https://doi.org/10.31891/2307-5732-2025-355-100
УДК 004.67

ХОМЕНКО СЕРГІЙ
Харківський національний університет радіоелектроніки
https://orcid.org/0000-0001-7185-9101
e-mail: serhii.khomenko@nure.ua
ШУБІН ІГОР
Харківський національний університет радіоелектроніки
https://orcid.org/0000-0002-1073-023X
e-mail: igor.shubin@nure.ua

МЕТОДИ ПІДТРИМКИ, УПРАВЛІННЯ ТА ІНТЕГРАЦІЇ РОЗРІЗНЕНОЇ ІНФОРМАЦІЇ

В роботі запропоновано онтологічну методологію семантичної інтеграції різнорідних освітніх ресурсів – текстів, презентацій, відео-матеріалів та лабораторних посібників на основі Resource Description Framework (RDF). Кожен фрагмент описується набором триплетів «суб'єкт-предикат-об'єкт», що дає змогу автоматично виявляти дублікати, термінологічні колізії й прогалини знань. Для кількісного контролю якості введено три показники: V_n – повнота анотацій, V_m – узгодженість термінів з онтологією, V_b – рівень відсутності протиріч. На практиці методика дозволяє викладачеві формувати персоналізовані курси й автоматично генерувати тестові та практичні завдання різного рівня складності. Життєздатність підходу перевірена рольовим експериментом «комп'ютер-з-комп'ютером», де дві програмні сутності (генератор і редактор) поетапно розширювали базу знань дисципліни «Основи програмування». Вимірювання показників V_n, V_m та V_b на кожному кроці підтвердило можливість своєчасного виявлення конфліктів і поступового підвищення узгодженості даних. Отримані результати свідчать про ефективність RDF-анотування для гнучкого конструювання навчальних програм та розвиток інтелектуальних освітніх систем.

Ключові слова: віртуальне навчання, онтологія, семантична інтеграція, освітні ресурси, коефіцієнт повноти, узгодженість знань.

KHOMENKO SERHII
SHUBIN IHOR
Kharkiv National University of Radio Electonics

METHODS OF SUPPORT, MANAGEMENT AND INTEGRATION OF HIGHLY DIVERSIFIED INFORMATION

The paper introduces an ontology-driven methodology for semantic integration of heterogeneous educational resources into a unified knowledge environment built on the Resource Description Framework (RDF). Printed textbooks, slide decks, video lectures, laboratory manuals and other learning artefacts are decomposed into elementary "subject–predicate–object" triples. This representation enables the platform to (i) identify duplicate or semantically equivalent notions (e.g., "two-dimensional array" vs. "matrix"), (ii) detect terminology conflicts across independently authored materials, and (iii) reconstruct implicit prerequisite chains between concepts.

To quantify annotation, we define three formal indicators. V_n measures coverage, i.e. the ratio of unique informative fragments captured by triples to the potential number of knowledge units in the source. V_m reflects term alignment by comparing the set of classes and properties used in an annotation with those already present in the domain ontology. V_b estimates logical consistency as the share of statements that do not contradict previously validated facts. These coefficients are calculated automatically after every ingestion or revision step and serve as immediate feedback for instructional designers.

The methodology is validated through a "computer-to-computer" role-playing experiment. Two autonomous agents interact within the domain "Introduction to Programming": the Generator assembles learning paths and produces candidate tests, while the Editor periodically adds new fragments or deliberately introduces terminology variations. A few cycles demonstrate how Vn steadily rises from 0.80 to 0.97 as the corpus expands, whereas temporary drops in Vm and V_b reliably signal emerging conflicts that are later resolved via ontology enrichment. The experiment confirms that the triple-based representation supports both incremental evolution of the knowledge graph and automated synthesis of personalized assessments without manual curation of item banks.

The proposed approach extends beyond static content organization: it lays the groundwork for adaptive e-learning systems capable of self-diagnosing knowledge gaps, recommending targeted resources, and continuously updating course structures in response to curricular changes.

Keywords: e-learning, ontology, semantic integration, educational resources, completeness coefficient, knowledge consistency.

Стаття надійшла до редакції / Received 29.05.2025
Прийнята до друку / Accepted 26.06.2025

Вступ та постановка проблеми у загальному вигляді

Розвиток сучасних цифрових освітніх технологій супроводжується стрімким зростанням кількості і різноманіття навчальних ресурсів. Спектр матеріалів розширився від друкованих підручників до різноманітних електронних ресурсів: посібників, відеокурсів, симуляторів та онлайн-статей. Все це різноманіття дає студентам і викладачам безмежні можливості для навчання, але одночасно породжує проблеми управління та координації: як систематизувати сотні і тисячі розрізнених фрагментів знань так, щоб вони утворювали логічно цілі та коректні навчальні програми?

Традиційні пошукові системи на основі ключових слів не можуть виявити семантичні зв'язки між навчальними об'єктами [1]. Відсутність єдиної термінологічної бази перешкоджає створенню персоналізованих рекомендацій, які б враховували не лише ключові слова, а й семантику знань [2]. Створення єдиного тезаурусу власноруч зазвичай виявляється занадто складним завданням.

Кількість матеріалів зростає. Відповідно невизначеність, дублювання та конфлікти змінюються арифметично (а подекуди – геометрично). Це призводить до відсутності гнучкості при формуванні персоналізованих навчальних програм та ускладнює впровадження автоматичних механізмів оновлення навчального контенту.

Рішенням цієї проблеми може стати підхід, коли замість простого індексування за ключовими словами ми переходимо до формалізації знань у вигляді триплетів «суб'єкт-предикат-об'єкт» [3-4]. Така модель дозволяє відобразити семантичні зв'язки між поняттями, фільтрувати дублікати та контролювати консистентність термінології навіть у разі масштабного зростання кількості ресурсів. На практиці це означає, що система сама розпізнає, коли «двовимірний масив» та «матриця» належать до єдиної категорії багатовимірних структур, та інформує про потенційні конфлікти або нові поняття [5].

Вчитель або адміністратор, володіючи онтологічною базою, може швидко збирати окремі компоненти (лекції, приклади, тестові завдання) в логічно пов'язану програму, впевнений в тому, що різні розділи та теми не суперечать один одному, а їх термінологія залишається єдиною або, принаймні, узгодженою. При цьому інформація про час додавання, а також про версії та поправки, може бути систематично фіксована, спрощуючи розвиток курсу та його подальшу адаптацію. Даний аспект також підтверджується дослідженнями щодо можливості застосування онтологій для семантичної інтерпретації запитів, демонструючи можливість забезпечення точності відповідей і уточнення запитань [6].

