

4. What is AI? A guide to artificial intelligence. *IBM*. URL: <https://www.ibm.com/cloud/learn/what-is-artificial-intelligence> (дата звернення: 05.05.2025).
5. 9 найкращих ШІ інструментів для CRM у 2025 році. *Chat GPT Academy*. URL: <https://www.chatgptacademy.online/sfery-vykorystannya-ai/obslugovuvannya-kliientiv/9-najkrashhyh-shi-instrumentiv-dlya-crm-u-2025-rocz/> (дата звернення: 03.05.2025).



УДК: 004.652.4:929:004.9

Артур Спектор, аспірант

Національний університет «Львівська Політехніка», м. Львів

ВИКОРИСТАННЯ NOSQL БАЗ ДАНИХ У КОНТЕКСТІ ЗБЕРІГАННЯ ТА ОБРОБКИ ГЕНЕАЛОГІЧНОЇ ІНФОРМАЦІЇ

Анотація. У роботі досліджено потенціал використання NoSQL баз даних у контексті збереження, організації та обробки архівних документів, що містять генеалогічну інформацію. Проаналізовано специфіку застосування документно-орієнтованих та інших нереляційних підходів до зберігання архівних даних, зокрема у випадках, коли інформація є структурно неоднорідною. Розглянуто ключові переваги NoSQL баз даних у порівнянні SQL базами, зокрема в аспектах масштабованості, гнучкості моделювання структури документів, можливостей зберігання слабо структурованих даних, а також ефективної реалізації пошукових запитів. Особливу увагу приділено питанням моделювання генеалогічних зв'язків та метаданих у цифровому архівному середовищі. Отримані результати засвідчують доцільність використання NoSQL технологій для побудови адаптивних інформаційних систем, орієнтованих на потреби генеалогічних досліджень.

Ключові слова: бази даних, NoSQL, архівні фонди, генеалогічні документи.

