



[https://doi.org/10.36273/2076-9555.2026.1\(354\).41-46](https://doi.org/10.36273/2076-9555.2026.1(354).41-46)
УДК 004.65NoSQL:[930.25:929.5]:(477)(045)

Артур Спектор,
аспірант кафедри соціальних комунікацій
та інформаційної діяльності НУ "Львівська політехніка"
artur.y.spektor@lpnu.ua
<https://orcid.org/0000-0003-4176-9177>

Використання NoSQL-баз даних для зберігання, опрацювання та аналізу архівних документів генеалогічного спрямування

У статті обґрунтовано потребу впровадження архітектурної моделі документно-орієнтованих NoSQL-баз даних у сфері зберігання, опрацювання й аналізу генеалогічних документів. Актуальність дослідження зумовлено інтенсивними процесами цифровізації архівних фондів в Україні та потребою технологічних рішень, що дають змогу ефективно працювати з великими масивами неструктурованої й неоднорідної інформації, як-от цифрові зображення, результати оптичного розпізнавання тексту (OCR) і супровідні метадані, структура яких варіюється залежно від наповнення архівного фонду. Метою дослідження є обґрунтування використання NoSQL-моделей як оптимального рішення для управління генеалогічними даними в архівній справі, що дасть змогу забезпечити централізоване ведення генеалогічних досліджень на основі різноманітних архівних джерел. Констатовано, що NoSQL-модель підтримує гнучку схему зберігання, котра допомагає інтегрувати всі дані про особу в єдиний документ. Ця функція має розв'язати проблему різноманітності та багатомовності даних для українських архівних фондів. Документно-орієнтований підхід забезпечує горизонтальну масштабованість для управління зростаючими обсягами Big Data та надає можливість використовувати гнучкі індекси, що сприяє ефективному контекстному пошуку й виявленню прихованих семантичних зв'язків між даними, розподіленими в різних фондах. За результатами аналізу прикладів використання AtlantoCracies зроблено висновок, що впровадження NoSQL в архівній справі є основою для створення стійких, високопродуктивних систем, здатних пришвидшити темпи оцифрування та підвищити доступність національної культурної спадщини для проведення досліджень.

Ключові слова: бази даних; архівні фонди; архіви; NoSQL; документно-орієнтована модель; генеалогічні дослідження; цифровізація; Big Data

Постановка проблеми. Інтенсивні процеси цифровізації архівних фондів зумовлюють потребу впровадження сучасних технологій збереження, опрацювання й аналізу неструктурованої інформації, як-от цифрові зображення, скановані копії документів, результати оптичного розпізнавання тексту (OCR) і супровідні метадані, що можуть відрізнитися за структурою документа. Ці процеси, зі свого боку, потребують впровадження баз даних (БД), здатних працювати з неструктурованими даними та забезпечити можливість централізованого ведення генеалогічних досліджень на основі різноманітних архівних джерел.

Аналіз досліджень і публікацій. У колективній праці М. Діас-Ордоньеса та ін. обґрунтовано використання NoSQL-баз даних як гнучкого та потужного інструменту для зберігання й аналізу історичних джерел, особливо під час роботи з напівструктурованими чи неструктурованими даними. На прикладі проєкту AtlantoCracies, що має на меті дослідження європейських аристократичних мереж, було продемонстровано, як NoSQL-бази даних допомагають ефективно моделювати складні родинні зв'язки й адаптуватися до змін у дослідницьких потребах. Завдяки гнучкості ці бази даних ефективніше пристосовуються до часто непередбачуваного характеру історичних джерел, не зменшуючи водночас простоти

їх використання чи аналітичного потенціалу. Перевагами NoSQL-систем є те, що у процесі проєктування не потрібно визначати жорстку схему на початку дослідження та можна вносити зміни в міру зростання знань про об'єкт дослідження й джерела [1].

