

РЕЦЕНЗІЯ

на дисертаційну роботу Василенка Олега Олеговича
«Розроблення засобів опрацювання авторитетних записів для автоматизації
зведеного бібліографічного каталогу»,
подану на здобуття наукового ступеня доктора філософії
за спеціальністю 123 «Комп'ютерна інженерія»

Актуальність теми дослідження

Стрімке зростання обсягів інформаційних ресурсів у бібліотеках з переходом на нові моделі обслуговування користувачів потребують сучасних засобів інтеграції метаданих. Дисертаційне дослідження присвячене розробленню засобів і методів автоматизації отримання, конвертування (зокрема з MARC21 до UNIMARC), очищення, узгодження та об'єднання авторитетних записів від різних джерел. Ці засоби і методи було застосовано автором для створення автоматизованої системи зведеного каталогу. Розроблені засоби і методи збагачують функціональні можливості автоматизованих бібліотечних інформаційних систем (АБІС), зокрема, широко розповсюдженої в українських бібліотеках АБІС Коґа, за рахунок автоматизації створення і поповнення авторитетних файлів АБІС. Питання національного авторитетного файлу та різноманіття АБІС породжують суттєві бар'єри для формування єдиних інформаційних просторів. Наведені методи накопичення коректних авторитетних записів є актуальними, оскільки дозволяють автоматизувати ведення зведеного каталогу в АБІС і зменшити кількість ручної праці бібліотекарів.

Зв'язок дослідження з науковими програмами й темами

Запропоноване дисертантом рішення щодо автоматизованого опрацювання авторитетних записів відповідає критичним викликам галузі та сприяє реалізації завдань «Стратегії розвитку бібліотечної справи на період до 2025 року “Якісні зміни бібліотек для забезпечення сталого розвитку України”» (розпорядження Кабінету Міністрів України № 219-р від 23 березня 2016 р.).

Рівень обґрунтованості наукових результатів

Програмні компоненти (гарвестер бібліографічних і авторитетних записів, гарвестер записів WikiData, гарвестер записів ISNI, модуль очищення записів, модуль кластеризації записів, та модуль об'єднання записів) описані формально

з наведеними схемами алгоритмів. Програмний код доступний в репозиторії GitHub. Експериментальна база сформована на реальних даних українських бібліотек, ISNI та Wikidata. Отже, висновки мають достатнє теоретичне й практичне підґрунтя.

Огляд змісту дисертації та відповідність вимогам до оформлення

Робота складається зі вступу, чотирьох розділів, висновків, списку літератури (180 джерел) і трьох додатків. Структура дотримується вимог до дисертацій на здобуття ступеня доктора філософії. Текст та ілюстрації подано українською та частково англійською мовами; нумерація рисунків і таблиць коректна. Окремої уваги заслуговує додаток В із детальною XSL-картою конвертації MARC21 → UNIMARC.

Практична значимість роботи

Розроблений програмний комплекс «КОБЗА» впроваджений в Українському гуманітарному інституті (м. Буча, Київська область) і забезпечує консолідацію каталогів декількох партнерських ЗВО. Середній час каталогізації одного документа зменшено втричі (від 12 до 4 хвилин). Цей комплекс із застосуванням методів GPU-прискорення має інтерес для великих бібліотек.

Наукова новизна результатів дисертації

У дисертації вперше розроблено методику добування авторитетних записів з використанням протоколів ILS-DI, OAI-PMH та REST API. Особливістю пропонованого підходу є відмова від протоколів, які потребують окремого сервера (Z39.50, SRU/SRW), що забезпечує корпоративну взаємодію бібліотек, які працюють в АБІС Koha. У дисертації вперше обґрунтовано й реалізовано повне XSL-перетворення форматів MARC21 → UNIMARC для авторитетних записів про осіб, без введення додаткових, не передбачених UNIMARC полів. Рішення є платформонезалежним та дає змогу оперативно імпортувати записи в чинні АБІС, зокрема Koha. У роботі удосконалено методику верифікації авторитетних записів шляхом зіставлення з національними авторитетними базами даних із використанням міжнародної бази імен ISNI та семантичної бази знань Wikidata. Новизна полягає у перевірці мовного варіанта заголовка й урахуванні різної послідовності компонентів особового імені. Уперше

запропоновано комплексний алгоритм кластеризації та об'єднання авторитетних записів без опори на національний нормативний файл: застосовано методи fuzzy matching (нечіткого паросполучення) для обробки верифікованих бібліотечних даних і відомостей з міжнародних ресурсів. Цей алгоритм враховує українську специфіку (наявність омогліфів, використання ініціалів) і забезпечує якісне формування кластерів авторитетних записів.

