

ISSN: 2658–5782

Номер 2

2025

МНОГОФАЗНЫЕ СИСТЕМЫ

mfs.uimech.org



ИМех

им. Р.Р. Мавлютова



Анализ возможности применения интеллектуальных систем машинного обучения для работы с научными документами

В.Н. Гокарев¹✉, А.Ф. Беззубов¹, Д.В. Березкин²

¹ 27 ЦНИИ, Москва

² МГТУ им. Н.Э. Баумана, Москва

E-mail: vgokarev@mail.ru

В предлагаемой статье описывается гибридный подход к автоматизации обработки научно-технических документов в исследовательском институте, сочетающий многоходовую нейросетевую архитектуру и интеграцию лингвистических методов. Такие методы позволяют значительно оптимизировать процессы работы с документами — ускорять выделение информации и снижать долю ошибок на первых этапах обработки данных, что в итоге оказывает экспоненциально положительное влияние на все последующие стадии работы с научными документами. В работе отмечается важность данных подходов для решения актуальных задач в различных научных сферах. Рассматриваются ключевые проблемы, которые требуют комплексного подхода к решению, и описываются перспективные направления дальнейших исследований. Предложена методология по повышению эффективности автоматизации, в которой особое внимание уделяется гибридным системам, объединяющим экспертные оценки, алгоритмические решения и нейросетевые модели для повышения эффективности автоматизации. Непрерывный рост вычислительных мощностей способствует развитию этих технологий и расширению их практического применения. Полученные результаты могут найти применение в оптимизации процессов автоматизации производства, а также в разработке масштабируемых систем, основанных на современных технологиях искусственного интеллекта.

Ключевые слова: автоматизация обработки документов, машинное обучение, экспертные оценки, научные документы, нейронные сети, обработка естественного языка (NLP), лингвистическая обработка текстов, классификация, гибридные методы

Analysis of the possibility of using intelligent machine learning systems to work with scientific documents

V.N. Gokarev¹✉, A.F. Bezzubov¹, D.V. Berezkin²

¹ The 27th Central Research Institute, Moscow, Russia

² Bauman Moscow State Technical University, Moscow, Russia

E-mail: vgokarev@mail.ru

The article is devoted to the analysis of the application of modern methods of automation of work with scientific documents using intelligent machine learning systems. The paper highlights the importance of these approaches for solving urgent problems in various scientific fields. The article examines the key problems that require an integrated approach to solving, and describes promising areas for further research. A methodology for improving automation efficiency is proposed, in which special attention is paid to hybrid systems that combine expert assessments, algorithmic solutions and neural network models to increase automation efficiency. The continuous growth of computing power contributes to the development of these technologies and the expansion of their practical application. The results obtained can be used in optimizing production automation processes, as well as in the development of scalable systems based on modern artificial intelligence technologies.

Keywords: document processing automation, machine learning, expert assessments, scientific papers, neural networks, natural language processing (NLP), linguistic text processing, classification, hybrid methods

1. Введение

В современных научно-исследовательских институтах (НИИ) автоматизация работы с научно-техническими документами приобретает особую значи-

мость в связи с возросшими требованиями к оперативности обработки информации и ее структурированию. В настоящее время бурное развитие информационных технологий и стремительное увеличение объема научных публикаций требуют разработки эффективных

методов автоматизации обработки и анализа научных документов. Применение интеллектуальных систем машинного обучения позволяет существенно повысить качество извлечения, структурирования и интерпретации информации, что является основополагающим фактором для научно-исследовательской деятельности. Применение искусственного интеллекта (ИИ) в данной области базируется на принципах обработки естественного языка (Natural Language Processing, NLP), методов глубокого обучения и алгоритмов кластеризации, что обеспечивает возможность выявления скрытых закономерностей в больших массивах данных. Систематизация и автоматизация процессов анализа научных текстов способствуют не только ускорению научного поиска, но и повышению объективности и воспроизводимости результатов исследований.

