

ВИКОРИСТАННЯ АКТИВНОСТІ КОРИСТУВАЧА ВІДКРИТИХ WIKI-РЕСУРСІВ ДЛЯ ПОПОВНЕННЯ ЙОГО ОНТОЛОГІЧНОЇ МОДЕЛІ

Ю.В. Рогушина, І.Ю. Гришанова

Предложен подход к построению персонафицированной онтологической модели предметной области, которая интересует пользователя. Эта модель формируется на основе обработки семантически размеченных Wiki-ресурсов, с которыми взаимодействовал пользователь, и использует семантические свойства, ссылки и категории для формирования классов и экземпляров онтологии. Такой подход позволяет обеспечить прикладные системы актуальными формализованными знаниями, представленными в Web.

Ключевые слова: Wiki-ресурс, онтология, RDF, семантический поиск.

Запропоновано підхід до побудови персоніфікованої онтологічної моделі предметної області, яка цікавить користувача. Ця модель формується на основі обробки семантично розмічених Wiki-ресурсів, з якими взаємодіяв користувач, і використовує семантичні властивості, посилання і категорії для формування класів і екземплярів онтології. Такий підхід дозволяє забезпечити прикладні системи актуальними формалізованими знаннями, представленими в Web.

Ключові слова: Wiki-ресурс, онтологія, RDF, семантичний пошук.

An approach to the construction of personalized ontological domain model that interests the user. This model is based on processing of semantically marked Wiki-resource which the user interacted with, and uses semantic properties, links, and search for the ontology classes and instances building. This approach allows us to provide relevant application formalized system of knowledge represented on the Web.

Key words: Wiki-resource, ontology, RDF, semantic search.

Вступ

Проблема пошуку в Web є складником найрізноманітніших інформаційних систем. Сьогоднішні прогнози в галузі інформаційних технологій засвідчують загальну тенденцію: найближчими роками обсяг інформації та потреба в ній збільшуватимуться експоненційно. Тому в процесі інформаційного пошуку (ІП) постає необхідність у переході від обробки великих обсягів даних до обробки знань – більш компактних, проте зі значно складнішою структурою.

Користувачам сучасних інформаційних систем потрібно надати персоніфіковані засоби семантичного пошуку. Вимоги часу в тому, що недостатньо просто знайти у Web певний документ або набір документів – необхідно видобути з цих документів знання саме про ті інформаційні об'єкти, що забезпечать користувачам виконання завдань, і пояснити шляхи та методи генерації саме таких результатів. Це зумовлює потребу в класифікації інформаційних об'єктів, а також засобів пошуку відомостей про них. До того ж, необхідно формалізувати інформаційні потреби користувачів так, щоб позбавити їх рутинних пошукових операцій, надавши, натомість, гнучкі засоби керування процесом семантичного пошуку.

Щоб найточніше сформулювати теорію інтелектуального інформаційного пошуку, необхідно чітко визначити основні компоненти його семантичної моделі, їх властивості та зв'язки між ними. Така модель містить опис користувача, інформаційних ресурсів та інших важливих для пошуку компонентів.

Системи семантичного пошуку

Система семантичного пошуку – це інформаційна система, що забезпечує пошук і розпізнавання інформаційних об'єктів (ІО) різних типів із використанням знань для зіставлення запиту з наявними інформаційними ресурсами (ІР) на семантичному рівні.

Системи семантичного пошуку (ССП) можна розглядати як певну інтелектуальну надбудову над традиційними інформаційно-пошуковими системами (ІПС) як загального призначення, так і спеціалізованими [1]. Можна сказати, що ССП є інтелектуалізацією ІПС. ССП дають змогу обробляти складніші інформаційні об'єкти і використовувати знання, що відповідають предметній області (ІО). Але для того, щоб більш коректно вести мову про інтелектуалізацію ІПС, необхідно формально визначити термін «інтелектуалізація» відносно інформаційних систем (ІС), а також інші терміни, пов'язані з цим процесом.

Відкрита ССП – це ССП, у якій використовуються не тільки внутрішні, а й зовнішні бази знань.

