ТЗ. №5(37)’2025 https://doi.org/10.32782/2786-9024/v3i5(37).344519

УДК 004.8:004.67: 004.8

ІНТЕЛЕКТУАЛЬНІ ТЕХНОЛОГІЇ АНАЛІЗУ, КЛАСИФІКАЦІЇ ТА РЕКОМЕНДАЦІЙ У СИСТЕМАХ УПРАВЛІННЯ КОЛЕКЦІЯМИ

Д.І. Поперешняк[1], С.В. Поперешняк[1, 2]

[1] Department of Software Engineering, State University of Information and Communication Technologies, Kyiv, Ukraine

[2] Department of Informatics and Software Engineering, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Kyiv, Ukraine

E-mail: spopereshnyak@gmail.com

АНОТАЦІЯ

Статтю присвячено дослідженню та розробці інтелектуальних технологій аналізу, класифікації та рекомендацій у системах управління колекціями. У роботі розглядається проблема організації та інтелектуальної обробки колекційних даних, які відзначаються постійним зростанням обсягів і різнорідністю, що ускладнює їх ефективне збереження, пошук та систематизацію. Показано, що традиційні системи управління здебільшого зосереджуються на ізольованих завданнях, як-от базове збереження чи пошук інформації, залишаючи поза увагою питання комплексної інтеграції алгоритмів аналізу подібності, класифікації та рекомендацій. Використання інтелектуальних підходів дає змогу подолати ці обмеження та сформувати універсальні системи, придатні для роботи з великими обсягами даних.

Метою дослідження є досягнення підвищеної ефективності та масштабованості управління колекціями шляхом побудови моделі інтелектуальної системи, яка інтегрує формалізацію мультимодальних ознак об’єктів, методи класифікації та виявлення дублікатів, а також рекомендаційні механізми. Для досягнення цієї мети проаналізовано існуючі підходи до роботи з колекціями, сформульовано модель представлення об’єктів, розроблено методику обчислення подібності та запропоновано архітектуру, що містить класифікаційні й рекомендаційні модулі.

Експериментальна перевірка показала, що інтеграція зазначених компонентів забезпечує високу точність класифікації, ефективне виявлення дублікатів і релевантність персоналізованих рекомендацій. Особливу увагу приділено масштабованості: система зберігає стабільний час відповіді навіть за умови значного збільшення кількості об’єктів. Практична цінність полягає в універсальності моделі, яка може застосовуватися як у приватних цифрових і мультимедійних колекціях, так і в корпоративних чи наукових інфраструктурах. Перспективи подальших досліджень пов’язані з упровадженням глибинного навчання для обробки мультимодальних ознак, удосконаленням рекомендаційних алгоритмів і посиленням інтеграції з механізмами безпеки та контролю доступу.

Ключові слова: *інтелектуальні системи, управління колекціями, аналіз подібності, класифікація, рекомендаційні системи, мультимодальні дані, масштабованість, IoT.*

Вступ

У сучасному інформаційному суспільстві обсяг цифрових даних зростає надзвичайно швидкими темпами, охоплюючи найрізноманітніші сфери – від приватних мультимедійних архівів до масштабних корпоративних і наукових сховищ. Це вже не лише бібліотеки електронних книг чи фотоархіви, які зберігають користувачі вдома. Ідеться про корпоративні бази, наукові сховища, мультимедійні колекції, що налічують мільйони елементів. Одночасно із цим відбувається ускладнення структури колекцій, які містять числові, категоріальні, текстові та мультимедійні дані. Чим більш різнорідними стають ці дані, тим складніше організувати їх зберігання та забезпечити швидкий доступ.

Більшість традиційних систем управління колекціями справді добре виконують базові функції: зберігають інформацію, дають змогу здійснити пошук. Однак, коли йдеться про масштабування чи інтеграцію з іншими середовищами, з’являються проблеми. Одна зі слабких сторін – практично повна відсутність інтелектуальних механізмів аналізу та рекомендацій. Окремим джерелом поповнення колекцій є потоки даних від IoT-пристроїв (сенсорика, мультимедійні

ТЗ. №5(37)'2025 https://doi.org/10.32782/2786-9024/v3i5(37).344519

фіди), що додатково ускладнює гетерогенність та вимоги до масштабованості.

Останні роки показали, що вирішити цю проблему можна шляхом використання інтелектуальних технологій. Методи аналізу даних, алгоритми класифікації та рекомендаційні підходи надають можливість автоматизувати роботу з колекціями, позбавитися зайвих дублікатів і створювати персоналізовані пропозиції для користувачів. Особливо ефективними такі підходи є у хмарних середовищах, де важливі масштабованість і доступність сервісів.

Таким чином, використання інтелектуальних методів у системах управління колекціями є одним із найперспективніших напрямів у сфері інформаційних технологій. Це питання має і практичний вимір, і теоретичну вагу, адже йдеться про створення нових моделей, здатних поєднати різні аспекти обробки даних у єдину архітектуру.

