

## Структурно-функціональна модель. Системи верифікації даних електронного каталогу

---

**ЯРМОЛЮК Руслан Сергійович,**

Асистент кафедри прикладної математики та соціальної інформатики Хмельницького національного університету  
e-mail: rywerner@gmail.com

---

### Анотація

*Запропонована структурно-функціональна модель системи верифікації даних електронного каталогу. Описано основні вимоги до системи верифікації. Вказано схему модулів системи верифікації та основні функціональні можливості кожного з модулів.*

**Ключові слова** – електронний каталог, система верифікації даних електронного каталогу, АБИС, теорія нечітких множин.

### Abstract

*The proposed structural and functional model of system data verification OPAC. We describe the basic requirements for system verification. Specified scheme verification system modules and the functionality of each module.*

**Keywords** - OPAC, the system data verification OPAC, LMS, the theory of fuzzy sets.

### Вступ

Електронний каталог, як основний довідниково-інформаційний модуль бібліотеки, є складною метаінформаційною системою. Від якості даних, що містяться в електронному каталозі на пряму залежить якість інформаційно-пошукових послуг бібліотеки. З іншого боку, на сучасній стадії розвитку інформаційних технологій на перше місце виходить завдання об'єднання

Третя науково-практична конференція «Сучасні проблеми діяльності бібліотеки в умовах інформаційного суспільства»,  
29 вересня 2011 р, Львів, Україна

локальних інформаційних ресурсів в рамках потужних інтегрованих інформаційних систем. Відповідно до концепції «Державної цільової національно-культурної програми створення єдиної інформаційної бібліотечної системи "Бібліотека-XXI"»[1] в Україні функціонує понад 40 тис. бібліотек різного підпорядкування і форми власності. У той же час доступність цієї інформації залишається на дуже низькому рівні. Можливим шляхом вирішення даної проблеми є створення розподіленої системи зберігання електронних бібліотечних ресурсів на основі єдиного центру каталогізації. При створенні такої потужної інформаційної системи одною з основних проблем є проблема якості даних, відсутності єдиного підходу до оцінки та моніторингу наповненості електронного каталогу бібліотеки.

Аналіз можливостей сучасних автоматизованих бібліотечних інформаційних систем (АБІС) показав, що розробники недостатню увагу приділяють інструментам пошуку та корекції помилкових записів баз даних електронного каталогу. Отже, розробка підходів до проектування систем пошуку, оцінки та корекції помилок (систем верифікації даних) в даних електронних каталогів бібліотек є досить актуальною.

### **Виклад основного матеріалу**

Засоби введення даних в електронний каталог мають різну структуру та природу [2], тому при розробці підходів до проектування систем верифікації даних в електронних каталогах потрібно враховувати типологію методів створення записів.

Система верифікації даних електронного каталогу (СВДЕК) – набір програмних засобів для пошуку, оцінки, виправлення та уточнення помилкових даних в електронних каталогах бібліотек. Відповідно до двох основних функцій електронного каталогу: інформаційно-пошукової та інтеграційної, записи електронного каталогу повинні задовольняти ряд вимог:

Третя науково-практична конференція «Сучасні проблеми діяльності бібліотеки в умовах інформаційного суспільства»,  
29 вересня 2011 р, Львів, Україна

Відповідати національним стандартам опису бібліографічної інформації.

Відповідати міжнародним стандартам формату обміну бібліографічними записами (ISO2709, MARC-сімейства, DublinCore).

У повному обсязі і достовірно відображати всі бібліографічні властивості джерела-інформації.

Перші дві вимоги функціонально задовольняє система стандартизації та нормалізації даних. У більшості сучасних АБІС вона наявна у програмних засобах імпорту-експорту даних у структурі СУБД електронного каталогу. Третя вимога реалізовується за допомогою засобів верифікації інформації в електронному каталозі. На даний час не існує єдиного підходу до розробки систем верифікації інформації.

