ISSN 2782-2001
Системы анализа и обработки данных
том 101, № 1, 2026, с. 59–70

http://journals.nstu.ru/vestnik
Analysis and data processing systems
Vol. 101, No. 1, 2026, pp. 59–70

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И ТЕЛЕКОММУНИКАЦИИ

INFORMATION TECHNOLOGIES AND TELECOMMUNICATIONS

УДК 004.451.075

DOI: 10.17212/2782-2001-2026-1-59-70

Разработка системы автоматического распознавания типа документа по прикрепленному сообщению в СЭД*

А.В. МОРГУНОВ

630102, г. Новосибирск, ул. Кирова, 86, Сибирский государственный университет телекоммуникаций и информатики

all122001@mail.ru

Проект направлен на создание автоматической системы, основанной на типовых документах, включающей документооборот и использующей методы машинного обучения и трансформерные нейронные сети. Основная цель разработки заключается в повышении качества и скорости обработки документации, снижении нагрузки на сотрудников и предотвращении ошибок, возникающих при индивидуальном подходе, для достижения которой поставлена задача по реализации комплексной архитектуры, включающая предобработку данных, дообучение моделей, а также измерение точности результатов по метрикам «точность» и «F1-мера», что позволяет объективно измерять эффективность метода.

Ключевой формулой системы является применение современных языковых моделей на основе BERT[1] и трансформерного соединения, ориентированного на контекстный анализ текста. Использование моделей GPT[2] обеспечивает возможность адаптации решений под различные форматы документов и специфику единого документооборота. Формирование эмбеддингов[3] осуществляется с учетом позиционных признаков обработки механизмом многоголового внимания, что обеспечивает получение контекстных представлений и последующую классификацию категории документа [1–3].

Предложенная система позволяет анализировать не только текстовое правило, но и такие параметры, как метаданные, стиль оформления и структурные элементы, которые повышают точность определения типа файлов при их использовании пользователями. Автоматизация данного процесса позволяет сократить время на регистрацию документов в СЭД, минимизировать возможные человеческие ошибки, обеспечить соответствие нормам организации и повысить эффективность управления информационными потоками.

---

* *Статья получена 16 октября 2025 г.*

[1] BERT (Bidirectional Encoder Representations from Transformers) – двунаправленные презентации кодировщика для трансформеров.

[2] GPT (Generative Pre-trained Transformer) – генеративный предобученный трансформер.

[3] Эмбеддинг (англ. *embedding*) – представление данных (текста, изображений, категорий) в виде плотных числовых векторов фиксированной размерности в многомерном пространстве.

Полученные результаты демонстрируют высокую перспективность применения трансформерных технологий в задачах интеллектуальной обработки документов и подтверждения возможности расширения функционала системы для управления крупными корпоративными и значимыми платформами.

Ключевые слова: распознавание документов, машинное обучение, трансформерные нейронные сети, автоматизация документооборота, классификационные тексты, предобработка данных, классификация документов, надежность, математическая модель, квадратичная модель, BERT, GPT, цифровой документооборот

ВВЕДЕНИЕ

Современные изменения в организациях приводят к резкому увеличению объемов электронных документов, циркулирующих между подразделениями и выездными контрагентами. Особую оригинальность приобретают процессы классификации документов, поскольку правильное определение их категории влияет на скорость дальнейшей обработки, полноту регистрации и соблюдение нормативных требований. Традиционные методы ручного введения типа документа оказываются неэффективными в условиях усиленного документооборота: сотрудники затрачивают время на рутинные операции, возможны ошибки, приводящие к нарушению бизнес-процессов и росту операционных рисков [2–4]. Поэтому актуальной становится проблема интеллектуальных систем, способов автоматического определения типа документа по содержанию, последовательностей и метаданных прикрепленных файлов. Применение алгоритмов машинного обучения и нейронных сетей открывает возможности для создания адаптивных классификаторов, возможностей изучения практики и учета доменных особенностей конкретной организации.

Особый интерес представляет структура трансформеров, продемонстрировавшая высокую эффективность в задачах обработки естественного языка (НЛП). Благодаря механизму самовнимания[4] модели такие возможности позволяют научиться сложным контекстным связям между элементами текста, что позволяет повысить качество категоризации документов. В рамках работы предложен вариант использования русскоязычных предобученных моделей модулей GPT и BERT, а также подходов к их дообучению на специализированных корпусах. Разработка и внедрение системы автоматического управления видами документов позволит минимизировать человеческий фактор, ускорить бизнес-процессы, снизить затраты на обработку данных и повысить информационную безопасность. Это делает данное исследование востребованным как для коммерческих структур, так и для государственных организаций [1, 5].