У цій статті ми описуємо розробку моделі за концепцією Resource Description Framework (RDF) для семантичної інтеграції різнорідних освітніх ресурсів (текстів, презентацій, відео, лабораторних посібників) у єдину систему. Ми пропонуємо набір формальних коефіцієнтів для оцінки повноти анотацій, узгодженості нових матеріалів із вже описаними концептами і демонструємо роботу методики на кількох експериментальних сценаріях. Застосування таких інструментів [7] дає змогу автоматизувати збір, анотування та перевірку контенту, створити основу для персоналізованих курсів і значно спростити адміністрування великих навчальних проектів.

Аналіз досліджень та публікацій

Проблема ефективного управління та інтеграції освітніх ресурсів за допомогою семантичних технологій знаходиться в центрі уваги багатьох сучасних наукових досліджень. Онтології та графи знань розглядаються як ключові інструменти для структурування інформації, забезпечення її узгодженості та автоматизації процесів, пов'язаних з обробкою та генерацією контенту.

Значний пласт робіт присвячений застосуванню онтологій для вирішення конкретних завдань у різних предметних областях. Наприклад, пропонується підхід до формалізації дизайну експериментів у роботі з великими даними, що включає декомпозицію та рекомбінацію систем для їх оцінки. Цей підхід до структурування та оцінки складних систем може бути адаптований для моделювання та аналізу освітніх програм, що складаються з різнорідних компонентів [8]. Подібним чином, робота розглядає семантичний аналіз продуктивності, пропонуючи інструментарій для моделювання подій та спостереження за ними важливим для відстеження навчального процесу [9]. Дослідження в галузі сервісів, залежних від контексту також демонструють, як онтології можуть використовуватися для адаптації контексту з метою створення релевантних персоналізованих навчальних середовищ [10].

Завдання вилучення знань з різноманітних джерел є центральним для багатьох досліджень. Так, у [5] представлена система KnowRex, яка використовує онтології для семантично насиченого вилучення інформації з однорідних неструктурованих даних. Цей підхід є надзвичайно актуальним для обробки навчальних матеріалів, які часто представлені у вигляді текстів, презентацій та інших подібних форматів. Можливість побудови системи для обробки запитів природною мовою та перетворення на граф знань демонструє потенціал для створення інтелектуальних інтерфейсів до освітніх баз [6]. Також необхідно пам'ятати про важливість конвертації даних між різними форматами для забезпечення інтероперабельності, що підкреслюється у [4], де для цієї ролі представлено RDF-транслятор.

Інтеграція онтологій для підтримки прийняття рішень та автоматизації процесів також є поширеною темою. Дослідження демонструють системи підтримки прийняття рішень, що базуються на онтології семантичної сенсорної мережі, різних правилах, фреймворках та великих мовних моделях (LLM) [7]. Хоча предметна область відрізняється, методологія інтеграції онтологій, правил та сучасних технологій обробки даних є релевантною для створення адаптивних освітніх систем. Подібним чином, робота [11] пропонує систему, засновану на знаннях, яка включає етапи збору знань, їх представлення за допомогою онтологій та правил дедукції, а також моделювання процесів. Такий підхід може бути застосований для формалізації та автоматизації створення навчальних курсів та програм.

Звіт ISWS 2019 детально висвітлює актуальну проблему еволюції та збереження графів знань [12]. Цей документ охоплює широкий спектр питань, актуальних для довготривалого управління освітніми онтологіями:

- автоматичне виявлення змін в графах знань на атомарному, локальному та глобальному рівнях;
- вивчення волатильності відношень та динаміки типів екземплярів;
- методи вимірювання синтаксичних, структурних та семантичних змін між різними версіями графів знань, а також питання перевірки фактів;
- дослідження еволюції графів знань з точки зору обробки природної мови, наприклад, на основі аналізу.

Наразі дослідники активно вивчають роль генеративного штучного інтелекту та LLM у моделюванні метаданих та вилученні знань. Частим у використанні стає підхід, що базується на співпраці людини та LLM, з метою подолання «концептуальної заплутаності» на різних рівнях представлення інформації (перцептивному, термінологічному, онтологічному тощо) [13]. Це відкриває нові можливості для автоматизації анотування та структурування освітніх ресурсів. Існують роботи, що досліджують використання генерації, доповненої пошуком (Retrieval-Augmented Generation, RAG), для допомоги у задачах обслуговування, що демонструє потенціал LLM для надання точних інструкцій в залежності від контексту на основі різнорідних даних [14]. Цей підхід може бути екстрапольований на створення освітніх асистентів, які генерують пояснення або завдання на основі анотованих матеріалів.

Окремим напрямком є використання онтологій для покращення якості конкретних завдань, таких як, наприклад кластеризація даних, де онтології використовуються для зменшення розмірності атрибутів та підвищення якості кластерів [15]. Хоча ці роботи безпосередньо не стосуються генерації освітнього контенту, вони демонструють ефективність онтологічного підходу до структурування та інтерпретації даних. Загалом, онтології виступають потужним інструментом для підвищення якості та ефективності освітніх систем [3].

Таким чином, аналіз останніх публікацій свідчить про активний розвиток методів та інструментів для створення, інтеграції, еволюції та використання онтологій і графів знань у різних сферах. Застосування цих підходів до галузі освіти, зокрема для семантичного анотування ресурсів та автоматичної генерації навчальних і тестових матеріалів, є перспективним напрямком, що дозволяє вирішити проблеми неоднорідності, несумісності та недостатньої структурованості освітнього контенту. Однак, як показує аналіз, залишаються відкритими питання щодо створення уніфікованих методів для роботи з освітніми даними, що постійно змінюються, та розробки інструментів, які були б одночасно потужними та простими у використанні для викладачів та методистів. Важливим аспектом є також забезпечення якості, повноти та узгодженості онтологій, особливо в контексті їх динамічної еволюції та інтеграції.

Визначення проблеми, мети, та цілей статті

Однією з ключових проблем лишається неоднорідність форматів та термінології. Матеріали існують у вигляді текстових документів, презентацій, відео та методичних матеріалів. Крім того, автори часто використовують різні формулювання, описуючи одні й ті ж поняття, що ускладнює автоматичний аналіз та порівняння інформації. Традиційні пошукові механізми, що базуються на ключових словах або простих метаданих, не забезпечують повноцінної перевірки узгодженості та повноти відомостей [1]. Наприклад, в базі може зберігатися безліч документів, пов'язаних з поняттям «масив», але для систематичного складання тестових питань потрібно однозначно виділяти типи масивів (одновимірні, двовимірні, динамічні) та способи роботи з ними. В свою чергу, онтологічний підхід орієнтується на створенні семантично насичених моделей, що підтримують глибшу перевірку узгодженості знань [3] що доведено рядом досліджень [9, 15].

