УДК 004.853, 004.55 https://doi.org/10.15407/pp2019.02.017

*Ю.В. Рогушина*

# ПРОБЛЕМИ ВИКОРИСТАННЯ ОНТОЛОГІЧНОГО АНАЛІЗУ ДЛЯ ПОДАННЯ ЗНАНЬ У WIKI-РЕСУРСАХ

Проаналізовано напрямки застосування Wiki-технологій для створення інформаційних ресурсів великого обсягу та складної структури, орієнтованих на функціювання у відкритому середовищі Web. Обґрунтовано потребу в інтелектуалізації Wiki-технології, а саме – доцільність використання Semantic MediaWiki для отримання необхідної виразної потужності у поданні семантики основних елементів Wiki-ресурсу, їх властивостей та зв'язків між ними. Розглянуто зв'язок між поданням знань у Semantic MediaWiki та стандартами Semantic Web, які використовуються для інтеграції інтелектуальних застосунків та баз знань у Web. Продемонстровано на прикладах доцільність застосування онтологічного аналізу та побудови онтологічної моделі для формального та однозначного відображення структури та контенту бази знань онлайн-версії «Великої української енциклопедії» (e-BYE). Запропоновано методи та засоби застосування цієї онтологічної моделі для створення контенту e-BYE та більш ефективного пошуку та навігації у цьому інформаційному ресурсі.

Ключові слова: онтологія, Semantic Web, Wiki-ресурс, Wiki-онтологія.

## Вступ

Збільшення обсягів інформації у Web, поширення Big Data викликають потребу у переході від збереження та накопичення даних до побудови та збереження знань, придатних для повторного використання в інформаційних системах різних виробників та різного призначення. Такі знання можуть стати основою для структурування даних, для систем машинного навчання та для семантичного пошуку. Це викликає потребу у створенні та колективному використанні розподілених баз знань, що базуються на загальноприйнятих стандартах подання інформації. На сьогодні в якості таких стандартів можна використовувати розробки Semantic Web, такі як OWL та RDF, але їх безпосередня підтримка вимагає від користувачів спеціалізованих навичок. Тому виникає необхідність у технологіях, які підтримують колективне накопичення знань та мають достатню виразну потужність для їх інтеграції зі складними інтелектуальними застосунками. Крім того, потрібно орієнтуватися на соціальні засоби подання інформації, які вже сьогодні широко розповсюджені та використовуються багатьма спільнотами користувачів.

## Технологія Wiki як основа для колективної побудови бази знань

Важливою рисою багатьох Web-сайтів є створення спільнот, де користувачі зі спільними інтересами є активними учасниками як споживання, так і створення інформації, тобто соціальних мереж. У сучасному розумінні термін «соціальна мережа» використовується для характеристики програмного забезпечення, призначеного для встановлення відношень між окремими особами, а інтелектуалізація Web-технологій визначають тенденцію переходу до семантичних соціальних мереж ($S^2N$ – The Semantic Social Network) [1]. Однією з широко відомих інформаційних технологій, що використовуються для цього, є Wiki, особливістю яких є відкритий підхід до створення контенту [2].

*Базові можливості та обмеження Wiki в обробці інформації*

*Wiki-система* – це форма Web-платформи для соціального програмного забезпечення, яка дозволяє спільно створювати, супроводжувати та знаходити цифровий контент. Найбільш вживане нині програмне забезпечення для Wiki-систем – MediaWiki, що базується на PHP і MySQL. MediaWiki використовують проекти Wikipedia, Wikidata, Wikibooks. MediaWiki відрізняється якісним та зручним редактором контенту.

У. Каннінгем, що створив Wiki у 1994 році, визначив цю технологію як *найпростішу онлайнову базу даних*, яка спроможна працювати [3]. Основними характе-

ристиками Wiki і надалі залишаються простота колаборативного використання та редагування сторінок, легке встановлення взаємозв'язків між сторінок всередині і поза Wiki-ресурсом.

*Wiki-ресурс* – це екземпляр соціального програмного забезпечення, який є сукупністю спільно створених статей, що містять структурований за допомогою Wiki-розмітки текст (призначений для читання та розуміння людьми) і гіперпосилання на інші Wiki-статті або зовнішні IP. Wiki-ресурси є основою для спільного створення та використання знань. Спільними рисами усіх Wiki-ресурсів є наявність Web-інтерфейсу, простий синтаксис для структурування контенту та можливість встановлення гіперпосилань на інші сторінки. Більшість Wiki-систем також забезпечують механізм відкату у випадку випадкових або небажаних змін.

Ефективність використання Wiki як основи для створення інформаційного ресурсу (наприклад, посібника, енциклопедії або довідника) базується на таких особливостях цієї технології, як:

- колаборативність – контент негайно стає доступним для всіх після публікації;
- простота створення документів та інформаційних взаємозв'язків між ними, що полегшує повторне використання інформації;
- тонка деталізація контенту – технологія забезпечує створення окремих сторінок для кожної теми, терміну або слова з довільним рівнем деталізації;
- засоби структурування контенту – за допомогою метаданих (шаблонів, категорій та семантичних властивостей).

Незважаючи на ефективність Wiki як інструменту співпраці, ця технологія має певні недоліки. Інформація, що накопичується у Wiki-ресурсах, є неструктурованою у тому розумінні, що для неї відсутня попередньо визначена модель даних, а її наявні структурні елементи не пристосовані для автоматизованої обробки без додаткових уточнень. Наприклад, природномовна (ПМ) складова контенту має певну лінгвістичну структурованість (поділ на частини мови, члени речення), але у загальному випадку кожна Wiki-сторінка має довільний розмір, складається з довільних частин і не має стандартизованої моделі подання, що викликає складнощі її аналізу.

Така інформація обробляється як неструктуровані дані (НСД) [4], що створює проблеми для менеджменту знань у Wiki-ресурсі. Сторінки не можуть бути автоматично відформатовані відповідно до того, яку інформацію вони містять. Сортування об'єктів, описаних у Wiki за довільними критеріями, потребує використання спеціальних шаблонів.

У традиційних Wiki-ресурсах досить складно імпортувати структуровані дані із зовнішніх баз даних і робити до них запити. Протилежна операція, тобто експорт довільного вибору контенту Wiki до зовнішнього програмного забезпечення, теж пов’язані з певними складнощами та не може бути виконаний автоматично.

Поширений підхід до вирішення цих проблем полягає в тому, щоб семантично структурувати інформацію (з використанням форм та шаблонів) – це робить контент Wiki-ресурсу набагато зрозумілішим для людей і більш доступним для комп'ютерних операцій, таких як пошук на основі структури, інтеграція даних і аналіз даних. Цей підхід називають семантизацією Wiki, і для його реалізації розроблено велику кількість програмного забезпечення, що різниться за своїми цілями, можливостями та вимогами до користувачів. Семантичні Wiki здатні автоматизовано обробляти дані з чітко визначеною семантикою, а це дає змогу істотно розширити функціональність таких систем.

*Програмне забезпечення для семантизації Wiki-ресурсів*

Для кращого розуміння того, що можна назвати семантизацією Wiki, доцільно порівняти можливості існуючих програмних систем, які більш детально розглянуті в [5, 6].

*Приклади семантичного розширення Wiki-технології*

*AceWiki (http://attempto.ifi.uzh.ch/acewiki)* [7] використовує обмежену підмножину англійської мови Attempto Controlled English (ACE). Формальні твердження є основним

вмістом самої Wiki. Таким чином, система намагається інтегрувати онтологію, правила і мову запитів. Редактор дозволяє користувачеві безпосередньо вводити висловлювання або використовувати керовану форму вибору з існуючої онтології. AceWiki концептуалізує кожну Wiki-сторінку як поняття і здійснює виведення OWL для базової онтології.

*Kiwi* (http://www.kiwi-project.eu/) [8] створює екземпляри даних на основі існуючої онтології, а також надає засоби для створення та редагування онтологій. Kiwi використовує інтерфейс, подібний до Mediawiki, та базується на таких розробках Semantic Web, як RDF і OWL. Формальна структура знань використовується для покращення навігації та надання рекомендацій під час редагування. У системі реалізоване логічне виведення для підтримки користувача у виконанні завдань. Система пропонує і WYSIWYG-редактор, що призначений для користувачів, які не мають досвіду роботи з Wiki-редактором.

*KawaWiki* [9] надає повну формальну структуру для даних на основі RDF і RDFS. Архітектура Wiki-ресурсу поділяється на три основні шари: 1) схему RDF, що визначає базову онтологію і використовується для перевірки шаблонів RDF; 2) шаблони RDF, які визначають тип сторінок Wiki, які можуть бути створені користувачем; 3) контент Wiki-ресурсу, який користувач отримує для редагування.

*OntoWiki* [10] призначена для розробки баз знань і спирається на подання даних у форматі RDF. Вона не надає Wiki-інтерфейс для вводу ПМ-тексту для подання понять, але підтримує декілька функцій спільної роботи та дозволяє встановлювати плагіни. OntoWiki реалізує кожну сторінку як ресурс, зберігаючи їх як твердження RDF. Знання в системі презентовані за допомогою «інформаційної карти», збагаченої зручними інтерфейсами для візуалізації й редагування контенту (WYSIWYG редактор для RDF, контроль версій, статистика, підтримка співтовариства тощо). Кожен вузол, поданий як сторінка системи, в інформаційній карті зв’язаний з відповідним цифровим джерелом.

*SMW* [11] є семантичним розширенням до Mediawiki, що дозволяє користувачам додавати відношення та властивості до Wiki-сторінок. Semantic Mediawiki зберігає семантичні дані в базі даних MySQL у Mediawiki, які також можуть бути експортована як RDF.

Основні семантичні поняття у Semantic MediaWiki – це категорії, що дозволяють користувачам класифікувати Web-сторінки (присутні й у звичайних Wiki-ресурсах), семантичні властивості, які дозволяють встановлювати зв’язки елементів контенту Wiki-сторінок з іншими сторінками Wiki ресурсу та з даними (дата, число, текстовий рядок, географічні координати тощо), визначаючи семантику цих зв’язків, та семантичні запити, в яких можуть використовуватися категорії та семантичні властивості. Результати семантичних запитів вставляються у відповідні Wiki-сторінки та автоматично оновлюються, якщо змінюється контент тих сторінок, з яких вони здобувають інформацію. Це забезпечує цілісність та актуальність створюваного Wiki-ресурсу.

*KnowWE* є розширенням JSPWiki, що додає до нього семантичну функціональність. Його механізми аналізу та розв'язування проблем також будуються на проекті d3web. Кожна сторінка є поняттям в контрольованій онтології. Семантика включена в Wiki-розмітку і надає три альтернативи: явну розмітку знань, семантичну анотацію та сегментований текст. Текст анотується онтологічним контентом. Крім того, знання для вирішення проблем в тексті можуть бути явними, з використанням, наприклад, правил. Крім того, забезпечується обмін знаннями за допомогою онтологій OWL.

*Tiki Wiki* CMS Groupware є одним з найбільш багатофункціональних пакетів систем керування контентом (Content Management System – CMS), який забезпечує визначення деякі семантичні відношень між Wiki-сторінками.

