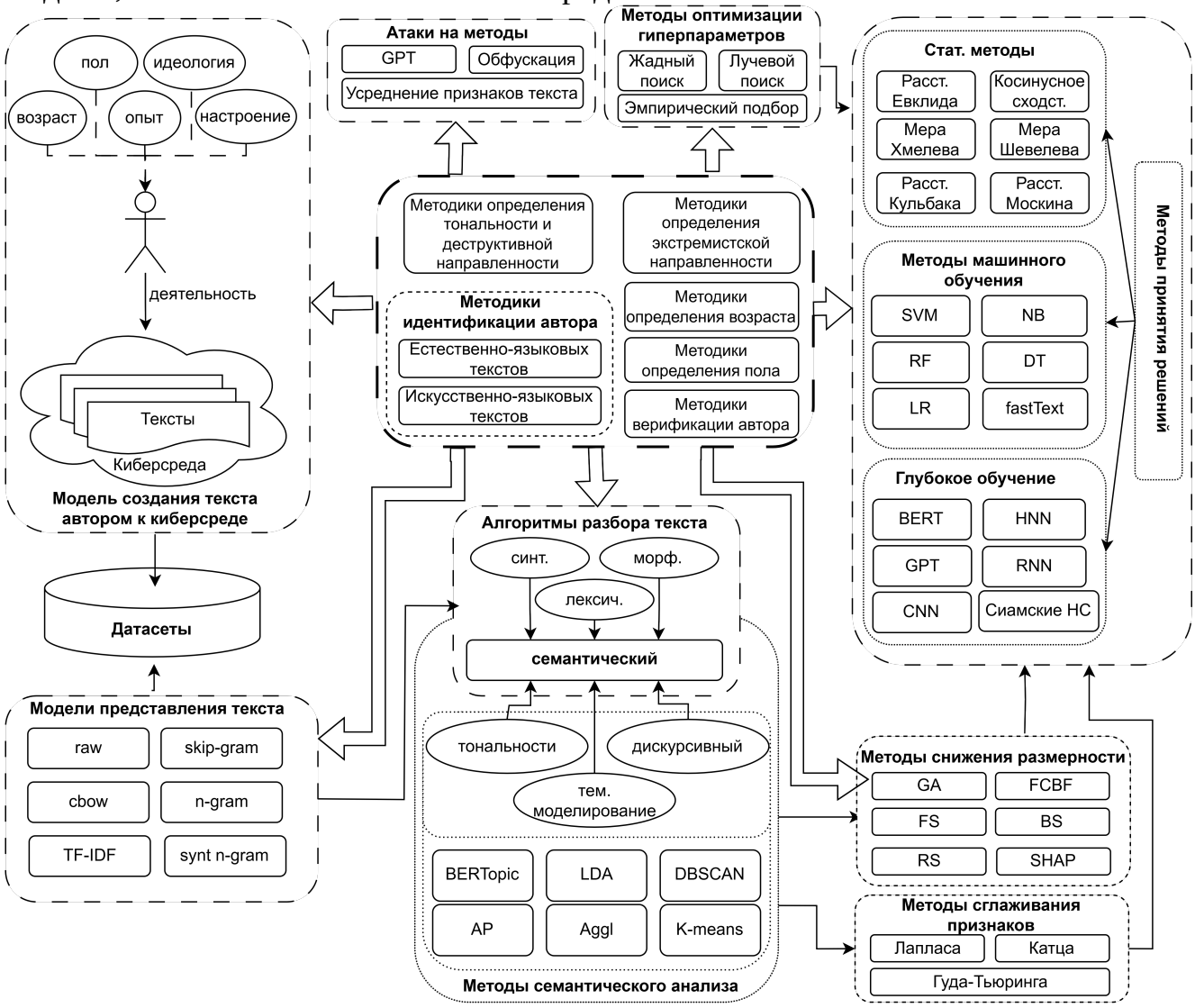


Рисунок 3 – Методология идентификации автора текстовой информации для решения задач кибербезопасности

В случае, когда текст t можно представить как объединение фрагментов $t = \bigcup_{i=1}^{N_t} t_i^{'}$, написанных несколькими авторами, будем говорить, что текст t «написан в соавторстве»:

$$\exists t_i^{'}, t_j^{'} \subseteq t, \exists a_l, a_m \in \mathbf{A}, a_l \neq a_m : (t_i^{'} R a_l) \wedge (t_j^{'} R a_m).$$

Случай, когда текст, написанный одним автором, подвергается изменению другим автором при сохранении общей семантики и тональности текста назовем «редактированием» и опишем в виде функции:

$$t^e = edit(t).$$

Текст, полученный в результате работы генеративного алгоритма, обученного на текстах $\mathbf{T}^a$ автора a, упрощенно опишем в виде:

$$t^g = gen(\mathbf{T}^a).$$

Обозначим факт вмешательства в процесс создания текста как:

$$inter \in \mathbf{INTER} = \left\{ inter_1, ..., inter_{|\mathbf{INTER}|} \right\},$$

где множеством активных действий по отношению к текстам является самостоятельное написание, редактирование, соавторство, применение генеративного алгоритма.

Каждый автор в момент создания текста представляется набором атрибутов:

$$\mathbf{C}^a = (id, age, gender, emo, status, action, pop) \in$$

$$\in \mathbf{ID} \times \mathbf{AGE} \times \mathbf{GENDER} \times \mathbf{EMOA} \times \mathbf{STATUSA} \times \mathbf{ACTION} \times \mathbf{POPA},$$

где $id \in \mathbf{ID}$ – идентификатор: любая последовательность символов, которой человек себя идентифицирует. В качестве идентификатора может выступать ФИО, псевдоним в социальных сетях или на портале сети Интернет и др.; $gender \in \mathbf{GENDER}$ – пол и гендерная идентичность: мужской, женский, представитель ЛГБТ*; $emo \in \mathbf{EMOA}$ – настроение, с которым автор писал текст: положительное, отрицательное, нейтральное. Возможна более детальная классификация, учитывающая оттенки агрессии и др.; $status \in \mathbf{STATUSA}$ – статус, показывающий имеет ли гражданин статус иностранного агента или экстремиста, или имеет отношение к организациям, имеющим эти статусы; $action \in \mathbf{ACTION}$ – деятельность автора по созданию текстовой информации. Можно выделить общение, творческую и профессиональную деятельность; $pop \in \mathbf{POPA}$ – категории популярности автора; $age \in \mathbf{AGE}$ – возраст автора.

Например, для решения практических задач кибербезопасности могут иметь значение возрастные интервалы согласно ст. 12 закона № 436-ФЗ от 29.12.2010 «О защите детей от информации, причиняющей вред их здоровью и развитию»: «0+», «6+», «12+», «16+», «18+» и др.

Рассмотрим текст с точки зрения деления на классы, к которым его можно отнести, с учетом контекста кибербезопасности.

Текст имеет семантическое описание $topic \in \mathbf{TOPICT}$. В свою очередь каждую тему можно представить списком ключевых слов:

$$topic_i = \left[keyword_1, ..., keyword_{|topic_i|} \right].$$

В контексте решения задач кибербезопасности в дальнейшем будем использовать абстрактное понятие «тип текста», которое учитывает вид, стиль, жанр, назначение текста и др. $type \in \mathbf{TYPET}$. Например, в качестве типов в зависимости от задачи будем понимать художественные, любительские и сообщения из социальных сетей; естественно-языковые и искусственно-языковые и т.д.

Текст может быть отнесен к экстремистским материалам или материалам иностранных агентов, доступ к которым запрещен на территории Российской Федерации. Действующие статусы можно представить как $status \in \mathbf{STATUST}$.

Каждый текст имеет эмоциональный окрас (тональность) $emo \in \mathbf{EMOT}$. Эмоциональный окрас в простом случае может принимать значения: положительный, негативный, нейтральный.

Текст может получить популярность у целевой аудитории. Категории популярности можно представить как $pop \in \mathbf{POPT}$.

---

* Запрещенная в Российской Федерации экстремистская организация.

Таким образом множество классов, к которым можно отнести текст, можно описать как декартово произведение вышеозначенных множеств, а конкретный класс, к которому относится текст, представить как:

$$\mathbf{C}^t = (type, status, topic, emo, pop, inter) \in$$

$$\in \mathbf{TYPET} \times \mathbf{STATUST} \times \mathbf{TOPICT} \times \mathbf{EMOT} \times \mathbf{POPT} \times \mathbf{INTER}.$$

Каждый элемент текста описывается вектором признаков, отражающим его свойства. Набор признаков текста можно представить как результат работы функции извлечения полного вектора характеристик из текста:

$$\mathbf{F} = extract(t) = \left[f_1, ..., f_{|\mathbf{F}|} \right],$$

где $f_i \in \mathbf{LEX} \cup \mathbf{MORPH} \cup \mathbf{SYNT} \cup \mathbf{SEM} \cup \mathbf{IDIO} \cup \mathbf{META} \cup \mathbf{EMB}$, **LEX** – лексические признаки, **MORPH** – морфологические признаки, **SYNT** – синтаксические признаки, **SEM** – семантические признаки, **IDIO** – идиосинкразические признаки, **META** – метаданные текста, **EMB** – некоторое векторное представление текста, полученное с помощью модели машинного обучения.

Множество информативных характеристик текста $\mathbf{F}^{inf}$ и их значения зависят от типа текста и атрибутов автора, рассматриваемых в рамках интересующей задачи идентификации. Получение информативных признаков можно представить в виде функции, принимающей на вход признаки, тип текста и атрибуты автора:

$$\mathbf{F}^{inf} = inf(\mathbf{F}, type, \mathbf{C}^a) = \left[f_1^{inf}, ..., f_{|\mathbf{F}^{inf}|}^{inf} \right].$$

Текстами и совокупностью векторов признаков текстов, написанных авторами, все или некоторые атрибуты которых совпадают, можно представить стиль определенной категории авторов.

$$\forall a_l, a_m \in \mathbf{A}, \mathbf{K}^{a_l} = \mathbf{K}^{a_m} = \mathbf{K}, \mathbf{K} \subseteq \mathbf{C}^{\mathbf{A}}:$$

$$\mathbf{C}^{\mathbf{K}} = \{\mathbf{F}^{inf_i^{\mathbf{K}}}\}_{i=\overline{1}}^{N_{\mathbf{T}^{\mathbf{K}}}} = \begin{Bmatrix} f^{inf_{1,1}^{\mathbf{K}}}, & ... & f^{inf_{1,|\mathbf{F}^{inf_1^{\mathbf{K}}}|}^{\mathbf{K}}} \\ ... & ... & ... \\ f^{inf_{N_{\mathbf{T}^{\mathbf{K}}},1}^{\mathbf{K}}} & ... & f^{inf_{N_{\mathbf{T}^{\mathbf{K}}},|\mathbf{F}^{inf_{N_{\mathbf{T}^{\mathbf{K}}}}^{\mathbf{K}}}|}^{\mathbf{K}}} \end{Bmatrix},$$

где **K** – подмножество совпадающих атрибутов авторов, $\mathbf{T}^{\mathbf{K}}$ – множество текстов, написанных авторами с одинаковыми атрибутами, а $N_{\mathbf{T}^{\mathbf{K}}}$ – количество текстов в этом множестве, $\mathbf{F}^{inf_i^{\mathbf{K}}}$ – множество признаков и эмбеддингов, информативных для определенной категории авторов, имеющих одинаковые атрибуты.

Стиль автора может меняться со временем, и значения характеристик **F** могут изменяться. Однако множество информативных признаков $\mathbf{F}^{inf}$ должно быть устойчивым к этим изменениям во времени и учитывать небольшое редактирование другими авторами.

Среда, в которой автор пишет и публикует текст, описывается атрибутами:

$$\mathbf{C}^e = (topic, type, status, rules, pop) \in$$

$$\in \mathbf{TOPICE} \times \mathbf{TYPEE} \times \mathbf{STATUSE} \times \mathbf{RULES} \times \mathbf{POPE},$$

где $topic \in \mathbf{TOPICE}$ – семантическое описание среды, т.е. список тематик, тексты, относящиеся к которым, размещаются в среде; $type \in \mathbf{TYPEE}$ – тип среды. Например, мессенджер, хостинг IT-проектов, Интернет-СМИ, сайт научного журнала, и др.; $status \in \mathbf{STATUSE}$ – действующий статус ресурса (разрешен или

запрещен) в «Едином реестре доменных имен, указателей страниц сайтов в информационно-телекоммуникационной сети Интернет и сетевых адресов, позволяющих идентифицировать сайты в информационно-телекоммуникационной сети Интернет, содержащие информацию, распространение которой в Российской Федерации запрещено»; $rule \in \mathbf{RULES}$ – правила размещения текста в среде и/или стандарты среды и необходимо ли им следовать. Например, к ним можно отнести соблюдение законодательства РФ, правил публикации статей в научном журнале, стандартов кодирования для исходных текстов программ на хостинге IT-проектов в компании, запрет публикации текстов, не соответствующих определенной возрастной категории и др.; $pop \in \mathbf{POPE}$ – популярность среды. Например, для группы в социальной сети или мессенджере – количество реальных подписчиков; для сайта научного журнала – его импакт-фактор и т.д.

Введем бинарное отношение «текст создается в среде» $Q \subset \mathbf{T} \times \mathbf{E}$ на декартовом произведении множеств **T** и **E** такое, что выполняется *tQe* если текст $t \in \mathbf{T}$ создается и размещается в среде $e \in \mathbf{E}$:

$$\exists t \in \mathbf{T} \exists e \in \mathbf{E} : (tQe).$$

Решение задач кибербезопасности (рисунок 4) сводится к отнесению текста *t* к классу $c \in \mathbf{C} = \mathbf{C}^t \cup \mathbf{C}^{\mathbf{K}}$ с учетом ограничений и особенностей среды $\mathbf{C}^e$ и авторских признаков $\mathbf{C}^a$ на основе текстов, класс для которых известен $\mathbf{T}' = \{t_1, ..., t_m\} \subseteq \mathbf{T}$, т.е. существует множество пар «текст-класс» $\mathbf{D} = \{(t_i, c_j)\}_{i=1}^{m}$. При этом каждый текст рассматривается как вектор признаков **F**. Целью является построение классификатора, решающего данную задачу, т.е. нахождение некоторой целевой функции $\Phi : \mathbf{T} \times \mathbf{C} \to [0,1]$. Значения функции интерпретируется как степень принадлежности объекта классу: 1 соответствует полностью положительному решению, 0 – отрицательному.

