

Деркач Ніна Валентинівна,

ORCID <https://orcid.org/0000-0002-6859-9254>,

молодша наукова співробітниця,

відділ комплексного опрацювання документів,

Національна бібліотека України імені В. І. Вернадського,

Київ, Україна

e-mail: derkach317@ukr.net

Застосування штучного інтелекту для каталогізації даних

У роботі розглянуто вплив цифровізації та впровадження застосунків штучного інтелекту на процеси каталогізації. Наголошено на потребі єдиного стандарту та формату бібліографічного опису, зрозумілого фахівцям у будь-якій країні світу.

Ключові слова : цифровізація бібліотечної сфери, нові стандарти в каталогізації, бібліотечні інновації, штучний інтелект, каталогізація.

Вперше термін «штучний інтелект» (ШІ) запропонував американський вчений Джон Маккарті на літньому семінарі 1956 року у Дартмутському коледжі. У 2012 році з'явилися нові технології, коли нейромережі навчилися розпізнавати зображення з високою точністю. А у 2022 році настала ера генеративного ШІ (поява систем на кшталт ChatGPT), здатного створювати тексти людською мовою, коди та зображення. Штучний інтелект перестав бути лише темою для наукової фантастики. Сьогодні він став невід'ємною частиною нашого повсякденного життя.

Бібліотеки завжди були центрами знань, зберігаючи та роблячи інформацію доступною для спільноти. Цифрова епоха принесла нові можливості для бібліотек. Зростання цифрової інформації призвело до зміни у способах надання послуг бібліотеками. Бібліотеки поступово стають мультимедійними му-

льтифункціональними центрами. Поява штучного інтелекту у бібліотеках – це значний технологічний прорив, який революціонізує способи роботи бібліотек.

У контексті інтелектуального управління ресурсами, технології ШІ дозволяють бібліотекам автоматизувати процеси каталогізації, оцифровувати матеріали та ефективно організовувати інформацію. Це покращує доступ до ресурсів, роблячи їх доступними для ширшої аудиторії.

Штучний інтелект отримує інформацію з тексту та зображень, автоматизує сортування та опис даних. Він перекладає назви, знаходить автора, дату, розмір або ціну, класифікує об'єкти, аналізує картинки або текст і сам визначає ключові слова без участі людини. Він автоматично переносить ці дані в потрібні поля бази даних. Це прискорює роботу, знижує кількість помилок і робить пошук у каталозі зручнішим. Щоб документ було простіше знайти, він перекладає паперові каталоги або фото карток у цифровий вигляд, розпізнаючи навіть рукописний текст. ШІ трансформує каталогізацію бібліотек, роблячи її більш ефективною, точною та зручною для користувача.

Хмарні системи у бібліотеках – це сучасні програми. Такі платформи допомагають зберігати документи, керувати каталогами та видавати електронні видання. Вони не потребують місця та затрат на великі сервери. Всі дані зберігаються на віддалених потужних комп'ютерах, які знаходяться в дата-центрах, пов'язаних між собою єдиною мережею. Основні дата-центри розташовані в США та країнах Європи. Програмне забезпечення (ПЗ) не управляється за допомогою комп'ютерів бібліотеки. Доступ до ПЗ надається через загальну мережу Інтернет, зазвичай на основі щомісячної або річної передплат. Тобто бібліотека повинна бути готова до витрат. Читачі можуть відкривати книги з телефону або комп'ютера в будь-який час.

Наприклад, автоматизована інформаційно-бібліотечна система (АІБС) «Коѡа» є першою бібліотечною системою з відкритим вихідним кодом, яка широко використовується публічними, шкільними та спеціальними бібліотеками. АІБС «Коѡа» забезпечує основні функції, які потрібні для обслуговування читачів: онлайн доступ до електронного каталогу; електронне замовлення доку-

ментів; каталогізація; база даних читачів бібліотеки; видача/повернення книг. У деяких країнах, до прикладу, у Казахстані вже використовується система Elibra.kz. Це сучасна хмарна бібліотечна система для автоматизації бібліотек. Система Elibra.kz запровадила штучний інтелект для прискорення процесу каталогізації книг у форматі MARC21. Вона пропонує наступні основні можливості та функції: допомагає швидко додавати книги до каталогу, вести облік читачів та збирати статистику, вміє розпізнавати текст з картинок і сама заповнює потрібні поля. Спеціальний інструмент допомагає швидко перевіряти наявність книг на полицях. Програма стежить за тим, хто взяв книгу і коли її потрібно повернути.

Виконання рутинної роботи з додавання даних до системи не звільняє бібліотекаря від процесу каталогізації. На сьогодні впровадження ІІІ в каталогізації потребує експертизи. Тому бібліотекам потрібні співробітники, які здатні розробляти, перевіряти інформацію, керувати та доопрацьовувати алгоритми ІІІ.

Як пише путівник по досягненнях науки і техніки «Scientific American», «Міжнародний комітет Червоного Хреста (МКЧХ) попереджає, що штучний інтелект вигадує наукові статті, журнали та архіви, вводячи користувачів в оману. Такі популярні моделі ІІІ як-от ChatGPT від OpenAI, Gemini від Google та Copilot від Microsoft генерують некоректні або сфабриковані архівні посилання. За словами керівниці відділу взаємодії з дослідниками Бібліотеки штату Вірджинія Сари Фоллс, близько 15 % електронних довідкових запитів, які надходять до бібліотеки, згенеровані ІІІ. Частина таких запитів містить вигадані цитати на опубліковані роботи та унікальні першоджерела, що значно ускладнює роботу бібліотекарів. Також системи ІІІ вимагають доступу до даних про відвідувачів для надання рекомендацій та покращення процесу каталогізації. Тож бібліотеки повинні приділяти пріоритетну увагу конфіденційності користувачів та безпеці даних, впроваджуючи надійні політики та заходи захисту.