Аналіз літературних даних і постановка проблеми

Сучасні системи управління колекціями стикаються з вибуховим зростанням обсягів та різнорідності даних, тож поряд із базовим збереженням і пошуком на перший план виходять автоматизована класифікація, інтелектуальний аналіз і рекомендації. Оптимізація зберігання передусім зводиться до дедуплікації: огляди показують, що навіть помірний рівень повторів суттєво погіршує ефективність сховищ, а підходи chunk-based знижують витрати пам'яті та прискорюють доступ [1; 2]. Паралельно досліджуються моделі безпеки, зокрема когнітивні парадигми, що поєднують криптографію та семантику для підвищення довіри до даних [3]. У рекомендаційних системах найкращі результати демонструють гібридні схеми, які поєднують контентний і колаборативний підходи, однак більшість рішень лишається домен-специфічною і погано переноситься на гетерогенні колекції [4]. Узагальнюючи, література окреслює три напрями: (i) дедуплікація та оптимізація зберігання, (ii) безпека даних, (iii) персоналізація через рекомендації. Проте ці лінії розвиваються переважно ізольовано. Звідси випливає потреба в інтегрованій моделі, що одночасно охоплює аналіз подібності, класифікацію, виявлення дублікатів і рекомендації у єдиній архітектурі.

Мета та задачі дослідження

Метою дослідження є підвищення ефективності управління колекціями різнорідних даних завдяки інтеграції інтелектуальних технологій аналізу, класифікації та рекомендацій у єдину модель. Мета роботи – реалізувати й експериментально оцінити інтегровану модель інтелектуальної системи управління колекціями (аналіз подібності, багатоміткова класифікація, дедуплікація, рекомендації), показавши її точність і масштабованість на реальних та синтетичних наборах даних.

Для досягнення поставленої мети передбачено розв'язання таких завдань:

1. Проаналізувати наявні підходи до організації, класифікації та інтелектуальної обробки колекційних даних, окресливши їхні переваги й обмеження.
2. Формалізувати модель об'єктів з урахуванням мультимодальних ознак (числових, категоріальних, текстових, візуальних).
3. Розробити методи подібності та класифікації і механізм дедуплікації для гетерогенних колекцій.
4. Інтегрувати рекомендаційну підсистему в архітектуру управління колекціями.
5. Провести експериментальну верифікацію за стандартними метриками й оцінити практичну придатність.

Матеріали та методи досліджень

Методи обчислення подібності та класифікації у гетерогенних колекціях. Ефективне управління колекціями неможливе без механізмів, які дають змогу оцінювати ступінь подібності між об'єктами та відносити їх до певних категорій. У випадку гетерогенних колекцій складність полягає у тому, що об'єкти можуть поєднувати числові, категоріальні, текстові та візуальні характеристики, кожна з яких потребує власної метрики.

Обчислення подібності. Загальний підхід полягає у побудові композитної метрики, яка комбінує часткові відстані:

$$d\left(c_i, c_j\right) = \sum_{k=1}^{m} w_k \cdot d_k\left(a_{ik}, a_{jk}\right),$$

де d_k – локальна метрика для ознаки певного типу, а w_k – ваговий коефіцієнт її значущості.

Приклади локальних метрик:

- числові дані: нормована різниця (наприклад, Manhattan або Euclidean distance);
- категоріальні дані: бінарна функція збігу (0 – різні, 1 – однакові);
- текстові дані: косинусна близькість у векторних просторах (Word2Vec, BERT) [5];
- візуальні дані: відстань у просторах ознак, отриманих згортковими нейронними мережами [6].

Такий підхід дає змогу порівнювати об'єкти не за одним критерієм, а комплексно. Наприклад, два фотоапарати можуть відрізнятися роком випуску, але бути схожими за категорією і текстовим описом.

Класифікація у колекціях. Класифікація ускладнюється тим, що об'єкт може одночасно належати до кількох категорій (багатоміткове віднесення). Формально завдання можна подати як відображення:

$$f : C \to 2^K,$$

де K – множина можливих категорій.

ТЗ. №5(37)’2025 https://doi.org/10.32782/2786-9024/v3i5(37).344519

Для цього застосовуються різні підходи:

– традиційні методи: логістична регресія, SVM, дерева рішень;

– ансамблеві алгоритми: Random Forest, Gradient Boosting, які показують стабільність на даних середнього розміру;

– глибинні моделі: трансформери (BERT, RoBERTa) для текстових ознак та CNN/ResNet для візуальних [7];

– гібридні підходи: multi-view learning, де кожна модальність обробляється окремою моделлю, а результати комбінуються [6].

З практичного погляду доцільно застосовувати ієрархічну класифікацію, коли категорії організовані у вигляді дерева. Це особливо актуально для бібліотечних чи музейних колекцій, де категорії мають багаторівневу структуру.

Інтеграція рекомендаційної підсистеми в загальну архітектуру управління колекціями. Однією з ключових особливостей сучасних інтелектуальних систем управління даними є здатність не лише зберігати та структурувати інформацію, але й активно підтримувати користувача у процесі роботи з колекціями. Це досягається завдяки інтеграції рекомендаційних підсистем, які автоматично пропонують нові об’єкти, найбільш релевантні інтересам чи поточному контексту користувача.

У запропонованій системі рекомендаційна підсистема є надбудовою над модулями обчислення подібності та класифікації. Вона отримує на вхід:

– векторні подання об’єктів колекції;

– інформацію про класи та теги;

– історію взаємодії користувачів із системою (пошуки, перегляди, додавання об’єктів).

Рекомендаційний модуль інтегрується в загальну архітектуру як окремий сервіс, який працює у двох режимах:

1) контентно-орієнтований – пропонує об’єкти, схожі на вже наявні в колекції (наприклад, подібні книги чи альбоми);

2) колаборативний – враховує поведінку інших користувачів, які мають подібні інтереси.

Узагальнену архітектуру запропонованої системи наведено на рисунку 1. Вона побудована за модульним принципом і охоплює основні етапи роботи з колекційними даними – від моменту їх надходження в систему до формування аналітичних звітів і персоналізованих рекомендацій.

