



## ОПТИМІЗАЦІЯ ОБСЯГУ МЕТАДАНИХ У СУЧАСНИХ ІНФОРМАЦІЙНИХ СИСТЕМАХ: МЕТОДИ, ІНСТРУМЕНТИ Й АЛГОРИТМІЧНІ ПІДХОДИ

**Вступ.** Дослідження питання оптимізації обсягу метаданих спрямовано на досягнення балансу між достатністю опису й ефективністю системи. Надмірний обсяг метаданих може перевантажувати системи, а недостатній – ускладнювати доступ до даних. Водночас зростаючий обсяг даних ускладнює роботу з метаданими, оскільки їхнє створення, зберігання та оброблення вимагають значних ресурсів. Оптимізація обсягу метаданих стала важливим завданням для організацій, які прагнуть досягти ефективного управління інформацією. Виклики, пов'язані з обсягом метаданих: надмірність, недостатність, дублювання, динамічність даних, невідповідність стандартам.

**Методи.** Розглянуто теоретичні основи, практичні методи, інструменти (Collibra, Apache Atlas, Talend Metadata Manager, AI-алгоритми) та переваги оптимізації обсягу метаданих (зменшення витрат, покращення продуктивності, підвищення якості даних, гнучкість і масштабованість, покращення аналітики).

**Результати.** Запропоновано комплексну стратегію оптимізації метаданих, адаптовану для IT-середовища. Показано, що використання систематичного підходу, включаючи аналіз, стандартизацію, автоматизацію та інтеграцію новітніх технологій, дає змогу значно зменшити витрати та покращити керування даними. Представлено алгоритм оптимізації обсягу метаданих, що може бути адаптований для різних сфер застосування, таких як бази даних, системи управління контентом чи великі дані.

**Висновки.** Запропонований алгоритм враховує: оцінювання корисності метаданих із використанням нормалізації метрик для уніфікації шкал оцінювання та з визначенням корисності кожного елемента метаданих; відбір метаданих (фільтрація, кластеризація); стиснення метаданих; автоматичну оптимізацію з використанням моделей машинного навчання та динамічного налаштування; перевірку й адаптацію. Алгоритм може бути розширений або змінений залежно від специфіки задачі.

**Ключові слова:** метадані, оптимізація, управління, інформація, динамічність, адаптивність, релевантність, продуктивність, якість, алгоритм, стиснення.

### Вступ

Метадані є структурованою інформацією, яка описує, пояснює, визначає або в інший спосіб полегшує пошук, використання і керування даними. У багатьох галузях, таких як інформаційні технології, бібліотечна справа, бізнес-аналітика й управління контентом, обсяг метаданих має вирішальне значення. Надмірний обсяг метаданих може перевантажувати системи, а недостатній – ускладнювати доступ до даних. Оптимізація обсягу метаданих спрямована на досягнення балансу між достатністю опису й ефективністю системи.

Метадані – ключовий компонент управління інформацією в цифрову епоху. Вони забезпечують опис, класифікацію та організацію даних, полегшуючи доступ і використання інформації. Водночас зростаючий обсяг даних ускладнює роботу з метаданими, оскільки їхнє створення, зберігання й оброблення вимагають значних ресурсів.

Оптимізація обсягу метаданих стала важливим завданням для організацій, які прагнуть досягти ефективного управління інформацією. У цій статті ми розглянемо теоретичні основи, практичні методи, інструменти та переваги оптимізації обсягу метаданих.

Аналіз попередніх досліджень показує, що проблема оптимізації метаданих розвивалася поступово та в різних напрямках.

У дослідженнях (Ghemawat et al., 2003), (Dean, & Ghemawat, 2004) увагу направлено на масштабовані системи зберігання й обробки інформації. Google File System та MapReduce створили передумови для управління величезними обсягами метаданих у розподілених обчислювальних середовищах. Сильна сторона цих робіт полягала у практичній придатності для "великих даних", тоді як слабким місцем стала складність інтеграції зі стандартними метамоделями.