*Knoodl* – це колаборативний редактор Web-онтології. Кожен ресурс є онтологією і має свою власну Web-сторінку, яка містить як структурований контент з онтології, так і неструктурований контент

у вигляді Wiki-тексту. Контент у Knoodl організований у спільноти, які можуть створюватися будь-якими користувачами. Онтології можна імпортувати і експортувати як OWL-файли, з пов'язаним з ними Wiki-текстом або без нього.

*Параметри порівняння*

Існує багато різних підходів до порівняння засобів семантизації Wiki. З методологічної точки зору при додаванні семантики до Wiki можна виділити два підходи:

- текстові підходи, які збагачують традиційне Wiki-середовище семантичними анотаціями (Semantic MediaWiki);
- логіко-орієнтовані підходи, що використовують семантичні Wiki як засіб для онлайнового редагування онтологій (OntoWiki, Knoodl).

Для того, щоб порівняти різні варіанти семантизації Wiki, в [12] запропоновано два виміри: 1) перспективи користувача – скільки технічних навичок користувач повинен мати для того, щоб використовувати Wiki та додати свій внесок до онтології; та 2) виразність знань – наскільки виразною є результуюча онтологія (рис. 1).

Перспектива користувача (вісь x) розрізняє такі категорії користувачів:

- *Повсякденний користувач* (Everyday user) – Користувач, який знайомий з використанням конкретних програм, без адміністративних навичок, моделювання або програмування;
- *Потужний користувач* (Power user) – користувач, який знайомий з використанням більш широкого кола програм та адмініструванням власного комп'ютера, але без навичок моделювання;
- *Професійний користувач* (Professional user) – досвідчений користувач, який обізнаний у використанні та адмініструванні різноманітного програмного забезпечення, має знання з моделювання та програмування, але не знає технологій Semantic Web (наприклад, OWL і RDF);
- *Онтолог* (Ontologist) – експерт з онтологічного аналізу, який знає, як використовувати технології Semantic Web, зокрема онтології.

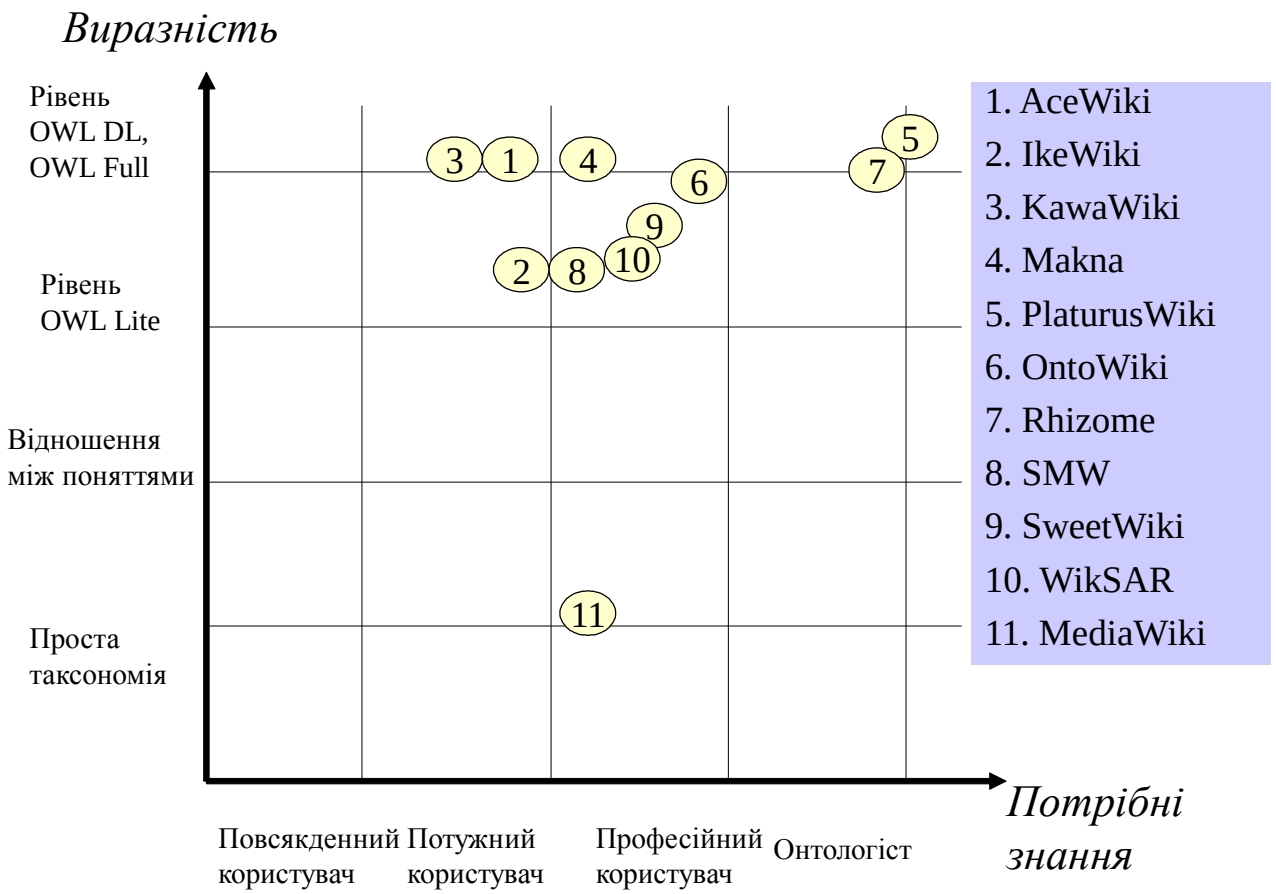


Рис. 1. Порівняння семантичних Wiki

Шкала експресивності (вісь у) виокремлює широкі рівні експресивності та формалізму, які відноситься до того, наскільки виразною є результуюча онтологія:

• *проста таксономія* – таксономічна класифікація Wiki-сторінок без доданої семантики щодо відношень між сторінками та термінами;

• *відношення між поняттями* – поняття в Wiki пов'язані між собою за допомогою семантичних анотацій;

• *рівень OWL Lite* – краща формалізація, ніж прості відношення, з деякими обмеженнями, подібними до функціональності, яку пропонує OWL Lite;

• *рівень OWL DL, OWL Full* – функціональність, подібна до функціональності OWL DL і OWL Full, що підтримує більшу виразність обмежень і відношень з характеристиками властивостей, таких як транзитивність, функціональність, відсутність перетину тощо.

Системи в лівій частині графіка (наприклад, AceWiki і KawaWiki) обмежують дії користувачів і не потребують від них складних навичок, а Wiki у правій частині графіка (KiwiWiki, KnowWE і до деякої міри OntoWiki) мають більшу виразність, але вони потребують від користувачів знання технологій Semantic Web або спеціального синтаксису, а система Knoodl, що знаходиться праворуч, потребує навичок професійного онтолога. Системи, що знаходяться в середній частині графіка, такі як SMW andTiki Wiki CMS Groupware, намагаються збалансувати достатню виразність і знання, необхідні для застосування цієї виразності.

Таке порівняння дозволяє робити припущення щодо знань, що зафіксовані в таких Wiki-ресурсах, та оцінити зусилля користувачів, потрібні для їх створення та підтримки. Більш детально результати порівняння розглянуто в [6].

Це порівняння між семантичними Wiki обґрунтовує вибір MediaWiki і Semantic MediaWiki як базової технології для вирішення питання взаємодії в гетерогенних середовищах, які потребують обміну знаннями та спільного використання:

- з точки зору користувача Semantic MediaWiki базується на тексті і легке в освоєнні і використанні;
- з точки зору виразності знань Semantic MediaWiki являють собою рішення, вдало збалансоване між виразністю та необхідними знаннями.

Semantic MediaWiki використовується у великій кількості як публічних, так і приватних Wiki-ресурсів.

*Модель даних у MediaWiki*

MediaWiki надає об'єктно-орієнтовану модель зберігання документів середньої деталізації (іменовані фрагменти – це Wiki-сторінки). Модель зберігання подібна в багатьох аспектах на поширені зараз noSQL, хоча попередні MediaWiki зазвичай використовують mysql.

Таким чином, знання створюються і підтримуються в Semantic MediaWiki еволюційно через Web-інтерфейс самою спільнотою користувачів. Загалом, Wiki можна порівняти з системами управління контентом (CMS – content-management systems), що створють та колективно використовують знання, але Wiki надають набагато простіший і більш відкритий інтерфейс для цього технологічного процесу.

*Двигун MediaWiki як основа створення Wiki-ресурсів*

MediaWiki базується на архітектурі LAMP (PHP на стороні сервера і Javascript на стороні клієнта, з базою даних MySQL). Ядро MediaWiki дозволяє створювати та редагувати сторінки, встановлювати їх категорії та взаємозв'язки, переглядати контент цих сторінок тощо. З точки зору інформаційної структури MediaWiki надає систему шаблонів і можливість категоризації сторінок.

Функціональні шари MediaWiki означають також багаторівневу розробку Wiki:

• *ядро* – базове програмне забезпечення та повний набір API, які надаються для зовнішнього програмного доступу до основних функцій MediaWiki;

• *розширення* – засоби додавання розширених функціональних можливостей до ядра MediaWiki: вдосконалені функції редагування, розширені інтерфейси корис-

тувача та семантичні функції. Зазвичай розширення розробляються мовою PHP, використовуючи механізм перехоплення (hook) для використання подій MediaWiki;

• *шаблони* – дозволяють створювати зразки Wiki з метою зменшення суперечностей у Wiki-ресурсі та гнучкого розвитку схеми;

• *створення контенту* – використання вихідних даних, отриманих від попередніх шарів, та ефективне створення екземпляра MediaWiki, тобто Wiki-ресурсу.

Ця багатошарова структура співпраці в розвитку MediaWiki дозволяє виділити різні групи людей, що приймають участь у створенні Wiki-ресурсу:

- розробники базового програмного забезпечення,
- розробники розширень,
- розробники шаблонів,
- розробники контенту.

З точки зору соціального програмного забезпечення, соціальним об'єктом у MediaWiki є безпосередньо MediaWiki.

З точки зору об'єктно-орієнтованого програмування (ООП), розробка контенту в MediaWiki є створенням екземпляру Wiki-ресурсу.

Основною одиницею Wiki-ресурсу, побудованого на основі MediaWiki, є Wiki-сторінка. У Wiki інформацію можна зберігати на неформальному, напівформальному або формальному рівні, залежно від рівня співпраці спільноти розробників контенту.

MediaWiki надає кілька механізмів для створення структури у Wiki. На цьому рівні дані у Wiki структуровані, напівструктуровані або неструктуровані. За замовчуванням не існує якоїсь обов'язково необхідної структури, але відсутність структурування робить контент дуже складним для навігації. Тому для Wiki-ресурсів, що використовують MediaWiki, зазвичай будується їх власна структура, що відповідає специфіці тієї предметної області, для якої розробляється цей інформаційний ресурс.

