



ЦИФРОВА ТРАНСФОРМАЦІЯ БІБЛІОТЕК, НАУКИ Й ЛОКАЛЬНИХ ІНІЦІАТИВ

DIGITAL TRANSFORMATION OF LIBRARIES, SCIENCE AND LOCAL INITIATIVES

УДК 004.932:[929.5:930.25

DOI: 10.31866/2617-796X.8.1.2025.335538

Дмитро Кузьменко,

бакалавр,

Харківський національний університет радіоелектроніки,

Харків, Україна

dmytro.kuzmenko@nure.ua

<https://orcid.org/0009-0006-1366-9530>

Ігор Гребеннік,

доктор технічних наук, професор,

завідувач кафедри системотехніки,

Харківський національний університет радіоелектроніки,

Харків, Україна

igor.grebennik@nure.ua

<https://orcid.org/0000-0003-3716-9638>

Інна Урняєва,

доктор філософії, доцент кафедри системотехніки,

Харківський національний університет радіоелектроніки,

Харків, Україна

inna.urniaieva@nure.ua

<https://orcid.org/0000-0001-9795-6954>

ІНФОРМАЦІЙНО-ПОШУКОВА СИСТЕМА ГЕНЕАЛОГІЧНИХ ДОСЛІДЖЕНЬ

Метою статті є розробка прототипу цифрової платформи для підтримки генеалогічних досліджень, яка об'єднує архівні джерела, ДНК-аналіз, обробку файлів, зокрема зображень, комплексний пошук за текстовими і візуальними даними, а також побудову родовідного дерева в єдиному інформаційному середовищі. Обґрунтовано необхідність створення національного сервісу, орієнтованого на потреби українських користувачів, з підтримкою пошуку як у внутрішніх, так і в зовнішніх джерелах через інтеграцію з відкритими архівами та державними реєстрами.

© Кузьменко Д. С.,

© Гребеннік І. В.,

© Урняєва І. А.

Методи дослідження. У роботі застосовано аналітичний метод для дослідження обмежень наявних рішень у сфері генеалогічного пошуку, а також метод системного аналізу для формування вимог до архітектури майбутньої цифрової платформи. Проведено порівняльне оцінювання функціоналу міжнародних і вітчизняних сервісів з метою виокремлення ключових компонентів, необхідних для побудови комплексної інформаційної системи.

Наукова новизна. Відмінністю запропонованої системи є об'єднання всіх ключових складників генеалогічного дослідження в межах єдиної цифрової платформи, що охоплює ведення архівних записів, повнотекстовий пошук, візуальний пошук за обличчями, ДНК-аналіз та побудову родоводу. На відміну від наявних сервісів, запропоновано інтеграцію не тільки з історичними архівами, а й з актуальними державними реєстрами. Система підтримує електронну автентифікацію користувачів через такий інструмент, як Дія. Підпис, і взаємодію з реєстрами через систему «Трембіта». Інноваційним є використання штучного інтелекту для автоматизованої обробки файлів, зокрема зображень, з метою розпізнавання тексту й ідентифікації облич. Крім того, реалізовано спрощений підхід до генетичного аналізу – система працює із «сирими» даними секвенування, що дає змогу користувачам зменшити витрати й уникнути обмежень, пов'язаних з конкретними лабораторіями.

Висновки. Розроблений прототип цифрової платформи демонструє ефективний підхід до об'єднання архівних, візуальних та генетичних даних, що дає змогу значно спростити генеалогічні дослідження. Система охоплює повний цикл: від оцифрування та обробки джерел до побудови родоводу й пошуку родичів. Дослідження має наукове й технічне значення, а також це вагомий внесок у збереження історичної пам'яті та культурної спадщини України, що є важливим для формування ідентичності й передачі знань майбутнім поколінням.

Ключові слова: генеалогія; родовід; цифрова спадщина; штучний інтелект; пошук за обличчям; розпізнавання тексту; архівні документи; ДНК-дані; мікросервісна архітектура.

Вступ. Дослідження власного родоводу є важливим аспектом збереження історичної пам'яті та сімейної спадщини. Однак цей процес супроводжується численними викликами, які значно ускладнюють пошук і систематизацію інформації.

Історичні документи, що містять інформацію про предків, розпорошені між численними джерелами: записами актів цивільного стану, ревізійними списками, церковними реєстрами, військовими звітами, переписами населення тощо. Однак ці документи створювали з різними цілями, тому вони не завжди мають взаємопов'язані дані, що ускладнює їх узгодження. Навіть у випадках, коли записи збереглися, вони часто розміщені в різних архівах без єдиного механізму пошуку та доступу (Котович, 2024).

Навіть якщо документи збереглися в архівах, їх відсутність у цифровому форматі суттєво ускладнює дослідження. Багато історичних матеріалів представлені у вигляді сканованих копій або фотографій, що не підтримують текстовий пошук. Це змушує дослідників переглядати кожен документ вручну, що значно збільшує час і трудовитрати (Family Tree, 2024).

Доступ до архівних документів часто є обмеженим через бюрократичні процедури. Багато архівів вимагають офіційних запитів, оплати за перегляд доку-

ментів або навіть особистої присутності дослідника в архіві. У деяких випадках документи зберігаються в закритих фондах. Їх можуть не видавати через фізичний стан. А особливо цінні документи можуть бути недоступними навіть для професійних істориків (Family Tree, 2024). Це створює додаткові бар'єри для пошуку інформації, особливо для людей, які перебувають в іншій країні або не мають можливості особисто відвідати архів. Якщо ж записи про предків містяться в обмежено доступних документах, установлення родинних зв'язків стає практично неможливим (Komisarenko, 2017).

Ще однією проблемою є те, що навіть після отримання доступу до архівного документа немає гарантії, що він містить необхідні відомості. Це означає, що дослідник може витратити багато часу та коштів на перегляд документів, які не принесуть результатів.

Частина архівних даних навіть у відсканованому вигляді залишається важко-доступною через необхідність додаткової обробки. Наприклад, і для професійних істориків рукописні документи можуть бути складними, адже їх важко прочитати через особливості почерку, використання старих орфографічних норм або поганий фізичний стан записів (вицвіле чорнило, пошкоджений папір) (Family Tree, 2024). Це унеможливує автоматичне витягування тексту стандартними методами та значно ускладнює роботу з великими обсягами архівних даних.

Ще одним викликом є складність самостійного ведення генеалогічного дерева. У міру додавання нових предків і нащадків генеалогічне дерево розростається в геометричній прогресії, що ускладнює його систематизацію та оновлення. Звичайні засоби ведення родоводу, такі як записи в блокнотах або створення схем вручну, є неефективними для збереження великої кількості даних. Окрім основних біографічних даних, до родоводу можуть додаватися супровідні матеріали, такі як листи, фотографії, спогади, що ще більше ускладнює їхню систематизацію та збереження в структурованому вигляді.

