

<https://doi.org/10.31516/2410-5333.065.16>¹

УДК 004.91

О. Д. Красинський

аспірант спеціальності «Інформаційна, бібліотечна та архівна справа»,
Національний університет «Львівська політехніка», м. Львів, Україна

oleksii.d.krasynskiy@lpnu.ua

<https://orcid.org/0000-0002-5993-8496>

О. В. Марковець

кандидат технічних наук, доцент, завідувач кафедри соціальних комунікацій
та інформаційної діяльності, Національний університет «Львівська політехніка»,
м. Львів, Україна

oleksandr.v.markovets@lpnu.ua

<https://orcid.org/0000-0001-8737-5929>

МОЖЛИВОСТІ ЗАСТОСУВАННЯ ТЕХНОЛОГІЙ OCR ВІД GOOGLE ДЛЯ РОЗПІЗНАВАННЯ ТА ЦИФРОВІЗАЦІЇ АРХІВНИХ ДОКУМЕНТІВ

Стаття аналізує потенціал використання технологій оптичного розпізнавання символів (OCR) від Google для ефективної цифровізації архівних документів. Автор описує, як інтеграція Google Cloud Vision API може пришвидшити перетворення фізичних документів у цифровий формат, знижуючи вартість і підвищуючи доступність оцифрування для архівних установ різного масштабу. Також розглядаються переваги збільшення точності та якості оцифрованих даних, що сприяє їх подальшому науковому та освітньому використанню.

Аналіз включає огляд потенційних технічних та організаційних проблем, які можуть виникати під час впровадження OCR-технологій, зокрема сумісність систем та вимоги до безпеки даних. Пропонується розробка вебзастосунку для кінцевих користувачів, який зробить передові можливості OCR доступнішими та зручнішими для використання.

Метою дослідження є підвищення ефективності збереження інформації та забезпечення швидкого доступу до неї, зважаючи на актуальні потреби управління архівними ресурсами. Стаття також надає огляд останніх досліджень і публікацій, що демонструють ефективність Google OCR порівняно до інших продуктів і підкреслює можливості подальшого розвитку цих технологій для оптимізації процесів оцифрування в архівах.

Ключові слова: *OCR, Google, Google Vision, HTШ, Google Cloud Vision API.*

1 This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

O. Krasynskyi

postgraduate student of the specialty "Information, library and archival affairs", Lviv Polytechnic National University, Lviv, Ukraine

O. Markovets

Candidate of Technical Sciences, Head of Department of Social Communications and Information Activities, Lviv Polytechnic National University, Lviv, Ukraine

POSSIBILITIES OF USING OCR TECHNOLOGIES FROM GOOGLE FOR RECOGNITION AND DIGITALIZATION OF ARCHIVE DOCUMENTS

The relevance of the article. This article examines the significant capabilities of Google's Optical Character Recognition (OCR) technologies, particularly Google Cloud Vision API, in transforming the way archival documents are digitized. The rapid evolution of digital technologies offers unprecedented opportunities not only for the preservation but also for the enhanced accessibility of historical documents, which are susceptible to degradation and potential loss over time.

The purpose of the article. The primary objective of this study is to critically analyze and demonstrate the efficiency and effectiveness of employing Google Cloud Vision API to digitize diverse types of archival documents.

The methodology. The approach taken involves a comprehensive evaluation of Google Cloud Vision API's capability to accurately process and recognize text from a variety of archival materials, such as printed books, handwritten manuscripts, detailed maps, and other complex formats. The study methodically assesses the accuracy, processing speed, and cost-effectiveness of the OCR technology, comparing it to traditional digitization methods. Additionally, the research delves into the challenges of developing a user-friendly web interface that supports the seamless application of OCR technology.

Conclusion. The study highlights significant advantages of integrating Google OCR technologies into archival practices. These technologies not only accelerate the digitization process but also considerably reduce operational costs and enhance the quality of the digital outputs. Furthermore, the research identifies the necessity for customized approaches to OCR processing, particularly for complex data types such as musical scores and detailed cartographic materials. This adaptability is crucial for preserving the nuances and accuracy of diverse archival contents. The findings suggest promising avenues for future research and development, indicating that further innovations in OCR technology could revolutionize the field of archival science.