Для побудови такої структури в MediaWiki можуть бути використані наступні механізми:

- *Простори імен* (*Namespaces).* Всі сторінки в MediaWiki поділяються на окремі простори імен, що надаються моделлю зберігання. Це дозволяє повторно використовувати базову модель для об'єктів контенту першого класу (наприклад, у е-ВУЕ це сторінки гасел енциклопедії), а також створювати об'єкти, що використовуються при використанні гіпертексту (це можуть бути різноманітні медіа-елементи (двійкові плюс метадані) в просторі імен «Файл», "Ілюстрація" або "Відео", блоки програмування та фрагменти структурованого тексту в просторі імен «Шаблон»);
- *Категорії* (Categories) – це сторінки простору імен Категорії, які можна використовувати для автоматичного групування сторінок. Ці механізми передбачають побудову таксономій у Wiki-ресурсі (наприклад, підкатегорії категорії "Галузь знань" в е-ВУЕ утворюють таксономічну структуру). Категорії Wiki відповідають визначенню класу в об'єктно-орієнтованій парадигмі, а сторінки Wiki-ресурсу, які належать до певних категорій, є екземплярами цих класів;
- *Шаблони* (*Templates).* MediaWiki надає простір імен, присвячений транзакції, який називається Шаблон. Сторінки у цьому просторі назв призначені для включення до інших статей повторюваних елементів. Цей механізм використовується також для уникнення суперечностей у Wiki та для більш уніфікованого подання інформації. Це поширений спосіб структурування даних. Наприклад, в е-ВУЕ створено шаблони для опису типових інформаційних об'єктів.

*Використання шаблонів структурування контенту Wiki-ресурсу*

Модель на основі шаблонів передбачає гнучку розробку схеми Wiki-ресурсу. Кожен шаблон визначає клас з вільно обраними атрибутами (еквівалент "тип сутності"), екземпляри яких можуть бути вільно вбудовані в інші об'єкти. Екземпляри шаблонів можуть бути пов'язані ієрархічно. Наприклад, у е-ВУЕ за допомогою набору ієрархічно організованої множини шаблонів подаються знання щодо типових інформаційних об'єктів (ІО).

Механізми шаблонізації контенту використовуються у Wiki для того, щоб дати авторам контенту можливість заповнити схему (модель) конкретними значеннями. Сторінки з такими шаблонами можуть використовуватися для опису подібних елементів.

Вивчення сучасних механізмів Wiki-шаблонів дозволяє зрозуміти, як саме вони можуть бути використані для створення семантично багатих Web-сторінок. При цьому можна виділити дві альтернативні моделі шаблонів, які можна знайти в сучасних Wiki-движках:

− *функціональні* шаблони (functional templating), коли шаблони викликаються за ім'ям і переданими параметрами;

− шаблони *створення* (creational templating), в яких шаблони просто копіюються як новий контент на початок історії перегляду сторінки.

На відміну від традиційних CMS, де шаблони розглядаються як спеціальні ресурси, у Wiki-технології шаблони та екземпляри розглядаються на одному рівні: це дозволяє користувачам керувати обома видами інформації подібним чином.

В обробці шаблонів беруть участь два види Wiki-сторінок:

- *сторінка шаблону* (наприклад, *Шаблон:Річка*): основна сторінка, інформацію з якої використовувати ті сторінки, на яких цей шаблон викликається;
- *сторінка екземпляру шаблону* (наприклад, *Дніпро*): сторінка, контент якої побудовано з використанням виклику шаблону з певними конкретними значеннями його параметрів.

У функціональних шаблонах слід відмітити кілька важливих характеристик:

- міцний і постійний зв'язок між шаблоном та його екземплярами (зміни, що вносяться на сторінку шаблону, відразу ж відображаються на усіх сторінках, де цей шаблон викликається);
- різний вихідний код шаблонів та їх екземплярів;
- екземпляри шаблонів є нелінійною розміткою щодо її форми інтерпретації.

Важливо, що за допомогою шаблонів дані, що містяться на Wiki-сторінках, можуть бути структуровані безпосередньо – шляхом їх зв’язування із семантичними властивостями та з категоріями. Незважаючи на це, такі дані, вбудовані в екземпляри шаблонів, складно вилучати і повторно використовувати, хоча різні дослідники пропонують для виконання цієї задачі досить ефективні евристичні методи. Наприклад, один з таких методів використовується у проекту Dbpedia [13], що спрямований на здобуття структурованої інформації з Вікіпедії та забезпечення доступності цієї інформації в Web-середовищі.

У Вікіпедії, яка не є семантизованим ресурсом, нині теж досить часто застосовуються шаблони. Найчастіше такі шаблони призначені лише для макетування інформації, однак відносна частка шаблонів, що забезпечують структурування контенту через елементи Wiki-розмітки, що відмінні від семантичних властивостей, постійно збільшується (приблизно від чверті до третини сторінок Вікіпедії нині містять структуровану інформацію, придатну для машинної інтерпретації). Наприклад, широко застосовується особливий тип шаблонів – інфорбокси (infoboxes), що призначені для створення послідовно форматованих блоків для певного контенту в статтях, які описують екземпляри певного специфічного типу (наприклад, типові ІО). Щоб розкрити семантику, закодовану в таких шаблонах, використовують різні методи здобуття знань. Застосування таких методів може бути корисним і для семантизованих Wiki-ресурсів, тому що вони забезпечують отримання тих знань, що неявно містяться в контенті, але не формалізовані повністю для автоматичного видобуття. Наприклад, в [14] запропоновано такий алгоритм здобуття інформації, що працює в п'ять етапів.

1. Обрати усі сторінки Wiki-ресурсу, які містять шаблони. Таку множину сторінок Wiki-ресурсу можна отримати за допомогою запиту SQL, який шукає входження роздільника шаблону "{{" в тексті бази даних MediaWiki. Запит SQL дозволяє також вибрати тільки сторінки певних категорій Wiki-ресурсу або сторінки, що містять певні шаблони.

2. Здобути з обраних сторінок важливі шаблони. Усі шаблони на сторінці Wiki-ресурсу можна отримати за допомогою рекурсивного регулярного виразу. Через те, що шаблони використовуються у Wiki-ресурсі для різних цілей, потрібно обрати тільки ті шаблони, що містять структуровану інформацію. Це можна зробити на основі такої евристики: ігнорувати шаблони з одним або двома атрибутами (такі шаблони зазвичай діють як ярлики для попередньо визначених розділів), а також ігнорувати шаблони, кількість використання яких не перевищує певний поріг (такі шаблони можуть бути помилковими або неважливими). Алгоритм здобуття може бути додатково сконфігуровано для ігнорування певних явно визначених шаблонів або груп шаблонів, наприклад, якщо ці шаблони пов'язані з іншими аспектами структурування інформації, такими як умови створення та використання контенту.

3. Виконується аналіз структури кожного обраного шаблону і створюються відповідні трійки "суб'єкт-предикат-об'єкт". URL-адреса, отримана з назви Wiki-сторінки, на якій знайдено шаблон, використовується як суб'єкт для шаблонів, які використовуються не більше одного разу на сторінці. Для шаблонів, що зустрічаються на сторінці більше за один раз, потрібно генерувати новий ідентифікатор, який використовується як суб'єкт. Кожен атрибут шаблону відповідає предикату трійки, а відповідне значення атрибута перетворюється на його об'єкт. Шаблони MediaWiki можуть бути вкладеними, тобто значення атрибута в шаблоні може знову бути шаблоном. У такому випадку генерується порожній вузол, що пов'язує значення атрибута з новостворюваним екземпляром для вкладеного шаблону.

4. Обробка значень об'єктів для створення відповідних URI-посилань або літеральних значень. Для гіперпосилань MediaWiki створюються відповідні URI-посилання, які забезпечують перехід на пов'язану Wiki-сторінку. Типізовані літерали генеруються для рядків і числових значень. Загальновживані одиниці виміру (наприклад, кг для кілограмів, с для секунд) виявляються і кодуються як спеціальні типи даних, однак перетворення між різними масштабами (наприклад, між см, м, км) не виконується автоматично. Крім того, розпізнаються списки, що розділяються комами, і, залежно від параметрів конфігурації, перетворюються на списки RDF або на окремі висловлювання.

5. Визначення належності Wiki-сторінки до класу. Сторінки Wiki-ресурсу організовані за категоріями. У деяких випадках ці категорії можуть бути інтерпретовані як класи, що включають екземпляри, описані сторінками Wiki-ресурсу з відповідній категорії. Крім того, назва шаблону може бути індикатором належності до класу. Самі категорії є сторінками Wiki-ресурсу і також можуть бути організовані у надкатегорії (такі відношення можуть відображатися через зв'язок «бути пов'язаними з»).

На основі використання цього методу здобуття знань з Wiki-ресурсів, що структуровано за допомогою шаблонів, його авторами розроблено набір рекомендацій для побудови Wiki-шаблонів. Дотримання цих рекомендацій не тільки корисно для семантичного здобуття інформації з Wiki-сторінок, але зазвичай також покращує відповідний шаблон у цілому, тобто він стає більш зручним для використання авторами Wiki-сторінок. Будуючи нові шаблони та модифікуючи існуючі, доцільно дотримуватися наступних правил:

- не визначати атрибути в шаблонах, які кодують інформацію для макетування сторінки – це правило відповідає відомому принципу відокремлення контенту від його подання, тобто HTML-розмітка повинна використовуватися в значеннях атрибутів тільки у разі необхідності;
- використовувати лише один шаблон для певного інформаційного об'єкта замість того, щоб використовувати окремий шаблон для кожного атрибута (це потребує створення більш складних шаблонів, але спрощує процес здобуття знань щодо структури цих об'єктів);
- кожен атрибут шаблону повинен мати точно одне значення в шаблоні статті (хоча таке значення може бути спи-

ском значень), але не слід змішувати декілька операторів (з точки зору RDF) в межах одного значення атрибута (тобто потрібно більш чітко відрізняти різні атрибути шаблона, виокремлюючи їх можливі значення);

- не бажано використовувати різні шаблони для однакової цілі (таке дублювання структури різними шаблонами не тільки ускладнює здобуття знань й потребує додаткових операцій з їх об'єднання, але й спричиняє накопичення застарілих шаблонів із некоректно поданою інформацією);
- не слід використовувати різні імена атрибутів для одного і того ж типу контенту (наприклад, в різних шаблонах називати подібний атрибут "Назва міста", "Назва селища" та "Позначення міста") і не доцільно використовувати одне ім'я атрибута для різних типів вмісту (наприклад, в шаблонах "Персоналія" та "Місто" використовувати атрибут "Ім'я");
- намагатися використовувати стандартні подання для елементів, щоб вони могли бути виявлені алгоритмом здобуття.