Наприклад, якщо для формалізованого опису сфери інформаційних інтересів користувача ССП використовує посилання на онтологію відповідної предметної сфери, а для опису структури ІО, який потрібно знайти, – посилання на зовнішню таксономію ІО, то така ССП є відкритою. Відкриті ССП орієнтовані на роботу у Web-середовищі й реалізують концепцію відкритого світу.

Використання онтологічного підходу для формування моделі ССП забезпечує інтероперабельне використання накопичених знань [2] – знання можуть імпортуватися в модель з інших інтелектуальних ІС (наприклад, із семантичних Wiki-ресурсів) і експортуватися в інші системи, з якими працюють ті ж користувачі (наприклад, можна переносити відомості про користувача із ССП у персоніфіковану систему дистанційного навчання або в систему оцінки кваліфікації)

© Ю.В. Рогушина, І.Ю. Гришанова, 2016

При семантичному пошуку формується інформаційна модель користувача, у якій відображаються [3]:

- персональні переваги користувача, його знання і спроможність до сприйняття інформації [4];
- інформаційні потреби користувача, особливості задачі, що стоїть перед ним, і тих інформаційних об'єктів, виявлення відомостей про які дозволить йому вирішити цю задачу;
- досвід взаємодії користувача з пошуковою системою;
- відомості, що характеризують відношення користувача до досвіду інших користувачів і визначають доцільність використання колаборативного пошуку.

Ця модель зіставляється з моделями ІР, у яких можуть міститися корисні для користувача відомості [5]. Модель ІР може включати як семантичну розмітку і метадані про ІР, так і оцінки цих ІР різними користувачами. Для семантичної розмітки застосовується онтологічна модель ПрО, що цікавить користувача, тобто така розмітка повинна виконуватися заново для кожної нової ПрО.

Використання онтологічного підходу забезпечує інтероперабельність і повторне використання знань – усі ці моделі можуть динамічно поповнюватися за рахунок інформації з зовнішніх онтологій і аналізу зовнішніх ІР.

Крім того, в онтологічній моделі взаємодії користувачів і ІР присутні такі класи, як *онтологія ПрО*, що описує область, до якої відносяться інформаційні потреби користувача; *лексична онтологія ПрО*, що дозволяє розпізнавати відомості про терміни онтології в природно-мовних текстах; *тезаурус задачі* – терміни онтології, сукупність яких характеризує ту конкретну задачу з ПрО, що у даний момент вирішує користувач, і їхня вага; *інформаційна потреба* користувача, для задоволення якої і виробляється пошук інформації; *запит* – явно передане користувачем повідомлення про наявність інформаційної потреби (як правило, це множина ключових слів, що характеризують одну з інформаційних потреб користувача, зв'язаних з конкретною задачею за допомогою тезауруса); *тема* – множина запитів, зв'язаних з однією інформаційною потребою різних користувачів, що дозволяє поєднувати семантично зв'язані запити; *результат запиту* – посилання на ІР і їхні оцінки; *інформаційний ресурс*; *рекомендація* – відомості, надані користувачу проактивно, унаслідок дій інших користувачів, що беруть участь у колаборативному пошуку; *елемент*, що рекомендується – посилання на ІР, що входить у рекомендацію; *інформаційне середовище* – сукупність усіх доступних ІР, їхніх властивостей (включаючи оцінки їх користувачами) і зв'язків між ними; і *група користувачів*.

Для того, щоб формалізувати опис елементів моделі і забезпечити його автоматизовану обробку, доцільно використовувати онтологічне подання знань про них. Таким чином, для опису інформаційної потреби користувача треба побудувати онтологічну модель, класи якої відповідають вказаним вище елементам, а відносини – зв'язкам між ними (рис. 1).