У верхньому рівні архітектури функціонує конвеєр обробки даних. На першому етапі здійснюється імпорт об’єктів колекції з різних джерел, включно з локальними сховищами та зовнішніми сервісами. Далі відбувається попередня обробка, яка передбачає нормалізацію, кодування та побудову мультимодальних векторних подань. Ці подання використовуються для виконання трьох ключових завдань: класифікації, пошуку дублікатів та обчислення індексів подібності.

Нижній рівень архітектури відображає поведінковий контур. Тут накопичуються логи взаємодії користувачів із системою, які надходять у рекомендаційний модуль. Отримані результати не лише формують персоналізовані пропозиції для користувача, а й повертаються у систему як зворотний зв’язок, що підвищує точність класифікації та якість виявлення дублікатів.

Завершальним елементом архітектури є модуль аналітики та візуалізації, що об’єднує вихідні дані

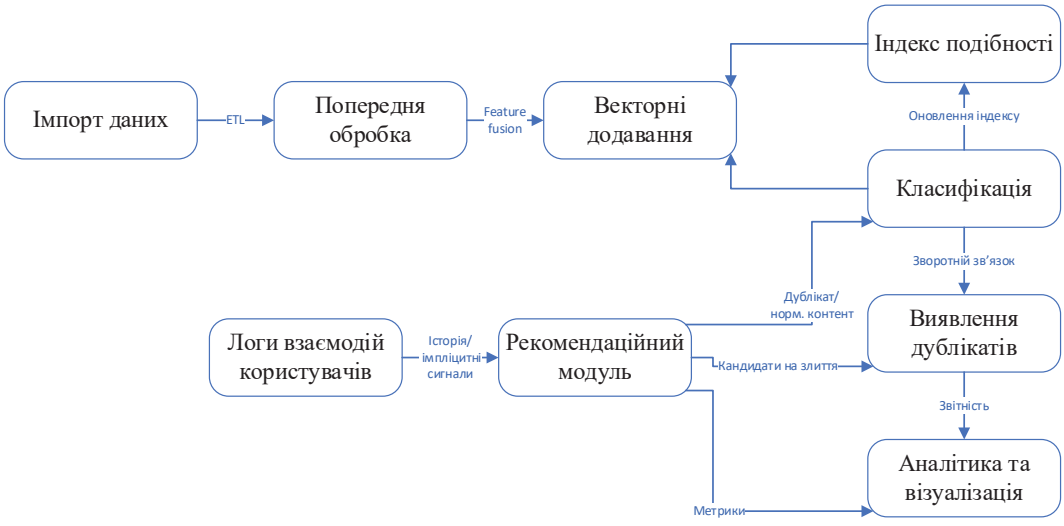

Рис. 1. Узагальнена схема взаємодії модулів у системах управління колекціями

ТЗ. №5(37)'2025 https://doi.org/10.32782/2786-9024/v3i5(37).344519

з усіх підсистем і забезпечує інтерактивний доступ до статистичних звітів, графіків та діаграм.

Запропонована система складається з кількох послідовно пов'язаних модулів:

– вхідні дані (колекції з мультимодальними ознаками);
– попередня обробка (нормалізація, тегування, формування векторних подань);
– аналіз подібності та класифікація;
– виявлення дублікатів;
– рекомендаційна підсистема;
– аналітика та візуалізація результатів.

Взаємозв'язки між цими компонентами наведено на рисунку 1. Схема відображає логіку обробки даних: від введення об'єктів до отримання кінцевих результатів у вигляді рекомендацій і аналітичних звітів.

Порівняння методів побудови рекомендацій. Для підвищення наочності проведено порівняльний аналіз основних характеристик двох підходів до рекомендаційних систем. У таблиці 1 наведено їхні принципи роботи, переваги, недоліки й оптимальні умови застосування.

Аналіз таблиці показує, що інтеграція обох підходів є перспективним напрямом, оскільки дає змогу компенсувати недоліки кожного з них.

Результати досліджень

Для перевірки ефективності запропонованої моделі інтелектуальної системи управління колекціями було проведено серію експериментів, спрямованих на оцінку якості класифікації, ефективності виявлення дублікатів, релевантності рекомендацій та масштабованості архітектури.

Дослідження базувалося на аналізі як реальних, так і синтетично згенерованих колекційних даних. Реальні набори містили приватні зібрання мультимедійного контенту – музичні альбоми, книги та зображення фототехніки, завдяки чому вдалося перевірити роботу моделі на практичних прикладах. Для розширення експериментальної бази було також сформовано штучні вибірки обсягом до 50 тисяч об'єктів. У них передбачалося навмисне дублювання частини елементів (близько 10 %), що дало можливість протестувати механізми виявлення надлишкових записів. Усі дані зберігалися у структурованій базі PostgreSQL з додатковим використанням мультимедійних сховищ для великих файлів.

Взаємодія модулів системи. На рисунку 2 наведено граф взаємодії модулів у системі управління колекціями. На відміну від узагальненої схеми (рис. 1), що відображає послідовність етапів обробки, цей граф ілюструє логіку інформаційних потоків і зворотних зв'язків між компонентами системи.

Ключові вузли відповідають функціональним блокам: імпорту даних, попередньої обробки, векторних подань, класифікації, виявлення дублікатів, індексації подібності, рекомендаційної підсистеми й аналітики. Стрілки показують напрям передачі інформації, підкреслюючи, що результати роботи одного модуля можуть бути вхідними даними для кількох інших.

