

Informatics and Information Technologies

DOI: <https://doi.org/10.15407/kvt214.04.004>

CC BY-NC

ГЛАДУН А.Я.¹, канд. техн. наук, доцент,
провід. наук. співроб. відд. комплексних досліджень
інформаційних технологій,
<https://orcid.org/0000-0002-4133-8169>, e-mail: glanat@yahoo.com

РОГУШИНА Ю.В.², канд. фіз.-мат. наук, доцент,
старш. наук. співроб. відд. автоматизованих інформаційних систем,
<https://orcid.org/0000-0001-7958-2557>, e-mail: ladamandraka2010@gmail.com

ПРИЙМА С.М.³, д-р. пед. наук, професор,
професор каф-ри комп'ютерних наук,
<https://orcid.org/0000-0002-2654-5610>, e-mail: pryima.serhii@tsatu.edu.ua

¹ Міжнародний науково-навчальний центр інформаційних технологій
та систем НАН України та МОН України,
пр. Акад. Глушкова, 40, м. Київ, 03187, Україна

² Інститут програмних систем НАН України
пр. Акад. Глушкова, 40, м. Київ, 03187, Україна

³ Таврійський державний агротехнологічний університет ім. Дмитра Моторного,
вул. Жуковського, 66, м. Запоріжжя, 69600, Україна

РЕПОЗИТОРІЙ СКЛАДНИХ ІНФОРМАЦІЙНИХ ОБ'ЄКТІВ ЯК КОМПОНЕНТ РОЗРОБЛЕННЯ СЕМАНТИЧНИХ АНАЛІТИКО-ІНФОРМАЦІЙНИХ ВЕБОРІЄНТОВАНИХ СИСТЕМ

***Вступ.** Сучасне створення різноманітних інформаційно-аналітичних систем передбачає отримання відомостей щодо предметної області з різноманітних зовнішніх інформаційних ресурсів, структура та контент яких можуть змінюватися. Для автоматизації експорту та структурування таких відомостей, доцільно застосовувати онтологічний підхід, який добре зарекомендував себе у створенні веборієнтованих застосувань різного призначення та рівня складності, а саме — отримувати відомості із зовнішніх онтологій. Саме онтології дають можливість встановлювати терміносистему для оброблення контенту матеріалів та встановлювати семантичні зв'язки з іншими джерелами інформації (енциклопедіями, нормативними документами, таксономіями спеціальностей та компетенцій тощо).*

© ВД «Академперіодика» НАН України, 2023

Метою статті є розроблення алгоритмів і методів використання формалізованих онтологічних знань предметної області для створення прикладних семантично-орієнтованих інформаційно-аналітичних систем, експорту знань із зовнішніх онтологій, створення репозиторію складних інформаційних об'єктів з розширеним функціоналом сервісів.

Результати. Розроблення концепції репозиторію складних інформаційних об'єктів для прикладних систем штучного інтелекту, яка забезпечує пошук екземплярів різних онтологічних класів, пов'язаних певними типами семантичних відношень. Удосконалення наявних функціоналів репозиторіїв онтологій за рахунок експорту знань про структуру СІО із зовнішніх джерел знань та семантично-розмічених документів. Розроблені алгоритми та методи створення репозиторіїв складних інформаційних об'єктів дають змогу аналізувати складні сукупності різних класів інформаційних об'єктів, пов'язаних між собою відношеннями обмеженнями та правилами для семантичних аналітико-інформаційних веборієнтованих систем. Сформовано базові вимоги до репозиторію та подано метод його поповнення. Отримані результати дають змогу створити оригінальні інтелектуальні інформаційні системи у сфері оброблення великих даних, кібербезпеки, аналізу компетентності під час створення професійних груп виконання інноваційного проєкту, керування людськими ресурсами, фінансами та бізнесом для компаній, які працюють з динамічно-змінюваним контентом документів.

Висновки. Запропонований оригінальний підхід, алгоритми та метод удосконалення репозиторію складних інформаційних об'єктів, розширення його функціоналу та забезпечення його поповнення за рахунок експорту знань із зовнішніх джерел (вікіпедії, енциклопедій, словників, сховищ наукових публікацій, довідників) та семантично розмічених документів і відстеження змін у цих джерелах та документах. Створено прототип семантичної веборієнтованої системи «e-Підручник», який забезпечує підбір релевантних підручників згідно з робочими програмами навчальних дисциплін для викладачів та студентів навчальних закладів Застосування онтологій та даних у системі «e-Підручник» базується на семантичному аналізі метаданих та визначенні семантичної подібності структурних моделей даних (онтологій, даних) і формуванні ранжованої множини пов'язаних онтологій для розв'язання поставлених завдань.

Ключові слова: вікі, знання-орієнтований інформаційний ресурс, онтологія, формальна модель онтології, інтелектуальна інформаційна система, репозиторій онтологій, складний інформаційний об'єкт.

ВСТУП

Сучасне створення різноманітних інформаційно-аналітичних систем (ІАС) передбачає отримання відомостей щодо предметної області (ПрО) з різноманітних зовнішніх інформаційних ресурсів (ІР), структура та контент яких можуть змінюватися. Щоб автоматизувати експорт та структурування таких відомостей, доцільно застосовувати онтологічний підхід, який добре зарекомендував себе у створенні веборієнтованих застосувань різного призначення та рівня складності, а саме — отримувати відомості із зовнішніх онтологій. Формальні моделі семантики інформаційних об'єктів потрібні для того, щоб формалізувати та забезпечити однозначну інтерпретацію їхнього змісту. Онтології можуть стати основою для однозначного та інтероперабельного визначення структури та взаємозв'язків тих інформаційних об'єктів, які дозволяють формалізувати властивості тих інформаційних об'єктів, що є предметом аналізу в ІАС і суттєвими для їх пошуку та порівняння. Саме онтології дають змогу встановлювати терміносистему для оброблення контенту матеріалів та встановлювати семантичні зв'язки з іншими джерелами інформації (енциклопедіями, нормативними документами, таксономіями спеціальностей та компетенцій тощо) [1].

СКЛАДНІ ІНФОРМАЦІЙНІ ОБ'ЄКТИ

Інформаційний об'єкт у найбільш загальному розумінні — це узагальнення та формалізація відомостей, за допомогою яких можна описати різні види матеріальних та нематеріальних об'єктів, які характеризуються різноманітними матеріальними або нематеріальними властивостями. Вибір характеристик та способів подання інформаційних об'єктів залежить від мети створення такого опису та можливостей їх оброблення. З точки зору онтологічного аналізу [2], *інформаційні об'єкти* (ІО) — це будь-які класи або екземпляри онтології. При цьому класи характеризуються своєю структурою — набором властивостей та їхніх характеристик, а також можливими відношеннями з іншими класами. У екземплярів класів онтології можуть також бути визначені значення (всіх або деяких) властивостей. Але у багатьох практичних завданнях потрібно аналізувати складніші сукупності інформації, в яких ІО пов'язані один з одним певними відношеннями та обмеженнями. Деякі з цих проблем відповідають визначенню задачі семантичного пошуку [3], якщо результатом аналізу інформації має бути побудова набору з кількох ІО одного класу, що відповідають набору умов (можливо, впорядкованих за рівнем цієї відповідності). Але у більш узагальненому випадку сам результат аналізу є множиною різних сукупностей ІО різних типів, кожна з яких відповідає певним умовам щодо структури відношень між ними та між їхніми властивостями. Надалі такі сукупності ІО будемо називати *складними інформаційними об'єктами* (СІО).

