

УДК 655.254.422

ТЕНДЕНЦІЇ У ТЕХНОЛОГІЯХ ОЦИФРУВАННЯ РАРИТЕТНИХ СТАРОДРУКОВАНИХ ВИДАНЬ

В. Б. Репета, О. Б. Гаврилишин, В. Я. Филь

*Національний Університет «Львівська Політехніка»
вул. Під Голоском, 19, Львів, 79020, Україна*

У цифрову епоху актуальним завданням наукових бібліотек і архівів є збереження та забезпечення доступу до культурної спадщини, зокрема до раритетних стародрукованих видань. Оцифрування таких матеріалів потребує спеціалізованих технологій, що враховують їх фізичну крихкість, специфіку друкарських технік та вимоги до якості цифрового зображення. Удосконалення процесів сканування включає адаптацію сучасного обладнання, застосування алгоритмів обробки зображень високої точності, а також створення умов для мінімального втручання у структуру документа. Оцифрування стародрукованих видань має низку викликів – від проблем зі збереженням палітурки та унікальних шрифтів до труднощів розпізнавання тексту, надрукованого за історичними орфографічними нормами. До основних ризиків належать пошкодження матеріалів у процесі сканування, зниження якості цифрових копій через зношення паперу або втрату контрасту.

У статті розглянуто технічні та організаційні аспекти удосконалення технології сканування, проаналізовано досвід провідних європейських бібліотек у сфері цифрової реставрації та описано практичні рішення щодо обробки дефектних аркушів, підвищення точності OCR (оптичного розпізнавання тексту), та метаданих. Показано, як інтеграція новітніх технологій – таких як штучний інтелект і нейромережі – може значно покращити результат, забезпечити автоматизовану класифікацію видань та відновлення втрачених фрагментів. Також проаналізовано вимоги до цифрових архівів, у тому числі дотримання стандартів сумісності та довгострокового збереження.

Ключові слова: сканування, оцифрування, стародруки, цифрова реставрація, цифрові архіви, збереження документів, інформаційні технології, штучний інтелект.

Постановка проблеми. У сучасних умовах цифрової трансформації провідним завданням бібліотек, архівів і наукових установ є збереження та забезпечення відкритого доступу до культурної спадщини. Особливої актуальності набуває питання оцифрування раритетних стародрукованих видань, які мають виняткову історичну, наукову та мистецьку цінність. Проте процес сканування таких джерел є надзвичайно складним з огляду на їх фізичну крихкість, особливості друкарських технік та відсутність єдиних стандартів цифрової репрезентації.

Традиційні технології оцифрування не завжди забезпечують належну якість цифрових копій, а також можуть призводити до часткового пошкодження артефактів у разі некоректного використання обладнання. Водночас низький рівень автоматизації обробки зображень та обмежені можливості розпізнавання текстів старою орфографією ускладнюють створення функціональних електронних ресурсів. Перевидання таких джерел є ускладненим і вимагає оновленого підходу до видавничого процесу на основі раритетних джерел.

Отже, виникає потреба в удосконаленні методів та технічних засобів сканування для забезпечення високої якості цифрових копій без шкоди для оригіналів. Це вимагає пошуку інноваційних рішень на стику інформатики, видавничої діяльності та реставраційної справи, що й зумовлює актуальність даного дослідження.

Аналіз останніх досліджень та публікацій. У науковій літературі оцифрування раритетних видань розглядається як один із ключових напрямів розвитку сучасної бібліотечно-інформаційної сфери та видавничої діяльності. Так, у роботі [1] акцентується увага на необхідності інтеграції новітніх сканувальних систем з алгоритмами попередньої обробки зображень для уникнення спотворень та втрати фрагментів тексту. У дисертаційній роботі [2] аналізується досвід українських бібліотек у сфері цифровізації фондів, підкреслюючи значущість міждисциплінарного підходу та необхідність розробки стандартів репрезентації цифрових копій. Також у роботі [4] акцентується увага на заходах як необхідності сьогодення: розробки та впровадження національної стратегії оцифрування історико-культурної спадщини, яка має охоплювати інтеграцію України та ЄС у культурному просторі.

У роботі [5] висвітлюється обмеженість функціоналу типових репозитаріїв і потребу у розробці спеціалізованих платформ для зберігання та представлення стародрукованих видань, що відповідають вимогам автентичності та повноти. Крім того, у публікації [6] розглядається приклад Національної бібліотеки Франції, яка впровадила інноваційні підходи до цифрового збереження документів, зокрема створення багаторівневих копій та метаданих.

У практикумі [7] акцентується увага на використанні мультимедійних технологій для візуального відтворення раритетних видань і зазначають потенціал гібридних моделей сканування з додатковими модулями обробки зображень.

