

Д. А. Сакович

*Белорусская сельскохозяйственная библиотека им. И. С. Лупиновича
Национальной академии наук Беларуси, Минск, Беларусь*

ВОЗМОЖНОСТИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В КАТАЛОГИЗАЦИИ ДОКУМЕНТОВ

Аннотация. В статье рассматривается использование искусственного интеллекта (ИИ) как перспективной технологии для автоматизации процессов каталогизации документов в библиотеках. Анализируется история развития ИИ, его междисциплинарный характер и влияние на различные сферы человеческой деятельности. Рассматриваются ключевые возможности применения ИИ на каждом этапе каталогизации, включая сбор данных, создание библиографических записей, составление аннотаций, определение тематики и индексирования. Подчеркивается, что интеграция ИИ в библиотечные процессы открывает новые горизонты для повышения эффективности работы библиотек, однако требует необходимости сбалансированного подхода к внедрению технологий для сохранения их социальной значимости.

Ключевые слова: искусственный интеллект (ИИ), каталогизация документов, автоматизация процессов, машинное обучение, обработка естественного языка.

Для цитирования. Сакович, Д. А. Возможности искусственного интеллекта в каталогизации документов / Д. А. Сакович // Библиотеки в информационном обществе: сохранение традиций и развитие новых технологий. Тема 2024 года – «Библиотечно-информационная деятельность в среде меняющихся социальных условий и технологических инноваций» : докл. VI Междунар. науч. конф., Минск, 5–6 дек. 2024 г. / Белорус. с.-х. б-ка им. И. С. Лупиновича Нац. акад. наук Беларуси ; редкол.: Ю. О. Каракулько (отв. ред.) [и др.]. – Минск, 2024. – С. 84–91.

D. A. Sakovich

*I.S. Lupinovich Belarusian Agricultural Library of the National Academy of
Sciences of Belarus, Minsk, Belarus*

THE POTENTIAL OF ARTIFICIAL INTELLIGENCE IN DOCUMENT CATALOGING

Abstract. This article discusses the use of artificial intelligence (AI) as a promising technology for automating document cataloging processes in libraries. It analyzes the history of AI development, its interdisciplinary nature, and its impact on various spheres of human activity. The key opportunities for applying AI at each stage of cataloging are examined, including data collection, bibliographic record creation, annotation drafting, subject determination, and indexing. It is emphasized that the integration of AI into library processes opens new horizons for enhancing the efficiency of libraries, however, it also requires the need for a balanced approach to the introduction of technologies to maintain their social significance.

Keywords: artificial intelligence (AI), document cataloging, process automation, machine learning, natural language processing.

For citation. Sakovich D. A. The potential of artificial intelligence in document cataloging. Libraries in the information society: preserving traditions and developing new technologies. The theme for 2024 – «Library and information activities in the environment of changing social conditions and technological innovations»: proceedings of the VI International scientific conference, Minsk, December 5–6, 2024. Minsk, 2024, pp. 84–91 (in Russian).

Искусственный интеллект (ИИ) представляет собой одну из наиболее перспективных технологий для решения задач, связанных с каталогизацией документов. Алгоритмы ИИ, такие как машинное обучение и обработка естественного языка, способны автоматизировать процессы обработки информации, что значительно упрощает работу с большими объемами данных. Применение этих технологий позволяет не только ускорить процесс каталогизации, но и повысить его точность, что в свою очередь ведет к улучшению качества поиска информации.

Таким образом, исследование использования ИИ в каталогизации документов актуально как для научного сообщества, так и для практической деятельности организаций, стремящихся оптимизировать свои информационные процессы. В данной статье рассматриваются основные возможности использования ИИ в каталогизации документов.

Искусственный интеллект (ИИ) зародился как научное направление в середине XX века и с тех пор претерпел значительные изменения, став важнейшей частью современных технологий. Он объединяет достижения различных наук, таких как нейробиология, математика, кибернетика, психология и философия, что делает его междисциплинарной областью исследований. Научные разработки в этой сфере позволили ИИ выйти за рамки простого автоматизированного решения задач и стать инструментом, способным к обучению, адаптации и анализу данных.

В 1950 году английский математик Алан Тьюринг предложил тест, который должен был ответить на вопрос, может ли машина мыслить. Этот тест стал одним из первых шагов в понимании ИИ как технологии, способной моделировать человеческие когнитивные способности. Тьюринг заложил основу для многих последующих исследований, как в области теории, так и в практическом применении ИИ.