Моделі СІО можуть:

- створюватися розробниками ІАС самостійно у відповідності до специфіки проблеми та власного уявлення про неї;
- будуватися вручну на основі знань з зовнішніх інформаційних джерел (онтологій та семантично розмічених ресурсів);
- бути отримані з репозиторіїв СІО, які містять достатньо виразні метадані, щоб забезпечити автоматизоване співставлення опису проблеми з елементами контексту.

Кожен екземпляр СІО — це набір з більше ніж одного екземпляра ІО, які пов'язані один з одним онтологічними відношеннями та відповідають вимогам щодо структури і значень властивостей ІО. Приклади СІО — організація, набір її співробітників та множина проектів, які виконуються в організації; кілька ієрархічно пов'язаних підрозділів, які виконують спільну задачу з використанням набору технічних засобів; інфраструктура населеного пункту, засоби її підтримки та особовий склад, що здійснює цю підтримку; навчальний заклад, набір спеціальностей, за якими проводиться навчання, та множина компетенцій, які цими спеціальностями забезпечуються [4].

Формальна модель СІО базується на онтології ПрО

$$O_{\text{domain}} = \langle T, R, F \rangle, \quad (1)$$

де T — скінчена множина термінів ПрО, яку описує онтологія O , що поділяється на: множину класів T_{cl} та множину екземплярів класів T_{ind} ; R — скінчену множину відношень між екземплярами класів заданої ПрО; F — скінчену множину функцій інтерпретації, заданих на термінах і/або відношеннях онтології O .

У моделі СІО використовується така непорожня підмножина елементів онтології $O(1)$, яка виокремлює набір екземплярів класів з T_{cl} , між екземплярами яких в СІО визначено певні семантичні відношення з R .

Модель СІО має таку структуру:

$$\begin{aligned} C = & \langle T_C = \{t_i, i = \overline{1, p}\} \subseteq T, N_C = \\ & = \{n_i, i = \overline{1, p}\}, R_C \subseteq R, \{(t_j, t_k, r_m), t_j \in T_{ind}, t_k \in T_{ind}, r_m \in R\} \rangle. \end{aligned} \quad (2)$$

Важливо, що в моделі СІО, на відміну від онтології ПрО, виокремлюються позиції класів та елементів класів для позначення структурних елементів СІО, яким надаються унікальні імена $N_C = \{n_i, i = \overline{1, p}\}$. З кожним таким елементом може пов'язуватися набір характеристик та обмежень (наприклад, елемент обов'язково присутній, може мати єдиний екземпляр або кілька екземплярів, не може збігатися з екземпляром певного іншого елемента СІО тощо). У тих випадках, коли таке виокремлення не є важливим для ПрО, кілька формально різних СІО розглядаються як один. Наприклад, якщо в структурі СІО присутні ІО класу «Особа», які позначаються назвами «Автор підручника», для яких вказано, що елемент обов'язково має бути присутній та може мати більше одного екземпляра, але їх порядок не є значущим для пошуку, тобто між СІО, що відрізняються тільки порядком відомостей про авторів, на змістовному рівні немає різниці.

ЗОВНІШНІ ДЖЕРЕЛА ЗНАТЬ ПРО СТРУКТУРУ СІО

Зараз вже створено велику кількість різноманітних онтологій, які характеризують різні аспекти предметних областей (ПрО) з тим ступенем виразності та деталізації, що визначається цілями розроблення таких онтологій. Відомості в таких онтологіях подані у відкритих форматах, а самі онтології досить часто є у відкритому доступі. Крім того, є розвинуті інструментальні засоби, що підтримують виконання запитів до таких онтологій, щоб забезпечити видобуток відомостей щодо структури та екземплярів певних СІО. Але проблема полягає в тому, що зазвичай метадані змістовно описують онтологію в цілому, тоді як ті СІО, які входять до її складу, можна шукати тільки за іменами. Але для багатьох завдань такого пошуку недостатньо, тому що потрібно знаходити не окремі класи онтології або їхні екземпляри, а набори екземплярів певних класів, що є один з одним у потрібних відношеннях. Це спричиняє потребу у створенні репозиторіїв СІО, що підтримують набір функцій, аналогічний репозиторіям онтологій [5] або документів, але підтримують пошук на більш детальному рівні. Важливо зазначити, що такі репозиторії можуть поповнюватися відомостями як із зовнішніх онтологій, так і з семантично розмічених документів (якщо формально визначено зіставлення елементів семантичної розмітки з елементами онтології). Крім того, інформація щодо СІО може вноситися або редагуватися вручну експертами відповідної ПрО.

Специфічними для репозиторіїв СІО є такі функції:

- пошук класів, що містять визначений набір об'єктних властивостей;
- пошук екземплярів класів, для яких визначено значення певних властивостей;

- пошук екземплярів вибраних класів, що є між собою у визначених відношеннях;
- перевірка наявності екземплярів класів, що відповідають визначеному набору умов;
- пошук семантично близьких класів та порівняння їхніх екземплярів (з використанням різних мір семантичної близькості та семантичної подібності).

Важливою умовою ефективного функціонування такої системи є її відкритість та можливість поповнення її контенту різними користувачами. Доцільно орієнтуватися на вимоги FAIR [6], вибираючи технологічне середовище для такої системи: хоча ці рекомендації розроблялися з метою покращення доступу та повторного використання результатів наукових досліджень, навчальні матеріали (підручники, посібники, довідники тощо) можуть розглядатися як підмножина та окремі випадки таких інформаційних ресурсів. У цілому FAIR є подібним до відкритих даних, але зазначимо ключову відмінність. Відкриті дані доступні кожному без яких-небудь ліцензійних обмежень, угод, авторських прав чи патентів, тоді як FAIR припускає можливість доступу до даних (метаданих) у певний час і на певних умовах. Такий підхід є більш гнучким і дає змогу характеризувати дані на кожному етапі їхнього життєвого циклу.

Тому основою для побудови такого довідника може бути семантично розширена вікітехнологія, програмні застосунки на основі якої відповідають умовам FAIR [7] та підтримують семантичний пошук за властивостями тих інформаційних об'єктів, що подано у ресурсі.

СФЕРА ЗАСТОСУВАННЯ РЕПОЗИТОРІЇВ СІО

Створення репозиторію СІО має забезпечити інфраструктуру та автоматизувати аналіз семантики онтологій.

Використання репозиторію СІО дає змогу:

- зменшити час на пошук інформації про структуру та властивості інформаційних об'єктів, що обробляються в застосовних інформаційних системах;
- узгодити термінологічний базис та підвищити інтероперабельність знань, що обробляються та створюються в таких системах;
- підвищити повторне використання раніше здобутих знань та підвищити пертинентність пошуку.

Порівняно з репозиторієм онтологій, репозиторій СІО має забезпечувати:

- пошук онтологічних класів, що близькі до вибраного (за структурою, назвою, екземплярами, надкласами та підкласами тощо);
- пошук всіх класів різних онтологій, до яких належить один або кілька вибраних екземплярів (пошук за зразком);
- пошук екземплярів подібних класів з різних онтологій;
- знаходження класів та екземплярів, які можуть бути значеннями властивостей екземплярів вибраного класу;
- знаходження онтологій, які містять екземпляри вибраних класів, пов'язаних певними відношеннями (прототипів СІО).

Цим переліком функцій репозиторію СІО не обмежуються, але більшість з них можна розглядати як різні підтипи семантичного пошуку, у

якому обмеження та результати задаються у термінах класів та екземплярів класів онтологій ПрО та відношень між ними.