Expanded applications and implications. The study's implications extend beyond mere digitization; it explores the potential for Google Cloud Vision API to facilitate advanced data analysis and retrieval techniques. By enabling more sophisticated search functionalities, such as context-aware searches and metadata generation, OCR technology can transform archival resources into dynamic databases that support academic research, historical inquiries, and public education.

The practical significance. Building on the current findings, future research could explore the integration of OCR technology with blockchain to enhance the security and traceability of digitized documents. Exploring adaptive OCR techniques that can learn from user corrections and feedback may enhance the accuracy and reliability of digitization efforts, especially for documents with unusual fonts or degradation.

Keywords: *OCR, Optical Character Recognition, Google Cloud Vision API, digitization, archival materials, digital preservation, user interface development.*

Актуальність теми дослідження. Актуальність теми дослідження, яка стосується підбору та використання потужних сервісів від Google для оптимізації процесу оцифрування, впливає з кількох важливих аспектів сучасних потреб управління архівними ресурсами. По-перше, інтеграція передових технологій від Google, таких як Google Cloud Vision API, може істотно пришвидшити процес оцифрування великих обсягів документів, зображень, карт та інших архівних матеріалів. Це зменшує час, необхідний для перетворення фізичних документів у цифровий формат, що є критично важливим для збереження інформації та швидкого доступу до неї.

Здешевлення процесу оцифрування також є ключовим фактором. Використання автоматизованих інструментів OCR від Google дозволяє знизити вартість обробки даних, оскільки зменшується потреба в ручній праці, яка традиційно потрібна для введення даних та первинної обробки. Це робить процес доступнішим та вигіднішим для архівних установ різних масштабів, від малих до великих організацій.

Крім того, застосування цих технологій сприяє значному підвищенню якості результатів оцифрування. Сучасні сервіси OCR здатні розпізнавати текст з високою точністю, зокрема різні шрифти та рукописні тексти, що покращує достовірність оцифрованих даних та їх подальше використання для наукових досліджень, освіти та збереження культурної спадщини.

Зрештою інтеграція google-сервісів у процес оцифрування вможливує подальше використання цих даних у різноманітних аналітичних інструментах, що сприяє глибшому аналізу, виявленню зв'язків та тенденцій, які можуть бути неочевидними при традиційних методах дослідження. Таким чином, актуальність використання передових технологій Google для оцифрування архівних матеріалів полягає в значному впливі на ефективність, вартість та якість збереження й доступу до історичних ресурсів.

Постановка проблеми. Вибір оптимальних інструментів для оцифрування архівних матеріалів, зокрема використання технологій OCR від Google, актуалізує перед дослідниками та архівістами комплексну проблему, яка потребує аналізу вартості, якості результатів, людських ресурсів та технічних можливостей. Врахування цих аспектів важливе для забезпечення ефективного процесу цифровізації, який повинен бути водночас економічно вигідним і технічно вдосконаленим.

Використання технологій OCR від Google може значно знизити вартість оцифрування, оскільки ці інструменти дозволяють автоматизувати обробку великих обсягів даних, що традиційно потребує значних людських ресурсів. Така автоматизація допомагає зменшити час на обробку даних та витрати на персонал. Однак важливо враховувати, що висока ефективність розпізнавання OCR від Google може варіюватися залежно від якості та типу

вихідних матеріалів, особливо якщо вони містять рукописний текст або застарілі шрифти.

Якість результатів оцифрування є критичною, особливо коли йдеться про історичні документи, що містять важливу культурну і наукову інформацію. Технології Google OCR зазвичай надають високу точність у розпізнаванні друкованого тексту, але можуть мати обмеження при роботі з особливо складними архівними матеріалами.