1. ПОСТАНОВКА ЗАДАЧИ

Целью исследования является разработка программного решения для определения типа документа на основе анализа текста прикрепленных файлов с применением методов глубокого обучения.

[4] Самовнимание (англ. *self-attention*) – это механизм, позволяющий нейронной сети оценивать важность различных частей входной последовательности по отношению друг к другу.

Для достижения поставленной цели необходимо решить следующие задачи.

1. Сформировать и подготовить обучающие данные, включающие большое количество документов различных типов, обеспечив репрезентативность выбора.

2. Выполнить предобработку: очистку текста, устранение спецсимволов, токенизацию[5], нормализацию и приведение структуры данных к формату, пригодному для обучения нейронных сетей [1].

3. Выбрать и адаптировать модели глубокого обучения, способы выполнения текстов классификации: модели модулей GPT для генерации обучающих примеров и расширения корпуса и модели BERT для решения задач классификации.

4. Обучить трансформерную архитектуру с использованием размеров типовых документов, обеспечив правильное обновление весов.

5. Повысить качество классификации и обеспечить его объективную оценку по метрикам точности и F1-меры.

6. Разработать прототип системы, автоматически определяющий название и вид документа при его использовании пользователем в СЭД [3–6].

Решение этих задач позволяет реализовать разумную интеллектуальную систему, способную сократить время обработки документации, уменьшить количество ошибок при классификации и обеспечить повышение эффективности всего оборота.

2. АЛГОРИТМ РАСПОЗНАВАНИЯ ОБРАЗОВ

Построение рассматриваемой функции реализуется в два этапа: генерация обучающей и тестовой выборок с использованием модели GPT [7–10] и последующая классификация на основе BERT [10–14], включая модель «gpt-4o-mini» и модель «ai-forever/ruT5-base». Архитектура трансформерной нейронной сети представлена на рис. 1.

Выбор трансформерной архитектуры обусловлен ее способностью эффективно обрабатывать длинные текстовые последовательности и учитывать контекст на уровне всего документа, что особенно важно для задач классификации служебной и организационно-распорядительной документации. В отличие от традиционных методов (TF-IDF[6], наивный байесовский классификатор[7]), трансформеры позволяют учитывать семантические и синтаксические зависимости между фрагментами текста, что повышает устойчивость классификации при вариативности формулировок документов.

---

[5] Токенизация (англ. *tokenization*) – процесс разбиения текста на токены (слова, подслова или символы), используемые в качестве входных единиц моделей обработки естественного языка.

[6] TF-IDF (англ. *Term Frequency – Inverse Document Frequency*) – статистический метод числового представления текста, отражающий важность слова в документе относительно корпуса документов.

[7] Наивный байесовский классификатор (англ. *Naive Bayes classifier*) – вероятностный алгоритм классификации, основанный на теореме Байеса и предположении независимости признаков.

