

Я. Р. Д'ЯКОВ

аспірант кафедри електронних обчислювальних машин
Харківський національний університет радіоелектроніки
ORCID: 0000-0002-6265-7329

МОДЕЛЬ ІЄРАРХІЧНОГО СЕМАНТИЧНОГО ГРАФУ ЗНАНЬ ДЛЯ БАГАТОРІВНЕВОГО ПРЕДСТАВЛЕННЯ ЗНАНЬ

Графи знань є базовою технологією для структурованого представлення знань, логічного виведення та виконання семантичних операцій у складних інформаційних системах. Водночас поширені моделі семантичних графів знань та графів, доповнених великими мовними моделями, працюють з фіксованим масштабом семантичної деталізації та не підтримують багаторівневу абстракцію. Моделі ієрархічних, мультимасштабних графів знань дозволяють конструювати топологічні ієрархії на основі структур мезо- та макрорівнів, водночас додаткових досліджень потребують їхні семантичні властивості та можливі операції над даними. У роботі запропоновано модель ієрархічного семантичного графа знань (ІСГЗ), що описується послідовністю рівнів із власними множинами сутностей та зв'язків і забезпечує багаторівневе представлення знань. Формально визначено структуру ІСГЗ, гібридну метрику семантичної зв'язності, яка поєднує косинусну подібність багатовимірних векторних представлень із структурною мірою Джаккара, а також ітеративний алгоритм ієрархічної агрегації сутностей у рівні-кластери. На основі цих компонентів реалізовано програмний прототип та проведено експериментальну оцінку на підмножинах репрезентативних наборів даних DbPedia50 та YAGO3-10, доповнених попередньо згенерованими багатовимірними векторними представленнями. Показано, що ІСГЗ забезпечує скорочення середнього часу виконання семантичних запитів (до 30–80 % відносно однорівневих моделей та моделей із Louvain/Leiden-кластеризацією) і формує кластери з оптимальнішими показниками семантичної узгодженості та компактності внутрішнього розподілу. Отримані результати підтверджують доцільність використання ІСГЗ та пов'язаних семантичних операцій для побудови масштабованих засобів семантичного аналізу, пошуку та логічного виведення у динамічних базах знань, систем генерації рекомендацій та інтелектуальних помічників.

Ключові слова: граф знань, семантичний граф знань, семантична метрика, багатовимірне векторне представлення, кластеризація.

YA. R. DIAKOV

Postgraduate Student at the Department of Electronic Computers
Kharkiv National University of Radio Electronics
ORCID: 0000-0002-6265-7329

A HIERARCHICAL SEMANTIC KNOWLEDGE GRAPH MODEL FOR MULTILEVEL KNOWLEDGE REPRESENTATION

Knowledge graphs are a fundamental technology for structured knowledge representation, logical inference and semantic operations in complex information systems. However, widely used semantic knowledge graph models and knowledge graphs augmented with large language models operate at a fixed level of semantic detail and do not support multi-level abstraction. Hierarchical, multiscale knowledge graph models allow constructing topological hierarchies based on meso- and macro-level structures, but their semantic properties and possible data operations require additional research. This paper introduces a hierarchical semantic knowledge graph (HSKG) model defined as a sequence of levels with their own sets of entities and relations, enabling multi-level knowledge representation. We formally specify the structure of HSKG, a hybrid semantic connectivity metric that combines cosine similarity of multi-dimensional embeddings with the Jaccard measure on graph neighborhoods, and an iterative hierarchical aggregation algorithm that groups entities into level-clusters. Based on these components, we implement a software prototype and conduct an experimental evaluation on subsets of the DbPedia50 and YAGO3-10 benchmark datasets with pre-generated multi-dimensional embeddings. The results demonstrate that HSKG reduces the average response time of semantic queries (by up to 30–80 % compared to single-level models and Louvain/Leiden-based clustering models) and yields clusters with optimized semantic connectivity and internal distribution compactness. These findings confirm the suitability of the proposed model and related semantic operations for building scalable tools for semantic analysis, search and logical inference in dynamic knowledge bases, recommendation generation systems and intelligent assistants.

Key words: knowledge graph, semantic knowledge graph, semantic metrics, embedding, clustering.

Постановка проблеми

Моделі графів знань (ГЗ) стали основою структурованого представлення знань у численних предметних областях, серед яких медицина, фінанси, штучний інтелект, біоінформатика тощо. Роль означених структур значно зросла з огляду на швидкий розвиток гібридних інтелектуальних систем та великих мовних моделей [1]-[3]. Основний принцип ГЗ – з'єднання фактичних сутностей за допомогою змістовних зв'язків – дозволяє проводити більш комплексний пошук, логічне виведення, моделювання та аналіз залежностей між даними тощо.

Протягом останніх років відбулися якісні зміни у побудові моделей з точки зору їх семантики та гібридизації. Типовими прикладами є семантичні графи знань (СГЗ) та графи знань, що доповнені великими мовними моделями (ГЗДВММ) [4]. Надбудова у вигляді багатовимірних векторного представлення (БВП) його сутностей та зв'язків суттєво покращила розуміння елементів ГЗ засобами електронно-обчислювальних машин: це забезпечило підвищену адаптивність та контекстну узгодженість модельних представлень. Здатність до виявлення семантичної подібності та векторних патернів у таких СГЗ дозволила використовувати у пов'язаних інформаційних технологіях методи пошуку знань, генерації рекомендацій, логічне виведення знань, семантичну кластеризацію тощо [5]. Подані гібридні підходи демонструють продуктивні результати у завданнях типу зв'язування сутностей, семантичного пошуку та багатокрокового логічного виведення.

Основною проблематикою моделей СГЗ та ГЗДВММ є їх фіксований масштаб семантичної деталізації. Так, усі сутності та зв'язки співіснують на певному концептуальному рівні. Така структура не є ефективною для задач багаторівневого представлення знань, де абстрактні концепції є наслідком семантичного узагальнення сутностей нижчих рівнів. У середовищах великих, динамічних баз знань такі обмеження ускладнюють або унеможливають виконання задач, що потребують масштабованих абстракцій, ієрархічного логічного виведення, адекватності моделі тощо.

Наразі останні дослідження зосереджені на ієрархічних, багаторівневих представленнях знань [6]-[7]. Більшість із них орієнтовані переважно на оптимізацію простору БВП та пов'язаних прикладних задач, проте без формального моделювання саме ієрархічних, семантичних моделей.

Для означення шляху до подолання означених обмежень, запропоновано модель ієрархічного семантичного графу знань (ІСГЗ). Внутрішні сутності згруповані у рівні-кластери відповідно до їхньої семантичної зв'язності. Також представлено алгоритм ієрархічної агрегації, що ітеративним шляхом зв'язує семантично подібні сутності у високорівневі абстракції. Сконструйовано метрику зв'язності, що обчислює концептуальну, семантичну близькість сутностей. Зазначені результати – дозволяють створювати та підтримувати цілісність багаторівневих представлень знань, що забезпечує адекватність структури на всіх рівнях абстракцій.