### **Основні вимоги до СВДЕК:**

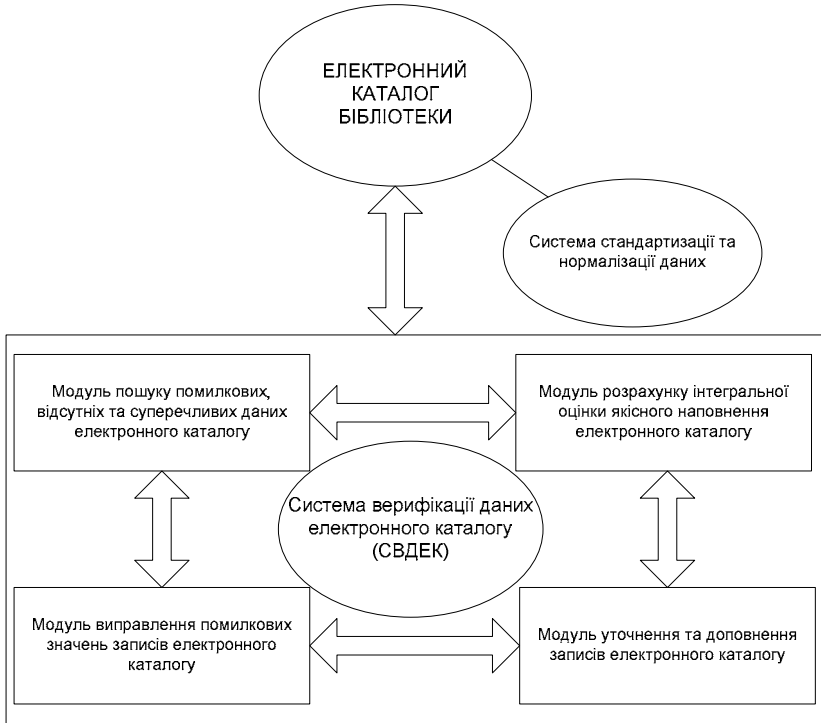
незалежність від апаратної платформи та СУБД, що використовується в АБІС;

наявність у складі програмних засобів для реалізації автоматичного пошуку всіх можливих недостовірних або неповних значень атрибутів запису в електронному каталозі;

наявність у складі програмних засобів для реалізації автоматичної та напівавтоматичної корекції та доповнення значень атрибутів запису в електронному каталозі;

наявність у складі програмних засобів для оцінки та моніторингу якісного наповнення електронного каталогу.

Отже, відповідно до функціональних вимог СВДЕК повинен складатись із 4-х модулів представлених на рисунку 1.



*Рис. 1. Функціональні модулі СВДЕК*

Модуль пошуку помилкових, відсутніх та суперечливих даних призначений для знаходження усіх можливих помилкових або відсутніх значень атрибутів запису. Модуль складається з 4-х основних програмних блоків:

- Блок пошуку відсутніх даних;
- Блок пошуку суперечливих даних;
- Блок пошуку помилкових даних;
- Блок он-лайн моніторингу.

*Блок пошуку відсутніх даних.* На основі перебору всіх записів по атрибутах знаходить записи, що не містять 1 або більше значень атрибутів. Звичайно, наявність значень деяких атрибутів

(зокрема назва та автор) є обов'язковими для однозначної ідентифікації джерела-інформації, тому при розрахунку інтегральної оцінки якісного наповнення каталогу це повинно бути враховано і відповідно кожному атрибуту надано певний ваговий коефіцієнт. Дані про записи, що містять відсутні значення атрибутів, передаються у модуль уточнення та доповнення записів каталогу.

*Блок пошуку суперечливих даних.* На основі певних закономірностей, як логічних (наприклад автор-рік видання, місце видання-назва видавництва, тощо), так і статистичних (за допомогою відповідних статистичних методів, що застосовуються до значень атрибутів усіх записів електронного каталогу) отримуються дані про записи, які мають суперечності у значеннях атрибутів. Відповідна інформація передається до модуля уточнення та доповнення записів. Окремою частиною блоку є програмний засіб контролю типу даних, значень атрибутів, якщо дана функція не реалізована у СУБД електронного каталогу.

*Блок пошуку помилкових даних.* Основним завданням даного блоку є пошук різного роду символічних спотворень у текстових значеннях атрибутів. На даний час існує клас ефективних алгоритмів для аналізу текстової інформації [3]. Окремим завданням у даному блоці є пошуку орфографічних помилок. Для даних потреб доступні програмні засоби на основі двох загальновідомих методів пошуку орфографічних помилок (словник та n-грама). Однак дані методи не є достатньо ефективні з точки зору часу опрацювання великих масивів текстових даних. Також не врахована особливість різноманітності засобів вводу даних в електронний каталог. Тому розвиток та застосування ідей теорії нечітких множин для пошуку та корекції орфографічних помилок [4] є досить перспективним. Відповідні дані про знайдені помилкові значення передаються у модуль виправлення помилкових значень.

Окреме місце у даному модулі займає блок *он-лайн моніторингу*, що на основі статистичної (пошуку записів з 0-показником користувацьких запитів) та користувацької (електронних повідомлень користувачів про знайдені помилки у процесі використання електронного каталогу) інформації передає «підозрілі» дані до відповідного модуля СВДЕК для перевірки.