Через історичні події, втрату документів або їх знищення, а також природний плин часу багато документів, що містять важливу інформацію про сімейну історію, втрачені або недоступні назавжди (Котович, 2024). У таких випадках дослідники намагаються відновити зв'язки, шукаючи родичів не тільки за документами, а й за допомогою ДНК-аналізу, який дає змогу виявити генетичні зв'язки між людьми, навіть якщо документальних підтверджень спорідненості немає (Аналізи ДНК в Україні, б.д.). Це допомагає знаходити не тільки прямих предків, а й бокові гілки родоводу – живих далеких родичів, які можуть зберігати важливі фрагменти сімейної історії, що були втрачені через час чи знищення матеріальних носіїв.

Однак наявні ДНК-сервіси працюють переважно в межах власних баз даних, що обмежує можливості пошуку. Якщо потенційний родич не проходив тестування в певній лабораторії, знайти його за генетичними збігами неможливо. Це змушує дослідників проводити тести в різних компаніях, що значно збільшує витрати і не гарантує успішного результату.

Розвиток інформаційно-пошукових систем для дослідження родоводу може значно спростити та прискорити процес встановлення сімейних зв'язків. Викори-

стання сучасних технологій дає змогу не тільки оцифровувати архівні документи, а й забезпечувати їхню автоматичну обробку й аналіз.

Одним з ключових напрямів є використання штучного інтелекту для розпізнавання рукописного тексту, облич на старих фотографіях й автоматичного встановлення зв'язків між історичними записами. Такі алгоритми здатні враховувати можливі помилки в написанні прізвищ, варіації імен і розбіжності в датах, що значно покращує якість пошуку та підвищує ефективність досліджень (Akdoğan, 2015).

Ще одним важливим кроком є створення єдиної централізованої бази генеалогічних даних, яка б інтегрувала інформацію з різних архівів, державних реєстрів та приватних колекцій. Це дасть змогу дослідникам отримати доступ до широкого масиву даних, уникаючи необхідності взаємодії з численними окремими джерелами.

Огляд аналогів. На сьогодні є низка платформ, що допомагають у дослідженні родоводу, проте жодна з них не забезпечує повноцінного комплексного підходу. Деякі сервіси спеціалізуються на історичних документах, інші – на ДНК-аналізі, але всі вони мають суттєві обмеження.

Одним з найпопулярніших міжнародних сервісів є MyHeritage, Ancestry та FamilySearch (n.d.). Вони пропонують великі бази історичних документів і можливість ведення родовідного дерева, ДНК-аналізу, що робить їх потужними інструментами для генеалогічних досліджень. Пріоритетною ділянкою роботи цих платформ є архіви Сполучених Штатів Америки та західної частини Європи, що значно ускладнює пошук для жителів України й сусідніх країн. Крім того, вони не мають доступу до українських архівів та не мають функцій розпізнавання облич, що обмежує можливості роботи з історичними фотографіями. Доступ до більшості функцій є платним, що створює фінансовий бар'єр для широкого загалу користувачів (Free Family Tree, n.d.; Ancestry | Family Tree, n.d.).

Українські системи, такі як Pra.in.ua і Ridni.org, надають доступ до оцифрованих архівних документів та метричних книг. Головна перевага – зосередженість на українських джерелах, що робить їх більш придатними для дослідників місцевої історії. Сервіс Pra.in.ua, що входить до генеалогічного товариства «Рідні», надає обмежений пошук по вже оцифрованих архівах, які доступні безкоштовно. У свою чергу Ridni.org має власні додаткові архіви, представлені переважно у вигляді відсканованих документів, що ускладнює автоматизований пошук, оскільки документи потрібно переглядати вручну (База жителів України 1650-1920 рр., б.д; Рідні – товариство українських родовідників, б.д.).

Перевагою Ridni.org є можливість пошуку у відкритих зовнішніх архівних ресурсах, що частково компенсує зазначені обмеження. Водночас обидві системи не підтримують ДНК-аналіз, що зменшує можливість встановлення генетичних зв'язків між родичами. Деякі архіви доступні лише на комерційній основі, що також знижує загальну доступність інформації (Рідні – товариство українських родовідників, б.д.).

Ridni.org має комерційну спрямованість на надання генеалогічних послуг, що впливає на функціонал для звичайних користувачів. Замість відкритої бази даних та можливості самостійного проведення досліджень, сервіс пропонує послуги

з побудови родоводу за оплату, що обмежує доступ до архівних матеріалів для незалежних дослідників (Рідні – товариство українських родовідників, б.д.).

Ще одним сегментом ринку є ДНК-лабораторії, зокрема медико-генетичний центр «Мама Папа». Вони спеціалізуються на аналізі генетичного матеріалу для визначення етнічного походження та пошуку родичів. Проте головним недоліком таких сервісів є те, що вони працюють лише в межах власних баз даних. Якщо потенційні родичі не проходили тестування в цій лабораторії, знайти їх буде неможливо. Крім того, такі платформи не інтегруються з архівами й історичними джерелами, що обмежує їхню ефективність у комплексному дослідженні родоводу (Аналізи ДНК в Україні, б.д.).

Загалом жоден з розглянутих варіантів не дозволяє виконати повноцінне дослідження родоводу українців, яке б об'єднувало автоматизований текстовий пошук (за місцевими архівами), пошук родичів на архівних фото, ведення родоводу та ДНК-аналіз. Міжнародні сервіси мають широкий функціонал, але не враховують потреб українських користувачів, тоді як локальні сервіси є менш численними та спеціалізуються на окремих аспектах досліджень. Українські архівні платформи мають обмежений пошук і не інтегруються з ДНК-базами, а генетичні лабораторії не забезпечують доступу до історичних джерел, що значно звужує можливості комплексного аналізу.

Постановка завдання. У зв'язку з виявленими обмеженнями наявних рішень постає необхідність у створенні єдиної цифрової платформи, яка забезпечить комплексну підтримку генеалогічних досліджень в українському контексті. Наявні сервіси фрагментовані: одні працюють лише з архівними документами, інші працюють виключно з ДНК-аналізом або спеціалізуються на побудові родовідного дерева. Це змушує дослідника самостійно поєднувати різні джерела, сервіси й інструменти, що потребує значних зусиль, часу та спеціальних знань у кількох суміжних галузях.

Основне завдання полягає в розробці єдиного інформаційного середовища, яке об'єднувало б в одному місці всі ключові компоненти для повноцінного дослідження родоводу. Зокрема, платформа має забезпечувати інтерактивну базу оцифрованих архівних документів, модулі автоматичного розпізнавання рукописного й друкованого тексту, інструменти ведення та візуалізації родовідного дерева, механізми аналізу генетичних даних користувачів, повнотекстовий пошук у внутрішніх і зовнішніх джерелах, а також технології розпізнавання облич на архівних фотографіях.

