КОМП'ЮТЕРНІ НАУКИ, ШТУЧНИЙ ІНТЕЛЕКТ ТА СИСТЕМНИЙ АНАЛІЗ

https://doi.org/10.31649/1997-9266-2026-186-3-27-34

УДК 004.414.23:004.89:658.84

С. Г. Московко[1]
Р. Н. Квєтний[1]

МЕТОД АВТОМАТИЗОВАНОЇ ІДЕНТИФІКАЦІЇ ОБМЕЖЕНИХ КОНТЕКСТІВ ПІД ЧАС ПРОЄКТУВАННЯ СИСТЕМ ЕЛЕКТРОННОЇ КОМЕРЦІЇ

[1]Вінницький національний технічний університет

Розглянуто проблематику автоматизації стратегічного етапу проєктування систем електронної комерції, фокусуючись на моделюванні складної бізнес-логіки та структуруванні предметної області. Сучасні системи електронної комерції характеризуються високим рівнем заплутаності бізнес-правил та великою кількістю взаємопов'язаних процесів. Ефективне управління цією складністю можливе завдяки застосуванню підходу предметно-орієнтованого проєктування, який передбачає декомпозицію системи на логічно відокремлені модулі — обмежені контексти. Якість виділення цих контекстів безпосередньо впливає на життєздатність системи, зрозумілість кодової бази та можливість її подальшого розвитку.

Традиційні підходи до ідентифікації меж контекстів базуються переважно на евристичних методах та експертних сесіях, які є ресурсомісткими та залежними від людського фактора. Наявні формальні методи, що ґрунтуються на структурному аналізі даних, часто не здатні коректно інтерпретувати семантичні нюанси бізнес-термінології. У роботі запропоновано інформаційну технологію, що використовує можливості генеративного штучного інтелекту та великих мовних моделей для автоматизованого аналізу простору задачі.

Основну увагу в статті приділено розробці методу семантичної кластеризації вимог, який дозволяє виявляти приховані лінгвістичні закономірності в описі бізнес-процесів. Запропонований підхід передбачає використання великих мовних моделей для аналізу єдиної мови проєкту та формування карти контекстів на основі семантичної близькості понять, а не лише їхніх технічних зв'язків. Описано алгоритм, який трансформує текстові специфікації та історії користувачів у формалізовані моделі доменів, визначаючи рекомендовані межі відповідальності для кожного модуля.

Результати дослідження демонструють, що застосування генеративного підходу дозволяє значно підвищити об'єктивність моделювання предметної області, мінімізувати когнітивне навантаження на архітекторів та забезпечити валідацію логічної цілісності системи на ранніх етапах проєктування. Розроблена технологія слугує інструментом інтелектуальної підтримки прийняття рішень, дозволяючи формувати гнучкі та адаптивні до змін бізнес-моделі електронної комерції.

Ключові слова: обмежений контекст, предметно-орієнтоване проєктування, великі мовні моделі, семантична кластеризація, карта контекстів, електронна комерція.

Вступ

Системи електронної комерції характеризуються складною та взаємопов'язаною бізнес-логікою, що охоплює керування каталогом і цінами, промоактивності, оформлення та супровід замовлень, оплати, доставку, повернення, а також взаємодію з клієнтськими сервісами [2]. На стратегічному етапі проєктування ключовим є коректне структурування предметної області, оскільки саме воно визначає розподіл відповідальностей у програмних модулях, зрозумілість моделі та керованість подальших змін [3]. У рамках предметно-орієнтованого проєктування (DDD, Domain-Driven Design) така структуризація реалізується через виділення обмежених контекстів і встановлення їх меж та інтеграційних взаємодій [4].

Актуальність зумовлена високою складністю та мінливістю бізнес-правил електронної комерції

і критичністю коректного виділення обмежених контекстів для керованості кодової бази та розвитку системи [2], [3]. Традиційні підходи є ресурсомісткими й суб'єктивними [4], а суто структурні методи аналізу не враховують семантику «Єдиної мови» [1]. Використання великих мовних моделей (LLM, Large Language Models) дає змогу автоматизувати інтерпретацію текстових вимог [7], [8] і виявляти контекстні межі на рівні змісту [9].

Метою статті є підвищення точності ідентифікації обмежених контекстів під час проєктування систем електронної комерції за допомогою розробки методу на основі семантичного аналізу вимог засобами великих мовних моделей.