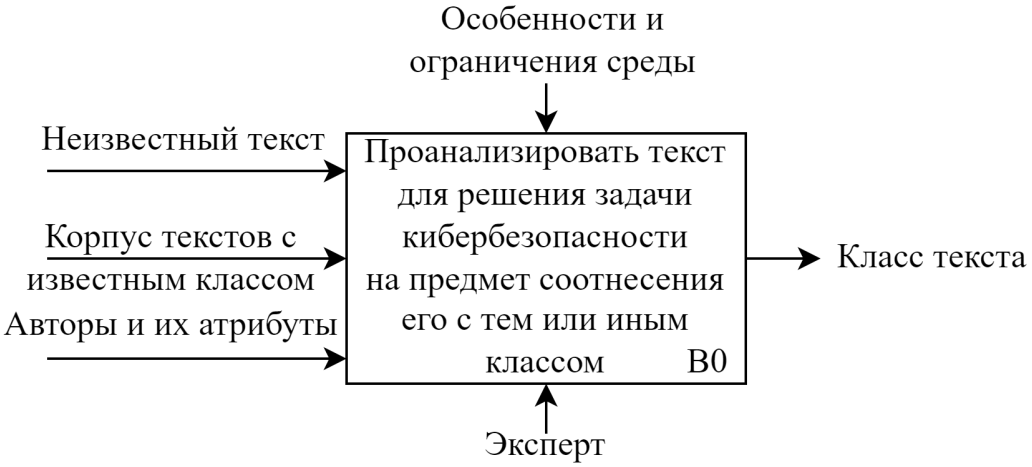


Рисунок 4 – IDEF0 диаграмма процесса анализа текста
для решения задач кибербезопасности

В **третьей главе** предложена методика идентификации автора естественно-языкового текста.

Идентификацию автора текста рассматривают в двух случаях: закрытое и открытое множество авторов. Различие заключается в том, что в случае открытого множества кандидатов истинного автора может вовсе не быть в списке кандидатов, в то время как в случае закрытого множества он обязательно присутствует.

При математической постановке задачи **закрытой атрибуции** процесс установления авторства текста задан тремя конечными множествами: $\mathbf{A} = \{a_1, a_2, ..., a_n\}$, $\mathbf{T}' = \{t_1, t_2, ..., t_m\}$, $\mathbf{T} = \{t_1, t_2, ..., t_s\}$ – авторов, текстов с известным авторством и анонимных текстов, соответственно. Каждому тексту, включая анонимные, соответствует вектор признаков.

Решение задачи сводится к вычислению целевой функции:

$$f(t_i) = [\, p(a_1), p(a_2), ..., p(a_n)\,],$$

где t_i – произвольный текст, $p(a_1), p(a_2), ..., p(a_n)$ – распределение вероятностей принадлежности текста авторам на основе вычисленных признаков текста.

Конечным ответом в вопросе разрешения авторства по анонимному тексту является выбор автора с максимальным значением вероятности $max(p_i)$.

Верификация авторства позволяет подтвердить или опровергнуть предположение о том, что тексты написаны одним и тем же автором. Верификация предполагает использование только образцов $\mathbf{T}_a$ целевого класса a. Решение задачи сводится к построению бинарного классификатора, определяющего относится ли образец *t* к классу a или отрицательному классу $\bar{a}$:

$$verification(t, a) \rightarrow \{0,1\},$$

где 0 – соответствует классу $\bar{a}$, 1 – соответствует классу a.

В случае возможного отсутствия истинного автора в списке претендентов, решается задача **открытой атрибуции**. Учитывается дополнительный нулевой класс a_{null}. Для того, чтобы его учесть можно использовать несколько подходов:

1) включить в образцы класса тексты как можно большего числа авторов $\mathbf{T}_{a_{null}} = \{t_1, ..., t_{a_{null}}\}$. Но, должны учитываться ограничения $\mathbf{T}_{a_{null}} \cap \mathbf{T}' = \varnothing$, $\left|\mathbf{T}_{a_{null}}\right| \approx \underset{i}{avg}\left(\left|\mathbf{T}_{a_i}\right|\right)$, $a_i \in A$, чтобы нулевой класс не включал образцы известных классов и обучающая выборка была сбалансирована по количеству образцов. В случае ограниченного количества образцов каждого автора этот подход не применим, потому что количество образцов нулевого класса превышает обучающее $\left|\mathbf{T}_{a_{null}}\right| >> \left|\mathbf{T}'\right|$;

2) использовать *n* одноклассовых классификаторов, решающих задачу верификации для каждого потенциального автора. Решение отнести текст к нулевому классу формируется на основе *n* отрицательных решений точных одноклассовых классификаторов:

$$\sum_{j}^{n} verification(t, a_j) = 0 \Rightarrow t \in \mathbf{T}_{a_{null}},$$

иначе решение о принадлежности образца известному классу принимается классификатором, обученным на известных текстах потенциальных авторов.

Методика для случая закрытой атрибуции представлена на рисунке 5.

При разработке методики рассматривались варианты использования как классических методов машинного обучения (случайный лес (RF), SVM, логистическая регрессия (LR), метод *k*-ближайших соседей (KNN), наивный байесовский классификатор (NB)), так и нейросетевых методов (fastText, RuBERT, MultiBERT,

CNN, НС долгой краткосрочной памяти (LSTM), BiLSTM, управляемые рекуррентные блоки (GRU), BiGRU).

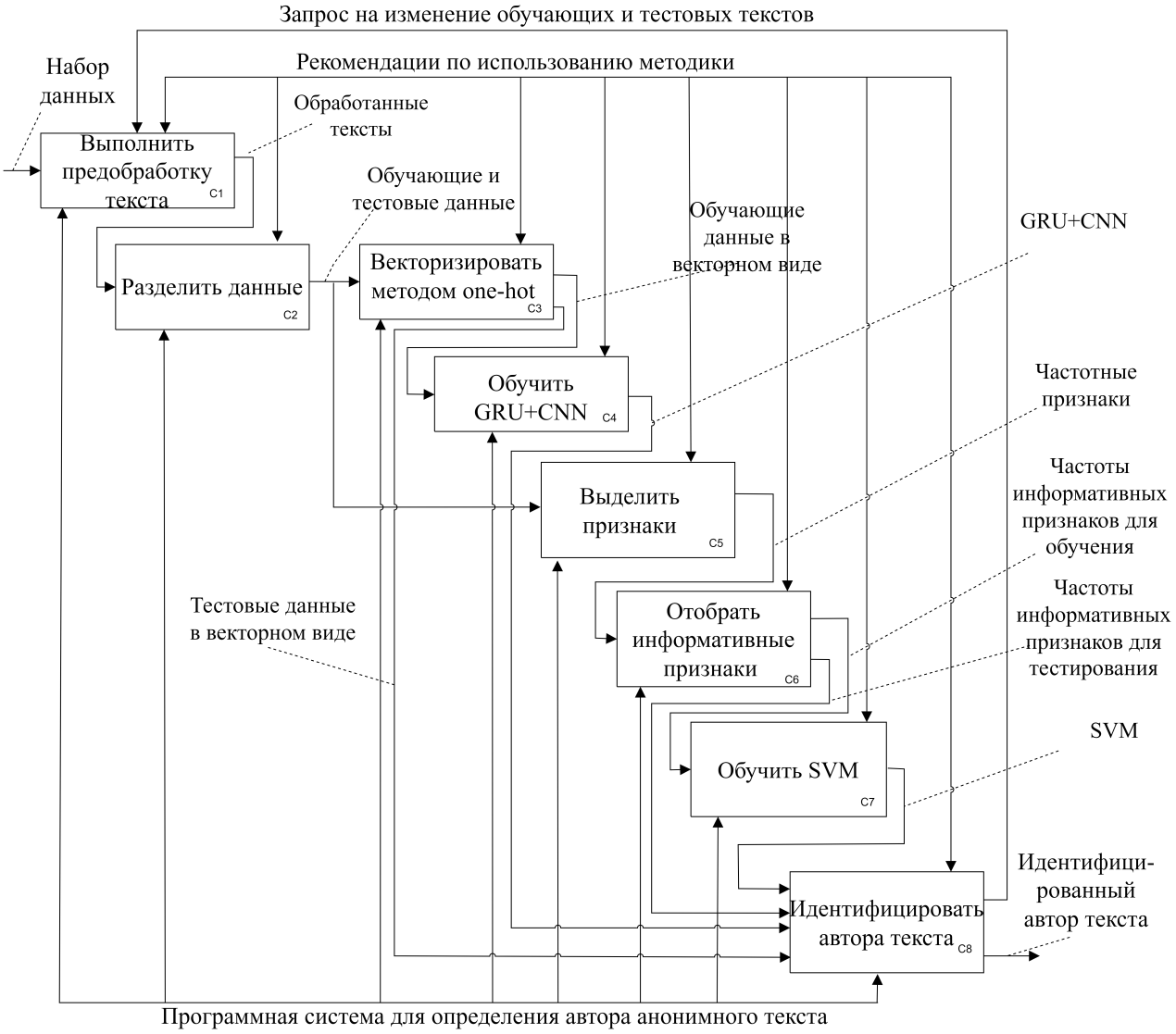


Рисунок 5 – Методика определения автора текста (закрытая атрибуция)

Дополнительно используются многослойные гибридные архитектуры НС на основе совместного использования рекуррентных и сверточных слоев с разными размерами фильтров. В данном исследовании рассмотрены комбинации LSTM+CNN, CNN+LSTM, GRU+CNN, CNN+GRU, BiGRU+CNN, CNN+BiGRU.

Для отбора информативных признаков применялись генетические алгоритмы (ГА), алгоритм на основе регуляризации (RS), прямой (FS) и обратный последовательный отбор (BS), аддитивное объяснение Шепли (SHAP).

Все результаты получены в рамках 10-фолдовой кросс-валидации. Для обучения используется 3 образца по 15000 символов в случае художественных и любительских текстов и 40 текстов длиной до 250 символов для комментариев пользователей социальных сетей. При тестировании использовался 1 текст и 10 сообщений соответственно.

Результаты исследования для случая закрытого множества приведены в таблице 1.

Лучшие результаты для художественных текстов были продемонстрированы моделями SVM+RS, GRU+CNN, fastText и CNN. Для перечисленных моделей были проведены тесты, направленные на оценку статистической значимости результатов (рисунок 6).

Таблица 1 – Результаты экспериментов по определению автора текста (закрытая атрибуция)

Метод	Количество авторов					Время (50 авторов), с.
	2	5	10	20	50	
Точность (%), художественные текст, закрытая атрибуция						
fastText	98,2±4,5	95,0±3,7	90,2±6,3	69,9±4,3	45,8±6,2	2372
SVM	95,4±1,7	94,6±2,1	81,9±4,6	63,3±4,8	37,7±6,3	118
SVM+RS	98,6±1,4	98,1±1,5	94,2±1,8	88,3±2,5	56,9±2,2	34264
SVM+ГА	98,8±0,8	98,2±1,6	92,3±2,0	80,8±2,2	56,6±3,2	31553
CNN	97,1±3,6	95,9±2,8	81,3±5,9	71,2±5,8	50,1±5,1	7319
GRU+CNN	97,6±1,4	95,9±5,0	88,5±4,2	70,1±4,0	56,9±4,2	8028
BiGRU+CNN	85,3±2,9	82,9±4,6	66,9±5,0	58,8±4,5	41,5±3,8	8071
CNN+BiGRU	95,4±5,8	87,6±2,7	75,2±3,4	62,7±4,1	46,5±3,7	7994
Точность (%), любительские тексты, закрытая атрибуция, кросс-тематический случай						
fastText	99,0±1,9	95,2±4,0	91,0±4,1	74,8±3,4	57,5±5,0	2372
GRU+CNN	96,6±2,9	95,7±6,9	92,6±5,8	74,2±4,8	61,8±6,1	8028
BiGRU+CNN	98,9±1,7	92,3±4,8	91,8±3,0	65,8±4,7	50,6±4,2	8071
CNN+BiGRU	96,1±2,8	95,0±2,1	83,3±3,8	69,1±3,5	56,9±6,1	7994
SVM	89,9±3,4	86,9±4,6	83,0±3,3	69,5±3,2	52,1±5,5	118
SVM+RS	98,9±1,2	96,3±1,8	90,1±2,1	74,7±3,3	61,6±2,8	73904
SVM+ГА	98,6±0,4	96,7±1,3	91,9±1,8	73,1±2,4	62,4±3,5	69875
Точность (%), любительские тексты, закрытая атрибуция, монотематический случай						
fastText	96,6±1,6	91,5±2,0	75,9±3,5	64,6±2,4	46,7±2,9	2280
GRU+CNN	96,8±1,8	93,9±3,1	78,4±3,9	68,2±3,0	51,3±3,4	8004
BiGRU+CNN	90,0±4,5	79,4±5,2	66,6±3,6	55,3±4,2	35,0±4,2	8142
CNN+BiGRU	97,9±3,0	86,6±2,9	75,8±2,0	66,9±2,9	45,9±4,3	8012
SVM	93,6±2,2	81,1±1,8	72,7±2,9	62,6±1,4	39,8±4,0	109
SVM+RS	93,1±1,2	90,5±0,9	77,6±1,7	61,5±3,4	46,4±2,1	62288
SVM+ГА	93,2±0,9	90,0±1,6	83,5±1,9	67,3±2,4	48,1±3,3	60136
Точность (%), короткие тексты, закрытая атрибуция						
GRU+CNN	96,3±5,5	93,5±6,2	78,1±5,4	73,7±2,5	56,0±5,5	3920
BiGRU+CNN	96,0±3,9	92,0±3,7	73,9±2,9	71,6±7,0	53,1±4,2	4157
CNN+BiGRU	96,0±3,5	93,4±5,9	78,0±4,8	66,7±4,9	47,5±3,8	3822
fastText	94,0±1,2	87,2±2,2	76,1±3,5	68,4±2,3	55,6±2,8	1017
SVM	82,2±4,0	69,9±3,5	66,3±3,8	55,3±3,1	32,1±3,9	79
SVM+RS	89,5±1,3	88,3±2,0	78,0±3,1	71,8±2,2	47,9±2,6	96284
SVM+ГА	91,3±0,6	88,5±0,3	78,2±2,0	70,8±2,5	50,2±2,0	98753

Для всех рассмотренных случаев *p*-значение в рамках теста Фридмана не превысило 0,05, следовательно, полученная разница является существенной. Согласно диаграмме Демшара, для 2, 5, 10 и 20 авторов SVM в сочетании с RS по-

казывает лучший результат. Решающую роль в данном случае сыграл выбор SVM с линейным ядром, которое сочетается с основным ограничением RS – применимостью исключительно к линейным моделям. Для 50 авторов статистическая разница между SVM+RS и GRU+CNN является незначительной, однако лучший результат демонстрирует комбинация из глубоких НС. Отметим также, что метод RS требует в среднем 3 раза больше времени на отбор признаков, чем обучение моделей НС.