Крім того, для більш зручного здобуття знань та коректного введення даних у Wiki-ресурсах доцільно мати наступні можливості:

- зафіксувати тип даних для значення атрибута шаблону. У Semantic MediaWiki це можна робити за допомогою типів значень семантичних властивостей, але тільки для базових типів даних. Наприклад, можна вказати у визначенні шаблону, що значення атрибуту "Рік народження" має бути числом, а "Місце народження" – посиланням на Wiki-сторінку. Але необхідно враховувати, що іноді виникає необхідність використовувати природномовні описи для значень атрибутів, наприклад, значення року народження може бути "у середині 3 століття" або "приблизно між 1100 та 1105 роками", що призводить до застосування менш жорстких обмежень на дані, наприклад, усі дані описуються як текстовий рядок;
- використовувати специфічні для певного Wiki-ресурсу типи даних, наприклад, забезпечити засоби введення тегів природної мови (якщо певне значення атрибута може бути подано різними мовами, наприклад, прізвище особи може бути подано українською та рідною мовою особи) та одиниць виміру для елементів, які можуть оцінюватися у різних одиницях (наприклад, у метрах та футах).

Для атрибуту, який визначено у шаблоні, MediaWiki дозволяє знайти шаблони, де цей атрибут вже використовується, та показувати інші характеристики атрибуту, щоб дати розробнику шаблону можливість перевірити, чи мають ці атрибути однакове семантичне значення. Це спрощує розробку шаблонів та дозволяє уніфікувати імена параметрів, не викликаючи суперечностей. Наприклад, в e-BYE атрибут "Площа" має однакові одиниці виміру та тип значення для шаблонів "Країна" та "Озеро", тому доцільно використовувати в них однакову назву. Але використання атрибуту "Ім'я" в шаблонах "Персоналія" та "Літературний твір" недоцільно, якщо в першому випадку цей параметр має значення типу "текстовий рядок" та характеризує особисте ім'я особи, а в другому – це посилання на прізвище автора, що має тип "Wiki-посилання", і тому краще замінити його на параметр "Автор".

*Модель даних Semantic MediaWiki*

*Semantic MediaWiki* – це семантично розширений Wiki-двигун, який дозволяє користувачам анотувати контент Wiki явно визначеною інформацією, придатною для автоматичної обробки (машиночитаною). Він підтримує додавання структурованої та семантично анотованої інформації у Wiki з використанням відповідного синтаксису.

Semantic MediaWiki (SMW) є надбудовою над інструментальним засобом побудови Wiki-сайту MediaWiki. Переваги SMW – це обробка інформації на семантичному рівні, наявність засобів групового керування знаннями, відносно висока виразна потужність, надійна реалізація і зручний інтерфейс користувачів, наявність документації та спільнот користувачів. SMW дозволяє інтегрувати інформацію з різних Wiki-сторінок, здійснюючи пошук

на рівні знань, та генерувати за Wiki-сторінками онтологічні структури, які можуть використовувати інші системи.

Використовуючи семантичні елементи, SMW вирішує такі основні проблеми сучасних Wiki:

- послідовність контенту (*Content consistency)*: інформація, що зустрічається на різних Wiki-сторінках, має бути несуперечною.
- доступність знань (*Knowledge Access*): пошук і порівняння інформації з різних сторінок у Wiki-ресурсі великого обсягу потребує відповідних засобів.
- повторне використання знань (*Knowledge reuse*): на відміну від жорсткого контенту на основі тексту у традиційних Wiki, який може використовуватися лише для читання сторінок у браузері або подібній програмі, семантизація є основою для доступу до Wiki-контенту з інших застосунків та експорт знань у поширені формати їх подання.

Крім категорій, в SMW для структурування інформації використовуються такі механізми, як *семантичні властивості*. Вони дозволяють семантично пов'язувати Wiki-сторінки як між собою, так і з різними даними. Кожна семантична властивість має тип, назву і значення, а також власну Wiki-сторінку в спеціальному просторі імен, яка дозволяє визначати її місце в ієрархії властивостей та документувати те, як цю властивість необхідно використовувати. Кожна семантична властивість має власну Wiki-сторінку. На цій сторінці можна явно визначити тип значень, що приймає ця властивість, використовуючи конструкцію [[Has type::xxx]]. Has type розпізнається SMW як спеціальна властивість, тобто значення такої властивості є посиланням на іншу Wiki-сторінку. В Semantic MediaWiki підтримуються наступні типи властивостей: String, Number, Annotation UR1, Boolean, Email, Text, Geographic Coordinates тощо.

Динамічно генерований контент може створюватися за допомогою вбудовування запитів у Wiki-сторінки. Імпорт зовнішніх даних з існуючих онтологій типу FOAF, SIOC або Doublin Core може здійснюватися шляхом зіставлення Wiki-анотацій з елементами цих словників.

Щоб зробити можливим зовнішнє повторне використання інформації з Wiki-ресурсу, що базується на Semantic MediaWiki, можна отримувати формальний опис для однієї або кількох статей через Web-інтерфейс у форматі OWL/RDF. Оскільки SMW чітко дотримується стандарту OWL DL, експортовану інформацію можна повторно використовувати в різних інструментах.

Синтаксис семантичних анотацій використовує мову сценаріїв для опису контенту сторінок Wiki, який розширюється в SMW за допомогою наступних трьох основних наборів семантичних анотацій.

*Визначення класів і властивостей:* SMW повторно використовує простір імен "Category" ("Категорія") для визначення класів. Наприклад, сторінка Wiki з назвою "Category: Images" призначена для подання класу всіх зображень, а також новий простір імен під назвою "Властивість" для визначення властивостей понять. Завдяки такому підходу, SMW підтримує як бінарні, так і n-арні властивості.

*Аксіоми*: SMW дозволяє декларувати відношення "клас-підклас" і "властивість-субвластивість". Наприклад, можна декларувати, що категорія "Зображення" є підкатегорією "Медіа", додавши на сторінку цієї категорії анотацію "[[Category: Медіа]]". Це дозволяє утворювати таксономію категорій та властивостей, яка є дуже корисною як для навігації у Wiki-ресурсі, так і для коректного створення нових категорій та властивостей. SMW також підтримує затвердження еквівалентності між двома Wiki-сторінками або між двома класами за допомогою механізму переспрямування (redirection) у MediaWiki. Для цього використовується конструкція «REDIRECT [[x]]. Наприклад, в e-BYE за допомогою цього механізму організовано відсильні статті, коли існує кілька загальновживаних формулювань того самого терміну.

*Твердження про екземпляри:* SMW дозволяє оголошувати сторінку екземп-

ляром класу або суб'єктом RDF-трійки. Наприклад, якщо сторінка Wiki «Зображення Хрещатику» містить анотацію «[[Категорія: Зображення]] та [[Type :: Фотографія]]», це означає, що Зображення Хрещатику є цифровим ресурсом і має тип «Фотографія». Також можна створити екземпляр n-арної властивості. Наприклад, на Wiki-сторінці "Зображення Хрещатику" анотація "[[Загальні ключові слова :: Зображення Хрещатику; пейзаж]]" означатиме, що Зображення Хрещатику відноситься до предметної категорії пейзажей.

З точки зору функціональності механізм семантичних анотацій відображає три основні аспекти функціональності:

• функціональність *гіпертекстових посилань* – успадкована від MediaWiki для встановлення зв'язків між Wiki-сторінками;

• функціональність *подання знань* у семантичну мережу;

• функціональність бази даних, що забезпечує подання n-арних кортежів властивостей в SQL-подібному стилі.

Шаблони в Semantic MediaWiki – це Wiki-сторінки в спеціальному просторі імен. Використання шаблонів корисне для структурованого введення інформації користувачами. Агрегація властивостей в шаблонах та агрегація шаблонів подані у семантичних формах.

Приклад форми:

Property: CollectionName | hasType: String Property: BestQualityURI | hasType: URL Property: Type| hasType: String

Приклад шаблону:

{{Metadata | CollectionName = | BestQualityURI = | Type = }}

Form: {{#forminput:Form-MetadataEntry

{{{for Metadata

{| class="formtable" | Collection Name: | {{{field | CollectionName}}}

...

{{{end template}}}

}}

Для пошуку Semantic MediaWiki використовує вбудовані запити.

Семантичні властивості та вбудовані запити разом із семантичними формами та шаблонами є потужним інструментом, який може підтримувати моделювання для різних потреб проектів.

## Структура семантизованого Wiki-ресурсу як основа для побудови його онтологічної моделі

Використання семантичних Wiki-технологій для створення розподілених інформаційних ресурсів не тільки дозволяє досить легко додавати структурування до неструктурованих даних (НСД), але й є джерелом фонових знань для аналізу довільних природномовних текстів відповідної предметної області. Створення е-ВУЕ як семантизованого Wiki-ресурсу дозволяє вдосконалити процес генерації таких знань. Використання онтологічного аналізу є основою для переходу від неструктурованого контенту [4] до розподіленої бази знань, придатної для повторного використання.

З точки зору онтологічного аналізу, кожна Wiki-сторінка являє собою онтологічний елемент, тобто елемент одного з RDF-класів – Thing, Class, ObjectProperty, DatatypeProperty, AnnotationProperty. Крім того, кожна стаття має власний URI, який дозволяє уникнути плутанини між поняттями і HTML- сторінками. Зазвичай, статті є екземплярами класів онтології OWL, категорії – класами, а відношення – об'єктними властивостями онтології.

Виходячи з цього, для будь-якої сторінки SMW за запитом може генерувати відповідний OWL/RDF-файл. Найпростіший спосіб отримати цей RDF – використати посилання "Переглянути як RDF" ("View as RDF"), що знаходиться в нижній частині кожної анотованої сторінки. Ця сторінка може виступати як кінцева точка (endpoint) для зовнішніх сервісів (зовнішньої точки доступу), які хочуть отримати доступ до семантичних даних SMW. На жаль, ця функція реалізована дуже невдало та підтримує надто мало опцій.

Оскільки SMW сумісна з моделлю знань OWL DL, то існує можливість використання в Wiki існуючих онтологій. Це можливо здійснити двома шляхами: імпорт онтології дозволяє створювати і модифікувати сторінки в Wiki для подання відношень, заданих в деякому існуючому OWL DL-документі; а повторне використання словника дозволяє користувачам відображати (задавати відповідності) Wiki-сторінки на елементи існуючих онтологій.

Функція імпорту онтології для читання RDF-документів використовує інструментарій RAP toolkit. Він витягує RDF-твердження, які можуть бути подані у Wiki. Найменування статей імпортованих елементів витягуються з їх міток (labels), або, в разі відсутності мітки, з ідентифікатора розділу їх URI. Основною метою імпорту є ініціалізація (первинне автоматичне завантаження) основи-шаблону для заповнення Wiki. Крім того, імпорт онтології вставляє спеціальні анотації, які генерують еквівалентні твердження в експорт OWL (тобто. owl:sameAs, owl:equivalentClass, або owl:equivalentProperty). Імпорт онтологій дозволений тільки для адміністраторів сайту, оскільки це може бути використано для спаму.

Імпорт словника дозволяє користувачам ідентифікувати елементи Wiki, вказавши зв'язок з елементами існуючих онтологій. Наприклад, Category:Person може безпосередньо експортуватися в клас foaf:Person словника Friend-Of-A-Friend. Wiki-користувачі можуть вирішувати, які сторінки Wiki повинні мати зовнішню семантику, проте набір наявних зовнішніх елементів управляється тільки адміністраторами. Вводячи в словник Wiki деякий новий елемент, вони повинні упевнитися в тому, що повторне використання словника співвідноситься з типами обмежень OWL DL. Наприклад, зовнішні класи, такі, як foaf:Person, не можуть бути імпортовані у відношення.