Для досягнення поставленої мети необхідно вирішити такі завдання:

1) розробити метод семантичного аналізу та видобування доменних знань з текстових вимог засобами великих мовних моделей;

2) розробити алгоритм кластеризації вимог та автоматизованої побудови карти контекстів на основі семантичної близькості;

3) провести експериментальну оцінку ефективності запропонованого методу.

Огляд літератури

Проблематика декомпозиції складних програмних систем є предметом активних досліджень в галузі програмної інженерії. Як зазначають Dragoni et al. [2], в епоху переходу до мікросервісної архітектури критичного значення набуває коректне визначення меж сервісів [3]. Автори підкреслюють, що помилки на цьому етапі, зокрема неправильне визначення меж відповідальності, призводять до виникнення антипатерну «розподілений моноліт» (distributed monolith), який нівелює переваги модульності та ускладнює підтримку системи. Це підтверджує тезу про те, що суто технічного поділу системи недостатньо — необхідне глибоке розуміння семантики предметної області.

Загальновизнаним підходом до вирішення цієї проблеми є предметно-орієнтоване проєктування. У ґрунтовному систематичному огляді літератури (SLR), проведеному Özkan et al. [4] у 2025 році, аналізується сучасний стан впровадження DDD. Автори підсумовують, що хоча DDD є ефективним інструментом структурування бізнес-логіки, його застосування на практиці залишається неоднорідним та часто позбавленим достатньої емпіричної бази. Ключовою проблемою, виділеною у [2], є висока залежність стратегічного проєктування від евристичних методів та експертних сесій (наприклад, Event Storming). Такий підхід характеризується значною ресурсомісткістю та суб'єктивністю («людський фактор»), через що процес ідентифікації обмежених контекстів важко відтворити і він є залежним від кваліфікації архітектора.

Наявна невідповідність між потребою в об'єктивній декомпозиції та суб'єктивністю традиційних методів стимулює пошук автоматизованих рішень. Традиційні методи NLP (Natural Language Processing), основані на частотному аналізі або синтаксичних зв'язках [5], часто виявляються неспроможними вловити тонкі семантичні нюанси «Єдиної мови» (Ubiquitous Language).

Проривним напрямком у цій сфері стає використання великих мовних моделей (LLM) [6]. Застосування LLM в інженерії програмного забезпечення набуває широкої популярності завдяки їхній здатності обробляти великі обсяги неструктурованих даних [7], [8]. Arulmohan, Meurs та Mosser [9] у своєму дослідженні демонструють ефективність генеративного штучного інтелекту для автоматизованого вилучення доменних моделей з текстових вимог. Порівнюючи LLM з класичними інструментами моделювання, автори доводять, що моделі здатні значно точніше ідентифікувати сутності, ролі та бізнес-правила з історій користувачів (User Stories), що відкриває шлях до автоматизації рутинних етапів аналізу.

Для задачі групування виділених елементів у контексти критично важливим є поняття семантичної близькості. Petukhova та Matos-Carvalho [10] досліджували застосування векторних представлень (embeddings), генерованих LLM, для задач кластеризації тексту. Їхні результати свідчать про те, що LLM-ембеддінги значно краще захоплюють контекстуальні значення та семантичні нюанси слів порівняно з традиційними векторними моделями. Це дає підстави стверджувати, що використання LLM дозволяє проводити кластеризацію вимог на основі їхнього змісту, а не лише лексичного збігу, що є необхідною умовою для коректного виділення Bounded Contexts.

Отже, аналіз джерел дозволяє констатувати наявність розриву між існуючими експертними методами DDD та можливостями сучасних генеративних технологій. Поєднання підходів стратегічного моделювання з аналітичною потужністю LLM для автоматичної ідентифікації меж контекстів в E-commerce системах потребує подальшого вивчення та розробки відповідних алгоритмічних рішень.

Результати дослідження

Основою запропонованого підходу є гібридна модель, що поєднує методи обробки природної мови (NLP) на базі великих мовних моделей з алгоритмами машинного навчання без учителя (Clustering). Методологія спрямована на подолання семантичного розриву між текстовим описом вимог та архітектурною моделлю системи шляхом трансформації неструктурованих даних у формалізовану карту контекстів.

Загальна схема запропонованої інформаційної технології подана на рис. 1.

