

ВІДГУК

офіційного опонента на дисертацію
Петренка Максима Віталійовича
«Методи інтеграції гетерогенних бібліографічних даних»,
що представлена до спеціалізованої вченої ради К 26.139.03
на здобуття наукового ступеня кандидата технічних наук
за спеціальністю 01.05.03 – математичне та програмне забезпечення
обчислювальних машин і систем

Актуальність теми дисертації.

Проблема інтеграції даних є актуальна впродовж тривалого часу. Важливим питанням залишається невідповідність між відомими методами і засобами інтеграції даних та принципами побудови відкритих інформаційних систем.

Процеси синтезу даних мають широку сферу застосування. Інформаційні ресурси таких систем передбачають одночасне застосування значної кількості різноманітних даних. Однією з основних проблем синтезу є створення єдиних правил і способів подання гетерогенних даних. Тому інтеграція даних бібліотек, як складової інформаційних ресурсів з метою створення інтегрованих інформаційних ресурсів бібліотек є актуальний напрямок наукових досліджень.

Теоретичний аспект проблеми передбачає розробку загальної методології інтеграції даних, створення нових концепцій опрацювання даних в інформаційних системах та розвиток теоретичних засад інтеграції інформаційних ресурсів.

Тому розробка і запровадження уніфікованих методів побудови інформаційних ресурсів, тобто інтеграція різнорідних даних є важлива і актуальна науково-практична задача.

Структура дисертаційної роботи.

Дисертаційна робота структурно складається з анотації, вступу, п'яти розділів, висновків, списку використаних літературних джерел та додатків.

Роботу викладено загальним обсягом 121 стор., вона містить 30 рисунків, 30 таблиць та список використаних літературних джерел з 91 найменувань та додатків.

У **вступі** обґрунтовано актуальність теми дисертації, зазначено її зв'язок із науковими програмами та темами, сформульовано мету і задачі досліджень, викладено наукову новизну та практичну значимість отриманих результатів, наведено дані щодо публікацій автора за темою дисертації, а також відомості щодо впровадження результатів дисертаційної роботи.

У **першому розділі** «Інформаційні системи публічних бібліотек - поточний стан та перспективи» розглянуто технічні моменти сучасних тенденцій впровадження інформаційних технологій в бібліотечну сферу та супутні проблеми. Описано структуру інформаційних систем публічних бібліотек м. Києва та особливості інформаційних процесів в них. Проаналізовано напрями вдосконалення зведених електронних каталогів, визначено проблеми та перспективи зведених електронних каталогів в Україні. Розглянуто особливості систем доступу до зведених даних бібліотек залежно від способу отримання даних.

У результаті виконаного аналізу проблеми зроблено висновок, що

найбільшу ефективність в інтеграції гетерогенних даних матимуть методи, які розроблені на основі реальних бібліографічних даних українських публічних бібліотек.

У другому розділі «Аналіз даних відкритих інформаційних систем» обґрунтовується використання даних публічних бібліотек Києва для аналізу, знаходження особливостей даних та розробки методів інтеграції гетерогенних даних, які б ґрунтувалися на реальному стані та особливостях даних публічних бібліотек України.

Підґрунттям дисертаційної роботи стали вибір та отримання даних. Інтеграція баз даних зведених каталогів низки бібліотек м. Києва надала можливість створити зведений каталог публічних бібліотек та аналізувати особливості гетерогенних бібліографічних даних і тестування розроблених алгоритмів на великих наборах реальних даних.

Виконано порівняльний аналіз статистичних характеристик значень полів бібліографічних записів різних баз.

У третьому розділі «Візуально непомітні помилки введення та їх вплив на якість та пошукову доступність бібліографічних даних» висвітлено проблему випадкового використання при введенні даних візуально схожих символів з іншою розкладкою клавіатури.

Не зважаючи на те, що такі помилки є відомими, використовувані в Україні АБІС допускають введення даних з помилками та не мають ефективних засобів пошуку та виправлення помилок у бібліографічних даних, а отже помилки накопичуються. Виявити їх лише силами операторів, без спеціального програмного забезпечення, неможливо. А через модулі електронних каталогів стає складно чи неможливо знайти сотні записів книг лише через цей клас помилок.

У роботі запропоновано і застосовано простий метод детектування (а отже і виправлення) цього виду помилок, який побудований на правилах та не потребує наявності статистичних даних. Пропонований метод є достатньо точним для виявлення помилок.