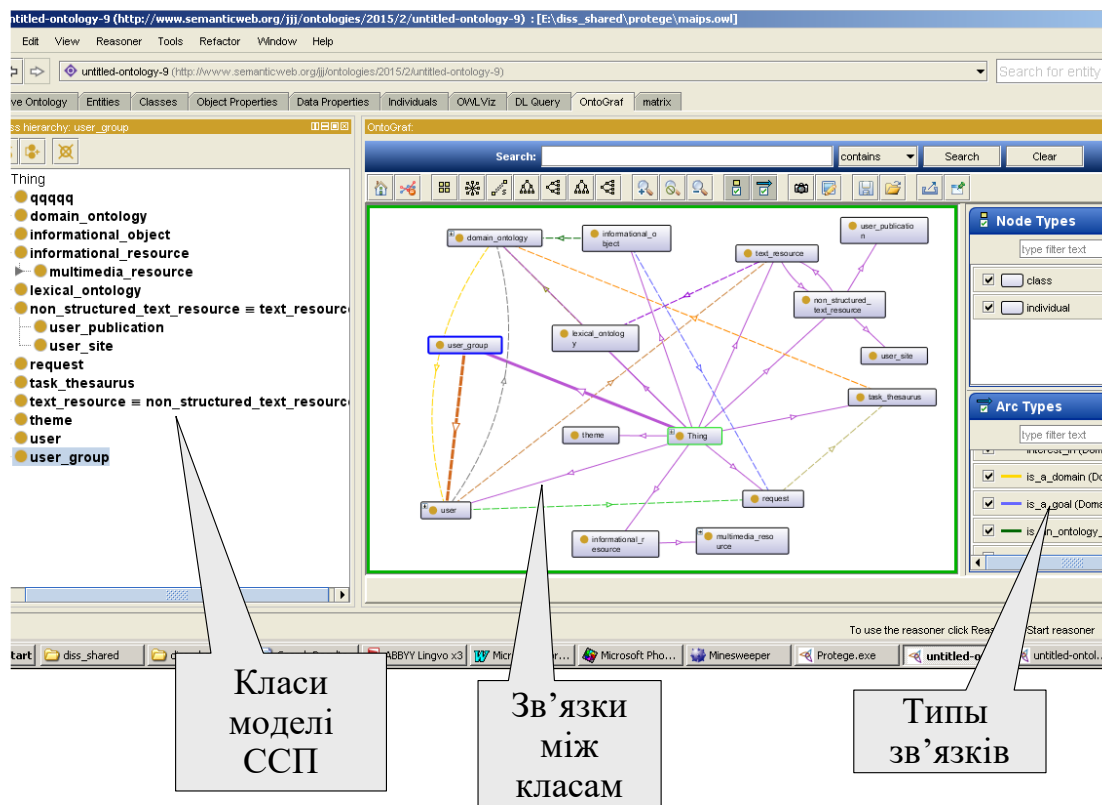


Рис. 1. Онтологічна модель взаємодії користувачів та ІР при семантичному пошуку

Для того, щоб описати знання й інтереси користувача, в онтологічну модель додаються два службових класи – «компетенція» і «атомарна компетенція». Набір екземплярів атомарних компетенцій може імпортуватися з різних онтологій Про, зв'язаних з навчальними закладами, працевлаштуванням і науково-дослідними організаціями.

Найбільш важливі класи онтологічної моделі ССП – це *користувач* – особа, яка потребує інформації, і *інформаційний об'єкт* – структурований образ тієї інформації, що користувач має намір отримати від ССП.

Інформаційна модель користувача

Розглянемо детальніше елементи класу онтологічної моделі “користувач”.

Інформаційна модель користувача відображає:

- 1) персональні переваги користувача, його знання та здатність до сприйняття інформації;
- 2) інформаційні потреби користувача, особливості задачі, що постала перед ним, і тих інформаційних об'єктів, знаходження відомостей про які дасть змогу йому розв'язати цю задачу;
- 3) досвід взаємодії користувача з пошуковою системою;
- 4) відомості, що характеризують ставлення користувача до досвіду інших користувачів і визначають доцільність використання колаборативного пошуку.

Цей клас має досить складну структуру. Його атрибути можна поділити на кілька груп:

Група 1. Реєстраційна інформація:

- ідентифікатор користувача;
- пароль для доступу до ІПС;

Група 2. Відомості, імпортовані з зовнішніх джерел (необов'язкові відомості, їх може й не бути):

- ідентифікатори користувача в соціальних мережах, що дають змогу динамічно оновлювати знання про нього;
- рейтинги користувача в соціальних мережах;
- адреса профілю акаунту користувача у Вікіпедії та інших Wiki-ресурсах;
- адреса сайту користувача;
- сфера компетенцій користувача (ключові слова, імпортовані з соціальних мереж);
- посилання на публікації користувача.

Група 3. Власні характеристики користувача: сфера компетенцій користувача (список ключових слів, що вводяться користувачем безпосередньо).

