УДК 004.89:004.912:331.5

# ІНТЕЛЕКТУАЛЬНА СИСТЕМА ДВОБІЧНОГО ПІДБОРУ ВАКАНСІЙ І КАНДИДАТІВ НА ОСНОВІ СЕМАНТИЧНОГО АНАЛІЗУ ТА ГЕНЕРАТИВНОГО ШІ

Сінгх А.А.[1]; Новікова Т.О.[2]
[1,2] Національний університет «Одеська політехніка»
[1,2] Україна, Одеса
[1] 7405629@stud.op.edu.ua; [2] novikova.t.o@op.edu.ua

***Анотація.*** *Розглянуто проєктування вебсистеми двобічного підбору вакансій і кандидатів на основі семантичного аналізу тексту та генеративного ШІ. Запропоновано гібридну архітектуру, алгоритм оцінювання відповідності резюме вакансії, модель даних і механізм перевірки фактів у супровідному листі. Результат — проєктна специфікація, готова до реалізації та оцінювання.*
***Ключові слова:*** *семантичний пошук; підбір персоналу; резюме; векторні представлення; генеративний штучний інтелект.*

**Вступ.** Онлайн-рекрутинг передбачає взаємодію двох сторін: кандидат шукає вакансію, що відповідає його компетентностям, а роботодавець — фахівця, який відповідає вимогам посади. Традиційний пошук за ключовими словами не завжди враховує синонімію та контекст, а ручне порівняння резюме з описами вакансій потребує значного часу. Семантичні векторні представлення дають змогу порівнювати тексти за змістовою близькістю, а генеративні мовні моделі — формувати структуровані дані й чернетки текстів [1, 2].

Використання ШІ в працевлаштуванні водночас створює ризики: модель може сформулювати непідтверджене твердження про досвід кандидата, а автоматизоване оцінювання може впливати на можливості працевлаштування. GDPR визначає принципи мінімізації персональних даних і права на їх видалення, а Регламент ЄС про штучний інтелект окремо охоплює низку систем ШІ у сфері зайнятості [3, 4]. Тому система має забезпечувати пояснюваність рекомендацій, обмеження обробки даних і обов'язкову участь людини.

**Мета дослідження.** Спроєктувати вебсистему двобічного підбору вакансій і кандидатів, яка використовує семантичне порівняння, формує пояснення відповідності та генерує супровідний лист із перевіркою його тверджень за даними профілю.

**Основна частина.** У межах проєктування проаналізовано предметну область онлайн-рекрутингу та публічні описи п'яти аналогів: Djinni, Work.ua, robota.ua, LinkedIn Hiring Assistant і Jobscan. Порівняння виявило, що окремі продукти пропонують пошук, рекомендації або генерацію супровідних листів, однак у розглянутих описах не простежується повний набір функцій: двобічний підбір, пояснення збігу та перевірка фактів згенерованого листа. Це спостереження стосується саме проаналізованих публічних описів, а не доводить відсутності таких можливостей у продуктах.

Для семантичного подання текстів обрано multilingual-e5-small — модель векторних представлень, що формує 384-вимірні ембеддинги [1]. Структуроване вилучення навичок, досвіду, освіти й мов із резюме та описів вакансій, а також генерацію пояснень і супровідних листів передбачено виконувати через Gemini API з використанням структурованого виводу [2]. Розроблення спроєктовано на Python 3.11 і FastAPI; Firebase Authentication використовується для автентифікації, Cloud Firestore — для зберігання профілів, вакансій, збігів, листів і подій [5].

Архітектура є гібридною: вебклієнт взаємодіє з REST API, сервер перевіряє Firebase ID-токен і роль користувача, після чого викликає відповідні сервіси. Доступ до Firestore здійснюється лише з серверної частини через Firebase Admin SDK. Резюме обробляється як

PDF із текстовим шаром; контактні дані маскуються, а оригінальний файл і його повний текст не передбачено зберігати в базі. Такий підхід зменшує обсяг персональних даних.

