УДК 62-50:519.49

В.М. Григор'єв, Д.С. Сліпуха

# РОЗРОБКА ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ДЛЯ КАТАЛОГІЗАЦІЇ ІНФОРМАЦІЇ В БІБЛІОТЕЧНІЙ СИСТЕМІ

*Анотація. Розроблено програмне забезпечення, що дозволяє створювати та використовувати бібліотечний каталог, та включає в себе рекомендаційну систему, що значно прискорює процес каталогізації бібліотеки.*
*Ключові слова: бібліотечний каталог, рекомендаційна система.*

**Актуальність теми.** Процес каталогізації інформації в бібліотеці завжди потребував значних зусиль з боку людини, оскільки його неможливо повністю автоматизувати. Він включає до себе такі процеси, як поповнення каталогу, систематизація, класифікація, доробка внутрішнього та зовнішнього оформлення, бібліографічна обробка, та багато інших. Проте завжди існує можливість частково автоматизувати процес, з ціллю зведення ручної праці до мінімуму. Ця задача ускладнена тим, що не існує загальних методів автоматизації процесів у слабоформалізованих бібліотечних системах, так як кожна з них потребує різних підходів, в залежності від типу бібліотечної системи. На даний момент фактично не існує повнофункціонального програмного забезпечення, що повністю дозволяє вирішити цю задачу.

**Аналіз останніх досліджень.** Інформаційну основою процесу каталогізації складають рекомендаційні системи. Векторна модель рекомендаційних систем полягає у тому, щоб представити бібліотечні об'єкти (книжки, каталоги та ін.) у вигляді багатовимірних векторів, де у якості координат використовується чисельне уявлення деяких їх властивостей. Тоді величиною, що характеризує схожість цих об'єктів буде косинус кута між векторами. У якості координат вектору можна використовувати будь-які параметри, по яких можна порівнювати об'єкти, та які можливо представити в чисельному вигляді. Популярним параметром є вага ключових слів.

Найбільш розповсюджені наступні системи каталогізації.

**Book Collector** – це досить потужна програма для каталогізації бібліотеки, з комплекту програм collectorz.com. Важливою перевагою цієї програми є модуль пошуку по базах даних Інтернет, що значно прискорює процес каталогізації та оформлення колекції. Також присутній модуль пошуку по книгах, що дозволяє шукати по ISBN, або назві книги. Проте слід зазначити, що ця програма більш призначена для керування приватною колекцією книг, аніж бібліотекою з багатою кількістю користувачів, та являється платною.

**БД Книги** – це доволі проста програма, призначена для впорядкованого зберігання та швидкого пошуку електронних книг. Вона є безкоштовною та не потребує реєстрації, проте має обмежену функціональність та не призначена для керування великою бібліотекою. Також вона не має засобів прискорення процесу каталогізації, і не призначена для сумісного користування бібліотекою.

**MyHomeLib** – це безкоштовна програма для керування колекціями електронних книг, що розповсюджується за ліцензією GPLv3. Вона має широкі можливості, такі як: необмежена кількість користувальницьких колекцій, автоматичний імпорт та експорт книг з будь-якими типами файлів, підключення користувальницьких скриптів для обробки книг. Проте, попри всі переваги, ця програма також більш орієнтована на приватну електронну бібліотеку, та не надає жодних можливостей для автоматизації процесу каталогізації.

Проведений аналіз існуючих систем каталогізації виявив, що жодна з існуючих систем не задовольняє водночас таким потребам, як:

- підтримка великих електронних бібліотек;
- підтримка видаленої роботи декількох користувачів;
- виявлення дублікатів книг у каталозі;
- наявність рекомендаційної системи, що прискорює процес каталогізації;
- наявність інтерфейсу адміністратора, що дозволяє керувати процесом каталогізації;
- індексація та пошук по книгах та їх змісту;

Для вирішення проблем каталогізації було прийняте рішення розробити таке програмне забезпечення у вигляді Інтернет-проекту, яке буде відповідати всім цим вимогам та надасть можливість

створення та використання бібліотечного каталогу, а також створення рекомендаційної системи, яка призначена полегшити процес каталогізації інформації у бібліотеці.

**Обґрунтування отриманих результатів.** Для користування розробленим програмним каталогом, насамперед необхідно синхронізувати його з існуючою бібліотекою. Процес синхронізації складається з чотирьох етапів:

1. Спочатку сканується структура бібліотеки та на її основі будується тимчасове дерево каталогу;
2. Виконується рекурсивний обхід створеного дерева, при якому в базу додаються нові категорії та книги;
3. Існуючі в каталогу книги перевіряються на наявність в бібліотеці, та видаляються з каталогу, якщо вони не присутні в бібліотеці;
4. Існуючі в каталогу категорії перевіряються на наявність в бібліотеці, та видаляються з каталогу, якщо вони не присутні в бібліотеці, або якщо вони не мають книг у собі.

Для виконання повнотекстового пошуку, а також для створення ключових слів для рекомендаційної системи, необхідно мати можливість працювати з текстовим вмістом книг. Для цього потрібно індексувати вміст книг у бібліотеці. Оскільки книги в бібліотеці зберігаються у різних форматах, потрібно забезпечити можливість вилучення тексту з найбільш поширених форматів книг. Це дозволяють зробити пошукові фільтри служби Windows Search, або сторонні програми. Служба каталогу використовує обидва методи, оскільки в залежності від формату даних, різні методи працюють з різною швидкістю. Тому для прискорення виконання процесу індексації використовуються як фільтри, так і сторонні програми, в залежності від того, який метод дає більшу швидкість.