Таким чином, проблема полягає в тому, що наявні моделі СГЗ та ГЗДВММ з фіксованим масштабом семантичної деталізації не забезпечують багаторівневого представлення знань та ієрархічного логічного виведення в умовах великих динамічних баз знань.

Аналіз останніх досліджень і публікацій

Сучасні дослідження, що стосуються графових представлень знань і логіко-семантичних операцій на їх основі, можна згрупувати за такими основними напрямками: сучасні моделі СГЗ та ГЗДВММ; ієрархічні, мультимасштабні графові моделі; онтологічні системи знань.

Використання БВП для ГЗ залишається основним підходом для представлення розподіленої семантики як сутностей, так і зв'язків між ними. Відомі дослідження [7] узагальнюють процес структурних модифікацій ГЗ, типових моделей БВП, а також оптимізації прикладних задач. А також акцентується увага на поступовий перехід до гібридних архітектур, що поєднують графові моделі знань, великі, завчасно натреновані мовні моделі, численні представлення БВП [7].

Паралельно існують інтеграційні моделі, що описують семантику між ГЗ та великими мовними моделями. Мета таких моделей корелює з рішеннями задач покращення точності, ефективності пошуку та здатності до оновлення вихідних даних великих мовних моделей. Ці підходи відкривають можливості до зменшення ефекту явища галюцинації [2][8] та пропонують практичні методи для динамічного оновлення пов'язаних ГЗ, їхнього узгодження з операціями лінгвістичного логічного виведення тощо. Показано важливість моделей графів знань, які є семантично насиченими та придатними до динамічного доповнення [9].

Так, більшість підходів до конструювання моделей СГЗ та ГЗДВММ дають у якості результату однорівневі представлення знань, тобто такі, де сутності та зв'язки знаходяться переважно на одному рівні семантичної деталізації. Таке обмеження не дозволяє ефективно переходити від фактів до високорівневих концепцій, візуалізувати знання різних рівнів, проводити багаторівневе логічне виведення з урахуванням масштабованості та адекватності [10].

Актуальним напрямом досліджень також є багатовимірні представлення знань у просторі БВП. Подібні моделі та підходи дозволяють покращувати типові векторно-семантичні операції, зокрема такі як семантичний пошук та прогнозування зв'язків [11]. Проте пріоритет надається ієрархіям у векторному просторі; відсутня структурна агрегація за семантичною зв'язністю елементів, що є невирішеною задачею.

Існує також напрям ієрархічних, мультимасштабних ГЗ для задачі виокремлення на площині графів знань мезо- та макроструктур. До мезоструктур відносяться: підграфи, патерни відношень, повторювані схеми тощо. А до макроструктур – загальний розподіл вузлів, центрів, дублікатів, зв'язності областей знань тощо. Класичні задачі пошуку сильнозв'язних кластерів сутностей та агрегація групи сутностей однією надсутністю (алгоритми Louvain та Infomap) дозволяють формувати топологічні ієрархії. Графові нейронні мережі з потоковою передачею повідомлень та їхні кластерно-чутливі варіанти у контексті зазначених задач створюють багаторівневі надбудови, що якісно впливають на масштабованість структур даних та забезпечують виконання операцій на великих топологічних відстанях. Загалом такі методи реструктуризації або доповнення графів прискорюють поширення повідомлень у нейронних мережах та чіткіше виокремлюють повторювані схеми [12].

Більшість моделей ієрархічних, мультимасштабних ГЗ, пов'язаних з ними графових нейронних мереж, та підходів у мезо- та макромасштабах, орієнтовані на структурні властивості знань, а саме модульність, векторну близькість. При цьому недостатня увага приділяється саме семантичним властивостям (семантична зв'язність, здатність до зсувів (дрейф)) і операціям (логіко-семантичне виведення, обчислення та контроль дрейфу) тощо.

Останні здобутки у розробці кубічних, регіональних моделей БВП відкрили можливості для формування ієрархічної семантики у векторному просторі [6]. Але вони націлені на оптимізацію векторних властивостей та операцій замість представлення формальної графової структури зі збереженням семантичної зв'язності усередині рівнів абстракцій.

Комплексні онтологічні моделі або онтології, наприклад Resource Description Framework Schema та Web Ontology Language, висувають чіткі таксономічні ієрархії сутностей та механізми логічного виведення для галузей інженерії знань, біомедицини [3] тощо. Нові гібридні підходи (наприклад, Onto2Vec) поєднують структуру онтології з шаром БВП для оптимізації задач прогнозування зв'язків, оцінки подібності елементів. Такі гібриди дозволяють увести семантичну компоненту до операцій над даними, але вимагають налаштувань з боку експертів предметних областей. Вони не дозволяють автоматично виводити високорівневі абстракції з даних, але при цьому вирішують задачі кластеризації, класифікації знань, пошук подібних сутностей [13].

Отже, наразі у наукових дослідженнях можна виокремити можливості для: представлення знань на основі СГЗ; покращеного логічного виведення за рахунок інтеграції великих мовних моделей; проведення структурно-ієрархічних операцій на площині мезо- та макромасштабних рівнів знань; впровадження онтологій, що представляють комплексні ієрархії сутностей вищого рівня. Проте бракує уніфікованої графової багаторівневої моделі знань, де кожен рівень зберігає семантичну зв'язність, підтримуючи можливості візуалізації та логічного виведення. Таким чином, постає потреба у формуванні моделі ІСГЗ, що поєднує формальну багаторівневу структуру з алгоритмом ієрархічної агрегації сутностей та метрикою семантичної зв'язності.

Таким чином, жоден із розглянутих підходів не забезпечує одночасно формально визначеної багаторівневої графової структури та метрик семантичної зв'язності для рівнів абстракції, що й обумовлює актуальність запропонованої в роботі моделі ІСГЗ.

Формулювання мети дослідження

Метою даного дослідження є побудова формальної моделі ІСГЗ, яка забезпечує багаторівневе представлення знань та підтримує ієрархічне логічне виведення в умовах динамічних баз знань. Для досягнення поставленої мети необхідно розв'язати такі наукові задачі:

- розробити формальну модель ІСГЗ, що розширює стандартну модель СГЗ шляхом впровадження багаторівневих абстракцій;
- визначити метрику, що кількісно оцінює структурно-семантичну подібність усередині рівнів-кластерів, а також підтримує ієрархічну агрегацію;
- розробити ітеративний алгоритм побудови ієрархії рівнів абстракції, що зберігають семантичну узгодженість елементів;
- провести тестування моделі на локальних даних та виконати емпіричну оцінку на підмножинах публічних наборів даних.