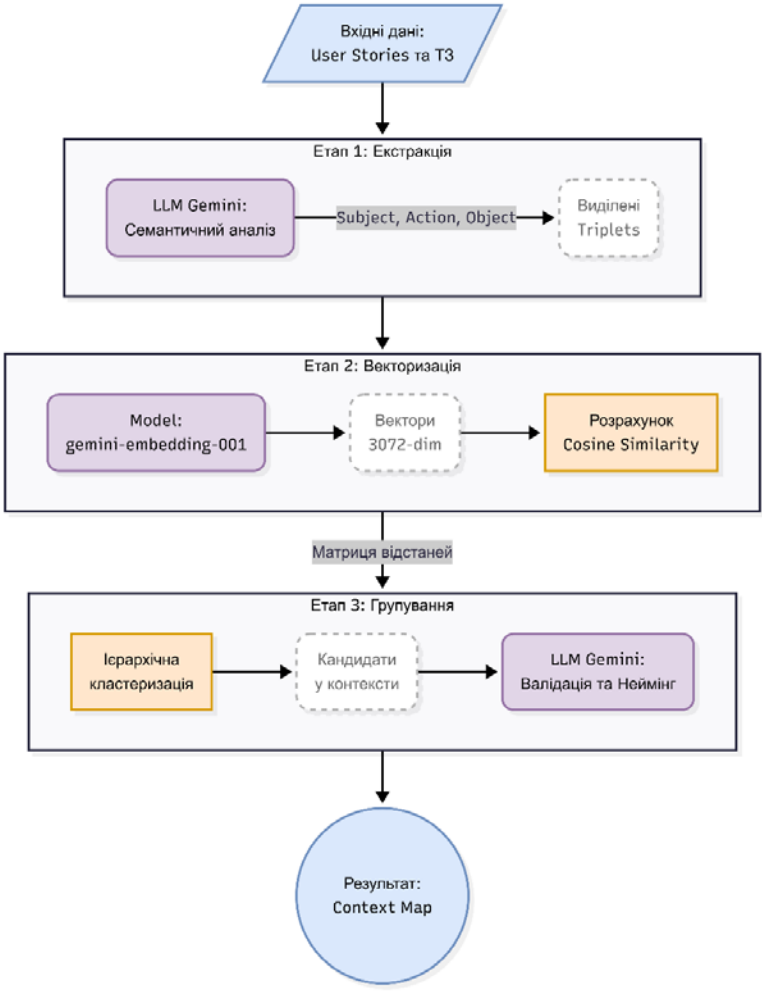

Рис. 1. Структурна схема методу автоматизованої ідентифікації обмежених контекстів

Процес ідентифікації складається з трьох послідовних етапів: екстракція доменних знань, векторизація та оцінка семантичної близькості, кластеризація та генеративна валідація меж.

Вхідними даними для системи є множина вимог $R=\{r_1,r_2,\ldots,r_n\}$, яка може бути представлена у вигляді історій користувачів, сценаріїв використання або технічного завдання.

Традиційні методи морфологічного аналізу не дозволяють коректно виділити бізнес-сутності через складність лінгвістичних конструкцій. Для розв'язання цієї задачі пропонується використання LLM сімейства Gemini [11], ефективність якого для екстракції доменних моделей з текстових вимог підтверджено в [9].

Задачу першого етапу можна формалізувати як функцію відображення $f_{LLM}: R \to E$, де E — множина структурованих доменних елементів. Для кожної вимоги r_i модель генерує набір триплетів

$$e_{ij} = \langle Subject, Action, Object \rangle, \quad (1)$$

де $Subject$ — актор або компонент, що ініціює дію; $Action$ — бізнес-операція або подія; $Object$ — сутність домену, над якою виконується операція.

Такий підхід дозволяє відфільтрувати допоміжні слова та сформувати чистий словник «Єдиної мови», що є критичним для DDD.

Для автоматичного групування виділених елементів необхідно перевести їх у метричний простір. Кожен елемент e_{ij} (або вихідна вимога r_i, збагачена контекстом) трансформується у векторне представлення $V_i \in R^d$ за допомогою моделі ембеддінгів:

$$V_i = Embed(r_i). \quad (2)$$

Підхід до отримання семантично змістовних векторних представлень речень обґрунтовано в [5]. У роботі використано модель gemini-embedding-001, яка оптимізована для роботи з багатомовними текстами та семантичним пошуком [10]. Розмірність вихідного вектора становить $d=3072$.

