

Ю.В. Рогушина

ВИКОРИСТАННЯ СИСТЕМ ОРГАНІЗАЦІЇ ЗНАНЬ НА ОСНОВІ ОНТОЛОГІЙ У WIKI-РЕСУРСАХ

У статті розглядаються теоретичні основи систем організації знань (СОЗ) в інтелектуальних застосуваннях на основі онтологій. Ціллю даного дослідження є аналіз застосування різних типів СОЗ для організації та вдосконалення бази знань семантизованих Wiki-ресурсів, які містять гетерогенний мультимедійний контент великого обсягу та мають складну структуру, що інтегрує знання із різних Про. Розглянуто діалекти мови подання онтологій OWL та їх виразність для подання окремих випадків онтологій, що використовуються у СОЗ. Проаналізовано критерії класифікації СОЗ та сфери їх застосування. Запропоновано формальну модель онтології семантизованого Wiki-ресурсу та засоби реалізації різних видів відношень між об'єктами у середовищі Semantic MediaWiki з використанням шаблонів, розглядаються проблеми доступу до інформації у цих ресурсах з точки зору СОЗ та наводяться методи й засоби вирішення цих проблем. Представлено реалізацію запропонованого підходу на прикладі портальної версії Великої Української Енциклопедії (e-BUE).

Ключові слова: система організації знань, онтологія, тезаурус, Wiki-ресурс.

Вступ

Характерною рисою сучасних інтелектуальних інформаційних систем (ІС) є використання знань – як внутрішніх, або зі задалегідь визначених джерел, так і зовнішніх, які генеруються іншими ІС, або створюються на основі аналізу зовнішніх інформаційних ресурсів (ІР) різного рівня структурування. На ефективність роботи ІС впливають як вибір форми та організації подання знань, що обробляються, так і джерела знань і методи їх здобуття. Сьогодні більшість Web-орієнтованих ІС використовують онтології різної складності та розміру. Крім того, все частіше ІС застосовують засоби інтеграції з великими даними (Big Data) для побудови на їх основі необхідних закономірностей та правил. Тому велике значення мають ті системи організації знань (СОЗ), що використовуються як концептуальна інфраструктура для підтримки цього процесу.

СОЗ забезпечують розуміння, інтеграцію та пошук знань, підготовку знань до застосування, надають можливості для виявлення нових зв'язків і узагальнень, для прогнозування, формулювання нових гіпотез та прийняття рішень на їх основі. Важливо розуміти характеристики, які

можна використовувати для опису та аналізу СОЗ. Ці характеристики поділяються на внутрішні, що характеризують типи та властивості знань у системі (наприклад, які підкласи онтологій застосовуються, які відношення між поняттями підтримуються), і зовнішні, що не стосуються внутрішньої природи СОЗ, але описують джерела та засоби її поповнення (наприклад, автоматизовані або ручні).

Онтологічне подання знань

Грубер визначає онтології як явні специфікації концептуалізації [2]. Онтології базуються на поданні знань як скінчної множини об'єктів (класів та індивідів) [3], що називається інтерпретацією понять. Як правило, онтології можна представити у вигляді орієнтованих графів, вузли яких представляють поняття предметної області (Про), а ребра – відношення між цими поняттями. Різні структури знань, що базуються на онтологіях, відрізняються за типами відношень між поняттями, за своїми властивостями і різними логічними характеристиками цих властивостей.

Формальну модель онтології О у найбільш узагальненому вигляді часто

представляють у вигляді впорядкованої трійки: $O = \langle X, R, F \rangle$ (1), де T – скінченна множина понять PrO , яку представляє онтологія O ; R – скінченна множина відношень між поняттями цієї PrO ; F – скінченна множина аксіом і функцій інтерпретації понять і відношень онтології O . Відношення визначають собою тип взаємодії між поняттями. Аксіоми використовуються для моделювання тверджень, які завжди є істинними для PrO . Ця формальна модель може бути уточнена відповідно до цілей розвитку онтології. Наприклад, деякі дослідження уточнюють X як $X = T \cup P$, де T – скінченний набір понять предметної області, а P – скінченний набір властивостей понять. Інші поділяють X підмножини класів і екземплярів класів. Особливі випадки онтології можуть бути визначені специфікаціями та деякими обмеженнями на X , R і F та їх підкласи. Окремі випадки онтологій, таких як глосарій, таксономія, каталог тощо, можуть бути визначені обмеженнями та специфікаціями моделі (1), які визначають можливі елементи X , R і F . Детальніше формальні моделі онтологій досліджено в [4].

Для практичного використання онтологій як джерела знань щодо PrO потрібно забезпечити їх широкомасштабну інтероперабельність та формалізоване спільне розуміння. Тому для представлення онтологій доцільно використовувати мови OWL (Web Ontology Language) та RDF, які розроблені Консорціумом World Wide Web (W3C) в рамках проєкту Semantic Web. Semantic Web надає засоби для перетворення Web на глобальну базу знань, що забезпечує взаємодію між системами через обмін даними та пошук інформації на рівні знань. OWL та RDF можна обробляти за допомогою мови запитів RDF SPARQL, яка забезпечує доступ до онтологічних знань, що містяться в них.

