УДК: 004.8:004.432.2

ПЕРСОНАЛІЗОВАНА ГЕНЕРАЦІЯ РЕЦЕПТІВ НА ОСНОВІ RAG ТА SUPABASE PGVECTOR

Петренко В.С.

Харківський національний автомобільно-дорожній університет,

Харків, Україна

Сучасні системи рекомендацій харчування не вміють генерувати персоналізований контент, який враховував би індивідуальні вподобання користувача. Традиційні підходи використовують статичні алгоритми фільтрації, які не адаптуються до зміни смаків та харчових потреб. Технологія Retrieval-Augmented Generation у поєднанні з векторними базами даних дає змогу створювати системи, які самостійно навчаються на основі взаємодії з користувачем. У роботі представлено реалізацію такої системи MealMatics, мобільний застосунок для персоналізованої генерації рецептів.

Система складається з чотирьох основних компонентів. Для перетворення рецептів у числові вектори використовується модель створення векторних представлень розмірністю 1536, яка кодує назву, тип кухні, опис та інгредієнти у семантичний вектор. Для зберігання та пошуку векторів застосовується PostgreSQL з розширенням pgvector та ієрархічним навігаційним індексом HNSW (параметри m=16, ef_construction=64), оптимізованим для пошуку за косинусною подібністю [1][2]. При генерації нового рецепту система шукає 3-5 найбільш схожих рецептів, які користувач раніше позначив як улюблені, розраховує коефіцієнт схожості та передає цю інформацію разом з коментарями користувача як контекст для великої мовної моделі. Для надійності реалізовано архітектуру з резервною моделлю через механізм автоматичного перемикання при недоступності основної [3][4].

Ключова особливість системи – здатність постійно навчатися на основі відгуків користувача. Після оцінки рецепту система створює векторне представлення та зберігає його разом з відгуком для швидкого пошуку.

Спеціальна функція бази даних виконує наближений пошук найближчих сусідів, розраховуючи косинусну відстань між векторами [2]. Знайдені схожі рецепти формують запит до мовної моделі, що дозволяє генерувати нові рецепти близькі за стилем до уподобань конкретного користувача.

Аналіз технічних характеристик показав високу продуктивність: формування контексту займає менше 200 мс, векторний пошук – менше 100 мс. Алгоритм HNSW забезпечує наближений пошук найближчих сусідів з високою точністю при складності O(log N) [1][2]. Тестування виявило, що при пороговому значенні подібності 0.7 система знаходить рецепти з подібними інгредієнтами, а при 0.85 – практично ідентичні за концепцією. З накопиченням відгуків (понад 10 оцінених рецептів) система починає генерувати рецепти, що краще відповідають індивідуальним вподобанням, що підтверджується підвищенням релевантності при збільшенні обсягу даних про користувача [3].

Для захисту персональних даних реалізовано відповідність вимогам GDPR:

• Ізоляція даних через механізм Row-Level Security у PostgreSQL;

• Право на забуття через автоматичне видалення векторних представлень;

• Шифрування AES-256 у стані спокою та TLS 1.3 при передачі;

• Псевдонімізація ідентифікаторів користувачів [5].

Розроблена система демонструє ефективність підходу доповненого пошуку для персоналізованих додатків.

Здатність до безперервного навчання – з кожним відгуком векторний простір збагачується новими представленнями, що покращує якість семантичного пошуку та релевантність рекомендацій. Наближений пошук найближчих сусідів забезпечує ефективне масштабування при збільшенні обсягу даних.

Література:

1. Datta S., Kumar A. (2024) Retrieval-Augmented Generation: research and implementation patterns. arXiv:2410.12837. https://arxiv.org/pdf/2410.12837.pdf
2. Jönsson P., Chen L. (2024) Deploying Supabase pgvector: HNSW index optimization. Supabase Technical Blog. https://supabase.com/blog/increase-performance-pgvector-hnsw
3. Zhang Y., Wang H., Liu M. (2024) Self-learning personalization systems with RAG and LLM. arXiv:2510.14629. https://arxiv.org/html/2510.14629v1
4. Brown T., Anderson K. (2025) Multi-LLM architecture patterns with fallback strategies. Portkey AI Technical Report. https://portkey.ai/blog/retries-fallbacks-and-circuit-breakers-in-llm-apps
5. Schmidt R., Mueller T. (2024) GDPR compliance in AI recommendation systems. EDPB Guidelines. https://www.edpb.europa.eu/system/files/2025-01/d2-ai-effective-implementation-of-data-subjects-rights_en.pdf