Дослідження (Bizer et al., 2009) у напрямі "зв'язаних даних" запропонували нові підходи до інтероперабельності метаданих завдяки семантичним вебтехнологіям. Їхньою перевагою є універсальність і масштабованість, однак неоліком виступає складність упровадження у традиційних системах.

Новий етап розвитку пов'язаний із застосуванням методів машинного навчання та нейронних мереж. Роботи (Mikolov et al., 2013) та (Joulin et al., 2017) показали, що автоматична генерація й відбір релевантних метаданих можуть істотно зменшити їхній обсяг, зберігаючи релевантність для аналітики. Перевагою цього підходу є висока адаптивність і можливість самонавчання, а неоліком – залежність від якості навчальних даних і ресурсомісткість обчислень. Новітні публікації, такі як (Korshun et al., 2023), інтегрують штучний інтелект у процеси керування метаданими в операційних системах, а (Turovsky et al., 2024) пропонують критерії для оцінювання ефективності методів роботи з інформацією, що демонструє міждисциплінарність і прикладну спрямованість сучасних досліджень.

У сучасних інформаційних системах метадані виконують функцію посередника між даними та користувачем, сприяючи інтеграції, пошуку й підвищенню якості інформаційних процесів.

Доречно виокремити такі основні функції метаданих:

- опис даних – інформація про тип, структуру та зміст даних;
- керування даними – полегшення зберігання, пошуку й оброблення великих масивів даних;



- інтеграція даних – підтримка сумісності між різними системами та форматами;
- аналітика – підвищення якості аналізу даних за рахунок доступності релевантної інформації.

З огляду на ці функції, важливо знайти баланс між достатністю метаданих для виконання їхніх завдань і їхнім обсягом, щоб уникнути перевантаження системи.

У процесі забезпечення ефективності використання метаданих виникають виклики, пов'язані з їхнім обсягом:

▪ надмірність – занадто великий обсяг метаданих призводить до перевитрати ресурсів на зберігання, оброблення та передачу інформації;

- недостатність – нестача метаданих ускладнює пошук і управління інформацією, знижуючи ефективність систем;
- дублювання – повторювані метадані створюють плутанину та збільшують вимоги до обчислювальних ресурсів;
- динамічність даних – постійне оновлення інформації вимагає гнучких підходів до управління метаданими;
- невідповідність стандартам – відсутність єдиної структури метаданих ускладнює інтеграцію систем.

Враховуючи основні функції метаданих і виклики щодо їхнього обсягу, сформуємо ключові принципи оптимізації обсягу метаданих:

- релевантність – метадані повинні відповідати конкретним потребам користувачів і завданням систем;
- стислість – скорочення обсягу метаданих до мінімально необхідного рівня для виконання їхньої функції;
- автоматизація – використання технологій для автоматичного створення, оновлення та видалення метаданих;
- стандартизація – впровадження єдиних стандартів метаданих для підвищення сумісності між системами;
- адаптивність – гнучкість у підходах до створення метаданих для різних типів даних і завдань.

Значущість оптимізації метаданих виявляється у її впливі на різні параметри інформаційних систем, основними з яких варто виділити такі:

1) продуктивність системи – занадто великий обсяг метаданих може уповільнювати оброблення запитів, індексацію та пошук; це особливо критично для великих баз даних або хмарних сховищ, де швидкість доступу має вирішальне значення;

2) зручність використання – метадані повинні бути зрозумілими та корисними для кінцевого користувача; надмірно детальний опис може заплутати, тоді як брак інформації може зробити об'єкт важким для ідентифікації;

3) вартість зберігання – хоча зберігання даних стало дешевшим, великий обсяг метаданих у масштабах великих організацій може суттєво збільшити витрати;

4) управління якістю даних – надмірна кількість метаданих ускладнює їхню перевірку й оновлення, що може призводити до помилок і дублювання.

