

УДК 004.6

ИСПОЛЬЗОВАНИЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ДОКУМЕНТООБОРОТЕ



ON THE QUESTION OF THE USE OF BIG DATA TECHNOLOGY IN MODERN EDUCATION

Багдасарян Р.Х.

кандидат технических наук,
доцент,
Краснодарский государственный институт культуры
rafael_555@mail.ru

Матвеева А.С.

кандидат педагогических наук,
доцент
Краснодарского государственного института культуры
kas_dok_04@mail.ru

Лосева Е.С.

магистрант 1 курса,
Краснодарского государственного института культуры
katyhka-losewa@mail.ru

Аннотация. В данной статье проведен обзор методов искусственного интеллекта, применяемых в документообороте. Рассмотрены методы машинного обучения, нейронные сети и обработки естественного языка, а также проанализированы преимущества и ограничения каждого подхода. В заключении указано, что использование искусственного интеллекта в документообороте является активно развивающейся областью, но все еще остаются открытыми вопросы, связанные с применением AI в условиях реального бизнеса и социальных организаций.

Ключевые слова: искусственный интеллект, документооборот, машинное обучение, нейронные сети, обработка естественного языка, категоризация документов.

Bagdasaryan R.Kh.

Candidate of Technical Sciences,
Associate Professor,
Krasnodar State Institute of Culture
rafael_555@mail.ru

Matveeva A.S.

Candidate of Pedagogical Sciences,
Associate Professor,
Krasnodar State Institute of Culture
kas_dok_04@mail.ru

Losewa E.S.

1st year Master Student,
Krasnodar State Institute of Culture
katyhka-losewa@mail.ru

Annotation. This article reviews the methods of artificial intelligence applied in document management. The methods of machine learning, neural networks, and natural language processing are considered, and the advantages and limitations of each approach are analyzed. In conclusion, it is noted that the use of artificial intelligence in document management is a rapidly developing field, but there are still open questions related to the application of AI in real business and social organizations.

Keywords: artificial intelligence, document management, machine learning, neural networks, natural language processing, document categorization.

Введение

Современный бизнес и государственные организации ориентированы на эффективное управление документами. Электронный документооборот стал неотъемлемой частью многих компаний и учреждений, и неизбежно возникают вопросы о том, как можно улучшить его процессы и сделать их более удобными и быстрыми. Одним из наиболее перспективных способов улучшения документооборота является использование искусственного интеллекта (ИИ). С его помощью можно автоматизировать многие процессы, связанные с обработкой документов, сократить время и уменьшить риски ошибок.

Целью данной научной статьи является рассмотрение применения искусственного интеллекта в документообороте и оценка его эффективности. **Актуальность** данной темы обусловлена растущим интересом к технологиям искусственного интеллекта в сфере управления документами и возрастающей необходимостью автоматизации процессов документооборота.

Использование машинного обучения в документообороте

Один из самых распространенных способов применения искусственного интеллекта в документообороте – это использование методов машинного обучения. Этот

подход позволяет автоматизировать процессы обработки документов, включая распознавание текста, классификацию документов и выделение ключевых фраз.

Методы машинного обучения включают в себя алгоритмы, которые используются для построения моделей на основе обучающих данных. Эти модели могут затем использоваться для автоматического анализа новых данных. Например, можно обучить модель распознаванию текста, которая будет автоматически переводить отсканированные документы в текстовый формат [2].

Преимущества использования машинного обучения в документообороте заключаются в уменьшении времени на обработку документов и уменьшении риска ошибок при ручной обработке. Однако, ограничения этого метода включают в себя необходимость больших объемов обучающих данных и возможность сбоев в работе при использовании сложных алгоритмов.

Одной из ключевых работ, посвященных применению машинного обучения в документообороте, является статья «Document Classification Using Machine Learning Techniques: A Review» авторов Muhammad Asif, Fazli Subhan, Javid Ali, Muhammad Noman and Zia ul Qamar (2019). В этой работе авторы провели обзор существующих методов классификации документов с использованием машинного обучения и проанализировали преимущества и ограничения каждого метода. Они также представили сравнительный анализ на основе различных метрик производительности, таких как точность, полнота и F1-мера [7].

Применение нейронных сетей в документообороте

Нейронные сети – это алгоритмы, которые моделируют работу человеческого мозга. Они используются для анализа текста, классификации документов и выделения ключевых фраз. Этот подход может быть особенно полезен для обработки больших объемов данных [1].