У дослідженні Ж.-П. Дедьє з'ясовано сутність баз даних для історичних і соціальних дисциплін, обґрунтовано потребу розділення етапу введення даних, що базується на мінімальній сегментації інформації на серію дій. Науковець розглядає бази даних не лише як сховища інформації, а і як критичні інструменти для трансформації масивів різноманітних історичних джерел у придатні для кількісного аналізу дані [2].

О. Артеменкова проаналізувала теоретичні та методологічні напрацювання провідних українських фахівців архівної справи щодо інформаційного інструментарію підвищення ефективності генеалогічних досліджень. Учена зауважує, що слід активізувати процеси впровадження інформаційних технологій на рівні держави для сталого розвитку архівної справи та проведення генеалогічних досліджень в архівах України [4].

І. Тюрменко, Л. Божук, І. Струк та Ю. Сєров, аналізуючи особливості оцифрування документів у регіональних державних архівах, що набули значної інтенсивності у 2016 р. і досягли піку в 2019—2020 рр., дійшли висновку, що попри наявність електронних

ресурсів, процес залишається несистемним. Основною причиною є брак єдиної державної політики, що призводить до хаотичного відбору документів і проблем у структурі архівних вебсайтів. Автори наголосили на критичній потребі розширення цифрових колекцій для підвищення доступності культурної спадщини нашої держави [5].

Мета статті полягає в обґрунтуванні використання моделей NoSQL в архівній справі; визначенні їхніх переваг перед традиційними БД, а саме: забезпечення гнучкої моделі зв'язків між документами, розширені можливості збереження та опрацювання даних, доступність до контекстної інформації документів.

Упровадження NoSQL в архівних установах дасть змогу пришвидшити темпи оцифрування та забезпечити доступність інформації для проведення генеалогічних досліджень.

Виклад основного матеріалу дослідження. NoSQL (Not only SQL) — це концепція організації зберігання даних, що відрізняється від класичних табличних структур реляційних БД. На відміну від останніх, NoSQL застосовує гнучкі схеми зберігання, зокрема документно-орієнтовані моделі (наприклад, JSON, BSON тощо), що дає змогу ефективніше масштабувати системи та опрацювати великі масиви неструктурованої інформації. Ця функція є особливо важливою в контексті роботи з генеалогічними даними, де структура записів (поля) може змінюватися залежно від типу документа чи історичних стандартів часу створення. Крім того, NoSQL-бази даних є розподіленими системами, які забезпечують реплікацію даних на кількох серверах, що підвищує їхню доступність, надійність та безпеку. Ця перевага позитивно впливає на ефективність роботи зі значними обсягами архівної інформації, яка надходить під час оцифрування документів, їх безпечного збереження, опрацювання та подальшого аналізу [5].

Порівнюючи продуктивність роботи з великими масивами неструктурованих даних між HarperDB, MongoDB і PostgreSQL, дослідники зауважують, що основними перевагами використання NoSQL є архітектурна придатність до сучасних вимог [3]. Це виявляється у двох основних аспектах: по-перше, NoSQL значно спрощує опрацювання неструктурованих даних, оскільки ці системи природно використовують моделі, як-от документи чи ключ-значення, на відміну від жорстких табличних схем RDBMS; по-друге, NoSQL забезпечує вищу продуктивність операцій читання / запису у сценаріях Big Data, що частково пояснюється ефективним використанням функції відображення пам'яті. Ця властивість робить їх особливо цінними для роботи з масивними обсягами інформації, з якими традиційні реляційні системи історично мали проблеми [6].

Головна перевага документно-орієнтованих БД над традиційними SQL-базами полягає у гнучкості та здатності ефективно працювати з непередбачуваною природою даних. На відміну від реляційних систем управління базами даних (СУБД), що потребують жорсткого визначення схеми на початку проекту, NoSQL дають змогу:

— використовувати напівструктуровані дані, дозволяючи досліднику не визначати фіксовану схему на початкових етапах проекту;

— вносити модифікації у структуру даних під час дослідження, що є важливим для роботи з архівними джерелами, які часто є багатовимірними, непередбачуваними та складними;

— уникнути проблем, пов'язаних із потребою декомпонувати складні об'єкти на множину таблиць й уникнути супутньої надмірної жорсткості реляційних моделей [1].