Ключовим аспектом кластеризації у DDD є визначення того, наскільки сильно пов'язані дві бізнес-

вимоги. Мірою семантичної близькості між двома векторами V_a та V_b вибрано косинусну подібність (Cosine Similarity) [15], яка інваріантна до довжини тексту і залежить лише від кута між векторами у семантичному просторі

$$Sim(V_a, V_b) = \cos(\theta) = \frac{V_a \cdot V_b}{|V_a||V_b|} = \frac{\sum_{k=1}^{d} V_{a,k} V_{b,k}}{\sqrt{\sum_{k=1}^{d} V_{a,k}^2} \sqrt{\sum_{k=1}^{d} V_{b,k}^2}}. \quad (3)$$

На основі розрахованих попарних відстаней будується матриця суміжності M розмірністю $n \times n$, де кожний елемент $M_{\text{ab}} = \text{Sim}(V_a, V_b)$ відображає силу зв'язку між вимогами.

Для декомпозиції системи на обмежені контексти задачу зведено до задачі виявлення спільнот (community detection) на зваженому неорієнтованому графі [16] $G = (V, E, W)$, який формально визначається так:

$$G = (V, E, W);\ V = R;\ E = \{(r_a, r_b) \mid r_a, r_b \in R,\ a \neq b\},\ W : E \to [0,1], \quad (4)$$

де V — множина вершин графа, що відповідає множині вимог R; E — множина ребер; $W(r_a, r_b) = M_{\text{ab}} = \text{Sim}(V_a, V_b)$ — вагова функція ребер, що відповідає значенню семантичної подібності між вимогами.

Для розв'язання задачі виявлення спільнот на графі G застосовано алгоритм агломеративної ієрархічної кластеризації [17] з методом зв'язку average linkage. Формально, задачу кластеризації можна представити як побудову розбиття

$$\mathcal{C} = \{C_1, C_2, \ldots, C_K\};\ \bigcup_{j=1}^{K} C_j = R;\ C_i \cap C_j = \varnothing \forall i \neq j, \quad (5)$$

де C — множина кластерів; K — задане число кластерів; $C_j \subseteq R$ — j-й кластер як підмножина вимог. Об'єднання двох кластерів C_i та C_j на ітерації агломеративного алгоритму виконується за критерієм мінімізації середньої відстані між їхніми елементами

$$d_{\text{avg}}(C_i, C_j) = \frac{1}{|C_i| \cdot |C_j|} \sum_{r_a \in C_i} \sum_{r_b \in C_i} (1 - \text{Sim}(V_a, V_b)), \quad (6)$$

де $|C_i|$ та $|C_j|$ — потужності відповідних кластерів. Алгоритм ітеративно об'єднує найближчі кластери, доки не буде досягнуто заданої кількості K.

Завершальним етапом є генеративна інтерпретація. Оскільки математичний кластер є лише набором векторів, для надання йому бізнес-змісту застосовується повторний запит до LLM Gemini. Для кожного кластера C_j, отриманого згідно з (5), формується промпт P_{gen}, побудований з використанням паттернів проєктування запитів [12]

$$P_{gen} = Analyze\ set\{r \in C_j\};\ Define\ Bounded\ Context\ Name\ and\ Responsibilities. \quad (7)$$

Модель аналізує сукупність вимог, що потрапили в один кластер, і формує опис контексту, перевіряючи його на логічну цілісність (Cohesion). Якщо LLM виявляє у кластері вимоги, що логічно суперечать одна одній, ініціюється процедура рефакторингу (розбиття кластера або переміщення елементів).

Отриманий набір іменованих контекстів та зв'язків між ними формує цільову карту контекстів, яка є основою для архітектури системи електронної комерції.

Приклад застосування

Для верифікації розробленого методу та оцінки його практичної придатності проведено експериментальне моделювання архітектури підсистеми електронної комерції. Експеримент реалізовано мовою програмування Python із застосуванням бібліотеки *scikit-learn* для виконання кластерного аналізу. Як інструмент векторизації текстових даних використано модель gemini-embedding-001, яка дозволяє перетворювати семантику вимог у 3072-вимірні вектори. Метою дослідження

виступала перевірка гіпотези щодо здатності LLM автоматично виявляти архітектурні межі на основі неструктурованих історій користувачів.

Вхідний набір даних складався з 13 історій користувачів, що описують різні аспекти функціонування онлайн-магазину: від роботи з каталогом до фінансової звітності. Вимоги охоплювали дії різних акторів (Адміністратор, Покупець, Менеджер, Система, Бухгалтер).