Запропонована система має стати єдиною точкою входу для дослідників, незалежно від їхнього рівня підготовки, місця проживання чи технічних навичок. Вона має забезпечити централізований, але гнучкий доступ до різних джерел і функціональних можливостей через зручний вебінтерфейс. Одним з ключових вимог до платформи є її здатність до масштабування як з технічного погляду (через мікросервісну архітектуру), так і з функціонального (через підтримку нових типів даних, мов інтерфейсу, інтеграцій із зовнішніми системами тощо).

Особливу увагу варто приділити безпеці персональних даних і правовому регулюванню доступу до чутливої інформації, а також валідації джерел, що дасть змогу зберігати наукову достовірність. Система має забезпечувати можливість простежити походження кожного запису, зокрема через збереження посилань на

цифрові копії документів. Крім цього, важливо передбачити можливість залучення волонтерів до процесу оцифрування та наповнення архіву – введення текстової інформації, розпізнавання документів та додавання супровідних матеріалів.

Отже, завданням розробки є створення національної генеалогічної платформи нового покоління, яка об'єднає зусилля державних архівів, громадськості, науковців та IT-фахівців задля збереження і дослідження родинної та історичної пам'яті українців у цифровому форматі.

Опис прототипу. У розробці акцентовано на оцифруванні архівних документів та інтеграції з уже наявними джерелами. Система має містити механізми, які дозволять не лише зберігати цифрові копії історичних матеріалів, але й забезпечувати їхню структуризацію та зручний пошук. Основою цього процесу є відсканування документів, що дає змогу зробити архівні матеріали доступними без необхідності фізичного контакту з оригіналами. Це критично важливо для збереження історичних джерел, оскільки сприяє мінімізації впливу зовнішніх факторів на їхній стан і продовжує термін зберігання.

Оцифровані документи підлягають подальшій обробці, яку здійснюють як адміністратори системи, так і волонтери, що вводять дані через вебінтерфейс. Проте цей процес є рутинним і вимагає значних витрат часу, особливо коли йдеться про старі документи з рукописним текстом (Komisarenko, 2017). Тому важливим етапом є використання технологій штучного інтелекту, зокрема програм для автоматичного розпізнавання тексту. Попередньо навчена модель здатна обробляти як друковані, так і рукописні тексти, ураховуючи особливості історичної орфографії та мови. Це значно прискорює процес введення даних, оскільки оператору достатньо лише перевірити правильність автоматичного розпізнавання та внести необхідні корективи.

Інтеграція з наявними архівними базами є ще одним ключовим аспектом системи. Передбачається взаємодія з державними установами, такими як архіви національного та регіонального рівня, а також з приватними та загальнодоступними ресурсами, наприклад платформою Pra.in.ua. Для цього необхідне укладення відповідних угод про обмін даними, що передбачатиме регулювання рівня доступу, захист інформації та дотримання законодавства у сфері обробки персональних даних.

Окрім текстових архівних даних, система має підтримувати збереження супутньої інформації у вигляді відсканованих документів, фотографій, листів та інших матеріалів. Важливим нововведенням є застосування технологій розпізнавання облич, що дає змогу автоматично знаходити всі фото, де одна й та сама людина може бути зображена як на індивідуальних, так і на групових знімках. Крім того, до кожного фото можуть бути додані супровідні відомості, зокрема інформація про людей, присутніх на знімку, про дату та місце зйомки, а також додаткові примітки чи описи, що допомагають краще зрозуміти контекст. Користувач може виконати пошук за фотографією, і система автоматично знайде всі можливі збіги. Якщо на зображенні є кілька людей, користувач отримує можливість знайти інформацію про інших осіб із групового фото та, якщо це можливо, встановити з ними зв'язок. Це відкриває додаткові можливості для дослідження: навіть якщо про саму особу інформації недостатньо, можна отримати відомості про людей,

які були з нею на фото, і спробувати через них відновити втрачені родинні зв'язки або дізнатися важливі біографічні деталі.

Наступним важливим елементом є підтримка ведення родовідного дерева. Система має надавати зручний редактор, який дає змогу додавати родичів, установлювати зв'язки між особами, такі як «батьки–дитина» або «подружжя», а також доповнювати профілі додатковими відомостями. Користувач може зберігати текстові нотатки, прикріплювати документи, фотографії, листи та інші матеріали, що мають значення для генеалогічного дослідження. Важливою функцією є можливість прив'язки архівних записів до відповідної особи в родоводі, що забезпечує систематизацію та підтвердження історичних фактів. Окремо варто зазначити, що завантажені зображення можуть містити портрети, тому система автоматично аналізує їх для визначення наявності облич і подальшого використання у візуальному пошуку.

Ще однією важливою функцією є підтримка імпорту й експорту даних у форматі GEDCOM, що є загальноприйнятим стандартом обміну генеалогічною інформацією (Gellatly, 2015). Це дозволяє користувачам переносити вже створені родоводи з інших платформ або, навпаки, експортувати свої дані для використання в інших системах. У межах запропонованої платформи передбачено розширену підтримку GEDCOM, яка дає змогу експортувати не тільки текстові відомості, а й пов'язані файли, що більшість наявних аналогів не підтримує.

Окремо слід розглянути можливості системи у сфері аналізу генетичних даних. Більшість сучасних ДНК-сервісів мають власні лабораторії, де відбувається повний цикл аналізу зразків, що значно збільшує вартість послуг. Натомість у запропонованій системі передбачається інший підхід – користувач самостійно проводить аналіз у будь-якій зручній для нього лабораторії та завантажує отримані первинні дані секвенування NGS до системи (Сирі дані секвенування, б.д.). Це дає змогу значно знизити витрати, оскільки «сирі» дані коштують дешевше, ніж повний аналіз у комерційних ДНК-центрах.

Після завантаження система обробляє ці дані та використовує їх для пошуку збігів серед інших користувачів, що дає змогу встановити можливі родинні зв'язки. Окрім цього, за допомогою аналізу генетичних маркерів система може визначати етнічне й територіальне походження користувача, що є важливим для тих, хто прагне дізнатися більше про своє коріння. Порівнюючи отримані результати з наявними генетичними базами даних, система може вказати на зв'язки з певними регіонами й етнічними групами, допомагаючи користувачам краще зрозуміти історію своєї родини (Wang, 2016).

Для підвищення точності аналізу система також використовує еталонні зразки геномних даних, отримані з відкритих дослідницьких проєктів, таких як «1000 Genomes Project». Ці зразки охоплюють широкий спектр популяційних груп і служать науково підтвердженою базою для порівняння. Завдяки цьому можна більш точно визначати генетичні зв'язки, особливості походження та регіональну приналежність користувачів (1000 Genomes Project, n.d.).

Останнім, але не менш важливим елементом системи є розширений пошук, який працює як з внутрішніми, так і з зовнішніми базами даних. У межах внутрішньої пошукової системи застосовується технологія Elasticsearch, що забезпечує

повнотекстовий пошук з підтримкою варіантів написання та можливістю виправлення орфографічних помилок. Це дає змогу користувачам знаходити необхідну інформацію навіть у разі неточного введення запиту (Akdoğan, 2015). Крім цього, реалізується можливість пошуку за зображеннями, що використовує алгоритми розпізнавання облич для знаходження схожих фотографій у базі.