З погляду опрацювання генеалогічних джерел, як-от метричні книги, акти цивільного стану, ревізькі казки й сімейні архіви, центральним об'єктом дослідження є особа, її життєві події та супровідна документація. Зазвичай дані про кожну особу є асиметричними та динамічними.

У SQL-базах даних головним обмеженням є завдання декомпозиції інформації, що передбачає створення окремих таблиць для подій і документів. Слід також підтримувати жорстку схему та зовнішні зв'язки. Якщо певна кількість осіб має унікальні поля, то для решти записів, де такого поля немає, відповідний зв'язок буде завжди порожнім. Попри це, система змушена підтримувати жорстку схему та зовнішні ключі, які рідко використовують. Щоб отримати повну інформацію про одну особу, потрібно виконати множинні JOIN-операції між головною та всіма допоміжними таблицями, що ускладнює запити й знижує продуктивність.

У NoSQL-базах даних усі дані та специфічні атрибути про особу інкапсульовано в одному документі. Нові специфічні атрибути, додані на пізньому етапі наповнення бази, вносять як нові поля до відповідних документів, без потреби змінювати глобальну схему чи створювати нову допоміжну таблицю. Під час моделювання родинних зв'язків (батько/дитина /шлюб) NoSQL-документ може або посилатися на ID інших документів в базі, або містити вбудовані дані (наприклад, список дітей), уникаючи складних багатотабличних зв'язків, характерних для SQL.

У дослідженні AtlantoCracies, зосередженому на вивченні іберійської аристократії в Іспанії й Америці, використано NoSQL-бази даних, що дало змогу дослідникам створити документ, який зберігається в колекції (документно-орієнтована модель) і містить потрібні метадані для здобуття повної інформації (рис. 1). Гнучкість такої системи надає можливість дотримуватися базової схеми в межах колекції, кожен документ може мати власну внутрішню структуру, яка повністю відрізняється від будь-якої іншої, що зберігається в тій самій колекції. Отже, модель даних у БД NoSQL є настільки ж динамічною, наскільки гнучкими й мінливими є історичні дані [1; 7].

У БД NoSQL є два основні способи зв'язку документів один з одним: за допомогою посилання чи вбудованих допоміжних документів, як зображено на рис. 2. Вбудовані документи є найкращим вибором для використання переваг високої продуктивності БД NoSQL [1].

```

{
  "_id": ObjectId("609cd66b9aab3a99799344bc"),
  "datebirth": "1716-01-21",
  "placebirth": "Ciudad de México",
  "pbirthUTM": "19.419444/-99.145556",
  "countryDeath": "México",
  "datedeath": "1786-08-29",
  "gender": "H",
  "histDeath": "Virreinato de Nueva España",
  "name": "Francisco Javier de Aristoarena de Lanz",
  "pDeathUTM": "25.816667/-100.133333",
  "pageFr": "Null",
  "personId": 117,
  "placeDeath": "Zacatecas",
  "source": "Ramón Maruri Villanueva Personal Database (American Spanish Nobility) ...",
  "sourceType": "Database",
  "titles": [
    {
      "continental": "América",
      "countryTit": "España",
      "dateTit": "1777",
      "histTit": "Virreinato de Nueva España",
      "nomTit": "Condado de Casa Fiel"
    }
  ],
  "relations": [
    {
      "idPerson": 710,
      "namePerson": "Josefa Tagle Bracho",
      "placeRel": "Ciudad de México",
      "placeRUTM": "19.419444/-99.145556",
      "treatment": "Doña",
      "typeRel": "Matrimonio"
    },
    {
      "idPerson": 769,
      "dateRel": "1770-12-03",
      "namePerson": "Maria Guadalupe de la Campa Cos",
      "placeRel": "Zacatecas",
      "placeRUTM": "25.816667/-100.133333",
      "treatment": "Doña",
      "typeRel": "Matrimonio"
    }
  ],
  "treatment": "Don",
  "countryBirth": "México",
  "hist30Birth": "Virreinato de Nueva España",
  "histBirth": "Virreinato de Nueva España",
  "continentBirth": "América"
}

