УДК 004.891.2

СПЕЦИФІКАЦІЯ ВИМОГ ТА ПРОЄКТУВАННЯ ПРОГРАМНОЇ СИСТЕМИ ДЛЯ ПРОДАЖУ АВТОРСЬКОЇ ТА САМОРОБНОЇ ЕЛЕКТРОНІКИ ЗІ ЗАСТОСУВАННЯМ АСИСТЕНТУ ШІ

Зіноватна С. Л., к.т.н., доц.[1]; Халюза. Я.Д.[2]
[1,2] Національний університет «Одеська політехніка»
[1,2]Україна, Одеса
[1]zinovatnaya.svetlana@op.edu.ua; [2]khaluza225@gmail.com

Анотація. *В цій роботі проведено специфікацію вимог та проєктування вебплатформи для продажу авторської та саморобної електроніки зі застосуванням асистенту ШІ. Основною метою створення платформи є підвищення зручності пошуку електронних пристроїв, виготовлених фізичними особами та малими підприємствами, шляхом застосування чат-асистента на основі LLM-агента у вебплатформі для їх продажу, що дозволить здійснити підбір пристрою відповідно до запиту, сформульованого природною мовою.*
Ключові слова: *ШІ-асистент; RAG; BM25; Embeddings; електронна комерція.*

Вступна частина. Зростання доступності електронних компонентів, поява дешевого малосерійного виробництва друкованих плат і засобів створення електронних пристроїв, сприяє появі малосерійних авторських розробок [1]. Частина таких розробок може мати потенціал для подальшого поширення та комерційної реалізації. Для їх авторів виникає проблема донесення інформації про власний виріб до потенційних користувачів. Наявні торговельні платформи та спеціалізовані майданчики не завжди забезпечують ефективний пошук вузькоспеціалізованих електронних пристроїв через обмежені можливості вбудованих пошукових інструментів. Особливо складним є пошук товарів за описом їхньої функціональності, оскільки пошукові механізми не завжди здатні враховувати призначення пристрою та його функціональні особливості, що ускладнює знаходження товарів, які відповідають конкретним потребам покупця. Додатковими факторами є використання різної термінології для опису аналогічних виробів, недостатня індексація спеціалізованих товарів у загальних торговельних сервісах та відсутність зручних механізмів пошуку за технічними характеристиками.

Сучасні великі мовні моделі (LLM) та технологія Retrieval-Augmented Generation (RAG) створюють можливості для семантичного зіставлення природномовних запитів користувачів із технічною інформацією про електронні пристрої. Тому актуальним є дослідження та розроблення програмного рішення на основі LLM і RAG, здатного інтерпретувати потреби користувача, знаходити відповідні пристрої та надавати релевантні рекомендації.

Метою роботи є визначення специфікації вимог та проєктування програмної системи пошуку електронних пристроїв, виготовлених фізичними особами та малими підприємствами, шляхом застосування чат-асистента на основі LLM-агента у вебплатформі для їх продажу, що дозволить здійснити підбір пристрою відповідно до запиту, сформульованого природною мовою.

Основна частина. Першим етапом розробки є визначення функціональних та нефункціональних вимог [2]. Системою користуються три типи акторів: адміністратор, покупець, виробник. Адміністратор відповідає за перевірку оголошень перед їх публікуванням. Покупець використовує платформу для пошуку, перегляду та придбання

товарів. Виробник створює та розміщує на платформі власні електронні пристрої для подальшого продажу. При цьому користувач може виконувати роль покупця та виробника.

Виробник може виконувати такі дії: створення пропозиції; редагування пропозиції; перегляд власних пропозицій; обробка замовлення. Дії покупця: перегляд каталогу; пошук товару; оформлення замовлення; консультація із ШІ-асистентом з можливістю пошуком за описом. Спільні дії для покупця та виробника: перехід в протилежний тип користувача (з виробника у продавця та навпаки), авторизація. Адміністратор може виконати модерацію оголошень та отримання метрик ранжування.