Мова подання онтологій OWL розширює можливості XML, RDF, RDF Schema та DAML+OIL. Онтологія OWL є послідовністю аксіом, фактів і посилань на інші онтології, а також компоненти для запису авторства та іншої подібної інформації.

Онтології OWL є документами Web, на які можна посилатися через URI. Онтології OWL зазвичай містять: 1. Класи, що визначаються у owl:Class; 2. Екземпляри, що визначаються у owl:Thing; 3. Властивості, що визначаються у owl:ObjectProperty та owl:DatatypeProperty; 4. Правила – твердження, які застосовуються для логічного виведення.

OWL має три діалекти, що різняться за виразністю та складністю обробки: OWL Lite; OWL DL; OWL Full. OWL Lite – найпростіший варіант, призначений для тих користувачів, які мають потребу класифікувати ієрархію й використовують прості обмеження. OWL Lite забезпечує швидку міграцію тезаурусів та інших таксономій. OWL DL є розширенням OWL Lite, а OWL Full – розширенням OWL DL. Як наслідок, будь-яка онтологія OWL Lite є онтологією OWL DL, а будь-яка онтологія OWL DL є онтологією OWL Full.

OWL DL орієнтований на тих користувачів, які потребують максимальної виразності без втрати повноти обчислень і гарантованого завершення всіх обчислень у визначений час. OWL DL містить усі мовні конструкції OWL з обмеженнями поділу типу (клас не може бути окремою властивістю, а властивість – індивідом або класом). Назва OWL DL пов'язана з його відповідністю дескриптивній логіці.

OWL Full призначається для користувачів, яким потрібна максимальна виразність і синтаксична потужність RDF без обчислювальних гарантій. Наприклад, у OWL Full клас може одночасно розглядатися і як сукупність екземплярів, і як екземпляр. Інша суттєва відмінність від OWL DL у тому, що owl:DatatypeProperty може бути позначена як owl:InverseFunctionalProperty. OWL Full припускає такі онтології, що розширюють склад визначеного словника RDF або OWL.

У виборі діалекту OWL для створення онтології потрібно проаналізувати, які саме виразні можливості необхідні для коректного відображення знань щодо PrO , та обирати менш складний з при-

датних. Наприклад, для подання тезаурусів доцільно застосовувати OWL Lite.

Мова OWL широко використовується у Web-орієнтованих ІС, і в процесі її практичного застосування виявилися певні обмеженість її виразних можливостей і недоліки технічного характеру (складність синтаксичного розбору, неможливість знайти помилки в іменах), які спричиняють багато проблем у створенні прикладних систем організацій знань. Це викликало потребу створення нової версії мови – OWL 2.0.

Якщо в мові OWL можна визначати лише симетричні й транзитивні властивості, то OWL 2. 0 дає змогу розширити спектр логічних характеристик властивостей рефлексивністю, антирефлексивністю й антисиметричністю. Також додається можливість декларування локальної рефлексивності, яка використовується тоді, коли для властивості рефлексивність не характерна, а для деяких класів об'єктів рефлексивність наявна. Якщо OWL давав змогу лише позначити класи як несумісні, то в OWL 2. 0 з'явилася можливість робити те саме й для властивостей. Оголошення деякої множини властивостей несумісними означає, що два екземпляри не можуть бути з'єднані більш ніж однією властивістю з цієї множини (приміром, два об'єкти можуть бути пов'язані або властивістю «знаходиться над», або властивістю «знаходиться всередині», але не обома одночасно).

OWL 2. 0 містить спеціальну конструкцію для визначення ключа – множини властивостей, яка дає змогу унікально ідентифікувати екземпляри заданого класу. Також в OWL 2. 0 розширено набір типів даних для властивостей атрибутів і введено можливість задавати деякі власні обмеження на діапазони значень. Ще одна важлива риса OWL 2. 0 – визначення нових властивостей через композицію інших властивостей (property chain inclusion).

Окремі випадки онтологій та їхні формальні моделі

Онтології є досить складним для обробки засобом подання знань, і тому для розв'язку багатьох практичних за-

вдань використовують їх окремі випадки, що містять різноманітні обмеження та припускають простіші методи аналізу.

Каталог – це сукупність понять X без формальних зв'язків між ними: $R = \emptyset$.

Семантика понять може бути описати анотаціями природної мови (ПМ). На відміну від відношень і властивостей, такі анотації припускають неоднозначні інтерпретації. Каталоги можна розглядати як найпростіший окремий випадок онтології, що представлений набором понять, котрі є екземплярами одного класу з єдиною властивістю “Анотація”, і порожнім набором відношень.

Глосарій – це впорядкований (наприклад, за абеткою) набір понять ПрО з визначеннями (анотаціями) цих термінів. Це окремий випадок онтології з набором відношень, що містить єдине значення $R = \{\text{"next"}\}$. Анотація поняття є єдиною властивістю екземплярів єдиного класу [5]

Таксономія – це схема ієрархічної класифікації, де поняття організовані в групи або типи: $R = \{\text{"subclass"}\}$. Таксономії можна використовувати для організації та індексації знань щодо документів, статей, відео тощо. Таксономія – це окремий випадок онтології з єдиним ієрархічним відношенням, яке має такі характеристики, як транзитивність та антирефлексивність.