Попри наявність окремих технічних напрацювань, більшість досліджень зосереджені на загальних питаннях цифрового збереження, тоді як проблема вдосконалення саме процесу оцифрування стародрукованих видань потребує подальшого ґрунтовного вивчення.

Мета статті – дослідження сучасних технологій оцифрування раритетних стародрукованих видань та перспектив модернізації процесу з урахуванням вимог до збереження, якості репрезентації та функціональності цифрових копій у сучасній видавничій діяльності.

Виклад основного матеріалу. Оцифрування стародрукованих видань можна розділити на два етапи: операція сканування і операція додаткової обробки отриманого цифрового зображення. Якщо якість сканування в певній мірі є вирішеною

саме на апаратному рівні завдяки застосування камер з високою роздільною здатністю і забезпечення рівномірності освітлення, то на етапі обробки зображення виникають проблеми з якістю цієї обробки, адже це в першу чергу залежить від кваліфікації і можливостей оператора. Удосконалення технологій оцифрування стародруків у сучасних умовах цифрової трансформації потребує впровадження інноваційних ІТ-рішень, що поєднують методи комп'ютерного зору, глибокого навчання та обробки великих даних. Цифрове відтворення раритетних джерел більше не обмежується створенням статичних зображень – воно перетворюється на цілісну систему розпізнавання, відновлення й аналітики. Ефективність таких систем істотно зростає завдяки попередній обробці зображень, яка усуває дефекти та підвищує розбірливість тексту. Наявність автоматизованих алгоритмів дозволяє значно зменшити навантаження на фахівців, які здійснюють ручну ретуш та верифікацію кожної сторінки.

Штучний інтелект, особливо нейронні мережі, активно використовується для виявлення дефектів сканованих сторінок, зокрема викривлень, плям і розмитостей. Завдяки глибокому навчанню системи здатні самостійно коригувати пошкодження, що виникають внаслідок зношеності паперу або порушення геометрії друку. Застосування інтелектуальних моделей дозволяє зберігати не лише змістовну, а й візуальну автентичність історичних джерел. Зокрема, у його роботі наведено приклади підвищення точності розпізнавання за рахунок поєднання гібридних моделей згладжування з адаптивною контрастною обробкою [2]. Завдяки таким рішенням можна зменшити кількість помилок розпізнавання навіть у складних багатошарових фоліантах із ручними дописами. Сканування стародруків з використанням штучного інтелекту набуває комплексного характеру [3], поєднуючи технічну, лінгвістичну та культурологічну аналітику. Разом з тим, застосування інтелектуальних технологій у скануванні має низку викликів. Насамперед, ефективність роботи алгоритмів залежить від якості навчальних даних і їх відповідності до типу джерел, що оцифровуються. У випадках із унікальними пам'ятками виникає потреба в ручному доопрацюванні результатів, що частково нівелює переваги автоматизації. Для повноцінного впровадження таких технологій потрібна узгодженість між програмними рішеннями, технічним обладнанням і реставраційними практиками.

Серйозною проблемою залишається розпізнавання тексту у виданнях із архаїчними шрифтами, лігатурами та староукраїнською орфографією. Стандартні OCR-системи демонструють низьку ефективність при роботі з такими матеріалами, оскільки не здатні ідентифікувати нестандартні графеми та скорочення. На відміну від них, сучасні AI-OCR модулі застосовують контекстуальне навчання на історичних корпусах, що дозволяє автоматично виправляти помилки розпізнавання на основі мовної моделі. Такі сучасні системи можуть не лише коректно розпізнати текст, а й реконструювати логіку верстки, що є критичним для наукових досліджень структури книги. Вказані підходи також забезпечують можливість аналітичної обробки тексту з використанням лінгвістичних інструментів.

Як узагальнення опрацьованої інформації у табл.1.1 показано порівняльну характеристику класичних і інтелектуальних рішень у сфері оцифрування стародруків.

Таблиця 1.1

**Порівняння традиційних і інтелектуальних підходів
до оцифрування стародруків**

Ознака	Класичні рішення	ІТ-рішення з використанням штучного інтелекту
Покращення якості зображення	Ручне налаштування	Автоматичне згладжування, реконструкція
Розпізнавання тексту	OCR з низькою точністю	Контекстуальний AI-OCR із навчанням
Виявлення дефектів	Візуальна перевірка	Автоматичне виявлення та виправлення
Генерація метаданих	Каталогізатор	Автоматичне формування та класифікація
Гнучкість роботи з нестандартними шрифтами	Низька	Висока

Удосконалення стосується і створення метаданих. Сучасні системи на базі штучного інтелекту здатні автоматично визначати назви документів, імена авторів, дати, мови, тип шрифту та інші структурні елементи видання. Це дає змогу суттєво скоротити час на каталогізацію й індексацію, що особливо важливо при роботі з великими бібліотечними або архівними фондами. Такі системи дедалі частіше інтегруються з платформами хмарного зберігання, що забезпечує швидку обробку зображень і доступ до цифрових копій у будь-який момент. Як засвідчує досвід низки європейських бібліотек, ефективна організація процесу цифрового збереження ґрунтується на поєднанні технічних та управлінських рішень [6]. Поєднання технологій автоматизованого розпізнавання та хмарної інфраструктури створює гнучке середовище для ефективного керування цифровими колекціями.

