– виявлення аномалій у даних;
– інтелектуального кешування;
– семантичного пошуку та класифікації.

Окрему увагу привертають автономні системи керування базами даних, як-от Oracle Autonomous Database, Google AlloyDB та Azure SQL з Copilot. Ці системи використовують вбудовані алгоритми ШІ, забезпечуючи роботу за принципом self-driving databases. Вони здатні автоматично обирати оптимальні індекси, оновлювати конфігурацію, прогнозувати навантаження та рекомендувати оптимальні плани запитів [5].

Отже, сучасний розвиток штучного інтелекту радикально трансформує можливості систем керування базами даних. Інтеграція інтелектуальних алгоритмів дає змогу створити СКБД нового покоління – автономні, високопродуктивні та здатні до самостійного аналізу власних процесів. Найближчими роками очікується повна автоматизація більшості адміністративних завдань.

Список використаних джерел

1. SQL Server and Azure SQL: AI-Powered Optimization. *Microsoft.* 2024. URL: https://learn.microsoft.com/sql/ (дата звернення: 24.11.2025).
2. Autonomous Database Technical Overview. *Oracle.* 2023. URL: https://www.oracle.com/autonomous-database/ (дата звернення: 24.11.2025).
3. AI-Driven Data Management Trends. *Google Cloud.* 2025. URL: https://cloud.google.com/blog/ (дата звернення: 24.11.2025).
4. Intelligent Database Systems Report. *IBM Research.* 2024. URL: https://research.ibm.com (дата звернення: 25.11.2025).
5. NoSQL and Machine Learning Integration Guide. *MongoDB Inc.* 2023. URL: https://www.mongodb.com/resources (дата звернення: 25.11.2025).
6. Степанюк О. С., Січко Т. В. Особливості використання реляційних та нереляційних баз даних в Big Data. *Комп'ютерні технології обробки даних: матеріали всеукр. наук.-практ. конф.* Вінниця, 2020. С. 103–106.

УДК 004.6:[004.056.53:004.9](043.2)

Байраківська В. В., здобувачка 3 курсу,
Січко Т. В., канд. техн. наук, доцент,
доцент кафедри інформаційних технологій

ВПРОВАДЖЕННЯ ПРОЦЕДУР ОЧИЩЕННЯ ДАНИХ (DATA CLEANSING) ДЛЯ ПІДВИЩЕННЯ ДОСТОВІРНОСТІ ІНФОРМАЦІЇ В ІНФОРМАЦІЙНОМУ ПРОСТОРІ

Донецький національний університет імені Василя Стуса, м. Вінниця

Вступ. Електронні реєстри є важливим елементом сучасних інформаційних систем, оскільки забезпечують зберігання, обробку та використання структурованих даних у державному, медичному, фінансовому чи корпоративному секторах.

Достовірність даних у таких системах визначається їх точністю, повнотою, узгодженістю та актуальністю. Саме ці характеристики дають змогу ухвалювати правильні управлінські рішення та формувати надійну аналітичну інформацію. Проте на практиці низька якість даних є однією з головних проблем цифрових реєстрів. Вона проявляється у вигляді дублювання записів, наявності неповних або застарілих відомостей, суперечливих полів і помилкових форматів. Такі дефекти часто призводять до спотворення статистики, хибних аналітичних висновків, а іноді – до значних фінансових втрат або неправомірних управлінських рішень.

Мета роботи. Проаналізувати підходи і технології очищення даних у електронних реєстрах для підвищення їх достовірності та ефективності функціонування. Робота спрямована на обґрунтування важливості системного підходу до забезпечення якості даних і визначення практичних методів виявлення та усунення помилок.