Легка (lightweight) онтологія — це онтологія, де поняття пов'язані загальними асоціаціями, а не строго визначеними формальними зв'язками. Досить часто легкі онтології розглядають як онтології, що складаються лише з набору базових таксономій. Тобто R містить кілька різних відношень типу “клас-підклас”. Термін «легка онтологія» широко використовується для позначення простих таксономій понять, організованих ієрархічно для семантичної взаємодії щодо термінології з групами користувачів. Проте деякі дослідники [6] розширюють це поняття, узагальнюють відношення «є частиною» до понять, що відповідають основним властивостям базових таксономій. Тобто у легкій онтології розширення поняття дочірнього вузла є підмножиною розширення концепції батьківського вузла.

Ширше визначення розглядає легкі онтології як онтології із обмеженими наборами відношень між концептами. Саме цим вони відрізняються від виразніших важких онтологій [7], але явно не визначається, які саме обмеження можуть за таких умов застосовуватися. В цьому випадку формальна модель легкої онтології — це підтип трійки (1), де $\langle X, R \rangle$ — кореневе дерево, а F — скінченний набір понять, виражених формальною мовою F , що належить до сімейства пропозиційної логіки опису (DL) мови без ролей. Кожну концепцію легких онтологій можна перевести у вираз DL. Приклади спрощених онтологій включають асоціативну мережу та багатомовні класифікації, але цей термін не використовується послідовно. В своїх попередніх дослідженнях [8] ми вважали, що онтологія є легкою, якщо для неї характерна наявність тільки наступних відношень:

- відношення “клас-підклас”, значення яких пов’язані транзитивно;
- об’єктні відношення синонімії, значення яких пов’язані симетрично;
- об’єктні відношення, що специфічні для ПрО, значення яких не мають бінарних властивостей.

Водночас аксіоми та правила в легкій онтології не застосовуються: $F = \emptyset$.

Таксономії, тезауруси, бізнес-каталоги, фасетні класифікації, Web-каталоги та класифікації користувачів можна розглядати як неформальні прототипи формальних легких онтологій. Як показано в [9], формальні легкі онтології можуть бути автоматично створені з неформальних класифікацій користувачів.

Легкі онтології поділяють за їх використанням на два основні типи: 1. *описові* легкі онтології: для визначення значення термінів, а також природи та структури домену; 2. *легкі онтології класифікації*: для опису, класифікації та доступу до великих колекцій документів або даних.

Легкі онтології можуть використовуватися для інтеграції даних. Цей процес об’єднує дані з різних джерел і надає користувачеві можливість уніфікованого погляду на ці дані. Часто таке джерело

даних може бути представлене у вигляді кореневого дерева (“rooted tree”), де вузли пов’язані з поняттями та їхніми природномовними мітками, а інтеграція даних може бути полегшена виявленням семантичних відношень між цими поняттями на основі даних. Знайдене семантичне відношення між двома вузлами може бути класифіковано як ієрархічне, еквівалентне або відношення неперетину. Такі відношення потім можуть бути використані для інтеграції різних СОЗ. У легких онтологіях семантичні відношення можуть бути визначені між елементами контрольованих словників, таксономій, тезаурусів, бізнес-каталогів, фасетних класифікацій тощо.

Системи класифікації та таксономії можуть бути перетворені на формальні системи, що описуються конструкціями DL замість природномовних текстів, які допускають неоднозначну та суб’єктивну інтерпретацію. Таксономії, тезауруси, Web-каталоги тощо можна використовувати як прототипи формальних легких онтологій, створення яких потребує участі експертів ПрО [10]. Будь-яка описова спрощена онтологія може бути використана як класифікаційна легка онтологія, але не навпаки.

Більш складні онтології надають ширші виразні засоби для подання знань і можуть використовуватися для ПрО з різноманітними специфічними властивостями, які не можуть бути представлені ієрархічними («клас-підклас») і мереологічними («є частиною») відношеннями [11]. Але обробка складних онтологічних структур потребує значно більше обчислювальних ресурсів і часу. Тому в практичних застосуваннях виникає потреба в їх редукції відповідно до вимог задач, але без втрати необхідних знань.

Онтології в СОЗ

Інтелектуальні програми використовують СОЗ різних типів (такі як схеми класифікації, тезауруси, тематичні карти, онтології) та різного обсягу. Термін «системи організації знань» позначає групу засобів, спрямованих на упорядкування інформації та підтримку управління зна-

ннями. Зазвичай, такі системи забезпечують ефективніший пошук та збереження знань у певному ІР.

