

УДК 004.65:681.3:06(477)

О.К.Жигаревич

Луцький інститут розвитку людини Університету «Україна»

ДОСЛІДЖЕННЯ СТАНДАРТНИХ СПОСОБІВ ОПИСУ ЗВ'ЯЗКІВ МІЖ ОБ'ЄКТАМИ ДАНИХ ЗА ДОПОМОГОЮ WEB ONTOLOGY LANGUAGE

Дослідження стандартних способів опису зв'язків між об'єктами даних за допомогою Web Ontology Language. Метою даної статті є опис та дослідження зв'язків між об'єктами даних за допомогою Web на основі OWL. Сценарії використання OWL. Web-портали і корпоративні сайти.

Ключові слова: онтологія, каталоги сайтів, мультимедія, представлення знань, формальна семантика, Web онтології.

Постановка проблеми

Синтаксична взаємодія мереж - необхідна умова для того, щоб множинні додатки могли посправжньому "розуміти" дані і працювати з ними як з інформацією. Це також необхідна умова для коректної перевірки даних. Синтаксична взаємодія мереж вимагає перетворення між термінами, для чого, у свою чергу, необхідний контент-аналіз.

Такий контент-аналіз вимагає формальних і докладних специфікацій моделей доменів, які визначають терміни і їх зв'язки, що використовуються. Подібні формальні моделі доменів іноді називаються онтологіями. Вони визначають моделі даних в термінах класів, підкласів і властивостей.

Онтологічна мова Web (Web Ontology Language), що рекомендується консорціумом W3C, допомагає у представленні онтологій. Робоча онтологічна мова (Ontology Working Language, скор. OWL) додає більше можливостей для опису властивостей і класів, ніж RDF або схема RDF. Зокрема, дозволяє описувати зв'язки між класами, потужність множини (наприклад, "рівно один"), рівність, більшу типологію властивостей і їх характеристики (наприклад, симетрію).

Онтологічна мова Web на основі OWL розроблена для використання додатками, які повинні працювати із змістом інформації, а не просто надавати її користувачу. OWL покращує можливості автоматичної інтерпретації вмісту інтернету в порівнянні з тими, що можуть забезпечити XML, RDF і схема RDF. Це відбувається завдяки тому, що OWL надає додаткові можливості разом з формальною семантикою. OWL включає три підмови: повний OWL (OWL Full), OWL DL і спрощений OWL (OWL Lite) . [1]

- Повна версія онтологічної мови Web на основі OWL OWL Full. Ця мова використовує всі базисні елементи мови OWL і дозволяє комбінувати їх випадковим чином з RDF і схемою RDF. Повний OWL сумісний "від" низу до верху з RDF, як синтаксично, так і семантичний: будь-який дозволений документ RDF є також дозволеним документом OWL Full. Маловірогідно, що які-небудь інтелектуальні програмні засоби здатні підтримувати всі можливості OWL Full, оскільки ця мова пропонує максимум виразних засобів і синтаксичної свободи RDF за відсутності обчислювальних гарантій.
- OWL DL призначений для тих користувачів, кому необхідний максимум виразних засобів без втрати обчислювальних можливостей. OWL DL - це підмова конструкцій мови OWL Full з деякими обмеженнями, такими як розділення типів (type separation) (наприклад, клас не може бути одночасний індивідуальним елементом або властивістю, а властивість не може одночасно бути індивідуальним елементом або класом).
- OWL Lite призначений для користувачів, яким необхідна класифікаційна ієрархія і прості обмежувальні можливості. Перевагою цієї мови є велика легкість його розуміння і упровадження в порівнянні з двома іншими. Але в той же час його виразні можливості набагато нижче. Наприклад, хоча OWL Lite і підтримує обмеження потужності множини, єдиними допустимими значеннями цього параметра є 0 або 1.

Прикладами онтологій є каталоги сайтів інтерактивних покупок, таких як Amazon.com, стандартні термінології тієї або іншої області діяльності, наприклад, UNSPSC - United Nations Standard Products and Services. [2]

Основні компоненти OWL включають класи, властивості і індивідуальні елементи. Класи - це основні блоки онтології OWL. Клас - це концепція в домені. Класи звичайно утворюють таксономічну ієрархію (тобто систему підклас-надклас).

Метою дослідження виступає наступне:

- визначення класів;
- спільне використання загального розуміння структури інформації серед користувачів;
- можливість повторного використання знань домену;
- можливість робити явними припущення домену;
- сценарії використання OWL;
- реалізації Семантичного Web;

Класи визначаються за допомогою елемента owl: Class. В мові OWL існує два наперед визначених класи: owl:Thing і owl:Nothing. Перший з них є самим загальним і включає все, другий - це порожній клас. Будь-який клас, визначуваний користувачем, є підкласом класу owl:Thing і надкласом класу owl: Nothing. Приклади класів в області банківської справи можуть включати класи Рахунків (Account) або Клієнт (Customer). [3]

Властивості включають дві основні категорії:

- властивості об'єкту (Object properties), які зв'язують індивідуальні елементи між собою;
- властивості типів даних (Datatype properties), які пов'язують індивідуальні елементи із значеннями типів даних, такими як цілі числа, числа з плаваючою комою і рядка. Для визначення типів даних OWL використовує схему XML.