Навчальні ресурси часто не мають чіткого логічного каркасу та тісних взаємозв'язків. У результаті, викладачеві або методисту часто доводиться власноруч порівнювати фрагменти та перевіряти їх сумісність. Це не лише трудомістко, а й підвищує ризик пропуску важливих тем та появи протиріч між елементами курсу. Крім того, в освітній практиці відсутня достатня автоматизація при створенні вправ та тестових завдань. Навіть в умовах обширної бази знань, формування тестів по новому предмету зазвичай відбувається вручну, що ускладнює швидку адаптацію навчального процесу та оновлення матеріалів [12].

Для подолання перерахованих складнощів необхідно використовувати підхід, що дозволяє семантично описувати кожен ресурс та враховувати його зв'язок з іншими матеріалами, а також автоматично виявляти потенційні протиріччя та прогалини в знаннях. Онтології, що поповнюються за рахунок анотування навчальних даних у форматі RDF-триплетів, відкривають шлях до більш глибокої інтеграції контенту. Схожі підходи використовуються в системах обробки природної мови, де текст розбивається на токени, після чого прив'язуються згадки про онтологічні сутності, встановлюються семантичні зв'язки між ними [6]. Це дозволяє зберегти структуру знань, зв'язувати поняття через чітко визначені відношення, та на основі цих зв'язків формувати навчальні програми та тестові завдання. Отже, дане дослідження звертається до найбільшого провалу електронного навчання – відсутності простих та зручних механізмів для семантичної інтеграції та подальшого використання різноманітних ресурсів в рамках єдиного інтелектуального середовища [13].

В зв'язку з цим формується мета дослідження: розробити та обґрунтувати методику, що дозволяє ефективно інтегрувати велику кількість різнорідних освітніх ресурсів в єдину онтологічну модель, а потім використовувати отримані структуровані дані для автоматичної генерації навчальних та тестових матеріалів. Реалізація такої методики вимагає всебічного аналізу існуючих методів анотування та оцінки перспективи їх застосування до різних типів контенту (від статичних текстових документів до інтерактивних презентацій та лабораторних посібників)[14]. Крім того, важливим етапом стає формальний опис структури навчальної області у форматі RDF-триплетів, створення онтологій, що відображають множину предметних термінів та рівнів знань, а також введення системи показників повноти, оброблюваності та узгодженості (прогнозовані коефіцієнти) [11].

Під час дослідження необхідно представити механізм, здатний генерувати вправи та тестові завдання з вже анотованого корпусу, враховуючи конкретну тему та рівень підготовки студентів. Це передбачає гнучкий вибір понять та їх властивостей з онтології, аналіз зв'язків «є частиною», «визначає», «вимагає уточнення», а також підключення алгоритмів формування правильних та неправильних відповідей.

Завершальний етап передбачає апробацію запропонованого рішення на прикладі конкретної предметної

області, такої як «Основи програмування», для перевірки того, наскільки отримані результати дозволяють виявляти дублювання або протиріччя визначень та сприяють створенню змістовних тестів та завдань.

Таким чином, ключові задачі дослідження зводяться до того, щоб детально проробити методологію анотування освітніх ресурсів, формалізувати структуру предметної області у вигляді онтології, ввести кількісні показники якості анотацій та узгодженості даних, а також представити інструменти для формування тестових завдань та вправ в умовах складного набору джерел, що до того ж може постійно оновлюватися. Очікується, що запропонований підхід спростить процес семантичної інтеграції та наблизить до створення інтелектуальних навчальних систем, здатних адаптуватися до потреб освітнього середовища, що постійно змінюється [5].

Матеріали та методологія дослідження

Побудова онтології базується на багаторівневій ієрархії понять, де терміни об’єднуються в категорії за принципом інклюзії, подібно до кластеризації числових даних [15]. Кожен ресурс – будь то навчальна стаття, розділ з посібника, презентація, практичне завдання або інструкція з лабораторного посібника – представляється у вигляді сукупності елементарних тверджень (триплетів) [9]. Основою методології є перетворення інформації у формат RDF-триплетів для представлення семантичних зв'язків між поняттями, їх властивостями та значеннями.

Завдяки цьому підходу стає можливим оцінювати повноту опису кожного ресурсу:

$$V_n = \frac{N_a}{N_p}, \tag{1}$$

де N_a – число унікальних фрагментів, відображених у анотації, а N_p визначає потенційну кількість інформативних одиниць у самому документі.

Наступний параметр що вводиться потрібен для аналізу, наскільки поняття та зв'язки анотації вписуються в уже існуючу онтологічну базу за принципом:

$$V_m = \frac{C_a}{C_o}, \tag{2}$$

де C_a – загальна кількість термінів (класів, властивостей, об'єктів) у анотації, а C_o – кількість термінів, що співпадають з сутностями в онтології.

Також можливим стає виявлення протиріч:

$$V_b = \frac{F_n}{F_o}, \tag{3}$$

де F_n – кількість фактів (тверджень), що не входять в конфлікт з іншими даними, а F_o – загальна кількість тверджень в анотованому фрагменті.

Як це зображено на рис. 1, процес перетворення початкових матеріалів в RDF-анотації охоплює декілька стадій лінгвістичного та семантичного аналізу. Це відкриває можливість не лише інтеграції матеріалів, але й подальшого використання механізмів для оцінки повноти та суперечності знань [3, 11].

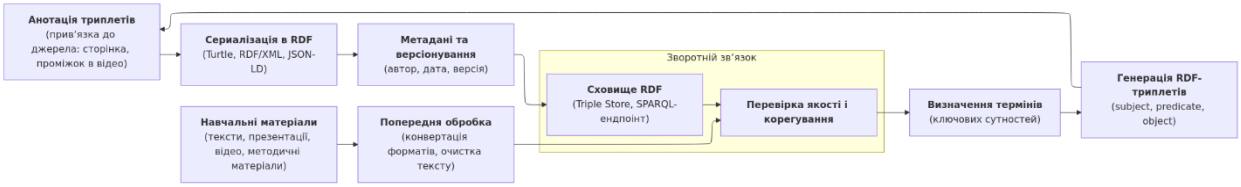

Рис. 1. Загальна схема формування RDF-анотацій з початкових навчальних матеріалів

Отже, запропонована методологія не лише визначає загальний підхід до семантичного опису матеріалів, але й встановлює кількісні критерії для оцінювання якості анотацій та можливостей їхньої інтеграції [10]. Для забезпечення формальної узгодженості знань застосовується принцип декомпозиції системи на окремі компоненти з подальшим формуванням семантичних тверджень на основі онтології, що відображають ключові властивості, взаємозв’язки та рівень обробки інформації на кожному етапі [8]. Складність і багаторівнева структура методики особливо чітко проявляються при розробці тестових завдань, вправ або практичних кейсів, побудованих на основі наявного корпусу знань [3]. У подальшому наводиться докладний приклад, який ілюструє запропонований підхід.