ПРАКТИЧНЕ ЗАСТОСУВАННЯ РЕПОЗИТОРІЮ СІО ДЛЯ ПОШУКУ НАВЧАЛЬНИХ МАТЕРІАЛІВ

Розглянемо на практичному прикладі ситуацію, коли в процесі розроблення інформаційно-аналітичної системи потреби у використанні зовнішніх баз знань не задовольнялися тими можливостями, які забезпечували репозиторії онтологій.

Веборієнтована інформаційна система «е-Підручник» призначається для пошуку та надання доступу студентам та викладачам українських університетів до навчальних матеріалів — підручників, довідників, методичних вказівок тощо.

Користувачі цієї системи поділяються на кілька груп:

- студенти, які шукають навчальну літературу, що відповідає їхнім поточним потребам (до певного навчального модуля, до курсової або дипломної роботи з певної спеціальності тощо);
- викладачі, які відбирають найбільш відповідні та актуальні підручники для своїх курсів або аналізують потребу у створенні власних навчальних матеріалів;
- інші працівники кафедр університету, які визначають відповідні підручники для кожного модуля навчальної програми (курсу), визначають потребу у підручниках у бібліотеці закладу, планують розробку оригінальних матеріалів;
- книговидавці (приватні та державні видавництва спеціалізованої літератури), які можуть подавати відомості про вже видані ними підручники або проаналізувати вже доступні для аналізу доцільності публікації нових видань.

Потреба у зовнішніх базах знань визначалася необхідністю однозначно та інтегровано визначити структуру основних об'єктів та суб'єктів ІАС, щоб уможливити обмін інформацією з іншими застосунками та її автоматизований експорт з різних джерел. Найважливіше було уніфікувати метадані про навчальні матеріали, які надано у системі.

Постачальниками контенту системи «е-Підручник» можуть бути:

- книговидавці, які можуть зареєструвати видані ними підручники та визначити умови доступу до них;
- викладачі, які можуть зареєструвати підручники зі своєї спеціальності (як підготовлені ними особисто або у співавторстві, так і знайдені у відкритому доступі);
- представники кафедр університету, які реєструють підручники, доступ до яких може надавати відповідний навчальний заклад;
- інші зацікавлені особи, які є авторами або володільцями відповідних інформаційних ресурсів.

Передбачається також можливість автоматизованого імпорту контенту з зовнішніх структурованих джерел (електронних бібліотек, репозиторіїв), що підтримують відповідні стандарти подання знань.

Основні СІО, що мають бути подані в системі «е-Підручник» — це «підручник», «автор», «видавець», «спеціальність», «компетенція», «навчальний курс (модуль)».

Зараз на різних освітніх порталах та в електронних бібліотеках навчальних закладів відомості про навчальні матеріали (СІО «Підручник», «Монографія» тощо) та області їх використання (СІО «Дисципліна», «Компетенція», «Спеціальність») подаються за допомогою різних наборів властивостей, а для подібних властивостей досить часто використовуються змістовно близькі, але різні назви.

Система підтримує пошук та зіставлення на семантичному рівні таких складних інформаційних об'єктів, як «Підручник», «Спеціальність», «Компетенція» тощо. Це дає змогу шукати підручники не тільки за назвою та спеціальністю, а з урахуванням значно більшої кількості параметрів (таких, як рік видання, мова подання, рівень складності та якість подання матеріалу). Система «e-Підручник» має бути організована як інформаційно-аналітичний веб портал, де всі поповнення та оновлення контенту автоматично відображаються на всіх пов'язаних сторінках. Семантична розмітка контенту має забезпечувати пошук та інтеграцію контенту та підтримувати навігацію на порталі на змістовному рівні. При цьому персональна інформація може бути доступна тільки групам осіб з відповідним рівнем доступу: наприклад, всім користувачам може бути доступна узагальнена оцінка певного підручника, але лише адміністратори порталу можуть бачити, хто з користувачів надав які оцінки.

БІБЛІОТЕКИ НАВЧАЛЬНИХ МАТЕРІАЛІВ ЯК ДЖЕРЕЛО КОНТЕНТУ

Більшість вищих навчальних закладів мають власні електронні бібліотеки, які містять різноманітні навчальні матеріали, що відповідають тим спеціальностям, які в них викладені. На жаль, всі вони використовують різні технологічні засоби, і це ускладнює їх інтеграцію та глобальний пошук.

Наприклад, інституційний репозиторій Київського університету імені Бориса Грінченка (<https://elibrary.kubg.edu.ua/>) містить електронні навчальні курси до дисциплін, персональні кабінети студента та викладача, журнал оцінок, каталог вибіркових дисциплін тощо. Наявний мобільний застосунок до системи дає змогу опрацьовувати та виконувати завдання в режимі офлайн, призначений для накопичення, систематизації та зберігання в електронному вигляді інтелектуальних продуктів наукового загалу університету, надання відкритого доступу до них засобами інтернет-технологій, поширення наукових матеріалів у світовому науково-освітньому просторі. Це ресурс відкритого доступу, розміщений на сервері університету в мережі «Інтернет» і доступний з будь-якого місця і у будь-який час. Цей репозиторій зареєстровано у реєстрі архівів відкритого доступу університетів та наукових організацій світу ROAR (Registry of Open Access Repositories) та у директорії репозиторіїв відкритого доступу OpenDOAR (The Directory of Open Access Repositories), а також індексується європейським пошуковим сервісом BASE (Bielefeld Academic Search Engine). Але пошук матеріалів в каталозі можна робити тільки за назвами і ключовими словами, тобто семантичний пошук за спеціальностями, рівнем складності та змістовною подібністю не реалізовано.

НТУ-КПІ імені Ігоря Сікорського пропонує кілька джерел наукових та освітніх ресурсів. Електронний каталог НТУ-КПІ імені Ігоря Сікорського (<https://kpi.ua/1338-1>) має пошукову систему для доступу до електронних

підручників. Одним з основних інструментів для поширення результатів досліджень у відкритому доступі є інституційний репозиторій — електронний архів наукових та освітніх матеріалів ELAKPI — <https://ela.kpi.ua/>. Підтримується пошук за авторами, за підрозділами, за напрямками, за роками видання та за ключовими словами. Але немає можливості виконувати комплексний пошук за набором параметрів. Крім того, програмна реалізація не дає можливості явно побачити структуру подання інформації про підручники та визначити зв'язки між матеріалами.

Отже, можна зазначити, що більшість бібліотек вищих навчальних закладів України є інтегрованими між собою, реалізовано на різних платформах, значно різняться за інтерфейсом і структурою та не підтримують семантичний пошук. Внаслідок цього студент або викладач, які хочуть отримати доступ до цінних матеріалів, мають знати їхню назву та виконувати пошук окремо на порталах різних навчальних закладів. Це значно ускладнює отримання інформації та її повторне використання у різних організаціях.

Інший приклад електронної бібліотеки навчальних матеріалів — бібліотека Таврійського державного агротехнологічного університету імені Дмитра Моторного (<http://www.tsatu.edu.ua/biblioteka/>), що є інтегрованою системою, яка охоплює наукові бібліотеки територіальних підрозділів. Електронну бібліотеку створено виключно для забезпечення навчально-освітньої діяльності університету та не призначено для використання з комерційною метою. Інституційний репозиторій бібліотеки (<http://elar.tsatu.edu.ua/>) містить більше 11890 документів, серед яких — навчально-методичні матеріали, монографії, матеріали конференцій та конспекти лекцій, які індексуються зовнішніми пошуковими системами. Пошук здійснюється за типом видання, тематикою, назвою, авторами та датою видання, але не підтримується пошук відразу за довільним набором параметрів. Доступ до повних текстів надається зареєстрованим у бібліотеці читачам у форматі pdf. Пошук за тематикою підтримується досить зручними можливостями, що містять вибір тематики, ключових слів та різноманітні додаткові фільтри. Бібліотека містить тільки матеріали, підготовлені співробітниками університету, та не передбачає інтеграцію з іншими бібліотеками.