Актуальность применения ИИ в работе с научными документами обусловлена несколькими факторами. Во-первых, непрерывное увеличение объема научно-технической информации требует создания автоматизированных систем, способных оперативно обрабатывать данные, что уже нашло отражение в ряде современных исследований и разработок. Так, многие научные коллективы и технологические компании, такие как Google Scholar, Elsevier, Springer Nature, а также ведущие университеты, активно разрабатывают и внедряют алгоритмы машинного обучения для классификации, анализа цитируемости и выявления тематических трендов. Во-вторых, автоматизация рутинных процессов позволяет сократить время на подготовку и обработку информации, что особенно важно для специалистов в области научных исследований, библиотекарей и аналитиков, работающих с большими базами данных [1].

Сферы применения подобных решений весьма разнообразны. Они охватывают как научно-исследовательскую деятельность, так и образовательные процессы, управление библиотечными фондами, информационное сопровождение научных проектов и мониторинг инновационных технологий. Применение интеллектуальных систем позволяет обеспечить быстрый и точный поиск релевантной информации, оптимизировать процессы рецензирования и даже способствовать выявлению новых направлений исследований [2, 3]. Таким образом, разработка и внедрение методов автоматизации обработки научных документов становится необходимостью для университетов, исследовательских институтов, издательств и организаций, занимающихся анализом научных трендов.

В настоящей статье проводится анализ современных методов и алгоритмов автоматизации работы с научными документами с использованием систем машинного обучения. Рассмотрение теоретических основ применения ИИ, анализ существующих решений и обзор сфер применения закладывают фундамент для дальнейшего детального изучения конкретных методов и алгоритмов, способных обеспечить комплексное автоматизированное решение задач в сфере научной информации.

2. Работа с научными документами

Динамичное развитие информационных технологий стимулирует появление междисциплинарных методов анализа больших данных. Современные исследования направлены на интеграцию статистических моделей с алгоритмами машинного обучения, что позволяет не только повысить качество извлечения информации, но и оптимизировать процессы обработки данных с учетом специфики каждой научной дисциплины.

Применение искусственного интеллекта в работе с научными трудами и документами позволяет перевести большую часть обработки документов в автоматизированный процесс, который не требует постоянного вмешательства человека. Это позволяет повысить эффективность функционирования организации и снизить человеческие трудозатраты [4].

В научно-исследовательских институтах работа с научно-техническими документами представляет собой многоэтапный и трудоемкий процесс, требующий значительного вклада человеческих ресурсов. Сложность процесса заключается в необходимости проведения различных операций, таких как анализ, ознакомление, обработка, распределение и контроль, что требует внимания, времени и усилий сотрудников организации. Кроме того, постоянный рост объема документов и увеличение скорости обмена информацией создают дополнительные сложности в управлении документами. Совокупность этих факторов делает необходимым для руководителей, организаторов и разработчиков поиск эффективных решений для автоматизации процессов работы с научными трудами.

Применение технологий ИИ в данной сфере ориентировано на обработку научно-технических текстов, содержащих сложные терминологические конструкции, формулы, таблицы и специализированную лексику. В отличие от гуманитарных и социальных наук, где текстовые данные часто являются неструктурированными и вариативными, научно-технические документы обладают более четкой формальной структурой, включающей аннотации, ключевые слова, графические элементы и математические выражения. Это открывает возможности для использования специализированных алгоритмов обработки естественного языка и технологий машинного обучения, настроенных под специфику научно-технического стиля.

3. Решаемые задачи

Научно-технические документы обладают рядом специфических характеристик, которые необходимо учитывать при их автоматизированной обработке. Во-первых, они часто включают сложные структурные элементы, такие как формулы, таблицы, схемы, рисунки, что затрудняет их анализ традиционными алгоритмами обработки текста. Во-вторых, документы могут содержать специализированную терминологию, различную для каждой предметной области. В-третьих, различие в форматах публикаций (например, исследовательская

статья, патент, технический отчет) требует гибкого подхода к анализу текстов, способного адаптироваться к разным типам данных.