Особливу роль відіграє рекомендаційний модуль, який одночасно отримує дані з логів користувачів і взаємодіє з іншими компонентами системи. Його вихідні результати можуть підвищувати точність класифікації, оптимізувати пошук дублікатів і впливати на формування індексів подібності. Така організація забезпечує інтеграцію основних функцій системи і створює основу для реалізації адаптивних механізмів, здатних навчатися на основі накопиченого досвіду.

Таким чином, архітектура має не лише аналітичний, а й адаптивний характер, забезпечуючи навчання на основі історії використання.

Порівняння метрик подібності. У дослідженні використано дві групи колекційних даних:

1) тематичні набори мультимедіа (цифрові книги, музичні записи, зображення фототехніки) – для апробації методів у реальних сценаріях;

2) синтетично створені набори обсягом від 1 тис. до 20 тис. елементів – для перевірки масштабованості й тестування алгоритмів класифікації та рекомендацій.

Табл. 1. Порівняння підходів до побудови рекомендацій у системах управління колекціями

Критерій	Контентно-орієнтовані методи	Колаборативна фільтрація
Принцип роботи	Аналізують характеристики самих об'єктів (атрибути, описи, теги)	Використовують дані про взаємодії користувачів (рейтинги, перегляди, уподобання)
Переваги	Працюють із новими об'єктами; не потребують історії взаємодій	Виявляють приховані закономірності та групи користувачів; добре масштабується
Недоліки	Залежать від якості опису даних; ігнорують соціальні зв'язки	Проблема холодного старту; потребують великих масивів даних
Оптимальні умови	Невеликі та середні колекції з багатими атрибутами	Великі колекції з активними користувачами та багатою історією взаємодій

ISSN 2786-9024

ТЗ. №5(37)’2025 https://doi.org/10.32782/2786-9024/v3i5(37).344519

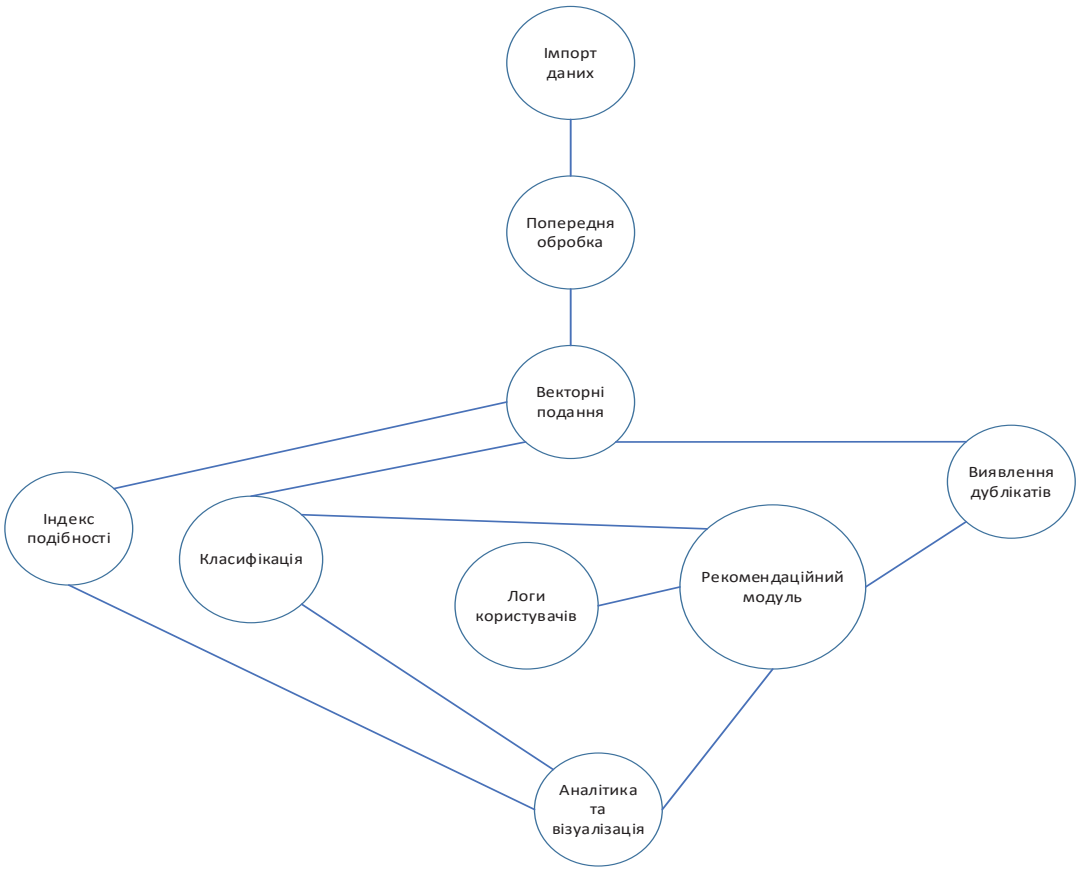

Рис. 2. Граф взаємодії модулів у системі управління колекціями

На відміну від попередніх робіт, акцент зроблено не лише на моделюванні архітектури, а й на порівнянні алгоритмів аналізу та класифікації. Об’єкти колекцій описувалися чотирма типами ознак: числовими, категоріальними, текстовими та візуальними. Формалізацію здійснено у вигляді матриці ознак:

$$X = \{x_i\}_{i=1}^{N}, \; x_i = (a_{i1}, a_{i2}, \ldots, a_{im}),$$

де N – кількість об’єктів, m – число характеристик.