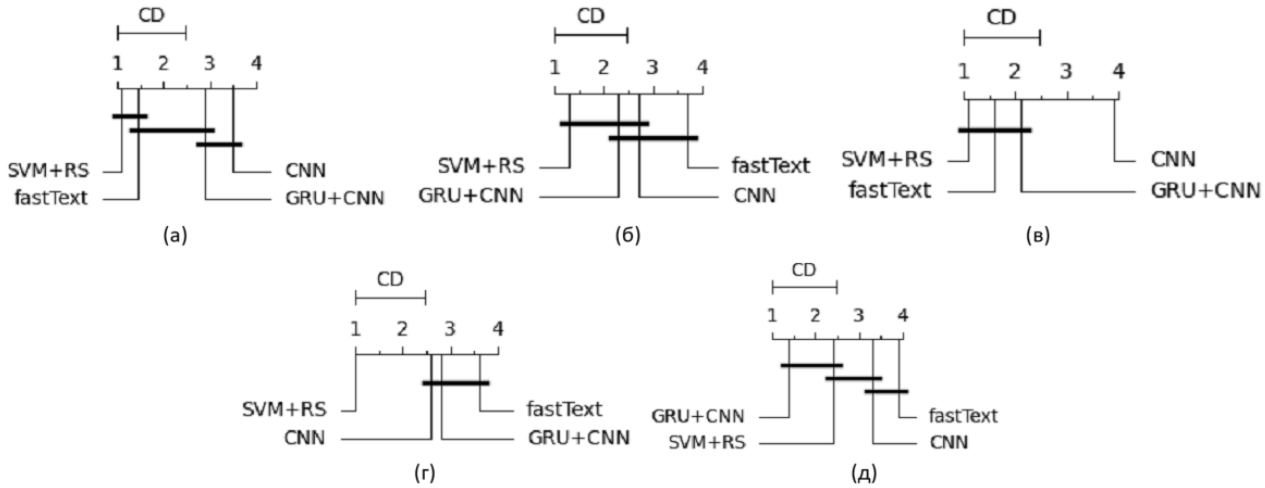


Рисунок 6 – Диаграмма Демшара для 2 (а), 5 (б), 10 (в), 20 (г) и 50 (д) авторов

Для любительских текстов рассматривается два случая:

1) кросс-тематический – тексты созданы на основе разных произведений.
2) монотематический – тексты созданы на основе одного произведения.

Среди нейросетевых подходов отмечены fastText и GRU+CNN. Классические методы с отбором не уступают нейросетевым моделям, однако подбор параметров с их помощью занимает на порядок больше времени, чем обучение глубоких моделей. Однако однозначного победителя в данном эксперименте выявлено не было. Для выбора оптимального варианта из всех альтернатив рассмотрены критерии принятия решений в условиях риска. По критериям Байеса-Лапласа, Вальда, Сэвиджа можно сделать вывод, что GRU+CNN является оптимальным выбором.

При той же постановке эксперимента, результаты определения автора в кросс-тематическом случае на 2–14% выше, чем в случае монотематической атрибуции, ввиду сложности последней. Все модели демонстрируют точность 90% и более при классификации двух авторов. Лучшие результаты во всех случаях, кроме бинарного, продемонстрировала модель GRU+CNN. Это подтверждает правильность ее выбора, как оптимальной альтернативы по критериям принятия решений в условиях риска.

В случае коротких текстов было установлено, что применение методов отбора улучшает результаты на 7–16%. Однако временные затраты на отбор затрудняют применение подходов при решении реальных практических задач, т.к. для каждого нового случая требуется проводить отбор заново. Использование НС позволяет добиться сопоставимой или большей точности, при этом временные затраты на порядок меньше. Для всех рассматриваемых случаев лучший результат был достигнут сочетанием архитектур GRU+CNN. Точность такой комбинации до-

стигает 96,3%. Сопоставимые метрики качества демонстрируют модели BiGRU+CNN, CNN+BiGRU и fastText.

Для оценки методики в случае текстов, созданных новейшими генеративными моделями, был проведен эксперимент на смешанном наборе данных, включающем оригинальные тексты авторов, а также сгенерированные на их основе образцы авторов-моделей. Результаты приведены в таблице 2.

Во всех случаях GRU+CNN показывает лучшие результаты классификации. Оригинальные тексты автора можно отличить от сгенерированных образцов с точностью 91,3% и 93,5% для версий 4 и 3.5 соответственно. Максимальная потеря точности для GRU+CNN составила 19%, а для SVM – 17%.

Таблица 2 – Результаты экспериментов по идентификации автора в случае использования генеративных алгоритмов

Количество авторов (классов)	Точность SVM+RS, %		Точность GRU+CNN, %	
	Оригинальные тексты + GPT-3.5	Оригинальные тексты+GPT-4	Оригинальные тексты + GPT-3.5	Оригинальные тексты + GPT-4
1 (2)	92,6±4,1	85,2±4,0	93,5±3,2	91,3±4,1
2 (4)	80,1±3,9	72,5 ±3,4	90,0±1,2	89,9 ±3,3
5 (10)	65,9±3,5	63,1 ±5,1	84,2±2,2	82,1 ±5,0
10 (20)	60,3±3,8	54,4±4,2	64,1±3,5	60,7±3,6
20 (40)	48,3±3,1	41,6±5,0	54,4±2,3	47,8±3,2
50 (100)	30,1±3,9	23,2±6,2	39,6±2,8	34,3±4,1

Для реализации **методики верификации авторства** предлагается использовать One-Class SVM, обученный на текстах автора.

При апробации метода (таблица 3) на вход каждой обученной модели подавались тексты как самого автора, так и образцы других авторов. Так, к тестовой выборке были добавлены образцы 1, 4, 9, 19 и 49 авторов. Для тестирования использовался 1 текст каждого автора для длинных и 10 для коротких текстов.

Таблица 3 – Результаты экспериментов по верификации автора текста

Количество авторов при тестировании	Полнота, %		
	Художественные	Любительские	Короткие
2	99,2±0,3	98,9±1,5	97,8±1,1
5	96,1±0,9	98,7±1,1	97,4±2,0
10	95,2±1,3	97,4±1,9	96,3±3,0
20	93,1±3,1	90,2±4,1	96,2±4,9
50	86,8±4,4	90,7±4,1	95,6±7,0
100	80,1±6,5	82,7±6,1	88,2±7,0

Полнота верификации автора короткого текста достигает 97,8%, что позволяет сделать выводы о том, что разработанный подход может применяться для продленной аутентификации пользователя на основе создаваемых им текстов.

Методика открытой атрибуции, представленная в виде IDEF0 диаграммы, изображена на рисунке 7.

В основе методики лежит комбинация одноклассового SVM и сочетания GRU+CNN, которая работает следующим образом:

1. Верификация авторства. Обучение отдельных One-Class SVM на текстах каждого известного автора. Совокупность обученных One-Class SVM используется как инструмент для верификации авторства и детектирования аномалий. При тестировании неизвестного образца, он подается на вход каждого из классификаторов. Если все одноклассовые классификаторы отнесут его к отрицательному классу, то принимается итоговое решение, что текст относится к null-классу.

2. По итогам работы всех One-Class SVM, все верифицированные образцы подаются на вход классификатора GRU+CNN, обученного только на текстах известных классов. Эта модель идентифицирует конкретного автора.

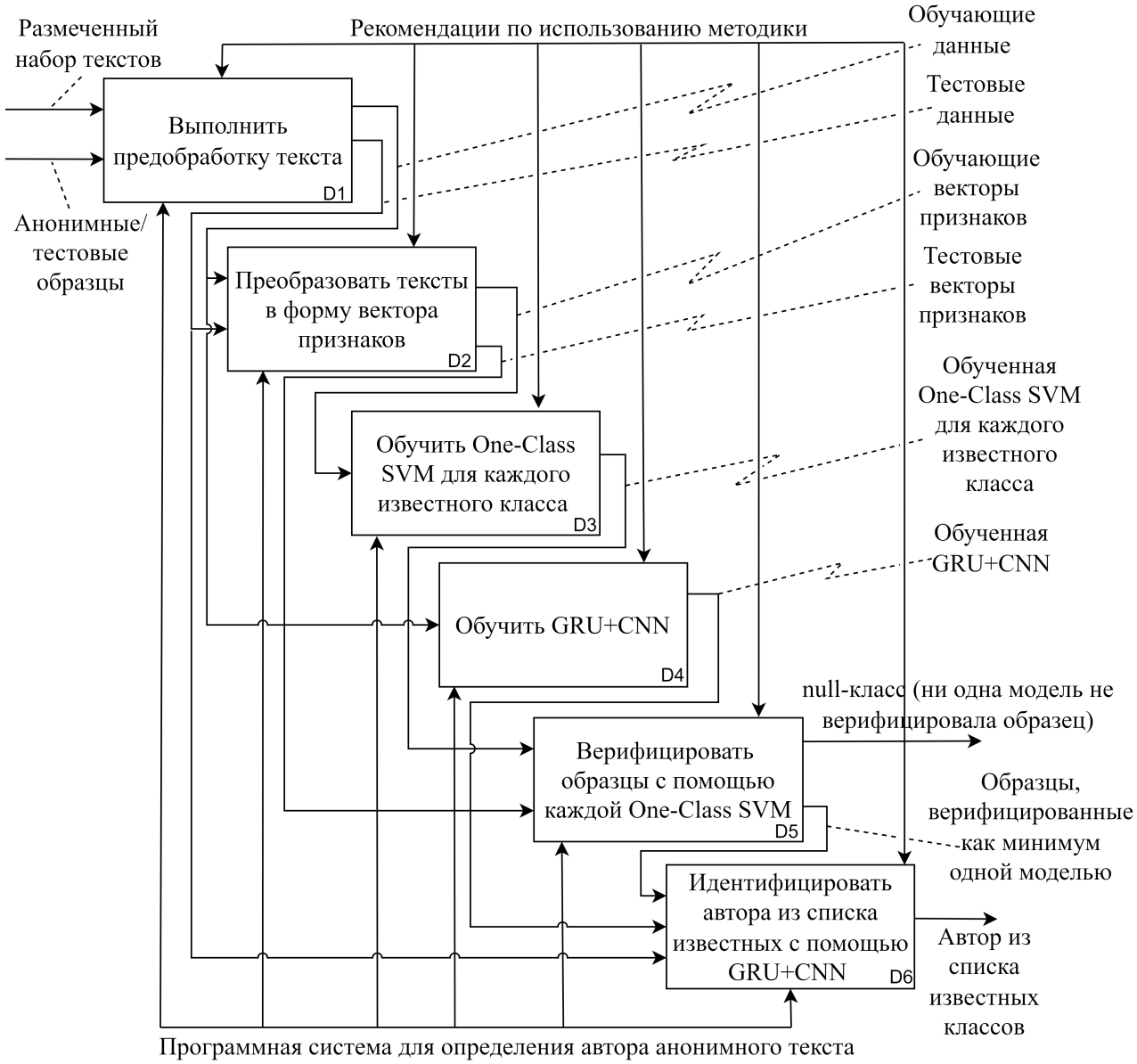


Рисунок 7 – Методика определения автора текста (открытая атрибуция)

Для открытой атрибуции обучение происходило на текстах N авторов, а при тестировании использовались тексты N+M авторов, где $M = \lceil N \times 0{,}3 \rceil = \min\{n \in \mathbb{Z} \mid n \geq (N \times 0{,}3)\}$ – количество авторов, тексты которых используются как

нулевой класс. Эксперименты проводились для 2+null, 4+null, 9+null, 19+null и 49+null авторов, где второе слагаемое означает нулевой класс.

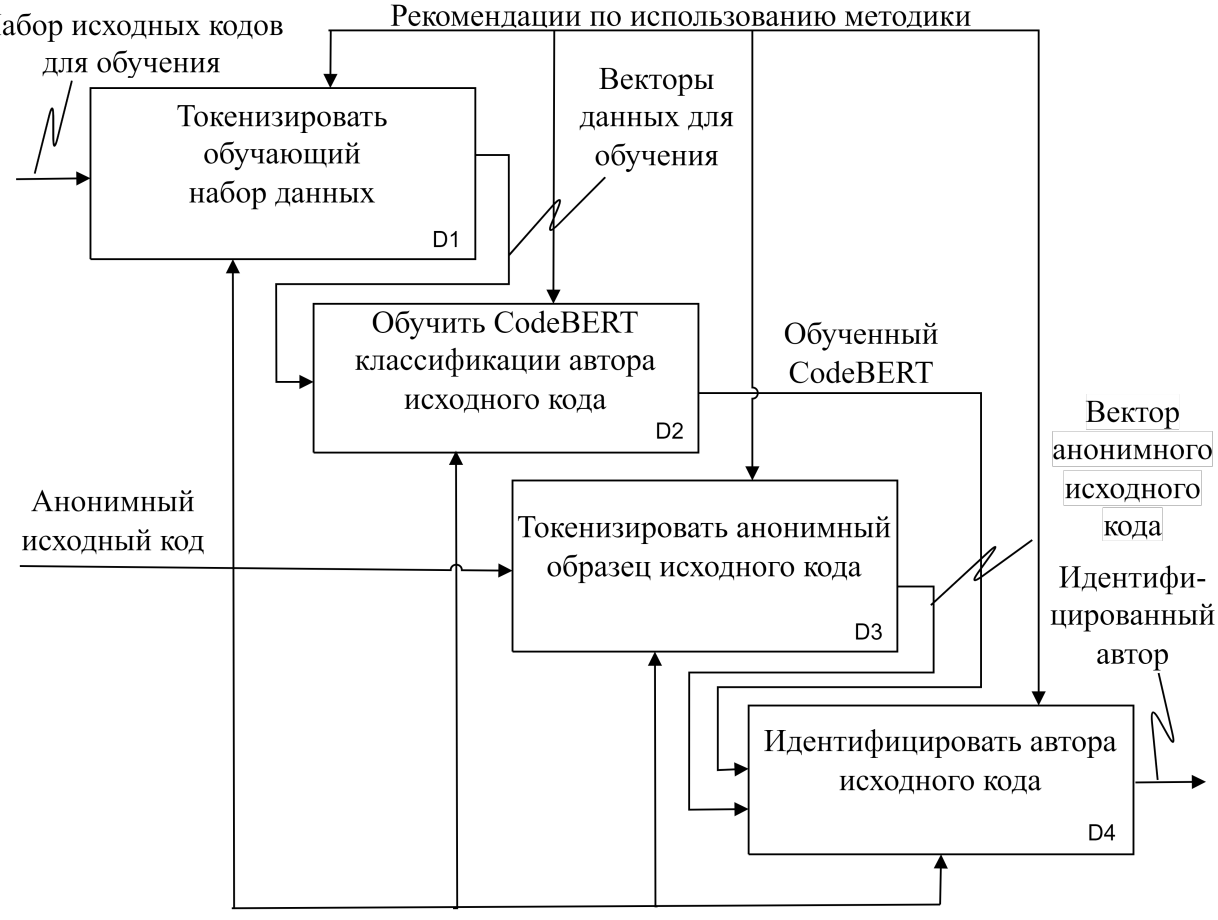
Результаты исследования методики приведены в таблице 4.