Крім того, існують технічні та організаційні виклики, пов'язані з інтеграцією OCR технологій від Google в наявні системи управління архівними даними. Потрібно зважати на питання сумісності, безпеки даних та приватності, особливо в контексті дотримання нормативних вимог.

Таким чином, вибір OCR-технологій від Google як інструменту для оцифрування архівних матеріалів має включати ретельне вивчення всіх згаданих факторів, аби забезпечити баланс між вартістю, ефективністю, якістю та технічною виконуваністю в межах цього інституційного контексту.

Аналіз останніх досліджень і публікацій. Дослідження Томаса Гегггеммера в статті “OCR with Tesseract, Amazon Textract, and Google Document AI: a benchmarking experiment” (Hegghammer, 2021) охоплює аналіз та порівняння продуктивності трьох провідних OCR продуктів — Tesseract, Amazon Textract та Google Document AI. Експеримент проводився на зображеннях англійського та арабського текстів, що були відтворені з різними видами штучного шуму для створення корпусу з 18,568 документів, що включав 51,304 обробки запитів. Результати засвідчили, що Google Document AI продемонстрував найкращі результати, особливо на шумних документах, тоді як Tesseract відставав порівняно з обробками, що базуються на сервері. Відповідно до дослідження, точність OCR для англійської мови була значно вищою, ніж для арабської. Висновки цього дослідження можуть допомогти вченим краще ідентифікувати найефективніші рішення OCR для своїх наукових потреб, а зібрані тестові матеріали зберігаються в вільно доступному “Noisy OCR Dataset” (NOD) для використання в майбутніх дослідженнях (Hegghammer, 2021).

Стаття Saavedra та Uribe, опублікована у серії “Communications in Computer and Information Science” (Saavedra & Uribe, 2022), досліджує використання Google Cloud Vision API для обробки зображень з використанням Raspberry Pi, наголошуючи на перевагах інтеграції цих технологій у контексті обмежених обчислювальних ресурсів. Автори демонструють, як застосування OCR може бути виконане на невеликих платформах, таких як Raspberry Pi, що зменшує витрати на апаратне забезпечення при одночасному забезпеченні високої точності обробки. Згідно з розвідкою, OCR точність обробки номерних знаків у Колумбії за допомогою Google Cloud Vision

на Raspberry Pi становила 100%, що свідчить про високу ефективність цього підходу (Saavedra & Uribe, 2022).

Це дослідження підкреслює можливості зниження витрат і збільшення доступності технологій OCR для малого бізнесу або індивідуальних користувачів, які не можуть дозволити собі дорогі обчислювальні системи. Використання віддалених серверів Google для обробки даних дозволяє уникнути необхідності інвестицій у потужне апаратне забезпечення, зробивши технологію OCR більш доступною і економічно вигідною. Такий підхід може мати значний вплив на розширення використання OCR у різних доменах, включаючи автоматизацію офісу, розумний дім, електронну комерцію та інші застосування, де потрібна ефективна обробка візуальних даних.

У контексті оцифрування архівних документів дослідження Р. Б. Кравця, П. О. Бойка та О. В. Марковця (Кравець та ін., 2023) підкреслює значення переходу на електронні архіви для оптимізації документообігу. Автори аналізують, як електронні архіви можуть сприяти швидшому та безпечнішому доступу до управлінської інформації, особливо у великих організаціях, таких як міська рада. Науковці стверджують, що впровадження сучасних інформаційних технологій дозволяє значно знижувати витрати та підвищувати ефективність збереження документів (Кравець та ін., 2023).

Мета статті — визначити потенціал та можливості використання технологій оптичного розпізнавання символів від Google для ефективного оцифрування та подальшого аналізу архівних матеріалів. А саме, як ці інноваційні технології можуть вирішити сучасні виклики, пов'язані зі збереженням, доступністю та аналізом великої кількості історичних документів, рукописів, фотографій та інших архівних ресурсів.