Повнота викладення результатів у публікаціях

Напрацювання, отримані в межах дисертаційного дослідження, знайшли відображення у трьох статтях, опублікованих у наукових виданнях, що входять до переліку фахових журналів України категорії «Б», додатково результати дисертації відображені в чотирьох інших публікаціях. Частина результатів була також представлена на шести міжнародних наукових конференціях. Така кількість і рівень публікацій відповідають вимогам МОН України до дисертацій на здобуття наукового ступеня доктора філософії за даною спеціальністю.

Зауваження та рекомендації

1. У розділі 3 назви підрозділів 3.5 і 3.6 дублюють слово «об'єднання».
2. Доцільно розглянути можливість імпорту метаданих з АБІС Ірбіс як такої, що все ще використовується у великих бібліотеках, а також дослідити Ірбіс з точки зору кібербезпеки.
3. Замість використання “ймовірнісний зв'язок записів” (с. 3, 120, 121, 132) краще використовувати термінологію “нечітке (fuzzy) зіставлення (паросполучення)”.
4. Слід обґрунтувати створення окремого стовпця MergedISNI в таблиці, яка використовується для збереження даних отриманих з бази даних ISNI.
5. Назва пункту «2.2.3. Типи сутностей» на с.70 не повністю відображає зміст цього пункту. Цей пункт доречніше назвати як «Тенденції розроблення єдиної моделі бібліотечних даних з урахуванням концептуальної моделі «сутність – зв'язок» і принципів семантичної мережі» або «Чотири рівні сутностей» (саме цей термін вживається на с. 71, а термін «типи сутностей» у тексті пункту не згадується взагалі). Зміст пункту 2.2.3 не дозволяє однозначно визначити, чи є цей пункт просто оглядом наукових робіт з проблеми «чотирьох рівнів

сутностей» (або з проблеми побудови єдиної моделі бібліотечних даних), чи у цьому пункті автор аналізує інформацію, яка суттєво використовується у розробленій ним програмній системі КОБЗА.

6. У назві рисунку 2.6 на с. 72 і у тексті роботи немає посилання на першоджерело, використане для підготування цього рисунка.

7. На с. 59 йде мова про метод перетворення авторитетних записів у бібліографічні, але не згадується про аналогічне перетворення авторитетних записів (MARC21) в авторитетні записи (UNIMARC). Можливо, на с. 59 просто допущена технічна помилка і мало би бути написано «...авторитетних записів в авторитетні ...»

8. Не подається однозначне і вичерпне визначення терміну «сервер» у тому сенсу, який розуміє автор в дисертаційній роботі. Аналізуючи вживання терміну «сервер» у дисертаційній роботі на сторінках 4, 20, 33, 34, 37–40, 42, 44–46, 48, 99–101, 104, 107, 108, 111, 114, 115–118, 128, 129–131, можна зрозуміти, що у більшості випадків автор має на увазі сервер бібліографічних послуг («сервер БП»), тобто програму, яка функціонує в певній АБІС, обробляє і дає відповіді на запити бібліографічного характеру, що надходять від користувачів тієї самої АБІС чи інших АБІС у мережі Інтернет. Однак цього не зазначається в тексті.

9. На с. 38 є речення «Архітектура ОАІ-РМН базується на клієнт-серверній моделі». Проте архітектура «клієнт-сервер» використовується не тільки для реалізації протоколу ОАІ-РМН, але й для реалізації практично кожного протоколу передачі метаданих, який застосовується для надання і отримання бібліографічних послуг.