До нефункціональних вимог розроблюваної програмної системи належать вимоги до продуктивності, масштабованості, сумісності, безпеки та місткості. Час відповіді серверної частини не повинен перевищувати 5 секунд за умови одночасного використання платформи 100 користувачами, а час формування відповіді ШІ-консультантом – 15 секунд. Система повинна бути сумісною із сервісами штучного інтелекту, що використовують структуру запитів і відповідей, сумісну з офіційним API OpenAI. Також необхідно забезпечити збереження історії листування з ШІ-консультантом обсягом щонайменше 100 останніх повідомлень. Довжина одного вхідного повідомлення не повинна перевищувати 1000 символів, а кількість повідомлень від одного користувача має бути обмежена 10 повідомленнями на годину. Задля безпеки, ШІ-консультанту надається виключно доступ на читання інформації про товари. Система повинна підтримувати безпарольну автентифікацію користувачів за допомогою облікових записів Google і GitHub із використанням протоколів OAuth 2.0 та OpenID Connect через офіційні SDK або перевірені програмні бібліотеки.

Відповідно до специфікації вимог, для розроблення вебзастосунку обрано клієнт-серверний підхід, у якій клієнтська частина відповідає за відображення та введення даних, а їх оброблення й зберігання здійснюються на сервері. Тому основну увагу під час проєктування приділено серверній частині системи.

Першим варіантом розглянуто модульний моноліт, у якому серверна частина реалізується як єдиний застосунок, поділений на функціональні модулі. Для зберігання даних використовується єдина реляційна СКБД. Перевагами є простота розроблення, розгортання, супроводу, транзакційна узгодженість даних і відсутність мережевих накладних витрат між модулями. Основні недоліки масштабування застосунку як єдиного цілого та конкуренція за ресурси СКБД між пошуковими й транзакційними операціями [3].

Другим варіантом є архітектура з виділеним ШІ-сервісом, за якої основна бізнес-логіка та ШІ-компонент розгортаються незалежно. Цей підхід дає змогу незалежно оновлювати ШІ-компонент і зберігати доступність основних функцій у разі його недоступності. Контроль доступу до даних через основний сервіс також сприяє підвищенню безпеки. Водночас така архітектура ускладнює розгортання, моніторинг і міжсервісну взаємодію та може спричиняти дублювання моделей даних.

Третім варіантом є сервіс-орієнтована архітектура, у якій кожен функціональний компонент реалізується як окремий сервіс. Вона забезпечує незалежне масштабування, проте потребує складніших механізмів узгодження даних, забезпечення ідемпотентності та моніторингу. З огляду на масштаб проєкту, обмежені ресурси розроблення й відсутність потреби в незалежному масштабуванні компонентів, цей підхід визнано надмірно складним на поточному етапі.

Отже, базовою архітектурою обрано модульний моноліт із чіткими межами модулів та абстракціями для взаємодії з пошуковим рушієм і мовною моделлю. Таке рішення спрощує реалізацію та супровід системи й водночас створює передумови для подальшого виділення

ШІ-компонента в окремий сервіс або переходу до сервіс-орієнтованої архітектури за зростання навантаження.

У межах обраної архітектури виділено модулі адміністративної панелі, каталогу, пошуку, замовлень, метрик, авторизації та ШІ-асистента. Модуль адміністративної панелі забезпечує модерацію оголошень і перегляд метрик адміністратором. Модуль каталогу відповідає за зберігання, отримання та керування інформацією про товари. Модуль пошуку реалізує повнотекстовий пошук, а також передає знайдені результати ШІ-асистенту для використання в RAG-конвеєрі. Модуль замовлень забезпечує створення, зберігання та перевірку замовлень покупців. Модуль метрик призначений для оцінювання ефективності пошуку й добору товарів відповідно до користувацьких запитів. Модуль авторизації здійснює перевірку облікових даних користувачів і керування доступом відповідно до їхніх ролей. Модуль ШІ-асистента забезпечує діалогову взаємодію з користувачем, надає рекомендації щодо вибору пристроїв, пояснює їхні характеристики та допомагає знаходити товари відповідно до потреб покупця.

Подальше дослідження акцентувалось на розробці модуля пошуку та ШІ-асистенту. Модуль пошуку забезпечує індексацію товарів і пошук за текстовими запитами користувачів. Він поєднує повнотекстовий пошук на основі алгоритму BM25 і векторний пошук за семантичною подібністю. Алгоритм BM25 був обраний, оскільки він враховує частотність термінів і довжину документів, забезпечуючи ефективніше ранжування порівняно з TF-IDF. Векторне представлення товару створюється на основі нормалізованого, за допомогою великої мовної моделі (LLM), опису товару. Для генерації векторних представлень модуль підтримує локальні моделі та зовнішні сервіси. Пошук семантично подібних товарів здійснюватиметься за косинусною мірою подібності, обчислення якої покладатиметься на СКБД із підтримкою векторних операцій. Для взаємодії з іншими модулями системи відкрито інтерфейси текстового й векторного пошуку.