```

Рис. 1. Приклад документа з БД AtlantoCracies [1]

У процесі оцифрування генеалогічних джерел формується великий обсяг різномірної інформації: цифрові зображення, скановані копії документів, результати оптичного розпізнавання тексту й супровідні метадані, які з'являються під час оцифрування документа. NoSQL-системи забезпечують збереження як основних файлів, так і пов'язаних із ними структурованих та напівструктурованих даних, як-от координати, часові мітки, іменовані сутності (осіб, місць, подій тощо).

Генеалогічна інформація може бути виявлена не лише у класичних джерелах на кшталт метричних книг, актів цивільного стану чи ревізьких казок, а й у сімейних архівах, списках виборців, поземельних книгах і судових справах. Ці документи мають нерівномірну структуру, змінну якість і часто потребують інтеграції даних, створених у різних правових і державних системах (Російська імперія, Австро-Угорщина, Польща, СРСР, сучасна Україна).

Традиційні SQL-бази даних стикаються з критичними обмеженнями в цьому контексті, оскільки їхня жорстка схема вимагає надмірної декомпозиції складного біографічного об'єкта на множину таблиць, що ускладнює запити та знижує продуктивність. Отже, з огляду на специфіку українських архівних фондів і документів, а саме їхню виняткову фрагментарність, багатомовність й асиметричність даних, використання БД NoSQL є оптимальним рішенням для ефективного управління такою інформацією.

Наведена модель зберігання дає змогу організувати механізми повнотекстового пошуку, автоматизованої класифікації, контекстного аналізу вмісту та формування семантичних зв'язків між даними. Приклад умовної схеми даних у NoSQL-базі з інформацією та посилання на суміжні документи в колекції подано на рис. 3.