Викладення основного матеріалу дослідження

Формальна модель ієрархічного семантичного графа знань

ІСГЗ розглядається як узагальнення стандартного семантичного графа знань, у якому знання подаються не в одному фіксованому семантичному шарі, а на послідовності рівнів абстракції. Така багаторівнева організація дає змогу одночасно працювати як з детальними фактами, що безпосередньо відповідають вихідним даним, так і з агрегованими узагальненнями, які відображають більш високі концептуальні структури предметної області. Іншими словами, ІСГЗ поєднує мікрорівень локальних взаємозв'язків між сутностями та макрорівень їхніх семантичних кластерів, формуючи узгоджену ієрархію представлень.

Формально модель ІСГЗ задається множиною рівнів, кожен з яких є окремим графом на власній множині сутностей та зв'язків. Модель ІСГЗ можна формалізувати наступним чином:

$$G = \{L_0, L_1, \dots, L_N\}, \quad (1)$$

де: G – ієрархічний семантичний граф знань; L_n – рівень графу, що відповідає певному ступеню абстракції знань ($0 \leq n \leq N$).

Кожен рівень L_n у (1) задається впорядкованою парою:

$$L_n = \{V_n, U_n\}, \quad (2)$$

де: V_n – множина сутностей (вершин) на рівні n ; U_n – множина зв'язків (ребер), що відображають семантичні відношення між елементами з V_n .

Таким чином, кожен рівень L_n (2) є однорівневим семантичним графом, а вся система рівнів $\{L_n\}$ задає ієрархію графів знань.

Основною вимогою до моделі є наявність для кожної сутності її багатовимірною векторного представлення в узгодженому просторі ознак. Для цього припускається, що

$$\forall v \in V_n \exists \varepsilon_v \in \mathbb{R}^d, \quad (3)$$

де v – окрема сутність на рівні n ; ε_v – БВП сутності v ; \mathbb{R}^d – d -вимірний векторний простір ознак.

Вектор ε_v (3) узагальнює семантичні характеристики сутності та забезпечує можливість кількісно оцінювати подібність між сутностями різних рівнів за допомогою стандартних метрик у \mathbb{R}^d . Розмірність d визначає ємність простору представлень – більші значення d дозволяють кодувати тонші семантичні відмінності, але збільшують обчислювальну складність обробки моделі.

Оскільки ІСГЗ є ієрархічною конструкцією, рівні пов'язані між собою відношеннями успадкування й узагальнення. Наступний рівень графу формується шляхом агрегування сутностей та зв'язків попереднього рівня, тобто:

$$L_{n+1} = \{V_{n+1}, U_{n+1}\}. \quad (4)$$

У (4) множина V_{n+1} інтерпретується як множина абстрактних сутностей (кластерів), утворених на основі елементів V_n . Множина U_{n+1} індукується зі зв'язків U_n з урахуванням обраної схеми агрегування. Таким чином, кожний елемент V_{n+1} відповідає підмножині сутностей з V_n , які є семантично пов'язаними за певним критерієм (наприклад, за метрикою подібності, що вводиться у наступних підрозділах). Аналогічно, зв'язки з U_{n+1} відображають узагальнені відношення між цими кластерами.

Рівень L_0 є базовим, множини V_0 та U_0 відповідають вихідному семантичному графу знань, побудованому безпосередньо за даними предметної області (онтологіями, базами фактів тощо). Вищі рівні L_1, \dots, L_N відображають послідовні кроки узагальнення структури L_0 : зі збільшенням n зменшується кількість сутностей, але зростає ступінь їхньої абстракції. Значення N визначає максимальний рівень ієрархії і залежить від властивостей первинного набору даних, необхідного масштабу семантичної деталізації та обмежень на складність моделі (допустиму кількість рівнів, вершин і зв'язків на кожному рівні).

Метрика семантичної зв'язності

Метрика семантичної зв'язності φ_n використовується при формуванні рівнів-кластерів ІСГЗ і на кожному рівні L_n задає міру близькості між парами сутностей. Формально для рівня n розглядається функція

$$\varphi_n : V_n \times V_n \rightarrow [0, 1],$$

яка для пари сутностей $v_p, v_q \in V_n$ задається наступною формулою:

$$\varphi_n(v_p, v_q) = \alpha \cdot \text{sml}_\varepsilon(\varepsilon_p, \varepsilon_q) + (1 - \alpha) \cdot \text{sml}_r(v_p, v_q), \quad (5)$$

де $\varepsilon_p, \varepsilon_q \in \mathbb{R}^d$ – відповідні багатовимірні векторні представлення сутностей v_p та v_q ; sml_ε – семантична міра косинусної подібності між векторами ε_p та ε_q [14]; sml_r – структурна міра Джаккара для множин суміжних вершин сутностей v_p та v_q ; $\alpha \in [0, 1]$ – константа, що визначає відносний внесок семантичної та структурної компонент у (5).

Поєднання структурної зв'язності зі зв'язністю на основі БВП дає змогу збалансувати вплив локальної топології графу та семантичного змісту сутностей на конкретному рівні L_n . За малих значень α переважає топологічний компонент, тоді як за α , близьких до одиниці, метрика більш чутлива до семантичної подібності векторних представлень.

Для чіткого визначення факту семантичної близькості двох сутностей вводиться порогове відношення

$$\varphi_n(v_p, v_q) = C_n, \quad (7)$$

де $C_n \in [0, 1]$ – гранична константа для рівня L_n , яка одночасно виступає як критерій семантичної деталізації цього рівня.

За більших значень C_n формуються більш щільні кластери з високою внутрішньою подібністю, тоді як менші значення C_n допускають більш різномірні групи сутностей.

Алгоритм ієрархічної агрегації

Алгоритм ієрархічної агрегації формалізує процес узгодження сутностей та зв'язків під час переходу від рівня L_n до рівня L_{n+1} . На інтуїтивному рівні він перетворює множину більш детальних сутностей V_n на множину абстрактніших сутностей V_{n+1} , а зв'язки U_n – на узагальнені зв'язки U_{n+1} , зберігаючи при цьому семантичну структуру графа.

Формально перехід між рівнями задається так:

$$L_{n+1} = f_\varepsilon(L_n, \varphi_n, \theta_n). \quad (6)$$

У (6) f_ε – математична формалізацією алгоритму, параметризованого сутностями та зв'язками попереднього рівня L_n , метрикою семантичної зв'язності φ_n та граничним коефіцієнтом близькості сутностей у кластері θ_n .

Алгоритм можна представити наступною послідовністю кроків.

Крок 1. Кластеризація сутностей на основі метрики семантичної зв'язності. Необхідно побудувати двовимірну матрицю результатів виконання семантичної метрики зв'язності над кожними двома сутностями графу G . Для цього застосовується алгоритм агломеративної кластеризації [15] до означеної матриці. Далі необхідно згрупувати сутності за кластерами відповідно до отриманих результатів.

Крок 2. Уніфікація сутностей та типів зв'язків наступного рівня на основі зібраних кластерів. Для кожного кластера необхідно визначити шляхом пошуку нове БВП та зважену типологію зв'язків. БВП обчислюється як середнє між усіма БВП сутностей кластера. Зважена типологія задається шляхом підрахунку зв'язків та їхніх типів між сутностями кластера.