Четвертий розділ «Пропоновані методи інтеграції гетерогенних бібліографічних даних» містить:

- аналіз помічених проблем якості бібліографічних даних, усунення яких дозволить краще інтегрувати гетерогенні дані;
- перевірку існуючих методів на реальних даних;
- пропозиції по вдосконаленню та розробці нових методів;
- опис та тестування розроблених автором методів.

Метою інтеграції гетерогенних бібліографічних даних є отримання єдиної однорідної за форматом, якістю та з мінімальною надлишковістю бібліографічної бази даних.

Виконано аналіз існуючих проблем основних бібліотечних полів. Запропоновано ефективний підхід до вирішення цих проблем, який полягає у розробці методу порівняння значень полів, альтернативний простому порівнянню чисел.

З метою ідентифікації аномалій у даних та мінімізації їх негативного впливу автором запропоновано методи автоматичного виявлення структури та генерації правил на основі емпіричних даних.

Автором розроблено метод розпізнавання даних, що ґрунтується на порівнянні тексту, стійкого до скорочень, який ґрунтується на правилах та особливостях утворення різних варіантів скорочень. Цей метод описано у вигляді взаємопов'язаних рекурентних функцій.

У п'ятому розділі описана «Програмна реалізація пропонованих методів та генерація інформаційних ресурсів». Всі методи були реалізовані у вигляді алгоритмів мовами програмування Delphi або Python.

У висновках сформульовано основні наукові і практичні результати, що отримано у дисертаційній роботі, а також стверджується, що на основі виконаних теоретичних і практичних досліджень запропоновано методи, які можуть мати універсальне значення для інтеграції бібліографічних ресурсів.

У додатках наведено акти впровадження результатів досліджень, програмна реалізація запропонованих алгоритмів інтеграції інформаційних даних.

Основні наукові результати.

В результаті дослідження найбільшого в Україні регіонального масиву бібліографічних даних публічних бібліотек (2 джерела, 141 публічна бібліотека, різні формати даних):

Оцінено вплив на пошукову доступність бібліографічних даних відсутності в АБІС засобів уникнення візуально непомітних помилок введення;

Знайдено та описано особливості та закономірності бібліографічних даних, які використані в розроблених методах.

Розроблено нові унікальні методи:

- метод порівняння значень поля «Сторінки» стійкий до різних підходів визначення кількості сторінок та, відповідно, різних значень поля;
- метод порівняння значень поля «Видавництво» стійкий до використання різних видів скорочень;
- метод виявлення аномалій даних поля ISBN;
- метод мінімізації впливу аномалій даних поля ISBN на процес інтеграції даних.

Наукове значення одержаних результатів. Наукові результати представленої дисертаційної роботи сприяють розвитку та вдосконаленню методів порівняння та очищення бібліографічних даних в цілому, та особливо у застосуванні до інтеграції гетерогенних бібліографічних даних.

Теоретичні результати дослідження, представлені у вигляді розроблених методів, дозволяють оцінити якість бібліографічних даних, порівнювати значення полів в умовах роботи з гетерогенними бібліографічними даними, аналізувати бібліографічні дані на наявність аномалій.

Практична цінність результатів дисертаційної роботи

В ході роботи було розроблено комплекс, що включає базу даних та ряд програм дослідницького та прикладного призначення:

1. спеціалізована пошукова програма-робот яка надає можливість індексування бібліографічної інформації в мережі Інтернет, з можливістю ефективного

- оновлення та доповнення індексу;
2. програмне забезпечення для:
- оцінки якості та пошуку закономірностей в бібліографічних даних;
 - очистки бібліографічних даних;
 - інтеграції бібліографічних даних з різних джерел;
 - шаблон сайту зведеного каталогу та власна спеціалізована CMS (система управління контентом) зведеного каталогу;
 - інформаційний ресурс публічних бібліотек м. Києва.

Результати дисертаційного дослідження впроваджені в роботу та пройшли апробацію у бібліотеках ЦБС Солом'янського району м. Києва, зокрема у центральній бібліотеці району ім. Достоевського (переможець всеукраїнського конкурсу «Бібліотека року 2020»), про що є довідка про впровадження.