10. Не завжди коректно і досить довільно вживається термін «сервер» у різних сенсах, зміст якого потребує контексту та ретельного аналізу тексту, наприклад: на с. 4 – «із серверів бібліотек» (із серверів бібліографічних послуг (БП) – прикладних програм, які надають послуги пошуку, збирання і систематизації бібліографічних даних);

на с. 4 – «розроблено методику ..., яка ... не використовує протоколи передачі метаданих, які потребують окремого сервера» (автором розроблена методика, яка

забезпечує використання одної програмної системи КОБЗА для отримання даних за декількома протоколами бібліографічних послуг);

на с. 33 – «Клієнтська програма може звертатися до кількох серверів баз даних» (розроблена автором система КОБЗА забезпечує отримання інформації з різних серверів баз даних без необхідності явно враховувати особливості тих чи інших типів СУБД – систем управління базами даних);

н с. 33 – «... клієнт може скопіювати результати отримані з різних серверів» (...програма-клієнт (система КОБЗА) може накопичувати інформацію, знайдену і отриману від декількох АБІС, використовуючи різні протоколи БП);

на с. 44 – «... підхід передбачає передачу ... файлів ... через... FTP-сервери» (бібліографічні файли передаються згідно з протоколом передачі файлів);

на с.107 – «Якщо сервер має графічний процесор, ...» (сервер як потужний комп'ютер з графічним процесором);

на с.107 – «...на прикладі оренди серверів Amazon Web Services» (закупівля обчислювальних послуг на серверах Amazon);

на с.108 – «на тестовому сервері Linux Intel Xeon E7-8867 v3 з 24 Гб оперативної пам'яті» (на тестовому комп'ютері);

на с.129 – «бібліографічних і авторитетних серверів» (Бібліографічний сервер (прикладна програма, сервер бібліографічних послуг) забезпечує пошук бібліографічних записів відповідно до параметрів запиту на певні бібліографічні записи. Авторитетний сервер (прикладна програма, сервер бібліографічних послуг) виконує аналогічну роботу щодо авторитетних записів).

Таким чином, термін «сервер» має у роботі широкий спектр значень, яке не завжди швидко можна визначити сенс у кожному випадку застосування терміну.

11. Очевидно, що система КОБЗА виконує функції сервера бібліографічних послуг і функції клієнта, який надсилає запити до інших АБІС. Ця обставина явно не відображена при описі функцій системи.

12. Не визначається, яке місце займає програмна система КОБЗА на шляху до створення єдиної моделі бібліотечних даних. Відповідь на це питання може посилити теоретичне і практичне значення отриманих результатів.

13. На с. 115, не визначається термін «клас» у таких реченнях: «Кожен клас, що підтримує збирання даних за певним протоколом може працювати як з бібліографічними, так і авторитетними записами. Клас, що відповідає за Koher API автоматично пробує використовувати ILS-DI у випадку, якщо в налаштуваннях сервера не вказано ключів доступу». На початку цієї с. 115 подано речення: «Розроблена система має модульну структуру (див. Рисунок 4.5) і спільні класи для підготовки даних до запису в базу даних...». Тоді не визначено, що мається на увазі під терміном «клас» у наведених реченнях.

14. На с. 99 у другому реченні пункту «3.4.1. ISNI» не є зрозумілим сенс терміну «члени», де автор пише: «Сервер може надавати відповідь використовуючи дві XML-схеми: базову вільнодоступну схему isni-b, чи розширену з доступом тільки для членів». Незрозуміло, що мається на увазі під «членами»: ними можуть бути суб'єкти, об'єкти, користувачі, програми-клієнти. Якщо це користувачі, то для користувачів слід вказати предмет користування.

15. Не дається однозначне і зрозуміле визначення термінів «архітектура зведеного каталогу», «архітектура каталогу». Це важливо тому, що на с. 4 є речення «Четвертий розділ описує архітектуру та функціональні можливості зведеного каталогу», а на с. 132 у списку основних результатів роботи останнім пунктом результатів є «Представлено архітектуру та функціональні можливості зведеного каталогу...», що є єдиним місцем у четвертому розділі (та всій роботі), де автор згадується термін «архітектура зведеного каталогу». Термін «архітектура зведеного каталогу» залишається не визначеним явно, хоча створення «зведеного каталогу» відповідно до запропонованої «архітектури зведеного каталогу» подається як один з головних результатів роботи.