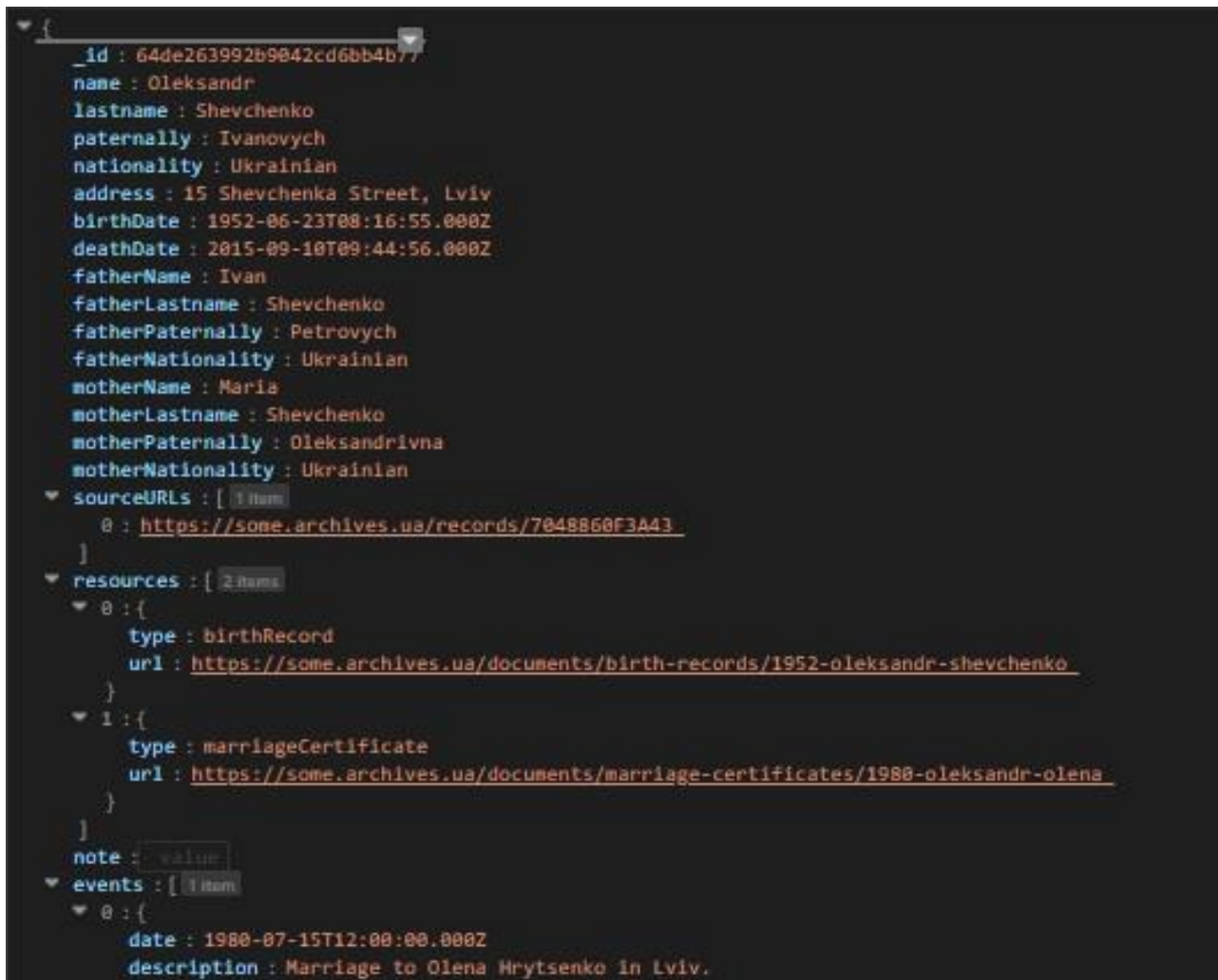
Інтенсивні процеси цифровізації архівних фондів зумовлюють потребу у впровадженні технологій, здатних зберігати та обробляти неструктуровану інформацію, що, у свою чергу, забезпечує можливість централізованого ведення генеалогічних досліджень на основі даних із різноманітних архівних джерел.

У роботі М. Діас-Ордоньеса, Д. С. Родрігеса Баєни та Юн-Касалільї обґрунтовується використання документно-орієнтованих баз даних як гнучкого та потужного інструменту для зберігання та аналізу історичних даних, особливо у процесі роботи з напівструктурованими або неструктурованими даними. На прикладі проекту AtlantoCracies, який досліджує європейські аристократичні мережі, демонструється, як NoSQL бази даних дозволяють ефективно моделювати складні родинні зв'язки та адаптуватися до змін у дослідницьких потребах [1].

NoSQL – це підхід до зберігання даних, що виходить за межі традиційних табличних структур реляційних баз даних. На відміну від останніх, NoSQL використовує гнучкі схеми, зокрема документно-орієнтовані моделі (наприклад, JSON,

BSON тощо), що спрощує масштабування та обробку великих обсягів неструктурованих даних, таких як генеалогічні дані, де структура (поля) документів можуть розрізнятися залежно від типу документу, стандартів часу створення написання документа. NoSQL бази є розподіленими системами, у яких дані реплікуються на кількох серверах, що підвищує надійність і доступність. Це актуально для обробки і взаємодії з великими обсягами архівної інформації, яка записується в базу під час оцифрування документів [2].

У процесі оцифрування генеалогічних джерел формується масив різномірної інформації: цифрові зображення, скановані копії документів, результати оптичного розпізнавання тексту (OCR) та супровідні метадані. NoSQL забезпечують збереження, як основних файлів, так і пов'язаних із ними структурованих та напівструктурованих даних, таких як, координати, часові мітки, іменовані сутності (осіб, місць, подій тощо). Подібна модель зберігання дає можливість організувати механізми повнотекстового пошуку, автоматизованої класифікації, контекстного аналізу вмісту та формування семантичних зв'язків між даними [3]. Приклад JSON запису у NoSQL базі з інформацією про документ та посилання на суміжні документи подано далі (див. рис. 1).