В качестве альтернатив для открытой атрибуции рассматривались методы на основе расстояния Евклида, косинусного сходства и сочетание GRU+CNN с пороговой функцией Softmax – результаты в этих случаях были ниже.

Таблица 4 – Результаты определения автора текста (открытая атрибуция)

Тексты	Количество авторов				
	2+null	4+null	9+null	19+null	49+null
	Точность, %				
Художественные	94,2±2,1	89,2±1,9	69,4±2,2	53,8±2,4	40,8±3,4
Любительские (кросс-тем.)	94,5±1,8	93,5±2,6	83,1±3	70,3±2,8	45,4±3,1
Любительские (монотем.)	98,2±1,1	91,8±1,2	82,2±1,3	70,9±1,3	43,1±2,1
Короткие сообщения	93,5±1,5	86,9±2,2	81,2±3,4	69,8±3,3	51,9±3,8

В **четвертой главе** предложена **методика идентификации автора исходного кода программы** (рисунок 8). Методика обеспечивает эффективную идентификацию автора-программиста в простых и сложных случаях. Простые случаи предполагают идентификацию автора на основе исходных кодов программ на одном из 13 языков программирования. Сложные случаи подразумевают факт использования обфускации, стандартов кодирования, командную разработку, владение двумя и более языками, а также искусственную генерацию кода.



Программная система для идентификации автора программного кода

Рисунок 8 – Обобщенная методика идентификации автора исходного кода

В качестве основы методики рассматривались такие подходы как SVM, обученный на множестве информативных признаков, fastText с экспериментально подобранными параметрами, а также мультиязыковой BERT и CodeBERT, модифицированный слоями регуляризации и классификации.

В рамках первого эксперимента (рисунок 9) рассматривался случай идентификации автора исходного кода среди 10 авторов-кандидатов. Качество оценивалось метрикой точности, вычисляемой в рамках 10-фолдовой кросс-валидации. Согласно полученным результатам, лучшая точность для всех языков программирования была достигнута моделью CodeBERT. Классификаторы BERT, fastText и SVM уступили в точности до 5%. CodeBERT позволил добиться точности 93% и был выбран в качестве основы для методики. В среднем методика достигает точности 93% для 5, 88% для 10 и 85% для 20 авторов в простых случаях.

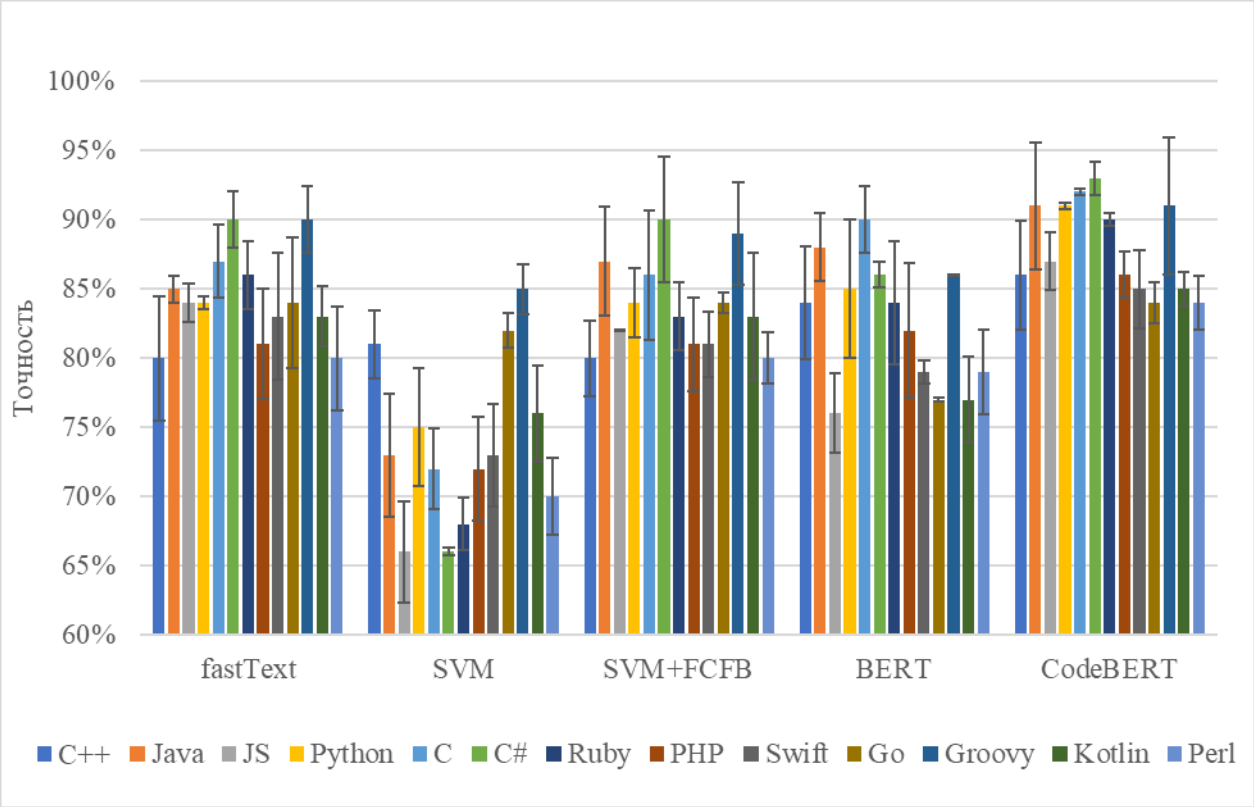


Рисунок 9 – Результаты экспериментов для разных моделей

Методика, основанная на CodeBERT, была протестирована для всех сложных случаев. Процесс проведения экспериментов представлен на рисунке 10. Образцы кодов из обучающего и тестового наборов соответствуют следующим критериям: длина кода не менее 1000 символов, отбирается не более 30 образцов исходного кода от одного автора, выбор образца случайный.

Установлено, что следование стандартам кодирования оказывает негативное влияние на процесс идентификации. Однако при наличии достаточного объема обучающих данных (30 файлов по 6000 символов) методика позволяет добиться приемлемой для решения прикладных задач точности в 80%.

Для случая обфускации проведены эксперименты с интерпретируемыми и компилируемыми языками. Простая лексическая обфускация не оказывает существенного влияния на эффективность методики. Более сложные модификации исходного кода, подразумевающие изменение байт-кода объектов, внедрение псевдокода и т.п. приводят к существенным потерям в точности, достигающим 30% при недостаточном числе обучающих образцов.

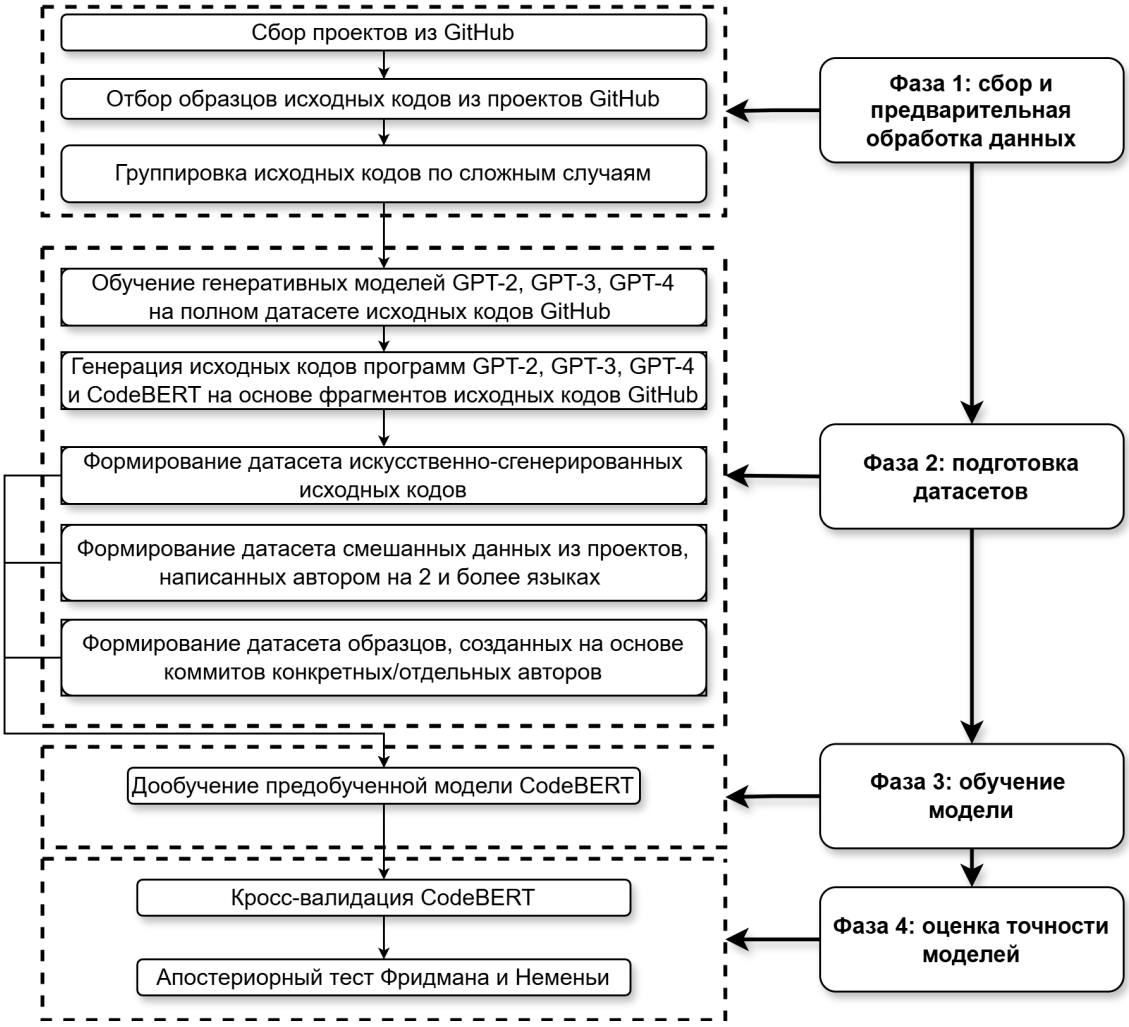


Рисунок 10 – Процесс проведения экспериментов для сложных случаев

Проведен эксперимент для случая, когда программист владеет несколькими языками программирования. Точность классификации для авторов, использующих два языка программирования, достигает 92%. Средняя точность для авторов, разрабатывающих на трех и более языках программирования, достигает 91%.

В рамках оценки точности идентификации автора искусственно-сгенерированного исходного кода программы было рассмотрено два случая. Первый – разграничение авторства между человеком и генеративной моделью (рисунок 11). В каждом эксперименте для 5, 10 и 20 авторов одним из авторов являлась генеративная модель. Средняя точность идентификации кода, созданного GPT-2 составила 96%, GPT-3 и GPT-4 – 92% и CodeBERT – 90%. Разница в точности обусловлена тем, что модели 3 и 4-го поколений демонстрируют лучшую генеративную способность, чем 2-го, а CodeBERT является специализированной моделью для исходных кодов.

Второй эксперимент был нацелен на оценку точности модели при разграничении авторства между четырьмя GPT-моделями. Целью работы методики являлось определение принадлежности анонимного исходного кода, также сгенерированного искусственно, к одной из языковых моделей. Эксперимент проводился для 5 языков программирования: Java, C++, Python, JavaScript и Go. Точность определения авторства исходного кода в этом случае для всех языков программирования составила более 90%.

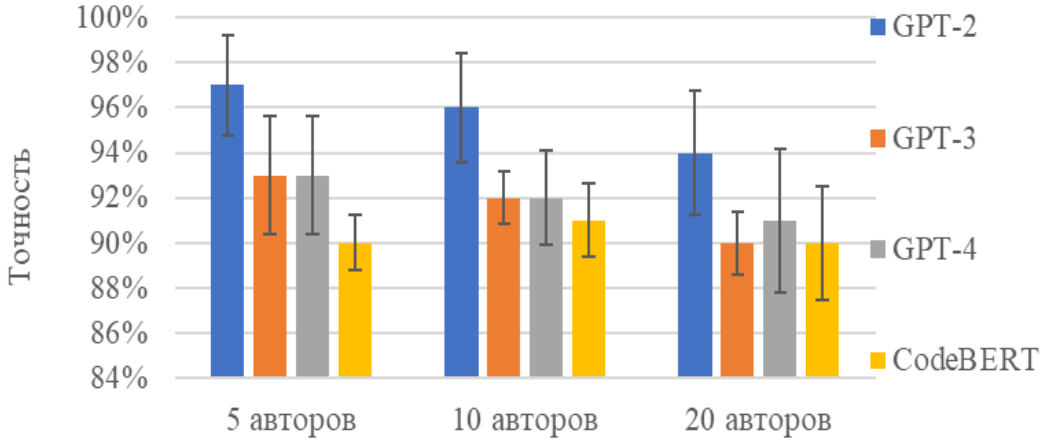


Рисунок 11 – Результаты эксперимента по разделению авторства между человеком и генеративной моделью

Последнее исследование посвящено идентификации автора изменений (коммита), сделанных разработчиками в одном и том же проекте. С целью разработки методических рекомендаций по применению методики для идентификации авторов-программистов, разрабатывающих коммерческое ПО и использующих системы контроля версий, было решено провести эксперимент для разного количества файлов – 5, 10 и 20, соответственно. Согласно результатам, для получения высокоточного (90%) инструмента принятия решений об авторстве следует использовать не менее 10 файлов исходного кода каждого автора.