У практичному аспекті активно розвиваються мобільні сканувальні рішення, які дають змогу здійснювати оцифрування в умовах польових експедицій або малих музеїв. Такі пристрої забезпечують оперативне сканування з подальшим передаванням зображень до хмарного середовища, де відбувається повна автоматизована обробка. Поєднання портативних систем із мультимедійною реконструкцією відкриває нові перспективи для представлення стародавніх джерел в інтерактивному форматі. Впровадження подібних підходів особливо актуальне для регіональних культурних установ, які мають обмежений доступ до централізованих ресурсів.

Таким чином, інтеграція сучасних ІТ-рішень у сферу сканування стародруків забезпечує якісно новий рівень збереження культурної спадщини. Поєднання інструментів штучного інтелекту, хмарних обчислень, мобільних сканувальних систем та мультимедійної реконструкції дає змогу не лише підвищити ефективність технічного процесу, а й розширити функціональність цифрових архівів і репозитаріїв.

Висновки. Отже, впровадження сучасних інформаційних технологій у процес сканування раритетних стародрукованих видань відіграє ключову роль у збереженні культурної спадщини, забезпечуючи її довготривале функціонування в цифровому середовищі. Завдяки інструментам комп'ютерного зору, глибокого навчання та обробки зображень стає можливим автоматичне покращення якості цифрових копій, виправлення дефектів, а також мінімізація втручання людини в процес оцифрування. Такі підходи особливо актуальні для великих бібліотечних та архівних фондів, що зберігають значні обсяги цінних документів. Цифрове сканування на основі інтелектуальних технологій поступово стає невіддільним етапом наукової, реставраційної та освітньої роботи.

Суттєвою перевагою сучасних ІТ-рішень є здатність виявляти дефекти, відновлювати пошкоджені фрагменти, а також розпізнавати тексти з архаїчними шрифтами та історичною орфографією. Нейронні мережі, натреновані на історичних корпусах, дозволяють досягати високої точності при роботі навіть зі складними друкованими структурами. Автоматизоване формування метаданих, а також інтелектуальна класифікація документів за змістовними або хронологічними ознаками істотно спрощують облік і пошук у цифрових репозитаріях. Це, у свою чергу, розширює можливості для міждисциплінарного дослідження історичних джерел.

Перспективи розвитку ІТ-рішень у сфері сканування раритетних видань залишаються вкрай позитивними. Інтеграція штучного інтелекту, хмарних платформ, мобільних сканерів і мультимедійної реконструкції створює потужний інструментарій для роботи з історичними документами в новому форматі. Це відкриває широкі можливості для розвитку цифрових гуманітарних наук, міждисциплінарної взаємодії та відкритого доступу до культурної спадщини. У найближчі роки очікується активне впровадження таких підходів у бібліотеках, архівах, музеях і наукових установах. Цифрове сканування на основі інтелектуальних технологій поступово формує новий етап у збереженні та репрезентації історико-культурної пам'яті, а також дозволяє виконувати якісний передрук раритетних джерел для широкого кола користувачів.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Гаврилишин О.Б., Репета В.Б., Миклушка І.З. Аналіз процесу оцифрування стародруків. Науковий вісник НЛТУ України. 2020. Т. 30, № 3. С. 106-110.
2. Шевченко М. О. Оцифрування фондів бібліотек України. Дисертація на здобуття наукового ступеня доктора філософії зі спеціальності 029 Інформаційна, бібліотечна та архівна справа. Міністерство культури та інформаційної політики України, Харківська державна академія культури, Харків, 2022. 267 с.
3. Automated Text Recognition with ChatGPT 4. URL: <https://www.annikarockenberger.com/2023-10-19/automated-text-recognition-with-chatgpt-4> (дата звернення 25.01.2025).
4. Shevchenko M. O. European Union experience in digitizing historical and cultural heritage: ways of implementation in Ukraine. *Science and Education a New Dimension. Humanities and Social Sciences*, VIII (38), June, 2020. P. 71-75.