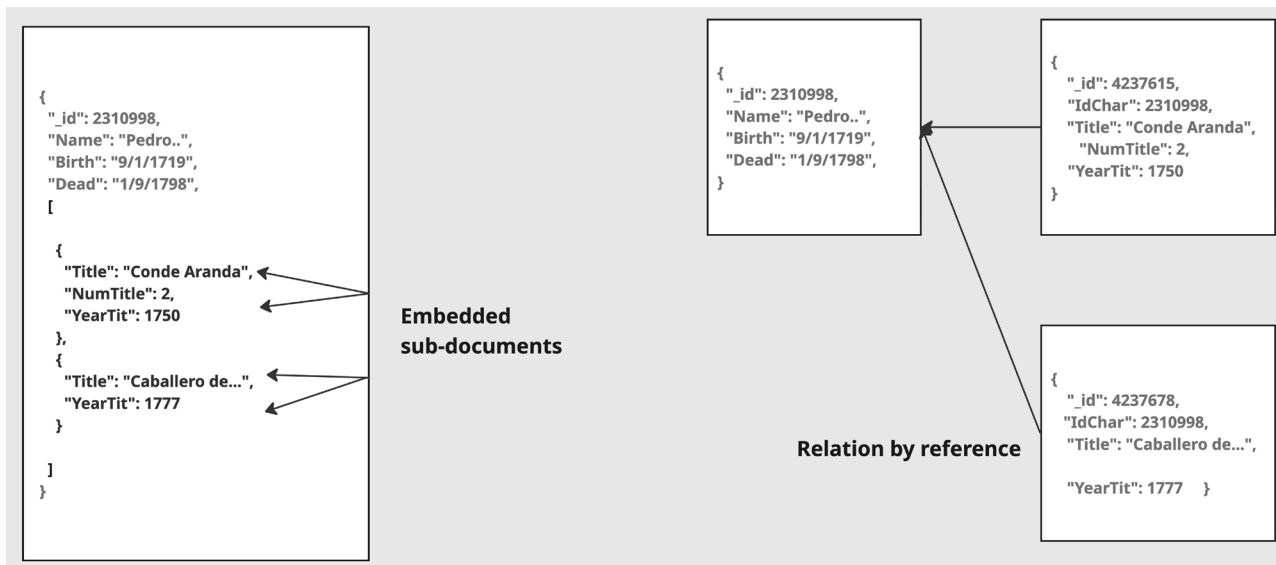


Рис. 2. Зв'язки в документно-орієнтованій БД в AtlantoCracies [1]