Предложенная методика показывает среднюю точность 93% в простых случаях идентификации автора исходного кода и позволяет добиться высокой точности для исходных кодов, осложненных обфускацией, стандартами кодирования, искусственной генерацией и командной разработкой.

В **пятой главе** приведены примеры применения разработанной методологии к решению важных народно-хозяйственных задач.

Методика определения текстов деструктивной и экстремистской направленности представлена на рисунке 12.

Эффективность решения задачи повышается за счет использования трансферного обучения и модели, способной решать смежную задачу. Ее использование совместно с отфильтрованными данными позволяет добиться лучших результатов при относительно небольшом обучающем датасете.

Пусть имеется смежная задача обучения $\mathrm{E}_s = \{\mathbf{Y}_s, P(\mathbf{Y}_s \mid \mathrm{X}_s)\}$ и домен $\Upsilon_s = \{\mathbf{X}_s \mid P(\mathrm{X}_s)\}$, где $\mathbf{X}_s$, $\mathbf{Y}_s$ – пространства признаков и меток. $\mathrm{X}_s = \{x_1, ..., x_s\} \in \mathbf{X}_s$ – частное распределение вероятностей. И есть интересующий домен данных $\Upsilon_t = \{\mathbf{X}_t \mid P(\mathrm{X}_s)\}$ и решаемая задача $\mathrm{E}_t = \{\mathbf{Y}_t, P(\mathbf{Y}_t \mid \mathrm{X}_t)\}$. Трансферное обучение направлено на улучшение обучения условного распределения ве-

роятностей $P(Y_t \mid X_t)$ в Υ_t за счет информации, полученной из Υ_s и E_s, где $\mathrm{E}_t \neq \mathrm{E}_s$ или $\Upsilon_t \neq \Upsilon_s$.

В нашем случае можно использовать модель для определения экстремизма в тексте, предназначенную для других языков $(\mathbf{X}_t \neq \mathbf{X}_s)$, $(\mathbf{Y}_t = \mathbf{Y}_s)$. Либо модель, обученную на русскоязычных текстах для определения негативного окраса $(\mathbf{X}_t = \mathbf{X}_s)$, $(\mathbf{Y}_t \neq \mathbf{Y}_s)$. Была использована модель Russian sensitive topics поскольку помимо стандартного разделения по эмоциональной окраске текста содержит потенциально опасные для физического и ментального здоровья тематики, характерные для запрещенных законодательством РФ текстов.

Рисунок 12 – IDEF0 диаграмма определения текста деструктивной и экстремистской направленности

Для выделения информативных признаков используется семантический анализ, созданный на базе BERTopic (рисунок 13).

Результаты экспериментов представлены в таблице 5. Экспериментально установлено, что добавление трансферного обучения позволяет получить прирост точности до 10%, 7% и 12% в случае закрытого и открытого множеств авторов и определения деструктивного контента, соответственно. Добавление семантической кластеризации – 9%, 5% и 9% для тех же случаев. Совместное использование двух инструментов дает совокупный прирост точности 12%, 14% и 20% для обоих случаев определения авторства и деструктивного контента по сравнению с базовой методикой, представленной в разд. 3.

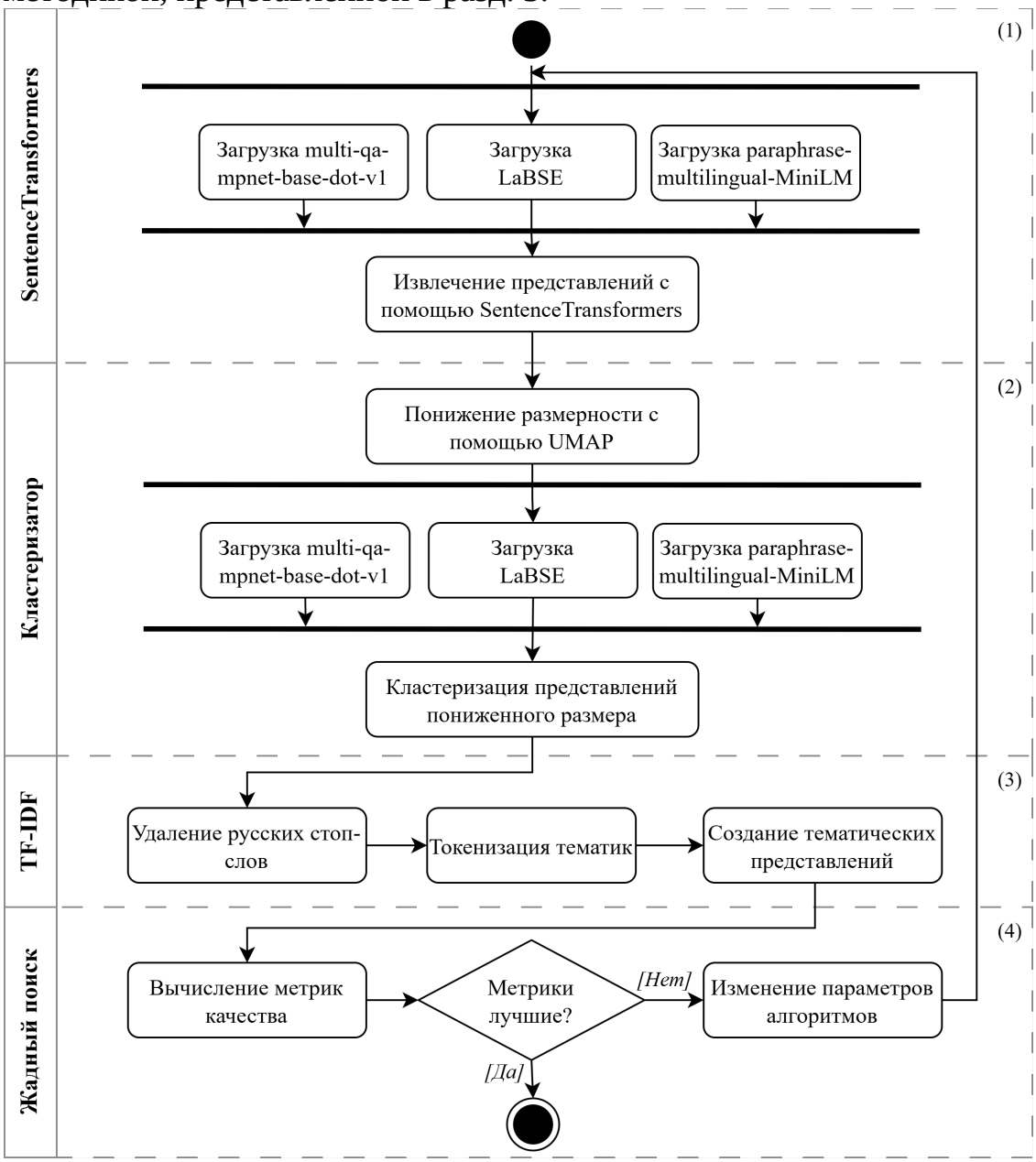


Рисунок 13 – Семантическая кластеризация при помощи BERTopic

Таблица 5 – Результаты экспериментов по определению деструктивного и экстремистского контента в тексте и его автора

Количество авторов	Точность, % (закрытая атрибуция)		
	Telegram экстремисты	KavkazChat*	Telegram иноагенты
2	94±4	87±4	95±3
5	90±3	80±9	91±3
10	80±4	72±6	77±4

20	71±6	54±9	67±5
Точность, % (открытая атрибуция)			
2+null	93,5±2,4	87±4	92±2
4+null	90,1±3,6	80±3	87±3
9+null	78,5±4,2	70±3,5	76±3
19+null	70,4±1,9	52±4	61±2
Точность, % (деструктивный и экстремистский контент)			
–	94±5	86±6	90±8

Методика определения возраста автора текстовой информации представлена на рисунке 14. Её ключевой особенностью является фильтрация недостоверных данных с помощью модели VGG-Face, которая включает несколько шагов. Сначала модель определяет возраст. Затем к предсказанному моделью возрасту прибавляется или вычитается 2 года. В датасет добавляются только тексты тех пользователей, чей указанный возраст совпадает с возрастом, определенным по фотографиям, не менее чем для 70% от имеющихся в профиле за последние два года. Сообщения пользователей, публикующих преимущественно старые фотографии (старше трех лет), в набор данных не включаются. Такой метод позволяет отобрать только достоверные данные для обучения.

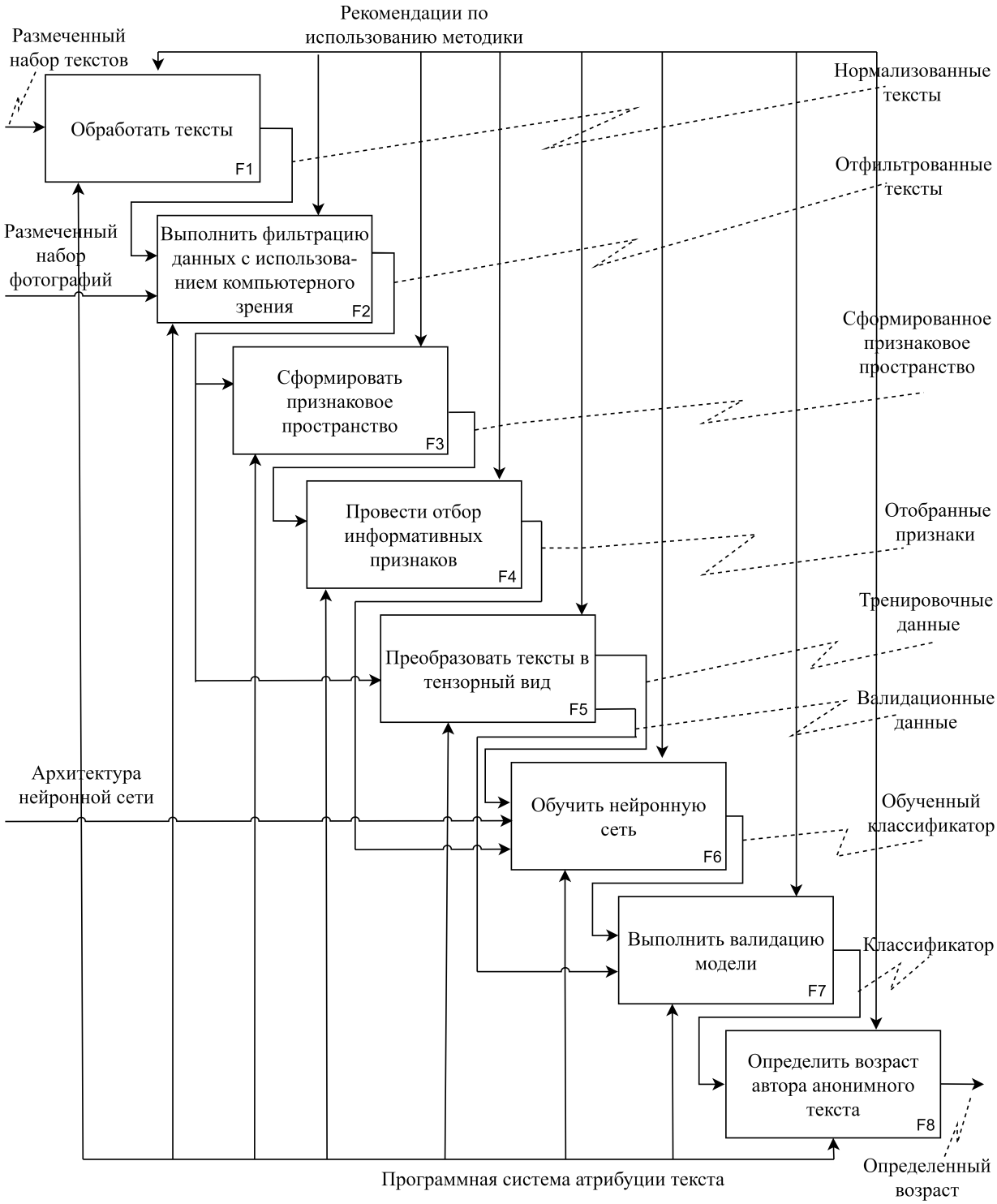


Рисунок 14 – Методика определения возраста автора

В рамках эксперимента модели обучались на двух наборах данных. В первый были включены данные, отсеянные в результате фильтрации, во второй – данные, отобранные VGG-Face как достоверные. Во второй набор также вошли сочинения школьников, отобранные с целью расширения набора с достоверными сведениями о возрасте. Образцы в наборах были разделены на возрастные группы «до 18 лет» и «старше 21 года», а также на группы «до 18 лет», «от 21 до 27 лет» и «старше 30». Полученные результаты представлены в таблице 6.

Таблица 6 – Результаты экспериментов по определению возраста

	Точность	Точность	Среднее время	Общее время

	2 возрастные группы, %		3 возрастные группы, %		на эпоху, с.		обучения, с.	
Модель	Исх.	Фильт.	Исх.	Фильт.	Исх.	Фильт.	Исх.	Фильт.
NB	52	60	40	43	6	4	299	193
SVM	60	68	57	62	7	4	308	206
CNN	43	47	49	52	9	7	120	88
RNN	60	75	42	58	10	9	129	98
fastText	66	82	45	63	8	5	326	201
CRNN	64	82	43	61	14	9	149	98
BERT	66	80	41	60	55	47	552	581
XLM	50	50	38	46	69	63	697	636
RoBERT	50	75	41	60	48	42	481	430

Использование верифицированного набора данных, отобранных фильтром, обеспечивает повышение точности в среднем на 12%. BERT и SVM, обученный на информативных признаках, отобранных методом ANOVA из базовых статистик, индексов лексического разнообразия и удобочитаемости уступают fastText. Точность предсказания возрастной категории в случае бинарной классификации составила 82%, в случае многоклассовой – 63%. Впервые учитываются тексты, написанные лицами младше 18 лет.

На рисунке 15 представлена **методика идентификации пола и гендера**, включая представителей ЛГБТ-сообщества*, основанная на принятии решений ансамблем SVM, CNN и BERT, обученных на признаковом пространстве, сформированном с помощью семантической кластеризации и частотных распределений триграмм символов, сглаженных методом Катца.

Было проведено 13 экспериментов, касающихся определения как половой принадлежности пользователей, так и их гендера:

1) определение пола автора текста среди гетеросексуальных мужчин и женщин;

2) определение принадлежности автора текста к ЛГБТ-сообществу*. В первый класс объединены все гетеросексуальные мужчины и женщины, во второй – все представители ЛГБТ*;

3) определение написан ли текст гетеросексуальным мужчиной, гетеросексуальной женщиной или представителем ЛГБТ*;

* Запрещенная в Российской Федерации экстремистская организация.