Щоб продемонструвати практичну реалізацію методики, обрали типову дисципліну «Основи програмування». Припустимо, у нас є наступний набір початкових ресурсів: навчальний посібник, що містить теоретичний опис базових конструкцій мови (змінні, цикли, умовні оператори), лекційні презентації, де розповідається про типові помилки при роботі з масивами, та лабораторні посібники з деталями по розв’язанню конкретних задач. Усі ці матеріали, розташовані у початковому вигляді (у PDF-файлах, текстових документах, слайдах), потребують перетворення в структуру триплетів.

На першому етапі методології відбувається лінгвістичний та семантичний аналіз тексту навчального посібника. Кожна інформаційна одиниця (наприклад, визначення «оператор циклу», опис «умовного блоку коду», правило «обчислення довжини масиву») перетворюється на елементарні твердження як це зображено на рис. 2.

До цих триплетів додаються посилання на конкретні сторінки посібника, час з лекційного відео або скріншоти з презентації, якщо потрібно детальне прив’язування до джерела. Аналогічно процедура анотування застосовується до абсолютно всього доступного матеріалу

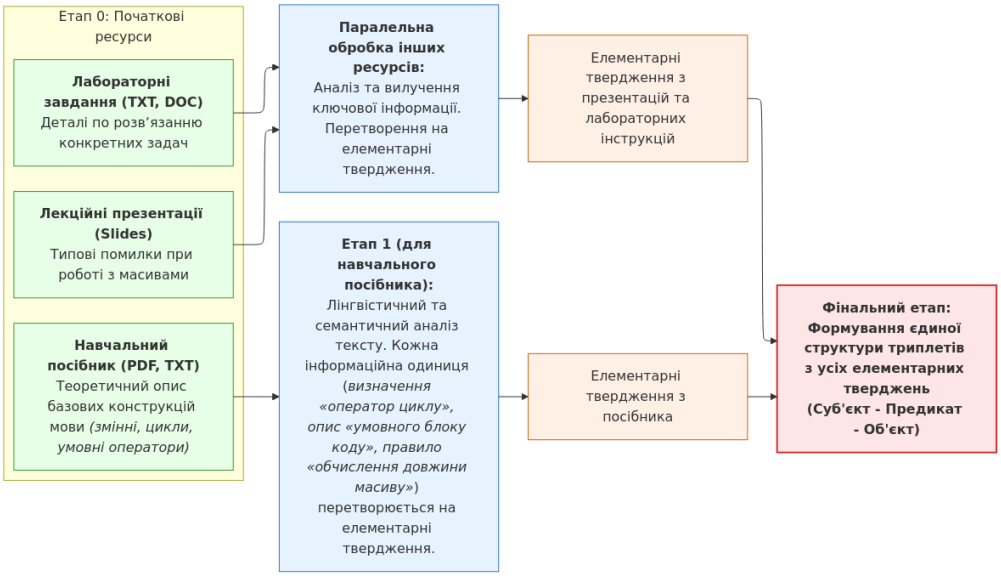

Рис. 2. Приклад трансформації навчальних матеріалів в триплети

Коли така анотація готова, система може визначати, які класи (поняття) та які властивості (відношення) вже відомі з загальної онтології з програмування. Якщо виявляється, що в анотаціях введено новий термін «межа масиву», якого немає в онтології, то коефіцієнт V_n понижується, і це вказує на необхідність або розширення онтології, або узгодження термінів.

Наступний етап методології стосується формування списку завдань. На ньому використовується ключова характеристика онтологічних моделей - автоматичне виявлення зв'язку між поняттями і отримання необхідних фрагментів знань.

Припустимо, що ми хочемо сформувати набір практичних вправ по темі «Операції з одновимірними масивами». Система вибирає триплети класу «Масив» і перевіряє, чи є в анотаціях згадка «приклад задачі» або «практична вправа» і вилучає відповідні ділянку тексту. Припустимо, в посібнику описана задача, де потрібно перебрати елементи масиву та підрахувати суму від'ємних чисел, а в презентації наводиться схожа задача, але з виділенням максимального елемента масиву. Обидві ці задачі зберігаються в анотованому вигляді, і обидві пов'язані з класом «Одновимірний масив». Система збирає їх в єдиний список і надає викладачеві або адміністратору можливість включення отриманих задач в новий навчальний модуль.

На схемі що відображена на рис. 3 показана спрощена частина онтологічного графа, яка ілюструє, як різні завдання та поняття пов'язані з концептом «Одновимірний масив».

Рис. 3. Приклад онтологічного графа, що демонструє зв'язок між поняттям та різними практичними завданнями

Найважливішою особливістю методології є те, що формально «Задача1» і «Задача2» пов'язані з більш загальним класом «Практичне завдання», а також з більш конкретним поняттям «Одновимірний масив». В онтологічному графі це може виглядати наступним чином: вершина «Практичне завдання» породжує дві дочірні вершини – «Задача1» і «Задача2», між якими і класом «Одновимірний масив» встановлені ребра «вимагає знань про», «використовує метод» тощо. Подібна структура автоматично формується в процесі анотування та аналізу анотацій [3].

Наступним аспектом є побудова тестових завдань на базі вже існуючих знань. Припустимо, що окрім практичних задач ми хочемо перевірити теоретичні знання студентів. При анотуванні навчального посібника та презентацій виділялися твердження, що описують базові поняття (цикл, умова, масив, функції тощо). З цих тверджень можна автоматизовано формувати тестові питання, виходячи з зв'язків «є частиною», «визначає», «вимагає уточнення».

Нижче на рис. 4 наведено загальний огляд модуля, що відповідає за генерацію тестових завдань з онтології. На ньому демонструється, як з пулу триплетів і класів обираються потенційні питання, проводиться фільтрація за рівнем складності, і підсумкові завдання надходять на перевірку викладачу.

Також важлива додаткова логіка, пов'язана з контролем знань різного рівня складності. Оскільки онтологія у своєму верхньому рівні може зберігати ієрархію рівнів матеріалу (елементарні поняття, середній рівень, розширені теми), система автоматично визначає, до якого рівня складності належить та чи інша концепція. Це означає, що під час генерації тестів можна обирати, які класи (поняття) потраплять до

підсумкового завдання. Якщо студент лише починає вивчати мову програмування, питання стосуватимуться базових конструкцій. Якщо ж ідеться про розширений курс, система включить питання з розділу «рекурсивні алгоритми», «динамічне програмування» та інші.