Нормативні стандарти задають вектор розвитку той чи іншої галузі. Коли ми створюємо бібліографічний запис, ми спираємося на чинний в Україні стан-

дарт ДСТУ ГОСТ 7.1:2006 («Бібліографічний запис. Бібліографічний опис. Загальні вимоги та правила складання»), але він вже не відповідає розвитку новітніх технологій у бібліотечної справі. Згідно ISBD (International Standard Bibliographic Description), бібліографічний опис (БО) повинен забезпечувати швидкий доступ до бібліографічних даних та сумісність при обміні бібліографічною інформацією. Бібліографічний опис мусить бути повним, універсальним і відповідати вимогам міжнародних стандартів. Англо-американські правила каталогізації (Anglo-American Cataloguing Rules, AACR) дедалі більше критикувалися за надмірну складність, занадто сильну залежність від концепцій застарілого карткового каталогу, великі труднощі щодо нових видів цифрових ресурсів. У 2010 році, після більш ніж шести років розробки, було опубліковано перше видання наступника Англо-американських правил каталогізації (AACR2) під назвою «Опис ресурсів та доступ» (RDA). Це новий стандарт каталогізації, який замінить 2-ге видання Англо-американських правил каталогізації. RDA розвиває попередні правила каталогізації, є посібником з каталогізації електронних ресурсів та приділяє більше уваги допомоги користувачам у пошуку, ідентифікації, відборі та отриманні необхідної їм інформації. RDA також сприяє кластеризації бібліографічних записів для подання зв'язків між творами та його творцями. RDA побудовано на двох концептуальних моделях, вироблених Федерацією бібліотечних асоціацій та інститутів (ІФЛА) – Функціональних вимогах до бібліографічних записів (FRBR) та Функціональних вимогах до авторитетних даних (FRAD). Це збірник правил та інструкцій для створення бібліографічних метаданих. По ньому бібліотеки, архіви та музеї мають опрацьовувати книги, фільми, вебсайти та інші матеріали. ІІІ навчають створювати та розуміти RDA-сумісні метадані для швидкого опрацювання величезних обсягів інформації. При розробці стандартів необхідно розуміти логіку прийняття рішень ІІІ, яка не завжди очевидна для людини. Тому зараз іде процес тестування стандарту, з метою виявлення помилок в роботі систем на основі ІІІ. Основна мета нового стандарту – спрощення, уточнення та модернізація правил бібліографічного опису. Ключовий момент у проєкті RDA – узгодження з кон-

цептуальними моделями для бібліографічних та авторитетних даних, розроблених Міжнародною федерацією бібліотечних асоціацій та установ (ІФЛА).

Основні відмінності ДСТУ від RDA. Внесено зміни до термінології, щоб відокремити зміст від носія. Ресурс – цей термін став головним для позначення об'єкта опису. ДСТУ орієнтований на створення чіткого, суворо відформатованого тексту для списків літератури наприкінці статей та дисертацій. Він диктує, де поставити крапку, тире та курсив. RDA орієнтований на створення даних (метаданих) для електронних каталогів. Він описує не зовнішній вигляд запису, а сутність самої книги, статті чи іншого ресурсу для легкого пошуку в мережі.

ДСТУ вимагає суворо дотримуватися розділових знаків між областями (наприклад, крапку і тире . —). RDA нейтральний до синтаксису і не прив'язаний до суворих розділових знаків. Запис RDA можна використовувати в різних форматах даних (наприклад, MARC 21).

ДСТУ часто обмежує кількість вказаних авторів (наприклад, якщо авторів четверо і більше, за ДСТУ вказують тільки перших трьох, а інші замінюються словами «та ін»).

ДСТУ вимагає масово скорочувати слова за суворими правилами (напр., С. – сторінка, Изд. – Видавництво). RDA прагне не використовувати скорочення, аби текст був зрозумілий звичайній людині без розшифровки.

ДСТУ використовує загальні позначення матеріалів, наприклад, [Текст] або [Електронний ресурс]. RDA поділяє опис на елементи: окремо описується форма носія (папір, диск, електронний ресурс) та окремо тип вмісту (текст, зображення, музика).

ДСТУ виходить із самого фізичного документа. RDA спирається на сучасні моделі FRBR та IFLA LRM, які пов'язують в єдину мережу твір (твір думки), його втілення (видання) та уявлення (фізичний екземпляр).

Впровадження ІІІ для класифікації даних може принести значні переваги у бібліотечну сферу. Стандартизована форма книги, каталогізація ресурсу, використання алгоритмів кластеризації полегшує пошук і перегляд у каталозі. Завдяки використанню можливостей моделей машинного навчання, процес

оновлення каталогу значно спрощується. Система може розрізняти нові або модифіковані ресурси даних і відповідно вносити необхідні зміни, гарантуючи точне представлення доступних ресурсів. Та головне – єдиний стандарт та формат бібліографічного опису зрозумілий фахівцям у будь-якій країні світу.

Nina Derkach,

ORCID <https://orcid.org/0000-0002-6859-9254>,

Junior Researcher,

Department of Integrated Processing of Documents,

V. I. Vernadskyi National Library of Ukraine

Kyiv, Ukraine,

e-mail: derkach317@ukr.net

Implementation of artificial intelligence for data catalogization

The paper examines the impact of digitalization and the introduction of artificial intelligence applications on cataloguing processes. It emphasizes the need for a single standard and format for bibliographic description that is understandable to specialists in any country in the world.

Keywords: digitalization of the library sphere, new standards in cataloguing, library innovations, artificial intelligence, cataloguing.