5. Баркова О.В. Оцифровані джерела з історії освіти: вимоги та підходи до цифрової репрезентації книжкових пам'яток в онлайн-ресурсах. 2019. URL: https://lib.iitta.gov.ua/id/eprint/717857/1/%D0%91%D0%B0%D1%80%D0%BA%D0%BE%D0%B2%D0%B0_%D0%9E%D1%86%D0%B8%D1%84%D1%80%D0%BE%D0%B2%D0%B0%D0%BDi.pdf (дата звернення 22.01.2025).
6. Халецька Л. П. Цифрове збереження електронних документів на прикладі діяльності Національної бібліотеки Франції. Бібліотекознавство. Документознавство. Інформологія. 2015. № 4. С. 26-34.
7. Золотухіна К.І., Бардовський Б.О. Мультимедійні технології відтворення раритетних видань. Комп'ютерний практикум. Київ, НТУУ «КПІ» ім. І. Сікорського, 2022. 71 с.

REFERENCES

1. Havrylyshyn, O. Repeta, V., & Myklushka, I. (2020). Analysis of the digitization process of early printed books. *Scientific Bulletin of the NLTU of Ukraine*, 30(3), 106–110.
2. Shevchenko, M. (2022). *Digitization of library collections in Ukraine* (Doctoral dissertation, Kharkiv State Academy of Culture). Ministry of Culture and Information Policy of Ukraine. 267 p.
3. Automated Text Recognition with ChatGPT 4. URL: <https://www.annikarockenberger.com/2023-10-19/automated-text-recognition-with-chatgpt-4/>.
4. Shevchenko M. O. (2020). European Union experience in digitizing historical and cultural heritage: ways of implementation in Ukraine. *Science and Education a New Dimension. Humanities and Social Sciences*, VIII (38), June, 71–75.
5. Barkova, O. (2019). Digitized sources on the history of education: requirements and approaches to the digital representation of book monuments in online resources. Retrieved from https://lib.iitta.gov.ua/id/eprint/717857/1/%D0%91%D0%B0%D1%80%D0%BA%D0%BE%D0%B2%D0%B0_%D0%9E%D1%86%D0%B8%D1%84%D1%80%D0%BE%D0%B2%D0%B0%D0%BDi.pdf.
6. Khaletska, L. (2015). Digital preservation of electronic documents based on the example of the National Library of France. *Library Science. Documentation Science. Informology*, (4), 26–34.
7. Zolotukhina, K., Bardovskyi, B. (2022). *Multimedia technologies for reproducing rare editions: Computer workshop*. Kyiv: NTUU “KPI” named after I. Sikorskyi. 71 p.

doi: 10.32403/1998-6912-2025-1-70-42-48

TRENDS IN TECHNOLOGIES FOR DIGITIZING RARE OLD BOOKS

V. B. Repeta, O. B. Havrylyshyn, V. Ya. Fyll

*Lviv Polytechnic National University,
19, Pid Holoskom, St., Lviv, 79020, Ukraine
olena.b.havrylyshyn@lpnu.ua*

In the digital age, preserving and providing access to cultural heritage – especially rare early printed books – has become a key task for academic libraries and archives. Scanning such materials requires specialized technologies that take into account their physical fragility, historical printing methods, and the need for high-quality digital images. Enhancing scanning processes involves adapting modern equipment, using high-precision image processing algorithms, and ensuring minimal physical impact on the materials. This not only preserves the original authenticity of documents but also provides users with accessible and functional digital versions.

Digitizing early printed books poses multiple challenges – from preserving delicate bindings and unique typefaces to difficulties in recognizing texts printed using outdated orthographic norms. Key risks include damage during scanning, reduced image quality due to paper degradation or low contrast, and limitations of conventional software when handling non-standard formats. Data security and backup, along with protection against unauthorized access, are also essential in the digital preservation process.

The article examines both technical and organizational approaches to improving scanning technologies. It analyzes best practices from leading European libraries in digital restoration and presents practical solutions for handling defective pages, enhancing OCR accuracy, and managing metadata. It also explores how artificial intelligence and neural networks can be used to improve outcomes, automate classification, and reconstruct missing fragments. Attention is given to compliance with archival standards and ensuring long-term digital preservation.

Modern approaches to scanning rare printed books are vital for national cultural heritage strategies. They enable the creation of online repositories, support academic research and education, and help promote historical sources to a broader audience. However, successful implementation requires interdisciplinary collaboration between IT specialists, bibliographers, conservators, and archivists to achieve a balance between technical precision and scholarly authenticity.

Keywords: *scanning, digitization, early printed books, digital restoration, digital archives, document preservation, information technology, artificial intelligence..*

Стаття надійшла до редакції 05.05.2025.

Received 05.05.2025.