

ПРОГРАМНИЙ ЗАСІБ ПОБУДОВИ СЕМАНТИЧНИХ ПОРТАЛІВ

Гончар Л.І.¹⁾, Ляхоцький О.С.²⁾

Тернопільський національний економічний університет

¹⁾ к.е.н., доцент; ²⁾ магістрант

I. Постановка проблеми

На сьогоднішній день у всесвітній мережі Інтернет знаходиться 1 099 511 627 776 гігабайт або 1 зеттабайт даних. Більша частина цих даних є не структурованою і, мало того, ще й зберігається в різних форматах даних, що погіршує їх обробку на програмному рівні і значно уповільнює фільтрацію. Для вирішення цієї проблеми було створено Semantic Web [1,2].

Semantic Web — це надбудова над сучасною Всесвітньою павутиною, яка покликана зробити інформацію, що розміщена в мережі, зрозумілішою для комп'ютерів. Відомо, що майже вся інформація в Інтернеті, знаходиться в текстовій формі. Не секрет також, що прогрес в галузі обробки людської мови (англ. Natural Language Processing) йде дуже повільно. Комп'ютери не можуть сприйняти й осмислити словесну інформацію, розміщену в Інтернеті, і в найближчий час, мабуть, не зможуть. Тому розробка веб-порталів для побудови семантичних моделей є надзвичайно актуальною задачею.

II. Мета роботи

У даній науковій роботі вирішується завдання побудови веб-семантичного порталу (Semantic Web), основна ідея якого полягає в тому, щоб зробити інформацію, передану в Web, більше формалізованою й зручною для машинного сприйняття, зокрема, для того щоб її можна було ідентифікувати й класифікувати [7]. На думку авторів технології Semantic Web, це може досягти за допомогою введення метаданих, які повинні супроводжувати будь-яку інформацію й розповідати про її походження, формат і багато іншого, що повинне радикальним способом полегшити пошук інформації в Web і її обробку.

III. Особливості програмної реалізації семантичного порталу

Наступні технології є основними в складі Semantic Web [8] :

- Глобальна схема імен (URI);
- Модель опису даних (RDF);
- Мова опису словників (RDFS);
- Засоби опису зв'язків між об'єктами даних (онтології, і мова їхнього опису OWL).

SPARQL (англ. Protocol And RDF Query Language) — нова мова запитів для швидкого доступу до даних RDF.

Для створення зрозумілого комп'ютеру опису ресурсу в семантичній павутині використовується формат RDF (англ. Resource Description Framework). RDF дозволяє об'єднати інформацію з довільних джерел. Формат RDF найбільш корисний у забезпеченні спільного використання інформації, зміст якої може однаково інтерпретуватися різними програмними агентами. Специфіка моделі даних RDF полягає в тому, що ресурси й властивості ідентифікуються за допомогою глобальних ідентифікаторів (URI). RDF описує предметну область у термінах ресурсів, властивостей ресурсів і значень властивостей. RDF - дані можна розцінювати як сукупність тверджень - суб'єкт, предикат і об'єкт твердження, і представляти у вигляді спрямованого графа, утвореного такими твердженнями.

На рисунку 1 у термінах відповідних сутностей і зв'язків зображена загальна схема моделі RDF. Тут під властивістю (Property) варто розуміти якийсь аспект, характеристику, атрибут або відношення, що використовується для опису ресурсу. Кожна властивість має свій специфічний зміст, припустимі значення, тип ресурсів, до яких воно може бути застосовано, а також відносини з іншими властивостями. Для забезпечення унікальності імен властивості дотримуються концепції URI, тобто властивість стає потенційним об'єктом для опису за допомогою RDF окремо від ресурсу, що характеризується наявним значенням.

Базовий будівельний блок моделі даних RDF - твердження, що представляє собою трійку: ресурс, іменована властивість і його значення. У термінології RDF ці три частини твердження називаються відповідно: суб'єкт (subject), предикат (predicate) і об'єкт (object) [3]. Ресурсом у цьому

випадку називають усе, що описується засобами RDF. Це може бути звичайна Web-Сторінка або якась її частина, наприклад, окремий елемент HTML розмітки.