Преимущества использования нейронных сетей в документообороте включают в себя возможность автоматического анализа больших объемов данных, а также возможность обучения нейронных сетей на основе небольшого количества обучающих данных. Однако, ограничения включают в себя сложность настройки алгоритмов и дороговизну этого подхода [4].

Применение методов обработка естественного языка

Применение методов обработки естественного языка (Natural Language Processing, NLP) в документообороте может значительно улучшить процессы работы с документами. Одним из способов применения NLP является автоматическое извлечение ключевых слов и фраз из документов. Это может быть полезно, например, при поиске документов, по ключевым словам, или при определении категории, к которой относится документ [3].

В работе «Automated Document Categorization Using Machine Learning and Text Mining Techniques» авторы использовали методы NLP для автоматической категоризации документов. Они использовали алгоритм Latent Dirichlet Allocation (LDA), который позволяет автоматически выделять темы из набора текстовых документов. LDA был использован для категоризации документов на основе их содержания [6].

Результаты работы показали, что комбинация методов NLP и машинного обучения может значительно улучшить производительность системы автоматической категоризации документов. Этот подход позволяет автоматизировать процесс категоризации документов и значительно сократить время, затрачиваемое на поиск и обработку документов [5].

Примеры успешного использования ИИ в документообороте. Рассмотрим несколько примеров компаний, которые успешно используют ИИ для обработки больших объемов документов. Эти компании демонстрируют, как использование ИИ позволяет ускорить процессы обработки документов, сократить время на выполнение задач и повысить производительность бизнеса.

1. JPMorgan Chase: Банк JPMorgan Chase использует ИИ для обработки большого объема документов, связанных с кредитными заявками и сделками на финансовых рынках. Использование ИИ позволяет банку быстрее и точнее анализировать документы, что уменьшает время на принятие решений и повышает производительность.

2. ABBYY: Компания ABBYY разрабатывает программное обеспечение для автоматического распознавания символов и обработки документов, используя технологии машинного обучения и ИИ. Это позволяет компаниям быстрее и точнее обрабатывать документы, уменьшая время на выполнение бизнес-процессов и повышая производительность.

3. DocuSign: Компания DocuSign использует ИИ для автоматической обработки документов, связанных с электронной подписью и автоматической аутентификацией. Использование ИИ позволяет компании упростить процесс подписания документов, уменьшить время на выполнение задач и повысить безопасность.

4. UiPath: Компания UiPath разрабатывает программное обеспечение для автоматической обработки документов, используя ИИ и машинное обучение. Это позволяет компаниям быстрее и точнее обрабатывать большие объемы документов, уменьшая время на выполнение бизнес-процессов и повышая производительность.

5. Amazon Web Services: Amazon Web Services предлагает инструменты для обработки документов, используя ИИ, такие как Amazon Textract и Amazon Comprehend. Эти инструменты позволяют компаниям автоматически анализировать документы, извлекать информацию из текста, классифицировать документы и т.д. Это упрощает процесс обработки документов и уменьшает время на выполнение задач.