Наразі є певна кількість інтегрованих застосовних систем пошуку навчальних матеріалів, які значно різняться функціоналом, моделями подання знань та обсягом й спрямованістю контенту. Наприклад, в [8] описується онлайн-сервіс Evdoxus, призначений для надання університетських підручників студентам університетів. Він надає: а) точний пошук та онлайн-інформацію про підручники, які доступні для кожного курсу/модуля; б) швидку доставку підручників студентам; в) ефективні механізми компенсації видавцям; г) паралельну роздачу безкоштовних електронних книг і застосунків; д) запобігання зловживанню державними ресурсами; та е) більше прозорості та менше бюрократії. Його користувачами є книговидавці, які реєструють свої підручники, викладачі, які шукають відповідні підручники для своїх курсів, персонал кафедр, які реєструють підручники, відібрані викладачами для кожного модуля навчальної програми (курсу), і, нарешті, студенти після реєстрації у студентській інформаційній системі свого університету та вибору модулів, які вони будуть відвідувати протягом кожного семестру.

ПОСТАНОВКА ПРОБЛЕМИ

Як показує порівняння наявних бібліотек навчальної літератури, вони використовують різні параметри для опису свого контенту, і тому виникає потреба у побудові більш універсальної структури цих СІО, які дають змогу інтегрувати для пошуку всю наявну інформацію, за можливості не втрачаючи важливі відомості.

Тому виникла необхідність проаналізувати наявні онтології, які описують такі СІО, та вибрати таку форму подання, яка повністю відповідає цілям розробки, але близька до існуючих усталених структур. Але виявилось, що пошук в репозиторіях онтологій структурних елементів, які відповідають обраним СІО, є не простою та нетривіальною задачею. У більшості випадків потрібно вручну аналізувати класи та підкласи онтологій обраної області, намагаючись відстежити відповідності до потрібних СІО.

ПОШУК ЗНАТЬ ПРО СТРУКТУРУ СІО В РЕПОЗИТОРІЯХ ОНТОЛОГІЙ ТА ГРАФАХ ЗНАТЬ

Передумовами розроблення репозиторію СІО є такі теоретичні та технологічні досягнення в сфері інтелектуального оброблення інформації:

- стандарти та інструментарій проєкту Semantic Web, які забезпечують інтероперабельні засоби подання та оброблення розподілених знань на основі онтологічного підходу (OWL, RDF, SPARQL);
- концепція зв'язаних даних Linked Data [9];
- концепція FAIR для надання доступу до даних та забезпечення їх пошуку та повторного використання;
- наявні онтологічні моделі СІО, які можуть використовуватися для формалізації навчальних матеріалів;
- графи знань;
- семантичні розширення вікітехнології, які забезпечують можливість розмітки природномовних текстів тегами довільної онтологічної структури і створення шаблонів подання екземплярів таких СІО, які надалі можуть використовуватися як джерело інформації щодо структури СІО.

Головними завданнями Semantic Web є створення мов семантичної розмітки електронних документів; словників, які описують структуру і семантику елементів семантичної розмітки, та реалізація засобів автоматичного генерування та оброблення цієї семантичної інформації. RDF (Resource Description Framework) — стандарт, який надає можливість формулювати твердження, які придатні для оброблення комп'ютером на основі моделі даних «об'єкт — атрибут — значення» для метаданих [10]. Він широко застосовується у різноманітних сферах, RDF Schema дає розробникам змогу визначити конкретний словник для даних RDF та види об'єктів, до яких можуть застосовуватися ці атрибути. Іншими словами, механізм RDF Schema надає базову систему типів для моделей RDF. На його основі стають загальнодоступними такі великі онтології і бази знань, як Dbpedia та Wikidata4. Крім того, такі пошукові системи, як Google і Bing, також підтримують RDF.

Онтологія OWL (Web Ontology Language) — це послідовність аксіом і фактів, а також посилань на інші онтології, на які можна посилатися через URI. OWL розширює можливості XML, RDF, RDF Schema та DAML+OIL [11].

Зв'язані дані (Linked Data) — це набір передових методів публікування та поєднання структурованих даних у веб. Ці найкращі практики прийнято багатьма постачальниками даних, що забезпечило створення простору глобальних даних Web of Data. Технологія Linked Data базується на двох фундаментальних вебтехнологіях: уніфікованих ідентифікаторах ресурсів URI і протоколі передачі гіпертексту (HTTP). На відміну від URL, використання URI надають більш узагальнений засіб ідентифікації будь-яких сутностей.

URI та HTTP доповнюються технологією, яка є критично важливою для Web of Data — RDF, поданою вище. Тоді як HTML надає засоби структурування та зв'язування документів у веб, RDF надає загальну модель даних на основі графів, за допомогою якої можна структурувати та зв'язувати дані, що описують різноманітні сутності. Модель RDF кодує дані у формі трійок «підмет-предикат-об'єкт». Суб'єкт і об'єкт трійки — це URI, кожен з яких ідентифікує ресурс, або URI та константа (рядковий літерал) відповідно. Предикат визначає, як суб'єкт і об'єкт пов'язані, і також поданий як URI.

Граф знань (ГЗ) — потужний спосіб подання структурованих знань та інтегрування інформації з різних джерел [12]. Вони складаються з вузлів, які подають сутності, і ребер, які відповідають зв'язкам між ними. ГЗ можна використовувати для підтримки різноманітних завдань, таких як аналізу даних, пошуку інформації, відповідей на запитання, рекомендацій, розуміння природної мови та комп'ютерного зору. Одним з ключових аспектів ГЗ є використання онтологій для забезпечення формального подання сутностей та їх зв'язків. Онтології забезпечують логічне виведення над ГЗ, а також перевірку їх узгодженості. Крім того, онтології також полегшують взаємодію та інтеграцію ГЗ з різних джерел, надаючи загальний словник і схему.

Використання ГЗ підтримує концепцію зв'язаних відкритих даних (linked data), пов'язуючи екземпляри інформаційних об'єктів системи з відповідними записами DBpe, використовуючи пошукову систему Wikipedia та DBpedia SPARQL. Такі ГЗ співвідносяться з легкими онтологіями, які можуть оброблятися багатьма відкритими інструментальними засобами.

Зв'язані відкриті дані (linked open data, LOD) [13] — це концепція, що є основою для ГЗ, спрямована на те, як зробити дані у Веб доступними, інтероперабельними та придатними для повторного використання. LOD базується на принципах зв'язаних даних, які використовують URI для іменування та ідентифікації сутностей, HTTP, щоб давати можливість шукати ці сутності, і RDF, щоб надавати структуровані дані за допомогою контрольованих словників [14]. LOD дає змогу об'єднувати дані з різних джерел і доменів та виконувати семантичні запити до них. Прикладами наборів даних LOD великого обсягу є DBpedia (<https://www.dbpedia.org/>), Wikidata (<https://www.wikidata.org/>) та GeoNames (<https://www.geonames.org/>).

Щоб автоматизовано інтегрувати дані з таких різних джерел, як бібліотеки окремих навчальних закладів та каталоги видавництва, виникає потреба в засобах їх перетворення у відкриті формати подання знань, наприклад, на відкритий ГЗ у форматі RDF.