Завдання дослідження. Основні завдання дослідження, які обговорюються, охоплюють аналіз можливостей технологій оптичного розпізнавання символів (OCR) від Google для ефективної цифровізації архівних документів. Вони передбачають вивчення того, як інтеграція Google Cloud Vision API може сприяти швидкому перетворенню фізичних документів у цифровий формат, зниженню вартості та підвищенню доступності оцифрування. Також метою є оцінка збільшення точності та якості оцифрованих даних, що може вплинути на їх наукове й освітнє використання. Дослідження також включає аналіз потенційних технічних та організаційних викликів, які можуть виникати під час впровадження OCR-технологій, зокрема сумісності систем і вимог до безпеки даних.

Виклад основного матеріалу дослідження. У сучасному світі, де інформація є основою розвитку та інновацій, процес оцифрування документів і матеріалів набуває особливої актуальності. Оцифрування дозволяє зберігати,

аналізувати та широко використовувати дані, які раніше були доступні лише у фізичному форматі. Розгляд різноманітних типів даних, які можуть бути перетворені в цифровий формат, демонструє широту потенційних застосувань і підкреслює значення цього процесу в сучасному інформаційному суспільстві.

Типи даних, які потребують оцифрування:

1. Текстові документи: оцифрування текстових документів, таких як книги, журнали і наукові статті, робить інформацію доступнішою та легкою для пошуку.
2. Рукописний текст: включає історичні рукописи та особисті нотатки, які часто мають унікальний стиль запису та орфографію.
3. Зображення: оцифрування зображень (фотографії чи картини) потребує збереження високої якості та деталей.
4. Формули: спеціалізовані наукові та математичні формули вимагають точного розпізнавання для правильного представлення в цифровій формі.
5. Музичні ноти: оцифрування музичних композицій потребує не тільки точності, а й здатності відтворювати музику відповідно до оригіналу.
6. Літературні твори: вірші та інші твори зі структурними особливостями, такими як віршована форма, потребують особливої уваги під час оцифрування.
7. Карти та плани: географічні карти та архітектурні плани містять велику кількість інформації та деталей, що мають бути точно відображені.
8. Наукові діаграми та графіки: елементи є важливими для зображення складних даних та аналітичних результатів.
9. Історичні та культурні артефакти: оцифрування таких об'єктів дозволяє зберегти їх для майбутніх поколінь та забезпечує ширший доступ.

Кожен із цих типів даних вимагає індивідуального підходу та використання спеціалізованих технологій для досягнення найкращих результатів у процесі цифровізації. Вибір правильних інструментів та методів для кожного типу даних є ключовим для успішного виконання процесу оцифрування, який відіграє критичну роль у збереженні та доступі до цінної інформації.

Процес оцифрування та розпізнавання даних відіграє вирішальну роль у збереженні та доступності інформації. За допомогою сучасних технологій, таких як оптичне розпізнавання символів від корпорації на кшталт Google, ми отримуємо можливість не тільки цифровізувати текст, а й розширювати його використання завдяки різноманітним обробкам.

Розпізнавання та обробка текстових даних. Текстові документи, такі як книги, наукові статті, історичні листи, після оцифрування відкривають широкі можливості для подальшої роботи. Оцифрований текст може бути

не лише прочитаним людиною, а й автоматично перекладеним на інші мови, що робить інформацію доступною для ширшої аудиторії. Більше того, створення абстрактів та анотацій дозволяє користувачам швидше зорієнтуватися в змісті документів, визначаючи їхню релевантність до потреб дослідження або пошуку.

Індексація та повноцінний пошук за змістом є ще однією перевагою, яка стає можливою завдяки технологіям розпізнавання тексту. Це дозволяє не просто знаходити документи за ключовими словами, а й аналізувати їхній зміст, забезпечуючи більш глибокий та контекстуальний пошук.

Оцифрування візуальних і графічних матеріалів. На відміну від текстових документів, візуальні та графічні матеріали, такі як фотографії, картини, карти й плани, потребують іншого підходу до оцифрування. Хоча ці об'єкти і зберігаються у вигляді графічних зображень, до них можна додати метадані, що значно розширює можливості пошуку та класифікації.

