

*Ю.В. Рогушина*

## СЕМАНТИЗАЦІЯ ВІКІРЕСУРСІВ: ЗАСОБИ, ПЕРЕВАГИ ТА ОСОБЛИВОСТІ РЕАЛІЗАЦІЇ

У роботі аналізуються основні переваги семантизації вікіресурсів та проблеми, які потрібно розв'язувати для її практичного впровадження. Розглянуто базові складові семантичного пошуку та конкретизовано цей парасольковий термін для вікітехнологій. Визначено засоби побудови структурованих запитів, що використовують семантичні властивості інформаційних об'єктів, що представлені у вікіресурсах. Проаналізовано критерії оцінювання пошукових мов та засобів семантичного структурування контенту ресурсів. Визначено, які саме додаткові можливості дає використання семантичного плагіну Semantic MediaWiki для ресурсів, що будуються на технологічній платформі MediaWiki.

Розглянуто виразні можливості, що дає Semantic MediaWiki для побудови семантичних запитів, визначено основні компоненти таких запитів та їх призначення.

Семантизація вже існуючих вікіресурсів значно відрізняється від розробки тих, що відразу розробляються як семантичні, і тому ми розглядаємо відмінності у порядку їх розробки. Обґрунтовано доцільність застосування онтологічної моделі бази знань семантизованого вікіресурсу, яка забезпечує однозначну інтерпретацію відношень між типовими інформаційними об'єктами, представленими у ресурсі, та їхніх властивостей. Запропоновані підходи до семантизації апробовано на трьох незалежних інформаційних ресурсах різних типів, які використовують технологічну платформу вікі для колаборативної обробки розподілених даних та знань.

У роботі використано досвід розробки бази знань порталної версії Великої української енциклопедії e-VUE, яка має великий обсяг, складну структуру та містить велику кількість різноманітних гетерогенних інформаційних об'єктів. Запропонований підхід може бути корисним для ухвалення рішень щодо доцільності семантизації інформаційних ресурсів різного обсягу й призначення та для визначення найбільш ефективних шляхів реалізації обраного рішення.

Ключові слова: вікітехнології, семантичний пошук, онтологія.

*J. Rogushina*

## SEMANTIZATION OF WIKI RESOURCES: TOOLS, ADVANTAGES AND IMPLEMENTATION SPECIFICS

Proposed article analyses advantages provided by the semantization of wiki resources and the problems that need in solving for its practical implementation. We considere basic functional components of semantic search and specify this umbrella term for wiki technologies. Such search is grounded on structured queries that use semantic properties of information objects represented into wiki resources and possibilities of search language that states such queries. The criteria for evaluating search languages and means of semantic structuring of resource content are analyzed.

We define additional opportunities that the use of the Semantic MediaWiki plug-in provides for the oresources built on the MediaWiki technological platform. The expressive possibilities provided by Semantic MediaWiki for building semantic queries are considered, the main components of such queries and their purpose are determined.

Semantization of already existing wiki resources differs from the development of semantic ones, and we compare main steps of these processes and specifics of use the ontological model in them. This model provides an unambiguous interpretation of the relations between typical information objects represented into the resource, their properties and restrictions. Proposed approaches to semantization are tested on three independent information resources of different types that use the wiki technological platform for collaborative processing of distributed data and knowledge. They can be useful for making decisions about the expediency of semantization of information resources with different scope and purposes and for determining the most effective ways of implementing the chosen solution. We use the experience of developing the knowledge base of the wiki-based portal version of the Great Ukrainian Encyclopedia e-VUE that has a large volume, a complex structure and contains a large number of various heterogeneous information objects.

Keywords: wiki technologies, semantic search, ontology.

## Вступ

Наразі існує велика кількість інформаційних ресурсів (IP), які базуються на вікітехнологіях, що обумовлюється їх відносною простотою для кінцевого користувача, підтримкою колаборативної роботи з контентом та вже перевіреною здатністю до подання великих обсягів інформації. Існує велика кількість програмних рішень для їх семантизації [1, 2], які забезпечують додаткові засоби для пошуку, перегляду та представлення контенту, але вибір конкретних моделей та інструментів залежить від призначення IP та від потреб у підтримці семантичного пошуку.