Для взаємодії із зовнішніми ресурсами передбачено два механізми: запити через API, якщо це підтримується зовнішніми системами, або автоматизований збір інформації зі сторінок відкритих джерел (парсинг даних). Результати пошуку поєднуються в єдиному інтерфейсі, що дає змогу користувачеві отримати комплексний набір даних з різних джерел.

Отже, запропонована система об'єднує кілька підходів до генеалогічного дослідження, забезпечуючи взаємодію з архівами, розширений текстовий та візуальний пошук, ведення родоводів та роботу з генетичними даними. Її ключова перевага – поєднання різних методів у межах єдиної платформи, що дає змогу здійснювати повноцінний аналіз родоводу, значно спрощуючи пошук і систематизацію інформації.

Окремої уваги потребує безпека даних, оскільки навіть обмежений доступ до державних архівів може створювати певні ризики для користувачів. Тому система має забезпечувати надійний захист інформації та дотримання принципів конфіденційності. Передусім це досягається через використання державних механізмів підтвердження особи, таких як Дія.Підпис. Цей метод автентифікації не тільки дозволяє однозначно ідентифікувати користувача, а й забезпечує юридично значуще підписання запитів (Інтеграція Дії, б.д.). Отримання офіційних даних з державних реєстрів, зокрема баз РАЦС, відбувається через інтеграційну систему «Трембіта» (б.д.), яка надає захищений та стандартизований API-доступ до таких реєстрів. Завдяки цьому система може автоматично виконувати рекурсивний пошук за підтвердженими родинними зв'язками (наприклад, від свідоцтва про народження до даних про батьків), дотримуючись чинного законодавства, яке обмежує доступ лише для найближчих родичів. Такий підхід значно полегшує процес збору родової інформації та гарантує правомірність обробки персональних даних.

Окрім захисту архівних даних, важливо також забезпечити приватність особистої інформації користувачів. Це реалізується за допомогою гнучких налаштувань доступу: користувач може обмежити перегляд своїх даних, тобто відкрити лише для авторизованих осіб або ж повністю приховати інформацію. Додатково передбачені параметри приватності для живих родичів у родовідному дереві: наприклад, можна відображати лише ПІБ та рік народження, приховуючи інші деталі, такі як місце проживання, контактні дані тощо. Це дає змогу зберігати баланс між відкритістю досліджень і захистом персональних відомостей.

Не менш важливим аспектом є постійне поповнення архівної бази даних. Основний акцент роблять на роботі з державними архівами, де зберігається широкий спектр історичних документів, проте це лише один з напрямів. Значний обсяг цінної інформації можна знайти у старих газетах, журналах, церковних книгах, списках населення та інших джерелах, які зберігаються в різних відомчих архівах.

Важливим складником процесу є волонтерська діяльність, що дозволяє значно розширити інформаційну базу завдяки ініціативі самих користувачів.

Волонтери можуть долучатися до процесу оцифрування документів, допомагаючи у введенні текстової інформації до бази, однак їхня роль не обмежується лише цим. Вони можуть займатися збиранням локальних даних, які мають історичну цінність. Наприклад, мешканці певного міста чи села можуть фотографувати пам'ятні таблички, меморіальні дошки або інші об'єкти, що містять імена та дати, важливі для генеалогічних досліджень. Такий підхід дає змогу відновлювати дані про людей, інформацію про яких могли втратити внаслідок знищення архівів або відсутності офіційних записів.

Окрім цього, до процесу збору даних можуть бути залучені продавці антикваріату, оскільки до них часто потрапляють старі фотографії, листи, документи чи навіть картини з підписами. Якщо в момент продажу антикварний предмет буде ідентифіковано (наприклад, до фото додадуть хоча б ПІБ людини, зображеної на ньому), це може допомогти доповнити чиєсь родове дерево. За умови, що продавець внесе відповідні дані до системи, така інформація стане доступною для пошуку в межах системи. Аналогічний підхід можна застосувати до приватних фотоархівів, що зберігаються в сімейних колекціях, музеях чи локальних історичних спільнотах.

За своєю архітектурою система побудована на мікросервісному підході, що дає змогу розподілити функціонал на окремі незалежні компоненти. Це забезпечує низьку зв'язаність між сервісами, зменшує вплив однієї частини системи на інші та дає змогу гнучко масштабувати окремі модулі залежно від навантаження (Newman, 2021).

Основу системи становлять вісім ключових сервісів (рис. 1), кожен з яких виконує свою специфічну функцію:

1. Сервіс користувачів забезпечує зберігання, управління та обробку даних про користувачів системи. Охоплює функціонал реєстрації, автентифікації, авторизації, управління профілями, налаштування доступу та адміністрування користувачів. Також підтримує інтеграцію з таким механізмом, як Дія.Підпис, для підтвердження особи.

2. Сервіс нотифікацій надсилає користувачам важливі повідомлення, зокрема підтвердження реєстрації, сповіщення про знайдені збіги в родоводі, завершення ДНК-аналізу тощо.

3. Сервіс файлового сховища відповідає за збереження цифрових файлів, що завантажують користувачі системи, охоплюючи архівні документи, фотографії, листи, нотатки та «сирі» дані ДНК-аналізу. Підтримує керування файлами, контроль доступу до них та їхню безпечну обробку.

4. Сервіс архіву забезпечує зберігання та обробку оцифрованих архівних документів, а також управління історичними записами. Дає змогу зберігати посилання на документ-першоджерело (відповідний файл у файловому сховищі), що підтверджує достовірність збережених даних.

5. Сервіс керування родовідними деревами відповідає за побудову, редагування та аналіз родовідних дерев користувачів. Дає змогу встановлювати родин-

ні зв'язки, прикріплювати документи, фото й архівні записи до конкретних осіб. Підтримує імпорт та експорт GEDCOM-файлів.

6. Сервіс аналізу ДНК працює з файлами сирих даних секвенування NGS, виконує їх попередню обробку, аналізує територіальне й етнічне походження користувача на основі порівняння з еталонними базами даних (наприклад, 1000 Genomes Project), а також забезпечує пошук потенційних родичів у межах системи.

7. Сервіс аналізу зображень здійснює автоматичне розпізнавання облич на фотографіях, розпізнавання рукописного та друкованого тексту на зображеннях, пошук схожих осіб серед архівних фото.

8. Сервіс пошуку відповідає за швидкий та ефективний пошук інформації у внутрішніх базах даних системи, використовуючи Elasticsearch для гнучкого повнотекстового пошуку з урахуванням орфографічних варіантів, морфологічного розбору та наближених відповідей. Також забезпечує інтеграцію із зовнішніми ресурсами, такими як державні реєстри, приватні або відкриті архіви, що дає змогу значно розширити можливості пошуку та забезпечити доступ до додаткових джерел інформації. Взаємодія із зовнішніми системами здійснюється через API або механізми автоматизованого збору даних, що допомагає користувачам знаходити відповідні записи навіть за межами внутрішньої бази системи.