Метадані можуть включати інформацію про автора, дату створення, географічне розташування, тему або будь-які інші дані, які допоможуть ідентифікувати та класифікувати зображення. Такий підхід не тільки полегшує пошук потрібних матеріалів серед великих архівів, а й дозволяє користувачам знаходити зображення за специфічними критеріями, такими як місцезнаходження або період часу.

Застосування сучасних технологій OCR та AI відкриває нові горизонти у сфері оцифрування та розпізнавання різних типів даних. Розробка спеціалізованих алгоритмів для розпізнавання складних наукових формул, музичних нот, а також інтерпретації візуальних та графічних матеріалів може значно покращити якість і доступність цифрових архівів.

Важливо підкреслити, що успіх оцифрування залежить не лише від використання передових технологій, а й від ретельного планування процесу та визначення відповідних метаданих для кожного типу даних. Це вимагає глибокого розуміння контексту матеріалів, що оцифровуються, та потреб користувачів, які будуть з ними працювати.

Google Cloud Vision API ("Cloud Vision API", 2024) представляє собою революційний інструмент для обробки та аналізу архівних даних, зокрема для цінних колекцій, як-от ті, що належать науковому товариству імені Шевченка (НТШ). Цей потужний сервіс машинного навчання може драматично трансформувати процеси цифровізації, розпізнавання та аналізу різноманітних архівних матеріалів:

- текстовий та рукописний матеріал: Google Cloud Vision API ефективно впорається з розпізнаванням як друкованого, так і рукописного тексту на зображеннях. Це відкриває широкі перспективи для оцифрування та подальшої обробки історичних документів, наукових праць, листування та інших текстових артефактів, що зберігаються в архівах НТШ.

- фотографії та візуальні матеріали: здатність API розпізнавати об'єкти, людей і сцени на фотографіях та інших візуальних матеріалах значно полегшує класифікацію, архівацію й вивчення історичних зображень. Це сприяє глибшому розумінню візуального контексту архівних колекцій.
- відомі місця та логотипи: автоматичне розпізнавання відомих місць, будівель і логотипів на зображеннях допомагає додати контекст до архівних фотографій та ілюстрацій, що може виявитися надзвичайно корисним для історичних і культурних досліджень.
- аналіз емоцій: хоча Google Cloud Vision API не здатен ідентифікувати конкретних осіб, він може аналізувати емоції на фотографіях, надаючи унікальну можливість вивчення емоційного стану та атмосфери, зафіксованої на історичних знімках.

Використання Google Cloud Vision API для обробки архівних даних, таких як колекції НТШ, не лише значно спрощує та прискорює процес оцифрування інформації, а й розкриває нові горизонти для аналізу та інтерпретації історичних матеріалів. Використовуючи передові можливості штучного інтелекту та машинного навчання, архіватори й дослідники можуть глибше зануритися в історію та культуру, збережену в документах і зображеннях, відкриваючи нові перспективи для вивчення минулого.

Хоча Google Cloud Vision API є надзвичайно потужним інструментом для розпізнавання та аналізу зображень, включаючи оцифрування тексту й візуальний аналіз, він сам по собі не має зручного користувачького інтерфейсу, який дозволив би кінцевим користувачам безпосередньо використовувати його можливості без спеціалізованих технічних знань. Тому для того, щоб зробити ці передові функції доступними широкому колу користувачів, таких як архівісти, дослідники, історики та ін., необхідно розробити та інтегрувати Google Cloud Vision API у вебзастосунок зі зручним та інтуїтивно зрозумілим інтерфейсом.

Такий вебзастосунок стане мостом між потужними аналітичними можливостями Google Cloud Vision API та кінцевими користувачами, дозволяючи їм легко оцифрувати, аналізувати та управляти великими обсягами даних без необхідності володіння спеціальними знаннями у сфері програмування чи машинного навчання. Така інтеграція не тільки спростить процес цифровізації та аналізу даних, а й забезпечить ширший доступ до цих технологій для сприяння дослідженням, освіті та збереженню культурної спадщини.

