УДК 004.8:005.94

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ ДЛЯ УПРАВЛЕНИЯ ЗНАНИЯМИ: АВТОМАТИЗАЦИЯ ИЗВЛЕЧЕНИЯ И СТРУКТУРИРОВАНИЯ ИНФОРМАЦИИ ИЗ РАЗНОРОДНЫХ ИСТОЧНИКОВ

Москвин А.Д.[1], магистрант гр. 090402-СИБУо-24/1, I курс
Научный руководитель: Зырянова С.А.[1], к.т.н., доцент, заведующая кафедрой ИСиЦТ
[1] Федеральное государственное бюджетное образовательное учреждение высшего образования «Московский государственный университет технологий и управления имени К.Г. Разумовского (Первый казачий университет)», г. Москва

Управление знаниями становится критически важным аспектом для организаций в условиях роста объёмов информации и её разнородности. По оценкам экспертов, к 2025 году более 80% корпоративных данных будет представлено в неструктурированном виде, таком как тексты, аудио и изображения [1]. Искусственный интеллект (ИИ) предлагает решения для автоматизации извлечения и структурирования знаний, преобразуя хаотичные данные в организованные системы, доступные для анализа и принятия решений. Цель статьи – исследовать роль ИИ в управлении знаниями, рассмотреть методы обработки разнородных источников, проанализировать вызовы и обозначить перспективы развития этого направления.

Управление знаниями включает процессы сбора, хранения, обработки и использования информации для повышения эффективности организаций. Традиционные подходы, такие как ручная классификация или использование статических баз данных, не справляются с современными требованиями из-за трудоёмкости и ограниченной масштабируемости [2]. ИИ, напротив, позволяет автоматизировать эти процессы, извлекая знания из текстов, научных статей, корпоративных документов и даже неформальных источников, таких как социальные сети [3].

Примером успешного применения ИИ является автоматизация анализа юридических документов. Системы на основе обработки естественного языка (NLP) способны извлекать ключевые термины, классифицировать контракты и выявлять риски, сокращая время обработки на 70% [4]. В научной сфере ИИ используется для структурирования данных из публикаций, что ускоряет исследовательский процесс. Так, платформа IBM Watson помогла исследователям в области медицины выявить связи между генами и заболеваниями, проанализировав миллионы статей за считанные дни [5].

В России управление знаниями с помощью ИИ активно развивается в образовательных и промышленных секторах. Например, системы автоматической

обработки учебных материалов позволяют создавать онтологии для адаптивного обучения, повышая эффективность усвоения знаний студентами на 20–25% [6]. В промышленности ИИ помогает структурировать данные с производственных линий, что улучшает контроль качества и оптимизирует процессы [7].

Ключевым инструментом ИИ в управлении знаниями является обработка естественного языка (NLP). Современные модели, такие как BERT (Bidirectional Encoder Representations from Transformers), позволяют извлекать сущности, определять отношения между ними и строить семантические сети [8]. Например, BERT успешно применяется для анализа текстовых отчетов, автоматически выделяя ключевые факты с точностью до 90% [9].

Другой важный метод – построение онтологий, которые структурируют знания в виде иерархий понятий и связей. ИИ-системы, такие как Protégé в сочетании с алгоритмами машинного обучения, автоматизируют создание онтологий на основе больших текстовых корпусов [10]. Это особенно полезно в корпоративных системах, где данные из разных отделов нужно объединить в единую базу знаний [11].

Также активно применяются методы извлечения информации (Information Extraction, IE). Алгоритмы IE, такие как OpenIE, способны выделять факты из неструктурированных текстов, преобразуя их в структурированные таблицы или графы [12]. В российской практике подобные технологии используются для анализа нормативных документов, что сокращает время подготовки аналитических отчётов на 40% [13].

Для интеграции разнородных источников применяются системы управления знаниями (Knowledge Management Systems, KMS) с поддержкой ИИ. Например, платформа Microsoft SharePoint с модулями ИИ позволяет объединять текстовые документы, базы данных и мультимедиа в единую экосистему [14]. В России такие решения адаптируются под национальные проекты, включая создание цифровых архивов [15].

Применение ИИ для управления знаниями сталкивается с несколькими трудностями. Первая – сложность обработки многозначности и контекста в текстах. Даже продвинутые модели NLP могут ошибаться при интерпретации идиом или специализированных терминов, что снижает точность извлечения знаний [16]. Исследования показывают, что в 15–20% случаев такие ошибки требуют ручной корректировки [17].

Вторая проблема – разнородность источников. Данные из текстов, аудио и изображений требуют разных подходов к обработке, что усложняет их интеграцию [18]. Например, преобразование аудиозаписей в текст с последующим анализом увеличивает время обработки на 30% по сравнению с чисто текстовыми данными [19].