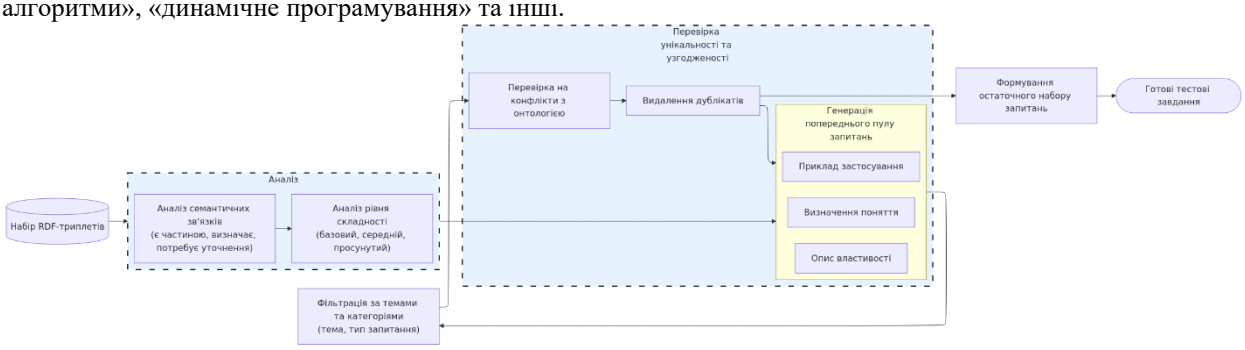

Рис. 4. Принцип автоматичної генерації тестових завдань з онтологічного графу

Процедура тестування в межах методології виглядає наступним чином. Спочатку викладач (або адміністратор) визначає набір класів (понять), які потрібно перевірити, та рівень складності завдань. Система видобуває з онтологічного графа всі пов'язані триплети, де фігурують дані поняття, відбирає відповідні фрагменти тексту та раніше збережені приклади питань. Потім формується пул потенційних тестових завдань [7].

На заключному етапі може проводитися перевірка на унікальність питань або на відсутність суперечностей. Для цього застосовується механізм оцінки узгодженості. Якщо система виявить, що частина питань або відповідей явно суперечать раніше затвердженій інформації, викладач отримає відповідне повідомлення і зможе відредагувати формулювання.

Таким чином, методологія анотування та інтеграції даних не обмежується лише перетворенням статичних ресурсів. Вона дає можливість породжувати нові навчальні та тестові матеріали, орієнтовані на різні рівні знань студентів. При цьому важливу роль відіграють: формальні описи (триплети), онтологічний граф з усіма успадкуваннями класів і зв’язками, а також система обчислення коефіцієнтів, яка інформує про повноту, оброблюваність і узгодженість об’єктів, що формуються.

У випадку, коли формування тестів відбувається на основі вже наявного, більш масштабного корпусу знань, алгоритм залишається тим самим, але розширюються можливості пошуку та фільтрації. Якщо множина предметів у корпусі велика, анотації відображають різноманіття термінів та зв’язків між ними. При генерації тестів з конкретної теми (наприклад, «багатовимірні масиви») система враховує глобальне онтологічне поле, але обирає лише ті вузли та ребра, які безпосередньо стосуються поняття «багатовимірний масив» і пов’язаних з ним концепцій (наприклад, «ітератори», «рекурсивні структури», «алгоритми обходу матриці»). У підсумку викладач бачить автоматично сформований пакет тестових питань, об’єднаних спільною тематикою, і може додатково згрупувати завдання, відредагувати формулювання або встановити вагу кожного питання в підсумковій оцінці.

Даний комплексний підхід показує, яким чином методологія дослідження дозволяє не просто впорядкувати розрізнені навчальні матеріали, але й отримати практичні інструменти для створення вправ і тестів будь-якої складності. Ключовим фактором успіху залишається якість онтологічного опису предметної області, оскільки саме онтологія стає «інтелектуальним ядром», яке переводить просте зберігання інформації у формат гнучкого конструювання освітніх програм і точного контролю знань [11].

Опис експерименту та його результати

Представлена методологія анотування та онтологічної інтеграції освітніх ресурсів спочатку була розроблена як теоретична модель, проте для часткової перевірки її працездатності та виявлення потенційних проблемних зон ми провели експеримент у вигляді рольової гри «комп’ютера з комп’ютером». Суть такого підходу полягала в тому, що два віртуальні «навчальні асистенти» (програмні модулі, що працюють на основі заздалегідь заданих сценаріїв) мали симулювати взаємодію в навчальному процесі: один «асистент» виступав у ролі системи, яка формує навчальні матеріали, а інший імітував запити викладача або зміни в корпусі даних [12].

У рамках експерименту ми створили спрощену базу знань, що включає кілька десятків «документів» (текстових блоків, що описують теми «Цикли в програмуванні», «Масиви», «Умовні оператори» тощо).

Кожен документ був попередньо анотований у форматі RDF-триплетів, щоб можна було перевірити логіку методології [4]. Для моделювання використовувалась стратегія спільної роботи людини і великої мовної моделі (LLM), де початкові RDF-анотації генерувалися автоматично з подальшою експертною перевіркою та уточненням [13]. «Асистент 1» (умовно назвемо його «ГЕНЕРАТОР») відповідав за формування навчального контенту і тестових питань на основі наявної анотованої бази, а «Асистент 2» (назвемо його «РЕДАКТОР») періодично додавав нові фрагменти або змінював термінологію, щоб змоделювати ситуації, коли викладач у реальному середовищі вносить правки в навчальний матеріал.

На початковому етапі редактор помістив у базу кілька спрощених описів тем, які не викликали конфлікту: змінні, умовні оператори, цикли while та цикли for.

«Генератор» формував із цих описів базовий навчальний план (просто перераховував теми) і кілька питань, використовуючи зв’язки по типу «визначає», «застосовується для» і «потребує уточнення». Оскільки

набір був невеликий і узгоджений, коефіцієнти V_n, V_m і V_b на цьому етапі залишалися в досить високому діапазоні. Потім редактор почав поступово вносити більш складні поняття (такі як «двовимірні масиви») і навмисно змінювати терміни в описах деяких уже існуючих документів, щоб перевірити, як система відреагує на появу потенційних конфліктів. У результаті кількох ітерацій «Асистент 1» щоразу перераховував дані, використовуючи внутрішній механізм узгодження триплетів з онтологією.

Згідно принципів визначення прихованих вузлів при побудові систем обробки природньої мови [6], або при роботі з великими даними [8], а також враховуючі підхід до валідації спостережних даних в системах високої складності [9], ми відстежували, як змінюються показники під час змін, і фіксували, наскільки швидко система виявляє термінологічні колізії [11] (наприклад, коли «двовимірні масиви» в одному документі називаються «матрицями»).