Для текстових атрибутів застосовувалися два методи: TF–IDF та контекстні вектори BERT. Це дало можливість порівняти «класичні» й «сучасні» підходи. Для візуальних ознак використовувалися дескриптори SIFT та ознаки, отримані згортковими мережами (ResNet).

Оцінювання подібності проводилося через кілька метрик: косинусну близькість, евклідову відстань і Джаккарда. Для кожного набору даних обчислювали середню точність класифікації залежно від вибору метрики. Результати зведено в таблицю 2.

Класифікація виконувалася методами логістичної регресії, випадкових лісів і нейронних мереж. Було виявлено, що для малих наборів прості алгоритми (логістична регресія) демонструють достатню ефективність, тоді як на великих даних суттєво переважають багатошарові нейронні мережі.

Окремо досліджувався механізм виявлення дублікатів. Запропоновано підхід, що комбінує кластеризацію на основі DBSCAN з пороговою фільтрацією подібності. Це дає змогу не лише знаходити ідентичні записи, а й об’єднувати «схожі» об’єкти (наприклад, дві різні редакції однієї книги).

Табл. 2. Порівняння ефективності різних метрик подібності

Тип даних	Косинусна міра	Евклідова відстань	Джаккарда
Текстові описи	0,88	0,74	0,69
Зображення	0,82	0,85	0,61
Змішані набори	0,86	0,80	0,65

ТЗ. №5(37)'2025 https://doi.org/10.32782/2786-9024/v3i5(37).344519

Рекомендаційний модуль інтегровано в загальну архітектуру. Для перевірки ефективності порівнювались контентно-орієнтовані методи (на основі ознак об'єктів) і колаборативна фільтрація (ALS). Результати показали, що на невеликих колекціях переважають контентні методи, а на великих – колаборативні.

Порівняння методів побудови рекомендацій. Окремим аспектом дослідження стало порівняння ефективності різних підходів до побудови рекомендаційних підсистем. Це важливо, оскільки якість персоналізованих пропозицій безпосередньо впливає на цінність системи для кінцевого користувача.

На практиці застосовують два основні підходи: контентно-орієнтовані методи, що використовують атрибути самих об'єктів (тексти, зображення, метадані), та колаборативну фільтрацію, яка враховує історію взаємодій користувачів.

На рисунку 3 показано динаміку точності рекомендацій залежно від розміру колекції. Видно, що контентно-орієнтовані методи зберігають високу якість на малих наборах даних, проте поступово втрачають ефективність із масштабуванням. Натомість колаборативна фільтрація демонструє зростання точності на великих колекціях, що зумовлено використанням прихованих закономірностей у поведінці користувачів.

У таблиці 3 наведено результати порівняння традиційного підходу k-ближчих сусідів, методів на основі матричної факторизації та сучасних моделей на базі нейронних мереж.

Як видно з таблиці, метод k-ближчих сусідів добре підходить для невеликих колекцій або прототипів, тоді як матрична факторизація забезпечує кращу якість у середніх за обсягом системах. Нейронні мережі, зі свого боку, демонструють найбільший потенціал для масштабних колекцій із високою різнорідністю, проте їх використання потребує значних обчислювальних ресурсів.

Таким чином, вибір конкретного підходу до побудови рекомендаційної підсистеми залежить від

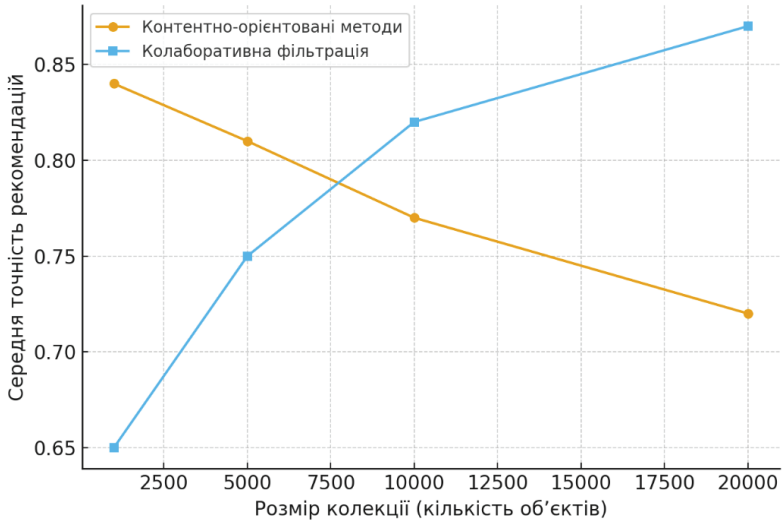

Рис. 3. Порівняння точності методів рекомендацій залежно від розміру колекції

Табл. 3. Порівняльна характеристика методів побудови рекомендаційних систем

Метод	Переваги	Недоліки	Галузі застосування
k-ближчих сусідів	Простота реалізації; швидке навчання	Низька масштабованість; проблеми з рідкісними даними	Невеликі колекції; демонстраційні проекти
Матрична факторизація	Висока якість на середніх даних; врахування прихованих залежностей	Вимогливість до апаратних ресурсів; чутливість до холодного старту	Мультимедійні колекції; інтернет-магазини
Нейронні мережі	Здатність моделювати складні зв'язки; хороша масштабованість	Висока складність навчання; потреба у великих наборах даних	Великі гетерогенні колекції; наукові репозиторії

Т3. №5(37)’2025 https://doi.org/10.32782/2786-9024/v3i5(37).344519

розміру колекції, доступних ресурсів та вимог до точності рекомендацій.