В рамках автоматизированной обработки научных документов в НИИ основное внимание уделяется решению следующих задач:

- Классификация научно-технических текстов.

Классификация является крайне важным аспектом обработки научной информации. Использование методов тематического моделирования и алгоритмов машинного обучения позволяет эффективно группировать документы по дисциплинам, направлениям исследований или типам публикаций. Например, можно выделить категории «физика», «химия», «инженерные науки» и далее детализировать их на подкатегории, такие как «квантовая механика», «органическая химия» и «авиационные технологии». Тематическое моделирование, в частности, позволяет выявить скрытые ассоциативно-семантические связи между терминами, документами и темами в корпусах текстов. Это достигается с помощью вероятностных моделей, таких как латентное размещение Дирихле (LDA), которое осуществляет «мягкую» кластеризацию, позволяя терминам относиться к разным темам в зависимости от контекстуального значения.

- Извлечение ключевых элементов текста.

Научно-технические документы часто содержат структурированную информацию, включающую аннотацию, ключевые слова и библиографические ссылки. Интеллектуальные системы обработки текста должны обеспечивать точное выделение этих элементов для создания метаданных и оптимизации научного поиска. Извлечение ключевых слов является важным аспектом этого процесса, поскольку оно позволяет суммировать содержание текста и распознавать основные темы, обсуждаемые в документе. Для этого используются различные методы, включая статистические, лингвистические и подходы машинного обучения.

- Генерация краткого содержания, аннотирование и резюмирование.

В связи с увеличением объемов публикаций исследователи сталкиваются с необходимостью быстрого ознакомления с большим количеством научных статей. Методы генерации кратких аннотаций на основе нейросетевых моделей, таких как BART и T5, позволяют автоматизировать процесс создания рефератов и увеличивать скорость анализа данных. Эти модели используются для создания кратких и информативных резюме, что облегчает ориентацию в огромных объемах научной информации. Кроме того, автоматическая генерация ключевых слов для аннотаций может существенно улучшить возможности поиска по статьям, особенно когда ключевые слова не указаны явно в тексте.

Ключевыми задачами автоматизации работы в НИИ является создание систем, способных классифицировать научные публикации по тематическим направлениям, типам документов (диссертации, отчеты, статьи, патенты, технические задания и т.д.), а также генерировать краткое содержание документа, сохраняющее общий смысл и отражающее ключевые темы. Нейросетевые модели, разрабатываемые в данном контексте, должны учитывать широкий спектр текстовых форматов, включая краткие аннотации, полнотекстовые научные статьи, технические отчеты и монографии. Кроме того, в рамках автоматизированных систем анализа необходимо учитывать сложность терминологического аппарата различных научно-технических дисциплин, что требует масштабных обучающих выборок и механизмов адаптации моделей к узкоспециализированным областям знания.

4. Существующие методы автоматизации работы с документами

Одним из ключевых аспектов автоматизации работы с любыми документами является распознавание текстового содержимого документов. При отсутствии текстового слоя пользователь лишается возможности выполнять базовые операции с документом, такие как выделение текста, копирование, поиск по словам, быстрая навигация. Это особенно актуально для отсканированных документов и изображений страниц, в которых текст представлен в виде графических данных. При высокой точности современных алгоритмов оптического распознавания символов (Optical Character Recognition, OCR), включая технологии на основе глубокого обучения, извлеченный текст из страниц или изображений позволяет в полной мере отразить содержание документа в цифровом формате. Это открывает возможность дальнейшей обработки текста с использованием других методов.