### Методи

Досягнення балансу між достатністю опису та продуктивністю системи можливе завдяки застосуванню комплексних методів оптимізації метаданих:

1) аналіз і класифікація даних – перший крок в оптимізації, що передбачає аналіз типів даних і визначення їхніх основних характеристик, для цього необхідно: визначити категорії даних (напр., текстові файли, зображення, відео), встановити ключові атрибути для кожної категорії (назва, дата створення, автор тощо);

2) стандартизація метаданих – використання загальноприйнятих стандартів, таких як Dublin Core, ISO 19115 або METS, що забезпечує єдину структуру даних і зменшує дублювання;

3) автоматизація процесів – сучасні інструменти, такі як AI і машинне навчання, дозволяють автоматично створювати метадані на основі аналізу вмісту файлів, наприклад: розпізнавання об'єктів на зображеннях, аналіз текстів для автоматичного визначення ключових слів;

4) видалення дублікатних даних – інструменти для пошуку та видалення дублікатних метаданих, такі як Talend або Apache Atlas, дозволяють скоротити обсяг інформації, що зберігається;

5) контроль рівня деталізації – диференційований підхід до опису даних, оскільки не всі об'єкти потребують однакового рівня деталізації, наприклад: документи для архівування можуть мати мінімальний набір метаданих, активно використовувані дані потребують детального опису;

6) видалення застарілих метаданих – регулярна перевірка й очищення систем дозволяють видаляти нерелевантну або застарілу інформацію.

Практичну реалізацію оптимізації метаданих забезпечують спеціалізованими інструментами, які автоматизують процеси аналізу, стандартизації та керування даними:

- Collibra – інтегрована платформа для керування метаданими;
- Apache Atlas – система для каталогізації, класифікації та стандартизації метаданих;
- Talend Metadata Manager – інструмент для аналізу й оптимізації метаданих у великих системах;
- AI-алгоритми – інструменти машинного навчання, які автоматизують процес створення й управління метаданими.

Впровадження методів оптимізації метаданих забезпечує значні переваги, пов'язані з продуктивністю, якістю й економічністю управління даними:

- зменшення витрат – скорочення витрат на зберігання й оброблення інформації;
- покращення продуктивності – швидкий доступ до даних і зменшення часу на виконання запитів;
- підвищення якості даних – зменшення кількості помилок і дублювань;
- гнучкість і масштабованість – легкість адаптації систем до нових вимог;
- покращена аналітика – забезпечення доступу до релевантної інформації для аналітичних завдань.

Практичні кейси застосування оптимізації метаданих підтверджують її значущість для корпоративних баз даних, хмарних сервісів та аналітичних платформ:

1) корпоративні бази даних – великі компанії, такі як Google чи Amazon, використовують оптимізацію метаданих для забезпечення швидкого пошуку й ефективного керування інформацією;

2) бібліотеки й архіви – національні бібліотеки впроваджують стандарти Dublin Core для каталогізації книг і цифрових ресурсів;



3) хмарні системи – Dropbox і Microsoft OneDrive автоматично створюють метадані на основі вмісту файлів, зменшуючи обсяг описів без втрати корисності;

4) медіаіндустрія – Netflix оптимізує метадані, щоб покращити рекомендації для користувачів, застосовуючи машинне навчання.

Майбутнє оптимізації метаданих відкриває нові горизонти для підвищення ефективності оброблення, зберігання й аналітики даних, створюючи умови для інтелектуальних систем нового покоління:

- інтеграція штучного інтелекту – застосування AI для автоматичного управління метаданими;
- гіперавтоматизація – поєднання кількох технологій (AI, RPA, NLP) для створення комплексних рішень;
- семантичні метадані – використання онтологій і семантичного вебу для покращення інтерпретації даних;
- енергоефективність – розроблення рішень, які знижують споживання енергії під час оброблення метаданих.

Метадані відіграють важливу роль у функціонуванні інформаційних систем, сприяючи ефективному управлінню даними, покращенню їхньої доступності та підтримці інтеграції між різними платформами. Проте обсяги даних, що постійно зростають, вимагають розроблення стратегій для оптимізації метаданих, які забезпечують баланс між їхньою повнотою, корисністю та продуктивністю систем.