Фрагмент переліку вимог подано в таблиці.

Вхідні дані (User Stories) для експерименту

ID	Актор	Зміст вимоги
US-01	адміністратор	додавання нових товарів з описом та фото до каталогу.
US-02	покупець	фільтрація товарів за ціною, брендом та характеристиками.
US-04	система	зменшення доступної кількості товару на складі після замовлення.
US-06	система	перевірка наявності товару перед підтвердженням замовлення.
US-08	покупець	отримання листа-підтвердження з деталями замовлення.
US-10	менеджер	перегляд статусу оплати замовлення в реальному часі.
US-11	система	ініціювання транзакції через платіжний шлюз.
US-13	бухгалтер	формування звіту по успішним транзакціям.

На етапі обробки кожну вимогу перетворено у векторне представлення. Для групування векторів застосовано алгоритм агломеративної кластеризації з використанням косинусної метрики відстані. У результаті роботи алгоритму було виділено чотири стійкі кластери (Cluster 0 – Cluster 3), розподіл вимог між якими продемонстрував специфічні особливості семантичного сприйняття моделі gemini-embedding-001. Візуалізація отриманого простору рішень методом багатовимірного шкалювання (MDS, Multidimensional Scaling) показана на рис. 2.

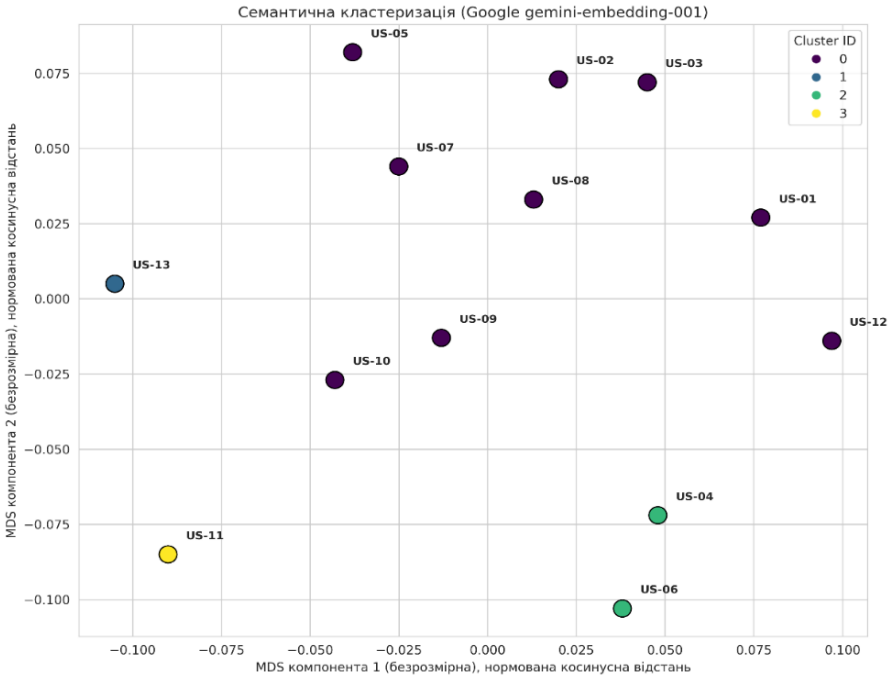

Рис. 2. Результат семантичної кластеризації вимог

Аналіз отриманих результатів дозволив виявити архітектурну закономірність. Найбільший кластер (Cluster 0) об’єднав дев’ять вимог (US-01, US-02, US-03, US-05, US-07, US-08, US-09, US-10, US-12). Характерною ознакою цього кластера є орієнтація на взаємодію людини з інтерфейсом системи. Модель згрупувала в один контекст дії покупців, адміністраторів та менеджерів, незалежно від того, чи стосуються вони каталогу, чи перегляду статусів оплати. З точки зору архітектури, це відповідає виділенню шару Front-Office або шару аплікаційних сервісів (Application Layer), де зосереджена логіка презентації та користувацьких сценаріїв.

Натомість, вимоги, де актором виступає «Система», чітко відокремлено в інші кластери. Cluster 2 об’єднав вимоги US-04 та US-06, які стосуються автоматичної перевірки залишків та зменшення кількості товару. Це дозволяє ідентифікувати цей кластер як Inventory Domain Service — сервіс,

що інкапсулює критичну бізнес-логіку управління запасами, приховану від користувача. Аналогічно, Cluster 3 (US-11), що відповідає за технічну транзакцію через шлюз, виділено окремо, формуючи ядро Payment Service. Вимога US-13 (звітність бухгалтера) сформувала Cluster 1, що логічно відповідає контексту аналітики (Reporting Context), відокремленому від операційних процесів.

