

УДК 004.41

М. Солтис; Д. Михалик, к. т. н., доц.

(Тернопільський національний технічний університет імені Івана Пулюя, Україна)

ВЕБСИСТЕМА ЦЕНТРАЛІЗОВАНОГО ЗБЕРІГАННЯ ФАЙЛІВ ІЗ ПІДТРИМКОЮ СЕМАНТИЧНОГО ПОШУКУ

UDC 004.41

M. Soltys; D. Mykhalyk, PhD., Assoc. Prof.

WEB SYSTEM FOR CENTRALIZED FILE STORAGE WITH SEMANTIC SEARCH SUPPORT

Ключові слова: семантичний пошук, централізоване зберігання даних, вебсистема, індексація файлів, векторні ембедінги, ElasticSearch.

Key words: semantic search, centralized data storage, web system, file indexing, vector embeddings, ElasticSearch.

Моя науково-дослідна робота присвячена розробці вебсистеми для централізованого зберігання файлів із підтримкою семантичного пошуку. Основною метою є створення інтелектуальної платформи, яка не лише зберігає дані, а й забезпечує їх змістовне пошукове опрацювання за допомогою сучасних алгоритмів векторного представлення тексту. Така система значно підвищує ефективність взаємодії користувачів із великими обсягами інформації, дозволяючи шукати файли не лише за назвою, а й за змістом.

На початковому етапі проводиться аналіз потреб користувачів та вимог до структури даних, визначаються ключові сценарії взаємодії із системою. На основі цього формується архітектура, яка включає сервіс обробки файлів, модуль індексації контенту та клієнтську частину з інтуїтивно зрозумілим інтерфейсом. Для клієнтської частини використовується фреймворк **Next.js**, тоді як серверна логіка реалізована на **NestJS** з використанням **TypeScript** та **PostgreSQL**.

Особлива увага приділяється реалізації семантичного пошуку. Для цього застосовується **ElasticSearch**, який зберігає індекси файлів і підтримує як повнотекстовий, так і векторний пошук. Векторизація вмісту виконується за допомогою **OpenAI Embeddings**, що забезпечує розуміння змістового контексту документів. Такий підхід дає можливість знаходити релевантні файли навіть тоді, коли користувач не знає точну назву документа або вводить запит у довільній формі.

Під час розробки системи значна увага приділяється безпеці даних, включаючи авторизацію на основі JWT-токенів, контроль доступу до файлів та обмеження прав для зовнішніх посилань на завантаження. Також реалізуються інструменти обробки файлів різних форматів (PDF, DOCX, зображення), що дозволяє індексувати текст із більшості поширених типів даних.

Завершальним етапом є тестування сервісу, яке включає перевірку коректності індексації, точності семантичного пошуку, швидкодії, масштабованості та стійкості системи до великих навантажень. Тестування дозволяє оптимізувати роботу алгоритмів та гарантує стабільність системи в реальних умовах.

Результати проведеної роботи свідчать про те, що поєднання централізованого сховища з інформаційним пошуком на основі векторних представлень значно підвищує ефективність роботи з файлами та відкриває широкі можливості для подальших досліджень у сфері інтелектуальних інформаційних систем.

Література

1. Effective Strategies for Managing Network-Attached Storage Systems [Електронний ресурс] – Режим доступу до ресурсу: <https://www.linkedin.com/advice/3/what-best-practices-managing-network-storage-qcpze>.
2. Data Security in Cloud Computing [Електронний ресурс] – Режим доступу до ресурсу: <https://cloud.google.com/learn/what-is-cloud-data-security>.