СОЗ є інструментами для опису контенту ІР і допомоги в доступі й пошуку документів та інформації [12]. У вузькому значенні КОС підтримують такі різноманітні види діяльності, як опис документів, індексація та класифікація в бібліотеках, архівах, бібліографічних базах даних тощо. У ширшому значенні такі системи використовуються для організації науково-освітніх установ, структури дисциплін та професій, поширення знань тощо. СОЗ може використовуватися як міст між інформаційними потребами користувачів та контентом ІР. Основні елементи більшості СОЗ можуть бути виражені в RDF за допомогою Simple Knowledge Organization System (SKOS) [13]. СОЗ забезпечує ідентифікацію тих інформаційних об'єктів (ІО), що можуть зацікавити користувачів, із використанням деяких додаткових знань щодо користувача. Серед СОЗ виділяють чотири основні групи, які можуть перетинатися:

- списки термінів;
- моделі, подібні до метаданих;
- класифікація та категоризація;
- моделі відношень.

Списки термінів містять:

- Списки: впорядковані скінченні множини термінів;
- Словники: алфавітні списки понять з варіантами визначень (зазвичай представлені природномовним текстом);
- Глосарії: алфавітні списки понять з єдиним визначенням для кожного відповідно до ПрО;
- Кільця синонімів: набори понять, які вважаються еквівалентними для пошуку в певній ПрО.

Моделі, подібні до метаданих ("Metadata-like Models") містять:

- нормативні файли ("Authority Files"): списки понять, які використовуються для керування варіантами термінів для об'єктів в обраній ПрО;
- директорії ("Directories"): списки імен і деяка контактна інформація, пов'язана із цими іменами;

- географічні довідники ("Gazetteers"): індекси із геопросторовими словниками назв та географічних об'єктів.

Класифікація та категоризація містять:

- Предметні рубрики ("Subject Headings"): схеми зі скінченим набором контрольованих понять для представлення тем (рубрик) для елементів колекції та набори правил, які об'єднують ці поняття в структуровані заголовки;
- Схеми категоризації: неформальні схеми для групування;
- Таксономії: класифікація предметів на групи або категорії на основі деяких обраних властивостей;
- Схеми класифікації: ієрархічна та фасетна систематизація кількісних або алфавітних позначень для подання теми.

Моделі відношень включають більш складні СОЗ:

- Тезауруси: скінченні набори відношень понять між термінами, що чітко відображені та ідентифіковані через стандартизовані відношення (включаючи відношення ієрархії, еквівалентності та асоціації);
- Семантичні мережі: множини понять, представлені вузлами мережі і з'єднані дугами, які представляють відношення між поняттями;
- Онтології: концептуальні моделі, що представляють різноманітні складні відношення між об'єктами, включаючи правила та аксіоми, відсутні в семантичних мережах.

Критерії класифікації СОЗ залежать від цілі такого дослідження. За такою критерій часто використовують семантичну потужність, яка визначається як набір семантичних відношень між поняттями (класами та екземплярами), що підтримуються в СОЗ.

Типи онтологій, які використовуються в СОЗ, проаналізовано в [14]. Вони представлені широким спектром артефактів, що задовольняють визначенню онтології Грубером [2]. Онтології, що використовуються в СОЗ, також можна класифікувати за багатьма параметрами. СОЗ можна класифікувати також відповідно до їхньої структури та функцій:

структура може варіюватися від плоскої до двовимірної та багатовимірної, а їхні функції можуть містити усунення неоднозначностей, контроль синонімів, встановлення ієрархічних та асоціативних зв'язків та представлення властивостей. Прості плоскі структури СОЗ представлені списками вибору та кільцями синонімів. Приклади двовимірних структур СОЗ використовують ієрархії, а багатовимірні структури СОЗ використовуються семантичними мережами на основі різних семантичних типів та онтологій.

Таксономія СОЗ, запропонована в [15], базується на тому, які з основних типів відношень між поняттями в них підтримуються:

- контрольовані синоніми;
- ієрархічні відношення;
- асоціативні відношення;
- представлення властивостей.

У твердженнях природної мови досить часто виникає неоднозначність, якщо слово або словосполучення мають більше одного значення. СОЗ надають різні способи усунення неоднозначності. Якщо СОЗ не підтримує відношення між поняттями, то можуть використовуватися досить прості способи уточнення семантики цих понять. Один із них заснований на використанні доменного імені для уточнення поняття. Наприклад, «Меркурій (метал)», «Меркурій (планета)» замість «Меркурій». Такий підхід широко використовується в онлайн-енциклопедіях і Wiki-словниках для подання багатозначних термінів. Інший підхід уточнює значення поняття, надаючи контекст для нього, наприклад, за допомогою «списку вибору» – скінченного впорядкованого (за алфавітом, хронологічно, географічно тощо) набору понять з однієї групи. Такі списки описують об'єкти класів з обмеженою кількістю елементів. Списки можна ефективно використовувати для перегляду та пошуку. Вони часто використовуються як найпростіший спосіб структурування та створення метаданих.