Запропоновано алгоритм ранжування, який поєднує семантичну близькість текстів і покриття навичок. Нормалізована семантична складова визначається за формулою:

$$sem = \min(\max((cos - 0{,}70) / 0{,}30; 0); 1), \quad (1)$$

де *cos* — косинусна подібність векторів профілю кандидата та вакансії. Частка покритих вимог обчислюється як $skill = |R \cap S| / |R|$, де $R$ — множина потрібних навичок вакансії, $S$ — множина навичок кандидата. Підсумкова оцінка має вигляд:

$$score = 100 \cdot (0{,}6 \cdot sem + 0{,}4 \cdot skill) \cdot k_{exp}, \quad (2)$$

де $k_{exp}$ дорівнює 1, якщо досвід кандидата не менший за мінімально потрібний, і 0,9 — в іншому разі. Ваги семантичної та навичкової складових, а також поріг 0,70 винесено в конфігурацію для подальшого уточнення за результатами оцінювання. Для пошуку кандидатів алгоритм застосовується симетрично до профілів із дозволеною видимістю.

Для пояснення результату система зберігає не лише загальний відсоток, а й окремі складові оцінки, перелік збіжних і відсутніх навичок. Топ-10 рекомендацій попередньо обчислюється та зберігається, що дає змогу показувати список без повторного виклику мовної моделі під час кожного перегляду. Векторний індекс активних вакансій і профілів передбачено підтримувати в пам’яті сервера з інкрементальним оновленням.

Окремим компонентом є шлюз перевірки фактів супровідного листа. Він перевіряє числові значення, назви технологій і власні назви, зіставляючи їх із даними профілю та вакансії. Якщо знайдено непідтверджені елементи, система до двох разів запитує повторну генерацію з переліком заборонених тверджень. Якщо невідповідності залишаються, чернетка зберігається з ознакою verified = false, а сумнівні фрагменти позначаються для перевірки користувачем. Детермінований шлюз не гарантує виявлення всіх семантичних вигадок, тому остаточний перегляд людиною є обов’язковим.

Модель даних Firestore містить сім колекцій: users, seekerProfiles, employerProfiles, vacancies, matches, coverLetters та events (рис. 1). Передбачено шість основних екранів для входу й реєстрації, рекомендацій, деталей вакансії та листа, профілю кандидата, форми вакансії й переліку кандидатів. REST API охоплює реєстрацію профілю, завантаження резюме, пошук і перегляд вакансій, генерацію листа, керування вакансіями та перегляд кандидатів.