СЕМАНТИЧНІ ІНФОРМАЦІЙНІ РЕСУРСИ ДЛЯ ПОБУДОВИ СІО

Для побудови структури бази знань будь-якої інформаційно-аналітичної системи доцільно використовувати зовнішні онтології, які уможливають визначення структури СІО відповідно до загально прийнятих правил. Наприклад, у пошуку навчальних матеріалів потрібно використовувати зовнішні онтології, які містять різні аспекти знань про навчальний процес, освіту та ті спеціальності, навчання яким потребує використання цих матеріалів. Але безпосередній пошук таких відомостей у пертинентних онтологіях не є простою задачею, яку можна виконувати автоматично.

Зараз є багато різних онтологій, які характеризують дослідницьку діяльність, освіту та наукові публікації [15], але більшість таких академічних онтологій пов'язані тільки з науковими дослідженнями та публікаціями, наприклад, Microsoft Academic Knowledge Graph [16] та Open Research Knowledge Graph [17], і лише деякі з них містять аспекти, що стосуються викладання та навчальних матеріалів до певних курсів, управління освітою та освітніх технологій [18].

Наприклад, репозиторій Linked Open Vocabularies (LOV — <https://lov.linkeddata.es/dataset/lov>), який забезпечує повторне використання інформаційних ресурсів в сфері наукових досліджень та освіти [19], зараз містить 782 онтології, з яких лише 6 описують освітню діяльність (решта зорієнтовані на дослідницьку та публікаційну діяльність). Проте не всі ці онтології містять інформацію щодо класів/властивостей англійською мовою (наприклад, Education Ontology — <https://schema.edu.ee>), а деякі відображають освітні системи тільки певної країни (наприклад, EduProgression Ontology — <http://ns.inria.fr/semmed/eduprogression/>).

Онтологія VIVO (<http://vivoweb.org/ontology/core>) в основному зосереджена на академічній науково-дослідницькій та публікаційній діяльності та відношеннях між дослідниками, а не на освіті, і тому вона тільки частково охоплює структуру університетів та навчальних модулів, однак не прив'язує ці модулі до навчальних програм і не зв'язує ієрархічно академічні підрозділи один з одним. Ця онтологія не залежить від конкретної сфери знань чи творчості [20]. Для книг (підручників) вона відповідає онтології BIBO — (<https://www.dublincore.org/specifications/bibo/bibo/>).

Онтологія AISO (<https://vocab.org/aiiso/schema-20080925.html>) характеризує внутрішню організаційну структуру академічної установи і надає класи та властивості для її опису. Інформація з AISO може використовуватися для опису структури університетів, кафедр, навчальних програм і курсів, але вона не охоплює підручники та їхній зв'язок із курсами та модулями.

Специфікація базового словника викладання (TEACH — <https://lov.linkeddata.es/dataset/lov/vocabs/teach>) – це спрощений словник, що надає терміни, які дають змогу викладачам пов'язувати предмети своїх курсів. TEACH базується на практичних вимогах, установлених шляхом надання описів семінарів і курсів як пов'язаних даних. Онтологія TEACH дає змогу описувати саме курси/модулі та навчальний матеріал (наприклад, підручники, книги), але не охоплює взаємовідношення між навчальними закладами.

Онтологія ReSIST (<https://lov.linkeddata.es/dataset/lov/vocabs/crsw>) подає освітні курси та ресурси і фокусується на внутрішній структурі навчальних модулів та програмному забезпеченні, що підтримує їх використання.

Болонська освітня онтологія (<https://gist.github.com/lsarni>) моделює академічне середовище, запропоноване Болонською реформою [21], і характеризує адміністративні процедури в європейських університетах та концепції для опису навчальних програм. Тому вона може використовуватись як джерело відомостей про структуру й властивості навчальних програм і спеціальностей.

Для опису спеціальностей та пов'язаних з вибором для них навчальних модулів доцільно використовувати онтології, які формалізують результати навчання, наприклад, онтологію Європейського багатомовного класифікатора навичок, компетентностей, кваліфікацій та професій ESCO (European Skills, Competences, Qualifications and Occupations) (<https://ec.europa.eu/esco/portal/home>) [22]. ESCO описує, визначає та класифікує професії, навички та кваліфікації, пов'язані з ринком праці та освіти і навчання в ЄС. ESCO може використовуватися як словник, що описує, визначає та класифікує професійні професії, навички та кваліфікації, які стосуються ринку праці та освіти та навчання в Євросоюзі. Зараз ESCO містить описи 2942 професій та 13 485 навичок, пов'язаних із цими професіями, перекладені на 27 мов (ESCO використовує всі офіційні мови Євросоюзу та деякі додаткові). Не менш важливим є той факт, що класифікатор ESCO об'єднує ринки праці країн-членів ЄС, даючи змогу здобувачам роботи та роботодавцям більш ефективно отримувати інформацію будь-якими європейськими мовами про навички, навчання та роботу.

У ESCO кожне поняття є пов'язаним щонайменше з одним терміном на всіх мовах ESCO. У багатьох випадках мова містить більш ніж один термін для позначення того ж самого або дуже схожих понять. Таким чином, ESCO може містити кілька термінів одного поняття. В рамках моделі даних ESCO кожен термін є окремим елементом і всі вони мають зв'язок з поняттям. ESCO публікується як Linked Open Data, і розробники можуть використовувати його в різних форматах (SKOS-RDF, CSV) у програмах, які надають такі послуги, як пошук вакансії, визначення профорієнтації та самооцінки. Користувачі можуть інтегрувати класифікатор ESCO в свої програми та послуги. Крім того, ESCO надає локальний API і API вебсервісів, щоб програми та вебсервіси могли запитувати інформацію з класифікатора в реальному часі.

Окрім спеціалізованих академічних та освітніх онтологій, можуть використовуватися й онтології загального призначення, таких як Schema.org, DBpedia та Wikidata, які містять інформацію про структуру та відношення між інформаційними об'єктами системи на верхньому рівні.

Словник Schema.org (<https://schema.org/docs/schemas.html>) підтримує схеми для структурованих даних. Зараз він складається з 797 типів і 1457 властивостей, що охоплюють сутності, відношення між сутностями та діями, але може бути легко розширений. Велика кількість сайтів використовує його для розмітки веб-сторінок. Багато програм від Google, Microsoft та інших уже використовують ці словники.

Онтологія DBpedia (<http://mappings.dbpedia.org/server/ontology/classes/>) — це міждомenna онтологія, створена вручну на основі найчастіше використовуваних інформаційних блоків у вікіпедії. Зараз це одна з найбільших онтоло-

гій загального призначення: вона містить 685 класів, які утворюють ієрархію та описуються майже 3000 різними властивостями. DBpedia є головним центром хмари зв'язаних відкритих даних, і вона може бути джерелом відношень між сутностями в тих випадках, коли інформація про них не знайдена у специфічніших онтологіях.

Wikidata [23] — це відкрита база знань, що використовується як центральна платформа керування даними вікіпедії та інших проєктів вікімедіа. Сховище віکیدаних складається здебільшого з елементів і тверджень про ці елементи. Елементи використовуються для подання різноманітних сутностей, включно з поняттями та об'єктами, і їм надається унікальний ідентифікатор, мітка та опис. Інструкції використовуються для запису даних про елемент і складаються принаймні з однієї пари «властивість-значення». Вони забезпечують з'єднання елементів один з одним, у результаті чого створюється пов'язана структура даних. Елементи поділяються на класи та екземпляри класів, що поєднуються відношенням приналежності. Класи можуть бути ієрархічно пов'язаними між собою таксономічним відношенням «підклас».