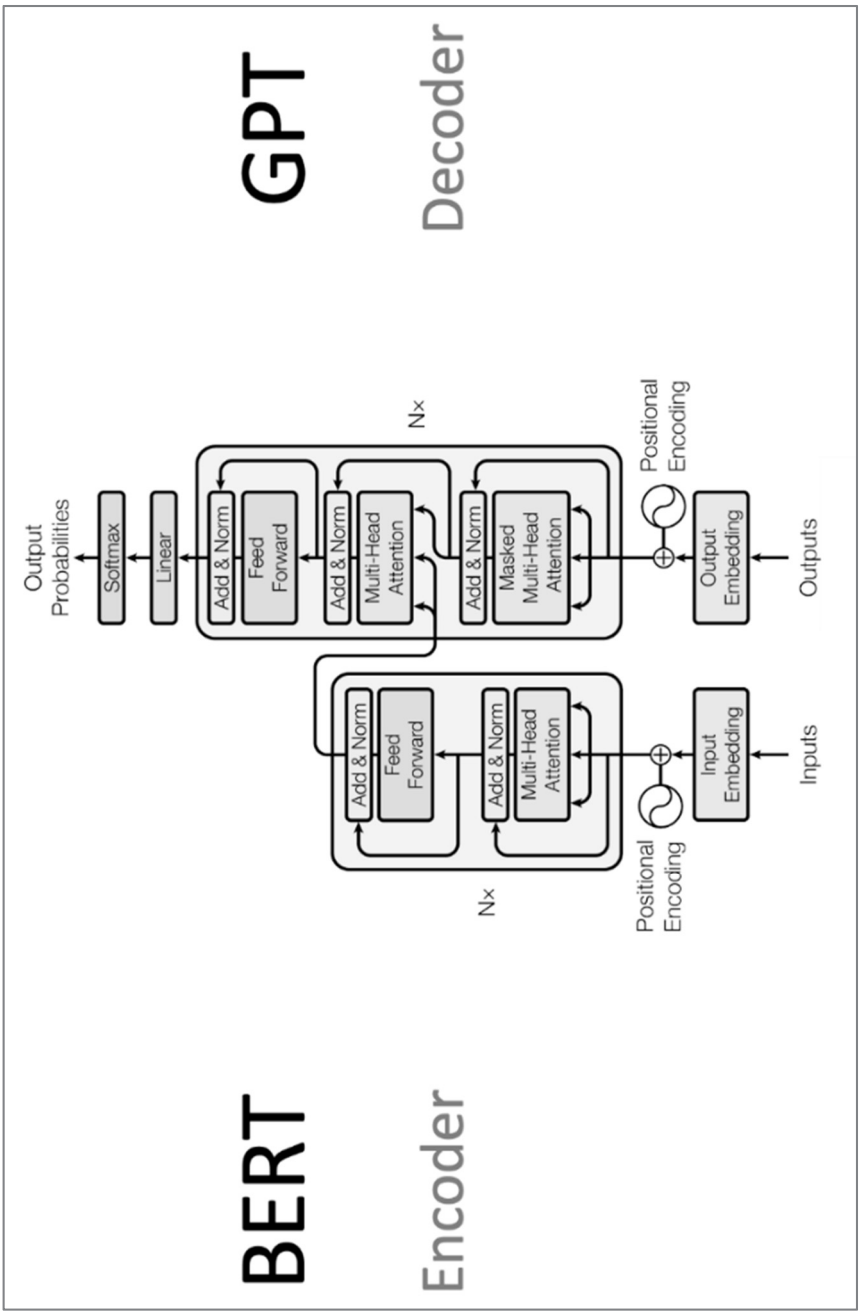

Рис. 1. Архитектура полноценной трансформерной нейронной сети

Fig. 1. Architecture of a full-fledged transformer neural network

Кодировщик

- Input Embedding: преобразует входные слова в векторы вложений, представляющие их семантическое значение.
- Multi-Head Attention: улавливает связи между словами в последовательности, взвешивая важность каждого слова для последующих позиций.
- Positional Encoding: добавляет информацию о порядке слов в последовательность вложений, так как порядок важен в естественном языке.
- Feed-Forward Network: выполняет нелинейные преобразования, извлекая более сложные паттерны из встроенных векторов.
- Add Norm: добавляет гауссовский шум к входящим данным для стабилизации обучения.

Декодировщик

- Masked Multi-Head Attention: аналогично кодировочному вниманию, но с маской, предотвращающей просмотр будущих слов при генерации.
- Multi-Head Attention на кодировке: учитывает соответствия между входной последовательностью и генерируемым текстом.
- Output Embeddings: выводит встроенные представления для входящих данных, используется в задачах классификации и рекомендаций для преобразования категориальных данных в цифровые представления.
- Feed-Forward Network: как в кодировщике.
- Add Norm: как в кодировщике.Linear (выполняет линейное преобразование входящих данных, умножая их на заданную матрицу весов и смещение). Используется для создания плотных слоев в нейронных сетях.
- Softmax: выводит вероятностное распределение, нормализуя выходы линейного слоя. Используется в задачах классификации для вычисления вероятностей принадлежности каждого примера к различным классам.

Алгоритм классификации

1. Ввод текста. Входной текст преобразуется в последовательность токенов[8] – элементарных текстовых единиц (слов или их частей), представленных в числовом виде и подаваемых на вход модели BERT.
2. Векторные представления токенов. Каждый токен преобразуется в числовой вектор (эмбеддинг), отражающий его смысловое значение.
3. Сложение позиций. К векторам токенов добавляется информация об их порядке в тексте, что позволяет модели учитывать структуру последовательности.
4. Трансформерные слои. Векторные представления токенов последовательно обрабатываются в нескольких слоях трансформера BERT. В этих слоях используется механизм самовнимания, позволяющий учитывать связи между словами в тексте.
5. Контекстные представления. После обработки формируются скрытые состояния – векторы, содержащие информацию о значении каждого слова с учетом его окружения.