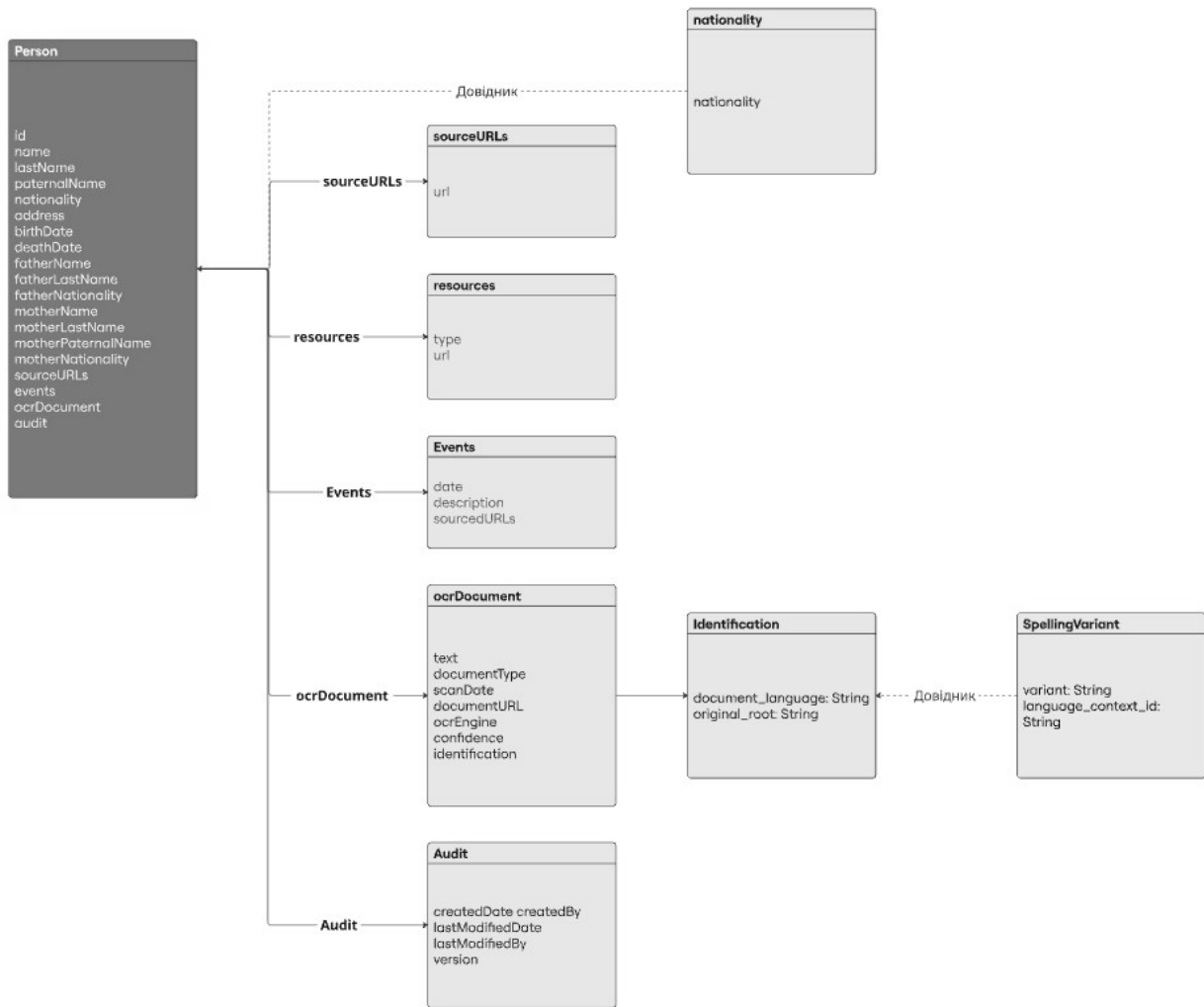


Рис. 3. Приклад умовної схеми генеалогічних даних у БД NoSQL

NoSQL допомагає інтегрувати в єдину модель дані з різномірних архівних фондів і поєднувати інформацію, яка б у реляційних базах потребувала значної декомпозиції таблиць. Наприклад, для особи (як центрального документа) можна об'єднати:

— базові структуровані дані (метрики):
 { 'birth_date': '1895-03-10', 'birth_place': 'с. Калинівка' };

— напівструктуровані біографічні дані (особові справи): { 'education': { 'institution': 'Київська політехніка', 'graduation_year': 1925, 'specialization': 'Інженер' } };

— майновий або економічний статус (поземельні книги): { 'property': { 'type': 'Estate', 'name': 'Квітневе', 'area_ha': 500, 'purchase_date': '1901-05-20' } } або: { 'property': { 'type': 'LAND_PLOT', 'area_desiatina': 5 } };

— мережеві зв'язки (судові справи, метрики): { 'social_roles': [{ 'role': 'witness_at_wedding', 'related_person_id': 'pid_456' }, { 'role': 'surety_in_court', 'related_person_id': 'pid_789' }] };

— еволюцію прізвища та багатомовність (метрики, списки): 'surname_main': 'Штерн', 'surname_original': 'Stern', 'surname_transliteration': 'Штерн', 'spelling_variants': [{ 'variant': 'Штернь', 'language_context': 'Російська' // Дореволюційний

варіант }, { 'variant': 'Sztern', 'language_context': 'Польська' // Типовий польський запис }], 'surname_OCR_error': ['Штерн', 'Штерн'] }.

Мовні й історичні зміни у фіксації прізвищ є одним з важливих елементів проведення генеалогічних досліджень. Територія України, як зазначалося, історично перебувала під впливом різних держав, що спричинило багатомовність записів (російська, польська, німецька, їдиш) і бюрократичну зміну стандартів. Одне прізвище може мати численні варіанти через транскрипцію, фонетичні відмінності чи асиміляцію (наприклад, Sztern — Штерн — Штайн).

NoSQL дає змогу вирішити цю проблему й згрупувати всі прізвища в єдиний атрибут (як, наприклад, spelling_variants); приклад умовного документа наведено на рис. 4. Ці функції забезпечують пошук певного прізвища чи імені в усіх можливих документах із різними варіаціями, що надає можливість одночасно підтримувати оригінальний текст (для архівістів), стандартизовану транскрипцію (для пошуку) й ідентифікувати потенційні помилки OCR, що є типовим викликом для дослідників. Таку практику вже використовують у різних архівах, наприклад у процесі пошуку в Arolsen Archives [8].