Внедрение инструментов распознавания текста является ключевой задачей для повышения эффективности современных информационных систем. Решение данной задачи позволяет существенно увеличить производительность обработки документов, сократить затраты времени и ресурсов, а также обеспечить более глубокую интеграцию с процессами управления, способствуя ускорению принятия решений. Такой подход особенно актуален в условиях цифровой трансформации организаций, где требования к оперативности и точности обработки данных становятся критически важными.

Другими полезными технологиями, автоматизирующими процессы работы с научными документами, являются интеллектуальный поиск и интеллектуальное сравнение. Технология интеллектуального поиска основана на использовании естественного языка и позволяющая пользователям формулировать свои поисковые запросы в наиболее удобной и интуитивной форме. Затем, с помощью алгоритмов обработки естественного языка, программа анализирует запрос и определяет документы, которые наиболее точно соответствуют описанным

в запросе условиям. Это обеспечивает более многогранный поиск информации, поскольку программа может понимать контекст и смысл запроса, а не просто искать отдельные ключевые слова.

Функциональность интеллектуального сравнения помогает находить различия между двумя версиями одного документа на всех этапах согласования в самом тексте документа, а также наличие или отсутствие визуальных признаков (картинок, подписей и прочего) [5]. Технология позволяет определить различия между документами разного формата, таких как doc, docx, pdf, djvu. Кроме того, она выявляет крупные структурные расхождения в документах — поля, отступы, колоннотулы и подобное. В результате сравнения формируется информативный отчет с результатами расхождений, позволяющий сотрудникам сосредоточиться только на измененных местах без необходимости вычитки всего текста. Это существенно экономит время и ресурсы организации, а также снижает риск ошибок и несоответствий в документах.

В условиях стремительного роста объема научных публикаций и разнообразия информационных источников перспективы дальнейших исследований в области автоматизации обработки научных документов представляют особую актуальность. Приоритетной задачей остается создание универсальных платформ, способных интегрировать разнородные данные и обеспечивать высокоточные алгоритмы классификации и семантического анализа. Важным направлением является разработка адаптивных систем, способных динамически обучаться на основе поступающих данных, используя методы активного обучения и механизмы обратной связи.

5. Ограничения современных методов

Обработка, анализ, управление документами, в том числе научными, служебными и другими, играют важнейшую роль в повседневной деятельности любых организаций, которые требуют значительных временных и человеческих ресурсов. С целью оптимизации этих процессов все чаще находят новые решения для повседневных и однообразных задач в работе с документами.

Однако, несмотря на существующие достижения в области автоматизации, остается ряд сложных технических потребностей, которые требуют особого внимания. К таким потребностям можно отнести: извлечение таблиц, которые не имеют четко выраженных границ ячеек; извлечение графических объектов, таких как рисунки и изображения, и анализ их содержания; распознавание рукописных текстов; устранение дефектов печатного текста, включая бледные отпечатки, надписи поверх текста или другие искажения. Эти задачи требуют применения гибридных технологий, включая машинное обучение, компьютерное зрение и алгоритмы обработки естественного языка, что подчеркивает важность комплексного подхода к автоматизации научного документооборота.

Ключевой нерешенной проблемой на сегодняшний день является отсутствие готового решения распределе-

ния научных документов по категориям рабочих проектов отделов, направлений или организаций без вмешательства оператора. Это связано с тем, что каждая организация имеет свои уникальные процессы и структуры работы с документами, которые не могут быть универсально охвачены существующими алгоритмами автоматизации.

Особое внимание следует уделить адаптации существующих алгоритмов под требования различных научных дисциплин. Методы, эффективные в технических и естественных науках, зачастую оказываются менее применимыми в гуманитарной сфере, где языковая неоднозначность и культурные контексты требуют специализированного подхода. Решение данной проблемы возможно через разработку универсальных алгоритмов, способных учитывать междисциплинарные особенности научных публикаций, что станет важным шагом для повышения качества обработки информации.