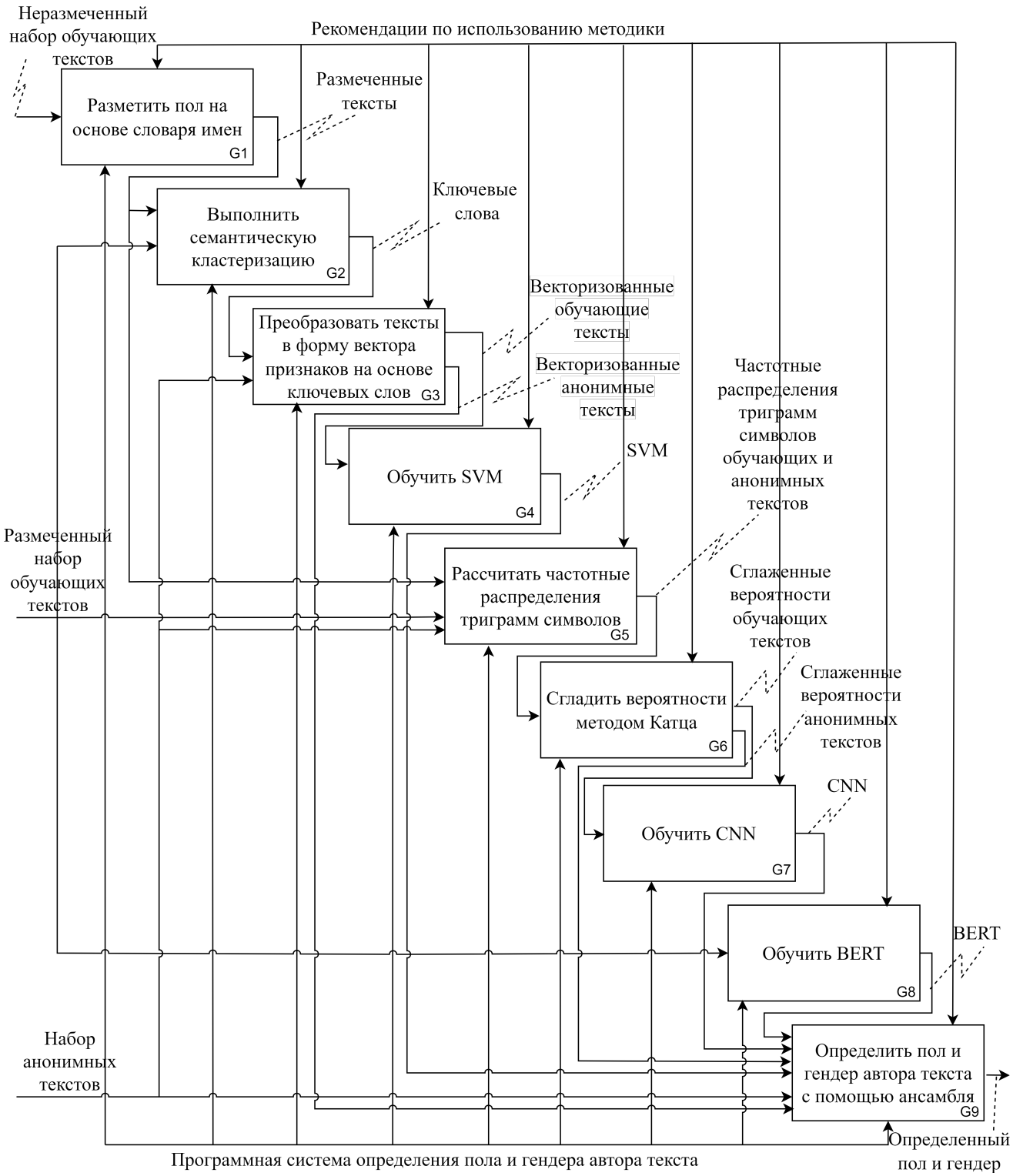


Рисунок 15 – Методика идентификации пола и гендера автора текста

4) определение написан текст гетеросексуальным мужчиной, гетеросексуальной женщиной, гомосексуальным мужчиной или гомосексуальной женщиной;

5) определение написан ли текст гомосексуальным мужчиной или гомосексуальной женщиной;

6) определение пола автора текста среди гетеросексуальных и гомосексуальных мужчин и женщин. В первый класс объединены все мужчины, во второй все женщины по биологическому полу;

7) определение написан ли текст гетеро- или гомосексуальным мужчиной;

8) определение написан ли текст гетеро- или гомосексуальной женщиной;

9) определение написан ли текст гетеросексуальным мужчиной, гетеросексуальной женщиной или гомосексуальным мужчиной;

10) определение написан ли текст гетеросексуальным мужчиной, гетеросексуальной женщиной или гомосексуальной женщиной;

11) определение написан ли текст гетеросексуальным мужчиной, гетеросексуальной женщиной, гомосексуальным мужчиной, гомосексуальной женщиной, бисексуалом, трансгендером;

12) конкретизация пола и гендера автора текста среди представителей ЛГБТ (гомосексуальный мужчина, гомосексуальная женщина, бисексуал, трансгендер);

13) определение написан ли текст бисексуалом или гетеросексуальным мужчиной или гетеросексуальной женщиной.

Результаты исследования приведены в таблице 7. В задачах разграничения по признаку биологического пола точность методики составляет 92% при классификации гетеросексуальных мужчин и женщин, и 88% при использовании смешанного набора, сочетающего гетеро- и гомосексуальных мужчин и женщин. Это подчеркивает факт установления паттернов речевого поведения, зависящих от биологического пола, а не ориентации. В случае разграничения людей одного биологического пола по признаку ориентации установлено, что гомо- и гетеросексуальные женщины используют менее схожую манеру письма, чем мужчины разной ориентации (точность 93% в случае женщин и 85% в эксперименте с мужчинами). В эксперименте, направленном на определение представителей ЛГБТ-сообщества* и гендерной идентичности (традиционная и нетрадиционная) на основе стиля письма, получена точность 93%. Отметим и эксперимент, направленный на предсказание гендерной группы, включая четыре класса ЛГБТ* и два класса, представленных традиционными ориентациями – точность модели составила 68%. Установлено, что ансамбль работает до 8% лучше, чем входящие в его состав модели по отдельности.

Таблица 7 – Результаты экспериментов по определению пола и гендера

Набор данных	Количество образцов в классе	Точность ансамблей, %		
		SVM+ CNN	CNN+ RuBERT	SVM+CNN +RuBERT
Муж./жен.	153 315	91±4	92±3	92±4
ЛГБТ*/муж.+жен.	336 316	91±3	93±3	92±3
ЛГБТ*/муж./жен.	153 315	74±3	82±2	82±3
Муж./жен./геи/лесб.	59 596	74±3	84±2	85±3
Геи/лесб.	59 596	84±3	89±3	90±3
Геи+муж. (гет.)/лесб+жен. (гет.)	229 372	88±2	86±2	88±3
Муж (гет.)/геи	76 057	83±2	84±4	85±2
Жен (гет.)/лесб.	59 596	86±3	93±3	91±2
Муж./жен./геи	76 057	70±4	75±3	74±4
Муж./жен./лесб.	59 596	74±4	80±2	79±3
Муж./жен./геи/лесб./бисекс./ транс.	36 166	59±3	68±4	68±3
Геи/лесб./бисекс./транс.	36 166	69±3	66±3	69±3

Муж./жен./бисекс.	36 166	76±3	75±3	77±3

Методика проверки однородности текста и поиска заимствований (рисунок 16) основывается на методике открытой атрибуции.

В качестве подходов для выявления неоднородных фрагментов текста предлагается использование:

1) метода накопительных сумм (QSUM) совместно с SVM для случаев, где приоритетной является скорость;

2) метода на основе сиамских нейронных сетей SimNN для случаев, где приоритетной является точность.

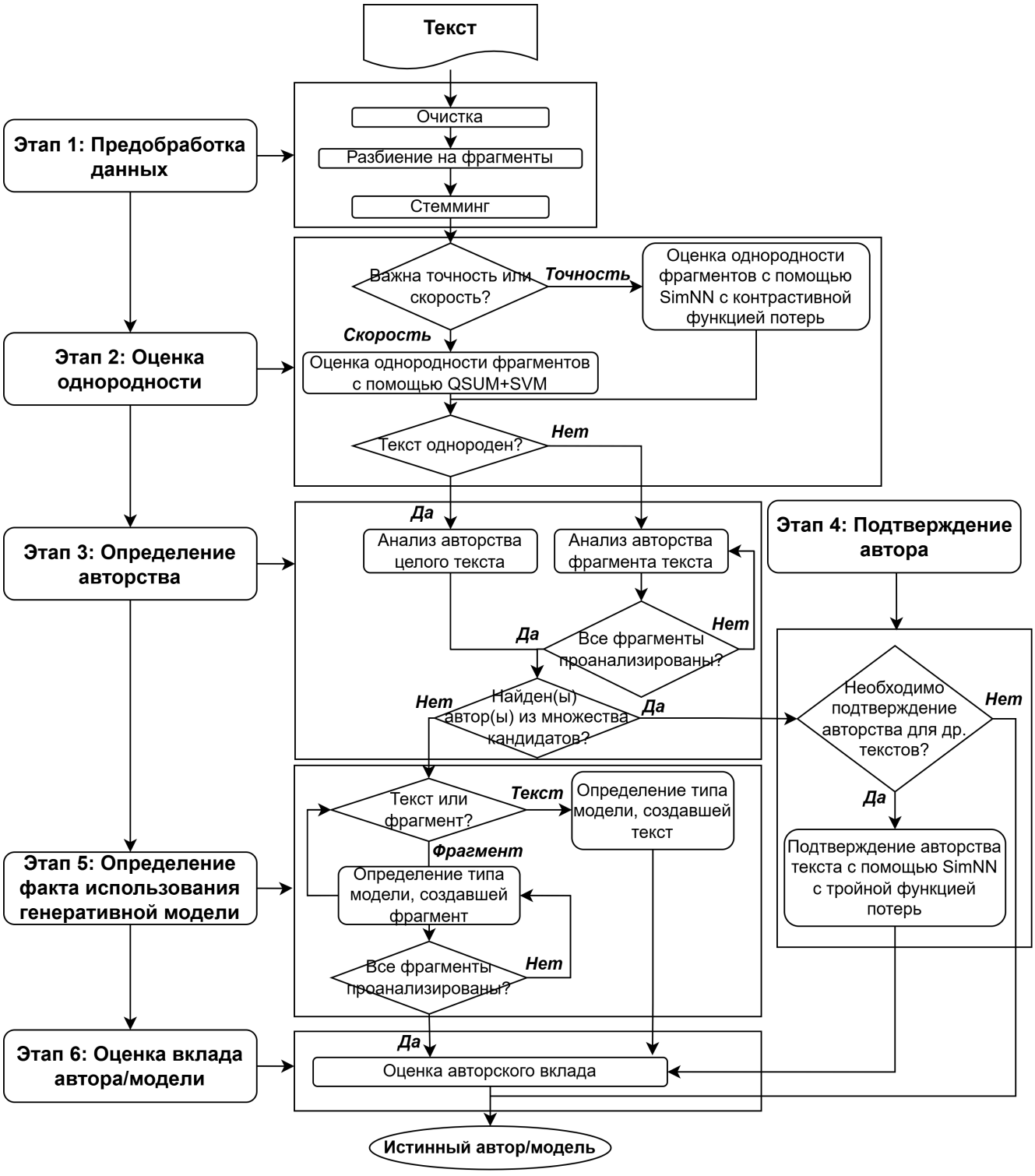


Рисунок 16 – Методика проверки однородности текста и поиска плагиата

Подход, отличающийся точностью, базируется на SimNN и состоит из двух основных шагов:

1. Обучение SimNN. Основой эффективного подтверждения автора является наличие эталона с достоверным авторством и качественных векторного представления. За основу взята модель контрастивного обучения, нацеленная на минимизацию различий между текстами одного автора и максимизацию между текстами разных авторов. Для эффективного обучения SimNN рекомендуется использование специфической функции потерь – тройной (*t_l*) или контрастивной (*c_l*). Особенность этих функций потерь состоит в обучении модели на тройке, состоящей из якоря (текста), позитивного примера (текста того же автора) и отрицательного примера (текста другого автора).

2. На основе эмбеддингов, выполняется обучение классификационного полносвязного слоя Dense, который принимает эти эмбеддинги в качестве входных данных и выносит финальное решение о принадлежности текстов автору и однородности. При оценке однородности данным способом производится сравнение каждого фрагмента с каждым.

На этапе определения авторства для неоднородных фрагментов происходит идентификация автора по открытому множеству кандидатов. Такой анализ оставляет допущение о том, что текст или фрагмент написан автором, отсутствующим в множестве кандидатов, или сгенерирован НС. В качестве метода принятия решения о принадлежности фрагмента автору используется комплекс из ГА, GRU+CNN и SVM.

Этап подтверждения автора выполняется для текстов, прошедших проверку однородности. Этап направлен на подтверждение принадлежности двух и более текстов одному и тому же автору. Подтверждение автора выполняется с помощью SimNN, однако вместо предложений используются полные текст с доподлинно известным автором или моделью.

В случае, если истинный автор не был найден на этапе подтверждения, проводится анализ, направленный на определение факта использования генеративных моделей при создании текста. В качестве метода принятия решения используется гибрид GRU+CNN.

Последним этапом методики является оценка вклада автора и (или) генеративной модели в создание анализируемого текста. Выполняется проверка авторства каждого фрагмента и суммирование частей, принадлежащих конкретному автору или модели.

В результате предварительных экспериментов установлено, что наиболее точным подходом является SimNN+*c_l* при использовании фрагмента в 25 предложений. Эффективность QSUM+SVM ниже на 8,5%. Однако этот подход работает на несколько порядков быстрее.

Подход на основе SimNN, обученного с *c_l*, был использован в рамках эксперимента, посвященного определению процента однородности в текстах научных статей и диссертаций. Диссертационные работы однородны на 85,8%. Такой результат обусловлен высокими требованиями к оригинальности работ и тем фактом, что диссертация пишется одним автором. Процент однородности для научных статей в случае, когда автор не имеет соавторов, составляет 81,3%. Чем боль-

ше соавторов у статьи, тем ниже однородность. Однако для 5 и более авторов разница перестает быть существенной.