Впровадження системного процесу очищення даних (Data Cleansing або Data Cleaning) є невід'ємним елементом забезпечення якості відомостей в інформаційному просторі. Під очищенням даних розуміють сукупність процедур, спрямованих на виявлення, виправлення або видалення некоректних, неповних, неточних чи дубльованих записів у базах даних [3]. На практиці виокремлюють кілька основних типів дефектів: неповнота (наявність пропусків або значень «null»), неточність (помилки введення, застарілі дані, неправильні формати), несуперечливість (розбіжності між записами про одну сутність) та дублювання (наявність кількох записів для одного об'єкта) [1; 6]. Кожен із цих дефектів знижує рівень довіри до джерела, ускладнює звітність і може спричиняти додаткові витрати на перевірку чи корекцію інформації.

Процес очищення даних зазвичай реалізується поетапно. Першим етапом є профілювання даних, тобто дослідження їх поточного стану з метою виявлення джерел і масштабів проблем. На цьому етапі аналізуються пропуски, аномалії, помилки форматів, частота дублювання записів та статистичні закономірності. Такий аналіз дає змогу визначити пріоритети подальших дій та оцінити, які саме ділянки потребують найглибшого втручання [5].

Другим етапом є розробка правил очищення, що визначає стандарти й бізнес-логіку корекції даних. До прикладу, можуть бути встановлені обов'язкові поля, стандартизовані формати дат або кодів, правила для виявлення та усунення дублікатів, а також алгоритми обробки пропущених значень (видалення, заповнення або позначення як невідомих). На цьому етапі важливо залучити експертів предметної області, аналітиків та IT-фахівців, щоб забезпечити узгодженість технічних і змістовних вимог [7].

Третім етапом є реалізація процедур очищення. Для цього застосовуються автоматизовані скрипти, ETL-інструменти (Extract → Transform → Load) та спеціалізовані програмні утиліти. Вони виконують перевірку відповідності правил, перетворення форматів, злиття або видалення дублікатів, виправлення значень за визначеними алгоритмами. Коли автоматизація не може забезпечити 100-відсоткову коректність результату, застосовується ручна верифікація даних. Особливо це важливо для контекстів, де потрібна логічна оцінка фахівця, наприклад, під час об'єднання близьких, але не ідентичних записів. Усі зміни повинні фіксуватися в журналах аудиту із зазначенням часу, виконавця та причини корекції [4].

Завершальний етап – моніторинг і аудит якості даних. Після первинного очищення необхідно впровадити систему постійного контролю, щоб запобігти появі нових дефектів. Для цього визначають ключові показники якості (наприклад, частка заповнених полів, рівень дублювання, кількість некоректних форматів), а також встановлюють регулярні перевірки та звітування. Як показують дослідження, впровадження безперервного моніторингу дає змогу скоротити кількість помилок до 70 % уже протягом перших шести місяців [2]. Окрім технічного контролю, важливим є й організаційний аспект – формування культури якості даних серед користувачів реєстру, навчання персоналу та призначення відповідальних осіб, які контролюють дотримання стандартів [5].

Технічний складник очищення даних включає використання різних методів і засобів. Найпоширенішими інструментами є ETL-платформи, утиліти для профілювання даних і системи моніторингу якості. Для перевірки правильності форматів застосовуються регулярні вирази (regex), для пошуку дублювань – нечітке зіставлення (fuzzy matching), що дає змогу знаходити схожі записи, навіть якщо вони містять орфографічні розбіжності [3; 4]. Важливо також реалізовувати стандартизацію даних: приведення назв, адрес, кодів або одиниць вимірювання до єдиного формату. Перед виконанням процедур очищення обов’язковим є резервне копіювання бази, щоб уникнути втрати інформації у разі помилкового видалення або зміни записів.

