

ВІДГУК

рецензента Волка Максима Олександровича
на дисертаційну роботу Даниленка Станіслава Дмитровича
на тему «Моделі та методи пошуку зображень у сховищах даних за векторним
представленням ознак»,
подану на здобуття наукового ступеня доктора філософії
з галузі знань 12 Інформаційні технології
за спеціальністю 121 Інженерія програмного забезпечення

1. Актуальність теми дисертації

Актуальність дисертаційної роботи зумовлена тим, що у сховищах великих даних для пошуку зображень використання підходів, що ґрунтуються на текстових описах або метаданих, вже не є достатнім. Замість цього використовується пошук за візуальними ознаками, представленими у вигляді векторних дескрипторів. Разом із тим, зі зростанням кількості зображень, розмірності дескрипторів та інтенсивності пошукових запитів виникають суттєві інженерні проблеми: збільшується час формування пошукових індексів, ускладнюється оновлення структури сховища дескрипторів, зростають вимоги до обчислювальних ресурсів, а наявні рішення не завжди забезпечують достатню адаптивність до різних прикладних задач.

Дисертаційна робота Даниленка С. Д. спрямована на розв'язання зазначених проблем шляхом розроблення моделей та методів побудови CBIR-систем. Вважаю, що робота має високий ступінь наукової та практичної значущості, а її тема є актуальною для розвитку інженерії програмного забезпечення, зокрема в частині проектування ефективних, масштабованих і конфігурованих CBIR-систем.

2. Оцінка обґрунтованості наукових результатів дисертації, їх достовірності та новизни

У дисертації чітко сформульовано проблему дослідження, визначено об'єкт, предмет, мету та завдання, що відповідають обраній темі. Здобувач застосував сучасний методичний апарат, зокрема методи системного аналізу та декомпозиції, теоретико-множинні та математичні методи, методи чисельної оптимізації, описової статистики, аналізу даних, а також експериментальні та порівняльні методи. Це дозволило розглянути задачу не лише з алгоритмічної, але й з інженерного погляду.

Наукові положення дисертації мають належний рівень обґрунтованості. Робота логічно побудована: спочатку виконано аналіз стану предметної галузі, далі сформульовано основні положення та гіпотези дослідження, після чого запропоновано відповідні моделі й методи, наведено їх програмну реалізацію та виконано експериментальну перевірку.

Основні наукові результати, що визначають новизну дисертаційної роботи, полягають у такому:

1. Вперше запропоновано модель розроблення програмного забезпечення пошуку зображень на основі вмісту у сховищах даних у вигляді багатовимірного куба, яка реалізує організацію векторних представлень ознак зображень шляхом розбиття простору ознак на підпростори з дискретизацією їх значень. Це зменшує обчислювальні витрати на етапі формування внутрішньої структури кластерів і пошукових індексів за рахунок уникнення складних ітеративних процесів та дозволяє, залежно від кількості та розмірності дескрипторів, пришвидшити цей процес від 4 до 22 разів порівняно з існуючими рішеннями у зіставних умовах.

2. Вперше запропоновано метод розроблення програмного забезпечення пошуку зображень на основі вмісту, що ґрунтується на хвильовому пошуку та реалізує поступове розширення області пошуку шляхом аналізу сусідніх

підпросторів. Використання структурної організації моделі зменшує потребу в додаткових обчисленнях для визначення належності елемента до кластера та дозволяє скоротити час виконання пошуку від 1,27 до 32 разів порівняно з існуючими методами у зіставних умовах із можливістю адаптації параметрів пошуку до характеристик дескрипторів та вимог якості за F1-мірою.

3. Отримав подальший розвиток метод розроблення програмного забезпечення пошуку зображень на основі вмісту завдяки введенню механізму, що поєднує статистично обґрунтоване коригування меж підпросторів та чисельну оптимізацію параметрів моделі. Це забезпечує формування кластерів з квазірівномірною наповненістю і, порівнюючи з рівномірно визначеними межами підпросторів, дозволяє збільшити кількість наповнених кластерів від 4,4 до 6,6 разів, кількість кластерів з очікуваною наповненістю – від 4,8 до 9,5 разів, підвищує рівномірність розподілу дескрипторів між кластерами та створює передумови для стабілізації трудомісткості ітерацій пошуку й показників якості пошуку.