В Советском Союзе значительный вклад в развитие ИИ внёс Дмитрий Александрович Поспелов. Он разработал первые в мире

подходы к принятию решений, основанные на логико-лингвистических моделях. Эти работы стали важной вехой для дальнейших исследований в области искусственного интеллекта, особенно в контексте автоматизации процесса принятия решений и обработки данных.

ИИ разделился на два крупных направления: кибернетику и нейрокибернетику. Кибернетики сосредоточились на функциональном равенстве реакций человека и машины, рассматривая ИИ как механизм, способный к эффективным решениям, не обращая внимания на внутреннее устройство процесса мышления. Это направление позволило разработать множество прикладных систем, способных имитировать человеческие решения [1].

Нейрокибернетика, в свою очередь, направлена на моделирование работы человеческого мозга. В рамках этого направления начались исследования, связанные с созданием искусственных нейронных сетей — систем, напоминающих структуру нервной системы человека. Эти исследования стали основой для глубокого обучения (deep learning), которое сегодня активно применяется в обработке данных, компьютерном зрении и других областях.

ИИ прошёл долгий путь от теоретических размышлений до создания реальных систем, способных решать сложные задачи. Развитие технологий, основанных на ИИ, продолжает ускоряться, открывая всё новые возможности для их применения в научных и практических областях.

Современные системы искусственного интеллекта (ИИ) демонстрируют резкий рост возможностей благодаря усовершенствованным алгоритмам и высокой скорости обработки данных. ИИ не просто имитирует человеческие когнитивные функции, но и выполняет их на значительно более высоких скоростях, используя огромные массивы данных. Одной из ключевых особенностей ИИ является способность анализировать и обрабатывать большие объёмы информации, выявляя закономерности и обучаясь на их основе, что делает его неотъемлемой частью многих современных технологий.

Системы ИИ, такие как нейронные сети, работают по принципу имитации обучения. Они анализируют множество примеров, создавая модель, которая затем способна решать аналогичные задачи. Эти системы обладают большой памятью и

используют её для перебора огромных массивов данных, а не просто поиска закономерностей. Это делает современные ИИ-системы мощными инструментами для анализа данных и принятия решений.

Существующие системы опираются на большие объемы данных и вычислительных мощностей для достижения своей эффективности. Принцип работы ИИ заключается в сочетании объема данных, интеллектуальных алгоритмов и быстрой обработки информации. Этот процесс позволяет программам автоматически обучаться, улучшая свои результаты на основе повторяющихся итераций и анализа скрытых закономерностей.

ИИ сегодня представляет собой комплексную научную дисциплину, включающую множество методик, теорий и технологий. Его развитие продолжает изменять различные сферы человеческой деятельности – от промышленности до медицины, управления данными и образования. ИИ, со своей способностью к самообучению, уже оказывает значительное влияние на технологический прогресс и неизбежно продолжит менять наш мир в будущем [2].

Специалисты в области библиотечного и книжного дела все активнее осознают важность внедрения искусственного интеллекта (ИИ). Одним из первых ученых, который затронул эту тему, является Вадим Константинович Степанов, доцент Московского государственного института культуры. В своей статье 1996 года под названием «Искусственный интеллект и возможности его применения в библиотеках» он отметил, что на тот момент некоторые элементы ИИ уже находили применение в библиотеках, интегрируясь в специализированные программные решения. Эти модули способствуют автоматизации процессов, связанных с соблюдением стандартов библиографической записи и созданием отчетных документов.

Современные программы способны самостоятельно генерировать заголовки библиографических записей, основываясь на числе авторов. Если в книге один или два автора, они оба отображаются в заголовке. При наличии трех или четырех авторов указывается только первый, а остальные перечисляются в поле ответственности. Если же авторов пять и более, программа формирует заголовок без указания всех имен, добавляя в отдельное поле имена первых трех авторов. Кроме того, при составлении инвентарной книги система автоматически предлагает указать инвентарный номер.

Тем не менее, данное применение ИИ является лишь первичным этапом его интеграции в библиотечное дело. На этом уровне разработанные компьютерные системы в основном направлены на оптимизацию существующих библиотечных технологий, что не приводит к кардинальным изменениям их сути. Эти технологии делают процессы более эффективными и экономичными, но не трансформируют саму природу библиотеки.

Переход на электронные технологии обещает глубокие изменения, которые могут поставить под угрозу существование библиотек как социальных институтов, если их не учитывать. В. К. Степанов подчеркивает, что системы ИИ должны стать краеугольным камнем работы библиотеки. Все технологические процессы, начиная с индивидуального приветствия читателей и напоминания о предыдущих взаимодействиях с документами до регулирования климатических условий в помещениях в зависимости от количества людей, должны выполняться в автоматическом режиме. На текущий момент мы наблюдаем лишь отдельные случаи внедрения элементов ИИ в библиотечную практику.