[8] Токен (англ. *token*) – минимальная единица текста (слово, часть слова или символ), используемая в качестве элемента входной последовательности в моделях обработки естественного языка.

6. Классификация. Полученные представления передаются в выходной классификационный слой (полносвязный слой с функцией softmax[9]), который определяет класс текста [13–17].

7. Обучение модели. Параметры модели BERT настраиваются на размеченной обучающей выборке путем минимизации функции потерь между предсказанными и истинными классами.

Дополнительно в рамках работы использовался подход дообучения предобученной модели BERT на специализированном корпусе организационно-распорядительных документов. Такой подход позволяет сохранить языковые знания, полученные на больших корпусах, и адаптировать модель к доменной терминологии конкретной организации. Генеративная модель GPT применялась для расширения обучающей выборки путем синтетической генерации текстов документов с контролируемыми типами, что позволило снизить проблему дисбаланса классов.

3. ТЕСТИРОВАНИЕ

В итоге на 15-й эпохе получились метрики, представленные на рис. 2.

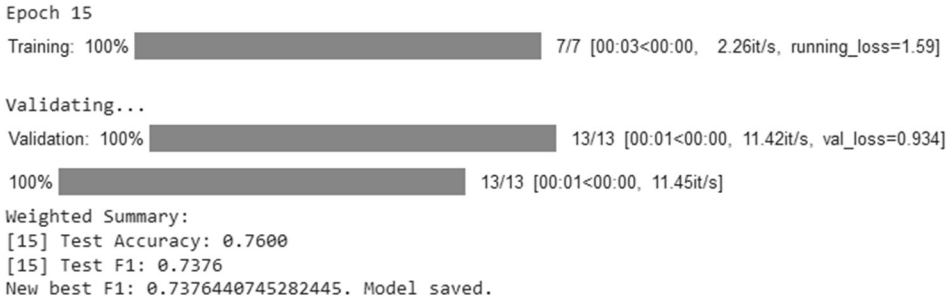

Рис. 2. Результат дообучение модели BERT

Fig. 2. BERT model retraining result

В качестве метрик были выбраны accuracy и F1-мера.

Точность (Accuracy) измеряет долю правильных предсказаний из всех предсказаний. Это простая и понятная метрика, которую легко интерпретировать. F1-мера учитывает как точность, так и полноту. Она полезна в несбалансированных наборах данных, где распределение классов неравномерно.

Использование F1-меры вместе с точностью предоставляет комплексный взгляд на производительность модели.

Точность показывает, насколько хорошо модель предсказывает правильные классы.

F1-мера показывает, насколько хорошо модель идентифицирует все экземпляры данного класса, включая как истинные положительные, так и истинные отрицательные.

[9] Softmax (англ. *softmax function*) – функция активации, преобразующая числовые значения выходного слоя нейронной сети в вероятностное распределение по классам, сумма вероятностей которого равна единице.

В ходе эксперимента обучение проводилось на размеченном корпусе документов, включающем несколько категорий (приказы, служебные записки, договоры, письма, акты и др.). Данные были разделены на обучающую, валидационную[10] и тестовую выборки в пропорции 70/15/15. Обучение осуществлялось с использованием оптимизатора AdamW[11] и линейного расписания скорости обучения с прогревом, что обеспечило стабильную сходимость модели.

По результатам тестирования на 15-й эпохе достигнуты следующие значения: точность = 0,76 и F1 = 0,7376, что свидетельствует о достаточно высоком качестве классификации типов документов. Полученные показатели подтверждают способность модели корректно различать классы документов при вариативности формулировок и структуры текстов. Выбранное число эпох обучения обеспечило достижение устойчивых метрик без признаков деградации качества на валидационной выборке.

Комбинируя эти две метрики, можно получить более полное представление о производительности модели и выявить потенциальные проблемы, такие как смещение в сторону доминирующего класса.

Кроме того, F1-мера особенно полезна для задач, где важна точная классификация нескольких классов, таких как обнаружение объектов, сегментация изображений и классификация естественных языков.

Полученные значения метрик показали, что трансформерная модель BERT после дообучения демонстрирует высокую способность к различению типов документов даже при схожести формулировок и структуры текстов. Это подтверждает применимость выбранного подхода для автоматизации классификации документов в системах электронного документооборота.