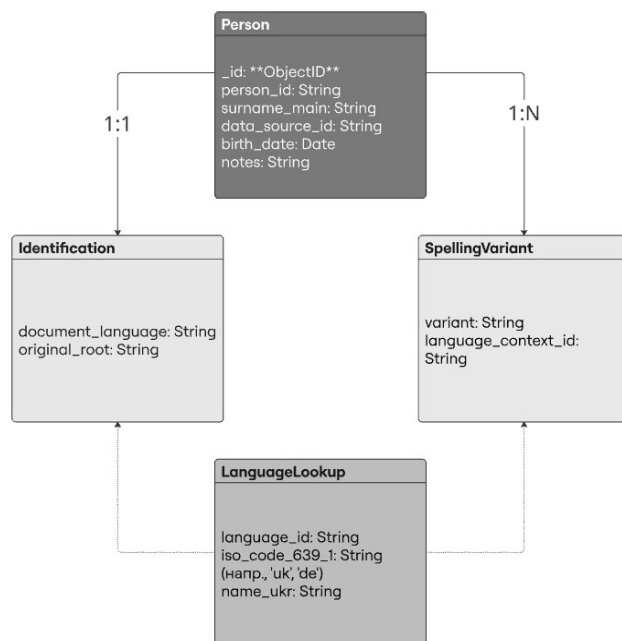


Рис. 4. Приклад схеми варіації прізвища в документі БД NoSQL

Централізоване застосування нереляційних БД в архівній сфері України має низку переваг, як-от підвищення ефективності доступу до інформації та якості науково-дослідної роботи. Упровадження гнучких схем і механізмів індексації дасть змогу забезпечити оперативний пошук, збирання та опрацювання інформації, яка розподілено зберігається в різних архівних фондах; значно скоротити час доступу до інформації й прискорити міжфондовий пошук. Гнучкість структури даних дозволить ефективно інтегрувати неструктуровані й напівструктуровані архівні дані, спрощуючи комплексні дослідження та формування складних запитів. Зокрема, важливим є підхід, який надає можливість зберігати результати оптичного розпізнавання символів, охоплюючи текстові шари, метадані й початкові зображення, в одному документі, що сприяє цілісності та подальшому опрацюванню даних. Використання внутрішніх зв'язків і гнучких індексів, зокрема графових чи повнотекстових, дає змогу не лише швидко знаходити інформацію, а й автоматично виявляти нові релевантні відомості для вже завершених чи поточних досліджень. Окрім того, важливо, що для національних архівних систем така архітектура забезпечить високу горизонтальну масштабованість, потрібну для управління зростаючими обсягами оцифрованих

архівних даних (Big Data), а також гарантує можливість інтеграції майбутніх технологій штучного інтелекту й машинного навчання.

Висновки. Застосування документно-орієнтованих баз даних у сфері зберігання та опрацювання архівних документів є одним із варіантів удосконалення розвитку архівної справи. Використання NoSQL дає змогу пришвидшити темпи оцифрування архівних фондів і спростити процес досліджень завдяки гнучким можливостям моделювання зв'язків усередині бази. На відміну від жорсткої структури традиційних SQL-баз, ця структура дає змогу зберігати великий обсяг інформації в одній колекції, як-от: текстові описи, результати роботи OCR, скановані копії зображень, метадані документа та ін. Використання індексів забезпечить можливість вдосконаленого контекстного пошуку, що є важливим для виявлення прихованих семантичних зв'язків між даними, які зберігаються в різних архівних фондах.

Отже, гнучкість NoSQL є оптимальним варіантом для побудови інформаційних систем, здатних ефективно підтримувати централізоване зберігання, опрацювання й аналіз генеалогічних досліджень, що, зі свого боку, дасть змогу забезпечити доступність культурної спадщини для дослідників і об'єднати різні заповнення архівні фонди.

Список бібліографічних посилань

1. A new approach for the construction of historical databases—NoSQL Document-oriented databases: the example of AtlantoCracies / Manuel Diaz-Ordoñez et al. *Digital Scholarship in the Humanities*. 2023. Vol. 38, Iss. 3 (September). P. 1014—1032. DOI: 10.1093/llc/fqad033.
2. Dedieu J.-P. Construire des bases de données. *Histoire & Mesure*. 2024. Vol. XXXIX-2. URL: <http://journals.openedition.org/histoiremesure/21642>. DOI: <https://doi.org/10.4000/140kh> (accessed 21.09.2025).
3. Performance Evaluation Between HarperDB, Mongo DB and PostgreSQL / D. Figueiredo et al. *Marketing and Smart Technologies. ICMarkTech 2022. Smart Innovation, Systems and Technologies*. 2024. Vol. 344. DOI: 10.1007/978-981-99-0333-7_7.