Ступінь обґрунтованості та достовірність наукових положень, висновків і рекомендацій дисертаційної роботи

Достовірність наукових положень і результатів, отриманих у дисертаційній роботі, підтверджується коректним використанням методів обробки даних. Достовірність також підтверджена відтворюваністю та можливістю перевірки теоретичних і експериментальних результатів та їх практичною реалізацією. Кількісні оцінки результатів обробки даних отримані в результаті дослідних експериментів.

Наукові положення дисертації, основні результати досліджень, висновки і рекомендації достатньо обґрунтовані та відображають особистий внесок здобувача в досягненні мети дослідження.

Повнота викладення наукових положень, висновків і рекомендацій, сформульованих у дисертації та опублікованих роботах автора

Результати дисертаційного дослідження опубліковано у 8 публікаціях за темою дисертації, з яких: 4 публікації у виданнях, що входять до переліку, затвердженого МОН України; 1 публікація у виданнях країн Європейського союзу.

Публікації автора задовольняють усім вимогам чинного «Порядку присудження наукових ступенів», затвердженого постановою Кабінету Міністрів України від 24 липня 2013 року № 567. Повнота відображення результатів дисертаційного дослідження та умови щодо кількості публікацій відповідає вимогам, які висуваються до докторських дисертацій з обраної спеціальності.

Оцінка мови, стилю та оформлення дисертації й автореферату

Дисертаційну роботу написано українською мовою грамотно, на хорошому стилістичному рівні. Тест дисертації викладено грамотною технічною мовою, логічно і послідовно. Стиль викладення доказовий.

Застосована в роботі наукова термінологія є загальновизнаною, стиль викладення результатів теоретичних і практичних досліджень, нових наукових положень, висновків і рекомендацій забезпечує доступність їх сприйняття та використання. Мова і стиль викладання дисертації та автореферату коректно висвітлюють отримані науково-практичні результати.

Зміст автореферату повністю відповідає змісту дисертаційної роботи. Автореферат містить основні положення, висновки і рекомендації, що наведено у дисертації, а також усю іншу необхідну для оцінки роботи інформацію. Оформлення дисертаційної роботи й автореферату повністю відповідає вимогам державних стандартів України.

Зауваження по дисертаційній роботі

1. Відсутній розділ «Постановка задач досліджень».
2. Відсутній список запроваджених у дисертаційній роботі скорочень та аббревіатур.
3. При аналізі даних відкритих інформаційних систем у недостатньому обсязі викладено концепції теорії множин.
4. Недостатньо приділено увагу впливу описаних помилок на процес інтеграції даних.
5. Відсутня математична формалізація описаних методів.

Але наведені зауваження не торкаються сутності виконаних досліджень, не знижують високий науковий рівень дисертаційної роботи і не впливають на її загальну позитивну оцінку.

Висновки щодо відповідності дисертації встановленим вимогам МОН України.

Дисертаційна робота Петренка Максима Віталійовича «Методи інтеграції гетерогенних бібліографічних даних» є завершеною науковою працею, в якій отримано нові, науково обґрунтовані та практичні результати, що в сукупності вирішують важливу науково-прикладну проблему створення методів і засобів інтеграції гетерогенних даних у відкритих інформаційних системах.

Автореферат у повній мірі відображає основний зміст, результати та висновки дисертаційної роботи і є ідентичним положенням дисертації. Наведені в авторефераті та дисертації публікації повністю висвітлюють основні наукові результати дисертаційного дослідження. Оформлення дисертації і автореферату відповідають вимогам МОН України. Тема, зміст та результати дисертації відповідають паспорту спеціальності 01.05.03 – математичне та програмне забезпечення обчислювальних машин і систем

Вважаю, що за актуальністю теми, високим науковим рівнем виконаних досліджень, новизною, прикладним значенням одержаних результатів дисертаційна робота «Методи інтеграції гетерогенних бібліографічних даних» відповідає вимогам пп. 9, 10, 12 «Порядку присудження наукових ступенів», затвердженого постановою Кабінету Міністрів України від 24 липня 2013 року № 567 на здобуття наукового ступеню кандидата технічних наук за спеціальністю 01.05.03 – математичне та програмне забезпечення обчислювальних машин і систем

Офіційний опонент

доктор технічних наук, доцент

доцент кафедри біомедичної кібернетики

КПІ імені Ігоря Сікорського

Кирило ЗЕЛЕНСЬКИЙ

Підпис доц. Зеленського К.Х. засвідчую.

Учений Секретар
КПІ імені Ігоря Сікорського

Валерія ХОЛЯВКО

«8» грудня 2021