Таким образом, проблема эффективной организации ведения документов в современных учреждениях не теряет своей актуальности, что объясняется следующими причинами: существенное отставание темпов научно-технического прогресса от темпов нарастания потоков информации вообще и потоков научного и служебного оборота в конкретном учреждении; способы и методы обработки документов, которые имеются в ряде учреждений, ежедневно осуществляющих распределение значительного количества документов, являются достаточно не совершенными [6].

6. Предлагаемые подходы

Для повышения эффективности обработки научных документов и снижения трудозатрат сотрудников организаций предлагаются как внедрение новых методов, так и модернизация уже существующих технологий. Современные подходы активно используют нейросетевые модели, которые обеспечивают высокую точность и скорость выполнения задач, связанных с анализом и обработкой текстовой информации.

Среди новых методов можно выделить генерацию краткого содержания документов, реализуемую с помощью модели MBART («mbart_ru_sum_gazeta») [7], которая способна автоматически создавать содержательные резюме больших текстов, значительно ускоряя процесс ознакомления с документами. Также еще одним важным направлением является выделение именованных сущностей, например, названий организаций, дат, мест и других ключевых данных с использованием модели SpaCy («ru_core_news_lg»).

Кроме того, модернизация существующих технологий включает адаптацию таких моделей к специфическим требованиям организаций, настройку параметров обработки данных под конкретные сценарии использования и интеграцию их с уже функционирующими системами документооборота. Такой подход не только снижает нагрузку на сотрудников, но и способствует улучшению качества управления информацией, повышая общую производительность рабочих процессов.

Для дальнейшего повышения эффективности автоматизации целесообразно внедрение ансамблевых моделей, объединяющих результаты нескольких алгоритмов в единое решение. Такой подход позволяет компенсировать слабые стороны отдельных методов за счет их синергетического эффекта. Интеграция статистических методов, нейросетевых архитектур и экспертных оценок создает основу для построения адаптивных систем, способных динамично реагировать на изменения в объемах и характере научных данных [8].

7. Метод экспертных оценок

Автоматизация процессов классификации документов требует комбинирования различных методов и технологий для достижения высокой точности и эффективности. Гибридный подход, основанный на интеграции экспертных оценок, алгоритмов и нейросетевых моделей, обеспечивает возможность более точного анализа и обработки текстов. Такой подход объединяет преимущества традиционных и современных методов, минимизируя их недостатки.

Метод экспертных оценок (МЭО) позволяет выбрать лучший вариант объекта (системы) из множества альтернативных. Для многокритериального выбора лучшего варианта из множества альтернативных возникает вопрос учета степени важности критериев. Для решения этого вопроса «обращаются» к МЭО, главными фигурами которого являются эксперты (эксперты — высококвалифицированные специалисты в конкретной предметной области, способные по предъявляемой им исходной информации выработать обоснованные рекомендации). На основе экспертных оценок можно задавать веса различным признакам, что особенно важно при работе с алгоритмами, где требуется начальная разметка данных. Кроме того, такие оценки помогают учитывать противоречивые и несводимые друг к другу аспекты эффективности системы, задавая их относительную важность через специальные коэффициенты.

Поскольку эффективность функционирования любой системы характеризуется различными, часто противоречивыми и несводимыми один к другому аспектами ее функционирования, имеющими различную физическую природу, направления изменения и масштабы измерения, то для определения степени важности частных показателей эффективности их необходимо «взвешивать» специальными коэффициентами важности.

Расширяя традиционную модель экспертных оценок, целесообразно использовать методы многокритериального анализа, такие как метод иерархий (АНП) или метод TOPSIS [9]. Эти инструменты позволяют учитывать не только субъективное мнение специалистов, но и объективные показатели эффективности алгоритмов, что приводит к более точной калибровке весовых коэффициентов. Такой гибридный подход повышает надежность автоматизированных систем за счет комплексного анализа параметров и согласования результатов между различными методологиями.