Група 4. Формальні дані про користувача (необов'язкові відомості, що дають змогу ІПС формувати групи користувачів зі схожими інформаційними потребами): місце проживання; вік; професія, освіта тощо.

Досвід взаємодії ІПС з користувачем:

- список онтологій, які користувач застосовував для опису своїх інформаційних інтересів;
- список тезаурусів, що користувач використовував у пошукових запитах;
- список раніше виконаних запитів;
- список результатів виконаних запитів з оцінками користувача для знайдених результатів.

Джерела поповнення онтологічної моделі відомостями про користувачів

Відомості про користувачів є основою для персоніфікації семантичного пошуку та створення рекомендацій. Отримання й аналіз знань про них дають змогу зробити інформаційний пошук значно ефективнішим.

Значну частину інформації про користувача він уводить сам під час реєстрації в МАІПС: чим детальніше користувач відповість на питання, тим більш пертинентними будуть результати пошуку. Так до онтології потрапляють відомості про місце проживання користувача, освіту, посаду, вік, знання мов тощо. Знання про сфери інтересів і області компетентності користувача можуть імпортуватися із зовнішніх джерел, на які вказує користувач (соціальних мереж, Wiki-ресурсів, особистих сторінок, природномовних ІР та зовнішніх онтологій), спеціалізованими сервісами (наприклад, для поповнення знань стосовно сфер інтересів користувача може бути використаний сервіс семантичного аналізу обраних ІР, а для визначення психофізіологічних особливостей користувача – спеціалізовані сервіси експрес-тестування).

Знання про користувача поповнюються в процесі його взаємодії з пошуковою системою. Деякі відомості про користувача заносяться в його профіль автоматично: щоразу, коли користувач вибирає якусь онтологію, чи створює або модифікує тезаурус, формує запит і виконує раніше створений запит, відомості про це фіксуються в його персональних даних. Крім того, користувач може сам вносити зміни у свій профіль, щоб він більш коректно відображав його інформаційні потреби: приміром, вказати користувачів, пошукова діяльність яких може бути цікавою для нього.

Для персоніфікації пошуку важливо правильно оцінювати й урахувати з метою подальшої роботи психофізіологічні властивості користувача, що впливають на його здатність до сприйняття інформації, наприклад, класифікація за швидкістю збудження й гальмування. Якщо користувач знає свої властивості, він може ввести їх або пройти експрес-тест.

Крім того, за допомогою індексу легкості читання користувач може оцінити, які IP достатньо зрозумілі та корисні для нього. Для цього він може або безпосередньо вказати індекс легкості читання (приміром, тексти для людей з середньою освітою), або вказати IP, що відповідають і не відповідають його вимогам, і тоді система обчислить цей індекс автоматично, враховуючи використані раніше користувачем онтології та тезауруси.

Область інтересів користувача визначається через онтологічні моделі певних ПрО. Тому джерела відомостей про онтології ПрО є також джерелами відомостей щодо користувача.

Підкреслимо, що для представлення знань у МАПС використовуються онтології, які можна поділити на дві окремі групи – *внутрішні* та *зовнішні*. Внутрішні онтології створюються безпосередньо розроблювачами МАПС і можуть поповнюватися в процесі взаємодії МАПС з користувачами. Основною особливістю таких онтологій є те, що розроблювачам цілком відома їх структура та зміст, тому можна прогнозувати кінцівку обчислень. Зовнішні онтології – це онтології, представлені на загальновідомих перевірених ресурсах Semantic Web [6, 7], список яких поповнюється та узгоджується, а інтерфейс підключення та формат обміну даними розроблюється в кожному випадку окремо.

Використання Wiki-ресурсів в якості зовнішніх онтологій

Якщо користувач виконує деяку діяльність щодо створення або редагування певних сторінок Wiki-ресурсу, то можна з певною ймовірністю вважати, що контент цієї сторінки цікавить даного користувача. Більш того, як правило, користувачі працюють з тими сторінками, які стосуються питань, в яких ці користувачі вважають себе експертами (ця переконаність може бути й помилковою, але вона відображає у першу чергу не знання користувача у певній ПрО, а його переконання та інтереси).