Експорт в OWL/RDF є засобом забезпечення зовнішнього повторного використання даних з Wiki, але тільки практичне застосування цієї функції може показати якість згенерованого RDF. З цією метою для видачі RDF розробники системи використовували ряд інструментів Semantic Web. SMW добре співпрацювало з найбільш відтестованими застосуваннями, такими, як FOAF Explorer, Tabulator RDF browser, або розширенням браузера Piggy Bank RDF.

Крім того, SMW надає сервіс для підтримки SPARQL запитів. Система базується на автономному (stand-alone) RDF сервері Joseki, який синхронізований з семантичним контентом Wiki. Синхронізація полягає у генерації RSS-фіду зі звітом про останні зміни, що відбулися у Wiki, для того, щоб швидко перезавантажити змінені статті. Таким чином, SPARQL-точка доступу (endpoint) демонструє можливість дзеркально відобразити RDF-контент Wiki за допомогою невеликих покрокових оновлень, і пропонує точку доступу для семантичних технологій, які повторно використовують ці дані.

*Формальна модель Wiki-онтології*

*Wiki-онтологія* – це онтологія, що відображає знання семантично розміченого Wiki-ресурсу (набору Wiki сторінок, що містять семантичну розмітку) [15]. Вона містить тільки ті знання, які можна безпосередньо здобути із семантичної розмітки. Тому в цій онтології відсутні, приміром, такі характеристики класів та властивостей, як еквівалентність, відсутність перетину тощо. Така онтологія може бути згенерована в результаті аналізу Wiki-ресурсу або, навпаки, створена до початку розробки самого ресурсу та використовуватися як основа для його семантичної розмітки, тобто її поняття та їх атрибути використовуються для позначення категорій та семантичних властивостей Wiki-ресурсу. На практиці зазвичай обидва варіанти використовуються ітеративно – спочатку генерується онтологія предметної області, потім вона використовується у розмітці, а в процесі наповнення контентом Wiki-ресурсу вносяться вдосконалення і в саму онтологію.

Для опису формальної моделі Wiki-онтології пропонується використовувати наступну формальну модель онтології $O = \langle X, R, F, T \rangle$ [16], що складається з наступних елементів:

- $X = X_{cl} \cup X_{ind}$ – множина концептів онтології, де
  - $X_{cl}$ – множина класів,
  - $X_{ind}$ – множина екземплярів класів, таких, що $\forall a \in X_{ind} \exists A \in X_{cl}, a \in A$;
- $R = r_{ier\_cl} \cup \{r_i\} \cup r_{ier\_prop} \cup \cup \{p_j\} \cup p_{ier\_prop}$ – множина відношень між елементами онтології, де
  - $r_{ier\_cl}$ – ієрархічні відношення між класами онтології – це структури часткового впорядкування з верхнім елементом *Thing*, що можуть встановлюватися між класами онтології і характеризується такими властивостями, як антисиметричність і транзитивність, $r_{ier\_cl} : X_{cl} \to X_{cl}$;
  - $\{r_i\}$ – множина об'єктних властивостей, що встановлюють відношення між екземплярами класів: $r_i(a, a \in X_{ind}) = b, b \in X_{ind}$, $r_i : X_{ind} \to X_{ind}$;
  - $r_{ier\_prop}$ – ієрархічні відношення між об'єктними властивостями класів онтології – це структури часткового впорядкування з верхнім елементом *topObjectProperty*, що можуть встановлюватися між класами онтології і властивостями класів і характеризується такими властивостями, як антисиметричність і транзитивність, $r_{ier\_prop} : \{r_i\} \to \{r_i\}$;
  - $\{p_j\}$ – множина властивостей даних, що встановлюють відношення між екземплярами класів і значеннями з T: $p_i(a, a \in X_{ind}) = t, t \in T$, $p_i : X_{ind} \to T$;
  - $p_{ier\_prop}$ – ієрархічні відношення між властивостями даних екземплярів класів онтології – це структури часткового впорядкування з верхнім елементом *topDataProperty*, що можуть встановлюватися між властивостями даних екземплярів класів онтології і характеризується такими властивостями, як антисиметричність і транзитивність, $p_{ier\_prop} : \{p_i\} \to \{p_i\}$;
- $F = \{F_{cl} \cup F_{prop}\}$ – множина характеристик, що можуть використовуватися для логічного виведення над онтологією:
  - $F_{cl}$ – множина характеристик класів онтології, що можуть застосовуватися для логічного виводу: еквівалентність, відсутність перетину тощо;
  - $F_{prop}$ – множина характеристик об'єктних властивостей екземплярів класів онтології, що можуть застосовуватися для логічного виводу: транзитивність, симетричність, антисиметричність, рефлексивність, антирефлексивність тощо);
- $T$ – множина типів даних (наприклад, рядок, ціле), значення з яких можуть приймати властивості даних класів онтології;
- $M$ – множина нелогічних правил ПрО (наприклад, рядок, ціле).

У Wiki-онтології $O_{wiki}$ множина концептів будується як поєднання таких елементів Wiki, як сторінки та категорії $X = X_{wiki\_categor} \cup X_{wiki\_page}$, пов'язаних різними видами відношень з $R = \{r_{ier\_cl}\} \cup \{r_{link}\} \cup \{r_{sem\_prop}\}$: множина класів – це множина категорій Wiki $X_{wiki\_categor}$, між якими існують ієрархічні відношення $r_{ier\_cl}$; множина екземплярів – множина Wiki-сторінок $X_{wiki\_page}$, між якими існують посилання $r_{link}$ та семантичні відношення $r_{sem\_prop_i}, i = \overline{0,m}$; множина типів даних доповнюється специфічним класом – "Wiki-сторінка". Ця модель може бути вдосконалена з урахуванням таких елементів Wiki, як шаблони, форми, спеціальні сторінки тощо.

У формальній моделі Wiki-онтології $O_{wiki}$ не семантизованого Wiki-ресурсу X ці множини мають наступний склад:

- X – це множини сторінок Wiki-ресурсу:

$$X_{wiki} = P_{user} \cup P_{categ} \cup P_{template} \cup P_{spec},$$

де

- $P_{user}$ – множина сторінок, створених користувачами,

- $P_{categ}$ – множина сторінок, що описують категорії,
- $P_{template}$– множина сторінок, що описують шаблони,
- $P_{spec}$ – множина інших спеціальних сторінок;

– $R = r_{ier\_categor} \cup R_{wiki}$

- $r_{ier\_categor}$ – ієрархічні відношення, що можуть встановлюватися між категоріями Wiki-ресурсу і характеризується такими властивостями, як антисиметричність і транзитивність,

$$r_{ier\_categor}: P_{categ} \to P_{categ};$$

- $R_{wiki} = \{"link","is\_a","use"\}$ – множина з трьох елементів, де "link" – відношення, що описує посилання однієї довільної Wiki-сторінки цього ресурсу на іншу Wiki-сторінку цього ресурсу (хоча в Wiki-ресурсах передбачені і посилання на інші види сторінок, у рамках даної моделі вони не враховуються), $"is\_a": X \to P_{categ}$ – відношення включення довільної сторінки до певної категорії, а $"use": X \to P_{template}$ – відношення, що формалізує включення Wiki-шаблону до довільної сторінки ресурсу;

– $T$ – множина типів даних, що підтримуються в MediaWiki (наприклад, рядок, ціле, число, дата, координата), значення з яких можуть приймати властивості даних класів онтології;

– інші множини такої Wiki-онтології є порожніми.

Формальна модель Wiki-онтології $O_{s\_wiki}$ для семантично збагачених Wiki-ресурсів є більш складною і включає ряд елементів, зв'язаних із семантичними властивостями [16]:

– X – це множини сторінок Wiki-ресурсу:

$$X_{sem\_wiki} = X_{wiki} \cup P_{s\_prop},$$

яка доповнюється множиною $P_{sem\_prop}$ – сторінками семантичних властивостей, деякі з який є семантично визначеними посиланнями на інші Wiki-сторінки:

$$P_{sem\_prop\_page} \subseteq P_{sem\_prop},$$

а інші зв'язують сторінки зі значеннями різних типів даних (ці типи даних визначаються на сторінках семантичних властивостей);

– $R = r_{ier\_cl} \cup R_{semant\_wiki}$, де

- $r_{ier\_cl} = r_{ier\_categor} \cup r_{ier\_property}$ розширюється порівняно із онтологією несемантизованого Wiki-ресурсу введенням окремого ієрархічного відношення $r_{ier\_property}$ для задання ієрархії семантичних властивостей;
- $R_{semant\_wiki} = R_{wiki} \cup \{"has\_property", "property\_value"\}$ – множина відношень, що доповнюється відношеннями, що пов'язані із семантичними властивостями Wiki-сторінок:

$$"has\_property": X \to P_{sem\_prop}$$

– відношення, що вказує на те, що довільна Wiki-сторінка має певну семантичну властивість, та

$$"property\_value": P_{sem\_prop} \to X \cup T$$

– відношення, що пов'язує семантичну властивість сторінки з її значенням, яке може бути як посиланням на іншу сторінку, так і літералом;

– $T$ – множина типів даних, що підтримуються в MediaWiki;

– інші множини такої Wiki-онтології є порожніми.

Вибір саме такої моделі онтології для даної задачі обумовлюється наступними причинами. По-перше, така модель має достатню виразність для розв'язання широкого класу інтелектуальних задач. По-друге, вона відповідає інтуїтивному уявленню про онтології, яке міститься в користувальницькому інтерфейсі редактора онтологій Protégé і тому легко поєднується з візуалізаціями елементів онтології в цьому програмному продукті. По-третє, ця модель досить легко інтегрується з різними інтелектуальними застосуваннями,

які підтримують семантичну обробку інформації (приміром, з семантичними Wiki-ресурсами, лексичними онтологіями).

## Онтологічна модель бази знань е-ВУЕ

Розглянемо проаналізовані вище засоби побудови та вдосконалення структури бази знань Wiki-ресурсу на прикладі науково-інформаційного наповнення портальної версії Великої української енциклопедії – е-ВУЕ (http://vue.gov.ua), що використовує вільне програмне забезпечення MediaWiki версії 1.29.1. та його семантичне розширення Semantic MediaWiki версії 2.5.5 – інноваційний проект зі створення національної енциклопедії на основі сучасних засобів подання знань.

Принциповими відмінностями е-ВУЕ від відомих онлайнових довідників та енциклопедій (приміром, від Вікіпедії) є: 1) орієнтація на авторські статті, тобто на оригінальний та якісний контент, підготовлений експертами відповідної області; 2) наявність складної системи семантичних відношень між гаслами, що базуються на наборах ієрархій категорій та семантичних властивостей; 3) рецензованість гасел, що забезпечує більший рівень об'єктивності та довіри до поданої інформації.