4. ПРИМЕР ВНЕДРЕНИЯ

В процессе документооборота допускается внесение дополненных атрибутов документа, включая его наименование или тип. Разработанная система позволяет решить заданную задачу по счетчику автоматического анализа содержания прикрепленного файла. После организации документа соответствующие поля формы регистрации закрепляются автоматически.

Интеграция разработанного классификатора в систему электронного документооборота осуществляется на уровне сервиса обработки вложений. После загрузки пользователем файла запускается модуль извлечения текста, далее текст передается в модель классификации, которая возвращает вероятности принадлежности документа к каждому типу. На основе максимальной вероятности система автоматически выбирает тип документа и заполняет атрибут «Вид документа» в карточке регистрации.

Такой подход обеспечивает повышение полноты данных, уменьшение количества вводимых ошибок и сокращение времени обработки документа в СЭД.

---

[10] Валидационная выборка (*validation set*) – часть размеченных данных, используемая для оценки качества модели и подбора ее параметров в процессе обучения.

[11] AdamW – оптимизатор градиентного спуска, модификация Adam с раздельным учетом весовой регуляризации (*weight decay*), применяемый при обучении нейронных сетей.

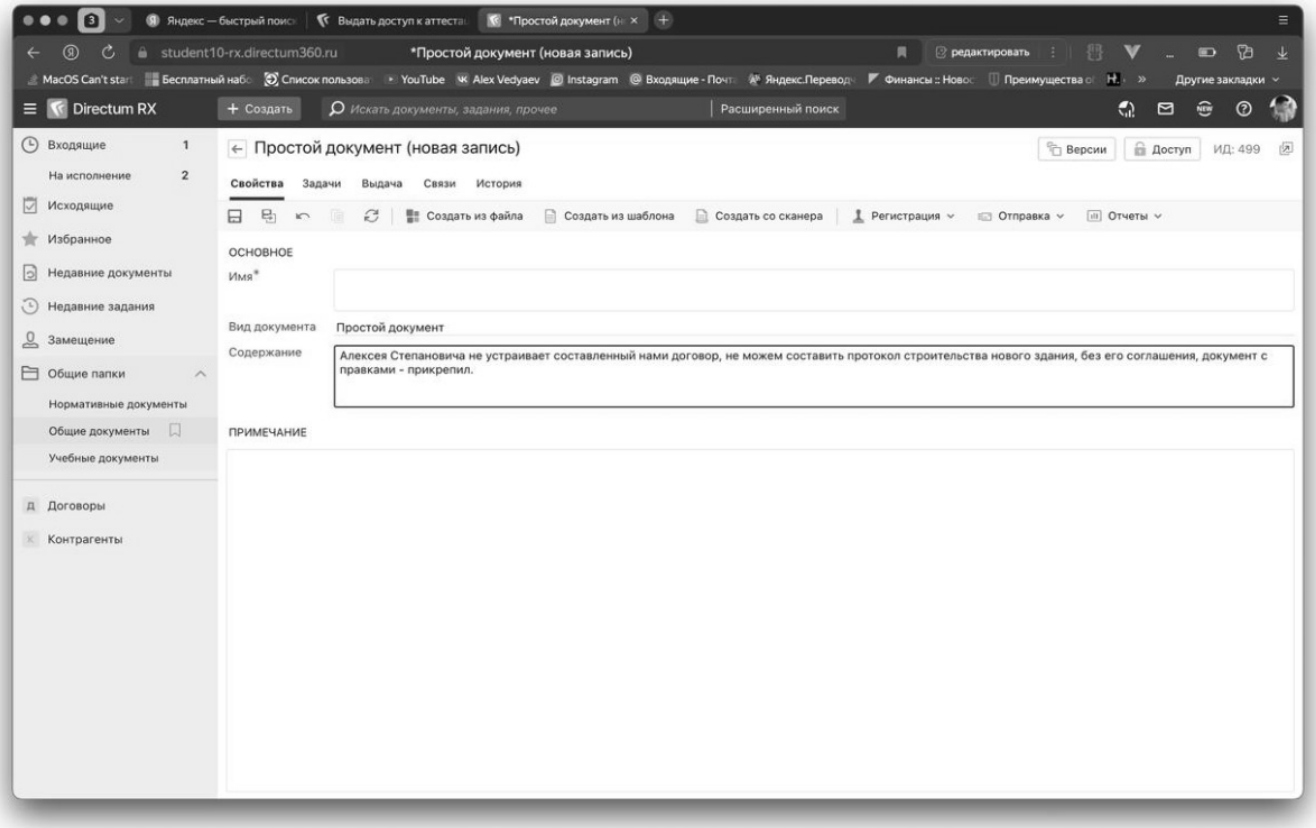

Рис. 3. Интерфейс создания отправки согласования документа

Fig. 3. Interface for creating a document approval submission

После загрузки файла достигается его автоматическая классификация, в форму автоматически подставляется документ с наименованием.