Найбільш показовим моментом було те, як генератор обробляв нові визначення і коригував список навчальних питань. Спочатку (при базовому корпусі) значення коефіцієнтів були близькі до наступного:

- $V_n \approx 0.80$, оскільки описи були відносно повними, але частина інформації залишалася узагальненою.
- $V_m \approx 0.88$, оскільки в онтології вже були присутні класи для змінних, циклів і умовних операторів.
- $V_b \approx 0.95$, що свідчило про відсутність реальних протиріч.

Потім редактор додав поняття «матриця» і «динамічний масив», при цьому об'єднавши деякі класи так, щоб «динамічний масив» фігурував як спадкоємець загальних властивостей поняття «масив», а «матриця» частково дублювала «двовимірний масив». На цьому етапі відбулося зниження V_m до ~0.78, оскільки частина нових термінів не була узгоджена з раніше існуючими в онтології властивостями (різні документи по-різному описували «матрицю»). Паралельно V_b упав приблизно до 0.88, адже система зафіксувала конфлікт: в одному документі «матриця» була визначена як «особливий тип двовимірного масиву», а в іншому - як логічно окреме поняття. Повнота анотації V_n при цьому підвищилася до 0.85 за рахунок збагачення корпусу додатковими відомостями про багатовимірні структури.

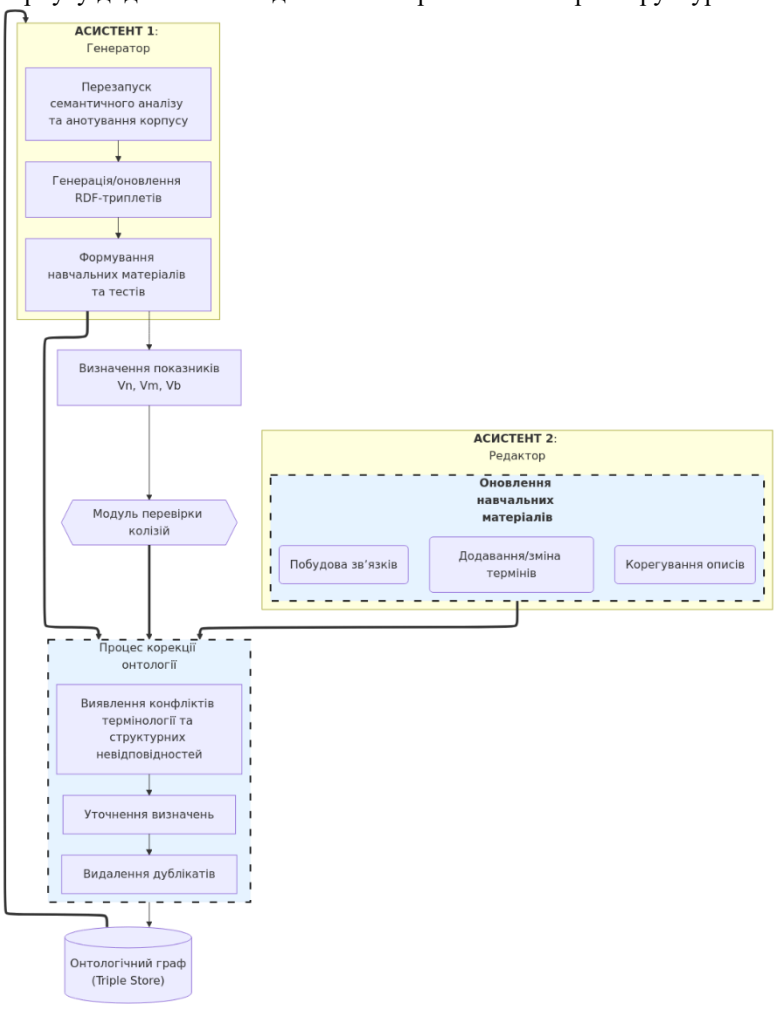

Рис. 5. Схема взаємодії двох асистентів

Протягом наступних циклів редактор впровадив часткові правки, щоб урегулювати колізію понять та розширив онтологію новими матеріалами. У результаті конфлікти були зведені до мінімуму, і в кінці рольової гри систему показників можна охарактеризувати приблизно так:

- загальне зростання повноти завдяки появі нових термінів і уточненню вже наявних;

- частина термінів все ж залишалася неоднозначною, але більшість описів уже збігалися з онтологією;
- удосконалення анотацій і об'єднання понять призвело до зниження протиріч.

Для більшої наочності нижче представлена схема ілюструє внутрішній цикл роботи двох асистентів і дозволяє зрозуміти, як у подібній рольовій грі оновлюються дані та перевіряється їхня узгодженість.

Як показано в таблиці 1, ключові показники якості анотацій, узгодженості та відсутності протиріч зазнають поступових змін відповідно до внесених коригувань у корпус знань. На початковому етапі з базовим набором тем значення всіх трьох коефіцієнтів залишалися на стабільно високому рівні. Структуризація лабораторних посібників та відеоматеріалів вносила додатковий семантичний шум, що призводило до незначних коливань коефіцієнтів. Проте поступове масштабування онтології, розширення атрибутів та введення версій сприяли відновленню та подальшому підвищенню узгодженості даних.

Таблиця 1

Значення показників впродовж експерименту

Етап	V_n	V_m	V_b	Причина змін
Початковий корпус	0.80 (–)	0.88 (–)	0.95 (–)	Базовий набір тем (змінні, умовні оператори, цикли).
Додавання «матриця» та «динамічний масив»	0.85 (+0.05)	0.78 (−0.10)	0.88 (−0.07)	Збагачення корпусу новими поняттями, конфліктні визначення
Цикли корекції термінів	0.90 (+0.05)	0.85 (+0.07)	0.93 (+0.05)	Узгоджено «матриця / двовимірний масив», уточнення сприяло зменшенню конфліктів
Інтеграція лабораторних посібників	0.92 (+0.02)	0.83 (−0.02)	0.90 (−0.03)	Деякі теги не збігалися з основною онтологією, трохи знизивши узгодженість
Додавання відеоматеріалів (транскриптів)	0.89 (−0.03)	0.84 (+0.01)	0.91 (+0.01)	Транскрипти відео додали нові фрагменти (V_n зменшився через шум у тексті), але допомогли краще зв'язати терміни з контекстом
Масштабування на тему «Функції»	0.93 (+0.04)	0.88 (+0.04)	0.89 (−0.02)	Додано нові сутності й атрибути для функцій; збільшилась повнота й узгодженість, але виникли поодинокі конфлікти у визначеннях
Введення додаткових властивостей (атрибути)	0.95 (+0.02)	0.85 (−0.03)	0.92 (+0.03)	Розширено онтологію новими атрибутами; V_m трохи впав через неоднорідність описів, але V_b покращився за рахунок зняття суперечностей у частині пов'язаних властивостей
Уточнення метаданих та версій	0.94 (−0.01)	0.87 (+0.02)	0.93 (+0.01)	Додано версії й часові мітки до триплетів; це трохи знизило повноту, але підвищило узгодженість при інтеграції
Оптимізація онтології з контролем дублікатів	0.96 (+0.02)	0.89 (+0.02)	0.94 (+0.01)	Значно зросли показники за рахунок оптимізації корпусу
Остаточна перевірка перед запуском	0.97 (+0.01)	0.90 (+0.01)	0.95 (+0.01)	Виконано фінальний аудит онтології і тестових завдань