Гибридный подход предполагает объединение

нескольких методов для достижения максимальной эффективности. Простейшие правила обработки, основанные на экспертных оценках, могут поддерживать работу высокоуровневой нейросети. Такой подход может быть реализован через механизмы, такие как «мягкое голосование» (soft voting), где решения различных моделей комбинируются для получения финального результата.

8. Нейронные сети с несколькими входами

Современные нейросетевые модели, такие как BERT или GPT, способны учитывать контекст, семантику и скрытые связи между словами, что значительно улучшает точность классификации. Гибридный подход, объединяющий экспертные оценки, традиционные алгоритмы и нейросетевые модели, представляет собой мощное решение для автоматизации классификации документов. Внедрение такого подхода должно обеспечить высокую точность и производительность, снизить трудозатраты и повысить качество работы организаций.

В работе [10] описываются разработка и применение многовходовой архитектуры нейронной сети. Моделью обрабатывается несколько входных данных различных форматов, несколько текстовых, графовый и формат даты времени. Архитектура разработанной модели дублирует архитектуру мультимодальных моделей с несколькими «головами», каждая голова принимает на вход свою структуру данных и на выходе выдает обработанное векторное представление. После этого все векторные представления объединяются посредством последовательного присоединения (конкатенации), затем объединенный вектор обрабатывается полносвязными слоями.

На рис. 1 показана архитектура многовходовой нейронной сети, в которой есть 4-е входа. Полный текст документа разбивается на два входа: ID токенов и маски внимания, затем добавляются входы с тегами зависимостей и метаинформацией. Полный текст документа разбивается на два входа: идентификатор токенов и маски внимания, затем добавляются входы с тегами зависимостей и метаинформацией.

Предварительно обученную модель BERT дообучали на наборе данных, состоящем из 5400 сообщений социальной сети Tweeter. Все описанные классы были разбиты строго равномерно по 900 сообщений в каждой категории. Разработанная структура позволяет комбинировать знания от предварительно обученной модели, семантических и лингвистических характеристик и метаинформации. За счет рассматривания текста с разных сторон и из разных модальностей модель способна замечать и выделять большее количество скрытых шаблонов в тексте. Проведенные эксперименты с использованием данной модели показали увеличение точности классификации от 3 % до 7 % по сравнению с аналогами.

Расширение архитектуры нейронной сети за счет интеграции дополнительных источников информации, например, структурированных метаданных и временных характеристик, позволит не только объединять дан-