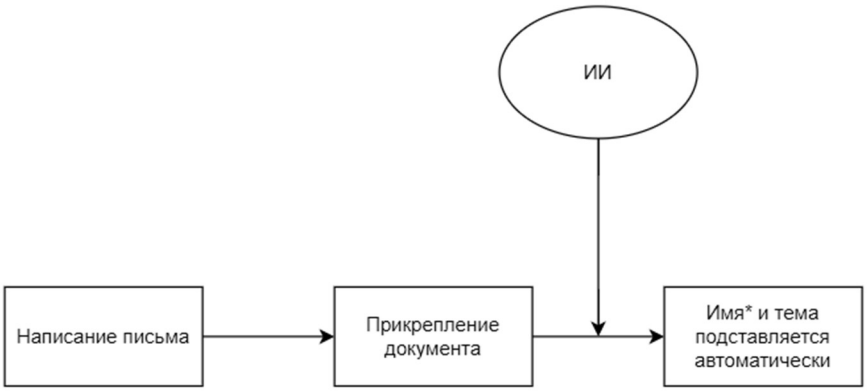

Рис. 4. Схема работы обновленного процесса с помощью ИИ

Fig. 4. A diagram of the updated AI-powered process

Предложенное решение позволяет использовать методы машинного обучения в процессах обеспечения документооборота с помощью значительных изменений. Автоматизация определения типов документов обеспечивает повышение скорости регистрации, формирует единообразие данных и обеспечивает соблюдение внутренних нормативов при обработке информационных материалов.

ЗАКЛЮЧЕНИЕ

Разработка системы принятия типа документа по прикрепленному сообщению доказала свою эффективность в условиях реального продолжения документооборота. Проведенные эксперименты подтвердили высокую точность классификации для различных форматов файлов, что позволяет использовать этот стандарт для практического применения.

Предложенная система выгодно отличается от известных решений и обеспечивает анализ содержания документа без предварительной ручной разметки его атрибутов, что позволяет снизить нагрузку на пользователя и уменьшить количество ошибок при регистрации. Применение трансформерных моделей обеспечивает адаптивность алгоритма и возможность к масштабированию при расширении перечня типов документов. Кроме того, архитектурное решение допускает интеграцию в СЭД без существенных изменений.

Таким образом, разработанный подход обеспечивает высокую перспективность для дальнейшего развития автоматизации документооборота, в том числе десятков интеллектуальных платформ и государственных информационных систем. Его дальнейшее совершенствование может включать расширение набора анализируемых параметров документа, внедрение технологий активного обучения и повышение устойчивости классификации для специализированных доменов.

СПИСОК ЛИТЕРАТУРЫ

1. *Бурков А.* Машинное обучение без лишних слов. – СПб.: Питер, 2020. – 192 с.

2. *Жерон О.* Прикладное машинное обучение с помощью Scikit-Learn, Keras и Tensor Flow: концепции, инструменты и техники для создания интеллектуальных систем. – СПб.: Диалектика, 2020. – 1040 с.

3. *Рашка С., Мирджалили В.* Python и машинное обучение: машинное и глубокое обучение с использованием Python, Scikit-Learn и TensorFlow. – СПб.: Диалектика, 2020. – 948 с.

4. *Грас Д.* Data Science. Наука о данных с нуля. – 2-е изд. – СПб.: БХВ-Петербург, 2021. – 416 с.

5. *Уатт Дж., Борхани Р., Катсаггелос А.* Машинное обучение: основы, алгоритмы и практика применения. – СПб.: БХВ-Петербург, 2022. – 640 с.

6. *Чжен Э., Казари А.* Машинное обучение. Конструирование признаков: принципы и техники для аналитиков. – М.: Бомбора, 2022. – 240 с.

7. *Плас Дж. В.* Python для сложных задач: наука о данных и машинное обучение / пер. с англ. И. Пальти. – СПб.: Питер, 2018. – 576 с.

8. *Мюллер А., Гвидо С.* Введение в машинное обучение с помощью Python: руководство для специалистов по работе с данными. – СПб.: Альфа-книга, 2017. – 480 с.