Отже, експеримент показав, що застосування моделі gemini-embedding-001 сприяє декомпозиції системи не лише за предметними областями, а й за архітектурними шарами. Такий результат є цінним для проєктування розподілених систем, оскільки дозволяє автоматично виявляти кандидатів на мікросервіси з високою цілісністю бізнес-правил, відокремлюючи їх від шару оркестрації користувацьких запитів.

Експериментальна оцінка

Для верифікації отриманих результатів проведено кількісне порівняння автоматизованої декомпозиції з еталонною архітектурою, побудованою експертом за класичним предметним принципом. Як метрику зовнішньої валідації кластеризації використано скоригований індекс Ренда (Adjusted Rand Index, ARI) [13] — міру подібності двох розбиттів множини, яка оцінює ступінь узгодженості між автоматичним та експертним групуванням з поправкою на випадковий збіг. Значення ARI знаходиться в діапазоні від –1 до 1, де 1 означає повний збіг, 0 — випадковий рівень, а від'ємні значення - анти кореляцію. Розрахункове значення ARI склало -0,023, що вказує на відсутність статистичної кореляції між машинним та традиційним людським підходами. Такий результат інтерпретується не як помилка кластеризації, а як виявлення альтернативної архітектурної парадигми: модель згрупувала вимоги не за іменниками предметної області (товар, кошти), а за типом поведінки та архітектурним шаром, що є нетривіальним завданням для людини-проєктувальника.

Як метрику внутрішньої валідації кластеризації використано коефіцієнт силуету (Silhouette Score) [14] — міру якості розбиття, що для кожного об'єкта оцінює наскільки він подібний до свого кластера порівняно з найближчим сусіднім кластером. Значення коефіцієнта силуету знаходиться в діапазоні від –1 до 1, де значення, близькі до 1, вказують на щільні та добре відокремлені кластери. Аналіз показав значення 0,202, що є задовільним показником для коротких текстових даних і свідчить про наявність чітких меж між виявленими контекстами. Це підтверджує, що алгоритм не просто об'єднав випадкові вимоги, а виявив стійку структуру в просторі ознак.

Ключовим показником ефективності методу став коефіцієнт внутрішньої семантичної зв'язності (Semantic Cohesion), який визначається як середнє значення косинусної подібності між усіма парами вимог всередині кожного кластера. Ця метрика відображає ступінь змістовної однорідності сформованих контекстів. Досягнуте значення 0,913 свідчить про виняткову однорідність сформованих контекстів. Це гарантує, що модулі, спроєктовані на основі такої декомпозиції, відповідатимуть принципу High Cohesion, що є критично важливим для мікросервісної архітектури [2]. Отже, експеримент довів, що запропонований підхід дозволяє створювати архітектурні моделі з вищою логічною щільністю, ніж традиційні евристичні методи, хоча й потребує подальшої інтерпретації архітектором для узгодження з бізнес-термінологією.

Висновки

Вирішено актуальну науково-прикладну задачу підвищення точності ідентифікації обмежених контекстів під час проєктування систем електронної комерції за допомогою розробки методу на основі семантичного аналізу вимог засобами великих мовних моделей.

Розроблено метод семантичного аналізу та видобування доменних знань з текстових вимог. Метод базується на використанні LLM для екстракції структурованих триплетів (Subject, Action, Object) з історій користувачів та їх подальшій векторизації за допомогою моделі ембеддінгів. Запропонований підхід дозволив об'єктивізувати процес декомпозиції предметної області, зменшивши вплив людського фактора та когнітивного навантаження на архітектора.

Розроблено алгоритм кластеризації вимог та автоматизованої побудови карти контекстів на основі семантичної близькості. Алгоритм використовує косинусну подібність векторних представлень для побудови матриці суміжності та агломеративну кластеризацію для виділення груп семантично пов'язаних вимог. Генеративна інтерпретація кластерів засобами LLM забезпечує валідацію та іменування обмежених контекстів.