К корпусу научных работ была применена методика идентификации автора для случая открытого множества кандидатов. Как к текстам в их полном объеме для случая, когда текст однороден, так и к фрагментам текста с признаками заимствований. При этом фрагменты соавторов генерировались при помощи моделей семейства GPT на основе стилей случайных авторов из набора данных, с учетом темы статьи-оригинала. Аналогичный эксперимент был проведен для полностью искусственно-сгенерированных текстов. Наилучшую разделительную способность классификатор демонстрирует при отсутствии или 1-2 соавторах. В этом случае точность достигает 98%.

В рамках экспериментов, посвященных подтверждению автора, оценивались различные подходы к контрастивному обучению и вычислению функции потерь. В качестве модели использовалась SimNN, в качестве функций потерь – *t_l* и *c_l*. Проверка проводилась на 2275 случайных парах текстов. В составе пары могли быть как тексты, написанные разными авторами, так и одним и тем же. Полученные результаты представлены в табл. 8. В отличие от результатов, полученных для этапа оценки однородности, лучшая точность была достигнута SimNN, обученной с *t_l*.

Таблица 8 – Результаты экспериментов по выбору подхода к обучению

Домен/Модель	Тройные потери	Контрастивные потери
	Точность, %	
Научные статьи	98±1	97,5±0,9
Диссертации	99,1±0,5	98±1,1
GPT-3.5	99,4±0,3	99,1±0,7
GPT-4	99±0,8	99±0,4

В **шестой главе** приводится описание разработанного алгоритмического и программного обеспечения, наборов данных для анализа естественно- и искусственно-языковых текстов в задачах кибербезопасности.

Структура программного комплекса для анализа естественно- и искусственно-языкового текста и решения задач кибербезопасности PyAuthorship представлена на рисунке 17. Были реализованы подсистемы сбора образцов текстов, хранения информации, анализа естественного и искусственного языка, выделения информативных признаков на основе семантического анализа, классификации.

На основе наиболее эффективных семантических методов извлечения информативных признаков комплекс определяет подмножество признаков, которое используется для исследований в рамках решения прикладных задач кибербезопасности:

1) Формирование «портрета» пользователя – стиля автора, пола, возраста.
2) Исследование социальных сетей, противодействие педофилии.
3) Противодействие терроризму и экстремизму.
4) Поиск авторов-вирусописателей.
5) Криминалистические экспертизы.
6) Плагиат в образовательном процессе и академической среде.

7) Продленная аутентификация по тексту.

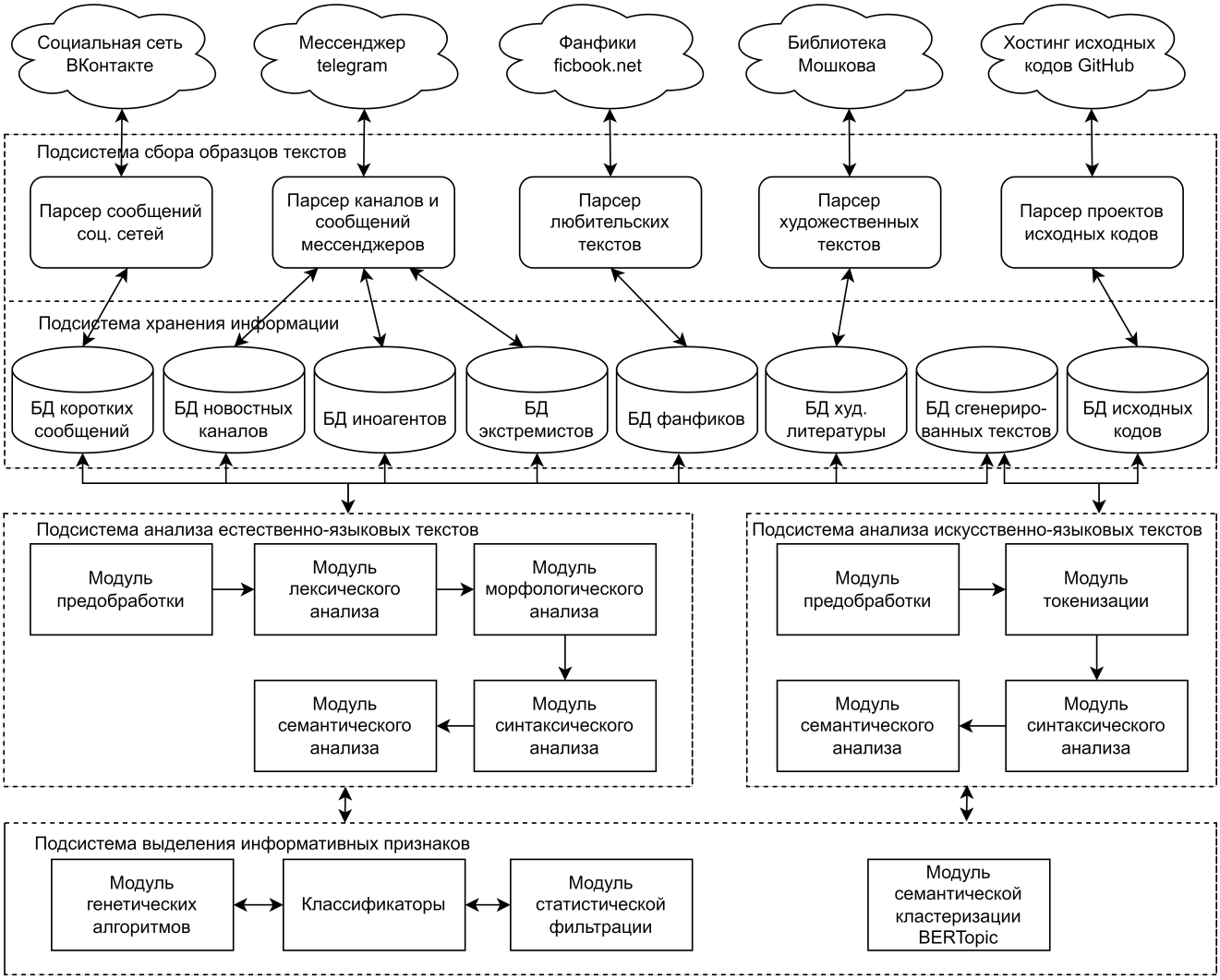


Рисунок 17 – Структура программного комплекса
для получения информативных признаков текста

На основе «PyAuthorship» созданы прикладные программы для решения задач кибербезопасности.

Разработано оригинальное алгоритмическое и программное обеспечение программной системы «Авторовед», позволяющее идентифицировать автора с помощью нейронных сетей и машины опорных векторов, выделять заимствованный текст, подтверждать или опровергать авторство текста.

Разработана программа прикладная программа для определения возраста автора текстового сообщения в социальных сетях и мессенджерах «Age Detector».

Разработано оригинальное алгоритмическое и программное обеспечение «CoDEtective», учитывающего в том числе сложные случаи (соавторство, обфускаторы, стандарты кодирования и др.), для идентификации автора программного кода, написанного на 13 наиболее популярных языках программирования.

Разработана прикладная программа для определения эмоциональной окраски текста «Moodorrude».

Разработан оригинальный программный комплекс «PyDestructiveDetector» и алгоритмическое обеспечение, предназначенное для идентификации текстов, со-

держащих пропаганду нетрадиционных ценностей, деструктивную и экстремистскую направленность, потенциально нарушающих законодательство РФ. А также идентификации лиц, распространяющих подобные текстовые материалы.

Разработан программный комплекс для определения пола и гендера автора естественно-языкового текста «ProGender».

Разработана структура баз данных для хранения текстов, созданных в результате повседневной, профессиональной и творческой деятельности автора, учитывающая особенности каждой группы текстов. Собраны объемные и релевантные датасеты, предназначенные для решения частных практических задач кибербезопасности (таблица 9).

Таблица 9 – Наборы данных для решения задач кибербезопасности

Набор данных	Количество авторов	Количество текстов	Количество символов
Художественные тексты	413	6132	332507700
Любительские тексты	686	6569	167332137
Сообщения пользователей соц. сетей	1114955	2815104	177683573
Тексты, сгенерированные НС	200	10742	1215330
Исходные коды программ	383	102908	212336673
Короткие сообщения, иноагенты	1921	33028	1051176
Короткие сообщения, экстремисты	1844	25336	678432
Короткие сообщения, пол	23054	336316	88035644
Короткие сообщения, ЛГБТ* (telegram)	20384	207193	43957226
Короткие сообщения, ЛГБТ* (X**)	395	162669	21615303
Научные тексты, диссертации	525	525	247743
Научные тексты, статьи	3971	4503	21401
Научные тексты, сгенерированные НС	500	2000	26358

В **заключении** сформулированы основные научные и практические результаты.

ЗАКЛЮЧЕНИЕ

Исследования, выполненные в диссертационной работе, позволили решить **научную проблему** – создание комплексной методологии идентификации автора текстовой информации, **имеющую важное хозяйственное значение, внедрение которой вносит значительный вклад** в развитие кибербезопасности, защиты авторского права, компьютерной лингвистики. Разработанная методология и инструментарий на ее основе может использоваться для защиты интеллектуальной собственности, идентификации автора сообщений в сети Интернет и продленной аутентификации пользователей социальных сетей, выявления признаков пропаганды ЛГБТ* и педофилии, определения автора-вирусописателя, определения эмоциональной окраски текста, деструктивных, а также текстов экстремистской направленности, запрещенных законодательством Российской Федерации.

Основные результаты диссертационной работы:

* Запрещенная в Российской Федерации экстремистская организация.

** Социальная сеть X (бывший Twitter) запрещена на территории Российской Федерации.

1. Проведен подробный анализ основных работ в области идентификации автора естественного и искусственного текста, методов профилирования автора на основе его профессиональной деятельности, возрастных и гендерных атрибутов, идеологии, тональности публикуемого текста. Рассмотрены работы, направленные на атаки методов путем внедрения образцов, созданными генеративными моделями, методов имитации и обфускации текстового материала.

2. Разработана комплексная методология идентификации автора текстовой информации для решения задач кибербезопасности, включающая широкий спектр методов, моделей и алгоритмов, основанных на анализе как естественно-языковых, так и искусственно-языковых текстов, предоставляющая универсальный инструментарий для классификации текстов. Методология предусматривает применение методов статистического анализа, машинного и глубокого обучения, что делает её адаптируемой к различным задачам и способной эффективно бороться с потенциальными атаками.

3. Предложена модель, учитывающая как внутренние компоненты процесса создания автором текста в киберсреде, так и внешние ограничения, накладываемые личностью автора, видом деятельности, в том числе профессиональной и досуговой, особенностями среды и семантики текста, что позволяет не только точно идентифицировать автора, но и учитывать специфику его взаимодействия с киберпространством.

4. Предложена новая методика идентификации автора русскоязычного текста. Особое внимание уделено методам отбора информативных признаков в сочетании с SVM, среди которых генетический алгоритм и регуляризационный отбор были признаны наиболее эффективными для классификации художественных текстов, для коротких текстов выбрана гибридная модель на основе комбинации GRU+CNN. Эксперименты показывают, что предложенная методика обеспечивает высокую точность классификации (до 99% для 2 и 5 авторов, 94% для 10 и 88% для 20), сопоставимую или превосходящую аналоги. Было установлено, что методика на основе одноклассового SVM и GRU+CNN обеспечивает высокую точность (от 98% до 75% для 3–20 классов) при работе с открытым множеством кандидатов. Методика применима к задаче верификации автора текста и способна решать ее с точностью 99,2%, 98,9% и 97,8% для художественных, любительских и коротких текстов, соответственно.

5. Разработана методика определения текстов деструктивной и экстремистской направленности, дополняющая базовую методику идентификации автора естественного текста за счет интеграции семантической кластеризации текста и трансферного обучения. Установлено, что добавление трансферного обучения позволяет получить прирост точности до 10% и 12% в случае идентификации автора и определения деструктивного контента, соответственно, а добавление семантической кластеризации – 9% для тех же случаев. Совместное использование инструментов дает совокупный прирост точности 14% и 20%, соответственно.

6. Разработана методика идентификации пола и гендера автора текстовой информации, впервые учитывающая особенности, характерные для текстов, созданных представителями ЛГБТ-сообщества*. Методика основана на принятии решений ансамблем SVM, CNN и BERT. Признаки для SVM сформированы на

* Запрещенная в Российской Федерации экстремистская организация.

основе частотных распределений ключевых слов, полученных с помощью семантической кластеризации, а признаки для CNN – на основе сглаженных вероятностей триграмм символов. Ансамбль при разграничении по половому признаку (мужчины и женщины) достигает точности 88–92%, точность определения признаков ЛГБТ-сообщества – 93%. Для наиболее сложного случая классификации шести гендерных групп точность составляет 68%.

7. Предложена методика определения возраста автора текста, основанная на модели fastText для анализа текстов и методе фильтрации пользовательских фотографий с помощью модели VGG-Face. Определение возраста по фотографии через VGG-Face и последующее сравнение полученных данных с возрастом, указанным в профиле пользователя, обеспечивает верификацию корректности пользовательских данных. Процедура фильтрации оказывает значительное влияние на точность распределения возрастных данных и качество итоговой модели. Для двух возрастных категорий (младше 18 лет и старше 21 года) точность составила 82%, а для трех категорий (младше 18, от 21 до 27 и старше 30 лет) – 63%.

8. Разработаны репрезентативные наборы данных для обучения и оценки методического и алгоритмического обеспечения. Набор данных для идентификации авторов естественно-языковых текстов в целях решения задач кибербезопасности включают художественные, любительские, научные тексты, тексты пользователей социальных сетей, а также искусственные тексты, созданные генеративными моделями на их основе. Набор данных для идентификации авторов искусственно-языковых текстов включают исходные коды программ на 13 языках программирования; обфусцированные исходные коды на 5 языках программирования; исходные коды авторов, владеющих 2 и более языками программирования; коммиты исходных кодов программ, написанных группой разработчиков; исходные коды, написанные по стандартам Linux Kernel; исходные коды, сгенерированные моделями семейства GPT.

9. Разработан классификатор на основе модели CodeBERT и методика идентификации автора исходного кода программы на его основе, учитывающая простые (точность до 93%) и сложные (точность до 85%) прикладные случаи. Полученные результаты подтверждают независимость методики от языка программирования автора, а также ее устойчивость к преднамеренным преобразованиям и следованию стандартам кодирования.

10. Разработана методика проверки однородности текста и поиска заимствований. Методика основана на комбинации сиамской сети SimNN, обученной с контрастивной и тройной функциями потерь, для подтверждения автора двух и более текстов/фрагментов, а также подхода к определению авторства неоднородных фрагментов. Точность методики для выявления неоднородных фрагментов составляет 94%, для определения авторства фрагментов – более 90%, для подтверждения автора – до 99%.

