

МЕТОДИ ТА ЗАСОБИ ПОШУКУ ПРОТИРІЧ У ЗАПИСАХ БІБЛІОГРАФІЧНОЇ БАЗИ ДАНИХ.

© Руслан Ярмолюк, 2012

У роботі піднімається актуальна проблема якості даних у бібліографічних базах даних (ББД) електронних каталогів бібліотек. Запропоновано підходи до пошуку логічних помилок та протиріч у записах ББД.

Ключові слова: електронний каталог, бібліографічна база даних, бібліографія, верифікація даних, система верифікації даних електронного каталогу.

The paper raised urgent problem of data quality in bibliographic databases (BDB) electronic library catalogs. Approaches to finding logical errors and contradictions in the records BBD.

Keywords: electronic catalog, bibliographic database, bibliography, verification of data, system data verification e-catalog.

Постановка проблеми.

Бібліографічна база даних (ББД) електронного каталогу бібліотеки (ЕКБ) є важливим компонентом, що задовольняє інформаційно-пошукові та аналітичні потреби користувачів бібліотеки. Від якості даних, що містяться у ББД напряму залежить результат пошуку певного джерела інформації. Під час функціонування інформаційної системи електронного каталогу у даних ББД можуть виникати різного роду помилки. Детальний аналіз та класифікація основних типів та джерел помилок запропонована у [1]. Пошук, аналіз та виправлення помилок різного роду є актуальною задачею для редакторів електронних каталогів сучасних бібліотек.

Класифікація помилок у ББД електронного каталогу.

Відповідно до характеру, помилки даних ЕК можна класифікувати наступним чином:

- помилка відсутності даних (відсутність даних у відповідних полях запису ББД);
- помилки уніфікації та стандартизації значень полів (відсутність єдиного підходу до структури запису у складних полях ББД, наприклад «автор», «видавництво», «назва джерела інформації»);

- орфографічні помилки у значеннях полів (орфографічні помилки, пропущені символи, перестановка символів у значеннях текстових полів ББД);
- помилка або протиріччя у полях певного запису ББД (даний тип помилки не дозволяє правильно ідентифікувати джерело інформації у ББД);
- помилка наявності дублів у ББД.

Для пошуку та виправлення вищезазначеного типу помилок у ББД ЕК автором у [2] запропоновано система верифікації даних електронного каталогу (СВДЕК).

Ціллю даного дослідження є розробка методів пошуку протиріч у ББД, що базуються на властивостях бібліографічного запису.

З властивостей бібліографічного запису та особливостей предметної області бібліографії було зроблено висновок про можливість розв'язання вищеописаної проблеми двома методами. Методом заснованим на протиріччі між полями запису та методом заснованим на статистичних законах бібліометрії.

Метод пошуку помилок заснований на протиріччі між значеннями полів запису. В основі методу лежить співставлення значень кожного поля бібліографічного запису з його можливо-допустимим значеннями, що визначаються з загальної сукупності усіх бібліографічних записів для кожного окремого запису. Метод заснований на інформаційній надмірності бібліографічних даних. Під інформаційною надмірністю будемо розуміти властивість бібліографічного запису, за якою однозначно ідентифікувати джерело інформації можна не використовуючи усі поля запису. Запропонуємо авторський алгоритм даного методу.

На першому етапі для кожного значення поля будується область допустимих значень усіх інших полів бібліографічного запису, за вибіркою з ББД. Під областю допустимих значень для значення *val_1* певного поля *atr_1* будемо розуміти множину усіх значень решти полів, що зустрічаються в одному записі з *val_1*.

На другому етапі отримані множини аналізуються на наявність аномальних значень, тобто значень, що виходять за межі середньо-допустимих. У випадку наявності аномальних значень у певній множині, редактору ЕК пропонується уточнити дане значення звіривши його з оригіналом джерела інформації.

Метод пошуку помилок заснований на бібліометричних законах. В основі методу лежать бібліометричні властивості записів ББД. До основних бібліометричних законів можна віднести: закон Лоткі, закон розподілу Дж.

Ципфа, закон розсіяння С.К. Бредфорда тощо. Огляд основних бібліометричних законів представлений у праці [3].

Кожен запис ББД та відповідне джерело інформації аналізуються на основі певного бібліометричного закону. У випадку відхилення від певних бібліографічних правил бібліографічний запис потребує перевірки редактором ЕК.

Висновки. У роботі піднімається актуальна проблема якості даних у ББД ЕКБ. Запропоновано підходи до пошуку логічних помилок та протиріч у записах ББД. Авторські підходи ґрунтуються на протиріччі між значеннями атрибутів та пошуку помилок заснований на бібліометричних законах. У подальшому робота буде проводитись над практичної реалізацією запропонованих методів. Вищезазначені методи будуть реалізовані у підсистемі пошуку протиріч СВДЕК.

Література

1. Ярмолюк Р.С. Основні типи та джерела помилок у записах електронного каталогу / Р.С. Ярмолюк // Вісник Національного Університету «Львівська політехніка» Інформаційні системи та мережі, - 2010. - № 689. – С. 348-357.
2. Ярмолюк Р.С. Структурно-функціональна модель верифікації даних електронного каталогу / Р.С. Ярмолюк // Суч. пробл. діяльн. бібл. в ум. інф. сусп. м. 3-ої наук.-прак. конф., 29 вересня 2011р., Львів/ Національний університет "Львівська політехніка"; - Львів : Видавництво НУ "Львівська політехніка", 2011. – С. 217-224.
3. Бредихин С.В. Методы библиометрии и рынок электронной научной периодики / С.В. Бредихин, А.Ю. Кузнецов – Новосибирск, Москва, 2012. – С. 255.

УДК 681.518.5

Олександр Шкіль, Дмитро Коротких
Харківський національний університет радіоелектроніки

АНАЛІЗ НАДІЙНОСТІ ТЕСТІВ В КОМП'ЮТЕРНІЙ СИСТЕМІ ТЕСТУВАННЯ ЗНАНЬ

© Олександр Шкіль, Дмитро Коротких, 2013

Розглянуто питання аналізу надійності тестів з використанням коефіцієнту α -Кронбаха для розріджених матриць результатів тестування. Запропоновано метод доповнення розріджених матриць та методика розрахунку статистичних показників з використанням табличного процесора Microsoft Excel.

Ключові слова – тестові завдання, надійність тестів, дисперсія, кореляція.