В роботі запропоновано комплексну стратегію оптимізації метаданих, адаптовану для IT-середовища:

1) оцінювання поточного стану метаданих – аналіз існуючих метаданих, мета – зрозуміти поточний стан управління метаданими, ідентифікувати надмірні, дубльовані або нерелевантні метадані, дії – провести аудит метаданих у базах даних, сховищах, системах управління контентом, виявити джерела дублювання та надлишкової деталізації, визначити ключові атрибути метаданих, що найчастіше використовуються, інструменти для аналізу – Apache Atlas, Collibra, Talend Metadata Manager;

2) визначення цілей оптимізації – пріоритети оптимізації (зниження витрат на зберігання й оброблення даних), підвищення продуктивності системи, полегшення пошуку й аналізу даних, підтримка сумісності між різними системами, оцінювання успіху (встановити KPI (ключові показники ефективності), наприклад: час виконання пошукових запитів), скорочення обсягу збережених метаданих, рівень задоволеності користувачів;

3) стандартизація метаданих – вибір стандартів (використання загальноновизнаних стандартів забезпечує узгодженість і спрощує інтеграцію даних), для загальних метаданих – Dublin Core, для просторових даних – ISO 19115, для бібліотек і архівів – MARC (Machine-Readable Cataloging), реалізація – розробити шаблони для метаданих, що відповідають стандартам, переконатися, що всі нові метадані відповідають обраній моделі;

4) автоматизація процесів – збір і генерація метаданих (автоматизація створення й оновлення метаданих, щоб мінімізувати ручну роботу), інструменти – AI для аналізу текстів, зображень і відео, системи ETL (Extract, Transform, Load) для автоматичного імпорту даних, моніторинг та управління (використання інструментів, що постійно перевіряють метадані на дублювання або нерелевантність), приклад – у системах CRM автоматизована генерація метаданих (напр., дата останньої активності клієнта) забезпечує актуальність інформації;

5) контроль рівня деталізації – принципи деталізації (мінімум для архівів – метадані, що забезпечують базову ідентифікацію (назва, дата, автор), розширена деталізація для активних даних – метадані з додатковими атрибутами, зокрема й теги, категорії, технічні характеристики), ієрархія метаданих – створення структури, що дозволяє динамічно додавати або прибирати метадані залежно від контексту використання;

6) видалення дубльованих і застарілих метаданих – очищення системи (регулярно проводьте аналіз метаданих для видалення дублікатної або нерелевантної інформації), автоматизовані інструменти (Talend Data Quality), ручний аудит (залучайте експертів для аналізу важливих наборів даних), приклад – в архівах електронної пошти дублювання тем та авторів листів можна усунути за допомогою інструментів очищення даних;

7) використання семантичних метаданих – це метадані, що не лише описують дані, але й додають контекстуальний сенс за допомогою онтологій або зв'язків між об'єктами, переваги – покращення пошуку за змістом, підвищення точності аналітики, інструменти – RDF (Resource Description Framework), OWL (Web Ontology Language), приклад – в системах електронної комерції семантичні метадані можуть допомогти встановлювати зв'язки між продуктами на основі категорій, рекомендацій або історії покупок;

8) інтеграція із хмарними технологіями – хмарне управління метаданими (хмарні платформи пропонують інструменти для автоматизації та оптимізації метаданих – AWS Glue (Amazon), Google Cloud Data Catalog, Azure Data Factory), переваги – глобальна доступність, масштабованість, мінімізація витрат на локальне зберігання, приклад – організація використовує Google Cloud Data Catalog для автоматизації збору метаданих у розподілених базах даних;

9) забезпечення безпеки метаданих – метадані можуть містити конфіденційну інформацію, наприклад, про користувачів або конфігурації системи, методи захисту – шифрування чутливих метаданих, обмеження доступу до метаданих за допомогою ролей і прав доступу, виявлення аномалій у доступі до метаданих, приклад – в системах управління даними фінансових установ доступ до метаданих обмежений лише для авторизованих співробітників;

10) навчання персоналу – розуміння важливості (персонал має розуміти важливість оптимізації метаданих і володіти базовими навичками управління ними), програми навчання – навчання роботи зі стандартами метаданих, курси з використання інструментів автоматизації, розвиток навичок роботи із семантичними технологіями, приклад – команда IT-відділу проходить тренінг із використання Apache Atlas для управління метаданими.