Через високу складність інформаційного наповнення е-ВУЕ виникає необхідність застосування засобів онтологічного аналізу для моделювання структури, взаємозв'язків та властивостей об'єктів, що складають контент цієї енциклопедії. Ці засоби дозволяють встановлювати співставлення між Wiki-онтологією (тобто тією онтологією, елементи якої лежать в основі семантичної розмітки Wiki-ресурсу) та довільною онтологією ПрО. Онтологічна модель знань е-ВУЕ дозволяє перетворити її на розподілену базу знань, що є джерелом корисної та перевіреної інформації як для людей, так і для інтелектуальних програмних засобів, необхідно створити цього енциклопедичного видання. Така модель формалізує відношення між її основними об'єктами, їх типами та властивостями.

*Категорії е-ВУЕ*

Кожне гасло е-ВУЕ може бути віднесено до довільної кількості категорій. Засоби Wiki-середовища дозволяють явно вказувати ієрархічні ("є підкатегорією") зв'язки між такими категоріями. Жодних зовнішніх обмежень на такі зв'язки в Semantic MediaWiki немає. Такі категорії можуть відображати різні аспекти, за якими можна класифікувати гасла енциклопедії. Деякі з них відображають специфіку предметної області енциклопедистики, інші можуть стосуватися, приміром, умов публікації та використання матеріалу.

Можна виділити кілька незалежних ієрархій категорій е-ВУЕ:

- напрямки знань та їх підкласи (приклад ієрархії підкатегорій: "Технічні науки" – "Радіотехніка та телекомунікації" – "Оптоелектронні системи");
- тип інформаційного об'єкта – "Персоналії", "Цивілізація" та "Природа", а також підкатегорії цих категорій;
- тип опублікування – категорії "е-ВУЕ" (всі сторінки гасел енциклопедії) та її підкатегорія "ВУЕ" (інформація, що опублікована у паперовій версії ВУЕ);
- мультимедійні матеріали;
- типи учасників побудови контенту – категорії "Автори ВУЕ", "Модератори".

З точки зору тематики найвищим рівнем класифікації в ієрархії категорій є поділ гасел на три категорії, що не перетинаються: "Персоналії", "Цивілізація" та "Природа". Усі повністю завершені гасла на порталі е-ВУЕ віднесено до прихованої категорії "е-ВУЕ", яка використовується в семантичних запитах і дозволяє не брати до розгляду та аналізу технічні й допоміжні сторінки, – це сторінки е-ВУЕ, на які можна посилатися із зовнішніх джерел.

*Шаблони е-ВУЕ*

Шаблони Semantic MediaWiki дозволяють автоматизувати введення відповідної інформації на сторінках та забезпечують уніфікацію імен категорій та семантичних властивостей. Крім того, вони дозволяють інтегрувати контент різних сторінок (з використанням убудованих запитів) та запобігати суперечностей у поданні інформації.

В процесі розробки енциклопедії виділено типові інформаційні об’єкти (ІО) – Wiki-сторінки з подібним набором розділів та близьким набором семантичних властивостей. Для таких типових ІО створено Wiki-шаблони, що дозволяють уніфіковано відображати відомості на екрані та забезпечують коректне введення значень семантичних властивостей.

На сьогодні в е-ВУЕ використовуються наступні шаблони базових типових ІО (рис. 2):

| | |
|---|---|
| Битва | Відзнаки |
| Війна | Група осіб |
| Історична подія | Книжкове видання |
| Країна | Материк |
| Мінерал | Місто |
| Модератор | Море |
| Музичний гурт | Озеро |
| Океан | Олімпійські_ігри |
| Організація | Періодичне видання |
| Персоналія | Порода |
| Рельєф | Річка |
| СМТ | Таксон |

*Семантичні властивості е-ВУЕ*

В е-ВУЕ використовуються семантичні властивості різних типів, які дозволяють як пов’язувати Wiki-сторінки на рівні знань, так і фіксувати значення їх атрибутів. Для кожного з типових ІО визначено набір таких властивостей та задано за допомогою шаблонів форму їх відображення. Наприклад, шаблон “Озеро” має такі семантичні властивості, як “Найбільша глибина” (число, єдине значення), “Країни” (посилання, можливо кілька значень), “Тип живлення” (текст) та “Площа” (число).

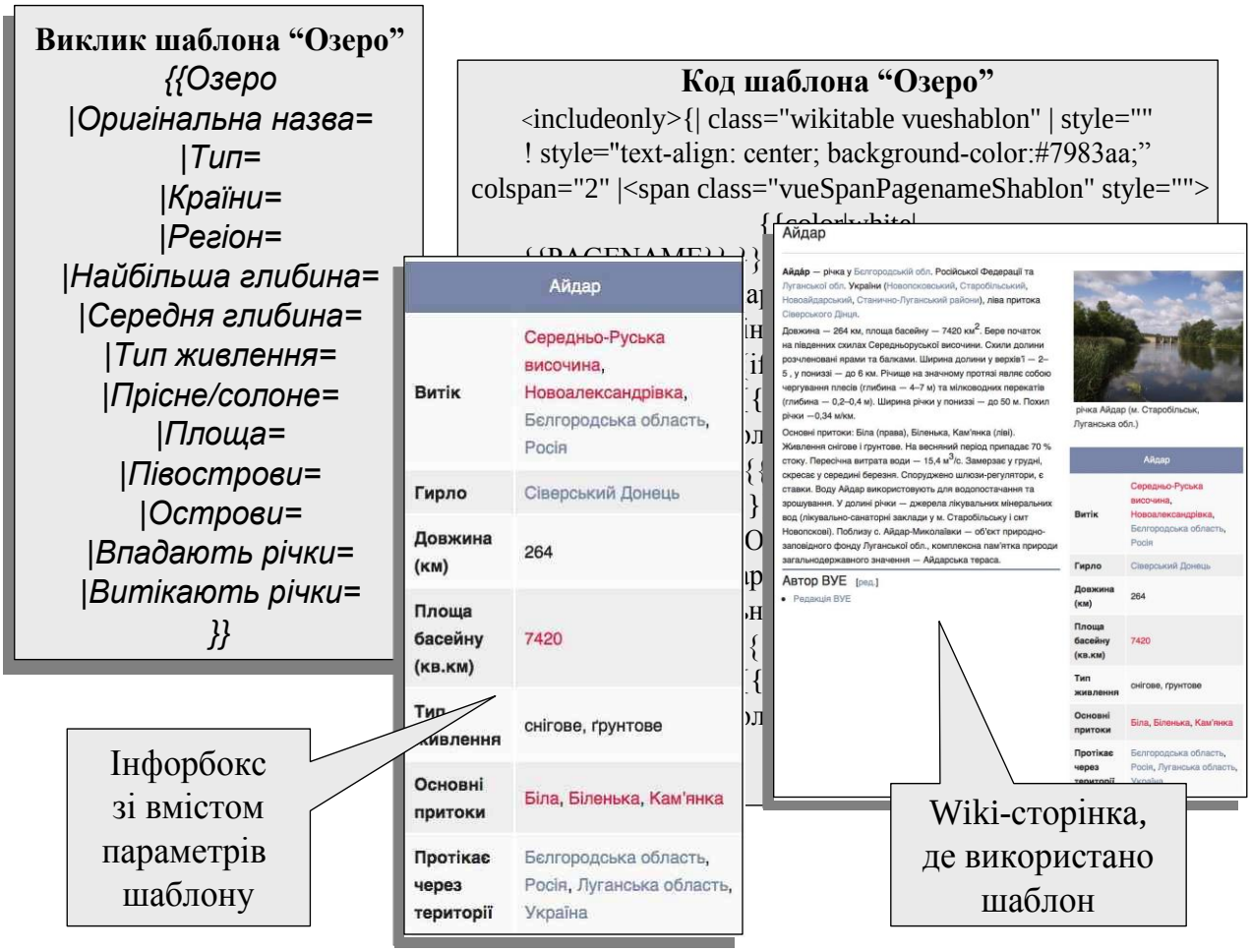


Рис. 2. Шаблони базових ІО в е-ВУЕ

Виразні можливості Semantic MediaWiki дозволяють відображати досить складну систему знань предметної області та відношень між її базовими поняттями, але їх недостатньо для формального подання таких описів, що мали б однозначну інтерпретацію користувачами. Це викликає доцільність застосування більш виразної онтологічної моделі, яка забезпечує вищий рівень формалізації знань.

*Wiki-онтологія e-BYE*

Wiki-онтологія e-BYE як модель знань цієї предметної області дозволяє коректно відображати його базові закономірності та обмеження. Використання Wiki-онтології як основи семантичної розмітки підтримує формування відповідного набору ієрархічно пов'язаних категорій, шаблонів типових інформаційних об'єктів, їх семантичних властивостей та запитів, що їх використовують.

Використання Wiki-онтології e-BYE є засобом явного визначення семантики типових ІО Wiki-ресурсу – їх характеристик та відношень з іншими ІО. Важливо, що таке онтологічне подання дозволяє виявляти та вирішувати неоднозначні інтерпретації та некоректне використання термінів, пов'язаних з описом ІО. Наприклад, це дозволяє розв'язати проблему уніфікації назв семантичних властивостей та категорій, які використовують різні розробники.

Крім того, онтологічне подання спрощує сприйняття знань користувачами, наприклад, дозволяє візуалізувати ієрархічні відношення між категоріями $r_{ier\_categor}$, описувати їх характеристики та надавати анотації. Для складної структури знань, що характерна для e-BYE, це дозволяє значно чіткіше описувати знання та запобігати повторного використання імен категорій з різними значеннями. Редактор онтологій Protégé забезпечує значно ширший функціонал для подання та візуалізації знань порівняно із вбудованими можливостями Semantic MediaWiki (рис. 3).