Крок 3. Індукція зв'язків. Для кожної пари центральних БВП кластерів необхідно знайти пріоритетний тип зв'язку та на його основі будується безпосередній зв'язок.

Застосування описаного алгоритму дає змогу з рівня більш конкретних сутностей та зв'язків L_n отримати рівень L_{n+1} , на якому знання подані в більш абстрактній формі, але зберігають вихідну семантичну організацію. Ітеративний характер процедури дозволяє повторно застосовувати f_ε до згенерованих рівнів, послідовно формуючи вищі рівні ієрархії та створюючи все більш високорівневі абстракції.

Умовою завершення роботи алгоритму виступає обмеження на чисельність кластерів найбільш абстрактного рівня або досягнення заданого критерію узагальнення структури графа.

Експериментальна оцінка

Експерименти проводились за допомогою скрипта на мові програмування Python та її бібліотек: numpy (швидкісні векторні обчислення), rukeyen (великі набори даних, методи тренування БВП), community (Louvain-кластеризація), leidenalg (Leiden-кластеризація), networkx (допоміжні структури), а також спеціалізовані пакети для тестування та візуалізації [16].

Для проведення експериментів було використано наступні великі набори даних:

– DbPedia50 (≈ 24000 сутностей та ≈ 350 зв'язків). Підмножина великої бази знань DbPedia, що створена з триплетів (суб'єкт, відношення, об'єкт). Зазначені триплети пов'язані з п'ятдесятьма типами сутностей, що зустрічаються найчастіше у DbPedia [17];

– YAGO3-10 (≈ 123000 сутностей та 37 зв'язків). Підмножина великої бази знань YAGO, що поєднує дані з WordNet (лексична база даних англійської мови) та GeoNames (база даних географічних місць планети Земля) та інших джерел даних, створена із сутностей, які беруть участь у більш ніж десяти відношеннях YAGO [18].

Означені набори даних є опорними для оцінки моделей БВП знань, семантичних операцій, зокрема пошуку, логічного виведення, динамічного поповнення тощо. Проте в них відсутні дані стосовно типових властивостей відношень (наприклад, комутативність), що знижує якість результатів кластеризації та семантичних операцій.

Для оптимізації часу ініціалізації програмних моделей, було взято підмножини поданих наборів даних з кількістю сутностей 1024 (DbPedia50 / 1024, YAGO3-10 / 1024) та 8096 (DbPedia50 / 8096, YAGO3-10 / 8096). Відповідно у графові структури було додано усі БВП та можливі зв'язки.

Особливості моделі ІСГЗ на програмному рівні дозволили впровадити деякі покращення, зокрема:

– введено метрику (косинусна збіжність) до механізму кластеризації для попередньої фільтрації пар сутностей, які гарантовано не пройдуть обмеження на граничну константу C_n метрики φ ;

– спрощено механізм обчислення структурної метрики Джакарда шляхом використання попередньо збереженої щільності відношень сутностей нижнього рівня графа.

Для порівняльного аналізу було розглянуто наступні моделі:

– класичний однорівневий ГЗ з шаром БВП (на метриках – Flat KG);

– СГЗ із використанням Louvain-кластеризації (алгоритм пошуку груп сутностей, що використовує евристичні метрики для виділення кластерів [19]);

– СГЗ із використанням Leiden-кластеризації (модифікація Louvain, що надає більшу масштабованість, зв'язність кластерів, але справляється із задачею за більший час [19]).

У контексті семантичних операцій здійснюються наступні перевірки:

– екстракція елемента ГЗ з його БВП. Класична атомарна операція, що надає інформацію стосовно сутності, її позиціонування у графі, та доцільна для пошуку, логічного виведення, кластеризації тощо;

– пошук сутностей на заданій відстані. Семантично-топологічна операція, створена для задач аналізу локального контексту сутностей, кластеризації, обґрунтування логічного виведення тощо;

– пошук найближчих сутностей за багатовимірними векторними представленнями (ПНСБВП). Фундаментальний елемент у семантичному пошуку, кластеризації, генерації рекомендацій тощо.

Вище перелічені операції дозволяють:

– проводити семантичний доступ до елементів СГЗ, а також загальну навігацію;

– аналізувати та розширювати контекст запиту до знань;

– формувати пояснення того чи іншого виведеного факту;

– оцінювати зв'язність графів та підграфів тощо.

Семантичні операції екстракції сутності з БВП та пошуку сутностей на заданій відстані були виконані однаково ефективно як за допомогою ІСГЗ, так і опорними моделями, зокрема похибка ефективності склала $\approx 2\%$. Отримання такого результату є логічним з огляду на повне використання алгоритмічно простих операцій та допоміжних конструкцій усіма моделями. У випадку ІСГЗ, виконання ієрархічних версій зазначених операцій показало аналогічну ефективність, що підкреслює властивість масштабованості моделі.

ПНСБВП продемонстрував суттєву перевагу моделі ІСГЗ у порівнянні з опорними моделями (Flat KG, Louvain, Leiden). На малих наборах даних (DbPedia50/1024, YAGO3-10/1024) середній час виконання запиту в ІСГЗ зменшився до 30 % відносно однорівневого ГЗ (Flat KG) та на 5–20 % – відносно моделей з Louvain- та Leiden-кластеризацією (рис. 1: а, б). Іншими словами, вже на відносно невеликих графах ієрархічне подання знань дає стабільний вииграш, не поступаючись за швидкодією жодній з порівнюваних схем.

На великих наборах даних (DbPedia50/8096, YAGO3-10/8096) перевага ІСГЗ стає ще виразнішою (рис. 2: а, б). Середній час виконання ПНСБВП скорочується до 80 % у порівнянні з однорівневим ГЗ та на 40–70 % – у порівнянні з моделями на основі Louvain- і Leiden-кластеризації. Це означає, що зі зростанням розміру графа ієрархічна організація ІСГЗ дозволяє істотно зменшити ефективний простір пошуку за рахунок того, що запит обробляється спочатку на узагальнених рівнях, а лише потім переходить до детальніших сутностей, тоді як Flat KG та класичні алгоритми кластеризації працюють фактично в одному масштабі деталізації.

Для дослідження узгодженості семантичних зв'язків було проаналізовано розподіл показника зв'язності всередині кластерів між ІСГЗ та опорною моделлю з Louvain-кластеризацією на множині DbPedia50/8096 (рис. 3). Інтервал між середнім арифметичним та медіаною розподілу для ІСГЗ становить 0.82–0.84, а для Louvain – лише 0.19–0.22. Крім того, значення показника для ІСГЗ згруповані значно компактніше, що вказує на меншу варіативність якості кластерів. Отже, кластери ІСГЗ є чисельно згуртованішими та семантично одноріднішими, тоді як Louvain формує суттєво більш різноманітні за змістом групи, орієнтуючись переважно на топологію графа.