11. На основе разработанной методологии и методик создано алгоритмическое и программное обеспечение для анализа естественно- и искусственно-языковых текстов в задачах кибербезопасности. В его состав входят: программный комплекс для получения информативных признаков на основе семантического анализа «PyAuthorship», программная система для идентификации автора письменной речи «Авторовед», чат-бот для определения возраста автора тек-

стового сообщения «Age Detector», система для идентификации автора исходного кода программы «CoDEtective», программа для тонального анализа текста «Moodorrude», программный комплекс для идентификации запрещенных законодательством РФ и деструктивных текстов «PyDestructiveDetector» и программный комплекс для определения пола и гендера автора естественно-языкового текста «ProGender».

ОСНОВНЫЕ ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

Публикации в журналах, рекомендованных ВАК

1. **Романов А.С.** Модель базы данных для хранения текстов и их характеристик / А.С. Романов // Доклады ТУСУР. – 2008. – № 1(17). – С. 70–73.

2. **Романов А.С.** Структура программного комплекса для исследования подходов к идентификации авторства текстов / А.С. Романов // Доклады ТУСУР. – 2008. – Ч. 1, № 2(18). – С. 106–109.

3. **Романов А.С.** Методика идентификации автора текста на основе аппарата опорных векторов / А.С. Романов // Доклады ТУСУР. – 2009. – Ч. 2, № 1(19). – С. 36–42.

4. **Романов А.С.** Обобщенная методика идентификации автора неизвестного текста / А.С. Романов, А.А. Шелупанов, С.С. Бондарчук // Доклады ТУСУР. – 2010. – Ч. 1, № 1(21). – С. 108–112.

5. **Романов А.С.** Методика проверки однородности текста и выявления плагиата на основе метода опорных векторов и фильтра быстрой корреляции / А.С. Романов, З.И. Резанова, Р.В. Мещеряков // Доклады ТУСУР. – 2014. – № 2(32). – С. 264–269.

6. Созинова И.С. Определение поискового спама с использованием метода опорных векторов / И.С. Созинова, **А.С. Романов.**, Р.В. Мещеряков // Труды СПИИРАН. – 2014. – Вып. 36. – С. 78–91.

7. Куртукова А.В. Моделирование архитектуры нейронной сети в задаче идентификации автора исходного кода / А.В. Куртукова, **А.С. Романов** // Доклады ТУСУР. – 2019. – Т. 22, № 3. – С. 37–42.

8. Куртукова А.В. Оценка влияния обфускации на процесс идентификации автора программного кода / А.В. Куртукова, Е.Е. Сваровская, **А.С. Романов** // Доклады ТУСУР. – 2020. – Т. 23, № 2. – С. 50–54.

9. Применение методов машинного обучения и отбора признаков на основе генетического алгоритма в решении задачи определения автора русскоязычного текста для кибербезопасности / А.В. Куртукова, **А.С. Романов**, А.М. Федотова, А.А. Шелупанов // Доклады ТУСУР. – 2022. – Т. 25, № 1. – С. 79–85.

10. Идентификация автора исходного кода программы на основе неоднородных данных для решения задач кибербезопасности / А.В. Куртукова, **А.С. Романов**, А.А. Шелупанов, А.М. Федотова // Моделирование, оптимизация и информационные технологии. – 2022. – № 10(3). – URL: https://moitvivt.ru/ru/journal/pdf?id=1227.

11. Архитектура интеллектуальной системы для идентификации автора исходного кода / А.В. Куртукова, **А.С. Романов**, А.А. Шелупанов, А.М. Федотова // Доклады ТУСУР. – 2022. – Т. 25, № 3. – С. 39–44.

12. Методика определения возраста автора текста на основе метрик удобочитаемости и лексического разнообразия / А.А. Соболев, А.М. Федотова, А.В. Куртукова, **А.С. Романов**, А.А. Шелупанов // Доклады ТУСУР. – 2022. – Т. 25, № 2. – С. 45–52.

13. Куртукова А.В. Разработка методики идентификации авторства бинарных и дизассемблированных кодов программы на основе ансамбля современных методов обработки естественного языка / А.В. Куртукова, **А.С. Романов**, А.А. Шелупанов // Доклады ТУСУР. – 2023. – Т. 26, № 4. – С. 53–60.

14. **Романов А.С.** Методы отбора признаков в задаче определения авторства в контексте кибербезопасности / А.С. Романов // Моделирование, оптимизация и информационные технологии. – 2024. – Т. 12, № 1. – URL: https://moitvivt.ru/ru/journal/pdf?id=1489.

15. **Романов А.С.** Идентификация автора текста для открытого множества кандидатов в контексте кибербезопасности / А.С. Романов // Моделирование, оптимизация и информационные технологии. – 2024. – Т. 12, № 1. – URL: https://moitvivt.ru/ru/journal/pdf?id=1510.

16. **Романов А.С.** Методология идентификации автора текста для решения задач информационной безопасности / А.С. Романов // Вопросы кибербезопасности. – М. : НПО «Эшелон», 2024. – № 3(61). – С. 120–128.

17. Методика идентификации автора текстовых материалов деструктивной направленности / **А.С. Романов**, А.М. Федотова, А.В. Куртукова, А.А. Шелупанов // Информационные технологии. – 2024. – Т. 30, № 12. – С. 632–640.

18. **Романов А.С.** Методика семантической кластеризации для выявления признаков экстремизма в текстовой информации / А.С. Романов // Доклады ТУСУР. – 2024. – Т. 27, № 4. – С. 141–149.

Публикации, включенные в библиографические базы Web of Science, Scopus

19. Анализ тональности текстов с использованием методов машинного обучения (Sentiment Analysis of Text Using Machine Learning Techniques) / **А.С. Романов**, М.И. Васильева, А.В. Куртукова, Р.В. Мещеряков // Proceedings of the R. Piotrowski's Readings in Language Engineering and Applied Linguistics. Saint Petersburg, Russia. November 27, 2017. – 2017. – P. 86–95.

20. Natural text anonymization using universal transformer with a self-attention / **A. Romanov**, A. Kurtukova, A. Fedotova, R. Meshcheryakov // Proceedings of the III International Conference on Language Engineering and Applied Linguistics (PRLEAL-2019)/ Saint Petersburg, Russia. 2019. – 2019. – P. 22–37.

21. Куртукова А.В. Идентификация автора исходного кода методами машинного обучения / А.В. Куртукова, **А.С. Романов** // Труды СПИИРАН. – 2019. – № 3(18). – С. 741–765.

22. Kurtukova A. De-Anonymization of the Author of the Source Code Using Machine Learning Algorithms / A. Kurtukova, **A. Romanov**, A. Fedotova // Conference Proceedings of 2019 International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON). – 2019. – P. 0612–0617.

23. Authorship identification of a russian-language text using support vector machine and deep neural networks / **A. Romanov**, A. Kurtukova, A. Fedotova [et al.] // Future Internet. – 2021. – Vol. 13(3). – Iss. 1. – 16 p.

24. Determining the Age of the Author of the Text Based on Deep Neural Network Models / **A.S. Romanov**, A.V. Kurtukova, A.A. Sobolev [et al.] // Information. – 2020. – Vol. 11. – Iss. 12. – 589. – 12 p.

25. Kurtukova A. Source Code Authorship Identification Using Deep Neural Networks / A. Kurtukova, **A. Romanov**, A. Shelupanov // Symmetry. – 2020. – Vol. 12. – Iss. 12. – 2044. – 17 p.

26. Authorship Attribution of Social Media and Literary Russian-Language Texts Using Machine Learning Methods and Feature Selection / A. Fedotova, **A. Romanov**, A. Kurtukova, A. Shelupanov // Future Internet. – 2022. – Vol. 14(1). – Iss. 4. – 24 p.

27. Age determination of the social media post's author using deep neural networks and facial processing model / **A. Romanov**, A. Kurtukova, A. Sobolev, A. Fedotova // CEUR Workshop Proceedings. – 2021. – Vol. 2842. – P. 51–59.

28. Complex Cases of Source Code Authorship Identification Using a Hybrid Deep Neural Network / A. Kurtukova, **A. Romanov**, A. Shelupanov, A. Fedotova // Future Internet. – 2022. – Vol. 14. – Iss. 287. – 20 p.

29. Authorship Identification of Binary and Disassembled Codes Using NLP Methods / **A. Romanov**, A. Kurtukova, A. Fedotova, A. Shelupanov // Information. – 2023. – Vol. 14. – Iss. 7. – 361. – 21 p.

30. Digital Authorship Attribution in Russian-Language Fanfiction and Classical Literature / A. Fedotova, **A. Romanov**, A. Kurtukova, A. Shelupanov // Algorithms. – 2023. – Vol. 16. – Iss. 1. – 13. – 32 p.

31. Semantic Clustering and Transfer Learning in Social Media Texts Authorship Attribution / A. Fedotova, A. Kurtukova, **A. Romanov**, A. Shelupanov // IEEE Access. – 2024. – Vol. 12. – P. 39783–39803.

Монографии

32. **Романов А.С.** Разработка и исследование математических моделей, методик и программных средств информационных процессов при идентификации автора текста : моногр. / **А.С. Романов**, А.А. Шелупанов, Р.В. Мещеряков. – Томск : В-Спектр, 2011. – 188 с.

Свидетельства о регистрации программ ЭВМ и баз данных:

33. Свидетельство о регистрации электронного ресурса № 151146 Российская Федерация. Программная система для идентификации автора письменной речи «Автороведъ» : заявлено 24.12.2009 : выдано 14.01.2010 / **Романов А.С.** Зарегистрировано в Объединенном фонде электронных ресурсов «Наука и образование».

34. Свидетельство о государственной регистрации программы для ЭВМ № 2020666001 Российскя Федерация. Программа для тонального анализа текста «Moodorrude» : № 2020665103 : заявлено 26.11.2020 : опубликовано 03.12.2020 / **Романов А.С.**, Васильева М.И., Шилов Л.С., Куртукова А.В., Шелупанов А.А. Зарегистрировано в реестре программ для ЭВМ.

35. Свидетельство о государственной регистрации программы для ЭВМ № 2021681140 Российскя Федерация. Система для идентификации автора исходного кода программы «CoDEtective» : № 2021667210 : заявлено 26.10.2021 : опубликовано 17.12.2021 / Куртукова А.В., **Романов А.С.** Зарегистрировано в реестре программ для ЭВМ.

36. Свидетельство о государственной регистрации программы для ЭВМ № 2022616780 Российскя Федерация. Чат-бот для определения возраста автора текстового сообщения «Age Detector» : № 2022616248 : заявлено 12.04.2022 : опубликовано 15.04.2022 / **Романов А.С.**, Соболев А.А., Федотова А.М., Куртукова А.В., Шелупанов А.А. Зарегистрировано в реестре программ для ЭВМ.

37. Свидетельство о государственной регистрации программы для ЭВМ № 2022667584 Российскя Федерация. Программный комплекс для получения информативных признаков автора текста «PyAuthorship» : № 2022667003 : заявлено 19.09.2022 : опубликовано 22.09.2022 / **А.С. Романов**, А.В. Куртукова, А.М. Федотова, А.А. Шелупанов. Зарегистрировано в реестре программ для ЭВМ.

38. Свидетельство о государственной регистрации базы данных № 2022622299 Российскя Федерация. База данных естественно- и искусственно-языковых текстов для определения авторства : № 2022622243 : заявлено 16.09.2022 : опубликовано 21.09.2023 / Куртукова А.В., **Романов А.С.**, Федотова А.М., Шелупанов А.А. Зарегистрировано в реестре баз данных.

39. Свидетельство о государственной регистрации программы для ЭВМ № 2023685210 Российскя Федерация. Программный комплекс для идентификации запрещенных законодательством РФ и деструктивных текстов «PyDestructive Detector» : № 2023682992 : заявлено 1.11.2023 : опубликовано 23.11.2023 / Куртукова А.В., **Романов А.С.**, Федотова А.М., Шелупанов А.А.. Зарегистрировано в реестре программ для ЭВМ.

40. Свидетельство о государственной регистрации базы данных № 2023623911 Российскя Федерация. База данных исходных и дизассемблированных кодов для идентификации авторства программ : № 2023623657 : заявлено 01.112023 : опубликовано 13.11.2023 / Куртукова А.В., **Романов А.С.**, Федотова А.М., Шелупанов А.А. Зарегистрировано в реестре баз данных.

41. Свидетельство о государственной регистрации базы данных № 2024624287 Российскя Федерация. База данных естественных текстов для идентификации пола и гендера автора текста : № 2024624158 : заявлено 05.10.2024 : опубликовано 14.10.2024 / Куртукова А.В., **Романов А.С.**, Федотова А.М., Шелупанов А.А. Зарегистрировано в реестре баз данных.

42. Свидетельство о государственной регистрации программы для ЭВМ № 2024685146 Российская Федерация. Программный комплекс для сбора данных о поле и гендере автора естественно-языкового текста : № 2024683353 : заявлено 10.10.2024 : опубликовано 24.10.2024 / Куртукова А.В., **Романов А.С.**, Федотова А.М., Шелупанов А.А. Зарегистрировано в реестре программ для ЭВМ.

43. Свидетельство о государственной регистрации программы для ЭВМ № 2025615680 Российскя Федерация. Программный комплекс для определения пола и гендера автора естественно-языкового текста «ProGender» : №2025613057 заявлено 20.02.2025 : опубликовано 06.03.2025 / Куртукова А.В., **Романов А.С.**, Федотова А.М., Шелупанов А.А. Зарегистрировано в реестре программ для ЭВМ.

Публикации в журналах, рекомендованных ВАК, по другим специальностям

44. Резанова З.И. О выборе признаков текста, релевантных в автороведческой экспертной деятельности / З.И. Резанова, **А.С. Романов**, Р.В. Мещеряков // Вестник Томского государственного университета. Филология. – Томск, 2013. – № 6(26). – С. 38–52.

45. Резанова З.И. Задачи авторской атрибуции текста в аспекте гендерной принадлежности (к проблеме междисциплинарного взаимодействия лингвистики

и информатики) / З.И. Резанова, **А.С. Романов**, Р.В. Мещеряков // Вестник Томского государственного университета. – 2013. – № 370. – С. 24–28.

Заказ № . Тираж экз.
Томский государственный университет
систем управления и радиоэлектроники
634050, г. Томск, пр. Ленина, 40.
Тел. (3822) 533018.