Таким чином, можна вважати, що діяльність користувача в Wiki є відображенням його інтересів (протилежне твердження не є правильним, тобто користувач може бути обізнаним та зацікавленим у тих питаннях, сторінки яких він не редагував).

Можна також казати про групи користувачів зі спільними інформаційними інтересами, якщо ці користувачі спільно працювали над декількома Wiki-сторінками [1]. Чим більше кількість таких сторінок (відносна та абсолютна), тим вище ймовірність того, що інтереси цих користувачів можуть збігатися.

Зрозуміло, що чим більш детально профіль користувача характеризує його потреби, інтереси та спроможності до сприйняття інформації, тим більш пертинентно буде задоволена його інформаційна потреба. Тому користувач зацікавлений у тому, щоб ця модель поповнювалася ефективно та динамічно, але з невеликими для нього затратами часу та зусиль.

Для спрощення задачі поповнення профілю користувача, ми пропонуємо автоматичне отримання знань з Wiki-ресурсу, з яким користувач працював та має історію такої взаємодії. Для того, щоб повідомити ССП щодо своїх інтересів, користувач може повідомити ССП адресу Wiki-ресурсу та свій логін (нік-ім'я) у цьому ресурсі. Маючи такі дані, ССП матиме змогу відслідковувати, які саме сторінки редагує цей користувач.

Але просто побудувати перелік таких сторінок не дуже корисно. Більш ефективним використанням буде здобути знання, які містять ці сторінки. Якщо Wiki ресурс є семантизованими, тобто містить семантичну розмітку, то виконати таку процедуру можна автоматизовано за допомогою спеціальних вбудованих засобів Wiki [8].

Приміром, такі можливості надає Semantic MediaWiki.

Для отримання знань з семантичних Wiki-джерел в меню “Спеціальні сторінки “ у підменю для “Semantic MediaWiki” потрібно обрати “Експорт сторінок у RDF”.

За множиною обраних сторінок Wiki-ресурсу буде побудовано онтологію, яка подається мовою RDF в стандартному форматі XML. Таку онтологію можна надалі використовувати для семантичного пошуку як онтологію ПрО (рис. 2).

Побудований таким чином файл в форматі RDF можна проглядати, обробляти та редагувати за допомогою редактора онтологій Protégé (рис. 3). При цьому певну проблему викликає лише кодування українських назв класів та екземплярів.

Побудована таким чином онтологія може використовуватися в МАПС для того, щоб охарактеризувати сферу інформаційних інтересів користувача. За термінами цієї онтології користувач може побудувати тезаурус, який відображає поточні інформаційні потреби користувача.