Призыв к полному отказу от традиционных методов может показаться чрезмерным, но недооценка потенциала ИИ чревата серьезными последствиями для всей библиотеки [3].

В контексте этих изменений важным аспектом остается и процесс каталогизации документов.

Вот как можно использовать искусственный интеллект на каждом этапе каталогизации документов:

1. Сбор данных:

Автоматизированный анализ метаданных: искусственный интеллект может использовать алгоритмы машинного обучения, такие как нейронные сети и модели обработки естественного языка, для извлечения и автоматического распознавания метаданных из различных форматов документов (например, PDF, Word, изображения). Для этого системы могут применять методы Optical Character Recognition (OCR) для извлечения текста из изображений и дальнейшего анализа полученного текста с помощью предобученных языковых моделей, что позволяет автоматически собирать заголовки, авторов, даты публикации и другие характеристики.

Оптимизация процесса ввода данных: системы на основе ИИ могут анализировать данные с использованием алгоритмов для обнаружения аномалий и предлагать исправления или дополнения на основе контекстного анализа, минимизируя ошибки ввода и ускоряя процесс сбора данных.

2. Создание библиографической записи.

Автоматическое формирование библиографических записей: ИИ может анализировать собранные данные с помощью алгоритмов структурированного извлечения информации и автоматически создавать библиографические записи в соответствии с заданными стандартами (например, MARC21). Это включает использование шаблонов для правильного форматирования записей и заполнение всех необходимых полей на основе заранее заданных правил и критериев.

3. Составление аннотаций.

Системы на основе ИИ могут автоматически генерировать аннотации к документам, применяя алгоритмы обработки естественного языка, такие как абстрактное и экстрактивное суммирование. Эти алгоритмы способны анализировать содержание документа, выделять ключевые идеи и темы, формулируя краткие и информативные аннотации, которые точно отражают содержание. Использование моделей трансформеров, таких как GPT, позволяет значительно сократить время, необходимое для ручного написания аннотаций, и обеспечивает их однородность [4].

4. Определение тематики и присвоение ключевых слов.

Автоматизированное присвоение ключевых слов: ИИ может анализировать содержание документа с использованием алгоритмов кластеризации и тематического моделирования (например, LDA) для автоматического определения его тематики. На основе анализа текста системы могут присваивать соответствующие ключевые слова с использованием предобученных моделей, таких как Word2Vec или GloVe, что помогает улучшить точность и релевантность поисковых запросов.

5. Индексирование.

Создание индексов на основе таблиц УДК и ББК: Искусственный интеллект может использовать таблицы Универсальной десятичной классификации (УДК) и Библиотечно-библиографической классификации (ББК) для автоматического присвоения индексных номеров документам.

Алгоритмы могут анализировать содержание материалов и сопоставлять его с соответствующими классами и под классами этих систем, обеспечивая точное индексирование.

Таким образом, интеграция ИИ в процессы библиотечного дела открывает новые возможности для повышения эффективности и адаптации библиотек к современным требованиям.

Искусственный интеллект представляет собой ключевую технологию для оптимизации процессов каталогизации документов в библиотеках. Благодаря алгоритмам машинного обучения и обработки естественного языка, ИИ позволяет автоматизировать сбор данных, создание библиографических записей, составление аннотаций, определение тематики и индексацию. Это не только ускоряет процесс обработки информации, но и повышает его точность, что в свою очередь улучшает доступ к знаниям. Внедрение ИИ в библиотечную практику способствует трансформации традиционных методов работы, делая их более эффективными и соответствующими современным требованиям. Однако важно учитывать необходимость продуманного подхода к интеграции этих технологий, чтобы сохранить социальную значимость библиотек как культурных институтов.

Список использованных источников:

1. Горохов, А. В. Искусственный интеллект / А. В. Горохов, В. А. Мартынов, В. А. Гаврин // Скиф. Вопросы студенческой науки. – 2022. – № 4 (68). – С. 159–162.
2. Утегенов, Н. Б. Искусственный интеллект на сегодняшний день / Н. Б. Утегенов // Universum: технические науки. – 2022. – № 7-1 (100). – С. 27–30.
3. Столяров, Ю. Н. Искусственный интеллект и книжная библиотечная отрасль: направления разработки проблемы / Ю. Н. Столяров // Научные и технические библиотеки. – 2022. – № 1. – С. 17–34. <https://doi.org/10.33186/1027-3689-2022-1-17-34>
4. Барышев, Р. А. Технологии искусственного интеллекта в библиотеке / Р. А. Барышев, Е. Н. Касянчук, И. С. Рзянкин // Информационный бюллетень РБА. – 2023. – № 101. – С. 54–58
5. Абдулина, Э. М. Искусственный интеллект: проблемы и перспективы / Э. М. Абдулина // Молодой ученый. – 2020. – № 1 (291). – С. 9–10.
6. Искусственный интеллект для науки и наука для искусственного интеллекта / К. В. Анохин, К. С. Новоселов, С. К. Смирнов [и др.] // Вопросы философии. – 2022. – № 3. – С. 93–105. <https://doi.org/10.21146/0042-8744-2022-3-93-105>