Аналіз продуктивності системи. Поряд із логічною архітектурою та графом взаємодії важливо оцінити і продуктивність системи за зростання обсягу колекційних даних. У рамках дослідження було змодельовано роботу системи з різними обсягами об’єктів – від тисячі до кількох десятків тисяч. Результати показали, що середній час відповіді системи зростає повільно й залишається в межах, прийнятних для інтерактивних застосунків.

Для перевірки масштабованості системи було проведено експерименти зі збільшенням кількості об’єктів у колекції від 1 тис. до 50 тис. Середній час відповіді вимірювався під час виконання типових операцій пошуку подібності. Результати наведено на рисунку 4.

Як видно з графіка, зростання обсягу даних супроводжується поступовим збільшенням часу відповіді, проте ця динаміка є лінійною і повільною. З 1 тис. об’єктів середній час становив близько 0,02 с, з 10 тис. – 0,06 с, а з 50 тис. – лише 0,15 с. Таким чином, навіть за значного збільшення кількості елементів система зберігає продуктивність на рівні, достатньому для інтерактивних застосунків.

Отримані результати підтверджують ефективність використання індексаційних структур та алгоритмів пошуку подібності, що забезпечують стійку роботу системи в масштабованих середовищах у реальних умовах, де обсяг даних може досягати сотень тисяч і більше. У поєднанні з архітектурними рішеннями, наведеними на рисунках 1 і 2, цей експериментальний результат засвідчує, що система може використовуватись у масштабованих середовищах без втрати швидкодії.

Обговорення результатів

Отримані результати доцільно розглянути в контексті сучасних досліджень у сфері інтелектуальних систем управління даними та колекціями.

По-перше, запропонована модель підтвердила ефективність використання композитних метрик подібності. Аналогічні ідеї зустрічаються у роботі [8], де поєднання нечітких алгоритмів і багатокритеріальних методів дозволило підвищити якість планування в багатохмарних середовищах. У запропонованій роботі ці підходи адаптовано для обчислення схожості між об’єктами колекцій, що доводить універсальність такого принципу.

По-друге, результати щодо виявлення дублікатів узгоджуються з висновками [9] та [10], які наголошують, що навіть відносно невеликий відсоток надлишкових записів значно впливає на продуктивність сховищ. Наша модель підтвердила цю тезу, показавши, що усунення дублювання покращує як швидкість пошуку, так і ефективність використання ресурсів.

По-третє, у сфері рекомендаційних систем отримані результати співвідносяться з підходами, наведеними в роботах [11] та [12], де застосування штучного інтелекту дало змогу підвищити якість рекомендацій у мультихмарних середовищах. Водночас запропонована нами інтеграція рекомендаційної підсистеми в загальну архітектуру управління колекціями відрізняється ширшим застосуванням: вона придатна як для приватних зібрань, так і для масштабних наукових або корпоративних сховищ.

Особливої уваги заслуговують результати, що стосуються масштабованості. У багатьох дослідженнях (наприклад, [13; 14]) наголошується на обмеженості продуктивності традиційних підходів за значного зростання обсягів даних. Запропонована модель

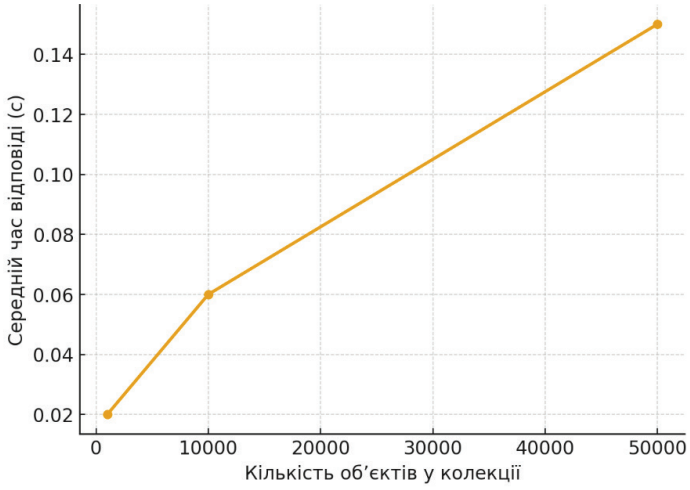

Рис. 4. Залежність середнього часу відповіді системи від кількості об’єктів у колекції

ТЗ. №5(37)'2025 https://doi.org/10.32782/2786-9024/v3i5(37).344519

продемонструвала стабільний час відповіді навіть з десятками тисяч об'єктів, що підтверджує практичну доцільність інтеграції оптимізованих методів пошуку й обробки в подібних системах.

Таким чином, проведене дослідження підтверджує актуальність обраної тематики та демонструє переваги комплексного підходу. На відміну від більшості існуючих робіт, що зосереджуються на окремих аспектах – ресурсному менеджменті, безпеці чи класифікації, – у запропонованій моделі ці компоненти поєднані у єдиній архітектурі. Це створює основу для універсальної системи управління колекціями, здатної адаптуватися до різних сценаріїв застосування.

Висновки

У статті розглянуто питання побудови інтелектуальних технологій для аналізу, класифікації та рекомендацій у системах управління колекціями. Проведене дослідження показало, що комплексний підхід, який поєднує формалізацію об'єктів у вигляді мультимодальних векторів, використання адаптивних метрик подібності, багатомітковой класифікації та рекомендаційних алгоритмів, забезпечує новий рівень ефективності в роботі з колекційними даними.