Таким чином, результати на рис. 1-3 демонструють, що ІСГЗ одночасно забезпечує прискорення семантичних запитів і підвищення когерентності кластерів, тобто покращує як обчислювальну, так і змістову якість обробки знань у порівнянні з однорівневим представленням та класичними алгоритмами кластеризації.

Обговорення результатів

Експериментальні дослідження дозволили виокремити кілька ключових чинників, які визначають якість моделі ІСГЗ та ефективність запропонованих методів в обмежувальних умовах:

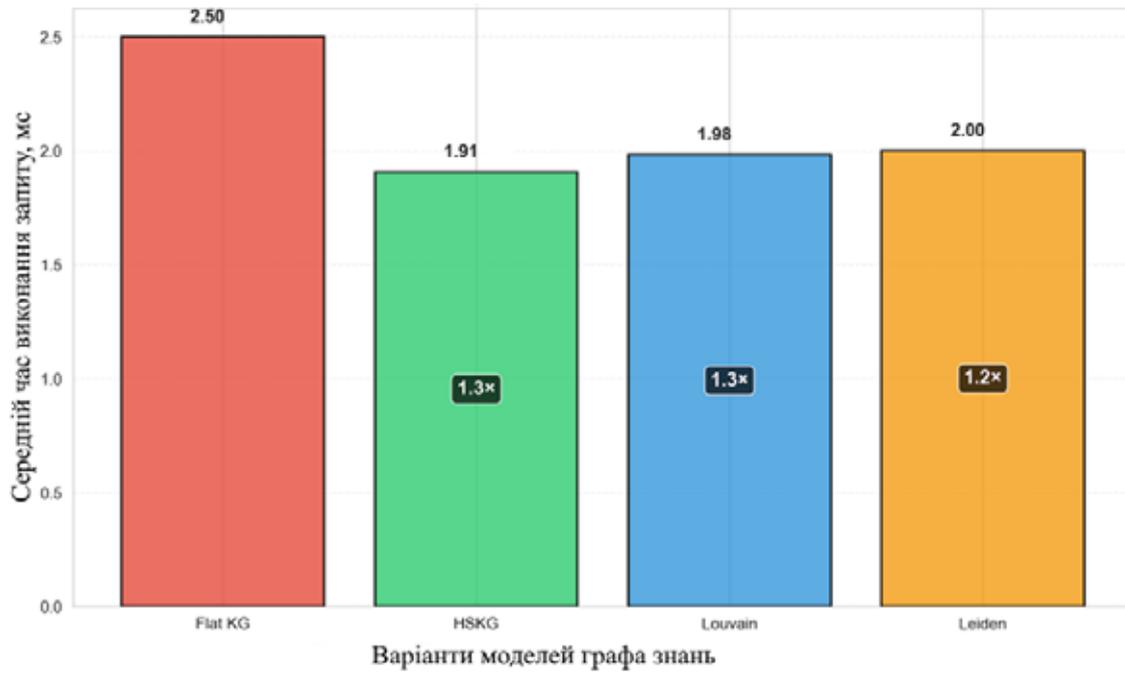
– попередній аналіз набору даних для гіперпараметризації ключових констант алгоритму, зокрема граничного коефіцієнта близькості сутностей у кластері θ_n та критерію семантичної деталізації рівня C_n ;

– обчислення й збереження агрегованих метрик нижчого рівня у кластерах вищих рівнів, що покращує локальний контекст для семантичних операцій та логічного виведення;

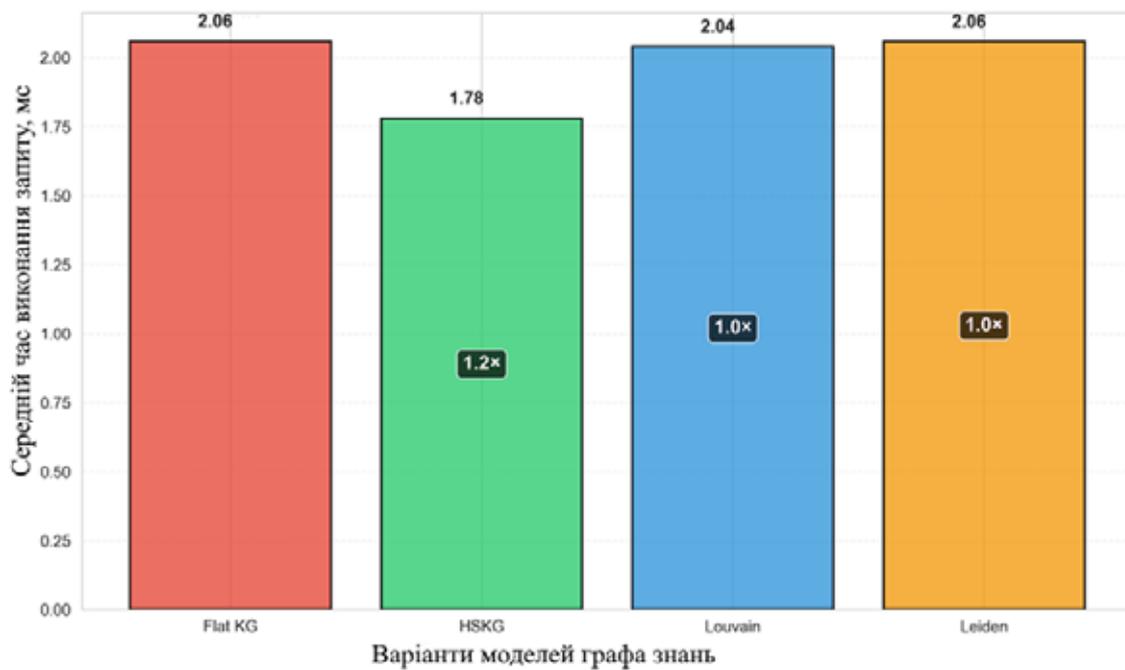
– висока розрідженість набору даних, за якої до механізму кластеризації доцільно вводити додаткові обмеження, щоб уникати зайвого обчислення базової метрики та прискорювати генерацію наступного рівня;

– досягнення граничної кількості сутностей на рівні, коли ефективними стають розосереджені методи кластеризації, що дозволяють пришвидшити формування рівнів-кластерів графа.

На експериментальні результати істотно вплинула також обмежена або повна відсутність комутативності у зв'язках опорних наборів даних. Використана в метриці семантичної зв'язності ϕ структурна міра Джаккара покриває топологічну компоненту графа знань і суттєво залежить від щільності зв'язків між сутностями. Якщо відсутні обернені зв'язки або дані щодо їх комутативності, це погіршує здатність алгоритму ієрархічної агрегації формувати семантично цільні кластери.