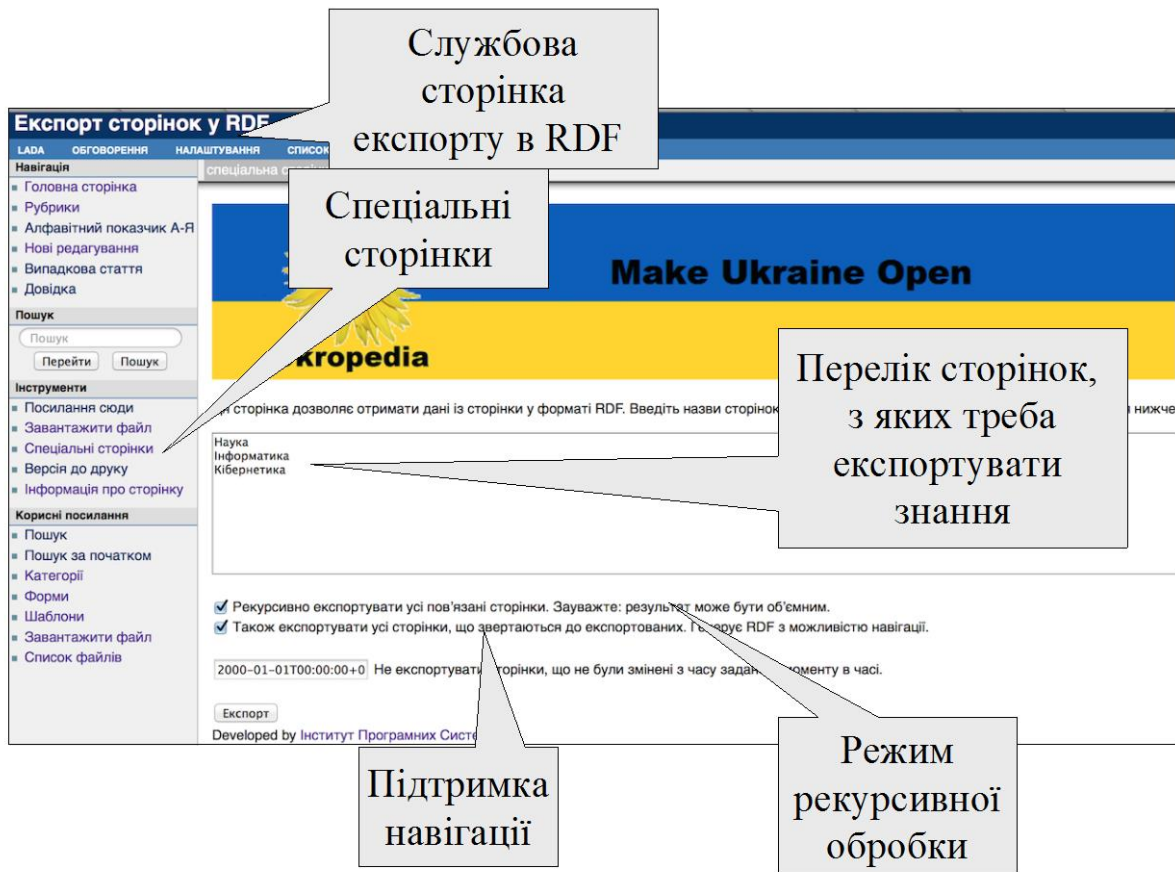


Рис. 2. Побудова RDF-файла за семантично розміченими Wiki-сторінками

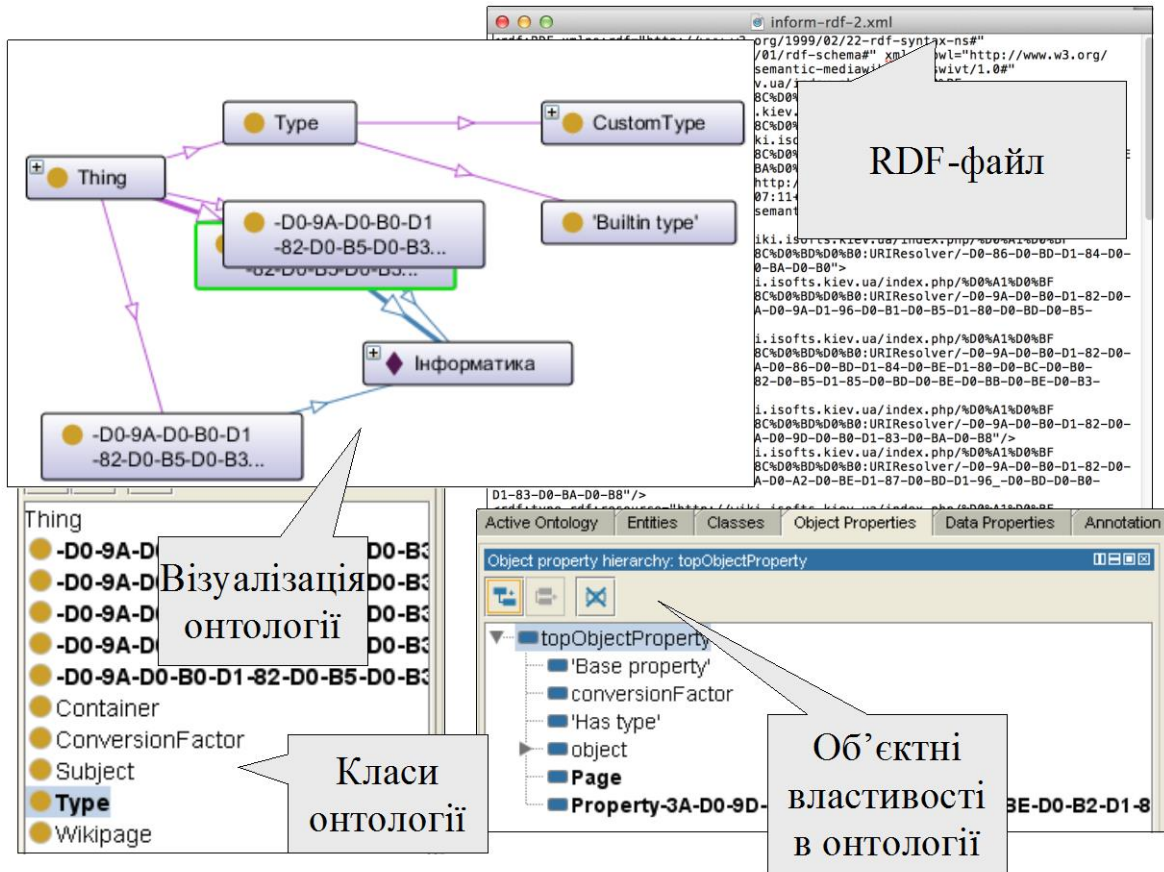


Рис. 3. Представлення згенерованого RDF-файлу в Protégé

Стандартний профіль користувача вікі-ресурсу

В звичайному стандартному пакеті ПО ВікіМедіа [9] профіль користувача поданий доволі стисло – лише

- логін (нік-ім'я) користувача;
- електронна адреса;
- пароль.

Окрім цієї загальної інформації, в Вікі-системі для адміністрування є система розподілення прав, яка встановлює групи користувачів та які права мають ці групи.

Для відстеження змін, ВікіМедіа зберігає всю історію редагування сторінок, з інформацією про те, який користувач які сторінки, категорії, класи, властивості, тощо, редагував.

Для отримання більш докладної інформації про користувача доволі цікавим видається можливість знати, які сторінки редагував певний користувач. Така інформація зберігається і її можливо отримати на спеціальній сторінці «Спеціальна:Внесок».

Таким чином, маючи інформацію про сторінки, які редагував користувач, та використовуючи можливість Семантичної надбудови ВікіМедіа для отримання RDF-даних, ми отримали можливість автоматизовано поповнювати профіль користувача МАІПС знаннями про Про його інтересів.

Висновки