Висновки. Очищення даних є не одноразовою технічною операцією, а системним і безперервним процесом, який має бути інтегрований у загальну стратегію управління якістю даних. Його реалізація забезпечує підвищення достовірності інформації, покращення точності аналітичних звітів, скорочення ризиків помилкових рішень та економію ресурсів, адже усунення помилок на ранніх етапах завжди обходиться дешевше, ніж їх наслідки [1; 2]. Ефективне впровадження Data Cleansing створює підґрунтя для формування надійного інформаційного середовища, що відповідає сучасним вимогам цифрового урядування, прозорості й довіри. У перспективі розвиток методів очищення даних передбачає застосування технологій машинного навчання для автоматичного виявлення дефектів і прогнозування ризиків появи помилкових записів. Отже, системне очищення є стратегічним інструментом підвищення достовірності даних, який безпосередньо впливає на ефективність функціонування електронних реєстрів та прийняття управлінських рішень.

Список використаних джерел

1. Azeroual O., Saake G., Abuosba M. Data Quality Measures and Data Cleansing for Research Information Systems. *Journal of Digital Information Management.* 2018. Vol. 16(1). URL: https://dspacecris.eurocris.org/handle/11366/632 (дата звернення: 12.11.2025).

2. Using a Data Quality Framework to Clean Data Extracted from Electronic Health Records / O. Dziadkowiec et al. *Applied Clinical Informatics.* 2016. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC4933574/ (дата звернення: 12.11.2025).

3. Data Cleansing: What Is It and How It Works. Experian Data Quality (EDQ). URL: https://www.edq.com/glossary/data-cleansing/ (дата звернення: 12.11.2025).

4. Data Cleansing – Data Cleaning. DataSilk. https://datasilk.com/data-cleansing/
5. 10 Essential Data Cleansing Best Practices. Procogia (blog). URL: https://procogia.com/data-cleansing-best-practices/ (дата звернення: 12.11.2025).
6. Алексюк В. В., Січко Т. В. Дослідження особливостей програмного забезпечення для аналізу даних. *Комп'ютерні технології обробки даних:* матеріали всеукр. наук.-практ. конф., Вінниця, 2022. С. 249–251.

УДК 004.001

Алексюк В. В., здобувач 2 курсу
ОС «Магістр» спеціальності
F3 Комп'ютерні науки,
Потапова Н. А., канд. екон. наук, доцент,
доцент кафедри інформаційних технологій

ЗАСТОСУВАННЯ БАЗ ДАНИХ NOSQL У ВЕБЗАСТОСУНКАХ ДЛЯ АНАЛІТИКИ ПРОДАЖІВ

Донецький національний університет імені Василя Стуса, м. Вінниця

У сучасних умовах цифровізації бізнес-процесів електронна комерція генерує великі обсяги даних, пов'язаних із транзакціями, клієнтською активністю, товарними позиціями та маркетинговими кампаніями. Ефективна обробка та аналіз цієї інформації є критично важливими для прийняття управлінських рішень, прогнозування попиту та оптимізації асортименту [1]. Вебзастосунки, що виконують функції аналітики, повинні забезпечувати швидкий доступ до даних, гнучке структурування та інтерактивну візуалізацію результатів.

Одним із ключових компонентів таких систем є база даних. У випадку роботи з напівструктурованими даними, які можуть змінюватися з часом, традиційні реляційні СУБД часто виявляються недостатньо гнучкими. Саме тому все більшого поширення набувають NoSQL-рішення, зокрема MongoDB, які дають змогу адаптувати структуру даних до потреб аналітичної системи.

NoSQL (Not Only SQL) – це клас систем управління базами даних, які не використовують традиційну реляційну модель. MongoDB, як представник документно-орієнтованих СУБД, зберігає дані у форматі BSON (розширений JSON), що дає змогу:

– зберігати документи з довільною структурою;
– легко додавати нові поля без зміни схеми;
– використовувати вкладені об'єкти та масиви;
– виконувати складні агрегаційні запити.

Ці властивості особливо корисні для аналітичних вебзастосунків, де структура даних може змінюватися залежно від бізнес-завдань [2].

Аналітичні вебзастосунки, орієнтовані на обробку комерційної інформації, зазвичай працюють з різними типами даних: транзакціями, інформацією про