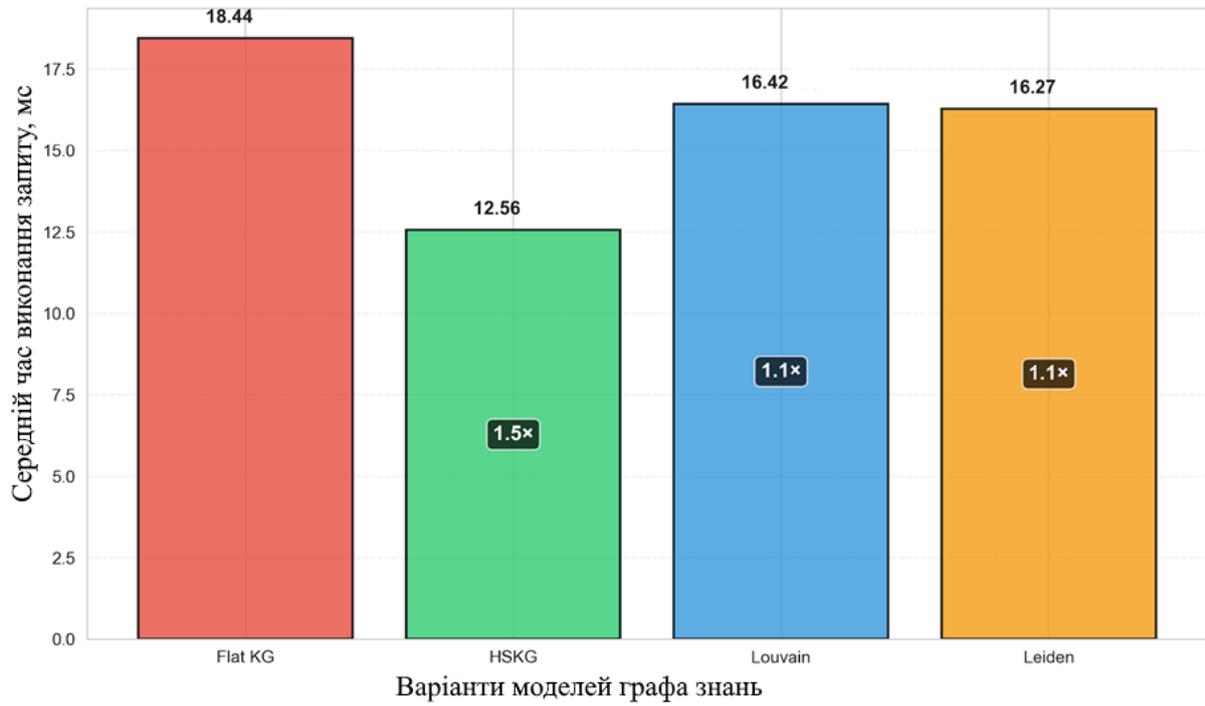


а)

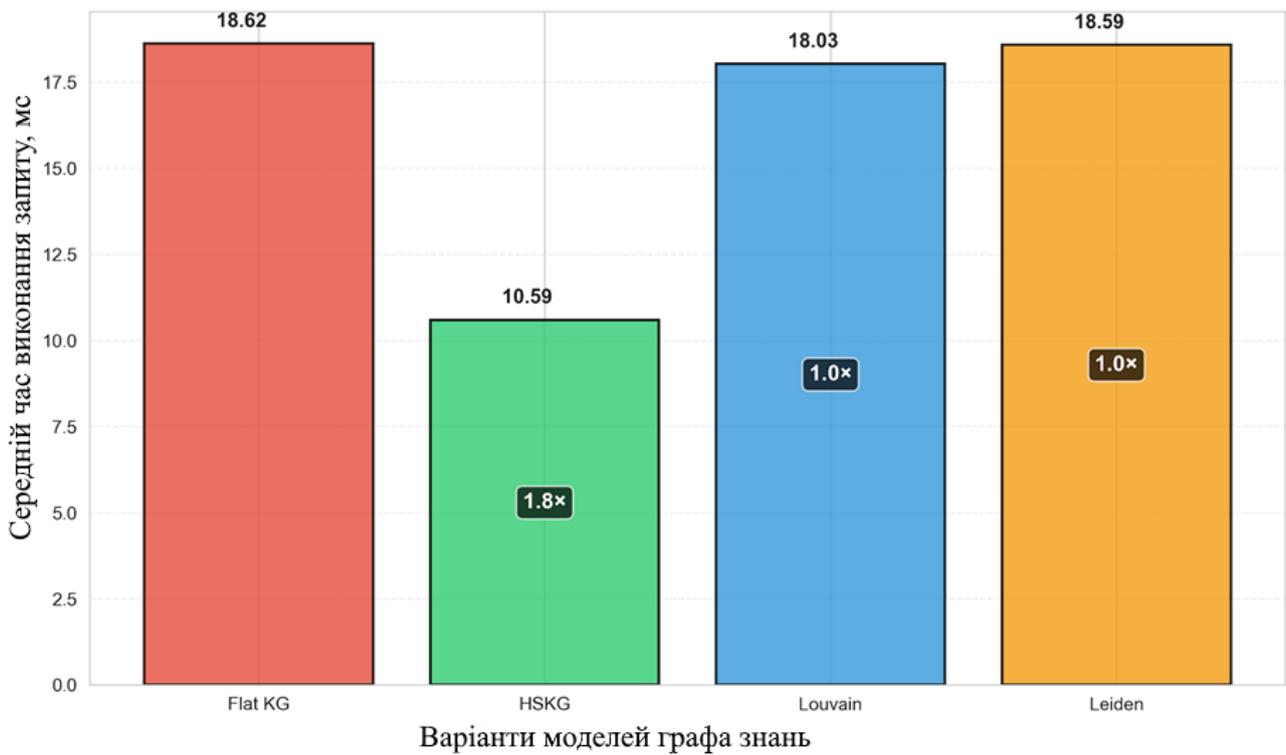


б)

Рис. 1. Усереднена ефективність виконання ПНСБВП на зрізах: а) DbPedia50 / 1024; б) YAGO3-10 / 1024



а)



б)

Рис. 2. Усереднена ефективність виконання ПНСБВП на зрізах: а) DbPedia50 / 8096; б) YAGO3-10 / 8096