Контрольовані синоніми (еквіваленти) використовуються для прийняття рішення в ситуаціях, коли поняття пред-

ставлено більш ніж одним терміном (має близькі синоніми), тобто його можна описати відмінною, але змістовно еквівалентною термінологією. Набір синонімів може містити терміни різних мов, акроніми та варіанти написання. Найпоширеніші проблеми стосуються близьких синонімів, значення яких зазвичай визначаються як різні, але можуть розглядатися як еквівалентні для певної ПрО. Наприклад, «машина» є синонімом «автомобіль» у ПрО транспорту, але у сфері інформатики це поняття є синонімом для «комп'ютер».

Ієрархічні відношення є найбільш поширеними відношеннями в СОЗ. Використання ієрархічних зв'язків розглядається як основний критерій для виокремлення таксономій і тезаурусів від простіших форм СОЗ, таких як списки та кільця синонімів. Ці відношення визначають рівні підпорядкування, які поділяють клас на підкласи, де кожен підклас є підмножиною вихідного класу. Класи одного рівня об'єднуються в класи вищого рівня.

Ієрархічні відношення охоплюють три різні групи відношень: 1. родові відношення ("клас-підклас"); 2. інстанційні відношення ("екземпляр-клас"); 3. мерелогічні відношення ("ціле-частина").

Родові відношення визначають зв'язки між класом і його підкласами. Цей тип відношення може пов'язувати деякий вузький термін з більш широким терміном.

Інстанційні відношення визначають зв'язок між загальною категорією предметів чи процесів і окремим екземпляром цієї категорії. Цю групу відношень можна узагальнити виразом природної мови «приклад».

Мерелогічні відношення охоплюють ситуації, коли одні поняття за своєю природою включені в інші, так що поняття можна організувати в логічні ієрархії. Усі ці відношення можна задати виразом природної мови «частина». Мерелогічні відношення [16] є транзитивними, але ці відношення поділяються на сім груп, і транзитивність діє лише всередині кожної групи і не прийнята між відношен-

нями різних груп: 1. Компонент-об'єкт; 2. Член-колекція; 3. Частина-об'єкт; 4. Матеріал-об'єкт; 5. Властивість-діяльність; 6. Стадія-процес; 7. Місце-район.

Асоціативні відношення характеризують зв'язки між поняттями, які не є ні еквівалентними, ні ієрархічними, але такі терміни семантично або концептуально пов'язані, і цей зв'язок має бути чітко визначений контрольованим словником. Деякі асоціативні відношення є предметно-специфічними і представляють відношення без їх логічних характеристик (таких як симетричність або транзитивність). Використання явних зв'язків між такими термінами надає додаткові можливості для індексації, пошуку чи обчислення семантичної подібності понять [17], які належать до різних ієрархій або термінологічних систем. Такі відношення можуть пов'язувати ІО різних типів. Приклади таких відношень – “є автором”, “має компетенцію”, “використовується для створення”.

Властивості представлення використовуються для подання знань щодо ПрО, складніших, ніж зв'язки між двома поняттями та визначеннями цих зв'язків. Онтології – це СОЗ, які використовують такі характеристики для явної специфікації концептуалізації домену. Онтології збагачують класифікаційну структуру, яку використовують таксономії та тезауруси: вони не лише відображають відношення між парами об'єктів ПрО, а й можуть містити також непорожні набори правил та аксіом, які визначають вимоги та обмеження щодо використання цих відношень та властивостей. Такий підхід забезпечує підтримку досить складного логічного висновку. Приклад правила – “Якщо об'єкти А та В знаходяться у відношенні Х1, а об'єкти А та С знаходяться у відношенні Х2, тоді об'єкти В та С знаходяться у відношенні Х3 та мають властивість Р зі значенням PPP”. Приклад обмеження – “Якщо об'єкти А та В знаходяться у відношенні Х1, а об'єкти А та С знаходяться у відношенні Х2, тоді об'єкти В та С не можуть знаходитися у відношенні Х3 та не мають властивості Р зі значенням PPP”.

Постановка задачі

Ціллю даного дослідження є аналіз застосування різних типів СОЗ для організації та вдосконалення бази знань семантизованих Wiki-ресурсів, які містять гетерогенний мультимедійний контент великого обсягу та мають складну структуру, інтегруючи знання із різних ПрО. Актуальність проблеми посилюється через потребу у якісних національних інформаційних ресурсах в умовах гібридної війни, коли визначальними факторами ефективності є як можливість отримання відповідей на складні інформаційні запити, так і достовірність та актуальність отриманої інформації. Це підвищує значення офіційних державних порталів, які мають інтегрувати дані з різних галузей знань та унеможлиблювати перекручення (як випадкові, так і зловмисні) інформації у ресурсах з відкритою генерацією контенту.

Wiki-технології та СОЗ

Зараз багато Web-орієнтованих ІР, створених в результаті колективної діяльності користувачів, базуються на технологіях Web 2.0 [18]. Контент таких ІР є більш динамічним та актуальним. Приклад найбільш успішних платформ Web 2.0, які надають механізми підтримки колаборативне створення контенту Web, – Wiki-технології [19], які забезпечують створення структурованих ІР великого обсягу. Однією з поширених реалізацій Wiki-технології є MediaWiki [20]. Існує велика кількість розробок на основі MediaWiki, найбільш відомими з яких є Вікіпедія, Wikibooks, Wiktionary, Wikidata.