Представлений алгоритм оптимізації обсягу метаданих може бути адаптований для різних сфер застосування, таких як бази даних, системи управління контентом чи великі дані. На рис. 1 представлено блок-схему алгоритму оптимізації обсягу метаданих.

1. Ініціалізація:

- 1) вхідні дані:
  - колекція метаданих

$$M = \{m_1, m_2, \dots, m_n\},$$

де  $m_i$  – окремий елемент метаданих;



- метрики якості метаданих Q, наприклад, релевантність, унікальність, необхідність;
  - обмеження ресурсів – пам'ять, час оброблення, обсяг даних;
- 2) цільова функція:
- зменшення обсягу метаданих |M|, із забезпеченням заданого рівня якості Qmin.
2. Оцінювання корисності метаданих:
- 1) для кожного елемента метаданих mi:
    - обчислити метрики Q(mi), включаючи:
      - релевантність – наскільки елемент метаданих відповідає цільовим завданням;
      - частоту використання – як часто елемент запитується чи використовується;
      - час створення / оновлення – чи є дані актуальними;
    - присвоїти ваговий коефіцієнт wi для кожного Q(mi);
  - 2) використати нормалізацію метрик для уніфікації шкал оцінювання;
  - 3) визначити корисність кожного елемента метаданих:

$$U(m_i) = \sum_{j=1}^k w_j \cdot Q_j(m_i),$$

де  $w_j$  – ваговий коефіцієнт, а  $Q_j(m_i)$  – оцінка за  $j$ -ю метрикою.



Рис. 1. Блок-схема алгоритму оптимізації обсягу метаданих

3. Відбір метаданих:
- 1) фільтрація:
    - видалити дублікати або надлишкові метадані, які мають низькі показники корисності:

$$U(m_i) < \epsilon,$$

де  $\epsilon$  – порогове значення.



- 2) кластеризація:
  - згрупувати схожі метадані за ключовими ознаками (напр., тематикою чи форматом);
  - залишити репрезентативний елемент для кожного кластера.
4. Стиснення метаданих:
  - 1) алгоритми стиснення:
    - використати методи компресії, такі як:
      - Huffman coding для текстових даних;
      - Run-length encoding (RLE) для повторюваних даних;
      - бінарне кодування для структурованих метаданих;
  - 2) видалення зайвих полів:
    - виключити необов'язкові атрибути, які зрідка використовуються.
5. Автоматична оптимізація:
  - 1) моделі машинного навчання:
    - використати класифікаційні алгоритми для передбачення корисності метаданих на основі їхніх характеристик;
    - застосувати нейронні мережі або дерева рішень для автоматичного відбору релевантних метаданих;
  - 2) динамічне налаштування:
    - оптимізувати обсяг метаданих залежно від запитів користувачів і доступних ресурсів.
6. Перевірка й адаптація:
  - 1) оцінити якість скорочених метаданих за метриками Q;
  - 2) якщо  $Q < Q_{\min}$ , повернутись до етапу фільтрації або коригування параметрів;
  - 3) забезпечити оновлення метаданих у реальному часі для підтримки актуальності.
7. Вихід – оптимізований набір метаданих  $M_{opt}$ , який задовольняє обмеження на обсяг і зберігає необхідний рівень якості. Цей алгоритм може бути розширений або змінений залежно від специфіки задачі (напр., для конкретних доменів чи типів метаданих).