Підводячи підсумки, варто зазначити що метод анотування з використанням RDF-триплетів і механізмів онтологічного моделювання заклав теоретичну основу, пояснивши, як саме конфігурувати навчальні матеріали та відстежувати колізії, що виникають. Тепер, описуючи рольову гру «комп'ютера з комп'ютером», ми продемонстрували, як ця методологія може працювати хоча б у спрощеному режимі. Незважаючи на те, що дослідження має багато в чому теоретичний характер, такий підхід дозволив отримати орієнтовні дані про те, як змінюються коефіцієнти при поетапному розширенні корпусу та зіткненні різних термінологічних систем.

Показники, отримані при такій роботі системи, не можуть вважатися доказом ефективності методології в реальних умовах, однак вони підтверджують її внутрішню узгодженість і демонструють основні механізми в дії. Особливо важливо те, що роль «Асистента 1» (генератора) не зводилася до автоматичного створення тестів на основі зовнішнього конструктора. Він лише формував списки потенційних питань, спираючись на контекст і корпуси даних. Це відповідає концепції, описаній раніше: питання можуть бути згенеровані з уже наявних триплетів, але остаточне формулювання може включати втручання викладача, який регулює зміст фінальних навчальних матеріалів.

Таким чином, результати експерименту у вигляді рольової гри «комп'ютера з комп'ютером» вказують на перспективність онтологічного підходу, а також на необхідність подальшого опрацювання і можливої реалізації прототипу, який дозволить протестувати дану методологію в умовах реального

освітнього середовища.

Висновки і перспективи подальшого розвитку у даному напрямі

Метод поки не охоплює весь спектр можливостей для взаємодії з учнем, що може стати предметом подальших робіт. Однак, проведене дослідження демонструє, що онтологічна модель семантичної інтеграції здатна не лише структурувати освітні ресурси, але й відкрити нові можливості для їхнього розвитку. Використовуючи формалізовані триплети, ми домоглися більш точного відображення властивостей та зв'язків між поняттями, що дозволило автоматично виявляти дублікати та розбіжності в термінології.

Особлива цінність методики проявляється в тому, що вона виходить за межі статичного впорядкування інформації. Процес анотування передбачає систематичне виявлення ключових сутностей та їхніх властивостей, завдяки чому забезпечується можливість виявлення неявних зв'язків.

Крім чисто теоретичної цінності, результатом є практичний механізм, який може інтегруватися в системи електронного навчання: викладачі та методисти отримують інструмент для оперативного формування курсів. Експериментальні показники демонструють стійке покращення якості бази знань при коригуванні термінів і анотуванні нових фрагментів.

Надалі слід зосередитися на практичній реалізації: створити прототип системи з часовими мітками для відстеження еволюції контенту та правилами для перевірки фактів. Також перспективним є об'єднання локальних онтологій у більші структури знань. Це забезпечить не лише уніфікацію термінології, але й сприятиме розвитку персоналізованих сервісів рекомендацій та глибинного аналітичного вивчення освітніх потреб.

У довгостроковій перспективі важливо врахувати інтеграцію штучного інтелекту для аналізу зворотного зв'язку від користувачів та автоматичного коригування онтологічної бази. Наприклад, алгоритми машинного навчання можуть виявляти нові патерни використання термінів і пропонувати нові зв'язки або уточнення існуючих триплетів, що підвищить адаптивність системи до змін у навчальному процесі.

Таким чином, запропонована методологія вносить вклад в теоретичну базу інтелектуальних освітніх систем, акцентуючи увагу на важливості формалізованого опису навчального контенту.

Розширення даного підходу, перехід до повноцінних прототипів та їх випробування в реальних умовах електронного освітнього середовища - важливі кроки, які дозволять остаточно підтвердити життєздатність і результативність онтологічного підходу в сфері віртуальної освіти.

Література

1. Melesko J. Semantic Technologies in e-Learning: Learning Analytics and Artificial Neural Networks in Personalised Learning Systems / J. Melesko, E. Kurilovas // WIMS '18: Proceedings of the 8th International Conference on Web Intelligence, Mining and Semantics. – 2018. – P. 1-7. – DOI: https://doi.org/10.1145/3227609.3227669.

2. George G. Review of ontology-based recommender systems in e-learning / Lal A.M. // Computers & Education. – 2019. – Volume 142. – DOI: https://doi.org/10.1016/j.compedu.2019.103642.

3. Confalonieri R. On the Multiple Roles of Ontologies in Explainable AI / Guizzardi G. // Neurosymbolic Artificial Intelligence. – 2023. – DOI: https://doi.org/10.48550/arXiv.2311.04778.

4. Stolz A. RDF Translator: A RESTful Multi-Format Data Converter for the Semantic Web / A. Stolz, B. Rodriguez-Castro, M. Hepp // 2013. – DOI: https://doi.org/10.48550/arXiv.1312.4704.

5. Adrian W.T. Semantic views of homogeneous unstructured data / W.T. Adrian, N. Leone, M. Manna // Web Reasoning and Rule Systems: 9th International Conference. – 2015. – P. 19-29. – DOI: https://doi.org/10.1007/978-3-319-22002-4_3.

6. Gorshkov S. Ontology-Based Question Answering over Corporate Structured Data / S. Gorshkov, C. Kondratiev and R. Shebalov. // 2021 International Symposium on Knowledge, Ontology, and Theory. – 2021. – P. 70-75. – DOI: https://doi.org/10.1109/KNOTH54462.2021.9685024.

7. Chandra R. Decision support system for Forest fire management using Ontology with Big Data and LLMs / R. Chandra, S. S. Kumar, R. Patra, S. Agarwal // 2024. – DOI: https://doi.org/10.48550/arXiv.2405.11346.

8. Zocholl M. Ontology-based Design of Experiments on Big Data Solutions / M. Zocholl, E. Camossi, A. L. Jousselme, C. Ray // 2019. – DOI: https://doi.org/10.48550/arXiv.1904.08626.