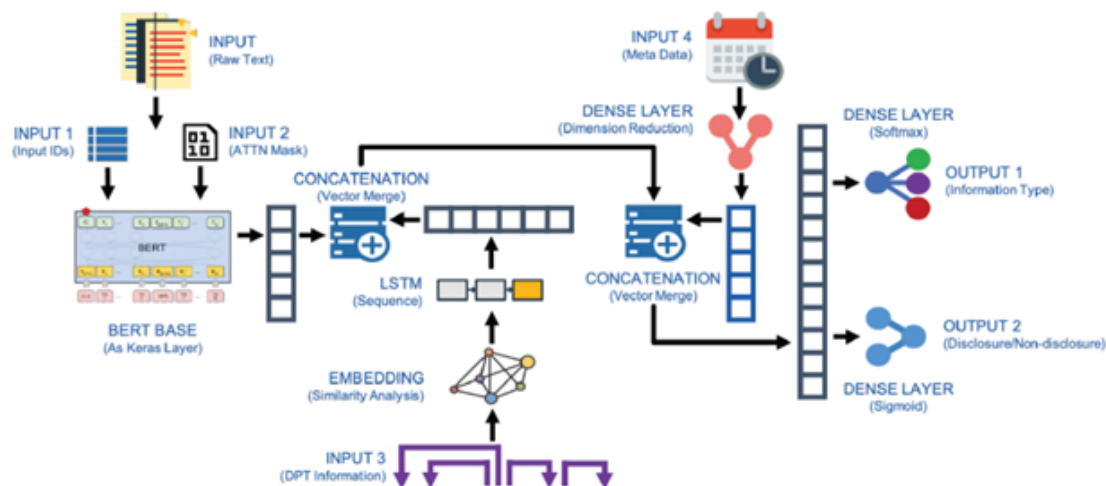


Рис. 1. Схема архитектуры многовходовой нейронной сети

ные различных типов, но и значительно повышать точность извлечения смысловых зависимостей из сложных научных текстов. Более того, применение механизмов внимания и динамического обновления параметров в сочетании с многовходовыми структурами позволит модели лучше адаптироваться к изменяющимся условиям и поспособствует более детальному анализу контекстных связей, что особенно важно при обработке длинных и структурно насыщенных документов. Такой подход обеспечивает более точное выделение ключевых информационных сегментов, что в итоге приводит к улучшению показателей эффективности работы с документами, а также результатов классификации и семантического анализа.

9. Проводимые исследования

В рамках исследования по возможности создания гибридной интеллектуальной архитектуры для автоматизации работы с научными документами планируется провести сравнительный анализ методов реферирования, основанных на классических лингвистических подходах, и современных нейросетевых методов. Целями данного анализа является выявление ключевых особенностей и определение преимуществ каждого из подходов в контексте автоматического создания кратких содержательных резюме научных текстов. Традиционные лингвистические методы, опирающиеся на формальные грамматические правила, синтаксический и морфологический разборы, зарекомендовали себя как надежные инструменты при работе с текстами, обладающими устойчивой структурой и предсказуемыми шаблонами изложения. Однако, несмотря на их теоретическую обоснованность и прозрачность алгоритмических решений, данные методы зачастую испытывают трудности при обработке больших объемов данных и при наличии сложных семантических конструкций, характерных для современных научных публикаций.

Современные нейросетевые методы, в свою очередь, используют возможности глубокого обучения, трансформеров и эмбедингов для извлечения семанти-

ческой информации, что позволяет учитывать контекст и скрытые взаимосвязи между фрагментами текста. Такие методы демонстрируют высокую адаптивность и способность обрабатывать как структурированные, так и неструктурированные данные, однако требуют значительных вычислительных ресурсов и обширных обучающих выборок для достижения оптимальной точности. В рамках исследования будет проведен детальный анализ эффективности обоих подходов с использованием метрик качества реферирования, таких как полнота, точность и когерентность получаемых резюме. Также планируется проведение опытов на различных предметных областях, чтобы определить условия, при которых один метод может быть предпочтительнее другого.

Анализ будет осуществляться на базе экспериментальных данных, полученных при обработке репрезентативного корпуса открытых научных публикаций.

10. Заключение

Подводя итоги, можно отметить, что интеграция междисциплинарных методов и технологий в области автоматизации работы с научными документами открывает широкие перспективы для дальнейших исследований. Применение нейросетевых моделей в сочетании с методами ансамблевого анализа и экспертных оценок способствует созданию адаптивных систем, способных к самокоррекции и обучению на основе новых данных. Стандартизация процессов и межсекторное сотрудничество станут ключевыми факторами для повышения эффективности научного документооборота и качества анализа информации в будущем.

Применение гибридного подхода автоматизации процессов работы с научными документами, который включает модель представления кластеров документов, должно обеспечить значительный прирост точности обработки. Использование данного подхода открывает возможности для автоматизации анализа больших объемов данных, минимизации ошибок, связанных с ручным вводом, и повышения прозрачности научной работы.