16. Не дається визначення терміну «архітектура отримання записів». Цей термін може бути важливим тому, що на с. 48 є речення «Початковим етапом автоматизації зведеного каталогу є побудова архітектури отримання записів», але подальших пояснень стосовно цього терміну не надається, а цей термін більше не зустрічається у роботі.

17. У пункті «2.2.3. Типи сутностей» (с. 70) друге речення не містить підмета і присудка: перші два речення на с. 70 мають об'єднуватися в одне «Sh. Wu [67]

пропонує використовувати дані бібліографічного покращення (bibliographic enhancement data, B.E.D.), проаналізувавши запити читачів до різних типів B.E.D., таких як рейтинги, огляди, пов'язані твори або посилання на інші формати (наприклад, аудіокниги, електронні книги), які розширюють можливості користувача, надають більш багатий контекст та інформацію про доступні матеріали».

18. У роботі зустрічаються граматичні помилки чи неточності. Наприклад:
на с. 70 слід (писати) «таких, як рейтинги» замість ««такі як рейтинги»;
на с.70 слід «вже зараз» замість ««в же зараз»;
на с.72 слід «дані, знайдені» замість «дані, знайдених»;
на с.76 слід «інші варіанти назви» замість «інші варіанти назву»;
на с.77 слід «заповнення поля «мова» » замість «заповнення поля мова»;
на с. 81 слід «недостатнє використання» замість «недостатнє використання»;
на с. 86 слід «дії дуже розрізняються» замість «дії дуже розрізняється»;
на с. 94 слід видалити одне з двох слів «бажано» або «логічно»;
на с.98 і на с.99 двічі повторюється той самий абзац, вище і нижче рисунка 3.7 на с. 98;
на с.104 двічі повторюється одне слово у назві підрозділу «3.5. Особливості об'єднання об'єднання записів»;
на с.126 слід «бібліотекарям» замість «бібліотекарам».

19. На с. 116 у рядках 4 і 5 є неповне речення, яке не має підмета, присудка, прийменника: «В залежності від протоколу записи MARCXML гілках відповіді сервера».

20. В українській мові слова «відноситися», «відносити» мають набагато вужчий зміст, ніж російські слова «относиться», «относится». Тому в українській мові бажано вживати (в залежності від контексту) інші слова, такі як «належати», «розподіляти», «стосуватися» тощо. Наприклад:

на с.78 замість «До такої категорії можуть відноситися» слід (писати) «До такої категорії можуть належати»;

на с.79 замість «що відносяться до заголовків» слід «що належать до заголовків»;

на с. 86 замість «...пари записів відносять до одного з трьох класів...» слід «...пари записів розподіляють на три класи...»;

на с.105 замість «Четвертий варіант, що відноситься до вилучення полів відносився до застарілих полів» слід «У четвертому варіанті, що стосується вилучення полів, розглянута обробка застарілих полів»;

на с.119 замість «...які відносяться до характеристик...» слід «... які належать до характеристик...»;

на с. 123 замість «...що відносяться до кожного імені» слід «...що належать до кожного імені».

Зауваження мають технічний характер і не знижують загальної високої якості роботи.

Загальний висновок

Дисертаційна робота Василенка Олега Олеговича є завершеним науковим дослідженням, що містить нові результати, важливі для розвитку автоматизованих бібліотечних систем. За актуальністю, рівнем обґрунтованості, науковою новизною, теоретичною й практичною цінністю та повнотою апробації вона відповідає вимогам «Порядку присудження ступеня доктора філософії та скасування рішення разової спеціалізованої вченої ради закладу вищої освіти, наукової установи про присудження ступеня доктора філософії» (затверджено постановою Кабінету Міністрів України від 12 січня 2022 р. № 44). Вважаю, що здобувач Василенко Олег Олегович заслуговує присудження ступеня доктора філософії за спеціальністю 123 «Комп'ютерна інженерія».

Рецензент

завідувач відділу інтелектуальних інформаційних технологій
Інституту кібернетики імені В. М. Глушкова НАН України
доктор фізико-математичних наук, професор

Василь ГОРБАЧУК

