

УДК 004.8:004.65

Прасол Н.С., здобувач,

Марчук Г.В., ст. викл.

Державний університет «Житомирська політехніка»

СЕМАНТИЧНИЙ ПОШУК ВАКАНСІЙ У ВЕБПОРТАЛАХ ПРАЦЕВЛАШТУВАННЯ НА ОСНОВІ AUGMENTED GENERATION

Сучасні вебпортали працевлаштування досі переважно використовують класичний пошук за ключовими словами. Такий підхід не враховує синоніми, контекст і семантичну близькість понять. У результаті кандидати пропускають вакансії, а роботодавці – кваліфікованих спеціалістів.

Щоб вирішити цю проблему, пропонується впровадження семантичного пошуку з використанням архітектури RAG. Архітектура Retrieval-Augmented Generation включає векторний пошук релевантних документів з генерацією відповіді за допомогою великої мовної моделі. На прикладі вебпорталу працевлаштування, користувач вводить запит українською або англійською мовами, система автоматично перетворює його на векторне представлення, знаходить семантично близькі вакансії в базі даних, додає їх до контексту промпту і генерує персоналізовану відповідь зі списком вакансій та поясненнями.

У межах MERN-проекту пропонується інтеграція RAG з використанням вбудованих можливостей MongoDB Atlas Vector Search [3], що дозволяє уникнути зовнішніх сервісів подібних до Pinecone [1]. На першому етапі до схеми Job додається поле embedding типу Array з векторним індексом. При створенні або оновленні вакансії на сервері генерується ембединг опису за допомогою відкритої моделі через API OpenAI [2]. Після цього налаштовується векторний індекс, який дозволяє знаходити вакансії за ступенем їхньої змістовної подібності. Коли користувач вводить свій запит, система створює таке ж математичне представлення запиту і виконує пошук, повертаючи найбільш близькі за змістом вакансії. Знайдені дані форматуються та додаються до промпту мовної моделі, яка генерує список вакансій із коротким описом до кожної. Такий підхід повністю інтегрується в існуючу серверну частину на Express та не вимагає значних змін у клієнтській частині вебзастосунку.

На рис. 1 представлена схема архітектури RAG для семантичного пошуку вакансій в проєкті.

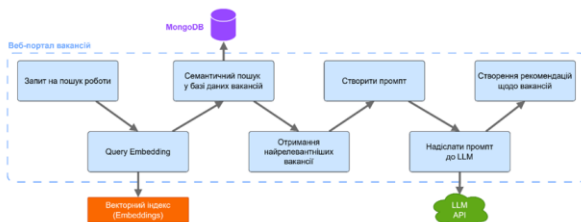


Рис.1. Архітектура RAG для семантичного пошуку вакансій

Запропонована архітектура має низку важливих переваг. По-перше, полегується інтерфейс, оскільки користувачу більше немає потреби заповнювати блоки фільтрації. По-друге, час відгуку залишається меншим за 200 мс для тисяч оголошень про роботу завдяки використанню векторного пошуку MongoDB Atlas [3]. По-третє, система масштабується без витрат на зовнішню векторну базу даних.

Водночас існують певні обмеження. По-перше, при додаванні нових вакансій потрібно додатково обробляти їхній текст, що вимагає додаткових обчислень на сервері. По-друге, відповіді мовної моделі іноді можуть трохи відрізнитися при повторних запитах, тому необхідно витратити час на налаштування моделі так, щоб її результати були максимально стабільними. По-третє, якщо кількість вакансій перевищить 50 тисяч, доцільно застосувати комбінований пошук - одночасно за змістовною подібністю та звичайними ключовими словами.

Таким чином, впровадження семантичного пошуку з використанням RAG перетворює звичайний MERN-портал працевлаштування на сучасну інтелектуальну платформу. Технологія, інтегрується в наявний проект завдяки вбудованому векторному пошуку MongoDB. Запропонований підхід підвищить точність підбору вакансій та покращить користувацький досвід.

Список використаних джерел

1. Pinecone - What is a Vector Database? [Електронний ресурс] / Pinecone Systems Inc. – 2025. – Режим доступу до ресурсу: <https://www.pinecone.io/learn/vector-database/>.

2. OpenAI API Reference - Embeddings [Електронний ресурс] / OpenAI. – 2026. – Режим доступу до ресурсу: <https://platform.openai.com/docs/api-reference/embeddings>.

3. MongoDB Atlas Vector Search Overview [Електронний ресурс] / MongoDB Inc. – 2026. – Режим доступу до ресурсу: <https://www.mongodb.com/docs/atlas/atlas-vector-search/vector-search-overview/>