### 1. Семантичний пошук та його характеристики

*Семантичний пошук* (СП) – парасольковий термін, який використовують для позначення цілої групи моделей та методів використання зовнішніх джерел знань, які у різний спосіб покращують традиційні підходи до пошуку в документах та їх наборах, використовуючи контекст і семантику як запиту користувача, так і тих ресурсів, в яких здійснюється пошук. Визначаючи доцільність його застосування, потрібно конкретизувати, які саме його елементи потрібні у конкретній реалізації. На виразність пошуку впливають засоби опису запиту користувача, що відображає його інформаційну потребу; засоби опису та структурування масиву даних, в яких здійснюється пошук, та способи подання користувачу результатів виконання пошукової процедури [3]. Структуровані запити дозволяють користувачам формалізувати вимоги до тих об'єктів, та визначити логічні відношення між його елементами та властивостями. Структуровані семантичні запити на основі онтологій можуть використовувати такі елементи онтології, як поняття класи та екземпляри класів і їхні властивості. На виразність запитів впливає виразність елементів семантичної розмітки, яку вони аналізують (можливість опису класів, екземплярів, властивостей тощо) [4]. Також доцільно брати до уваги складність використання мови розмітки та зручність її за-

стосування, наявність інструментів для її введення та аналізу.

### 2. Постановка задачі

Мета даного дослідження – визначити переваги виконання пошуку та навігації у семантизованих вікіресурсах та шляхи її реалізації. Для цього досліджуються складові семантичного пошуку та їх представлення у Semantic MediaWiki (SMW) – семантичному розширенні MediaWiki.

### 3. Семантичні вікіресурси та онтології

Семантичний пошук в SMW працює на основі явного структурування контенту довільним набором тегів розмітки, які можуть бути здобуті з онтології релевантної предметної області (PrO). Формальна семантика структурованих даних у SMW може бути відображена в онтології OWL [5]. Звичайні вікісторінки статей відповідають екземплярам класів онтології OWL, категорії відповідають класам, а семантичні властивості SMW відповідають властивостям: властивості SMW зі значеннями типу “Сторінка” зіставляються з об'єктними властивостями онтології, а властивості з іншими типами даних зіставляються з онтологічними властивостями даних. Відповідно значеннями властивостей можуть бути екземплярами онтології або константами. Категорії вікісторінок визначають їхній клас в OWL. MediaWiki підтримує ієрархічну організацію категорій, і SMW можна налаштувати так, щоб інтерпретувати це як ієрархію класів OWL.

SMW підтримує мову ASK для виконання структурованих запитів [6], які можна вбудовувати у шаблони або безпосередньо виконувати на спеціальній вікісторінці «Спеціальна:Ask». Найчастіше для побудови вбудованих запитів використовують функція аналізатора #ask та #show. Параметри пошуку дозволяють визначити: критерії вибору сторінок, які мають відображатися в результатах запиту; порядок подання результатів та набір властивостей, значення яких потрібно відобразити; кіль-

кість результатів, які потрібно надати користувачам; формати подання результатів. Основна перевага SMW – можливість обирати за значеннями властивостей сторінок та використовувати набір умов в одному запиті (на відміну, наприклад, від пошуку за категоріями у Вікіпедії). У запитах можна використовувати так звані магічні слова MediaWiki, такі як змінні, що повертають відомості про поточну сторінку, час і середовище, та функції аналізатора, що надають відомості про довільні сторінки. Запити у SMW дозволяють відображати не весь контент сторінок, а лише потрібну користувачеві інформацію щодо значень обраного набору властивостей.

#### 4. Переваги семантизованого вікіресурсу

Наведений вище аналіз базових можливостей SMW дозволяє виділити наступні переваги, що дає застосування цього плагіну (рис.1):

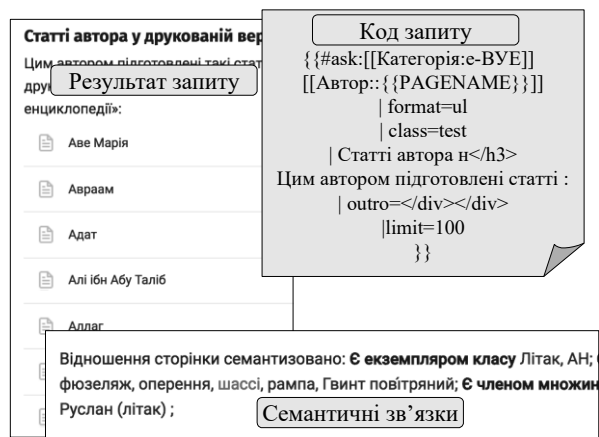


Рис.1. Результати використання можливостей SMW

*Можливість явно визначити зміст посилань між вікісторінками:* корисно як для автоматизованої обробки, так і для розуміння інформації користувачами.