Семантичні розширення Wiki-технології спрямовані на додавання змісту елементам Wiki-ресурсу, що робить їх придатними для автоматизованої обробки та аналізу на рівні знань. Воно дозволяє визначати та знаходити інформаційні об'єкти зі складною структурою, що є типовими для певної предметної області. Існує багато підходів до семантизації Wiki-технологій, більшість з яких базується на стандартах проєкту Semantic Web. Для них вже існують

формальні моделі, мови подання, методи обробки та програмні засоби. Одним із них є Semantic MediaWiki – семантичне розширення MediaWiki (www.mediawiki.org/wiki/MediaWiki), яке забезпечує інтелектуальну організацію та пошук контенту IP [20]. Прикладом складного IP на основі Semantic MediaWiki є e-BUE [21] – портална версія Великої української енциклопедії, яка використовує MediaWiki версії 1.34.0 та семантичний плагін Semantic MediaWiki версії 3.1.5.

Semantic MediaWiki забезпечує структуроване подання знань та можливість їх пошуку на змістовному рівні. Але, якщо такі IP, – приміром, енциклопедії національного рівня – мають великий обсяг та складну структуру, то вони потребують використання сучасних методів менеджменту розподілених знань та систем, що забезпечують ефективну організацію таких знань. У багатьох випадках для цього доцільно застосовувати СОЗ, різні типи яких дозволяють враховувати як специфіку Wiki-технологій, так і особливості організації багатогалузевих енциклопедій та довідників.

Wiki-онтології

Wiki-онтологія є окремим випадком онтології. Вона формалізує знання, представлені в IP, що розроблений на основі технології Wiki та її семантичних розширень [30]. Виразність онтології Wiki має деякі обмеження, оскільки така онтологія містить лише ті знання, які можна отримати безпосередньо з розмітки Wiki. Наприклад, вона не може визначати характеристики для властивостей об'єкта та властивостей даних, таких як еквівалентність і можливість перетину. В багатьох випадках семантичні розширення Wiki-технологій мають вбудовані засоби для автоматичної або автоматизованої генерації таких онтологій. Скажімо, в технологічному середовищі Semantic MediaWiki онтологія Wiki може бути створена автоматично на основі будь-якої колекції Wiki-сторінок).

З іншого боку, формування Wiki-онтології (або хоча б її структури) може передувати розробці самого Wiki-

ресурсу. В цьому випадку певна еталонна онтологія, що створюється експертами та інженерами зі знань, задає базові поняття ПрО та визначає коректні відношення між ними. В процесі розробки IP за його контентом генеруються поточні Wiki-онтології, які порівнюються з еталонною для того, щоб перевірити правильність подання знань у ресурсі. Чим складніше онтологія ресурсу, тим точніше можна відобразити ПрО, але тим складніше її аналізувати та співставляти з іншими онтологіями. Тому для ефективної розробки семантичного IP проблема вдалого вибору рівня складності СОЗ є важливим фактором. Від цього залежить і складність Wiki-онтології.

Слід розрізнити Wiki-онтології, які можуть бути згенеровані за звичайними (не семантизованими) IP та за IP із семантичною розміткою. Надалі розглядається семантизація Wiki-технології MediaWiki на основі Semantic MediaWiki.

Wiki-онтологія $O_{wiki_no_semant}$ для несемантичного ресурсу Wiki містить такі компоненти: $X = X_{cl} \cup X_{ind}$ – це набір понять онтології, де X_{cl} – набір класів, що співпадає з набором категорій Wiki, представлених в обраному наборі сторінок; X_{ind} – це набір екземплярів класів, що створюється як об'єднання імен обраних Wiki-сторінок $P = P_{user} \cup P_{template} \cup P_{spec}$, де P_{user} – набір сторінок, створених користувачами, $P_{template}$ – набір сторінок, що описують шаблони Wiki, P_{spec} – набір інших спеціальних сторінок, які явно відібрані для генерації онтології (як-от, сторінок семантичного пошуку); $R = L \cup \{r_{ier_cl}\} \cup \{r_{class_individual}\}$ – набір відношень між елементами онтології, де $L = \{link\}$ – набір з одного елемента, який описує посилання з однієї Wiki-сторінки цього ресурсу на іншу; r_{ier_cl} є ієрархічним відношенням між категоріями Wiki-ресурсу, що визначається у процесі створення нових категорій, $r_{class_individual}$ є ієрархічним відношенням між категоріями та сторінками Wiki-ресурсу, що належать до цих категорій;

$F = \{f_{equ}\}$ – це одноелементна множина, що містить відношення, яке можна використовувати для логічного виведен-

ня в онтології, – відношення еквівалентності між сторінками Wiki, яке пов’язує відсилні Wiki-сторінки. Інші елементи онтологічної моделі цієї Wiki-онтології представлені порожніми множинами.