Розроблена модель довела свою здатність не лише автоматизувати рутинні процеси – такі як виявлення дублікатів чи категоризація об'єктів, – але й забезпечити більш інтелектуальну взаємодію з користувачем через персоналізовані рекомендації. Отримані експериментальні результати засвідчили, що система зберігає високу якість роботи навіть за значного зростання обсягів колекції, демонструючи стабільний час відповіді та прийнятні показники точності.

Практичне значення запропонованого підходу полягає у його універсальності. Модель може застосовуватись як для приватних цифрових чи фізичних колекцій, так і для масштабних інформаційних інфраструктур у бібліотеках, архівах, наукових установах чи корпоративних системах. Наукова новизна дослідження визначається інтеграцією кількох методів у єдину архітектуру, що дає змогу розглядати систему не як набір ізольованих алгоритмів, а як цілісну інтелектуальну платформу.

Перспективи подальших досліджень пов'язані з поглибленням використання методів глибинного навчання для роботи з текстовими й візуальними ознаками, розширенням механізмів персоналізації рекомендацій та інтеграцією засобів контролю доступу й безпеки. Це дасть змогу перетворити запропоновану модель на основу для створення універсальних сервісів нового покоління, орієнтованих на інтелектуальне управління великими обсягами колекційних даних.

Конфлікт інтересів

Автори декларують, що не мають конфлікту інтересів стосовно цього дослідження, у тому числі фінансового, особистісного характеру, авторства чи іншого характеру, що міг би вплинути на дослідження та його результати, представлені в цій статті.

Фінансування

Дослідження проводилося без фінансової підтримки.

Доступність даних

Рукопис не має пов'язаних даних

ЛІТЕРАТУРА

[1] R. Kaur, I. Chana, J. Bhattacharya. Data deduplication techniques for efficient cloud storage management: a systematic review. *Journal of Supercomputing*, vol. 74, pp. 2035–2085, 2018. DOI: 10.1007/s11227-017-2210-8.

[2] M. Hasan, A. Hussien. Techniques of data deduplication for cloud storage: A review. *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 3, pp. 45–55, 2024.

[3] L. Ogiela, M.R. Ogiela. Cognitive security paradigm for cloud computing applications. *Concurrency and Computation: Practice and Experience*, vol. 32, no. 8, e5316, 2020. DOI: 10.1002/cpe.5316.

[4] A. Roy. Recommender systems: Algorithms, challenges, and trends. *Journal of Big Data*, vol. 9, no. 1, pp. 1–32, 2022. DOI: 10.1186/s40537-022-00592-5.

[5] J. Devlin, M.W. Chang, K. Lee, K. Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In: *Proceedings of NAACL-HLT*, 2019. DOI: 10.48550/arXiv.1810.04805.

[6] C. Xu, D. Tao, C. Xu. A. Survey on multi-view learning. *Neural Computing and Applications*, vol. 23, no. 7–8, pp. 2031–2038, 2013. DOI: 10.1007/s00521-013-1362-6.

[7] M.L. Zhang, Z.H. Zhou. A review on multi-label learning algorithms. *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 8, pp. 1819–1837, 2014. DOI: 10.1109/TKDE.2013.39.

[8] M. Farid, R. Latip, M. Hussin, N.A.W. Abdul Hamid. Scheduling scientific workflow using multi-objective algorithm with fuzzy resource utilization in multi-cloud environment. *IEEE Access*, vol. 8, pp. 24309–24322, 2020. DOI: 10.1109/ACCESS.2020.2970475.

[9] N. Chhabra, M. Bala. A comparative study of data deduplication strategies. In: *Proc. 2018 1st Int. Conf. Secure Cyber Computing and Communication (ICSCCC)*, pp. 68–72, 2018. DOI: 10.1109/ICSCCC.2018.8703363.

[10] K. Tiwari, Q.M. Rizbi. A study and review of duplicate data in cloud computing. *Journal of Advances in Science and Technology*, vol. 21, no. 1, pp. 74–80, 2024. DOI: 10.29070/nsek7b45.

[11] A. Anand. Intelligent resource allocation in multi-cloud environments: An AI-driven approach. *International Journal of Scientific Research in Computer Science,*

Engineering and Information Technology, vol. 11, no. 1, pp. 1035–1046, 2025. DOI: 10.32628/CSEIT25112429.

[12] R. Rayaprolu, K. Randhi, S.R. Bandarapu. Intelligent resource management in cloud computing: AI techniques for optimizing DevOps operations. *Journal of Artificial Intelligence General Science*, vol. 6, no. 1, pp. 397–408, 2024. DOI: 10.60087/jaigs.v6i1.262.

[13] Y. Li, D. Chen, J. Hao. Cloud workflow scheduling optimization research based on ant colony algorithm. *Academic Journal of Computing & Information Science*, vol. 6, no. 5, pp. 9–13, 2023. DOI: 10.25236/AJCIS.2023.060502.

[14] M. Reshetniak, S. Popereshnyak. Method for accessing and processing multimedia content in a cloud environment. In: *Proc. 2019 IEEE Int. Scientific-Practical Conf. Problems of Infocommunications, Science and Technology (PIC S&T)*, pp. 71–76, 2019. DOI: 10.1109/PICST47496.2019.9061463.