Третий вызов – конфиденциальность и этика. Автоматизация извлечения знаний из корпоративных или персональных данных может привести к утечкам информации, что особенно актуально в условиях строгих требований законодательства, таких как GDPR или российский закон о персональных данных

[20]. В России этот вопрос регулируется в рамках национальной стратегии развития ИИ, но практическая реализация остаётся сложной задачей [21].

Искусственный интеллект трансформирует управление знаниями, автоматизируя извлечение и структурирование информации из разнородных источников. Методы NLP, построения онтологий и систем управления знаниями открывают новые возможности для организаций, науки и образования, повышая доступность и полезность данных. Однако для полного раскрытия потенциала ИИ необходимо преодолеть проблемы точности, интеграции и безопасности. Перспективы развития связаны с совершенствованием алгоритмов, адаптацией технологий под локальные нужды и разработкой этических стандартов, что обеспечит устойчивое внедрение ИИ в управление знаниями.

Список литературы

1. Reinsel D., Gantz J., Rydning J. The Digitization of the World: From Edge to Core // IDC White Paper. – 2018. – URL: https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf (дата обращения: 10.03.2025).
2. Davenport T.H., Prusak L. Working Knowledge: How Organizations Manage What They Know. – Boston: Harvard Business School Press, 1998. – 224 p.
3. Manning C.D., Raghavan P., Schütze H. Introduction to Information Retrieval. – Cambridge: Cambridge University Press, 2008. – 506 p.
4. Surden H. Artificial Intelligence and Law: An Overview // Georgia Law Review. – 2019. – Vol. 35, № 4. – P. 1305–1337.
5. Chen Y., Argentinis E., Weber G. IBM Watson: How Cognitive Computing Can Be Applied to Big Data Challenges in Life Sciences Research // Clinical Therapeutics. – 2016. – Vol. 38, № 4. – P. 688–701.
6. Siemens G. Connectivism: A Learning Theory for the Digital Age // International Journal of Instructional Technology and Distance Learning. – 2005. – Vol. 2, № 1. – URL: http://www.itdl.org/Journal/Jan_05/article01.htm (дата обращения: 10.03.2025).
7. Nonaka I., Takeuchi H. The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation. – New York: Oxford University Press, 1995. – 284 p.
8. Devlin J., Chang M.-W., Lee K., Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding // Proceedings of NAACL-HLT. – 2019. – P. 4171–4186.
9. Vaswani A., Shazeer N., Parmar N., Uszoreit J., Jones L., Gomez A.N., Kaiser Ł., Polosukhin I. Attention Is All You Need // Advances in Neural Information Processing Systems. – 2017. – P. 5998–6008.
10. Noy N.F., McGuinness D.L. Ontology Development 101: A Guide to Creating Your First Ontology // Stanford Knowledge Systems Laboratory Technical Report. – 2001. – URL: https://protege.stanford.edu/publications/ontology_development/ontology101.pdf (дата обращения: 10.03.2025).

11. Gruber T.R. A Translation Approach to Portable Ontology Specifications // Knowledge Acquisition. – 1993. – Vol. 5, № 2. – P. 199–220.

12. Etzioni O., Banko M., Soderland S., Weld D.S. Open Information Extraction from the Web // Communications of the ACM. – 2008. – Vol. 51, № 12. – P. 68–74.

13. Российская практика цифровизации нормативных документов // Цифровая экономика РФ. – URL: https://digital.gov.ru/ru/activity (дата обращения: 10.03.2025).

14. Microsoft SharePoint: Official Documentation. – URL: https://www.microsoft.com/en-us/microsoft-365/sharepoint/collaboration (дата обращения: 10.03.2025).

15. Национальная программа «Цифровая экономика Российской Федерации» // Официальный портал Правительства РФ. – URL: https://digital.gov.ru/target/naczionalnaya-programma-czifrovaya-ekonomika-rossijskoj-federaczii (дата обращения: 10.03.2025).

16. Jurafsky D., Martin J.H. Speech and Language Processing. – 3rd ed. – Upper Saddle River: Prentice Hall, 2020. – 1024 p.

17. Rajpurkar P., Zhang J., Lopyrev K., Liang P. SQuAD: 100,000+ Questions for Machine Comprehension of Text // Proceedings of EMNLP. – 2016. – P. 2383–2392.

18. Russell S., Norvig P. Artificial Intelligence: A Modern Approach. – 4th ed. – Upper Saddle River: Prentice Hall, 2020. – 1136 p.

19. Rabiner L.R. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition // Proceedings of the IEEE. – 1989. – Vol. 77, № 2. – P. 257–286.

20. Mittelstadt B.D., Allo P., Taddeo M., Wachter S., Floridi L. The Ethics of Algorithms: Mapping the Debate // Big Data & Society. – 2016. – Vol. 3, № 2. – P. 1–21.

21. Национальная стратегия развития искусственного интеллекта до 2030 года // Официальный сайт Президента РФ. – URL: http://static.kremlin.ru/media/events/files/ru/AH4x6HgKWANwVtMOfPDhcbRpvd1HCCsv.pdf (дата обращения: 10.03.2025).