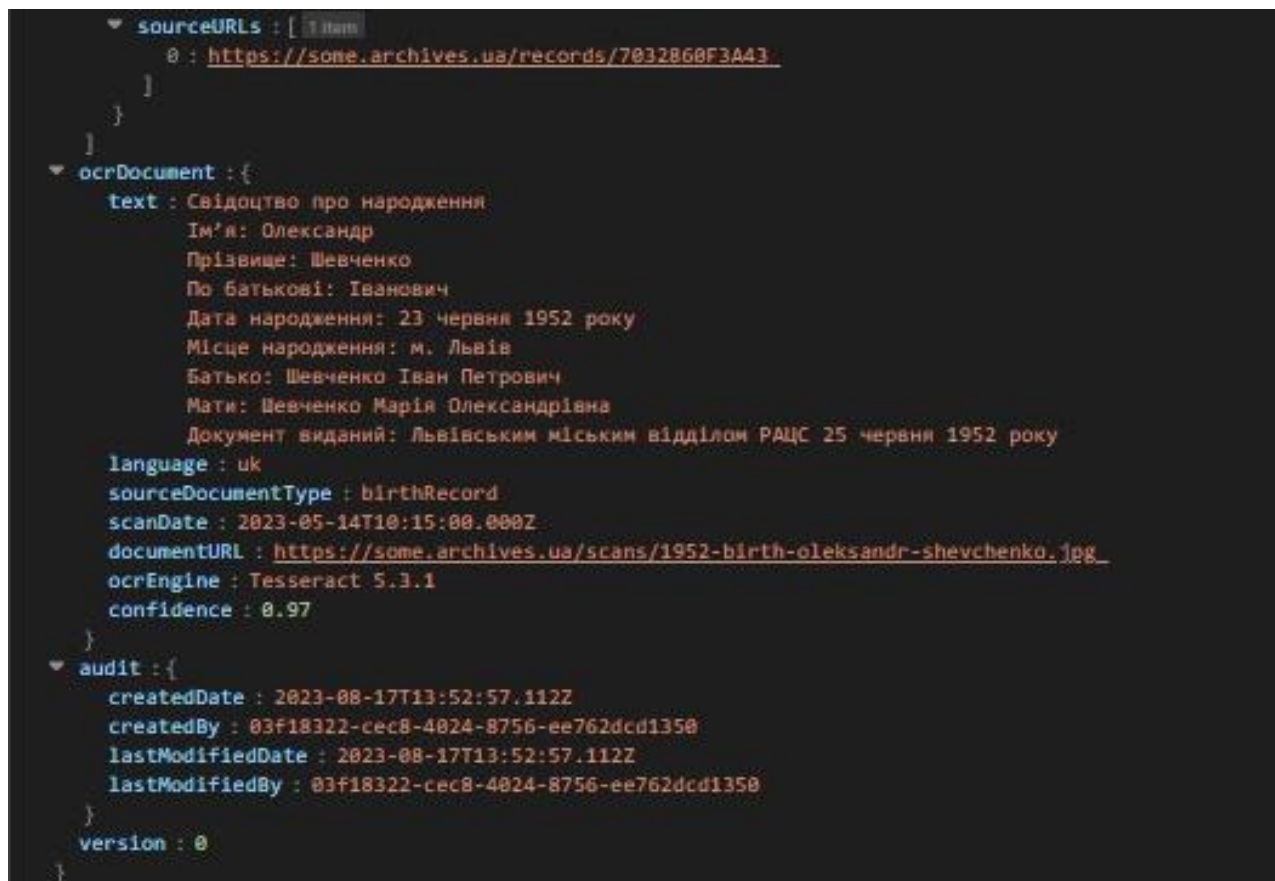


Рисунок 1 – Приклад структури генеалогічних даних в NoSQL базі

Генеалогічна інформація часто має вигляд дерева або графа (взаємозв'язки між особами, родами, поколіннями). У NoSQL, зокрема у графових базах (наприклад, Neo4j), такі структури моделюються напям, без необхідності створення складних зв'язків через зовнішні ключі, як у реляційних СУБД. Це спрощує запити типу «знайти всіх нащадків особи X до 4-го покоління» або «визначити родинні зв'язки між двома особами» [3]. Основні відмінності в контексті зберігання генеалогічних документів між SQL і NoSQL баз даних представлено в табл. 1 [4].

Таблиця 1 – Основні відмінності SQL і NoSQL баз з огляду на потреби генеалогічного моделювання

Критерій	SQL	NoSQL
Модель зберігання	Таблиці з фіксованими рядками і стовпцями. Зручно використовувати для фіксованих сутностей (особа, подія, локація)	Документи (JSON, BSON), ключ-значення, графи, колонки. Дає можливість гнучко моделювати джерела, події, сімейні зв'язки у вкладеній формі
Тип даних	Структуровані дані. Труднощі з обробкою архівних текстів та сканів	Неструктуровані, напівструктуровані та структуровані дані. Дає можливість одночасно зберігати мультимедіа, тексти OCR тощо
Схема	Жорстка, суворо визначена схема. Складно адаптувати структуру під	Гнучка, динамічна схема, можливість зберігання різнотипних структур.

Критерій	SQL	NoSQL
	нові типи подій чи нестандартні зв'язки	Можливе поступове розширення родоводу без реконфігурації схеми
Масштабування	Вертикальне, розширення ресурсів одного сервера (CPU, RAM, диски)	Горизонтальне, додавання вузлів у кластер, розподілене середовище
Керування транзакціями	Централізоване, ефективне за умови єдиного вузла	Ускладнене у розподілених системах, потребує додаткових механізмів синхронізації