Як показує вищенаведений аналіз, Вікі-сторінки, з одного боку, є результатом інтелектуальної діяльності певних осіб, інформація щодо чого фіксується системою підтримки Вікі-середовища.

З іншого боку, семантичні Вікі-ресурси є джерелом для побудови різноманітних легковажних онтологій, які формалізують сферу інтересів та компетентності як окремих осіб, так і спільнот, які складаються з груп осіб.

Таким чином, з точки зору семантичного пошуку, Вікі-ресурси можуть використовуватися для забезпечення наступних задач:

- розширення та поповнення профілю користувача ССП в онтологічній моделі семантичного пошуку, а саме – його компетенцій;
- автоматизованої побудови персоніфікованих онтологій Про, на яких базуються користувацькі запити до ССП;
- формування спільнот користувачів зі спільними інформаційними потребами для проактивної побудови рекомендацій щодо IP, що використовуються;
- побудови та уточнення тезаурусів задач, які можуть формуватися як множина термінів з побудованої за множиною Вікі-сторінок RDF-конструкції.

Доцільність такого застосування Вікі ресурсів та пов'язаних з ними методів обробки доведено в процесі використання сторінок з прототипу Української енциклопедії [10] в системі семантичного пошуку МАІПС, коли категорії та семантичні властивості сторінок, які редагував певний користувач, використовувалися для збагачення його профілю та вдосконалення пошуку інформації.