Модуль уточнення та доповнення записів призначений для аналізу, перевірки та заповнення значень атрибутів електронного каталогу, що були промарковані модулем пошуку помилкових, відсутніх та суперечливих даних. Звичайно, аналіз, перевірка, а тим паче доповнення записів електронного каталогу, не може відбуватись повністю автоматично, без відповідальної особи. Однак, деякі функції модуля можливо автоматизувати. Модуль складається з 2-х блоків:

Блок доповнення даних;

Блок уточнення даних.

*Блок доповнення даних.* На основі даних, отриманих з блоку пошуку відсутніх даних, відбувається аналіз можливості автоматичного доповнення NULL-значень відповідних атрибутів запису електронного каталогу. Доповнення запису можливе: на основі автоматичної генерації значення (наприклад, значення атрибуту «Авторський знак» чи індексів УДК та ББК можливо присвоїти на основі значення поля «Автор» або переліку ключових слів джерела-інформації); на основі запозичення відповідних значень атрибутів з інших джерел (інший бібліографічний опис, перелік літературних джерел, база даних видавництва, Інтернет, тощо). У другому випадку для реалізації запозичення формулюється пошуковий запит і в залежності від джерела запозичення застосовується відповідний програмний синтаксичний аналіз (parsing) з відокремленням необхідного значення атрибуту.

*Блок уточнення даних.* Основною функцією даного блоку є перевірка та коригування даних блоку пошуку суперечностей.

Звичайно, основну роль тут відіграє перевірка людиною «підозрілих» даних. Однак застосування методу запозичень спрощує дану процедуру.

Модуль виправлення помилкових значень призначений для автоматичного коригування символічних спотворень та опрацювання орфографічних помилок. Методи та алгоритми, що застосовуються в даному модулі тісно пов'язані з блоком пошуку помилкових значень. Дані питання добре висвітлені у працях таких вчених, як Бабко-Малая О.Б., Гельбух А.Ф., Damerau F.J., Левенштейн В.И., Nielsen R., Ballard T., Вершинин М.И. та інші. Однак при розробці даного модуля обов'язково потрібно враховувати принцип персоналізації введення записів в електронний каталог та технічні характеристики потоково-карткових сканерів при ретроспективній конверсії.

Модуль розрахунку інтегральної оцінки якісного наповнення електронного каталогу. Функції даного модуля полягають у моніторингу якісного стану даних у електронному каталозі бібліотеки. Даний модуль призначений для розрахунку, як узагальненої (інтегральної) оцінки всього каталогу, так і оцінки якості окремих записів (кортежів) електронного каталогу. Метод оцінювання полягає в урахуванні всіх типів помилок (тип помилки, можливість автоматичного виправлення, можливість виправлення без наявності оригінала джерела-інформації, тощо) в значеннях атрибутів електронного каталогу, так і ролі даного атрибуту для забезпечення основних функцій електронним каталогом (вагові коефіцієнти кожного атрибуту бібліографічного запису). Дана оцінка може бути застосовна для об'єктивного оцінювання стану електронного каталогу при фінансових операціях та у процесах створення корпоративних електронних каталогів міста, району, області, тощо.

### Висновки

Отже, запропонована модель СВДЕК повністю забезпечує електронний каталог усім потрібним програмним інструментарієм для верифікації даних. Звичайно, повністю автоматизувати всі процеси верифікації інформації в електронних каталогах бібліотек неможливо, особливо в процесах редагування записів потрібне втручання відповідальної особи. Однак представлена система на порядок полегшить роботу коректорів та каталогізаторів у бібліотеках та дасть можливість у майбутньому створювати великі корпоративні електронні каталоги з високим якісним наповненням інформацією.

### Список використаних джерел:

Концепція «Державної цільової національно-культурної програми створення єдиної інформаційної бібліотечної системи "Бібліотека - XXI"».- Режим доступу: [http://www.nbuv.gov.ua/law/09\\_bib\\_21.html](http://www.nbuv.gov.ua/law/09_bib_21.html)

Ярмолюк Р.С. Основні типи та джерела помилок у записах електронного каталогу / Р.С. Ярмолюк // Вісник Національного Університету «Львівська політехніка», - 2010. - № 689: Інформаційні системи та мережі. – С. 348-357.

Ярмолюк Р.С. Задача аналізу текстових атрибутів в електронному каталозі / Р.С. Ярмолюк // Вісник Хмельницького Національного Університету. Серія Технічні науки. - 2011. - № 3 – С.225-228.

Вершинин М.И. Электронный каталог проблемы и решения / М.И. Вершинин. – СПб. : Профессия, 2007. – 233с.