Гнучкість схеми	Складно оновлювати схему в реальному часі	Підтримка додавання нових атрибутів без перезапуску або міграцій

З огляду на основні відмінності між SQL і NoSQL баз можна зазначити, що NoSQL мають суттєві переваги в контексті зберігання, обробки та представлення генеалогічних даних, особливо коли йдеться про архівні документи зі складною внутрішньою структурою та мультимедійним наповненням.

Висновки. Застосування NoSQL баз даних у сфері зберігання та обробки архівних документів генеалогічного спрямування відкриває широкі можливості для комплексного представлення інформації, пов'язаної з окремим джерелом. Зокрема, це дає змогу інтегрувати текстові описи, зображення, гіперпосилання, метадані та відомості про доступ і зміни (аудит) в єдину інформаційну модель. У контексті цифрової трансформації архівної справи та розбудови інтерактивних платформ для дослідження родоводу, NoSQL технології забезпечують як гнучкість у моделюванні складних і неоднорідних структур даних, так і ефективну масштабованість інформаційних систем.

Список використаних джерел

1. Díaz-Ordóñez M., Rodríguez Baena D. S., Yun-Casalilla B. A new approach for the construction of historical databases—NoSQL Document-oriented databases: the example of AtlantoCracies. *Digital Scholarship in the Humanities*. 2023. Vol. 38, № 1. DOI: 10.1093/lle/fqad033.
2. NoSQL database. *IBM*. URL: <https://www.ibm.com/think/topics/nosql-databases> (дата звернення: 29.05.2025).
3. What is NoSQL. *AMAZON*. URL: <https://aws.amazon.com/nosql/> (дата звернення: 29.05.2025).
4. Understanding SQL vs NoSQL Databases. *MongoDB*. URL: <https://www.mongodb.com/resources/basics/databases/nosql-explained/nosql-vs-sql> (дата звернення: 29.05.2025).