1. *Leuf B., Cunningham W.* The Wiki way: collaboration and sharing on the Internet, 2001. – Режим доступу: <http://www.citeulike.org/group/13847/article/7659081>.
2. *Gruber T.R.* Toward Principles for the Design of Ontologies Used for Knowledge Sharing // International Journal of Human-Computer Studies. – 1995. – V. 43, Issues 5–6. – P. 907–928.
3. *Розушина Ю.В.* Использование онтологической модели при семантическом поиске информационных объектов // Онтология проектирования. – 2015. – Том 5, № 3(17). – С. 336–356. – http://agora.guru.ru/scientific_journal/files/Ontology_Of_Designing_3_2015_shot.pdf.
4. *Розушина Ю.В., Гладун А.Я.* Семантическая Википедия как источник онтологий для интеллектуальных поисковых систем // В кн.: Advanced Research in Artificial Intelligence. International Book Series «Information Science and Computing». – ITHEA, Sofia, 2008. – P. 172–178.
5. *Гришанова Л.Ю.* Аналітичний огляд методів і засобів інформаційного пошуку в Semantic Web, Проблеми програмування. – 2016. – № 1. – С. 51–72.
6. *Introduction to Semantic Web.* – <http://www.mphasis.com/knowledge-center/white-papers-all.asp>.
7. *Davies J., Fensel D., van Harmelen F.* Towards the Semantic Web: Ontology-driven knowledge management // John Wiley & Sons Ltd, England. – 2002. – 288 p.
8. *Krotzsch M., Vrandečić D., Volkel M.* Semantic MediaWiki. – <http://c.unik.no/images/6/6d/SemanticMW.pdf>.
9. *Barrett D.J.* MediaWiki (Wikipedia and Beyond). O'Reilly Media, 2008. – <http://www.nhmnc.info/wp-content/uploads/fbpdfs2014/MediaWiki-Wikipedia-and-Beyond-by-Daniel-J-Barrett-Good-Basic-Primer.pdf>.
10. *Методичні рекомендації з підготовки, редагування та оформлення статей до Великої української енциклопедії (проект) / За ред. д. і. н., проф. Киридон А.М.* – К.: ДНУ «Енциклопедичне видавництво», 2015. – 120 с.

References

1. Leuf B. The Wiki way: collaboration and sharing on the Internet, 2001 / B. Leuf, W. Cunningham. — Режим доступу: <http://www.citeulike.org/group/13847/article/7659081>.
2. Gruber T. R. Toward Principles for the Design of Ontologies Used for Knowledge Sharing / T. R. Gruber // International Journal of Human-Computer Studies. — 1995. — V. 43, Issues 5-6. — P. 907—928.
3. Rogushina J. Use of the ontological model for semantic search of information objects // Ontology for design, V. 5, No3(17), 2015. — P.336-356. — http://agora.guru.ru/scientific_journal/files/Ontology_Of_Designing_3_2015_shot.pdf.
4. Rogushina J. Semantic Wikipedia as a source of ontologies for intelligent retrieval systems / J.Rogushina, A.Gladun // Advanced Research in Artificial Intelligence. International Book Series «Information Science and Computing». — ITHEA, Sofia, 2008. — P. 172—178.
5. Grishanova I.Y. Analytical review on information retrieval methods and applications in the Semantic Web// Problems in Programming, 2016, №1, p.51-72
6. Introduction to Semantic Web. — <http://www.mphasis.com/knowledge-center/white-papers-all.asp>.
7. Davies J. Towards the Semantic Web: Ontology-driven knowledge management / Davies J., Fensel D., van Harmelen F. — John Wiley & Sons Ltd, England, 2002. — 288 p.
8. Krotzsch M., Vrandečić D., Volkel M. Semantic MediaWiki. — <http://c.unik.no/images/6/6d/SemanticMW.pdf>.
9. Barrett D.J. MediaWiki (Wikipedia and Beyond). O'Reilly Media, 2008. — <http://www.nhmnc.info/wp-content/uploads/fbpdfs2014/MediaWiki-Wikipedia-and-Beyond-by-Daniel-J-Barrett-Good-Basic-Primer.pdf>.
10. Methodical recommendations for preparing, editing and execution of articles for Big Ukrainian Encyclopedia (project) / Ed.by Kiridon A.M. — Kiev.: DNU «Enciclopedichne vidavnistvo», 2015. — 120 p.

Про авторів:

Рогущина Юлія Віталіївна,

кандидат фізико-математичних наук,
старший науковий співробітник.

Кількість наукових публікацій в українських виданнях — 110.

Кількість наукових публікацій в іноземних журналах — 28.

Індекс Гірша — 2 (Scopus), 11 (Google Scholar).

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

Гришанова Ірина Юріївна,

науковий співробітник,

Кількість наукових публікацій в українських виданнях — 16.

<http://orcid.org/0000-0003-4999-6294>.

Місце роботи авторів:

Інститут програмних систем НАН України,

03181, Київ-187, Проспект Академіка Глушкова, 40.

Тел.: (066) 550 1999.

E-mail: ladamandraka2010@gmail.com