Оцінка вартості використання. Для оцифрування середньої книги на 300 сторінок за допомогою Google Cloud Vision API і з урахуванням можливості повторних запитів через недостатню якість деяких сторінок, ми можемо розрахувати вартість, виключивши безкоштовний рівень. Це надасть

нам розуміння загальних витрат при використанні цієї служби на більш значному масштабі.

Припущення:

- вибрана функція для оцифрування: Document Text Detection (Geewax, 2018), оскільки вона найкраще підходить для обробки документів з щільним текстом.
- припускаємо, що 10% сторінок потребуватимуть повторної обробки у зв'язку з поганою якістю зображення або іншими проблемами з розпізнаванням, що загалом додасть 30 додаткових сторінок до обробки.

Розрахунок вартості:

1. Основна кількість сторінок: 300 сторінок
2. Додаткові сторінки для повторної обробки: 30 сторінок
3. Загальна кількість сторінок для обробки: 330 сторінок

Згідно з тарифами станом на квітень 2024 (“Cloud Vision pricing”, 2024), вартість “Document Text Detection” становить \$1,50 за кожні 1000 запитів після перших 1000 (які ми ігноруємо, оскільки перші 1000 сторінок є безкоштовні, а нас цікавить об'єктивна вартість).

Для 330 сторінок вартість розраховується так: оскільки вартість вказана за 1000 сторінок, а нам потрібно обробити лише 330, розрахунок буде наступним:

$$\frac{330 \text{ сторінок}}{1000 \text{ сторінок}} \times \$1.50 = \$0.495 \approx 20 \text{ грн}$$

Таким чином, вартість оцифрування однієї середньої книги на 300 сторінок з додатковим запасом для повторної обробки становить приблизно 50 центів за книгу.

Ціна виявляється доволі доступною навіть для великих масштабів оцифрування, що робить Google Cloud Vision API економічно вигідним варіантом для проектів із цифровізації документів. Така низька вартість може значно спростити вирішення бюджетних питань для організацій, що планують масштабні проекти оцифрування. Врахування додаткових витрат на можливі повторні обробки забезпечує реалістичне планування бюджету, що є ключовим для успішної реалізації проектів.

Оцінка вартості устаткування та розробки сервісу для інтеграції з Google Cloud Vision API включає кілька ключових компонентів, кожен із яких має свої фінансові наслідки. Зокрема, потрібно врахувати витрати на сканувальне обладнання та на проектування й підтримку програмного забезпечення.

Сканери, необхідні для оцифрування документів, можуть мати широкий діапазон цін — від кількох сотень до кількох тисяч доларів, залежно від

їх продуктивності та якості. Важливо обрати таке обладнання, яке відповідатиме вимогам проекту, забезпечуючи оптимальне співвідношення ціни та якості.

Розробка програмного забезпечення для інтеграції з Google Cloud Vision API може бути значною статтею витрат. Ці витрати включають проектування, реалізацію, тестування та підтримку системи. Вартість може значно коливатися, залежно від складності проекту, кількості інтегрованих функцій та тривалості процесу розробки.

До інших витрат можуть входити навчання персоналу, витрати на інфраструктуру та потенційні ліцензійні збори. Усе це разом формує загальну картину витрат, яка повинна бути ретельно спланована для забезпечення економічної ефективності проекту. Навіть з урахуванням можливості повторного сканування у зв'язку з проблемами з якістю, загальні витрати можуть бути контрольованими за умови точного планування та управління ресурсами.

Висновки. На основі нашого дослідження, стаття висвітлює значні переваги використання технологій оптичного розпізнавання символів (OCR) від Google для оцифрування архівних матеріалів. Використання Google Cloud Vision API значно пришвидшує процес перетворення фізичних документів у цифровий формат, що є критично важливим для збереження інформації та швидкого доступу до неї. Це також дозволяє знизити загальні витрати на оцифрування, мінімізуючи потребу в ручній праці й роблячи процес більш доступним і вигідним для архівних установ різного масштабу.