Онтологія віکیدаних WikiProject (https://www.wikidata.org/wiki/Wikidata:WikiProject_Ontology) — складається з кількох класів верхнього рівня та властивостей, спрямованих на підтримку широкої семантики взаємодії між іншими відомими онтологіями, такими як DOLCE, BFO, SUMO, Lemon, RDA тощо, та інтеграцію основних гілок дерева понять ядра даних вікі. Така онтологія може використовуватися як джерело знань для узгодження різних терміносистем близьких предметних областей та підтримки пошуку.

Як показують ці приклади, недостатньо знайти онтології, які пов'язані з предметною областю проєктованої системи (у цьому випадку, з освітою або науковими дослідженнями), необхідно вручну проаналізувати контент кожної з таких онтологій, визначити ті класи, які можуть бути використані як основа для структури СІО, та виявити наявність їхніх зв'язків з іншими СІО системи.

ПРАКТИЧНА РЕАЛІЗАЦІЯ СИСТЕМИ «Е-ПІДРУЧНИК»

Наразі розроблення системи «е-Підручник» перебуває на етапі створення прототипу, а саме:

- визначено основні суб'єкти та об'єкти системи, формалізовано їхню структуру та побудовано модель взаємодії між ними;
- визначено повноваження груп суб'єктів щодо можливостей редагування та перегляду інформації;
- визначено набір базових властивостей СІО, що використовуються в системі, та знайдено онтології, які дають змогу їх уточнювати та вдосконалювати;
- вибрано програмне середовище на основі Semantic MediaWiki для реалізації цієї моделі;
- створено приклади запитів-інтеграторів контенту, що дають змогу автоматизовано генерувати відомості на сторінках.

Наступним етапом має бути введення реальних даних щодо навчальних матеріалів, уточнення та розширення моделей СІО та створення більш дружнього користувацького інтерфейсу для виконання індивідуальних семантичних запитів.

Стосовно цього дослідження, створення прототипу системи «е-Підручник» демонструє потребу у наявності репозиторію СІО, який полегшуватиме пошук таких об'єктів у складі пертинентних онтологій, надасть змогу імпортувати відомості щодо структури цих об'єктів та підтримує інформування про зміни в тих онтологіях, що використовувалися для побудови бази знань системи.

СТРУКТУРА СІО ТА ЇХ ПОДАННЯ У СЕРЕДОВИЩІ SEMANTIC MEDIAWIKI

Для формалізації структури СІО використано апарат семантичних властивостей Semantic MediaWiki, який дає змогу створювати шаблони, параметрами яких є семантичними властивостями тих вікісторінок, на яких здійснюється виклик цих шаблонів. Semantic MediaWiki, додатково до звичайних властивостей вікіресурсів, дає змогу автоматично інтегрувати інформацію з різних сторінок, генерувати відповіді на складні семантичні запити та візуалізувати їхні результати. Ця технологія дає змогу будувати онтологічні бази знань, які відображають знання щодо семантичних властивостей та категорій складних інформаційних об'єктів і виконувати логічне виведення над ними. Тому розроблення інформаційно-аналітичного порталу на основі цієї технологічної платформи відповідає вимогам до створення відкритого довідника навчальних ресурсів. Ці можливості базуються на таких елементах, як семантичні властивості, категорії, шаблони та форми, які використовуються для структурування інформації, та семантичні запити.

Semantic MediaWiki дає змогу давати посиланням між сторінками такі імена, що пов'язують їх з певними поняттями ПрО, перетворюючи їх на так звані «семантичні властивості» сторінки, а також розмічати такими поняттями й інші фрагменти тексту, прив'язуючи до сторінки дані різних типів. Семантичні властивості і їх типи є головними інструментами введення семантичних даних у Semantic MediaWiki. Така семантична розмітка тексту забезпечує значно більшу виразність подання інформації (у порівнянні з традиційними категоріями, що використовуються у вікіпедії). Інформація стає доступною не тільки для читання, але і для автоматизованого машинного оброблення.

У Semantic MediaWiki підтримується мова запитів SMW-QL, що забезпечує можливості для семантичного пошуку у вікіресурсах.

Мова запитів SMW-QL дає змогу фільтрувати сторінки за заданими критеріями та виводити тільки потрібну інформацію, а не весь контент вікісторінки. Це дає змогу динамічно інтегрувати актуальну інформацію і надавати її у зручному для користувача форматі: як діаграми, географічні карти, таблиці і схеми тощо. Якщо сторінки, з яких видобуваються потрібні дані, будуть змінюватися, то результати запитів також автоматично оновлюватимуться, забезпечуючи несуперечність і узгодженість даних. Часто такі запити вбудовують у текст шаблонів.

Шаблони — це спеціальні вікісторінки, вміст яких призначено для вбудовування в інші сторінки. Основне призначення таких шаблонів — це забезпечити однотипне подання інформації та прискорити створення нових вікісторінок.

Базовий СІО системи е-Підручник — це «Підручник» (він може описувати й інші типи навчальних матеріалів), який пов'язаний з іншими СІО різними семантичними відношеннями, що визначаються на основі його властивостей (рис. 1). Інформація про екземпляри цих СІО повністю вводиться безпосеред-

ньо у систему (на відміну від відомостей про екземпляри інших СІО, де частина даних може генеруватися автоматично за допомогою семантичних запитів).

```

{{Підручник
|Автор=
|Назва=
|Видавництво=
|Рік видання=
|Тип=
|Умови доступу=
|Формат електронної версії=
|Обсяг=
|Рівень складності=
|Видавництво=
|Використовується в=
|Мова=
|Спеціальність=
|Код спеціальності=
|Компетенція=
|Посилання=
}}
    
```

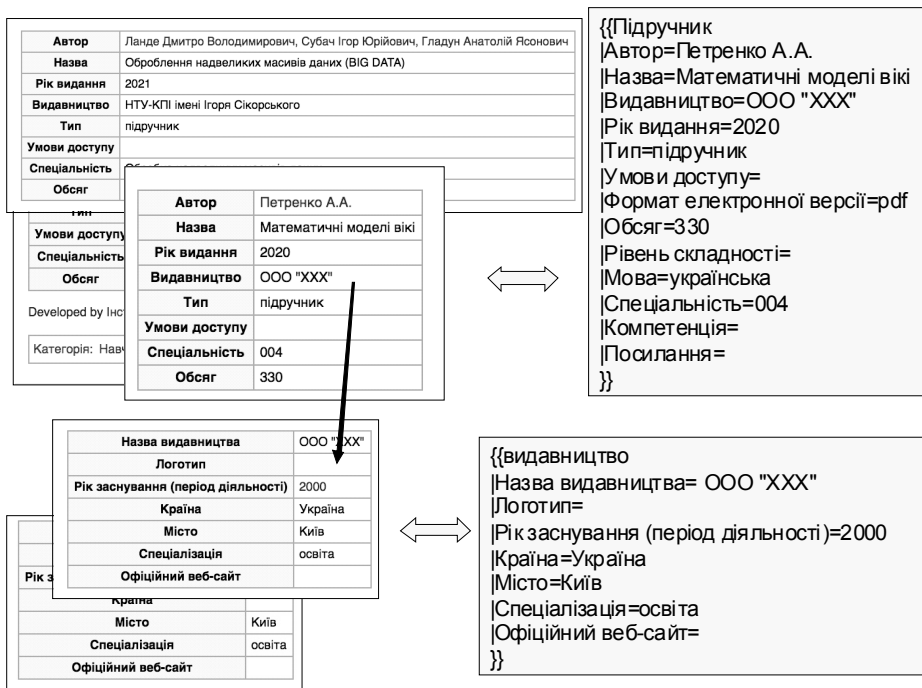


Рис. 1. Структура СІО в системі «е-Підручник» та їх відношення (фрагмент)

На перший погляд може здатися, що така структура містить надлишкову інформацію, але це забезпечує простішу автоматизовану інтеграцію відомостей з різних джерел (наприклад, в деяких джерелах є відомості про код спеціальності, для вивчення якої використовується підручник, а в інших — про назву цієї спеціальності, і перетворення одних відомостей в інші є досить простим, але на практиці доцільніше автоматично вносити всі наявні відомості, а подальші перетворення виконувати під контролем людини-експерта).