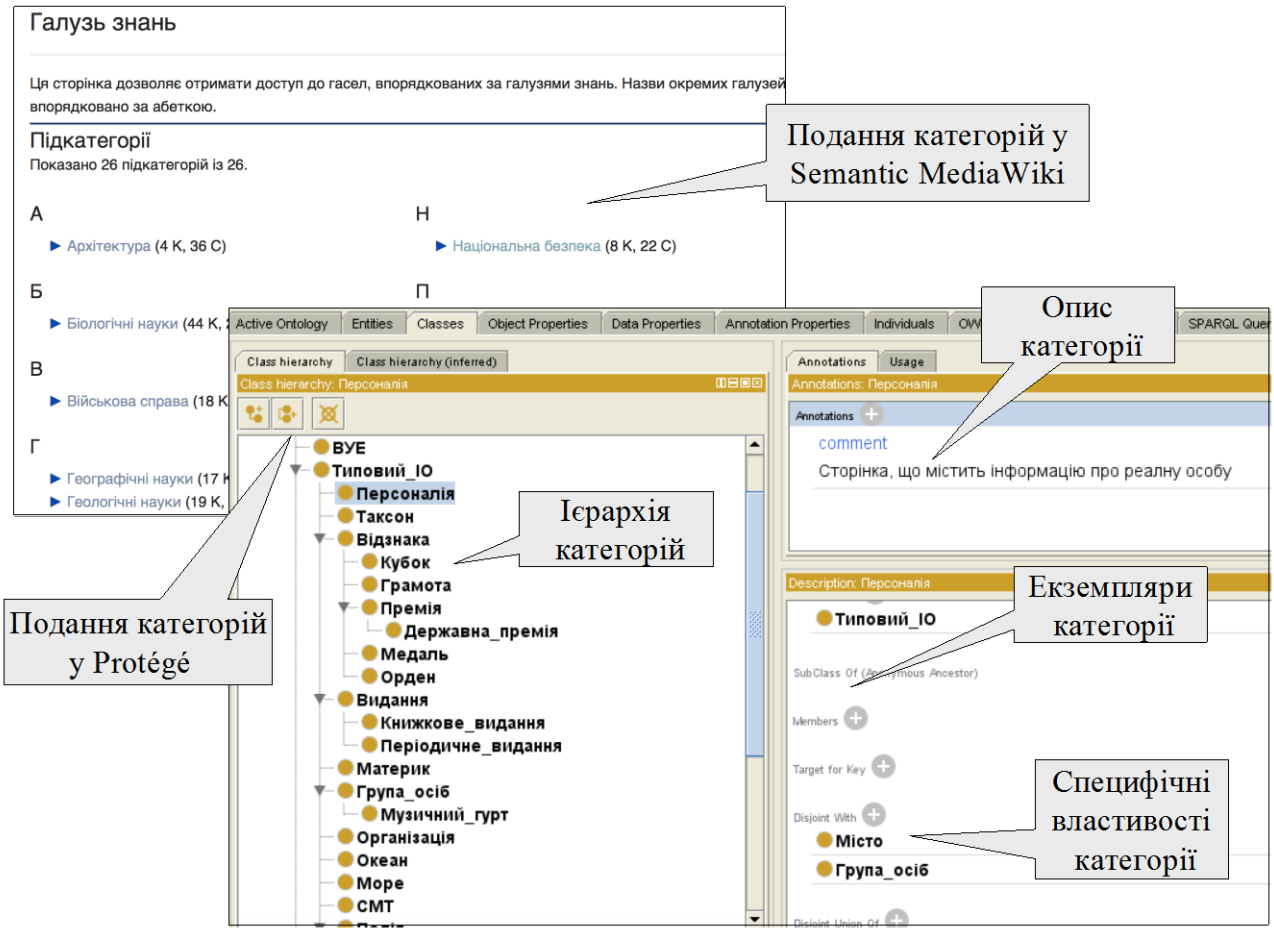


Рис. 3. Класи Wiki-онтології e-BYE в Protégé

Семантичні властивості в е-ВУЕ описуються у Wiki-онтології за допомогою об'єктних властивостей (для значень типу "Сторінка") та властивостей даних (для значень інших типів.

Особливо важливо формально описати семантичні властивості, що пов'язують Wiki-сторінки (в онтології для цього застосовуються об'єктні властивості). Наявність такого опису дозволяє полегшити семантичну розмітку природномовного тексту та перетворення НСД на структуровану базу знань. Використання онтологічної моделі дозволяє чітко визначити область значення та область визначення семантичних властивостей, які характеризують Wiki-сторінку, що пов'язана з певним типом інформаційного об'єкта за допомогою шаблону. Така формалізація семантики сторінок дозволяє запобігти некоректних посилань. Слід відмітити, що в Protege можна формально зафіксувати характеристики властивостей, щоб використовувати їх в е-ВУЕ тільки відповідним чином (рис. 4).

Редактор онтологій дозволяє явно вказати, що певна властивість є симетричною чи транзитивною. В е-ВУЕ такі характеристики має, наприклад, семантична властивість "Співпраця". Крім того, в Protege як об'єктні властивості, так і властивості даних створюються як підкатегорії інших властивостей, і ця ієрархія візуалізується у простій та зрозумілій формі. У середовищі Semantic MediaWiki можна також вказувати ієрархічні відношення між властивостями – за допомогою тверджень [[Subproperty of::dc:date]] на сторінці властивості, але відстежувати ці взаємини має сам користувач вручну. За наявності великої кількості як властивостей, так і користувачів, що мають право їх створювати (а саме така ситуація характерна для е-ВУЕ) це може призвести до некоректності у структурі бази знань енциклопедії.

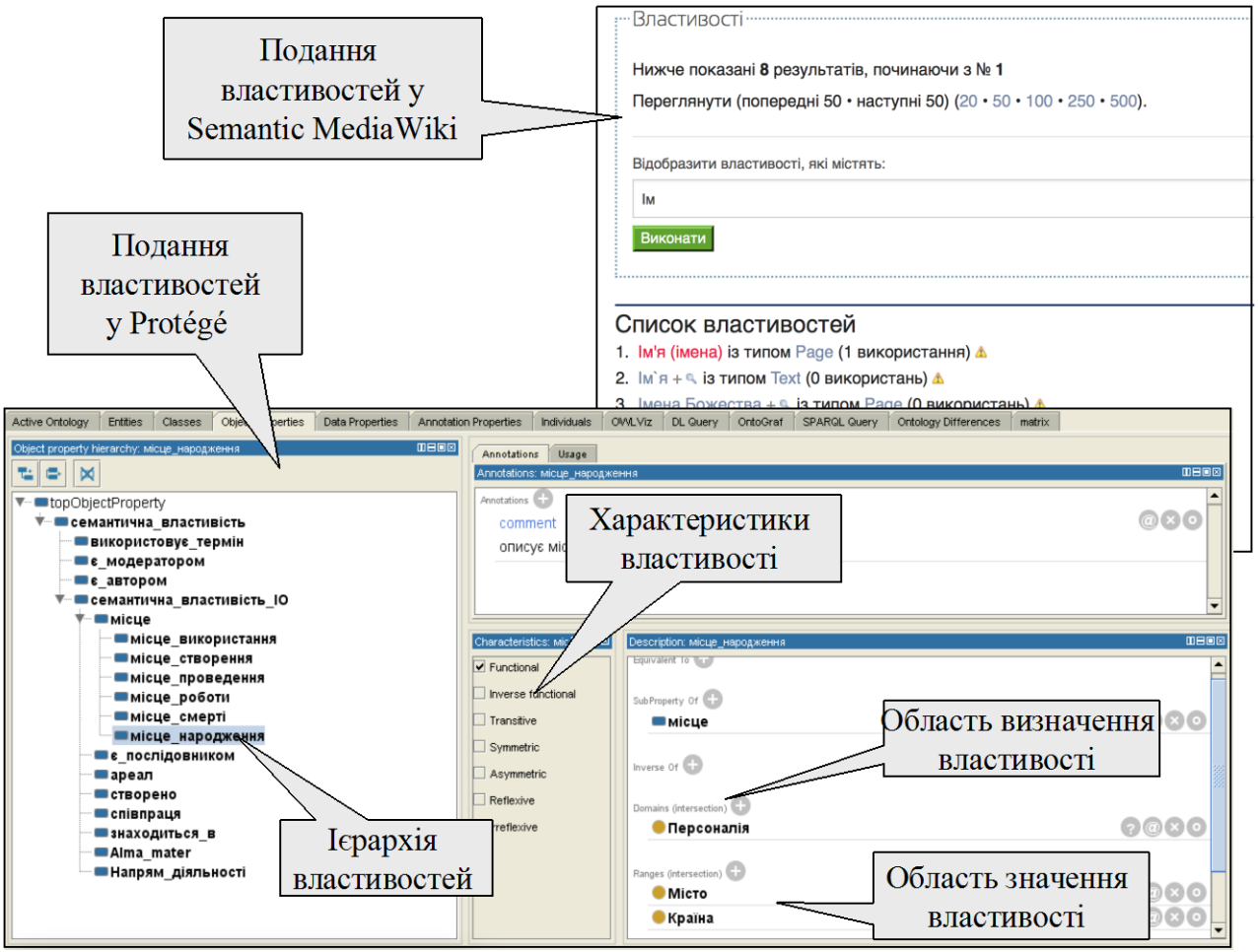


Рис. 4. Створення властивостей Wiki-онтології е-ВУЕ в Protégé

Важливо відмітити, що засоби Semantic MediaWiki дозволяють вказати тип властивості, поділяючи їх на об'єктні властивості, що посилаються на інші сторінки, та на властивості даних, що посилаються на конкретні значення (текст, ціле, дата тощо). Але це не дозволяє визначати тип інформаційного об'єкта, на який таке посилання є релевантним. Наприклад, за змістом "Місце народження" може пов'язувати сторінки категорії "Персоналія" зі сторінками категорій "Місто", "Країна", "Регіон", але формально визначити це засоби Semantic MediaWiki не дозволяють (можна тільки явно вказати припустимі значення, а цього недостатньо).

Wiki-онтологія дозволяє формалізувати набір властивостей кожної Wiki-сторінки як екземпляра відповідного класу.

Коли структура бази знань ресурсу містить багато різних типів інформаційних об'єктів з різними наборами властивостей, що перетинаються, але мають різні області значення та визначення, велику роль грає наявність засобів візуалізації цієї інформації у зручній для користувача формі. На жаль, середовище Semantic MediaWiki не може розглядатися як повноцінна система менеджменту розподілених знань: незважаючи на достатню для широкого класу задач виразність, вона не містить достатніх засобів для аналізу та візуалізації таких знань (інформація видається у формі абеткового переліку, через набори властивостей окремих Wiki-сторінок або як результати запитів тощо, але її важко сприймати у такому поданні).