Сучасні технології OCR від Google забезпечують високу точність розпізнавання тексту, що покращує достовірність оцифрованих даних та розширює можливості їх подальшого використання. Інтеграція передових технологій Google також відкриває нові можливості для глибшого аналізу зібраних даних, що може виявити неочікувані зв'язки та тенденції.

Однак упровадження OCR-технологій потребує уваги до сумісності систем, безпеки даних та дотримання нормативних вимог. Також існує потреба в розробці інтуїтивно зрозумілого вебінтерфейсу, що дозволить широко використовувати ці технології без спеціальних технічних знань. Такі кроки забезпечать інтеграцію передових технологій OCR у сучасні архівні практики, значно поліпшуючи доступ, ефективність та якість збереження культурної спадщини.

Перспективи подальших досліджень. Подальші дослідження у сфері оцифрування архівних матеріалів за допомогою технологій оптичного розпізнавання символів від Google мають зосередитись на розробці універсального вебзастосунок, що інтегрується з Google Cloud Vision API. Такий застосунок повинен надавати інтуїтивно зрозумілий інтерфейс, що дозволяє користувачам легко завантажувати та обробляти зображення, покращувати

їх якість і виконувати різноманітні аналітичні завдання. Важливим аспектом також є розробка спеціалізованих функцій для обробки складних типів даних, таких як музичні ноти, картографічні матеріали та наукові діаграми, які вимагають високої точності й специфічних підходів до розпізнавання.

Окрім технічного розвитку, науковцям варто звернути увагу на оптимізацію процесів колаборації та управління доступом, що зробить веб-застосунок корисним інструментом для групової роботи над проектами. Поліпшення методів OCR для складних даних може значно підвищити якість та доступність цифровізованих ресурсів, зумовлюючи збереження та вивчення культурної спадщини. Це також сприятиме розвитку нових напрямів досліджень і допоможе організаціям ефективно управляти великими обсягами цифрових даних, забезпечуючи їхню безпеку та легкий доступ.

Список посилань

- Кравець, Р. Б., Бойко, П. О., & Марковець, О. В. (2023). Електронний архів як засіб швидкого доступу до управлінської інформації. *Бібліотекознавство. Документознавство. Інформологія*, 4, 14–21.
- “Cloud Vision API” (2024, Квітня 15). <https://cloud.google.com/vision?hl=en>
- “Cloud Vision pricing” (2024, Квітня 15). <https://cloud.google.com/vision/pricing#prices>
- Geewax, John J. (2018). *Google Cloud Platform in Action*. Manning Publications.
- Hegghammer, T. (2021). “OCR with Tesseract, Amazon Textract, and Google Document AI: a benchmarking experiment”. *Journal of Computational Social Science* (2022), 23, Червень 2021.
- Saavedra, S. V., & Uribe, A. L. (2022). “Google Cloud Vision and Its Application in Image Processing Using a Raspberry Pi”. *Communications in Computer and Information Science*, 17, Жовтень 2022.

References

- Kravets, R. B., Boiko, P. O., & Markovets, O. V. (2023). Electronic archive as a means of quick access to management information. *Bibliotekoznavstvo. Dokumentoznavstvo. Informolohiia*, 4, 14–21. [In Ukrainian].
- “Cloud Vision API” (2024, April 15). <https://cloud.google.com/vision?hl=en>
- “Cloud Vision pricing” (2024, April 15). <https://cloud.google.com/vision/pricing#prices>. [In English].
- Geewax, John J. (2018). *Google Cloud Platform in Action*. Manning Publications. [In English].
- Hegghammer, T. (2021). “OCR with Tesseract, Amazon Textract, and Google Document AI: a benchmarking experiment”. *Journal of Computational Social Science* (2022), 23, June 2021. [In English].
- Saavedra, S. V., & Uribe, A. L. (2022). “Google Cloud Vision and Its Application in Image Processing Using a Raspberry Pi”. *Communications in Computer and Information Science*, 17, October 2022. [In English].

Надійшла до редколегії 15.03.2024