7. Зулунов, Р. М. Искусственный интеллект-от мифологии до машинного обучения / Р. М. Зулунов, А. М. Каюмов // *Proceedings of International Educators Conference*. – 2022. – Vol. 1, № 2. – P. 25–30.
8. Нуждова, Д. А. Нейросети в библиотечном деле: опыт проекта «Новые библиотекари» / Д. А. Нуждова // *Корпоративные библиотечные системы: технологии и инновации (КорФор-2023)* : материалы Междунар. науч.- практ. конф., 19–23 июня 2023 г. / С.-Петерб. политехн. ун-т. – СПб., 2023. – С. 59–64. <https://doi.org/10.18720/SPBPU/2/k23-6>
9. Тимошенко, И. В. Искусственный интеллект в библиотечных технологиях. Уже пора? / И. В. Тимошенко // *Румянцевские чтения – 2019* : материалы Междунар. науч.-практ. конф. (23–24 апр. 2019) : в 3 ч. / Рос. гос. б-ка ; редкол.: В. В. Дуда [и др.]. – М., 2019. – Ч. 3. – С. 153–157.

References:

1. Gorokhov A. V., Martynov V. A., Gavrin V. A. Artificial Intelligence. *Skif. Voprosy studencheskoi nauki = Sciff. Questions of Students Science*, 2022, no. 4 (68), pp. 159–162 (in Russian).
2. Utegenov N. Artificial intelligence today. *Universum: tekhnicheskie nauki* [Universum: Technical Sciences], 2022, no. 7-1 (100), pp. 27–30 (in Russian).
3. Stolyarov Yu. N. Artificial intellect and the book library industry: vectors for problem development. *Nauchnye i tekhnicheskie biblioteki = Scientific and Technical Libraries*, 2022, no. 1, pp. 17–34 (in Russian). <https://doi.org/10.33186/1027-3689-2022-1-17-34>
4. Baryshev R. A., Kasyanchuk E. N., Rzyankin I. S. Artificial intelligence technologies in the library. *Informatsionnyi byulleten' RBA = Information Bulletin of the RLA*, 2023, no. 101, pp. 54–58 (in Russian).
5. Abdulina E. M. Artificial intelligence: problems and prospects. *Molodoi uchenyi* [Young Scientist], 2020, no. 1 (291), pp. 9–10 (in Russian).
6. Anokhin K. V., Novoselov K. S., Smirnov S. K., Efimov A. R., Matveev Ph. M. AI for science and science for AI. *Voprosy filosofii*, 2022, no. 3, pp. 93–105 (in Russian). <https://doi.org/10.21146/0042-8744-2022-3-93-105>
7. Zulunov R. M., Kayumov A. M. Artificial Intelligence: from mythology to machine learning. *Proceedings of International Educators Conference*, 2022, vol. 1, no. 2, pp. 25–30 (in Russian).
8. Nuzhdova D. A. Neural networks in librarianship: the experience of the «New Librarians» project. *Korporativnye bibliotечnye sistemy: tekhnologii i innovatsii (KorFor-2023): materialy Mezhdunarodnoi nauchno-prakticheskoi konferentsii, 19–23 iyunya 2023 g.* [Corporate library systems: technologies and innovations (KorFor-2023): proceedings of the International scientific-practical conference, June 19–23, 2023]. St. Petersburg, 2023, pp. 59–64 (in Russian). <https://doi.org/10.18720/SPBPU/2/k23-6>
9. Timoshenko I. V. Artificial intelligence in library technologies. Is it time yet? *Rumyantsevskie chteniya – 2019: materialy Mezhdunarodnoi nauchno-prakticheskoi konferentsii (23–24 aprelya 2019) = The Rumyantsev readings – 2019: proceedings international scientific and practical conference (April 23–24, 2019)*. Moscow, 2019, pt. 3, pp. 153–157 (in Russian).

Поступила в редакцию 16.10.2024
Received 16.10.2024