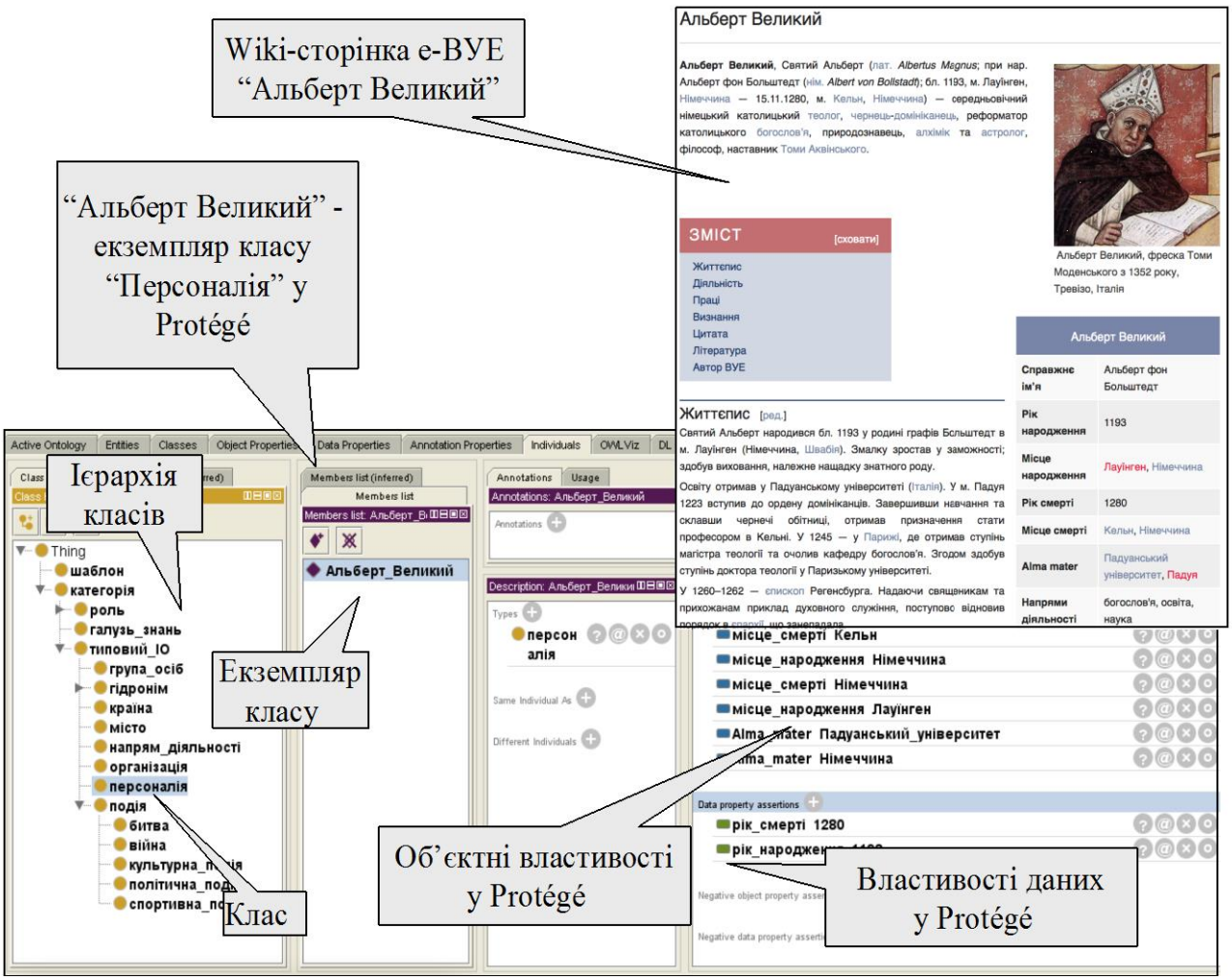


Рис. 5. Подання знань щодо ІО у e-BYE та у Wiki-онтології

## Висновки

Проаналізовано доцільність застосування Wiki-технологій для створення складних інформаційних ресурсів великого обсягу та складної структури, орієнтованих на функціювання у відкритому середовищі Web, до яких відноситься портальна версія Великої української енциклопедії.

Обгрунтована потреба семантичного розширення Wiki-технології, а саме – використання Semantic MediaWiki, для отримання необхідної виразної потужності для подання семантики основних елементів е-ВУЕ, їх властивостей та зв'язків між ними. Розглянуто зв'язок між поданням знань у Semantic MediaWiki та стандартами Semantic Web, які використовуються для інтеграції інтелектуальних застосунків та баз знань у Web.

Доведено на прикладах доцільність застосування онтологічного аналізу та побудови онтологічної моделі е-ВУЕ для формального та однозначного відображення та аналізу структури та контенту бази знань онлайн-версії «Великої української енциклопедії».

Запропоновано методи та засоби застосування цієї онтологічної моделі для створення контенту е-ВУЕ та більш ефективного здійснення пошуку та навігації у цьому інформаційному ресурсі.

## Література


1. Breslin J., Passant A., Deker S. The Social Semantic Web, Springer. 2009. http://staff.um.edu.mt/cabe2/lectures/webscience/docs/S2N.pdf.
2. Hagedorn, G., Weber, G., Plank, A. Giurgiu, M. Homodi, A., C. Veja An online authoring and publishing platform for field guides and identification tools. Proc. of Conf. Tools for Identifying Biodiversity: Progress and Problems, Paris, September 2010. P. 13–18.
3. Leuf B., Cunningham W. The Wiki Way Quick Collaboration on the Web. 2001. https://archive.org/details/isbn_9780201714999.
4. Рогушина Ю. В. Засоби та методи аналізу неструктурованих даних. *Проблеми програмування.* 2019. № 1. С. 57–77. http://pp.isofts.kiev.ua/ojs1/article/view/348/346.
5. Veja C.F., Vaida M.-F., Hagedorn G. Sharing the knowledge: semantic mediawiki. *Acta Technica Napocensis*. 2011. 52(2), N.2. http://users.utcluj.ro/~atn/papers/ATN_2_2011_2.pdf.
6. Рогушина Ю.В., Прийма С.М., Строкань О.В. Створення та використання семантичних Wiki-ресурсів: навчальний довідник. Мелітополь, ФОП Однорог Т.В. 2017. 169 с.
7. Kuhn T. Acewiki: A natural and expressive semantic wiki. 2008. https://arxiv.org/pdf/0807.4618.pdf.
8. Schaffert S., Eder J., Grünwald S., Kurz T., Radulescu M. KiWi – a platform for semantic social software. European Semantic Web Conference. 2009. P. 888–892. https://link.springer.com/content/pdf/10.1007/978-3-642-02121-3_76.pdf.). Springer, Berlin, Heidelberg.
9. Kawamoto K., Kitamura Y., Tijerino Y. Kawawiki: A semantic wiki based on rdf templates. IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology Workshops. 2006. P. 425–432. https://ist.ksc.kwansei.ac.jp/~kitamura/papers/src/iwi06.pdf.
10. Auer S., Dietzold S., Riechert T. OntoWiki–a tool for social, semantic collaboration. International Semantic Web Conference. 2006. P. 736–749. https://link.springer.com/content/pdf/10.1007/11926078_53.pdf.
11. Krötzsch M., Vrandečić D., Völkel M. Semantic mediawiki. International semantic web conference. 2006. P. 935–942. https://link.springer.com/content/pdf/10.1007/11926078_68.pdf.
12. Kousetti Ch., Millard I., Yvonne H. A Study of Ontology Convergence in a Semantic Wiki. Based Infrastructure. University of Southampton. UK, 2008.
13. Bizer C., Lehmann J., Kobilarov G., Auer S., Becker C., Cyganiak R., Hellmann S. Dbpedia – A crystallization point for the Web of Data. Web Semantics: science, services and agents on the world wide web. 2009. 7(3). P. 154–165. https://gesispanel.gesis.org/preprints/index.php/ps/article/download/164/162.
14. Auer S., Lehmann J. What have innsbruck and leipzig in common? extracting semantics from wiki content. European semantic web conference. 2007. P. 503–517. https://link.springer.com/content/pdf/10.1007/978-3-540-72667-8_36.pdf.
15. Rogushina J. Semantic Wiki resources and their use for the construction of personalized

ontologies. CEUR Workshop Proceedings 1631. 2016. P. 188–195.
16. Рогушина Ю.В. Теоретичні засади застосування онтологій для семантизації ресурсів Web. *Проблеми програмування*. 2018. № 2-3. С. 197–203.

## References

1. Breslin J., Passant A., Deker S. The Social Semantic Web, Springer, 2009. http://staff.um.edu.mt/cabe2/lectures/webscience/docs/S2N.pdf.
2. Hagedorn, G., Weber, G., Plank, A. Giurgiu, M. Homodi, A., C. Veja An online authoring and publishing platform for field guides and identification tools // Proc. of Conf. Tools for Identifying Biodiversity: Progress and Problems, Paris, September 2010. P. 13–18.
3. Leuf B., Cunningham W. The Wiki Way Quick Collaboration on the Web, 2001. https://archive.org/details/isbn_9780201714999.
4. Rogushina Y.V. Tools and methods of unstructured data analysis // Problems in Programming, № 1, 2019. P. 57–77. http://pp.isofts.kiev.ua/ojs1/article/view/348/346. [in Ukrainian]
5. Veja C.F., Vaida M.-F., Hagedorn G. Sharing the knowledge: semantic mediawiki // Acta Technica Napocensis, 52(2), N. 2, 2011. – http://users.utcluj.ro/~atn/papers/ATN_2_2011_2.pdf.
6. Rogushina Y.V., Priyma S.M, Strokan O.V. Creating and use of the Semantic Wiki resources: tutorial. Melitopol, FOP Odinorog T.V., 2017. 169 p. [in Ukrainian]
7. Kuhn T. Acewiki: A natural and expressive semantic wiki. 2008. https://arxiv.org/pdf/0807.4618.pdf.
8. Schaffert S., Eder J., Grünwald S., Kurz T., Radulescu M. KiWi – a platform for semantic social software. European Semantic Web Conference, 2009. P. 888–892. https://link.springer.com/content/pdf/10.1007/978-3-642-02121-3_76.pdf.). Springer, Berlin, Heidelberg.
9. Kawamoto K., Kitamura Y., Tijerino Y. Kawawiki: A semantic wiki based on rdf templates. IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology Workshops, 2006, P. 425–432. https://ist.ksc.kwansei.ac.jp/~kitamura/papers/src/iwi06.pdf.
10. Auer S., Dietzold S., Riechert T. OntoWiki–a tool for social, semantic collaboration // International Semantic Web Conference, 2006, P. 736–749. https://link.springer.com/content/pdf/10.1007/11926078_53.pdf.
11. Krötzsch M., Vrandečić D., Völkel M. Semantic mediawiki. International semantic web conference. 2006. P. 935–942. https://link.springer.com/content/pdf/10.1007/11926078_68.pdf.
12. Kousetti Ch., Millard I., Yvonne H. A Study of Ontology Convergence in a Semantic Wiki. Based Infrastructure. University of Southampton. UK, 2008.
13. Bizer C., Lehmann J., Kobilarov G., Auer S., Becker C., Cyganiak R., Hellmann S. Dbpedia – A crystallization point for the Web of Data. Web Semantics: science, services and agents on the world wide web. 2009. 7(3). P. 154–165. https://gesispanel.gesis.org/preprints/index.php/ps/article/download/164/162.
14. Auer S., Lehmann J. What have innsbruck and leipzig in common? extracting semantics from wiki content. European semantic web conference. 2007. P. 503–517. https://link.springer.com/content/pdf/10.1007/978-3-540-72667-8_36.pdf.
15. Rogushina J. Semantic Wiki resources and their use for the construction of personalized ontologies. CEUR Workshop Proceedings 1631. 2016. P. 188–195.
16. Rogushina Y.V. Theoretical means of use of ontologies for semanization of the Web resources. Problems in Programming. № 2-3. 2018. P. 197–203. [in Ukrainian]

Одержано 22.04.2019

***Про автора:***

*Рогушина Юлія Віталіївна,*
кандидат фізико-математичних наук,
старший науковий співробітник.
Кількість наукових публікацій в українських виданнях – 140.
Кількість наукових публікацій в зарубіжних виданнях – 30.
Індекс Хірша – 3.
http://orcid.org/0000-0001-7958-2557.

***Місце роботи автора:***

Інститут програмних систем
НАН України,
03181, Київ-187,
проспект Академіка Глушкова, 40.
Тел.: 066 550 1999.
E-mail: ladamandraka2010@gmail.com