

## **ПРОГРАМНА РЕАЛІЗАЦІЯ СИСТЕМИ ПЕРЕВІРКИ ОРФОГРАФІЇ В БІБЛІОГРАФІЧНИХ БАЗАХ ДАНИХ**

*Основною метою цього дослідження є програмна реалізація системи перевірки орфографії в бібліографічних базах даних на основі розроблених підходів до проектування підсистеми перевірки орфографії на основі технології перевірки орфографії за словником Hunspell.*

*The main purpose of this study is programmatic implementation of spelling bibliographic databases based on the developed approaches to the design of the subsystem Spelling technology-based spell checking dictionary for Hunspell.*

### **Постановка проблеми**

Система електронного каталогу має важливе значення у забезпеченні основних функцій сучасної бібліотеки. Від якості даних, що містяться в електронному каталозі, безпосередньо залежить якість інформаційно-пошукових та інтеграційних можливостей бібліотеки. Під час роботи з даними електронного каталогу бібліотеки можуть виникнути виняткові ситуації, що призводять до появи різного роду помилок. Основні атрибути запису електронного каталогу (автор, назва, видавництво, анотація джерела) мають текстовий тип. Відповідно, основними помилками в текстових атрибутах є символні спотворення або орфографічні помилки, що зумовлені як об'єктивними, так і суб'єктивними чинниками. Від того, наскільки ефективного відбувається пошук та корекція орфографічних помилок в текстових атрибутах записів електронного каталогу, залежить якість вихідної інформації при пошукових запитах [1].

### **Формулювання цілей та мети дослідження**

Основною метою даного дослідження є розробка програмного продукту, який забезпечить можливість перевірки орфографії за трьома словника Hunspell (українська, російська, англійська).

### **Виклад основних матеріалів дослідження**

Бібліографічна база даних – це база даних, що повністю складається з бібліографічних описів. Бібліографічний опис – це записана за певними правилами сукупність бібліографічних даних, що дають можливість ідентифікувати документ. У свою чергу,

бібліографічні дані – це конкретні відомості про назву, автора твору, місце і рік видання та багато інших.

### Можливості програмного продукту

Після запуску програми потрібно вказати шлях до бази даних (Access), в іншому випадку програма завершить свою роботу. Після аналізу бази даних, програма заповнює меню випадаючого списку назвами таблиць, що містяться в ній (рисунок 1).

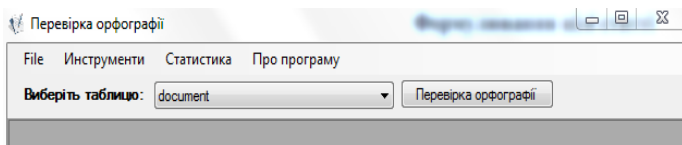


Рисунок 1. Головне меню програми

Перед початком перевірки орфографії можна сформувати кортежну, атрибутну статистику, попередньо вибрати жоду перевірки. Також загальну статистику, яка відображає кількість пустих та оригінальних значень. Усі дані можна зберегти в форматі “.xls”. Наприклад, на рисунку 2 відображено результат атрибутної статистики.

	Назва атрибуту	Кількість помилок в кортежах	Кількість помилкових кортежів
1	doc_type	0	0
2	name	46722	24049
3	publisher	80370	33090
4	reg_date	0	0
5	publ_place	51970	51306
6	in_date	0	0
7	attr	0	0
8	doc_id	0	0

Рисунок 2. Атрибутна статистика

На рисунку 3 відображається результат кортежної статистики.

	Процент	Формально	Кількість
1	0	Помилки відсутні	0
2	3	1 помилки з 30	1
3	6	2 помилки з 30	474
4	9	3 помилки з 30	8767
5	12	4 помилки з 30	20123
6	15	5 помилки з 30	21021
7	18	6 помилки з 30	18713
8	21	7 помилки з 30	10545

Рисунок 3. Кортежна статистика

На рисунку 4 відображається статистика, яка відображає кількість пустих та оригінальних значень.

	Назва атрибуту	Кількість пустих значень	Кількість оригінальних значень
13	in_user	6	138
14	name_prefix	8	41216
15	last_edit_user	19	118
16	item_qty	362	346
17	item_present	362	346
18	sizem	25771	10261

Рисунок 4. Загальна статистика

Перед початком перевірки орфографії програма запропонує вибрати унікальний атрибут таблиці, який в подальшому буде потрібний для внесення змін до бази даних. Після чого потрібно вказати атрибути, які потрібно відобразити в сітку відображення даних і відповідно відобразити їх. Приклад відображених даних зображено на рисунку 5.

Вибрати атрибути для відображення:

- isbn
- lang
- name
- item\_present
- last\_edit\_dat
- name\_prefix
- item\_qty
- last\_edit\_use
- parent\_id

Відобразити вміст

Вибрати мову перевірки: англійська Почати перевірку

doc_id	lang	name
16	<Не визначено>	Materia Medica. Гоме...
15	Українська	Основи архітектурної ...
16	Українська	Естетика
17	Українська	Зовнішньо-економічна...
18	Українська	Мікроелектроніка

Рисунок 5 Дані для перевірки

З рисунку видно, що унікальний атрибут завжди відображається. Після чого потрібно вказати мову, за якою будуть перевірятись дані. Результат перевірки даних за словником Hunspell (українська), зображено на рисунку 6.

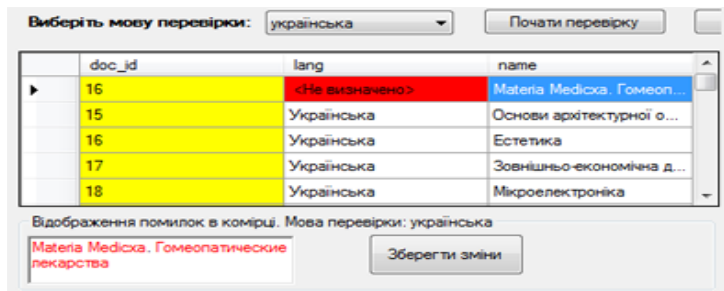


Рисунок 6. Результат перевірки за словником Hunspell (українська)

Після перевірки, комірки, в яких виявлені помилки заливуються червоним кольором. При виборі конкретної комірки в текстовому полі відображається вміст цієї комірки та помилки, які присутні в ній. Дане текстове поле можна редагувати та вносити зміни до бази даних (кнопка зберегти зміни).

Для зручності редагування в програмі присутня функція скриття правильних кортежів. Також є функція сортування даних.

### Висновки

Отже, дана наведена програма дає можливість перевірки бібліографічних баз даних за трьома словниками (англійський, російський, український) та можливості внесення змін до бази даних. Звичайно, теперішня реалізація внесення змін до бази даних не є універсальною, оскільки потрібно запропонувати користувачу можливо-правильний варіант, який потрібно внести до бази даних. Також можливість формування атрибутної, кортежної та загальної статистики.

### Література

Ярмолюк Р. С. Підсистема перевірки орфографії електронного каталогу бібліотеки на основі технології HUNSPELL / Р. С. Ярмолюк // Вісник Національного Університету «Львівська політехніка», - 2013. - № 743: Інформаційні системи та мережі. – С. 219-230.