Інші СІО системи дають змогу визначити структуру об'єктних властивостей цього СІО (таких, як «Автор», «Видавництво», «Спеціальність» тощо). З точки зору вікітехнології, такі властивості мають тип «Посилання», а семантика таких посилань визначається за допомогою тегів розмітки (або параметрів шаблонів).

Наприклад, шаблон СІО «Автор підручника», екземпляри якого є значеннями об'єктної властивості «Автор» екземплярів СІО «Підручник», має наразі таку структуру:

```
{{Автор підручника
|Прізвище=
|Ім'я=
|По батькові=
|Електронна пошта (приховано)=
|Місце роботи=
|Посада=
|Науковий ступінь=
|Спеціальність ВАК=
|Вчене звання=
|Напрями діяльності=
|ORCID=
|Google Scholar=
}}
```

СІО «Освітня організація» та «Видавництво» є підкласами СІО «Організація». Але їхня структура значно відрізняється відповідно до специфічних потреб цього застосунку.

Шаблон СІО «Освітня організація» має таку структуру:

```
{{Організація
|Назва=
|Абревіатура=
|Назва мовою оригіналу=
|Абревіатура мовою оригіналу=
|Тип=
|Рівень акредитації=
|Спеціальності, що викладаються=
|Підпорядкування=
|Рік створення=
|Очільник=
|Розташування=
|Веб-сайт=
}}
```


ПОШУК НАВЧАЛЬНИХ МАТЕРІАЛІВ У СЕРЕДОВИЩІ SEMANTIC MEDIAWIKI

Засоби пошуку у семантичних вікіресурсах можна класифікувати на кілька груп, які дають змогу враховувати специфіку застосунку та використовувати структуру та характеристики СЮ:

- за ключовими словами (для системи «e-Підручник» це пошук за назвою сторінок, які відповідають екземплярам СЮ «Підручник», «Автор», «Спеціальність», «Компетенція» тощо або за початковими літерами назви);
- за тематикою предметної області (для системи «e-Підручник» це пошук за категоріями та підкатегоріями СЮ — наприклад, «Лабораторні роботи», «Монографії», цей тип пошуку дублює семантичний пошук СЮ, але забезпечує швидшу навігацію у ресурсі);
- за типом інформаційного об'єкта (для системи «e-Підручник» це пошук СЮ, який розширює можливості попереднього варіанту шляхом аналізу значень властивостей СЮ);
- за семантикою інформаційного об'єкта (для системи «e-Підручник» це пошук за його властивостями — їхніми значеннями, діапазоном тощо, який не обмежується пошуком серед екземплярів одного СЮ: наприклад, це може бути пошук певної спеціальності серед усіх навчальних матеріалів та освітніх закладів, що поданих у ресурсі).

Інші види пошуку — такі, як знаходження семантично подібних об'єктів, де зіставлення здійснюється за певним набором властивостей та з урахуванням знань Про щодо близькості певних значень цих властивостей — не підтримується засобами Semantic MediaWiki, але може бути інтегровано до системи як окремі програмні модулі API та спеціалізовані сервіси. Прикладом такого пошуку може бути знаходження навчальних матеріалів для певної компетенції, яка не вказана у цих матеріалах явно, але є підкласом або надкласом вказаних, що визначається за допомогою таксономії таких компетенцій.

Основою для всіх цих видів пошуку є відомості про структуру СЮ, і тому для створення складних пошукових запитів користувачі мають отримувати доступ до цих відомостей. Безпосередньо надавати доступ до вікі-шаблонів недоцільно: по-перше, це може знизити захищеність самої системи, а по-друге, такі шаблони містять багато технічної інформації, яка є непотрібною та незрозумілою пересічним користувачам. Альтернативним рішенням може бути надання доступу до онтологічної моделі системи (або її візуального подання). Але практика показує, що досить часто така модель не є актуальною і не відображає всі останні зміни у структурі бази знань. Тому ефективніше було б надавати користувачам посилання на відповідні елементи репозиторію СЮ, які відповідають саме поточному стану бази знань системи.

ВИСНОВКИ

В процесі створення системи «e-Підручник» була визначена потреба у репозиторії складних інформаційних об'єктів, який підтримує пошук екземплярів різних онтологічних класів, пов'язаних певними типами семантичних відношень. Такий функціонал не підтримується наявними репозиторіями онтологій, і це значно ускладнює розроблення застосовних систем, які потребують використання зовнішніх джерел знань та можливості відстеження змін у цих джерелах.

В роботі наводяться базові вимоги до такого репозиторію, аналізуються технології, які можуть застосовуватися для його поповнення, та пропонуються деякі приклади сфер його практичного використання. Розроблений метод та модуль може бути використано у системах штучного інтелекту для оброблення великих даних, кібербезпеки, аналізу компетентності під час створення колективу для виконання проєкту, керування людськими ресурсами, у сфері фінансів та бізнесу, для компаній, які працюють з динамічно-змінюваним контентом документів (юриспруденція, фінанси, стандартизація, органи державної влади тощо)

REFERENCES

1. Rogushina J.V., Gladun A.Ya. The use of ontological knowledge for multi-criteria comparison of complex information objects. *Problems of programming*. 2022, №2–3. P. 249–259. URL: pp.isoftware.com.ua/ojs1/article/view/526/523. doi: <https://doi.org/10.15407/pp2022.03-04.249> (In Ukrainian)
2. Guarino N. Formal Ontology and Information Systems. Formal Ontology in Information Systems. Proceedings of FOIS'98, by N. Guarino (ed.). Trento. Italy, Amsterdam, IOS-Press. 1998. P. 3–15 <https://klevas.mif.vu.lt/~donatas/Vadovavimas/Temos/OntologiskaiTeisingasKonceptinisModeliavi-mas/papildoma/Guarino98-Formal%20Ontology%20and%20Information%20Systems.pdf>
3. Rogushina J.V. Classification of means and methods of the Web semantic retrieval. *Problems of programming*. 2017. № 1. P. 30–50. DOI: 10.15407/pp2017.01.030 (In Ukrainian)
4. Rogushina J., Priyma S. Use of competence ontological model for matching of qualifications // Chemistry: Bulgarian Journal of Science Education, Volume 26, Number 2, 2017. P.216–228. <http://elar.tsatu.edu.ua/bitstream/123456789/3181/1/2.pdf>.
5. Gladun A.Ya., Rogushina J.V. Ontologies repository as a method to knowledge reuse for information objects recognition. *Ontology of design*, № 1 (7), 2013. P. 35–50.
6. FAIR_data. https://en.wikipedia.org/wiki/FAIR_data.
7. Rogushina, Y.V. (2023). Use of ontologies and semantic mediawiki for representation and retrieval of scientific data in the FAIR paradigm. CEUR Workshop Proceedings. Vol. 2866. P. 61–73.
8. Bassiliades N. EvdoGraph: A Knowledge Graph for the EVDOXUS Textbook Management Service for Greek Universities. Accepted for presentation at, 15th International Conference on Knowledge Engineering and Ontology Development (KEOD 2023), 13–15 Nov 2023, Rome, Italy. <https://intelligence.csd.auth.gr/wp-content/uploads/2023/08/EvdoGraph-CR.pdf>
9. Bizer, C., Heath, T., & Berners-Lee, T. (2023). Linked data-the story so far. Linking the World's Information: Essays on Tim Berners-Lee's Invention of the World Wide Web (pp. 115–143).
10. Wylot M., Hauswirth M., Cudré-Mauroux P., Sakr, S. RDF data storage and query processing schemes: A survey. *ACM Computing Surveys (CSUR)*. 2018, 51(4), 1–36.
11. Antoniou G., Van Harmelen F. Web ontology language: Owl. Handbook on ontologies. Springer Berlin Heidelberg, 2004, pp. 67–92.
12. Hogan A., Blomqvist E., Cochez M., D'amato C., Melo G., Gutierrez C., Zimmermann A. Knowledge graphs. *ACM Computing Surveys*. 2022, 54(4), pp. 1–37. doi:10.1145/3447772.
13. Yu, L., & Yu, L. (2011). Linked open data. A Developer's Guide to the Semantic Web. 2011, pp. 409–466.
14. Bizer C., Heath T., Berners-Lee T. Linked data the story so far. *International journal on semantic web and information systems*. 2009, 5(3), pp. 1–22. doi:10.4018/jswis.2009081901.