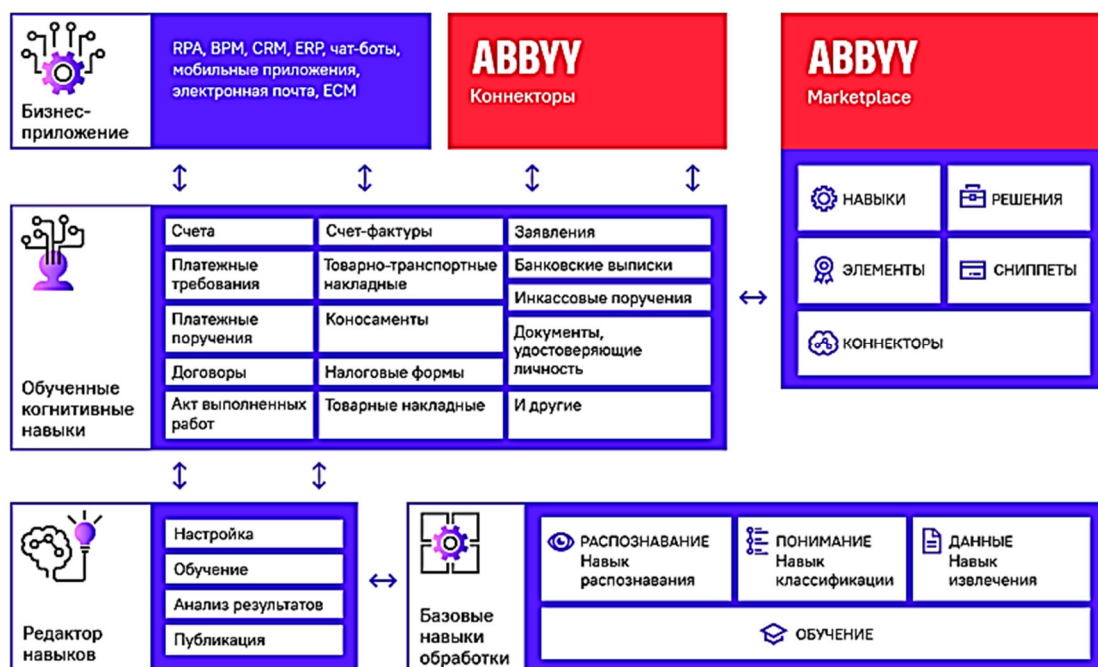


Схема 1 – Принцип работы программного обеспечения от ABBYY

Заключение

В данной статье был проведен обзор существующих подходов к применению искусственного интеллекта в документообороте. Мы рассмотрели методы машинного обучения, нейронных сетей и обработки естественного языка, а также проанализировали преимущества и ограничения каждого подхода. Методы машинного обучения позволяют автоматизировать процессы обработки документов, включая распознавание текста, классификацию документов и выделение ключевых фраз. Нейронные сети могут быть особенно полезны для обработки больших объемов данных. Методы обработки естественного языка могут помочь в автоматической категоризации документов и извлечении ключевых слов и фраз.

Результаты исследований показали, что комбинация различных методов и подходов может значительно улучшить производительность системы автоматической обработки и категоризации документов. Это позволяет сократить время, затрачиваемое на поиск и обработку документов, и повысить эффективность бизнес-процессов.

Литература

1. Бонч-Осмоловская А.А. Прикладной анализ данных: методы и практика с примерами на Python / А.А. Бонч-Осмоловская, К.В. Воронцов. – М. : ДМК Пресс, 2017.
2. Маннинг К. Введение в информационный поиск / К. Маннинг, П. Рагхаван, Х. Шютце. – М. : Техносфера, 2008.
3. Назаренко А. Что такое NLP или обработка естественного языка? – 2019.
4. Раджу П.Н.Н. Обзор алгоритмов и методов классификации текстов / П.Н.Н. Раджу, К.В.Н. Сунита, К. Равиндра // Журнал Advanced Computing: An International Journal (ACIJ). – 2019. – № 10(1). – № 1–8.
5. Фридман Б. Глубокое обучение на Python / Б. Фридман, Э. Фримен. – СПб. : КУДИЦ-Пресс, 2012.
6. Kumar S. Automated Document Categorization Using Machine Learning and Text Mining Techniques / S. Kumar, N. Kaushik, K. Sharma // Journal of Advanced Research in Dynamical and Control Systems. – 2019. – № 11. – P. 179–186.
7. Document Classification Using Machine Learning Techniques: A Review / M. Asif [et al.] // Journal of Intelligent & Fuzzy Systems. – 2019. – № 37(2). – P. 1571–1586.

References

1. Bonch-Osmolovskaya A.A. Applied Data Analysis: Methods and Practice with Python Examples / A.A. Bonch-Osmolovskaya, K.V. Vorontsov. – M. : DMK Press, 2017.
2. Manning K. Introduction to information retrieval / K. Manning, P. Raghavan, H. Schutze. – M. : Technosfera, 2008.
3. Nazarenko A. What is NLP or natural language processing? – 2019.
4. Raju P.N.N. Review of algorithms and methods for classifying texts / P.N.N. Raju, K.V.N. Sunita, K. Ravindra // Advanced Computing: An International Journal (ACIJ). – 2019. – № 10(1). – № 1–8.
5. Friedman B. Deep learning in Python / B. Friedman, E. Freeman. – SPb. : KUDITs-Press, 2012.
6. Kumar S. Automated Document Categorization Using Machine Learning and Text Mining Techniques / S. Kumar, N. Kaushik, K. Sharma // Journal of Advanced Research in Dynamical and Control Systems. – 2019. – № 11. – P. 179–186.
7. Document Classification Using Machine Learning Techniques: A Review / M. Asif [et al.] // Journal of Intelligent & Fuzzy Systems. – 2019. – № 37(2). – P. 1571–1586.