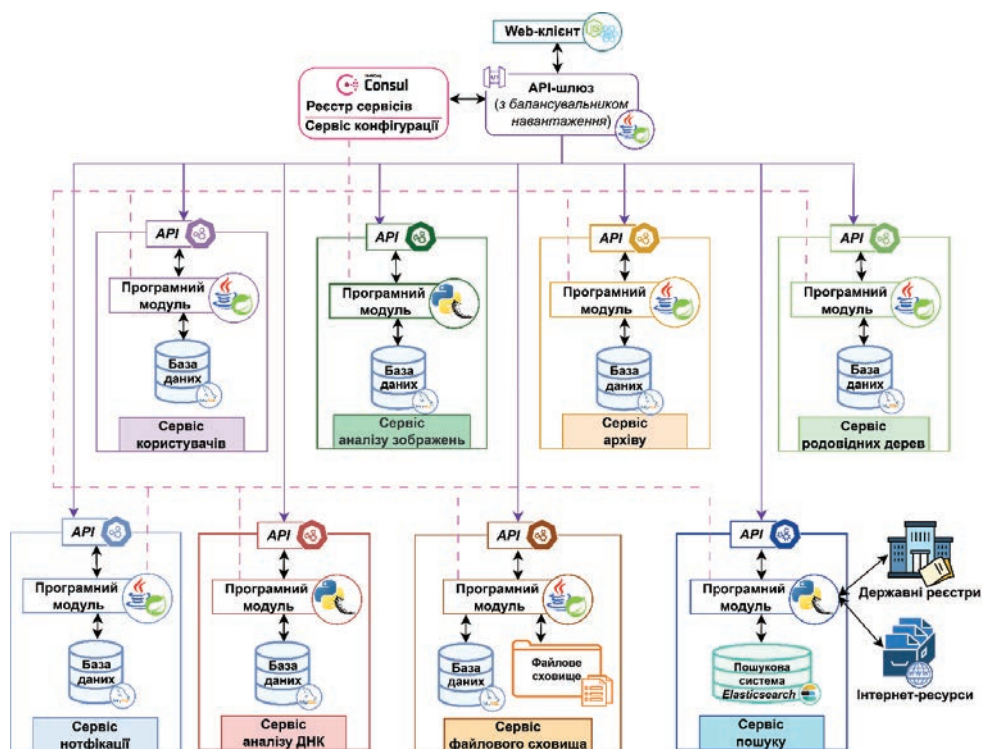


Рис. 1. Мікросервісна архітектура інформаційно-пошукової системи родоvodu

Використання мікросервісного підходу надає низку важливих переваг:

1. Гнучке масштабування. Кожен сервіс може бути розгорнутий окремо, що дає змогу збільшувати потужність лише тих компонентів, на яких підвищене навантаження (Newman, 2021). Наприклад, у разі активного використання пошуку або аналізу зображень можна масштабувати лише відповідні сервіси без впливу на інші частини системи.

2. Стійкість до збоїв. Незалежність мікросервісів забезпечує роботу системи навіть у разі відмови одного з компонентів (Newman, 2021). Наприклад, якщо сервіс ДНК-аналізу тимчасово недоступний, інші функції, такі як пошук чи управління родоходом, продовжать працювати без змін.

3. Розподіл навантаження. Окремі сервіси можуть бути розміщені на різних серверах або в хмарних середовищах, що покращує продуктивність і швидкодію системи (Newman, 2021).

4. Глобальне розширення. У разі популяризації системи за межами України мікросервісна архітектура дає змогу створювати локальні кластери в різних регіонах, зберігаючи централізовану координацію між усіма вузлами. Це означає, що під час зростання бази користувачів можна поступово додавати нові регіональні сервери, збільшуючи охоплення без критичних навантажень на основну інфраструктуру (Grebennik et al., 2019).

5. Можливість відкриття або інтеграції компонентів у зовнішні системи. У разі необхідності окремі сервіси можуть бути відкриті для використання з боку сторонніх розробників або інтегровані в інші платформи через API. Це дає змогу розширювати екосистему системи, підтримувати взаємодію з державними реєстрами, комерційними сервісами або науковими базами даних, що покращує функціональні можливості та розширює аудиторію користувачів.

Результати дослідження. У межах реалізації запропонованої платформи особливу увагу приділено сервісу аналізу зображень, який вирізняє систему серед аналогічних рішень. Цей сервіс реалізує три незалежні, але комплементарні напрями: розпізнавання облич, розпізнавання тексту на зображеннях, а також пошук за обличчями серед наявних записів у базі.

Розробку сервісу виконано з використанням мови Python, яка забезпечує високу ефективність у процесі роботи з візуальними й текстовими даними (Antao, 2015). Для розпізнавання облич застосовано бібліотеку `face_recognition`, яка надає функціонал для виявлення та аналізу облич на зображеннях. У процесі обробки автоматично виділяються маркери облич, зокрема очі, ніс, контур, та генерується їх цифрове представлення (Elgendy, 2020). На рис. 2 продемонстровано приклад успішного розпізнавання: система ідентифікує дві особи на користувацькому зображенні. Отримані вектори ознак облич зберігаються у спеціалізованій базі даних, що функціонує незалежно від інших модулів системи. Завдяки використанню UUID забезпечується унікальність кожного запису, а також сумісність і масштабованість у межах кластерної архітектури.

Другим напрямом є розпізнавання тексту на зображеннях, що є особливо актуальним для обробки історичних джерел, таких як рукописні документи, фото листів, пам'ятні таблички з портретами та супровідними відомостями.

ми. Для реалізації цього модуля використано модель TrOCRProcessor разом із VisionEncoderDecoderModel, які належать до екосистеми Hugging Face та дають змогу здійснювати трансформерне розпізнавання тексту на зображеннях.



Рис. 2. Розпізнавання облич на зображенні

У процесі розробки виявлено, що готові моделі не забезпечують надійної роботи з рукописним та друкованим текстом одночасно, з підтримкою як кирилиці (української мови), так і латиниці (англійської). Тому ухвалено рішення доопрацювати наявну модель на спеціально підібраних даних. Використано відкриті датасети з рукописним і друкованим текстом (зокрема, Cyrillic-MNIST, Cyrillic Handwriting Dataset), а також згенеровані синтетичні зображення з використанням історичних українських шрифтів і стилізованого письма. Це дало змогу створити гібридну модель, яка досягає точності розпізнавання на рівні 72 %. Цей показник є задовільним для завдань автоматичної обробки зображень і може бути підвищений у майбутньому.

Приклад роботи модуля з виявлення тексту на зображенні подано на рис. 3. Тут продемонстровано локалізацію текстових блоків на архівному матеріалі. Система змогла виявити майже всі текстові фрагменти на зображенні, проте деякі розпізнані літери потребують уточнення через неточності, пов'язані з якістю зображення та налаштуваннями моделі розпізнавання.