*Підтримка запитів за довільними комбінаціями категорій та значень семантичних властивостей:* виразність таких запитів значно більша за виразність пошуку за категоріями або ключовими словами, і це значно зменшує час для доступу до потрібних відомостей.

*Можливість автоматизованої генерації контенту вікісторінок за вбудова-*

*ними запитами:* зменшується час для введення контенту, забезпечується його інтегрованість та узгодженість.

*Можливість застосування параметрів шаблонів для створення семантичної розмітки:* SMW інтерпретує значення параметрів шаблонів як значення відповідних семантичних властивостей тієї сторінки, з якої викликався шаблон, і це значно спрощує внесення семантичної розмітки ресурсу.

Якщо серед цих переваг є такі, що можуть стати пріоритетними для розробників вікіресурсу, то тоді семантизація є доцільною. Але потрібно брати до уваги, що семантизація ресурсу потребує додаткових зусиль від розробників вікіресурсу та передбачає наявність у них додаткових компетенцій. Наприклад, для того, щоб вбудовувати семантичні запити у сторінки, потрібно знати синтаксис мови запитів або вміти користуватися інструментами для створення запитів. А для того, щоб вказувати, які саме властивості сторінок треба аналізувати у запитах, необхідно знати структуру бази знань. Тому для забезпечення переваг семантизації необхідно виконувати наступні дії:

- створювати сторінки семантичних властивостей та визначати їх тип, що має бути релевантним значенням, які вводяться, а також описувати їхню семантику та сферу застосування;
- слідкувати, щоб не створювалися різні семантичні властивості з однаковими чи подібними назвами, а також однакові за змістом властивості з різними назвами;
- відстежувати поточний набір семантичних властивостей, розуміти їх значення та знати правила створення семантичної розмітки, а також аналізувати пертинентність їх застосування до конкретних елементів контенту (ці дії не можуть бути повністю автоматизовані через неоднозначність природної мови);

- відстежувати наявність відповідних шаблонів для опису типових інформаційних об'єктів (ТІО), їхню поточну структуру та типи параметрів, знати правила їх виклику на вікісторінках та обмеження щодо використання форматувань у значеннях шаблонів;
- слідкувати за необхідністю створювати нові шаблони для опису ТІО та перевіряти, щоб їхні параметри не конфліктували з наявним набором семантичних властивостей;
- забезпечувати коректне адміністрування для індексації бази даних ресурсу, тому що несвоєчасна індексація може призвести до невідповідності схеми бази даних (набору семантичних властивостей) та контенту сторінок, що викликає помилки у виконанні запитів;
- забезпечити створення вбудованих семантичних запитів та аналізувати доцільність додавання результатів пошуку до вікісторінок.

Потрібно брати до уваги, що складність семантизації зростає нелінійно зі збільшенням кількості семантичних властивостей та шаблонів, які їх використовують. Збільшення кількості звичайних вікісторінок впливає на це лише лінійно – семантизація кожної окремої сторінки займає приблизно однаковий час (цей параметр незначно збільшується через довший пошук коректних посилань на інші вікісторінки у більш довгому їх переліку).

Якщо ресурс містить велику кількість різних інформаційних об'єктів (ІО), для опису яких використовується багато семантичних властивостей, то його база знань потребує формалізації, – наприклад, за допомогою онтології ПрО, яка фіксує семантику зв'язків між типами ІО та їхніми шаблонами, семантичними властивостями цих об'єктів, їх категоріями тощо [7] (рис.2). Онтологічне подання забезпечує

однозначну інтерпретацію цих відомостей, а наявність загальноприйнятих форматів (OWL, RDF) та зручних інструментів для роботи з ними (таких як Protégé [8]) спрощує взаємодію з користувачами та іншими системами, а також підтримує повторне використання інформації.



Рис.2. Елементи онтологічної моделі семантичного вікіресурсу

Візуалізація онтології (рис.3) допоможе користувачам додавати семантичну розмітку та заповнювати шаблони, але необхідно підтримувати синхронізацію онтології з поточними змінами в його базі знань.

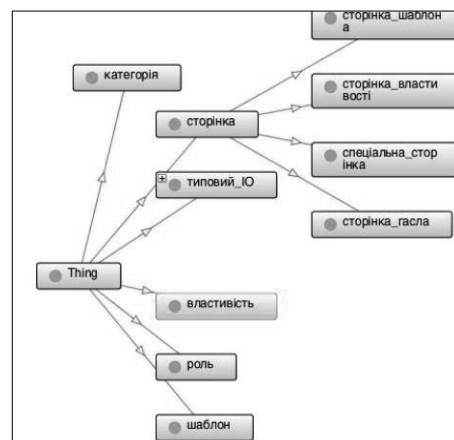


Рис.3. Візуалізація онтологічної моделі семантичного вікіресурсу

Важливо розрізняти дії, що виконуються у тому випадку, коли ресурс відразу розробляється як семантичний, від дії, необхідної семантизації вже існуючого вікіресурсу, для якого створено контент великої кількості сторінок.