9. *Дейтел П., Дейтел Х.* Python: Искусственный интеллект, большие данные и облачные вычисления. – СПб.: Питер, 2020. – 864 с.

10. *Харрисон М.* Машинное обучение: карманный справочник: краткое руководство по методам структурированного машинного обучения на Python. – СПб.: Диалектика, 2020. – 320 с.

11. *Элбон К.* Машинное обучение с использованием Python. Сборник рецептов. – СПб.: БХВ-Петербург, 2019. – 384 с.

12. *Шалев-Шварц Ш., Бен-Давид Ш.* Идеи машинного обучения: от теории к алгоритмам. – М.: ДМК Пресс, 2019. – 436 с.

13. *Deisenroth M.P., Faisal A.A., Ong Ch.S.* Mathematics for machine learning. – Cambridge: Cambridge University Press, 2020. – 398 p.

14. *Фофанов О.Б.* Алгоритмы и структуры данных: учебное пособие / Томский политехнический университет. – Томск: Изд-во ТПУ, 2014. – 126 с.

15. Алгоритмы. Построение и анализ / Т. Кормен, К. Штайн, Р. Ривест, Ч. Лейзерсон. – М.: Диалектика, 2019. – 1324 с.

16. *Кормен Т.* Алгоритмы: вводный курс. – М.: Вильямс, 2014. – 208 с.

17. *Бабичев С.Л.* Лекции по алгоритмам и структурам данных. – М.: МАКС Пресс, 2019. – 344 с.

Моргунов Александр Владимирович, кандидат технических наук, доцент кафедры математического моделирования и цифрового развития бизнес-систем Сибирского государственного университета телекоммуникаций и информатики E-mail: all122001@mail.ru

Morguov Aleksandr V., PhD (Eng.), Associate Professor, Department of Mathematical Modeling and Digital Development of Business Systems, Siberian State University of Telecommunications and Informatics. E-mail: all122001@mail.ru

DOI: 10.17212/2782-2001-2026-1-59-70

Development of a system for automatic document type recognition based on attached messages in the EDMS[*]

A.V. MORGUNOV

Siberian State University of Telecommunications and Informatics, 86 Kirov Street, Novosibirsk, 630102, Russian Federation

all122001@mail.ru

Abstract

The project aims at creating an automated system based on standard documents, including document flow, and employing machine learning methods and transformer neural networks. The primary goal of the development is to improve the quality and speed of document processing, reduce the workload on employees, and prevent errors that arise with individual approaches. To achieve these objectives, a comprehensive architecture is being implemented, including data preprocessing, additional model training, and measurement of the accuracy of results using the Precision and F1-score metrics, enabling objective measurement of the method's effectiveness.

The key formula of the system is the use of modern language models based on BERT (Bidirectional Encoder Representations from Transformers) and a transformer connection focused on contextual text analysis. The use of gpt-4o-mini and ai-forever/ruT5-base models enables the adaptability of solutions to various document formats and the specifics of a unified document flow. During the process of assessing the quality of embeddings, the addition of positional features, processing by a multi-headed attention mechanism, evokes contextual representations, and subsequent determination of the document category. The proposed system analyzes not only the text rule but also parameters such as metadata, formatting style, and structural elements, which improve the accuracy of file type detection when used by users. Automating this process reduces the time required to register documents in the EDMS, minimizes potential human errors, ensures compliance with organizational standards, and improves the efficiency of information flow management.

These results demonstrate the high potential of using transformer technologies in intelligent document processing and confirm the potential for expanding the system's functionality to manage large corporate and high-value platforms.

Keywords: document recognition, machine learning, transformative neural networks, document automation, classification texts, data preprocessing, document classification, reliability, mathematical model, quadratic model, BERT, GPT, digital document management

REFERENCES

1. Burkov A. *Mashinnoe obuchenie bez lishnikh slov* [Machine Learning Without the Babbles]. St. Petersburg, Piter Publ., 2020. 192 p.

2. Géron A. *Prikladnoe mashinnoe obuchenie s pomoshch'yu Scikit-Learn, Keras i Tensor Flow: kontseptsii, instrumenty i tekhniki dlya sozdaniya intellektual'nykh sistem* [Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: concepts, tools, and techniques to build intelligent systems]. St. Petersburg, Dialectika Publ., 2020. 1040 p. (In Russian).