4. Удосконалено уніфіковану модель проектування та оцінки якості програмного забезпечення пошуку зображень на основі вмісту у сховищах даних, яка базується на компонентно-орієнтованому підході та формалізованому описі конфігурацій системи. На відміну від відомих рішень, запропонована модель описує систему як простір типізованих компонентів із визначеними контрактами взаємодії. Це забезпечує ширше покриття операцій адаптації системи до прикладних задач, дає змогу формувати працездатні конфігурації без модифікації архітектурного ядра та порівнювати вплив компонентів на показники функціонування системи за вибраними метриками на ранніх етапах розроблення.

Достовірність результатів підтверджується проведенням експериментальних досліджень із використанням різних наборів даних і типів дескрипторів, а також порівнянням запропонованих рішень з альтернативними підходами.

Практична цінність дослідження підтверджується впровадженням результатів у приватному підприємстві «РІІТ», а також використанням отриманих результатів у межах держдоговірної теми. Запропоновані моделі й методи можуть бути використані для розроблення програмного забезпечення пошуку зображень на основі вмісту з конфігурацією, адаптованою до вимог конкретної предметної галузі.

Таким чином, дисертаційна робота вирізняється належним ступенем обґрунтованості, достовірності та новизни наукових результатів. Поставлене науково-практичне завдання виконано повністю, а здобувач продемонстрував володіння методологією наукової діяльності у сфері інженерії програмного забезпечення.

3. Оцінка змісту дисертації, її завершеність та дотримання принципів академічної доброчесності

За своїм змістом дисертаційна робота Даниленка С. Д. відповідає спеціальності 121 Інженерія програмного забезпечення та напрямам досліджень відповідно до освітньо-наукової програми у Харківському національному університеті радіоелектроніки.

Дисертаційна робота є завершеною науковою працею і свідчить про наявність особистого внеску здобувача у розвиток теорії та практики в галузі інформаційних технологій за спеціальністю 121 Інженерія програмного забезпечення.

Дисертація складається з анотацій українською та англійською мовами, вступу, чотирьох розділів, висновків, списку використаних джерел і додатків. Загальний обсяг роботи становить 189 сторінок, з яких 145 сторінок основного тексту. Робота містить 25 рисунків, 21 таблицю, список використаних джерел із 135 найменувань та 5 додатків.

Аналіз змісту дисертаційної роботи та публікацій здобувача дає підстави вважати, що робота є результатом самостійного наукового дослідження. Використані ідеї, результати й тексти інших авторів мають належні посилання на відповідні джерела. Ознак академічного плагіату, фальсифікації, фабрикації чи компіляції у роботі не встановлено.

4. Мова та стиль викладення результатів

Дисертаційна робота написана українською мовою з дотриманням вимог наукового стилю викладення. Тексту притаманні послідовність, логічність і зрозумілість. Сформульована проблема дослідження подана чітко, а зв'язок між метою, завданнями, гіпотезами, моделями, методами та експериментальною перевіркою простежується протягом усієї роботи.

Автор використовує загальноприйняту науково-технічну термінологію у сфері пошуку зображень, векторних дескрипторів, індексації, кластеризації, програмної архітектури та інженерії програмного забезпечення. У роботі наведено перелік скорочень, умовних позначень, одиниць і термінів, що полегшує сприйняття матеріалу.

Дисертація містить достатню кількість рисунків, таблиць, UML-діаграм і результатів експериментів, які ілюструють основні положення роботи та допомагають зрозуміти архітектуру запропонованих рішень, структуру моделей, хід експериментів і отримані кількісні результати.

Загальні висновки узагальнюють результати дослідження, відображають досягнення поставленої мети та відповідають змісту розділів дисертації. Дисертаційна робота оформлена відповідно до вимог наказу Міністерства освіти і науки України від 12 січня 2017 р. № 40 «Про затвердження вимог до оформлення дисертацій».

5. Оприлюднення результатів дисертаційної роботи

Основні результати дисертаційного дослідження достатньо повно відображені у 8 наукових публікаціях здобувача. Було опубліковано 6 статей у наукових журналах, включених до Переліку наукових фахових видань України, з яких 3 – у виданнях, що входять до наукометричної бази Scopus. Також опубліковано 2 тези доповідей у матеріалах міжнародних наукових конференцій: 2022 IEEE 9th International Conference on Problems of Infocommunications, Science and Technology та MoDaST 2025: Modern Data Science Technologies Doctoral Consortium, що входять до наукометричної бази Scopus.