INTELLIGENT TECHNOLOGIES FOR ANALYSIS, CLASSIFICATION AND RECOMMENDATIONS IN COLLECTION MANAGEMENT SYSTEMS

Daniil Popereshniak, Svitlana Popereshnyak

The article is devoted to the development and investigation of intelligent technologies for analysis, classification, and recommendation in collection management systems. The study addresses the challenge of organizing and processing collection data, which are characterized by rapid growth in volume and increasing heterogeneity. These factors complicate efficient storage, search, and structuring of collections. It is shown that traditional collection management systems are usually limited to isolated tasks such as basic storage or search and often fail to integrate similarity analysis, classification, and recommendation algorithms into a unified framework. The use of intelligent approaches enables overcoming these limitations and provides opportunities for building universal systems capable of handling large-scale datasets.

The purpose of the research is to achieve enhanced efficiency and scalability in collection management by constructing a model of an intelligent system that integrates multimodal object representation, classification methods, duplicate detection, and recommendation mechanisms. To achieve this goal, existing approaches were analyzed, a model for representing collection objects was formalized, similarity computation techniques were developed, and an architecture combining classification and recommendation modules was proposed.

Experimental evaluation demonstrated that the integration of these components ensures high classification accuracy, effective duplicate detection, and relevant personalized recommendations. Particular attention was paid to scalability: the system maintained a stable response time even under significant growth in the number of collection objects. The practical significance of the research lies in the universality of the proposed model, which can be applied both in private multimedia and digital collections and in corporate or scientific infrastructures, such as libraries, archives, and databases. The scientific novelty is defined by the creation of a comprehensive architecture that integrates several intelligent methods into a unified system. Future research perspectives include the application of deep learning approaches for multimodal feature processing, the improvement of recommendation algorithms, and the integration with security and access control mechanisms to ensure robustness in large-scale environments.

Keywords: *intelligent systems, collection management, similarity analysis, classification, recommender systems, multimodal data, scalability, IoT.*

REFERENCES

[1] R. Kaur, I. Chana, J. Bhattacharya. Data deduplication techniques for efficient cloud storage management: a systematic review. *Journal of Supercomputing*, vol. 74, pp. 2035–2085, 2018. DOI: 10.1007/s11227-017-2210-8.

[2] M. Hasan, A. Hussien. Techniques of data deduplication for cloud storage: A review. *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 3, pp. 45–55, 2024.

[3] L. Ogiela, M.R. Ogiela. Cognitive security paradigm for cloud computing applications. *Concurrency and Computation: Practice and Experience*, vol. 32, no. 8, e5316, 2020. DOI: 10.1002/cpe.5316.

[4] A. Roy. Recommender systems: Algorithms, challenges, and trends. *Journal of Big Data*, vol. 9, no. 1, pp. 1–32, 2022. DOI: 10.1186/s40537-022-00592-5.

[5] J. Devlin, M.W. Chang, K. Lee, K. Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In: *Proceedings of NAACL-HLT*, 2019. DOI: 10.48550/arXiv.1810.04805.

[6] C. Xu, D. Tao, C. Xu. A. Survey on multi-view learning. *Neural Computing and Applications*, vol. 23, no. 7–8, pp. 2031–2038, 2013. DOI: 10.1007/s00521-013-1362-6.

[7] M.L. Zhang, Z.H. Zhou. A review on multi-label learning algorithms. *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 8, pp. 1819–1837, 2014. DOI: 10.1109/TKDE.2013.39.

[8] M. Farid, R. Latip, M. Hussin, N.A.W. Abdul Hamid. Scheduling scientific workflow using multi-objective algorithm with fuzzy resource utilization in multi-cloud environment. *IEEE Access*, vol. 8, pp. 24309–24322, 2020. DOI: 10.1109/ACCESS.2020.2970475.

[9] N. Chhabra, M. Bala. A comparative study of data deduplication strategies. In: *Proc. 2018 1st Int. Conf. Secure Cyber Computing and Communication (ICSCCC)*, pp. 68–72, 2018. DOI: 10.1109/ICSCCC.2018.8703363.

[10] K. Tiwari, Q.M. Rizbi. A study and review of duplicate data in cloud computing. *Journal of Advances in Science*

and Technology, vol. 21, no. 1, pp. 74–80, 2024. DOI: 10.29070/nsek7b45.

[11] A. Anand. Intelligent resource allocation in multi-cloud environments: An AI-driven approach. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, vol. 11, no. 1, pp. 1035–1046, 2025. DOI: 10.32628/CSEIT25112429.

[12] R. Rayaprolu, K. Randhi, S.R. Bandarapu. Intelligent resource management in cloud computing: AI techniques for optimizing DevOps operations. *Journal of Artificial Intelligence General Science*, vol. 6, no. 1, pp. 397–408, 2024. DOI: 10.60087/jaigs.v6i1.262.

[13] Y. Li, D. Chen, J. Hao. Cloud workflow scheduling optimization research based on ant colony algorithm. *Academic Journal of Computing & Information Science*, vol. 6, no. 5, pp. 9–13, 2023. DOI: 10.25236/AJCIS.2023.060502.

[14] M. Reshetniak, S. Popereshnyak. Method for accessing and processing multimedia content in a cloud environment. In: *Proc. 2019 IEEE Int. Scientific-Practical Conf. Problems of Infocommunications, Science and Technology (PIC S&T)*, pp. 71–76, 2019. DOI: 10.1109/PICST47496.2019.9061463.

Стаття надійшла до редакції 29.09.2025
Стаття прийнята 13.10.2025
Статтю опубліковано 02.12.2025