У першому випадку порядок виконання дій такий:

- створюється узагальнена онтологічна модель бази знань ресурсу;
- визначаються типові інформаційні об'єкти ресурсу та їхні властивості;
- визначаються типи цих властивостей, їх можливі значення, припустимість множинності та невизначеності;
- створюються сторінки відповідних семантичних властивостей;
- семантичні властивості використовуються для розмітки контенту вікісторінок – як-от, для визначення семантики посилань на інші сторінки;
- визначається структура шаблонів для типових інформаційних об'єктів та форма представлення інформації про них користувачам;
- розробляються та тестуються шаблони типових об'єктів;
- ці шаблони викликаються на сторінках екземплярів відповідних типових об'єктів та заповнюються конкретними значеннями;
- розробляються семантичні запити, що здобувають відомості з контенту таких сторінок;
- генерується уточнена онтологічна модель бази знань ресурсу, яка заповнюється екземплярами шаблонів та семантичних властивостей.

У другому випадку будемо вважати, що на момент прийняття рішення щодо семантизації у вікіресурсі вже визначені певні групи типових інформаційних об'єктів, що об'єднані у категорії, і для подання цих об'єктів уже використовуються відповідні шаблони. Крім того, між сторінками вже існують посилання без визначення їхньої семантики. Тому процес семантизації має містити наступні дії:

- встановити плагін Semantic MediaWiki (за потребою – інші допоміжні семантичні плагіни);
- проаналізувати зміст посилань між вікісторінками, і у разі ви-

явлення достатньої (залежно від загального обсягу ресурсу та вимог розробників) кількості посилань з однаковою чи близькою семантикою, то створити семантичну властивість типу “Сторінка” з релевантною назвою, а відповідні анонімні посилання між сторінками замінити на семантичні;

- проаналізувати існуючі шаблони подання типових об'єктів та їхні параметри, створити семантичні властивості відповідних типів (слід враховувати, що за замовчанням після встановлення SMW усі параметри шаблонів інтерпретуються як властивості типу “Сторінка”, і це може призвести до некоректного відображення параметрів інших типів);
- протестувати шаблони у новому середовищі, за потреби внести зміни. Наприклад, для значень параметрів, що мають бути посиланнями на інші сторінки, не варто використовувати подвійні дужки, достатньо правильно визначити тип властивості;
- перевірити, чи не перетинаються некоректно назви семантичних властивостей різних типів (для імен параметрів у несемантизованому вікіресурсі це було несуттєво);
- створити потрібні семантичні запити, додати їх до відповідних сторінок та перевірити коректність їх виконання;
- зафіксувати побудовану структуру бази знань ресурсу у вигляді онтологічної моделі.

Якщо значна частина вікісторінок несемантизованого ресурсу містила шаблони, призначені для уніфікованого представлення екземплярів типових інформаційних об'єктів, то це значно полегшує процес семантизації – самі сторінки майже не потребують внесення додаткової семантичної розмітки – необхідно додавати се-

мантичні теги такої розмітки тільки до посилає на інші сторінки.

Незалежно від способу додавання семантичної складової до вікіресурсу, створення різноманітних сторінок-інтеграторів може продовжуватися відповідно до потреб користувачів на основі знань, збережених в онтологічній моделі. Основна перевага наявності такої моделі полягає в тому, що вона дозволяє отримувати відомості щодо бази знань ресурсу та семантики його елементів без безпосереднього контакту з її розробниками. Наприклад, для побудови вбудованих запитів потрібно не тільки знати коректні імена семантичних властивостей певної категорії сторінок, а й розуміти їхній зміст та правила введення значень.

Важливо враховувати, що створення семантичних властивостей та їх використання мають бути оброблені у базі даних вікіресурсу, і тому у результатах семантичних запитів наслідки семантизації будуть відображатися не відразу, а тільки після їх повної індексації, швидкість якої залежить від довжини черги задач та від обраної політики, і тому проміжні результати можуть бути некоректними.

