

*Огороднік М. О., здобувач вищої освіти,  
Римар П. В., старший викладач кафедри інформаційних технологій,  
Донецький національний університет імені Василя Стуса*

## **РОЗРОБКА ІНФОРМАЦІЙНОЇ СИСТЕМИ З ПІДТРИМКИ ІНДЕКСУВАННЯ ТА ПОШУКУ ІНФОРМАЦІЇ**

*Анотація. Розроблена інформаційна система підтримує індексування та пошук документів, використовуючи C#, .NET та PostgreSQL. Система забезпечує автоматичне створення індексів, швидкий пошук та ефективне керування великими обсягами даних.*

*Ключові слова: індексування, пошук інформації, C#, .NET, PostgreSQL, інформаційна система.*

**Вступ.** Процес індексування та пошуку інформації є одним з фундаментальних елементів сучасних інформаційних систем, що знаходять застосування в різноманітних галузях – від науки й техніки до бізнесу та щоденного життя. Зі зростанням кількості даних та різноманіттям їх джерел інформаційні системи повинні забезпечувати користувачам можливість швидкого доступу до необхідної інформації, що вимагає оптимізації процесів індексації та пошуку. В умовах сучасної цифрової ери, коли інформаційні ресурси збільшуються в геометричній прогресії, ефективно управління цими процесами стає стратегічно важливим завданням. Це особливо важливо для сфер наукових досліджень, бізнес-аналітики, управління даними в організаціях та іншими галузями, де швидкий доступ до релевантної інформації може визначати успіх діяльності [1].

Індексація інформації передбачає організацію даних у спосіб, що дає змогу швидко знайти потрібну інформацію за певними критеріями. Цей процес включає аналіз та структурування інформації, створення індексів, які полегшують процес пошуку. Пошук інформації використовує ці індекси для швидкого виявлення релевантних документів або даних на основі заданих запитів [2]. Процес індексації включає кілька основних етапів: збір даних, аналіз даних, створення індексів та підтримку актуальності індексів.

1. Збір даних. На першому етапі відбувається збір даних з різних джерел. Це можуть бути тексти, зображення, відео, аудіо та інші типи даних. Джерелами можуть бути вебсторінки, бази даних, документи, архіви, соціальні мережі та інші джерела інформації. Важливо забезпечити повноту та точність зібраних даних, оскільки від цього залежить ефективність подальших етапів індексації.

2. Аналіз даних. Після збору даних відбувається їх аналіз та підготовка до індексації. На цьому етапі здійснюється попередня обробка даних, яка включає видалення зайвих символів, нормалізацію тексту, видалення стоп-слів та інших елементів, які не несуть значущої інформації. Аналіз даних також може включати розпізнавання мов, визначення ключових слів, семантичний аналіз та інші методи обробки даних.

3. Створення індексів. Основний етап індексації включає створення індексів, які представляють структуру даних у вигляді, що дає змогу швидко знайти

необхідну інформацію. Це можуть бути інвертовані індекси, векторні моделі, деревоподібні структури даних та інші методи. Важливо забезпечити оптимальне співвідношення між швидкістю пошуку та обсягом індексів.

4. Підтримка актуальності індексів. Після створення індексів важливо забезпечити їх актуальність та оновлення у разі зміни даних. Це включає періодичне оновлення індексів, додавання нових даних, видалення застарілих даних та інші методи підтримки актуальності [3].

У процесі розробки інформаційної системи з підтримки індексування та пошуку інформації були використані сучасні технології, зокрема C#, .NET та база даних PostgreSQL (рис. 1).