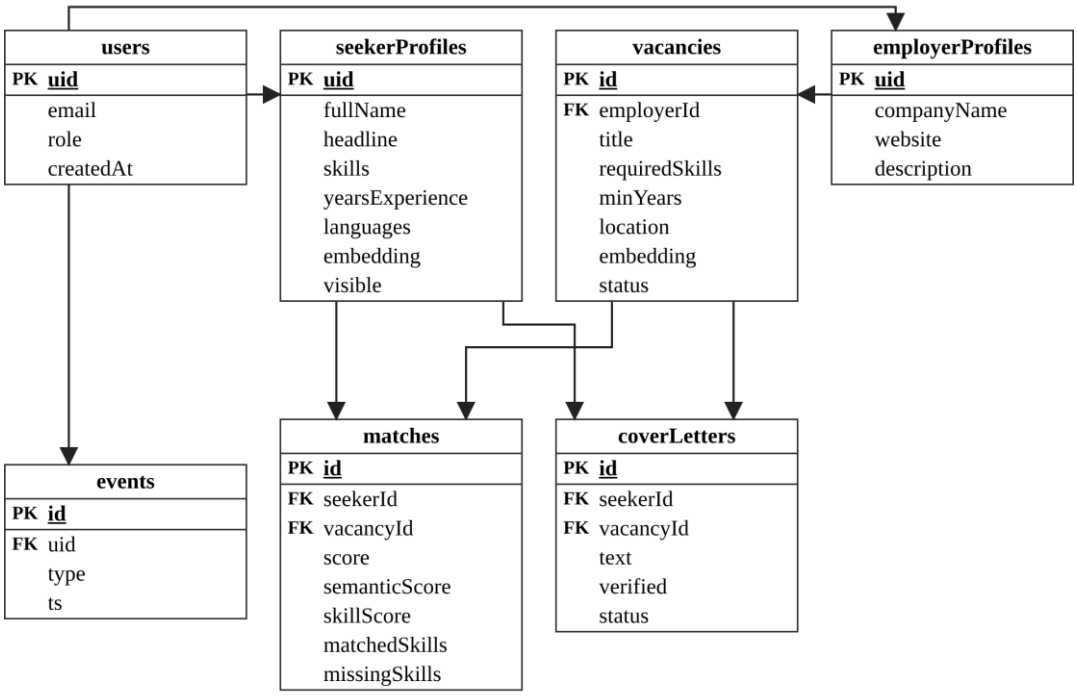



Рисунок 1 – Структура даних Firestore

Вимоги до приватності реалізуються через згоду на обробку даних, керування видимістю профілю (за замовчуванням вимкнена), приховування контактів від роботодавця та видалення пов'язаних даних під час закриття облікового запису. Ранжування має рекомендаційний характер: воно не замінює рішення роботодавця. Цільові показники якості, зокрема частку підтверджених тверджень листа не менше 98% на еталонному наборі з 50 пар, визначено як критерії майбутнього тестування, а не як уже отримані експериментальні результати.

**Висновки.** У результаті роботи спроєктовано вебсистему двобічного підбору вакансій і кандидатів, яка використовує семантичне порівняння, формує пояснення відповідності та генерує супровідний лист із перевіркою його тверджень за даними профілю. Визначено гібридну архітектуру FastAPI, Firebase Authentication, Cloud Firestore, multilingual-e5-small і Gemini API; сформульовано алгоритм оцінювання відповідності з вагами 0,6 для семантичної складової та 0,4 для навичкової; описано шлюз перевірки тверджень супровідного листа, модель даних із сімома колекціями та набір основних сценаріїв API.

Практична цінність проєкту полягає в об'єднанні пояснюваного підбору для обох сторін ринку праці з генерацією листа, обмеженою даними профілю. Водночас описані результати є результатами аналізу та проєктування. Реалізація, тестування алгоритму на еталонних наборах, вимірювання точності й навантаження мають бути виконані на наступному етапі.

**Використання інструментів ШІ.** Під час підготовки публікації використано мовні моделі ChatGPT (GPT-6) та Claude Sonnet 5.5 (Anthropic) для перекладу англомовної назви, анотації та ключових слів (розділ «Abstract»), перевірки пунктуації й орфографії та формулювання думок. Автор перевірив технічний зміст, наведені дані й остаточну редакцію тексту та несе відповідальність за зміст публікації.

## СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ


[1] Wang, L., Yang, N., Huang, X., Yang, L., Majumder, R., & Wei, F. (2024). *Multilingual E5 Text Embeddings: A Technical Report*. arXiv. https://doi.org/10.48550/arXiv.2402.05672

[2] Google. (n.d.). *Gemini API: Structured output*. https://ai.google.dev/gemini-api/docs/structured-output

[3] European Parliament and Council of the European Union. (2016). *Regulation (EU) 2016/679 (General Data Protection Regulation)*. https://eur-lex.europa.eu/eli/reg/2016/679/oj

[4] European Commission. (n.d.). *AI Act Service Desk: High-risk AI systems (Annex III)*. https://ai-act-service-desk.ec.europa.eu/en/node/464

[5] Google. (n.d.). *Firebase Admin SDK: Verify ID tokens*. https://firebase.google.com/docs/auth/admin/verify-id-tokens



Singh A.A.[1], Novikova T.O.[2]


**Intelligent system for two-sided job and candidate matching based on semantic analysis and generative AI**


*Abstract. The design of a web system for two-sided matching of vacancies and candidates based on semantic text analysis and generative AI is considered. A hybrid architecture, a résumé–vacancy matching score, a data model, and a fact-checking mechanism for generated cover letters are proposed. The result is a design specification ready for implementation and experimental evaluation.*



***Keywords:*** *semantic search; recruitment; resume; text embeddings; generative artificial intelligence.*