15. Stancin K., Poscic P., Jaksic D. Ontologies in education - state of the art. *Education and Information Technologies*. 2020, 25(6), pp. 5301–5320. doi:10.1007/s10639-020-10226-z.
16. Färber M. The Microsoft Academic Knowledge Graph: A Linked Data Source with 8 Billion Triples of Scholarly Data. *ISWC 2019*, LNCS. 11779, pp. 113–129. Springer. doi:10.1007/978-3-030-30796-7_8.
17. Jaradeh M. Y., Oelen A., Farfar K. E., Prinz M., D'Souza J., Kismihók G., Auer S. Open research knowledge graph: Next generation infrastructure for semantic scholarly knowledge. *KCAP 2019* (pp. 243–246). ACM. doi:10.1145/3360901.3364435.
18. Abu-Salih, B. Domain-specific knowledge graphs: A survey. *Journal of Network and Computer Applications*. 2021, 185. doi:10.1016/j.jnca.2021.103076.
19. Vandenbussche P.-Y., Ateazing G. A., Poveda-Villalón M., Vatan B. Linked Open Vocabularies (LOV): A gateway to reusable semantic vocabularies on the Web. *Semantic Web*. 2017, 8(3), pp. 437–452. doi:10.3233/SW-160213.
20. Corson-Rikert J., Mitchell S., Lowe B., Rejack N., Ding Y., Guo C. The VIVO Ontology. *VIVO, Synthesis Lectures on Data, Semantics, and Knowledge*. 2012, pp. 15–33. Springer, Cham. doi:10.1007/978-3-031-79435-3_2.
21. Demartini G., Enchev I., Gapany J., Cudré-Mauroux P. The Bowlogna ontology: Fostering open curricula and agile knowledge bases for Europe's higher education landscape. *Semantic Web*. 2013. 4(1), pp. 53–63. doi:10.3233/SW-2012-0064.
22. ESCO (the European Multilingual Classifier of Skills, Competences, Qualifications and Occupations). <https://ec.europa.eu/esco/portal/home>.
23. Vrandečić D., Krötzsch K. Wikidata: a free collaborative knowledgebase. *Communications ACM*. 2014, 10, pp. 78–85. doi:10.1145/2629489.

Received 30.08.2023

Gladun A.Ya.¹, PhD (Engineering), Associate Professor,
Leading Researcher of the Department of Complex Research
of Information Technologies,

<https://orcid.org/0000-0002-4133-8169>, e-mail: glanat@yahoo.com

Rogushina J.V.², PhD (Phys&Math), Associate Professor,

Senior Researcher of the Automated Information Systems Department,

<https://orcid.org/0000-0001-7958-2557>, e-mail: ladamandraka2010@gmail.com

Pryima S.M.³, DSc (Pedagogy), Professor,

Professor of the Computer Science Department,

<https://orcid.org/0000-0002-2654-5610>, e-mail: pryima.serhii@tsatu.edu.ua

¹ International Research and Training Center for Information
Technologies and Systems of the National Academy of Sciences
of Ukraine and the Ministry of Education and Science of Ukraine,
40, Acad. Glushkov av., Kyiv, 03187, Ukraine

² Institute of Software Systems of the National Academy of Sciences of Ukraine,
40, Acad. Glushkov av., Kyiv, 03187, Ukraine

³ Dmytro Motornyi Tavria State Agrotechnological University,
66, Zhukovskogo st., Zaporizhzhia, 72312, Ukraine

COMPLEX INFORMATION OBJECTS REPOSITORY AS A COMPONENT OF THE SEMANTIC ANALYTIC-INFORMATION WEB-ORIENTED SYSTEMS DEVELOPMENT

Introduction. *The paper examines the issue of reusing ontological knowledge in semantic analytical and informational web-oriented systems and analyzes the problems that arise in the process of searching for and exporting such knowledge from external ontologies. It proposes to create a repository of complex information objects, which should expand the functionality of services provided by ontology repositories, and provide opportunities to search for elements of such ontologies at the content level, taking into account the semantics of the relationships between them. The work states*

the basic requirements for such a repository, analyzes the technologies that can be used to replenish it, and offers some examples of areas of its practical use. The proposed approach considers on a practical example of the creation of a semantic directory for finding educational materials, which is oriented towards functioning in an open web environment and exporting information from external sources. The prototype of the system is implemented on the basis of the semantic extension of wiki technology, and the elements of the structure of complex information objects processed in the system are obtained from relevant external ontologies.

The purpose of the paper is to develop algorithms and methods of using formalized ontological knowledge of the subject area for the creation of applied semantically oriented information and analytical systems, to export knowledge from external ontologies, to create a repository of complex information objects with extended functionality of services.

The results. Development of the concept of a repository of complex information objects for applied systems of artificial intelligence, which provides a search for instances of various ontological classes connected by certain types of semantic relations. Improvement of existing functionalities of ontology repositories due to export of knowledge about the structure of CIO from external sources of knowledge and semantically marked documents. The developed algorithms and methods of creating repositories of complex information objects make it possible to analyze complex collections of different classes of information objects, interconnected by relationships, restrictions and rules for semantic analytical and informational web-oriented systems. The basic requirements for the repository are formed and the method of its replenishment is presented. The obtained results make it possible to create original intelligent information systems for artificial intelligence in the field of big data processing, cyber security, competence analysis when creating professional groups for the implementation of an innovative project, human resources management, finance and business, for companies that work with dynamically changing content of documents (jurisprudence, standardization, state authorities), national security, defense and military spheres.

Conclusions. The proposed original approach, algorithms and method for improving the repository of complex information objects, expanding its functionality and ensuring its replenishment due to the export of knowledge from external sources (Wikipedia, encyclopedias, dictionaries, repositories of scientific publications, directories) and semantically marked documents and tracking dynamic changes occurring in these sources and documents. A prototype of the semantic web-oriented system "e-Textbook" is created, which ensures the selection of relevant textbooks for teachers and students of educational institutions for work programs of educational disciplines. The application of ontologies and data in the "e-Textbook" system based on the semantic analysis of metadata and the determination of the semantic similarity of structural data models (ontologies, data) and the formation of a ranked set of related ontologies to solve the tasks.

Keywords: wiki, knowledge-oriented information resource, ontology, formal ontology model, intelligent information system, ontology repository, complex information object.