Проведено експериментальну оцінку ефективності запропонованого підходу. Досягнутий показник внутрішньої семантичної зв'язності на рівні 0,913 свідчить про формування винятково гомо-

генних контекстів, що є передумовою для побудови надійних мікросервісних архітектур з низьким рівнем зачеплення. Виявлена низька кореляція результатів роботи алгоритму з традиційним експертним баченням (ARI = –0,023) вказує на здатність штучного інтелекту пропонувати альтернативні, архітектурно обґрунтовані варіанти декомпозиції, що базуються на поведінкових патернах системи, а не лише на лінгвістичній схожості термінів.

Наукова новизна роботи полягає у поєднанні DDD та LLM для формування меж обмежених контекстів за семантичною близькістю доменних понять, а не лише за технічними зв'язками. Запропоновано метод семантичної кластеризації текстових вимог і алгоритм їх трансформації у формалізовані доменні моделі з автоматизованими рекомендаціями щодо декомпозиції.

Практична цінність дослідження полягає у створенні інструментарію, який дозволяє на ранніх етапах розробки трансформувати неструктуровані вимоги у валідну карту контекстів, що значно прискорює процес моделювання складних E-commerce платформ. Подальші дослідження доцільно спрямувати на розробку механізмів зворотного зв'язку для інтерактивного уточнення меж контекстів архітектором у режимі реального часу.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

[1] W. K. G. Assunção, S. R. Vergilio, and R. E. Lopez-Herrejon, "A survey on microservices extraction approaches," *arXiv preprint arXiv:2104.09278, 2021*. https://doi.org/10.48550/arXiv.2104.09278 .

[2] N. Dragoni, et al., "Microservices: Yesterday, today, and tomorrow," in *Present and Ulterior Software Engineering*, M. Mazzara and B. Meyer, Eds. Cham: Springer, 2017, pp. 195-216. https://doi.org/10.1007/978-3-319-67425-4_12 .

[3] H. Vural, M. Koyuncu, and S. Guney, "A systematic literature review on microservices," 2017, pp. 203-217. https://doi.org/10.1007/978-3-319-62407-5_14 . (Scopus)

[4] O. Özkan, Ö. Babur, and M. van den Brand, "Domain-Driven Design in software development. A systematic literature review on implementation, challenges, and effectiveness," *Journal of Systems and Software*, vol. 207, Art. no. 112537, 2025. https://doi.org/10.1016/j.jss.2025.112537 . (Scopus, WoS)

[5] N. Reimers, and I. Gurevych, "Sentence-BERT: Sentence embeddings using Siamese BERT-networks," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*, Hong Kong, 2019, pp. 3982-3992. https://doi.org/10.18653/v1/D19-1410 . (Scopus)

[6] A. Vaswani, N. Shazeer, and N. Parmar, "Attention is all you need," in *Advances in Neural Information Processing Systems*, vol. 30, 2017, pp. 5998-6008. https://doi.org/10.48550/arXiv.1706.03762 . (Scopus)

[7] X. Hou, et al., "Large language models for software engineering: A systematic literature review," *ACM Transactions on Software Engineering and Methodology*, 2024. https://doi.org/10.1145/3643677 . (Scopus, WoS)

[8] Q. Zhang, et al., "A survey on large language models for software engineering," *arXiv preprint arXiv:2312.15223,* 2023. https://doi.org/10.48550/arXiv.2312.15223 .

[9] S. Arulmohan, M.-J. Meurs, and S. Mosser, "Extracting domain models from textual requirements in the era of large language models," in *IEEE/ACM International Conference on Model Driven Engineering Languages and Systems Companion (MODELS-C)*, Västerås, 2023, pp. 1-9. https://doi.org/10.1109/MODELS-C59198.2023.00096 . (Scopus)

[10] A. Petukhova, J. P. Matos-Carvalho, and N. Fachada, "Text clustering with large language model embeddings," *International Journal of Cognitive Computing in Engineering*, vol. 6, pp. 1-12, 2025. https://doi.org/10.1016/j.ijcce.2024.11.004 . (Scopus)

[11] Gemini Team, R. Anil, et al., "Gemini: A family of highly capable multimodal models," *arXiv preprint arXiv:2312.11805,* 2023. https://doi.org/10.48550/arXiv.2312.11805 .

[12] J. White, et al., "A prompt pattern catalog to enhance large language model performance," *arXiv preprint arXiv:2302.11382*, 2023. https://doi.org/10.48550/arXiv.2302.11382 .