У поєднанні обидва типи аналізу – розпізнавання облич і тексту – формують автономний механізм обробки вхідних зображень. Користувачеві достатньо лише завантажити файл, після чого система автоматично виконує повний цикл обробки. Результати зберігаються у вигляді структури з прикріпленим текстовим вмістом та візуалізованими обличчями. На рис. 4 наведено приклад відображення результатів обробки доданого файлу. З урахуванням того, що розпізнавання

не завжди є ідеальним, користувач має змогу в режимі редагування скоригувати текст у разі появи артефактів або неточності розпізнавання.

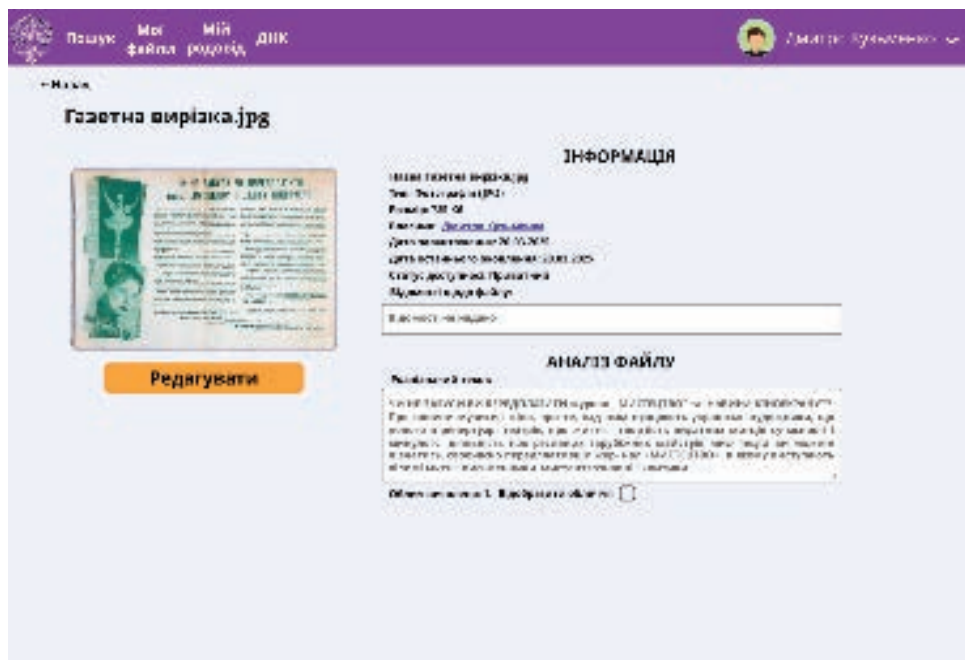


Рис. 3. Результати пошуку за обличчям

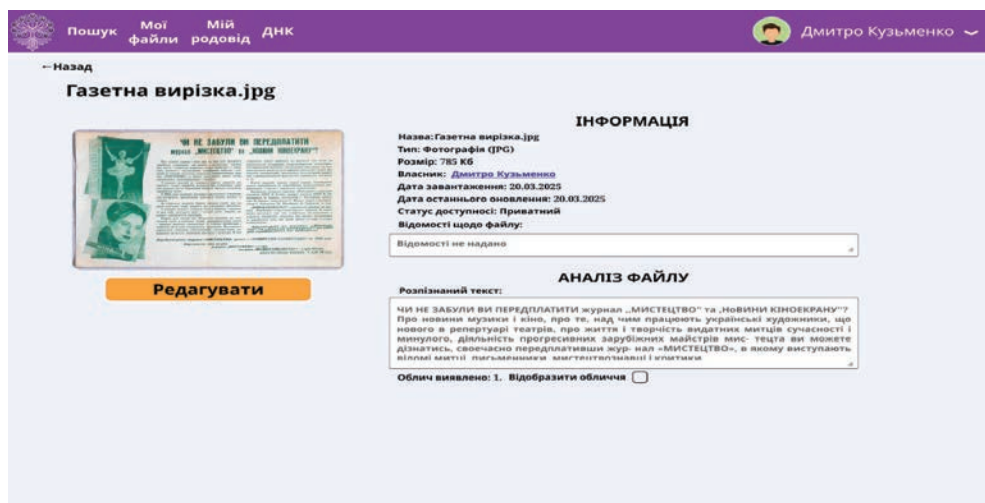


Рис. 4. Збережені відомості доданого файлу

Окремою функцією є можливість пошуку за розпізнаним обличчям. Для цього достатньо обрати зображення, після чого система здійснює пошук серед усіх збере-

жених записів. На рис. 5 представлено приклад результату такого пошуку. У розглянутому прикладі з газетною вирізкою відсутні текстові згадки про зображену особу, однак пошук за обличчям дав змогу знайти інше джерело з ідентичним обличчям та супровідними біографічними відомостями. Для цього знайдено один результат із зазначенням, що на фото зображено особу, схожу на Еліну Бистрицьку (рис. 6).



Рис. 5. Результати пошуку за обличчям

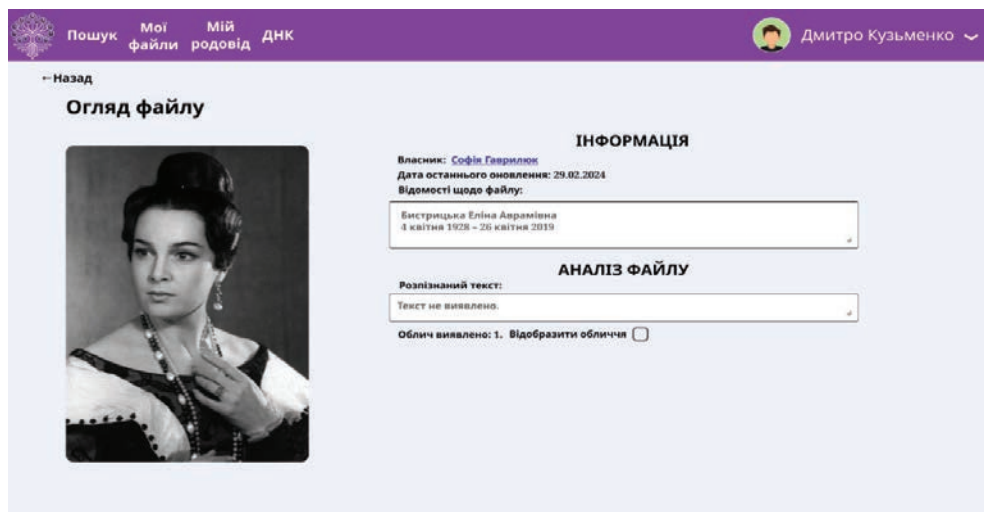


Рис. 6. Деталі знайденого файлу

Варто зауважити, що ідентифікація за обличчям не гарантує 100 % точності й не підтверджує однозначно особу чи родинний зв'язок. Це лише засіб визначення потенційної схожості, який дає змогу вийти на інші джерела інформації. Особа на зображенні може бути знайомою або родичем інших осіб, що мають профілі в системі. У такому разі в користувача є шанс установити додаткові зв'язки, зокрема через інші родини, які можуть володіти інформацією про спільного предка чи події родинної історії.