Список литературы / References

- [1] de la Torre-López J, Ramírez A, Romero JR. Artificial intelligence to automate the systematic review of scientific literature. *Computing*. 2023;105:2171–2194. DOI: 10.1007/s00607-023-01181-x
- [2] Мирошниченко МА, Абдуллаева АА, Вовк МА. Тенденции развития технологий искусственного интеллекта в управлении документами организации. *Естественно-гуманитарные исследования*. 2024;(51):398–403. Miroshnichenko MA, Abdullaeva AA, Vovk MA. Development trends of artificial intelligence technologies in organizational document management. *Natural-Humanitarian Studies*. 2024;(51):398–403 (in Russian). EDN: ypbwdd
- [3] Кубарский А В. Технологии искусственного интеллекта при построении самообучающейся системы электронного документооборота. *Эпомен*. 2021;(61):49–55. Kubarskiy AV. Intelligence technologies in the construction of a self-learning electronic document management system. *Epomen*. 2021;(61):49–55 (in Russian). EDN: vjtnhp
- [4] Перова МВ, Сибилева АА. Искусственный интеллект в системах электронного документооборота. *Тенденции развития науки и образования*. 2022;(81-2):33–36. Perova MV, Sibileva AA. Artificial Intelligence in Electronic Document Management Systems. *Trends in Science and Education Development*. 2022;(81-2):33–36 (in Russian). DOI: 10.18411/trnio-01-2022-50
- [5] Белов ИИ. Автоматизация функций систем электронного документооборота посредством применения технологий искусственного интеллекта. *Вестник архивиста*. 2022;(3):772–783. Belov II. Automation of Electronic Document Management Systems Functions by Means of Artificial Intelligence Technologies. *Herald of an archivist*. 2022;(3):772–783 (in Russian). DOI: 10.28995/2073-0101-2022-3-772-783
- [6] Павкина НН. Организация документооборота в современных учреждениях. *Актуальные проблемы современности: наука и общество*. 2020;(1(26)):12–16. Pavkina NN. Document management in the present-day institutions. *Current issues of our time: science and society*. 2020;(1(26)):12–16. EDN: cleivn
- [7] IlyaGusev/mbart_ru_sum_gazeta Hugging Face. [online] URL: https://huggingface.co/IlyaGusev/mbart_ru_sum_gazeta (Accessed 1.11.2024).
- [8] Zhao J, Guo Y, Zhanget H, et al. Improved machine learning estimation of surface turbulent flux using interpretable model selection and adaptive ensemble algorithms over the Horqin Sandy Land area. *Atmospheric Research*. 2025;316:107952. DOI: 10.1016/j.atmosres.2025.107952
- [9] Menon RR, Ravi V. Using AHP-TOPSIS methodologies in the selection of sustainable suppliers in an electronics supply chain. *Cleaner Materials*. 2022;5:100130. DOI: 10.1016/j.clema.2022.100130
- [10] Mehdy AKMN, Mehrpouyan HA. Multi-Input Multi-Output Transformer-Based Hybrid Neural Network for Multi-Class Privacy Disclosure Detection. *2nd International Conference on Machine Learning Techniques and NLP (MLNLP 2021), September 18–19, 2021, Copenhagen, Denmark*. CS & IT Conference Proceedings. 2021. p. 221–241. DOI: 10.5121/csit.2021.111419

Сведения об авторах / Information about the Authors

Вадим Николаевич Гокарев

27 ЦНИИ, Москва, Россия

Vadim N. Gokarev

The 27th Central Research Institute, Moscow, Russia

vgokarev@mail.ru

ORCID: 0009-0002-4082-2925

Александр Федорович Беззубов

кандидат технических наук

27 ЦНИИ, Москва, Россия

Alexander F. Bezzubov

Ph.D. (Techn.)

The 27th Central Research Institute, Moscow, Russia

b17a52f@yandex.ru

Дмитрий Валерьевич Березкин

кандидат технических наук

МГТУ им. Н.Э. Баумана, Москва, Россия

Dmitry V. Berezkin

Ph.D. (Techn.)

Bauman Moscow State Technical University, Moscow, Russia

dmitryb2007@yandex.ru

ORCID: 0000-0001-7149-6072