Формальна модель семантично розмічених Wiki-ресурсів O_{wiki_semant} є складнішою за $O_{wiki_no_semant}$ і містить низку елементів, пов’язаних із семантичними властивостями:

$R = L \cup \{r_{ter_cl}\} \cup \{r_{class_individual}\} \cup L_{sem_prop}$, де до R додано набір семантичних властивостей L_{sem_prop} із областю значень у множині Wiki-сторінок;

T – це набір типів даних (наприклад, «текст», «число») для значень властивостей даних.

Шаблони Semantic MediaWiki

Можна виділити три групи шаблонів, котрі застосовуються в складних IP на основі Semantic MediaWiki (наприклад, в e-BUE), які найбільше впливають на складність Wiki-онтології та СОЗ, що використовується для цього: 1. шаблони типових ІО, які дозволяють досліджувати область значень об’єктних властивостей Wiki-онтології і типи відношень між Wiki-сторінками; 2. шаблони для встановлення змісту відношень між довільними Wiki-сторінками, які групують семантичні властивості як за їх логічними характеристиками, так і за сферою використання (це може бути загальний шаблон “Відношення” або відповідний спеціалізований шаблон – “Відношення еквівалентності”, “Мереологічні відношення” тощо); 3. шаблони для визначення однотипних відношень (ієрархічних, синонімічних тощо), характерних для найпростіших окремих випадків онтологій (рис.1).

Приклади першої групи шаблонів у e-BUE – “Персоналія” (відношення “Місце народження”), другої – шаблон “Мереологічні відношення” (відношення “Є складовою”), третьої – шаблон “Багатозначний термін” (відношення “Значення1”). У розробці шаблонів для конкретного IP слід враховувати, як реалізовані в них семантичні відношення вплинуть на рівень складності СОЗ, яка забезпечить керування знаннями.

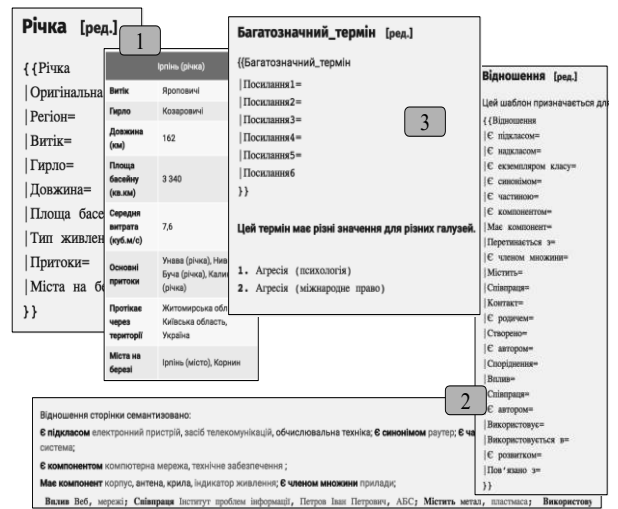


Рис.1. Приклади шаблонів e-BUE для різних груп відношень.

Висновки

Складність Wiki-онтології визначається тим набором семантичних властивостей типу “сторінка”, що використовуються у відповідному Wiki-ресурсі і, таким чином, шаблонами, що використовуються для введення їх значень. Із цього випливає значущість набору таких властивостей та потреба у його повноті та несуперечності: з одного боку, необхідно створити семантичні властивості для всіх тих відношень ПрО, які має відображати Wiki-ресурс, а з іншого – доцільно зробити цей набір найбільш компактним та зрозумілим для тих розробників ресурсу, що мають застосовувати ці властивості для структурування Wiki-контенту. Для цього пропонується розглядати типи відношень, що підтримуються у різних СОЗ, починаючи від найпростіших. Якщо відношення, що розглядається, є релевантним для контенту IP, то потрібно визначити його ім’я, описати його характеристики та створити у середовищі Semantic MediaWiki відповідну семантичну властивість типу “сторінка”. Крім того, потрібно проаналізувати, для яких категорій сторінок може застосовуватися це відношення, і, якщо ці сторінки належать до одного типового ІО (або до групи типових ІО), то додати відповідну семантичну властивість до шаблонів цих типових ІО. Якщо відношення

може застосовуватися досить часто, але його не вдасться зв'язати з певними типовими ІО, то відповідну семантичну властивість доцільно додати до одного зі спеціалізованих шаблонів, що створюються саме для встановлення змісту відношень між довільними сторінками.