3. Raschka S., Mirjalili V. *Python i* m*ashinnoe obuchenie: mashinnoe i glubokoe obuchenie s ispolzovaniem Python, Scikit-Learn i TensorFlow* [Python Machine Learning: machine and deep learning with Python, Scikit-Learn, and TensorFlow]. St. Petersburg, Dialectika Publ., 2020. 948 p. (In Russian).

4. Grus J. *Data Science. Nauka o dannykh s nulya* [Data science from scratch: first principles with Python]. 2nd ed. St. Petersburg, BHV-Petersburg Publ., 2021. 416 p. (In Russian).

[*] *Received 16 October 2025.*

5. Watt J., Borhani R., Katsaggelos A. *Mashinnoe obuchenie: osnovy, algoritmy i praktika primeneniy* [Machine learning refined: foundations, algorithms, and applications]. St. Petersburg, BHV-Petersburg Publ., 2022. 640 p. (In Russian).

6. Zheng A., Casari A. *Mashinnoe obuchenie. Konstruirovanie priznakov: printsipy i tekhniki dlya analitikov* [Feature engineering for machine learning: principles and techniques for data scientists]. Moscow, Bombora Publ., 2020. 240 p. (In Russian).

7. Plas J.V. *Python dlya slozhnykh zadach: nauka o dannykh i mashinnoe obuchenie* [Python data science handbook: essential tools for working with data]. St. Petersburg, Piter Publ., 2018. 576 p. (In Russian).

8. Müller A., Guido S. *Vvedenie v mashinnoe obuchenie s pomoshu* Python*: rukovodstvo dly specialistov po rabote c dannymi* [Introduction to machine leaning with Python: a guide for data scientists]. St. Petersburg, Al'fa-kniga Publ., 2017. 480 p. (In Russian).

9. Deitel P., Deitel H. *Python: Iskusstvennyi intellekt, bol'shie dannye i oblachnye vychisleniya* [Python: artificial intelligence, big data, and cloud computing]. St. Petersburg, Al'fa-kniga Publ., 2020. 864 p. (In Russian).

10. Harrison M. *Mashinnoe obuchenie: karmannyi spravochnik: kratkoe rukovodstvo po metodam strukturirovannogo mashinnogo obucheniya na Python* [Machine learning: a pocket guide to structured machine learning methods in Python]. St. Petersburg, Dialectika Publ., 2020. 320 p. (In Russian).

11. Albon C. *Mashinnoe obuchenie s ispol'zovaniem Ruthon. Sbornik retseptov* [Machine learning with Python cookbook]. St. Petersburg, BHV-Petersburg Publ., 2019. 384 p. (In Russian).

12. Shalev-Shwartz Sh., Ben-David Sh. *Idei mashinnogo obucheniya: ot teorii k algoritmam* [Understanding machine learning: from theory to algorithms]. Moscow, DMK Press, 2019. 436 p. (In Russian).

13. Deisenroth M.P., Faisal A.A., Ong Ch.S. *Mathematics for machine learning*. Cambridge, Cambridge University Press, 2020. 398 p.

14. Fofanov O.B. *Algoritmy i struktury dannykh* [Algorithms and data structures]. Tomsk, TPU Publ., 2014. 126 p.

15. Cormen T., Stein K., Rivest R., Leiserson Ch. *Algoritmy. Postroenie i analiz* [Introduction to Algorithms]. Moscow, Dialektika Publ., 2019. 1324 p. (In Russian).

16. Cormen T. *Algoritmy: vvodnyi kurs* [Algorithms unlocked]. Moscow, Williams Publ., 2014. 208 p. (In Russian).

17. Babichev S.L. *Lektsii po algoritmam i strukturam dannykh* Lectures on Algorithms and Data Structures]. Moscow, MAKS Press, 2019. 344 p.

Для цитирования:

Моргунов А.В. Разработка системы автоматического распознавания типа документа по прикрепленному сообщению в СЭД // Системы анализа и обработки данных. – 2026. – № 1 (101). – С. 59–70. – DOI: 10.17212/2782-2001-2026-1-59-70.

For citation:

Morgunov A.V. Razrabotka sistemy avtomaticheskogo raspoznavaniya tipa dokumenta po prikreplennomu soobshcheniyu v SED [Development of a system for automatic document type recognition based on attached messages in an EDMS]. *Sistemy analiza i obrabotki dannykh = Analysis and Data Processing Systems*, 2026, no. 1 (101), pp. 59–70. DOI: 10.17212/2782-2001-2026-1-59-70.