4. Артеменкова О. М. Інформаційний інструментарій генеалогічних досліджень в архівах України. *Вісник Харківської державної академії культури*. 2022. Вип. 61. С. 81—93. DOI: 10.31516/2410-5333.061.08. URL: <http://v-khsac.in.ua/article/view/261398> (accessed 12.10.2025).
5. Цифрові документовані колекції національної культурної спадщини на вебсайтах українських регіональних державних архівів / І. Тюрменко [та ін.]. *Developments in information & knowledge management for business applications*. 2022. Vol. 421. С. 449—470. DOI: 10.1007/978-3-030-97008-6_20. URL: https://link.springer.com/chapter/10.1007/978-3-030-97008-6_20 (accessed 12.10.2025).
6. NoSQL database. *IBM*. URL: <https://www.ibm.com/think/topics/nosql-databases>. (accessed 11.10.2025).
7. Atlantocracies. URL: <https://atlantocracies.com/> (accessed 11.10.2025).
8. Arolsen Archives — Міжнародний центр документації переслідувань націонал-соціалістами. URL: <https://arolsen-archives.org/uk/> (accessed 13.10.2025).

Artur Spektor,

Postgraduate Student

at the Department of Social Communications

and Information Activity,

Lviv Polytechnic National University

The use of NoSQL databases for the storage, processing, and analysis of genealogical archival documents

The article examines the relevance and substantiates the architectural model for the application of document-oriented NoSQL databases in the field of storage, processing, and analysis of genealogical documents. The relevance of the study is determined by the intensive processes of digitalization of archival collections in Ukraine and the need for technological solutions capable of efficiently handling large volumes of unstructured and heterogeneous information, such as digital images, optical character recognition (OCR) results, and accompanying metadata, the structure of which varies depending on the content of a particular archival collection. The purpose of the study is to develop and justify the use of NoSQL models as an optimal solution for managing genealogical data in archival science, which will enable centralized genealogical research management based on diverse archival sources. It is noted that the NoSQL model provides a flexible storage schema that allows the integration of all personal data into a single document. This approach addresses the problem of data diversity and multilingualism typical of Ukrainian archival collections. Moreover, the document-oriented approach ensures horizontal scalability for managing the growing volumes of Big Data and allows the use of flexible indexing mechanisms that facilitate advanced contextual search and the identification of hidden semantic relationships between data distributed across various collections. Based on the example of AtlantoCracies, it has been established that the implementation of NoSQL in archival science serves as a foundation for creating robust, high-performance systems that will significantly accelerate digitization processes and improve access to the national cultural heritage for research purposes.

Keywords: *databases; archival funds; archives; NoSQL; documentary-oriented model; genealogical research; digitalization; Big Data*

Надійшла до редакції 16.10.2025



ІСТОРИЧНІ РОЗВІДКИ

[https://doi.org/10.36273/2076-9555.2026.1\(354\).46-52](https://doi.org/10.36273/2076-9555.2026.1(354).46-52)

УДК 655.1:331.101.232-055.2]:070.484Czcionka=162.1(477.8-89Гал)"1872/1875":305(045)

Олег Гончар,

аспірант кафедри медійних технологій,

інформаційної та книжкової справи

Інституту поліграфії та медійних технологій

НУ "Львівська політехніка"

alegka83@ukr.net

<https://orcid.org/0000-0004-8660-5437>

Жіноча праця у друкарській справі крізь призму газети "Czcionka" (1872—1875)

Статтю присвячено аналізу дискусії щодо жіночої праці у друкарській галузі на сторінках польсько-мовної фахової газети "Czcionka", що виходила у Львові в 1872—1875 рр. і була першим профспілковим друкованим органом фахівців поліграфічного виробництва у Східній Галичині. Увагу зосереджено на