Модуль ШІ-асистента надає для кожного покупця одна сесію чату зі збереженням історії повідомлень і контролем кількості запитів. Асистент побудований на базі LLM доступ до якої реалізується через уніфікований інтерфейс, що підтримує локальне та віддалене розгортання. Асистент може надавати відповіді з ціллю консультації, а також виконувати пошук товарів за їх описом. Для пошуку товарів застосовується механізм tool-calling, за допомогою якого LLM може викликати публічні функції модуля пошуку для виконання текстового та семантичного пошуку.

Результати пошуку товарів відображатимуться в порядку спадання релевантності до запиту користувача. Для оцінювання ефективності пошукових моделей і спостереження за змінами якості пошуку використовуються метрики ранжування, а саме Precision@k для оцінювання частки релевантних товарів серед перших k результатів, Recall@k для визначення повноти знаходження релевантних товарів, Mean Reciprocal Rank (MRR) для оцінювання позиції першого релевантного результату та Normalized Discounted Cumulative Gain (NDCG@k) для оцінювання якості ранжування з урахуванням позиції та ступеня релевантності товарів [4]. Релевантність визначатиметься на основі поведінкових даних користувачів, зокрема кількості переходів до сторінок товарів і здійснених покупок. Відповідні дії фіксуватимуться та зберігатимуться в базі даних для подальшого аналізу й формування оцінок релевантності. Метрики використовуватимуться для порівняння текстового й векторного пошуку, оцінювання впливу змін моделей і параметрів пошуку, а також моніторингу якості роботи системи в процесі її розвитку.

Висновки. У результаті дослідження сформовано специфікації функціональних і нефункціональних вимог до системи для продажу авторської та саморобної електроніки із використанням ШІ-асистента. Нефункціональні вимоги, в більшості, охоплюють основні аспекти роботи модулів пошуку та ШІ-асистенту. На основі сформованих вимог обґрунтовано архітектуру програмної системи в межах клієнт-серверної моделі взаємодії. Основну увагу архітектурного проєктування приділено модулям пошуку та ШІ-асистенту, оскільки інші модулі системи реалізують типові для систем електронної комерції функції, тому основну увагу зосереджено на специфічних компонентах системи.

Отримані результати створюють основу для реалізації програмного продукту, який буде використовувати ШІ-асистента, що може сприяти підвищенню ефективності пошуку товарів і поліпшенню взаємодії з покупцями, що потенційно позитивно впливатиме на продажі.

Використання інструментів ШІ. Інструмент ШІ Google Gemini використовувався для пошуку джерел інформації та для проведення швидкого аналізу змісту знайдених джерел.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

[1] von Platen, V., & Kitani, Y. (2023). A designer's understanding of the maker movement. *Proceedings of the International Conference on Engineering Design*, *1*, 3433–3442. https://doi.org/10.1017/pds.2023.344.

[2] Gamma, E., Helm, R., Johnson, R., & Vlissides, J. (1994). *Design Patterns: Elements of Reusable Object-Oriented Software*. Boston, MA: Addison-Wesley Professional.

[3] Yakovenko, V., Ulianovska, Y., & Yakovenko, T. (2022). Analysing features of e-commerce systems architecture. *Technology Audit and Production Reserves*, *1* (4 (63)), 27–31. https://doi.org/10.15587/2706-5448.2022.253932.

[4] Harman, D. (2022). *Information retrieval evaluation*. Morgan & Claypool Publishers.

Zinovatna S.L., Khaliuza Y.D.

The Software Requirements Specification and System Design for a Custom and Handmade Electronics E-Commerce Platform with an AI Assistant

Abstract. *In this paper, the software requirements specification and system design of a web platform for selling custom and handmade electronics using an AI assistant were performed. The main goal of developing the platform is to improve the user experience in searching for electronic devices manufactured by individuals and small businesses by integrating an LLM agent-based chat assistant into the e-commerce web platform, which enables device selection based on natural language queries.*

Keywords: *AI assistant; RAG; BM25; Embeddings; e-commerce*