### Висновки

Семантизація вікіресурсів потребує використання методів керування розподіленими знаннями та елементів онтологічного аналізу для моделювання Про, що залежать від цілей семантизації та стану ресурсу на момент її початку. Вибір пертинентної моделі та її коректна програмна реалізація забезпечують не тільки зручну навігацію у ресурсі, а й більш складні пошукові та аналітичні функції. Наведені відмінності у способах семантизації вікіресурсів були апробовані на трьох незалежних прикладах – порталі Великої української енциклопедії е-ВУЕ (vue.gov.ua), тестовій версії Української електронної енциклопедії освіти УЕЕО (uee.gs4cms.com.ua) та довідковому вікіресурсі Інституту програмних систем НАНУ (<http://wiki.isoftware.kiev.ua/>). Усі три проекти використовують MediaWiki та семантичний плагін Semantic MediaWiki, але на них встановлені різні версії цього програмного

забезпечення, а розробка системи та семантизація контенту виконувалися на різних методологічних засадах. Тому можна вважати, що виявлені закономірності характерні якщо не для всіх, то для багатьох вікіресурсів, що створюються у такому технологічному середовищі.

### Література

1. D. Vrandečić, M. Krötzsch, Semantic MediaWiki, *Semantic Knowledge Management: Integrating Ontology Management, Knowledge Discovery, and Human Language Technologies*, 2009, pp. 171-179.
2. P. Haase, D. Herzig, M. Musen, T. Tran, Semantic wiki search, *The Semantic Web: Research and Applications: 6th European Semantic Web Conference*, 2009, pp. 445-460.
3. J. Rogushina, V., A three-dimensional model of semantic search: queries, resources, and results. *Problems in programming*, (4), 2023, pp.39-55.
4. Y. Gil, V. Ratnakar, A Comparison of (Semantic) Markup Languages. *FLAIRS*, 2002, pp. 413-418.
5. M. Völkel, M. Krötzsch, D. Vrandečić, H. Haller, R. Studer, Semantic wikipedia. *Proc. of the 15th international conference on World Wide Web*, 2006, pp. 585-594.
6. Semantic search. Accessed: 11.02.2024. [semantic-mediawiki.org/wiki/Help:Semantic\\_search](https://semantic-mediawiki.org/wiki/Help:Semantic_search).
7. Ю.В. Рогущина, А.Я. Гладун, В.В. Осадчий, С.М. Прийма, Онтологічний аналіз у Web. – Мелітополь: МДПУ ім. Богдана Хмельницького, 2015. – 407 с.
8. T. Tudorache, N. F. Noy, S. Tu, M. Musen, Supporting collaborative ontology development in Protégé, *The Semantic Web-ISWC 2008: 7th International Semantic Web Conference*, 2008, pp. 17-32.

### References

1. D. Vrandečić, M. Krötzsch, Semantic MediaWiki, *Semantic Knowledge Management: Integrating Ontology Management, Knowledge Discovery, and Human Language Technologies*, 2009, pp. 171-179.
2. P. Haase, D. Herzig, M. Musen, T. Tran, Semantic wiki search, *The Semantic Web:*

- Research and Applications: 6th European Semantic Web Conference, 2009, 445-460.
3. J. Rogushina, V., A three-dimensional model of semantic search: queries, resources, and results. Problems in programming, (4), 2023, 39-55.
  4. Y. Gil, V. Ratnakar, A Comparison of (Semantic) Markup Languages. FLAIRS, 2002, 413-418.
  5. M. Völkel, M. Krötzsch, D. Vrandečić, H. Haller, R. Studer, Semantic wikipedia. Proc. of the 15th international conference on World Wide Web, 2006, 585-594.
  6. Semantic search. Accessed: 11.02.2024. [semantic-mediawiki.org/wiki/Help:Semantic\\_search](http://semantic-mediawiki.org/wiki/Help:Semantic_search).
  7. J. Rogushina, A. Gladun, V. Osadchii, S. Pryma, Ontological analysis in the Web. – Melitopol: MDPU, 2015. (in Ukrainian).
  8. T. Tudorache, N. F. Noy, S. Tu, M. Musen, Supporting collaborative ontology development in Protégé, The Semantic Web-ISWC 2008: 7th International Semantic Web Conference, 2008, 17-32.

Одержано: 28.03.2024

Внутрішня рецензія отримана: 07.04.2024

Зовнішня рецензія отримана: 18.04.2024

***Про автора:***

*Рогущина Юлія Віталіївна,*

Кандидат фізико-математичних наук,

доцент,

старший науковий співробітник

ORCID

<http://orcid.org/0000-0001-7958-2557>.

***Місце роботи автора:***

Інститут програмних систем НАН

України,

тел. (+38) 066 5501999,

e-mail: [ladamandraka2010@gmail.com](mailto:ladamandraka2010@gmail.com)