References

1. Soergel, D. (2009). Knowledge organization systems: overview. www.dsoergel.com/UBLIS514DS-08.2a-1Reading4SoergelKOSOverview.pdf.
2. Gruber T.R. (1993) A translation approach to portable ontology specifications. *Knowl. Acquis.*, 5(2), 199–220.
3. Baader F., Calvanese D., McGuinness D., Nardi D., Patel-Schneider P. (2003) *The Description Logic Handbook: Theory, Implementation and Applications*. Cambridge University Press.
4. Soergel, D. (2009) Knowledge organization systems: overview. www.dsoergel.com/UBLIS514DS-08.2a-1Reading4SoergelKOSOverview.pdf.
5. Gruber T.R. (1993) A translation approach to portable ontology specifications. *Knowl. Acquis.*, 5(2), 199–220.
6. Baader F., Calvanese D., McGuinness D., Nardi D., and Patel-Schneider P. (2003) *The Description Logic Handbook: Theory, Implementation and Applications*. Cambridge University Press.
7. Rogushina J., Gladun A., Osadchiy V., Priyma S. (2015) Ontological analysis in the Web. – Melitopol State Pedagogical University Bohdan Khmelnytsky. (in Ukrainian)
8. Navigli R., Velardi P. (2008) From Glossaries to Ontologies: Extracting Semantic Structure from Textual Definitions, *Ontology Learning and Population: Bridging the Gap between Text and Knowledge*, Series information for *Frontiers in Artificial Intelligence and Applications*, IOS Press, 71-87.
9. Giunchiglia F., Zaihrayeu I. (2007) Lightweight ontologies. <http://eprints.biblio.unitn.it/1289/1/071.pdf>.
10. Nikonenko A.A. (2009) Overview of knowledge bases of the ontological type. <http://dspace.nbu.gov.ua/bitstream/handle/123456789/8144/27-Nikonenko.pdf?sequence=1>. (in Russian)
11. Rogushina J. (2018) Theoretical bases of application of ontologies for semanticization of Web resources. *Problems of programming*, (2-3), 197-203. (in Ukrainian)
12. Zaihrayeu I, Sun L., Giunchiglia F., Pan W., Ju Q., Chi M., Huang X. (2007) From web directories to ontologies: Natural language processing challenges. 6th International Semantic Web Conference (ISWC 2007). Springer.
13. Giunchiglia F., Marchese M., Zaihrayeu I. (2006) Encoding Classifications into Lightweight Ontologies. *The Semantic Web: Research and Applications, ESWC 2006*, 80-94. http://www.science.unitn.it/~marchese/pdf/P4_eswc06_Encoding.pdf.
14. Rogushina J., Gladun A. (2021) Task Thesaurus as a Tool for Modeling of User Information Needs. In *New Perspectives on Enterprise Decision-Making Applying Artificial Intelligence Techniques*, , 385-403. Springer, Cham. https://doi.org/10.1007/978-3-030-71115-3_17. https://link.springer.com/chapter/10.1007/978-3-030-71115-3_17.
15. Hjørland B. (2008) What is knowledge organization (KO)? *KO Knowledge Organization*, 35(2-3), 86-101. https://www.researchgate.net/profile/Birger-Hjørland/publication/277803483_What_is_Knowledge_Organization_KO/links/55d823260aed6a199a6afce/What-is-Knowledge-Organization-KO.pdf
16. SKOS Simple Knowledge Organization (2004) System. <https://www.w3.org/2004/02/skos/>.
17. Biagetti M. T. (2020) Ontologies (as knowledge organization systems). *ISKO Encyclopedia of Knowledge Organization*. <https://www.isko.org/cyclo/ontologies>.
18. Zeng M. L. (2008) Knowledge organization systems (KOS). *Knowledge Organization*, 35 (2-3), 160-182.
19. Gladun A., Rogushina J. (2010) Mereological aspects of ontological analysis for thesauri constructing. *Buildings and the Environment*, 301-308.
20. Rogushina J. (2019) Use of Semantic Similarity Estimates for Unstructured Data Analysis // XIX International Scientific and Practical Conference «Information Tech-

- nologies and Security» (ITS 2019). CEUR Vol-2577. 246-258. <http://ceur-ws.org/Vol-2577/paper20.pdf>.
21. Hendler J. A., Golbeck J. (2008) Metcalfe's law, Web 2.0, and the Semantic Web. *Web Sem.*, 6 (1): 14-20.
 22. Wagner C. (2004) Wiki: A technology for conversational knowledge management and group collaboration *The Communications of the Association for Information Systems*, 13(1), 264-289.
 23. Völkel M., Krötzsch M., Vrandečić D., Haller H., Studer R. (2006) Semantic wikipedia. *Proc.e of the 15th international conference on World Wide Web*, 585-594.
 24. Andon P.I., Rogushina J.V., Grishanova I.Y., Reznichenko V.A., Kyrydon A.M., Aristova A.V., Tyschenko A.O. (2021) Experience of Semantic Technologies Use for Development of Intelligent Web Encyclopedia. *UkrPROG, CEUR Workshoop Proc.*, Vol-2866, 246-259. http://ceur-ws.org/Vol-2866/ceur_246-259andon24.pdf.
 25. Rogushina J. (2019) Problems of ontological analysis use for knowledge representation of wiki-resources. *Problems in programming*, 2, 2019, 17-37. <https://doi.org/10.15407/pp2019.02.017>. (in Ukrainian)
- Отримано: 23.03.2022
- Про автора:**
- Рогущина Юлія Віталіївна,
Канд.фіз.-мат.наук, с.н.с Інституту програмних систем НАН України, публікації в українських виданнях – 200, публікації в іноземних журналах – 40, ORCID <http://orcid.org/0000-0001-7958-2557>.
- Місце роботи автора:**
- Інститут програмних систем
НАН України, 03181, Київ-187,
проспект Академіка Глушкова, 40,
e-mail: ladamandraka2010@gmail.com,
066 550 1999.