9. Klenik A. Adding semantics to measurements: Ontology-guided, systematic performance analysis / Klenik A., Pataricza A. // 2021. – DOI: https://doi.org/10.48550/arXiv.2112.11270.

10. Fissaa T. How Can Semantics and Context Awareness Enhance the Composition of Context-aware Services? / Fissaa T., Guermah H., Hafiddi H., Nassar M. // Proceedings of the 17th International Conference on Enterprise Information Systems. – 2015. – Volume 2. – P. 640-647. – DOI: https://doi.org/10.5220/0005381706400647.

11. Rajsiri V. Knowledge-based system for collaborative process specification / Rajsiri V., Lorré J. P., Bénaben F., Pingaud H. // Computers in Industry. – 2010. – Volume 61. – Issue 2. – P. 161-175. – DOI: https://doi.org/10.1016/j.compind.2009.10.012.

12. Nacira A. Knowledge Graphs Evolution and Preservation: A Technical Report from ISWS 2019 / Nacira A., Alghamdi K., Alinam M., Alloatti F., Amaral G., d'Amato C., Asprino L., Beno M., Bensmann F., Biswas R., Cai

L., Capshaw R., Carriero V., Celino I., Dadoun A., De Giorgis S., Delva H., Domingue J., Dumontier M., Xu W. // 2020. – DOI: https://doi.org/10.48550/arXiv.2012.11936.

13. Bagchi M. A Generative AI-driven Metadata Modelling Approach // 2024. – DOI: https://doi.org/10.48550/arXiv.2501.04008.

14. Nagy A. Cross-Format Retrieval-Augmented Generation in XR with LLMs for Context-Aware Maintenance Assistance / A. Nagy, Y. Spyridis and V. Argyriou // 2025 IEEE International Conference on Artificial Intelligence and eXtended and Virtual Reality (AIxVR). – 2025. – P. 355-361. – DOI: https://doi.org/10.1109/AIxVR63409.2025.00067.

15. Heiyanthuduwage S. R. Enhancing Cluster Quality of Numerical Datasets with Domain Ontology. / Heiyanthuduwage S. R., Rahman M. A., Islam M. Z. // 2022 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE). – 2022. – P. 1-6. – DOI: https://doi.org/10.1109/CSDE56538.2022.10089260.

References

1. Melesko J. Semantic Technologies in e-Learning: Learning Analytics and Artificial Neural Networks in Personalised Learning Systems / J. Melesko, E. Kurilovas // WIMS '18: Proceedings of the 8th International Conference on Web Intelligence, Mining and Semantics. – 2018. – P. 1-7. – DOI: https://doi.org/10.1145/3227609.3227669.

2. George G. Review of ontology-based recommender systems in e-learning / Lal A.M. // Computers & Education. – 2019. – Volume 142. – DOI: https://doi.org/10.1016/j.compedu.2019.103642.

3. Confalonieri R. On the Multiple Roles of Ontologies in Explainable AI / Guizzardi G. // Neurosymbolic Artificial Intelligence. – 2023. – DOI: https://doi.org/10.48550/arXiv.2311.04778.

4. Stolz A. RDF Translator: A RESTful Multi-Format Data Converter for the Semantic Web / A. Stolz, B. Rodriguez-Castro, M. Hepp // 2013. – DOI: https://doi.org/10.48550/arXiv.1312.4704.

5. Adrian W.T. Semantic views of homogeneous unstructured data / W.T. Adrian, N. Leone, M. Manna // Web Reasoning and Rule Systems: 9th International Conference. – 2015. – P. 19-29. – DOI: https://doi.org/10.1007/978-3-319-22002-4_3.

6. Gorshkov S. Ontology-Based Question Answering over Corporate Structured Data / S. Gorshkov, C. Kondratiev and R. Shebalov. // 2021 International Symposium on Knowledge, Ontology, and Theory. – 2021. – P. 70-75. – DOI: https://doi.org/10.1109/KNOTH54462.2021.9685024.

7. Chandra R. Decision support system for Forest fire management using Ontology with Big Data and LLMs / R. Chandra, S. S. Kumar, R. Patra, S. Agarwal // 2024. – DOI: https://doi.org/10.48550/arXiv.2405.11346.

8. Zocholl M. Ontology-based Design of Experiments on Big Data Solutions / M. Zocholl, E. Camossi, A. L. Jousselme, C. Ray // 2019. – DOI: https://doi.org/10.48550/arXiv.1904.08626.

9. Klenik A. Adding semantics to measurements: Ontology-guided, systematic performance analysis / Klenik A., Patericza A. // 2021. – DOI: https://doi.org/10.48550/arXiv.2112.11270.

10. Fissaa T. How Can Semantics and Context Awareness Enhance the Composition of Context-aware Services? / Fissaa T., Guermah H., Hafiddi H., Nassar M. // Proceedings of the 17th International Conference on Enterprise Information Systems. – 2015. – Volume 2. – P. 640-647. – DOI: https://doi.org/10.5220/0005381706400647.

11. Rajsiri V. Knowledge-based system for collaborative process specification / Rajsiri V., Lorré J. P., Bénaben F., Pingaud H. // Computers in Industry. – 2010. – Volume 61. – Issue 2. – P. 161-175. – DOI: https://doi.org/10.1016/j.compind.2009.10.012.

12. Nacira A. Knowledge Graphs Evolution and Preservation: A Technical Report from ISWS 2019 / Nacira A., Alghamdi K., Alinam M., Alloatti F., Amaral G., d'Amato C., Asprino L., Beno M., Bensmann F., Biswas R., Cai L., Capshaw R., Carriero V., Celino I., Dadoun A., De Giorgis S., Delva H., Domingue J., Dumontier M., Xu W. // 2020. – DOI: https://doi.org/10.48550/arXiv.2012.11936.

13. Bagchi M. A Generative AI-driven Metadata Modelling Approach // 2024. – DOI: https://doi.org/10.48550/arXiv.2501.04008.

14. Nagy A. Cross-Format Retrieval-Augmented Generation in XR with LLMs for Context-Aware Maintenance Assistance / A. Nagy, Y. Spyridis and V. Argyriou // 2025 IEEE International Conference on Artificial Intelligence and eXtended and Virtual Reality (AIxVR). – 2025. – P. 355-361. – DOI: https://doi.org/10.1109/AIxVR63409.2025.00067.

15. Heiyanthuduwage S. R. Enhancing Cluster Quality of Numerical Datasets with Domain Ontology. / Heiyanthuduwage S. R., Rahman M. A., Islam M. Z. // 2022 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE). – 2022. – P. 1-6. – DOI: https://doi.org/10.1109/CSDE56538.2022.10089260.