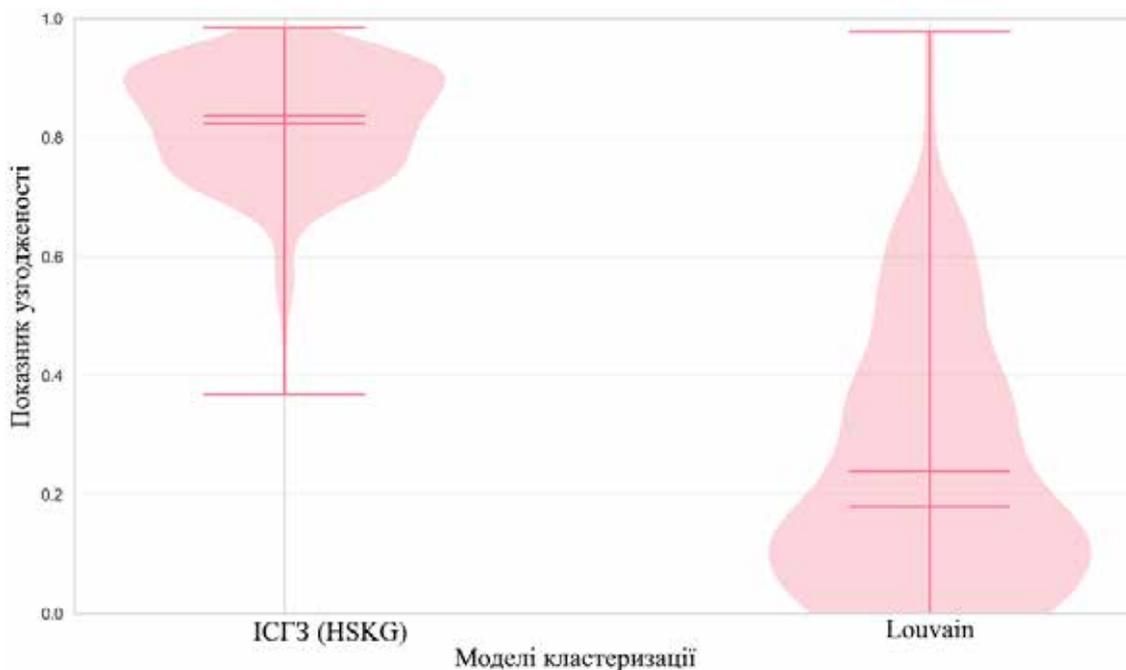


Рис. 3. Порівняння розподіленості кластерної зв'язності на множині DbPedia50 / 8096 між ІСГЗ та опорною моделлю з Louvain-кластеризацією

Висновки

У результаті проведеного дослідження набуло подальшого розвитку розв'язання проблемного питання побудови моделей семантичних графів знань, зокрема запропонована модель у вигляді ієрархічного семантичного графа знань, що описується послідовністю рівнів L_n з множинами сутностей V_n та зв'язків U_n . Поєднання багатовимірних векторних представлень із формальною ієрархією рівнів абстракції дозволило виконувати семантичні операції в різних масштабах деталізації без втрати узгодженості знань.

Запропонована метрика семантичної зв'язності ϕ_n , яка комбінує косинусну подібність БВП та структурну міру Джаккара з пороговими константами C_n і θ_n , забезпечила побудову ітеративного алгоритму ієрархічної агрегації f_n . Це дозволило формувати рівні-кластери з вищою структурно-семантичною цілісністю, ніж у класичних однорівневих СГЗ та моделях з Louvain/Leiden-кластеризацією.

Експериментальна оцінка на підмножинах DbPedia50 та YAGO3-10 показала, що ІСГЗ забезпечує скорочення середнього часу семантичних запитів (до 30–80 % відносно Flat KG і 5–70 % відносно Louvain/Leiden) та значно більш однорідні кластери (інтервал 0.82–0.84 проти 0.19–0.22 для Louvain). Це підтверджує перевагу ієрархічної організації знань з обчислювальною та зі змістовою точок зору.

Подальші дослідження доцільно спрямувати на урахування властивостей відношень (комутативність, симетричність, транзитивність), автоматизоване налаштування порогів C_n та θ_n й інтеграцію ІСГЗ з моделями глибокого навчання. Практичні перспективи моделі пов'язані з семантичними пошуковими системами, системами рекомендацій і підтримки рішень [20], де потрібні масштабований доступ до знань.

Список використаної літератури

1. ESCARGOT: An AI agent leveraging large language models, dynamic graph of thoughts, and biomedical knowledge graphs for enhanced reasoning / N. Matsumoto et al. *Bioinformatics*, Volume 41, Issue 2, 2025. DOI: <https://doi.org/10.1093/bioinformatics/btaf031>.
2. Unifying Large Language Models and Knowledge Graphs: A Roadmap / S. Pan et al. *IEEE Transactions on Knowledge and Data Engineering*. 2024. P. 1–20. DOI: <https://doi.org/10.1109/tkde.2024.3352100>.
3. Large language models for intelligent RDF knowledge graph construction: results from medical ontology mapping / A. Mavridis et al. *Frontier in Artificial Intelligence*, Volume 8, No 1546179, 2025. DOI: <https://doi.org/10.3389/frai.2025.1546179>.
4. Zichao Li, Zong Ke, Puning Zhao. Injecting Structured Knowledge into LLMs via Graph Neural Networks. In *Proceedings of the 1st Joint Workshop on Large Language Models and Structure Modeling (XLLM 2025)*, Vienna, Austria. Association for Computational Linguistics. C. 16–25. URL: <https://aclanthology.org/2025.xllm-1.3.pdf>

5. Large Language Models for Knowledge Graph Embedding: A Survey / B. Liu et al. *Mathematics*, Volume 13, Issue 14, 2025. DOI: <https://doi.org/10.3390/math13142244>.
6. Comprehensive survey on hierarchical clustering algorithms and the recent developments / X. Ran et al. *Artificial Intelligence Review*. 2022. DOI: <https://doi.org/10.1007/s10462-022-10366-3>.
7. GeOKG: geometry-aware knowledge graph embedding for Gene Ontology and genes / C. Jeong et al. *Bioinformatics*, Volume 41, Issue 4, 2025. DOI: <https://doi.org/10.1093/bioinformatics/btaf160>.
8. Detecting hallucinations in large language models using semantic entropy / S. Farquhar et al. *Nature*. 2024. Vol. 630, no. 8017. P. 625–630. DOI: <https://doi.org/10.1038/s41586-024-07421-0>.
9. Zihao Xu, Zhejun Shen, Qunzhi Zhou, Petar Ristoski. Knowledge Graph-Enhanced Retrieval Augmented Generation for E-Commerce / *RAGE-KG 2024: Retrieval-Augmented Generation Enabled by Knowledge Graphs*, November 11, 2024, Baltimore, Maryland. URL: https://2024.rage-kg.org/papers/RAGE-KG_2024_paper_6.pdf
10. Knowledge Graph Embedding for Link Prediction / A. Rossi et al. *ACM Transactions on Knowledge Discovery from Data*. 2021. Vol. 15, no. 2. P. 1–49. DOI: <https://doi.org/10.1145/3424672>.
11. Linders J., Tomczak J. M. Knowledge graph-extended retrieval augmented generation for question answering. *Applied Intelligence*. 2025. Vol. 55, no. 17. DOI: <https://doi.org/10.1007/s10489-025-06885-5>.
12. Zhong Z., Li C., Pang J. Hierarchical message-passing graph neural networks. *Data Mining and Knowledge Discovery*, Volume 37, 2023. P. 381-408. DOI: <https://doi.org/10.1007/s10618-022-00890-9>.
13. Smaili F. Z., Gao X., Hoehndorf R. Onto2Vec: joint vector-based representation of biological entities and their ontology-based annotations. *Bioinformatics*, Volume 34, Issue 13, 2018. P. i52–i60. DOI: <https://doi.org/10.1093/bioinformatics/bty259>.
14. Information retrieval using cosine and jaccard similarity measures in vector space model / A. Jain et al. *International Journal of Computer Applications*, Volume 164, No 6, 2017. P. 28-30. DOI: <https://doi.org/10.5120/ijca2017913699>.
15. Hierarchical clustering that takes advantage of both density-peak and density-connectivity / Y. Zhu et al. *Information Systems*. 2022. Vol. 103. P. 101871. URL: <https://doi.org/10.1016/j.is.2021.101871>.
16. ProjectHSKG: Project for Hierarchical Semantic Knowledge Graph core & experiments demonstration. URL: <https://github.com/YANZAAAN/ProjectHSKG> (дата звернення: 18.11.2025).
17. Open-world knowledge graph completion for unseen entities and relations via attentive feature aggregation / B. Oh et al. *Information Sciences*. 2022. Vol. 586. P. 468–484. DOI: <https://doi.org/10.1016/j.ins.2021.11.085>.
18. Mahdisoltani F., Biega J., Suchanek F. M. YAGO3: A Knowledge Base from Multilingual Wikipedias. *Proceedings of the 7th Conference on Innovative Data Systems Research*, Asilomar, CA, USA, 2015. URL: <https://imt.hal.science/hal-01699874v1>.
19. Traag, V., Waltman, L., van Eck, N. From Louvain to Leiden: guaranteeing well-connected communities. *Scientific Reports*, Volume 9, Article number 5233, 2019. DOI: <https://doi.org/10.1038/s41598-019-41695-z>.
20. Information and analytical support of gender audit of educational management system / T. Fesenko et al. *Bulletin of NTU "KhPI". Series: Strategic management, portfolio, program and project management*. 2022. No. 1(5). P. 70–79. DOI: <https://doi.org/10.20998/2413-3000.2022.5.9>.