Висновки. У межах дослідження проведено аналіз ключових труднощів, з якими стикаються дослідники родоводу: фрагментарність джерел, відсутність централізованих систем, складність у роботі з рукописними матеріалами та недостатня інтеграція із сучасними цифровими інструментами. Розглянуто наявні вітчизняні та міжнародні рішення, виокремлено їхні переваги та недоліки, що дало змогу сформулювати перелік функціональних вимог до майбутньої системи.

Окрему увагу приділено розробці та демонстрації модуля обробки зображень, здатного виконувати розпізнавання облич і тексту з історичних фото й документів. Цей модуль може функціонувати як незалежний компонент або інтегруватися з уже наявними платформами. Використання технологій штучного інтелекту дало змогу значно розширити функціональні можливості системи: автоматизувати рутинні етапи, оптимізувати аналіз вхідних даних і забезпечити точніше зіставлення інформації, а також надати можливість візуального пошуку.

Перспективи подальшого розвитку платформи охоплюють розширення джерел даних, інтеграцію з міжнародними архівами, масштабування користувацької бази, а також популяризацію рішення в інших країнах. Важливу роль може відіграти залучення громадських і волонтерських ініціатив для наповнення бази даних та верифікації інформації. Запропонована система не тільки вирішує технічні й методологічні завдання, а й створює фундамент для збереження національної культурної спадщини, сприяє відновленню родинних зв'язків і передачі історичної пам'яті наступним поколінням.

СПИСОК ПОСИЛАНЬ

- Аналізи ДНК в Україні, б.д. *Mama Papa Medical genetic center*. [online] Доступно: <<https://www.testdnk.com.ua/eng/>> [Дата звернення 29 березня 2025].
- База жителів України 1650-1920 рр. народження, б.д. *PRA.IN.UA*. [online] Доступно: <<https://pra.in.ua/uk>> [Дата звернення 29 березня 2025].
- Інтеграція Дії. Зручні сервіси цифрового документообігу для бізнесу та держустанов, б.д. *Дія*. [online] Доступно: <<https://integration.dii.gov.ua/>> [Дата звернення 29 березня 2025].
- Котович, С., 2024. Анатолій Хромов: про оцифрування, збереження й відкритість архівів. *Ukrainer*, [online] 04 травня. Доступно: <<https://www.ukrainer.net/anatoliy-khromov/>> [Дата звернення 03 квітня 2025].
- Рідні – товариство українських родовідників*, б.д. [online] Доступно: <<https://ridni.org/>> [Дата звернення 29 березня 2025].

- Сирі дані секвенування, б.д. *Mama Papa Medical genetic center*. [online] Доступно: <<https://www.testdnk.com.ua/analizy-dnk/raw-sequencing-data>> [Дата звернення 29 березня 2025].
- Трембіта, б.д. *Система електронної взаємодії державних електронних інформаційних ресурсів*. Міністерство цифрової трансформації України. [online] Доступно: <<https://trembita.gov.ua/>> [Дата звернення 29 березня 2025].
- Akdoğan, H., 2015. *Elasticsearch Indexing*. Birmingham: Packt Publishing.
- Ancestry | Family Tree, Genealogy & Family History Records, n.d. *Ancestry*. [online] Available at: <<https://www.ancestry.com/>> [Accessed 29 March 2025].
- Antao, T., 2015. *Bioinformatics with Python Cookbook: Learn how to use modern Python bioinformatics libraries and applications to do cutting-edge research in computational biology*. Birmingham: Packt Publishing.
- Elgendy, M., 2020. *Deep Learning for Vision Systems*. Shelter Island: Manning.
- Family Tree, 2024. *Генеалогія в Україні – Навіщо сканувати архівні документи*. [online] 11 листопада. Доступно: <<https://www.history-family.com/genealogy-scans/>> [Дата звернення: 29 березня 2025].
- FamilySearch, n.d. Available at: <<https://www.familysearch.org/en/>> [Accessed 29 March 2025].
- Free Family Tree, Genealogy, Family History, and DNA Testing, n.d. Available at: <<https://www.myheritage.com>> [Accessed 29 March 2025].
- Gellatly, C., 2015. Reconstructing Historical Populations from Genealogical Data Files. In: G. Bloothoof, P. Christen, K. Mandemakers and Schraagen M., eds. *Population Reconstruction*. [e-book] Cham: Springer, pp.111-128. https://doi.org/10.1007/978-3-319-19884-2_6
- Grebennik, I., Khriapkin, O., Ovezgeldyyev, A., Pisklakova, V. and Urniaeva, I., 2019. The Concept of a Regional Information-Analytical System for Emergency Situations. In: Y. Murayama, D. Velez and P. Zlateva, eds. *Information Technology in Disaster Risk Reduction*. Second IFIP TC 5 DCITDRR International Conference, ITDRR 2017, Sofia, Bulgaria, October 25-27, 2017. [e-book] Cham: Springer, Vol.516, pp.55-66. https://doi.org/10.1007/978-3-030-18293-9_6
- Комисаренко, А., 2017. Доступ до архівів як частина права на інформацію. Частина перша: Конспект-репортаж з конференції. *Вікімедіа Україна*, [online] 30 листопада. Доступно: <<https://blog.wikimedia.org.ua/2017/11/30/dostup-do-arkhiviv-yak-chastyna-prava-na-informatsiiu-chastyna-1/>> [Дата звернення 29 березня 2025].
- Newman, S., 2021. *Building Microservices: Designing Fine-Grained Systems*. 2nd Ed. O'Reilly.
- 1000 Genomes Project, n.d. *Coriell Institute for Medical Research*. Available at: <<https://www.coriell.org/1/NHGRI/Collections/1000-Genomes-Project-Collection/1000-Genomes-Project>> [Accessed 29 March 2025].
- Wang, X., 2016. *Next-Generation Sequencing Data Analysis*. [e-book] Boca Raton: CRC Press. <https://doi.org/10.1201/b19532>

REFERENCES

- Akdoğan, H., 2015. *Elasticsearch Indexing*. Birmingham: Packt Publishing.
- Analizy dnk v ukraini [DNA testing in Ukraine], n.d. *Mama Papa Medical genetic center*. [online] Available at: <<https://www.testdnk.com.ua/eng/>> [Accessed 29 March 2025].
- Ancestry | Family Tree, Genealogy & Family History Records, n.d. *Ancestry*. [online] Available at: <<https://www.ancestry.com/>> [Accessed 29 March 2025].