Слід зазначити, що для дуже великої кількості книг у бібліотеці, цей процес може зайняти декілька годин. Проте, оскільки при повторному запуску індексації будуть оброблятися лише не індексовані книги, існує можливість проведення повної індексації за декілька підходів.

Для подолання проблеми «холодного старту» рекомендаційної системи необхідно надати їй початкові дані, на основі яких вона зможе працювати. Такими даними виступають набори ключових слів,

що характеризують категорії бібліотеки. Набори ключових слів розподіляються на два типи:

1. Набори ключових слів з назв книг;
2. Набори ключових слів зі вмісту книг.

Створювати набори ключових слів з назв книг можна одразу після синхронізації каталогу з бібліотекою, оскільки при синхронізації назви книг зберігаються у базі даних.

Створення наборів ключових слів зі вмісту книг є можливим лише після проведення повної індексації бібліотеки.

Для кожної категорії бібліотеки створюється свій набір ключових слів, що засновується на тих книгах, які знаходяться в даній категорії.

Загальний алгоритм створення наборів ключових слів:

1. З наданих текстів (назви книг чи їх вміст) вилучаються усі слова;
2. З отриманого набору слів вилучаються стоп-слова;
3. Для кожного слова з набору підраховується скільки воно зустрічається в наданих текстах;
4. Слова, які зустрічаються менше зазначеної кількості разів вилучаються з набору;
5. Слова, які зустрічаються найчастіше, позначаються як ключові та зберігаються;
6. На основі відносної частоти зустрічі ключових слів у категорії обчислюється вага кожного ключового слова;
7. Вага ключових слів, що перетинаються у декількох категоріях зменшується пропорційно кількості таких категорій.

Для додання нових книг до каталогу слід перемістити їх до директорії не відсортованих книг, шлях до якої вказано в файлі конфігурації каталогу. Після цього, при наявності хоча б одного типу набору ключових слів, можна скористатися рекомендаційною системою каталогу. Рекомендаційна система намагається визначити, до якої категорії відноситься кожна не відсортована книга.

Алгоритм роботи рекомендаційної системи є наступним:

1. Завантажуються набори ключових слів, у формі, зручній для використання рекомендаційною системою;
2. З назви та вмісту оброблюваної книги вилучаються всі слова;

3. Підраховується ймовірність відношення книги до кожної категорії, базуючись на наявності ключових слів у імені книги;

4. Підраховується ймовірність відношення книги до кожної категорії, базуючись на наявності ключових слів у вмісті книги;

5. Отримані ймовірності для кожної категорії додаються;

6. Категорії, що в результаті мають найбільшу ймовірність надаються користувачеві для вибору.

Також рекомендаційна система відстежує, чи не знаходиться вже у бібліотеці книга, що додається. Якщо вона вже там знаходиться, то дублікат цієї книги буде переміщено до директорії дублікатів.

Слід зазначити, що точність рекомендаційної системи, що працює за таким алгоритмом, буде в першу чергу залежати від якості сортування вже існуючої бібліотеки.

Після обробки не відсортованих книг рекомендаційною системою, необхідно перейти на сторінку додання нової книги, де для кожної книги можна обрати категорію, з запропонованих рекомендаційною системою.

**Висновки.** Було розроблено програмне забезпечення, що дозволяє створювати та використовувати бібліотечний каталог, та включає в себе рекомендаційну систему, що значно прискорює процес каталогізації бібліотеки.

Умовно розроблене програмне забезпечення можна поділити на дві частини: служба каталогу та веб-додаток. Служба каталогу обробляє бібліотечні дані, створюючи на їх основі каталог та рекомендаційну систему. Веб-додаток надає адміністратору бібліотеки користувальницький інтерфейс до служби каталогу, а звичайним користувачам – доступ до бібліотеки.

Розроблена служба каталогу може виконувати наступні завдання:

- Синхронізація каталогу з бібліотекою;
- Пошук та видалення дублікатів книг;
- Індексація бібліотеки;
- Створення ключових слів;
- Надання рекомендацій, щодо каталогізації нових книг.

Точність роботи рекомендаційної системи збільшується зі збільшенням книг у каталозі.

Розроблене програмне забезпечення дозволяє спростити доступ до бібліотеки користувачам, та значно прискорює процес додавання нових книг у бібліотечний каталог.

## ЛІТЕРАТУРА

1. Application of Dimensionality Reduction in Recommender System - A Case Study / Badrul M. Sarwar, George Karypis, Joseph A. Konstan, John T. Riedl - University of Minnesota
2. Przemyslaw Kazienko, Pawel Kolodziejski Personalized Integration of Recommendation Methods for E-commerce - Wroclaw University of Technology, 2006
3. Ф.С. Воройский Основы проектирования автоматизированных библиотечно-информационных систем, Москва, ФИЗМАТЛИТ, 2002
4. David Mertz Text Processing in Python - Addison Wesley, 2003
5. Adrian Holovaty, Jacob Kaplan-Moss The Definitive Guide to Django: Web Development Done Right – Apress, 2008

Отримано 15.10.2009р.