#### Дискусія і висновки

Оптимізація обсягу метаданих є ключовим фактором успішного управління інформацією в сучасних системах. Баланс між обсягом, якістю та релевантністю метаданих дозволяє досягти ефективності, зменшити витрати та забезпечити зручність для користувачів. У майбутньому розвиток технологій, таких як штучний інтелект і семантичний веб, відкриє нові можливості для вдосконалення процесів оптимізації.

Оптимізація метаданих в ІТ є критичним процесом, який забезпечує ефективність, продуктивність та безпеку інформаційних систем. Використання систематичного підходу, включаючи аналіз, стандартизацію, автоматизацію та інтеграцію новітніх технологій, дозволяє значно зменшити витрати та покращити керування даними.

Впровадження цієї стратегії сприятиме зростанню бізнес-ефективності та конкурентоспроможності організацій у цифрову епоху.

**Внесок авторів.** Ольга Ткаченко – концептуалізація; Андрій Лемешко – аналіз джерел, підготовка огляду літератури.

**Джерела фінансування.** Це дослідження не отримало жодного гранта від фінансової установи в державному, комерційному або некомерційному секторах.

#### Список використаних джерел

- Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data – The story so far. *International Journal on Semantic Web and Information Systems*, 5(3), 1–22. <https://doi.org/10.4018/jswis.2009081901>
- Dean, J., & Ghemawat, S. (2004). MapReduce: Simplified data processing on large clusters. *Communications of the ACM*, 51(1), 107–113. <https://doi.org/10.1145/1327452.1327492>
- Ghemawat, S., Gobioff, H., & Leung, S.-T. (2003). The Google file system. In *Proceedings of the 19th ACM Symposium on Operating Systems Principles (SOSP '03)* (pp. 29–43). Association for Computing Machinery. <https://doi.org/10.1145/945445.945450>
- Joulin, A., Grave, E., Bojanowski, P., & Mikolov, T. (2017). Bag of tricks for efficient text classification. In M. Lapata, P. Blunsom, & A. Koller (Eds.), *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Vol. 2, Short Papers* (pp. 427–431). Association for Computational Linguistics. <https://doi.org/10.18653/v1/E17-2068>
- Korshun, N., Myshko, I., & Tkachenko, O. (2023). Automation and management in operating systems: The role of artificial intelligence and machine learning. In *Proceedings of the 20th International Scientific Conference Dynamical System Modeling and Stability Investigation (DSMSI 2023: Mathematical Foundations of Information Technologies)* (pp. 59–68). Igor Sikorsky Kyiv Polytechnic Institute; published online in CEUR Workshop Proceedings. <https://ceur-ws.org/Vol-3687/>
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient estimation of word representations in vector space*. <https://doi.org/10.48550/arXiv.1301.3781>
- Turovsky, O., Tkachenko, O., Ghno, G. S. N., & Abed, A. M. (2024). Selection and substantiation of the system of criteria for evaluating the effectiveness of steganographic methods of hiding information in the image. In *Proceedings of the 35th Conference of Open Innovations Association (FRUCT)* (pp. 755–763). Open Innovations Association (FRUCT). <https://doi.org/10.23919/FRUCT61870.2024.10516392>

#### References

- Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data – The story so far. *International Journal on Semantic Web and Information Systems*, 5(3), 1–22. <https://doi.org/10.4018/jswis.2009081901>
- Dean, J., & Ghemawat, S. (2004). MapReduce: Simplified data processing on large clusters. *Communications of the ACM*, 51(1), 107–113. <https://doi.org/10.1145/1327452.1327492>
- Ghemawat, S., Gobioff, H., & Leung, S.-T. (2003). The Google file system. In *Proceedings of the 19th ACM Symposium on Operating Systems Principles (SOSP '03)* (pp. 29–43). Association for Computing Machinery. <https://doi.org/10.1145/945445.945450>
- Joulin, A., Grave, E., Bojanowski, P., & Mikolov, T. (2017). Bag of tricks for efficient text classification. In M. Lapata, P. Blunsom, & A. Koller (Eds.), *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Vol. 2, Short Papers* (pp. 427–431). Association for Computational Linguistics. <https://doi.org/10.18653/v1/E17-2068>
- Korshun, N., Myshko, I., & Tkachenko, O. (2023). Automation and management in operating systems: The role of artificial intelligence and machine learning. In *Proceedings of the 20th International Scientific Conference Dynamical System Modeling and Stability Investigation (DSMSI 2023: Mathematical Foundations of Information Technologies)* (pp. 59–68). Igor Sikorsky Kyiv Polytechnic Institute; published online in CEUR Workshop Proceedings. <https://ceur-ws.org/Vol-3687/>



Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient estimation of word representations in vector space*. <https://doi.org/10.48550/arXiv.1301.3781>  
Turovsky, O., Tkachenko, O., Ghno, G. S. N., & Abed, A. M. (2024). Selection and substantiation of the system of criteria for evaluating the effectiveness of steganographic methods of hiding information in the image. In *Proceedings of the 35th Conference of Open Innovations Association (FRUCT)* (pp. 755–763). Open Innovations Association (FRUCT). <https://doi.org/10.23919/FRUCT61870.2024.10516392>

Отримано редакцією журналу / Received: 25.06.25  
Прорецензовано / Revised: 20.09.25  
Схвалено до друку / Accepted: 20.09.25

Olga TKACHENKO, DSc (Engin.), Prof.  
ORCID ID: 0000-0001-7983-9033  
e-mail: olga.tkachenko@knu.ua  
Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

Andriy LEMESHKO, PhD, Assoc. Prof.  
ORCID ID: 0000-0001-8003-3168  
e-mail: andriy.lemeshko@knu.ua  
Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

## OPTIMIZATION OF METADATA VOLUME IN MODERN INFORMATION SYSTEMS: METHODS, TOOLS AND ALGORITHMIC APPROACHES

**B a c k g r o u n d .** *The study of metadata volume optimization aims to achieve a balance between the sufficiency of description and system efficiency. Excessive metadata can overload systems, and insufficient metadata can complicate data access. At the same time, the growing volume of data complicates work with metadata, since its creation, storage and processing require significant resources. Metadata volume optimization has become an important task for organizations seeking to achieve effective information management. Challenges associated with metadata volume: redundancy, insufficiency, duplication, data dynamics, non-compliance with standards.*

**M e t h o d s .** *The paper considers the theoretical foundations, practical methods, tools (Collibra, Apache Atlas, Talend Metadata Manager, AI algorithms) and the benefits of metadata volume optimization (cost reduction, improved productivity, improved data quality, flexibility and scalability, improved analytics).*

**R e s u l t s .** *The paper proposes a comprehensive metadata optimization strategy adapted for the IT environment. It is shown that the use of a systematic approach, including analysis, standardization, automation and integration of the latest technologies, allows to significantly reduce costs and improve data management. An algorithm for optimizing the volume of metadata is presented, which can be adapted for various application areas, such as databases, content management systems or big data. The proposed algorithm takes into account the assessment of metadata usefulness using metric normalization to unify evaluation scales and determine the usefulness of each metadata element; metadata selection (filtering, clustering); metadata compression; automatic optimization using machine learning models and dynamic tuning; verification and adaptation. The algorithm can be expanded or modified depending on the specifics of the task.*

**C o n c l u s i o n s .** *The proposed algorithm takes into account: metadata usefulness assessment using metric normalization to unify evaluation scales and determine the usefulness of each metadata element; metadata selection (filtering, clustering); metadata compression; automatic optimization using machine learning models and dynamic tuning; testing and adaptation. The algorithm can be expanded or modified depending on the specifics of the task.*

**K e y w o r d s :** *metadata, optimization, management, information, dynamism, adaptability, relevance, performance, quality, algorithm, compression*

Автори заявляють про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження; у зборі, аналізі чи інтерпретації даних; у написанні рукопису; в рішенні про публікацію результатів.

The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; in the decision to publish the results.