- Antao, T., 2015. *Bioinformatics with Python Cookbook: Learn how to use modern Python bioinformatics libraries and applications to do cutting-edge research in computational biology*. Birmingham: Packt Publishing.
- Baza zhyteliv ukrainy 1650-1920 rr. Narodzhennia [Database of Ukrainian residents 1650-1920 years of birth], n.d. *Pra.in.ua*. [online] Available at: <<https://pra.in.ua/uk>> [Accessed 29 March 2025].
- Elgendy, M., 2020. *Deep Learning for Vision Systems*. Shelter Island: Manning.
- Family Tree, 2024. *Henealohiia v Ukraini – Navishcho skanuvaty arkhivni dokumenty* [Genealogy in Ukraine – Why scan archival documents?]. [online] 11 November. Available at: <<https://www.history-family.com/genealogy-scans/>> [Accessed 29 March 2025].
- FamilySearch, n.d. Available at: <<https://www.familysearch.org/en/>> [Accessed 29 March 2025].
- Free Family Tree, Genealogy, Family History, and DNA Testing, n.d. Available at: <<https://www.myheritage.com>> [Accessed 29 March 2025].
- Gellatly, C., 2015. Reconstructing Historical Populations from Genealogical Data Files. In: G. Bloothoof, P. Christen, K. Mandemakers and Schraagen M., eds. *Population Reconstruction*. [e-book] Cham: Springer, pp.111-128. https://doi.org/10.1007/978-3-319-19884-2_6
- Grebennik, I., Khriapkin, O., Ovezgeldyyev, A., Pisklakova, V. and Urniaieva, I., 2019. The Concept of a Regional Information-Analytical System for Emergency Situations. In: Y. Murayama, D. Velez and P. Zlateva, eds. *Information Technology in Disaster Risk Reduction*. Second IFIP TC 5 DCITDRR International Conference, ITDRR 2017, Sofia, Bulgaria, October 25-27, 2017. [e-book] Cham: Springer, Vol.516, pp.55-66. https://doi.org/10.1007/978-3-030-18293-9_6
- Intehratsiia dii. Zruchni servisy tsyfrovoho dokumentoobihu dlia biznesu ta derzhustanov [Diia integration. Convenient services for using digital documents for business and government institutions], n.d. *Diia*. [online] Available at: <<https://integration.diia.gov.ua/>> [Accessed 29 March 2025].
- Komisarenko, A., 2017. Dostup do arkhiviv yak chastyna prava na informatsiiu. Chastyna persha: Konspekt-reportazh z konferentsii [Access to archives as part of the right to information. Part one: Summary report from the conference]. *Wikimedia Ukraine*, [online] 30 November. Available at: <<https://blog.wikimedia.org.ua/2017/11/30/dostup-do-arkhiviv-yak-chastyna-prava-na-informatsiiu-chastyna-1/>> [Accessed 29 March 2025].
- Kotovych, S., 2024. Anatolii khromov: pro otsyfruvannia, zberezhennia y vidkrytist arkhiviv [Anatoliy Khromov: on digitization, preservation and openness of archives]. *Ukrainer*, [online] 04 May. Available at: <<https://www.ukrainer.net/anatoliy-khromov/>> [Accessed 03 April 2025].
- Newman, S., 2021. *Building Microservices: Designing Fine-Grained Systems*. 2nd Ed. O'Reilly.
- Ridni – tovarystvo ukrainskykh rodovidnykiv [RIDNI is a genealogy company], n.d. [online] Available at: <<https://ridni.org/>> [Accessed 29 March 2025].
- Syri dani sekvenuvannia [Raw sequencing data], n.d. *Mama Papa Medical genetic center*. [online] Available at: <<https://www.testdnk.com.ua/analizy-dnk/raw-sequencing-data>> [Accessed 29 March 2025].
- 1000 Genomes Project, n.d. *Coriell Institute for Medical Research*. Available at: <<https://www.coriell.org/1/NHGRI/Collections/1000-Genomes-Project-Collection/1000-Genomes-Project>> [Accessed 29 March 2025].
- Trembita, n.d. *Systema elektronnoi vzaємodii derzhavnykh elektronnykh informatsiynykh resursiv* [Electronic interaction system Trembita]. Ministry of Digital Transformation of Ukraine. [online] Available at: <<https://trembita.gov.ua/>> [Accessed 29 March 2025].
- Wang, X., 2016. *Next-Generation Sequencing Data Analysis*. [e-book] Boca Raton: CRC Press. <https://doi.org/10.1201/b19532>

UDC 004.932:[929.5:930.25***Dmytro Kuzmenko,****Bachelor,**Kharkiv National University of Radio Electronics,**Kharkiv, Ukraine**dmytro.kuzmenko@nure.ua**<https://orcid.org/0009-0006-1366-9530>****Ihor Hrebennik,****Doctor of Science in Engineering, Professor,**Head of the System Engineering Department,**Kharkiv National University of Radio Electronics,**Kharkiv, Ukraine**igor.grebennik@nure.ua**<https://orcid.org/0000-0003-3716-9638>****Inna Urniaieva,****PhD, Associate Professor**at the System Engineering Department,**Kharkiv National University of Radio Electronics,**Kharkiv, Ukraine**inna.urniaieva@nure.ua**<https://orcid.org/0000-0001-9795-6954>*

INFORMATION AND SEARCH SYSTEM FOR GENEALOGICAL RESEARCH

The purpose of the article is to develop a prototype digital platform to support genealogical research, which combines archival sources, DNA analysis, file processing, including images, comprehensive search by text and visual data, as well as the construction of a family tree in a single information environment. The necessity of creating a national service focused on the needs of Ukrainian users, with support for searching both internal and external sources through integration with open archives and state registers, is justified.

Research methodology. The work uses an analytical method to study the limitations of existing solutions in the genealogical search field and a system analysis method to form requirements for the architecture of the future digital platform. A comparative evaluation of international and domestic services' functionality was carried out to identify the key components necessary for building a comprehensive information system.

Scientific novelty. The proposed system is distinguished by integrating all key components of genealogical research within a single digital platform that covers archival records, full-text search, visual search by faces, DNA analysis, and family tree construction. Unlike existing services, the proposed system integrates historical archives and current state registers. The system supports electronic user authentication through a tool such as Diya – signature and interaction with registers through the Trembita system. The use of artificial intelligence for automated file processing, including images, for text recognition and face identification is innovative. In addition, a simplified approach to genetic analysis has been implemented – the

system works with raw sequencing data, allowing users to reduce costs and avoid restrictions associated with specific laboratories.

Conclusions. The developed prototype of a digital platform demonstrates a practical approach to integrating archival, visual and genetic data, which significantly simplifies genealogical research. The system covers the entire cycle: from digitising and processing sources to building family trees and searching for relatives. The research is of scientific and technical importance. It significantly contributes to preserving Ukraine's historical memory and cultural heritage, which are essential for forming identity and transferring knowledge to future generations.

Keywords: genealogy; family tree; digital heritage; artificial intelligence; face search; text recognition; archival documents; DNA data; microservice architecture.

Надійшла 10.04.2025

Прийнята 29.05.2025

Стаття була вперше опублікована онлайн 18.07.2025



This is an open access journal, and all published articles are licensed under a Creative Commons Attribution 4.0.