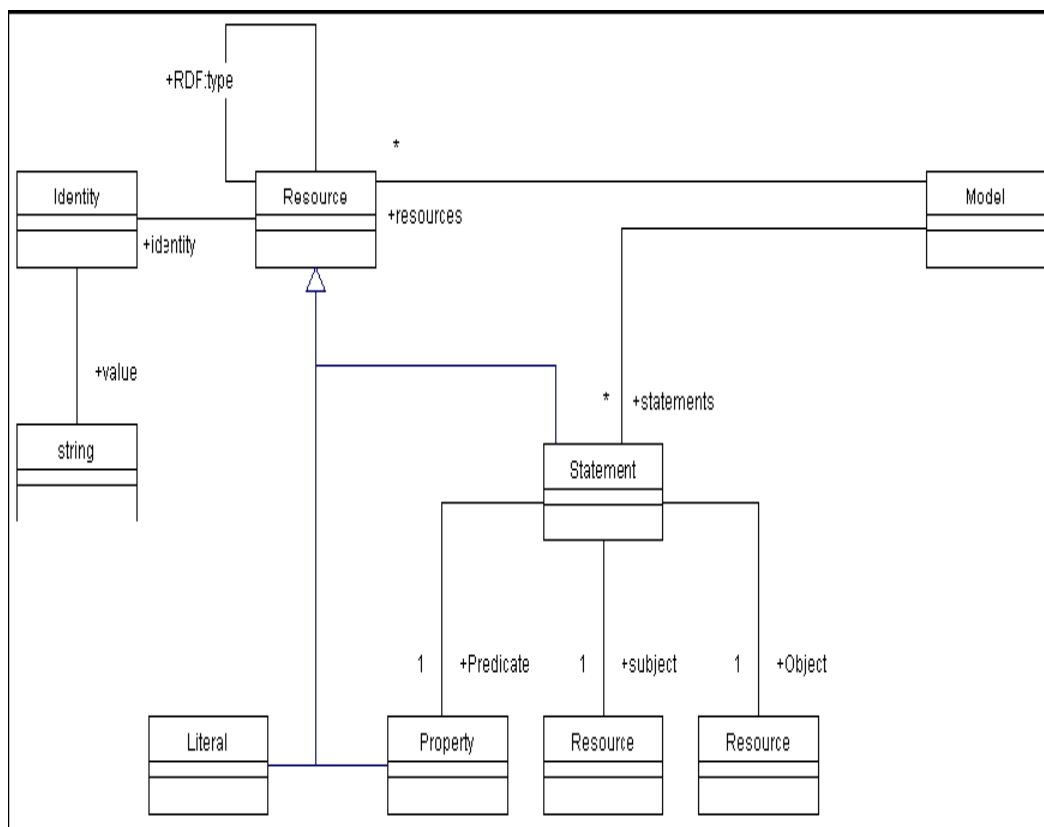


Рисунок 1 - Модель RDF

Висновок

У науковій роботі було здійснено аналіз існуючих підходів до розробки порталів знань на основі технологій Semantic Web. Встановлено, що для структуризації й класифікації інформаційних ресурсів необхідно використовувати підхід Topic Maps.

Здійснено реалізацію автоматичного рубрикування на основі семантичного аналізу вмісту інформаційних ресурсів. Даний підхід базується на певних способах подання знань про предметну область і текстової інформації. Він забезпечує значно кращу якість автоматичної класифікації ресурсів за рахунок того, що семантичний аналіз змісту ресурсу забезпечує більш достовірну оцінку приналежності ресурсу до тієї або іншої тематичної рубрики.

Список використаних джерел

1. <https://www.w3.org/standards/semanticweb/>
2. <http://www.semanticweb.narod.ru/>
3. <http://www.alik.su/articles/semantic-web/semantic-web-2.pdf>
4. "Semantic Web for the Working Ontologist, Second Edition: Effective Modeling in RDFS and OWL" by Dean Allemang, May 20, 2011
5. "Learning SPARQL" by Bob DuCharme
6. "Programming the Semantic Web" by Toby Segaran and Colin Evans, Jul 24, 2009
7. "Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL" by Dean Allemang and James Hendler, May 9, 2008
8. http://semanticweb.org/wiki/Main_Page.html