[13] L. Hubert, and P. Arabie, "Comparing partitions," *Journal of Classification*, vol. 2, no. 1, pp. 193-218, 1985. https://doi.org/10.1007/BF01908075 . (Scopus, WoS)

[14] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53-65, 1987. https://doi.org/10.1016/0377-0427(87)90125-7 . (Scopus, WoS)

[15] A. Singhal, "Modern information retrieval: A brief overview," I*EEE Data Engineering Bulletin*, vol. 24, no. 4, pp. 35-43, 2001. [Electronic resource]. Available: https://www.researchgate.net/publication/292766142_Modern_Information_Retrieval_A_Brief_Overview .

[16] S. Fortunato, "Community detection in graphs," *Physics Reports*, vol. 486, no. 3-5, pp. 75-174, 2010. https://doi.org/10.1016/j.physrep.2009.11.002 . (Scopus, WoS)

[17] D. Müllner, "Modern hierarchical, agglomerative clustering algorithms," *arXiv preprint arXiv:1109.2378*, 2011. https://doi.org/10.48550/arXiv.1109.2378 .

[18] S. Moskovko, and R. Kvyetnyy, "The method of structured transformation of a business process activity diagram into a context map of domain-driven design," *Bulletin of Cherkasy State Technological University*, vol. 30, no. 2, pp. 33-43, 2025. https://doi.org/10.62660/bcstu/2.2025.33 .

Рекомендована кафедрою автоматизації та інтелектуальних інформаційних технологій ВНТУ

Дата надходження 6.04.2026
Дата прийняття до друку після рецензування 30.05.2026
Дата публікації 7.07.2026

Московко Сергій Геннадійович — аспірант кафедри автоматизації та інтелектуальних інформаційних технологій, e-mail: smoskovko@icloud.com . https://orcid.org/0009-0001-4376-0187 ;

Квєтний Роман Наумович — член-кореспондент НАПН України, д-р техн. наук, професор, професор кафедри автоматизації та інтелектуальних інформаційних технологій, e-mail: rkvetny@vntu.edu.ua . https://orcid.org/0000-0002-9192-9258.

Вінницький національний технічний університет, Вінниця

S. G. Moskovko[1]
R. N. Kvyetnyy[1]

Method for the Automated Identification of Bounded Contexts in the Design of E-Commerce Systems

[1]Vinnytsia National Technical University

The article examines the issue of automating the strategic stage of e-commerce system design, focusing on modeling complex business logic and structuring the domain. Modern e-commerce systems are characterized by a high level of complexity in business rules and a large number of interconnected processes. Effective management of this complexity is possible through the application of the Domain-Driven Design approach, which involves decomposing the system into logically separated modules known as Bounded Contexts. The quality of identifying these contexts directly affects the system's viability, the clarity of the codebase, and its potential for further evolution.

Traditional approaches to identifying context boundaries are based mainly on heuristic methods and expert sessions, which are resource-intensive and dependent on the human factor. Existing formal methods that rely on structural data analysis are often unable to correctly interpret the semantic nuances of business terminology [1]. This paper proposes an information technology approach that uses the capabilities of generative artificial intelligence and large language models (LLMs) for automated analysis of the problem space.

The main focus of the article is the development of a method for semantic clustering of requirements, which makes it possible to identify hidden linguistic patterns in the description of business processes. The proposed approach involves using LLMs to analyze the project's Ubiquitous Language and to form a Context Map based on the semantic similarity of concepts rather than only their technical relationships. An algorithm is described that transforms textual specifications and user stories into formalized domain models, determining the recommended boundaries of responsibility for each module.

The research results demonstrate that applying a generative approach makes it possible to significantly increase the objectivity of domain modeling, minimize the cognitive load on architects, and ensure validation of the system's logical integrity at the early stages of design. The developed technology serves as an intelligent decision-support tool, enabling the creation of flexible and business-change-adaptive e-commerce models.

Keywords: Bounded Context, Domain-Driven Design, large language models, semantic clustering, Context Map, e-commerce.

Moskovko Serhii G. — Post-Graduate Student of the Chair of Automation and Intelligent Information Technologies, e-mail: smoskovko@icloud.com . https://orcid.org/0009-0001-4376-0187 ;

Kvyetnyy Roman N. — Corresponding Member of the National Academy of Pedagogical Sciences of Ukraine, Dr. Sc. (Eng.), Professor, Professor of the Chair of Automation and Intelligent Information Technologies, e-mail: rkvetny@vntu.edu.ua . https://orcid.org/0000-0002-9192-9258