

## АЛГОРИТМИ ІНТЕЛЕКТУАЛЬНОГО ПОШУКУ В СЕРВІСІ ДЛЯ ОНЛАЙН-БІБЛІОТЕКИ

Шевченко К.Г.

Керівник: Беседіна С.В.

*E-mail: pochta13524@gmail.com*

*Черкаси, Черкаський національний університет імені Богдана Хмельницького*

У час бурхливого розвитку ІТ-технологій, створено багато різномітних систем, у тому числі і для онлайн-бібліотек, які допомагають знайти книгу для читання за деяким заданим параметром. Однак це і є найбільшим недоліком таких систем, оскільки в них застосовуються стандартні алгоритми пошуку – за назвою книги, за ім'ям автора, за входженням слова або частини слова до назви книги тощо. Тому актуальним залишається створення сервісу для онлайн-бібліотеки з елементами інтелектуального пошуку.

Було розроблено сервіс для онлайн-бібліотеки, який застосовує алгоритми інтелектуального пошуку із використанням засобів PostgreSQL для повнотекстового пошуку та створення таблиці синонімів для пошуку за синонімічними словами у пошуковому запиті.

Повнотекстовий пошук – автоматизований пошук документів, який здійснюється не за назвами документів, а за їх вмістом. У системі керування базою даних PostgreSQL для цього використовуються процедури `to_tsvector` та `to_tsquery`. Процедура `to_tsvector` застосовується для нормалізації рядка, за яким здійснюватиметься пошук. Нормалізація рядка означає видалення стоп-слів, вирізання закінчень слів, приведення запиту до нижнього регістру тощо. Першим аргументом процедури вказується мова пошукового запиту. Другим аргументом процедури вказується поле таблиці бази даних, в якому буде здійснюватись пошук. Процедура `to_tsquery` також має два аргументи – першим вказується мова пошукового запиту, а другим аргументом є власне сам пошуковий запит. Для зіставлення результатів цих процедур застосовується оператор `@@` [1].

Також для пошуку в реалізованому сервісі використовується пошук з використанням таблиці синонімів. Для реалізації такого пошуку було створено та заповнено даними дві таблиці в базі даних – власне таблиця слів та таблиця синонімів. Такий підхід до обробки пошукового запиту дає кращі результати в комбінації з алгоритмом повнотекстового пошуку засобами PostgreSQL, так як під час пошуку враховуються не лише слова з пошукового запиту, але й синоніми до нього, що розширює відповідні до запиту результати пошуку.

У перспективі до вже застосованих та розроблених алгоритмів у сервісі можна додати такий алгоритм нечіткого пошуку, як алгоритм на основі знаходження відстані Левенштейна. Він є найбільш поширеною метрикою, яка застосовується в алгоритмах нечіткого пошуку і дає змогу правильно обробити навіть ті пошукові запити, які містять помилки. Відстань Левенштейна (редакційна відстань, дистанція редагування) – це мінімальна кількість операцій вставки або видалення одного символу та заміни його на інший, що необхідна для перетворення одного рядка в інший [2].

Отже, було проаналізовано та розглянуто основні алгоритми інтелектуального пошуку на прикладі створеного сервісу для онлайн-бібліотеки, які можна застосувати при розробці і інших систем. Застосування та реалізація алгоритмів інтелектуального пошуку дає можливість здійснювати в будь-яких системах більш гнучкий пошук інформації, що надає перевагу таким системам над іншими аналогічними системами, які застосовують стандартні алгоритми пошуку.

### Література

[1] Основы полнотекстового поиска в PostgreSQL [Електронний ресурс]. – Режим доступу до ресурсу: <https://eax.me/postgresql-full-text-search/>

[2] Відстань Левенштейна [Електронний ресурс]. – Режим доступу до ресурсу: [https://uk.wikipedia.org/wiki/Відстань\\_Левенштейна](https://uk.wikipedia.org/wiki/Відстань_Левенштейна)