Тематика публікацій відповідає змісту дисертаційної роботи.

Таким чином, наукові результати, описані в дисертаційній роботі, належним чином оприлюднені у наукових публікаціях здобувача.

6. Недоліки та зауваження до дисертаційної роботи

Визнаючи високий рівень проведених досліджень, маю зазначити наступні недоліки та зауваження до дисертаційної роботи.

1. У першому розділі розглянуто різні підходи до наближеного пошуку найближчих сусідів. У пункті 3.2.1 наведено пояснення відмінностей MDC від цих підходів. Водночас позиціонування запропонованої моделі MDC у межах класу ANN/ANNS-рішень доцільно було б подати більш формалізовано, наприклад у вигляді окремої класифікаційної таблиці з порівнянням характеристик.

2. Експериментальна перевірка уніфікованої моделі проектування та оцінки якості CBIR-систем у пункті 4.1.7 підтверджує можливість заміни компонентів без модифікації архітектурного ядра. Водночас подальше розширення кількості перевірених конфігурацій, типів сховищ, дескрипторів та прикладних сценаріїв дозволило б повніше оцінити практичну універсальність запропонованої моделі.

3. Оскільки програмний прототип, описаний у розділі 4.1, реалізовано з використанням Java, Spring Boot та Docker, доцільно було б детальніше описати параметри середовища виконання, зокрема версію JDK/JVM, налаштування пам'яті JVM, параметри збирача сміття, обмеження ресурсів Docker-контейнерів та режим запуску, оскільки ці чинники можуть впливати на час пошуку, формування індексів і використання ОЗП.

4. У пункті 4.4.4 наведено умови проведення експериментів: використання декількох ітерацій і середніх значень за ітераціями, а для навантажувального тестування використано Locust. Водночас доцільно було б детальніше описати протокол бенчмаркінгу: наявність або відсутність етапу прогріву, правила усереднення, обробку відхилень, розкид результатів та можливість використання спеціалізованих бібліотек для мікробенчмаркінгу.

5. У пункті 4.2.2 проведено порівняння різних СУБД для розміщення MDC у базі даних, результати якого наведено у таблиці 4.6. Водночас не деталізовано, чи виконувалося додаткове налаштування параметрів СУБД, зокрема параметрів пам'яті, кешування, індексації та виконання запитів. Аналогічне уточнення було б корисним і для реалізацій, що використовують розширення pgvector.

Вважаю, що зазначені зауваження не є визначальними, не зменшують загальну наукову новизну та практичну значущість результатів і не впливають на позитивну оцінку дисертаційної роботи.

7. Висновок про дисертаційну роботу

Вважаю, що дисертаційна робота здобувача ступеня доктора філософії Даниленка Станіслава Дмитровича на тему «Моделі та методи пошуку зображень у сховищах даних за векторним представленням ознак» є завершеним науковим дослідженням, виконаним на високому науковому рівні, не порушує принципів академічної доброчесності. Сукупність теоретичних та практичних результатів роботи розв'язує актуальне науково-практичне завдання, що має істотне значення

для галузі знань 12 Інформаційні технології. Дисертаційна робота за актуальністю, практичною цінністю, науковою новизною, обґрунтованістю та достовірністю отриманих результатів повністю відповідає вимогам чинного законодавства України, що передбачені в п. 6–9 «Порядку присудження ступеня доктора філософії та скасування рішення разової спеціалізованої вченої ради закладу вищої освіти, наукової установи про присудження ступеня доктора філософії», затвердженого Постановою Кабінету Міністрів України від 12 січня 2022 р. № 44.

Здобувач Даниленко Станіслав Дмитрович заслуговує на присудження ступеня доктора філософії в галузі знань 12 Інформаційні технології за спеціальністю 121 Інженерія програмного забезпечення.

Рецензент,
доктор технічних наук, професор,
професор кафедри
електронних обчислювальних машин
Харківського національного
університету радіоелектроніки

Максим ВОЛК