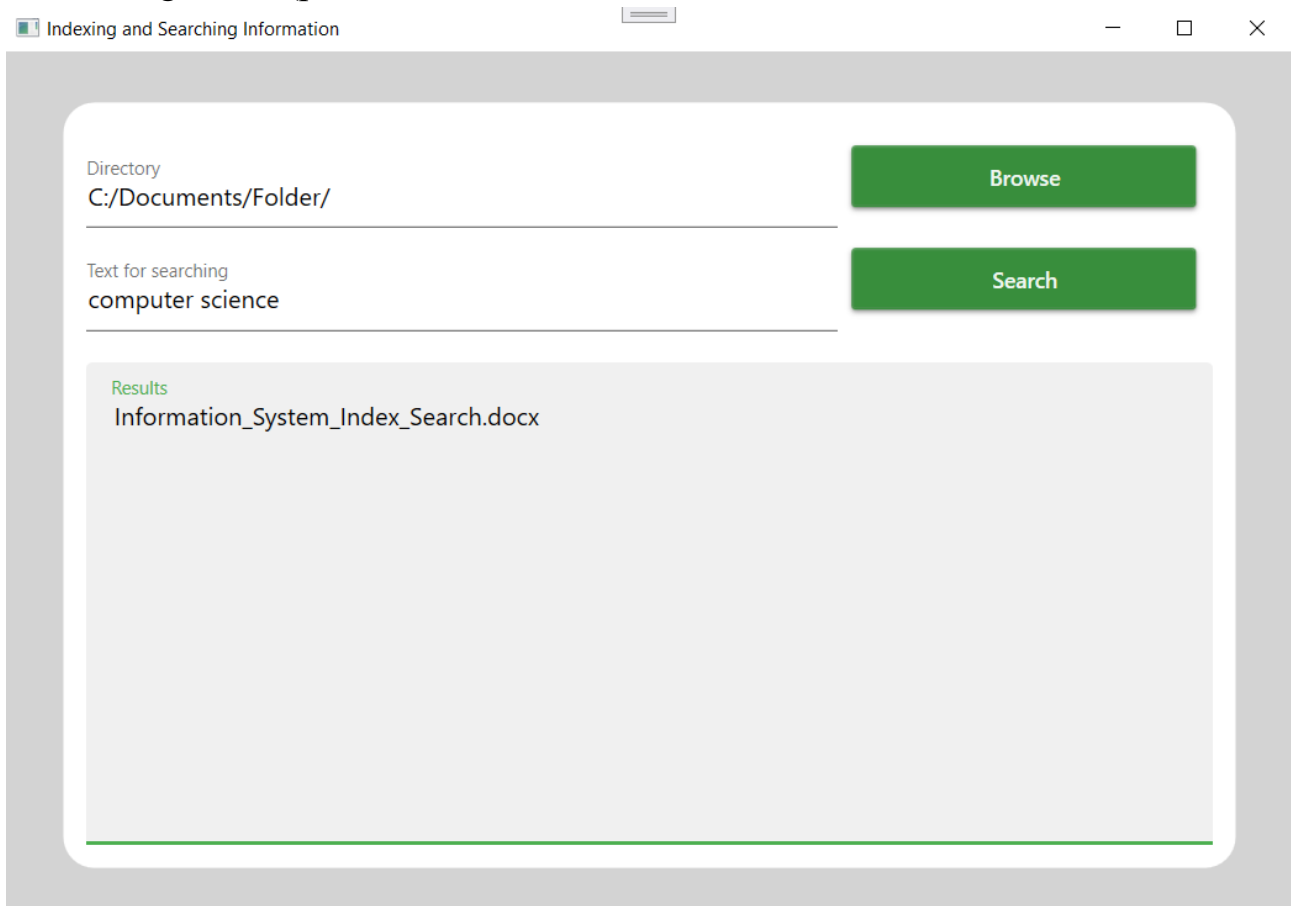


Рисунок 1 – Інтерфейс інформаційної системи з індексування та пошуку інформації

Розроблена система забезпечує ефективне зберігання, обробку та швидкий пошук інформаційних об'єктів, які індексуються під час завантаження нових файлів або документів. Основний функціонал системи включає такі процеси:

1. Індексція даних – автоматичне створення індексів для завантажених файлів, що дає змогу швидко знаходити необхідну інформацію на основі ключових слів або метаданих.

2. Пошук за індексами – система дає змогу виконувати високошвидкісний пошук за раніше проіндексованими документами.

3. Відображення результатів – результати пошуку відображаються у вигляді списку. Користувачі можуть отримати повний доступ до потрібної інформації.

Особливістю системи є інтеграція з базою даних PostgreSQL, що дає змогу зберігати великі обсяги даних і ефективно управляти ними. Завдяки архітектурі на основі .NET система є високопродуктивною та стійкою до великих навантажень. Унікальність системи полягає у поєднанні зручного інтерфейсу для користувача та потужних алгоритмів обробки даних, що забезпечує високий рівень продуктивності і точності пошуку. До того ж система підтримує можливість масштабування для обробки ще більших обсягів інформації в майбутньому.

**Висновки.** Розроблена інформаційна система з підтримки індексування та пошуку інформації забезпечує ефективну обробку та пошук великих обсягів даних, що є важливим аспектом в умовах стрімкого зростання інформаційних ресурсів у сучасному світі. Використання технологій C#, .NET та PostgreSQL дало змогу створити стійке, продуктивне та масштабоване рішення, яке може бути застосоване у різних сферах, як-от наукові дослідження, бізнес-аналітика та управління інформаційними потоками в організаціях. Завдяки використанню сучасних алгоритмів та технологій система демонструє високу продуктивність та точність під час пошуку інформації, що робить її надійним інструментом для роботи з великими інформаційними масивами. Отже, розроблена система є важливим інструментом для забезпечення швидкого та ефективного доступу до інформації, що відповідає сучасним вимогам до інформаційних технологій в умовах постійного зростання обсягів даних.

#### Список використаних джерел

1. What is Document Management (DMS)? *aiim*. URL: <https://www.aiim.org/what-is-document-imaging> (дата звернення: 22.10.2024).
2. Document Indexing Explained. Types, Steps & Outsourcing. *Quantanite*. URL: <https://www.quantanite.com/blog/document-indexing-explained-types-steps-outsourcing/> (дата звернення: 22.10.2024).
3. Document Indexing. *Nanonets*. URL: <https://nanonets.com/blog/document-indexing/> (дата звернення: 22.10.2024).