References

1. Matsumoto, N., et al. (2025). ESCARGOT: An AI agent leveraging large language models, dynamic graph of thoughts, and biomedical knowledge graphs for enhanced reasoning. *Bioinformatics*, 41(2). <https://doi.org/10.1093/bioinformatics/btaf031>.
2. Pan, S., et al. (2024). Unifying large language models and knowledge graphs: A roadmap. *IEEE Transactions on Knowledge and Data Engineering*, 1–20. <https://doi.org/10.1109/TKDE.2024.3352100>.
3. Mavridis, A., et al. (2025). Large language models for intelligent RDF knowledge graph construction: Results from medical ontology mapping. *Frontiers in Artificial Intelligence*, 8, 1546179. <https://doi.org/10.3389/frai.2025.1546179>.
4. Li, Z., Ke, Z., & Zhao, P. (2025). Injecting structured knowledge into LLMs via graph neural networks. In *Proceedings of the 1st Joint Workshop on Large Language Models and Structure Modeling (XLLM 2025)* (pp. 16–25). Association for Computational Linguistics. <https://aclanthology.org/2025.xllm-1.3.pdf>.
5. Liu, B., et al. (2025). Large language models for knowledge graph embedding: A survey. *Mathematics*, 13(14). <https://doi.org/10.3390/math13142244>.
6. Ran, X., et al. (2022). Comprehensive survey on hierarchical clustering algorithms and the recent developments. *Artificial Intelligence Review*. <https://doi.org/10.1007/s10462-022-10366-3>.
7. Jeong, C., et al. (2025). GeOKG: Geometry-aware knowledge graph embedding for Gene Ontology and genes. *Bioinformatics*, 41(4). <https://doi.org/10.1093/bioinformatics/btaf160>.
8. Farquhar, S., et al. (2024). Detecting hallucinations in large language models using semantic entropy. *Nature*, 630(8017), 625–630. <https://doi.org/10.1038/s41586-024-07421-0>.

9. Xu, Z., Shen, Z., Zhou, Q., & Ristoski, P. (2024, November 11). Knowledge graph-enhanced retrieval augmented generation for e-commerce. In *RAGE-KG 2024: Retrieval-Augmented Generation Enabled by Knowledge Graphs*, Baltimore, MD, United States. https://2024.rage-kg.org/papers/RAGE-KG_2024_paper_6.pdf.
10. Rossi, A., et al. (2021). Knowledge graph embedding for link prediction: A comparative analysis. *ACM Transactions on Knowledge Discovery from Data*, 15(2), 1–49. <https://doi.org/10.1145/3424672>.
11. Linders, J., & Tomczak, J. M. (2025). Knowledge graph-extended retrieval augmented generation for question answering. *Applied Intelligence*, 55(17). <https://doi.org/10.1007/s10489-025-06885-5>.
12. Zhong, Z., Li, C., & Pang, J. (2023). Hierarchical message-passing graph neural networks. *Data Mining and Knowledge Discovery*, 37, 381–408. <https://doi.org/10.1007/s10618-022-00890-9>.
13. Smaili, F. Z., Gao, X., & Hoehndorf, R. (2018). Onto2Vec: Joint vector-based representation of biological entities and their ontology-based annotations. *Bioinformatics*, 34(13), i52–i60. <https://doi.org/10.1093/bioinformatics/bty259>.
14. Jain, A., et al. (2017). Information retrieval using cosine and Jaccard similarity measures in vector space model. *International Journal of Computer Applications*, 164(6), 28–30. <https://doi.org/10.5120/ijca2017913699>.
15. Zhu, Y., et al. (2022). Hierarchical clustering that takes advantage of both density-peak and density-connectivity. *Information Systems*, 103, 101871. <https://doi.org/10.1016/j.is.2021.101871>.
16. Diakov, Y. ProjectHSKG: Project for hierarchical semantic knowledge graph core & experiments demonstration [GitHub repository]. GitHub. Retrieved November 18, 2025, <https://github.com/YANZAAN/ProjectHSKG>.
17. Oh, B., Seo, S., Hwang, J., Lee, D., & Lee, K.-H. (2022). Open-world knowledge graph completion for unseen entities and relations via attentive feature aggregation. *Information Sciences*, 586, 468–484. <https://doi.org/10.1016/j.ins.2021.11.085>.
18. Mahdisoltani, F., Biega, J., & Suchanek, F. M. (2015). YAGO3: A knowledge base from multilingual Wikipedias. In *Proceedings of the 7th Conference on Innovative Data Systems Research (CIDR 2015)*. <https://imt.hal.science/hal-01699874v1>.
19. Traag, V., Waltman, L., & van Eck, N. (2019). From Louvain to Leiden: Guaranteeing well-connected communities. *Scientific Reports*, 9, 5233. <https://doi.org/10.1038/s41598-019-41695-z>.
20. Fesenko, T., et al. (2022). Information and analytical support of gender audit of educational management system. *Bulletin of NTU “KhPI”. Series: Strategic Management, Portfolio, Program and Project Management*, 1(5), 70–79. <https://doi.org/10.20998/2413-3000.2022.5.9>.

Дата першого надходження рукопису до видання: 16.11.2025
Дата прийнятого до друку рукопису після рецензування: 15.12.2025
Дата публікації: 31.12.2025