Властивість може включати домен і деяку область, пов'язану з ним. Будь-яка властивість потрапляє в одну з наступних категорій:

- функціональна: для будь-якого об'єкту властивість може приймати тільки одне значення (наприклад, вік, рост або вага людини);
- зворотно-функціональна: два різні індивідуальні елементи не можуть мати одного і того ж значення. Наприклад, у кожної людини свій унікальний номер банківського рахунку або так званий SSN (social security number)¹;
- симетрична: якщо властивість зв'яже елемент А з елементом В, то з цього можна зробити висновок, що воно також зв'яже елемент В з елементом А. Приклади симетричних властивостей включають вирази типу "є братом (сестрою)" або "такий же, як";
- транзитивна: якщо властивість зв'яже елемент А з елементом В, а елемент В з елементом С, то можна припустити, що воно також зв'яже елемент А з елементом С. Наприклад, якщо А вище В, а В вище С, то А вище С.

До класів і властивостей можуть застосовуватися різні обмеження. Наприклад, обмеження потужності множини указують на число зв'язків, в яких може брати участь клас або індивідуальний елемент. [4]

Сценарії використання OWL. Web-портали і корпоративні сайти. Портал дозволяє користувачам, зацікавленим в конкретній тематиці, одержувати новини, інтерактивно спілкуватися один з одним, створювати віртуальні співтовариства і т.п., причому простій індексації порталу по рубриках і тематиках для точного знаходження потрібної інформації за запитом може бути недостатньо. Для більш інтелектуального узагальнення розділів інформації порталам необхідно визначити свою онтологію, яка повинна описувати термінологію, що використовується у вмісті порталу, і аксіоми, задаючі правила використання цих термінів в контексті інших термінів. Аксіоми також можуть описувати нові терміни, користуючись вже певними примітивними термінами з поточної або якої-небудь іншої онтології. В поєднанні з якими-небудь фактами з наочної області з цих визначень і аксіом можна вивести нові факти, які, у свою чергу, можуть бути отриманий користувачем при такому інтелектуальному пошуку по порталу, але не могли бути отриманий раніше при звичайному пошуку. Наприклад, якщо зареєстрований користувач порталу шукає інформацію про мишей, і в його реєстраційних даних

вказано, що людина займається комп'ютерами, то при добре спроектованій онтології в процесі пошуку більший пріоритет матимуть документи, що містять інформацію про комп'ютерних мишей, а не про гризунів. Зрозуміло, ефективність запропонованої технології залежить від авторів порталу, що публікують інформацію, оскільки їм доведеться занотувати всі сторінки метаінформацією, записаною на мові онтологій.[5]

Онтології використовують при систематизації даних на корпоративному порталі для індексації і зручного пошуку — не дивлячись на те, що багато організацій мають власну таксономію для організації внутрішньої інформації, цього звичайно недостатньо. Проста класифікація сильно обмежує можливості пошуку і індексації, оскільки багато документів можуть підпадати під різні категорії, тому пошук по різних критеріях буде набагато ефективніший, ніж звичайний пошук за ключовими словами. Крім того, типова проблема користувачів корпоративного порталу полягає в неможливості користування єдиною термінологією. Наприклад, фахівець з комп'ютерного відділу використовує для опису проблеми свій професійний жаргон, а фахівець з відділу продажів погано розуміє отриманий документ через те, що звик користуватися іншою термінологією. Зрозуміло, ця проблема розв'язується іншими засобами, але майже завжди створює додатковий головний біль. У випадку з OWL достатньо створити декілька онтологій термінів для різних типів користувачів, але з посиланнями на відповідні терміни з інших онтологій, щоб одні терміни могли транслюватися в інші автоматично.[6]

Мультимедіа і розробка документації. OWL добре підходить для опису різних нетекстових об'єктів, наприклад, колекцій зображень, відео- або аудіо-файлів. Очевидно, що для машини задача автоматичного витягання семантичної метаінформації з мультимедіа-файлів поки що неможлива. За допомогою OWL можна спроектувати систему опису вмісту мультимедіа-файлів, і потім здійснювати пошук, скажімо, серед відеофайлів по їх вмісту.

Документація буває різних типів, кожний з яких має свою ієрархічну структуру, використовує свій набір перехресних посилань, який може бути заданий як явно, так і неявно. Онтології в цьому випадку можна застосовувати для побудови єдиної інформаційної моделі, щоб в будь-якій документації використовувалися чітко певні терміни. Треба відмітити, що онтологія і таксономія (ієрархічна класифікація) не обов'язково незалежні, але цілком можуть розроблятися паралельно.[7]

Реалізації Семантичного Web. Вже сьогодні практично всі відомі компанії рівня IBM, Adobe або Sun Microsystems, активно використовують технологію Семантичного Web в своїх продуктах для вирішення задач управління даними.

Компанія Microsoft інвестує сотні мільйонів доларів в проект взаємодіючих мережних ресурсів NET, який відображає їх уявлення про найближчий майбутній Internet. Створювана система дозволяє проводити автоматизований обмін мережними ресурсами між окремими програмами, додатками, базами даних, користувачами, ґрунтуючись на XML, як на ключовій технології. [8]

В Європі ведеться проект, подібний Семантичному Web, - "Мережа знань", Knowledge Web (<http://kw.dia.fi.upm.es/semanticportal/jsp/frames.jsp>). Ця мережа орієнтована на потреби інформаційних технологій в промисловості, науці і освіті, а Семантичний Web (підтримуваний і в Європі - SWAD-Europe, www.w3.org/2001/sw/Europe/) більше розрахований на електронну комерцію і спрощення роботи користувачів мережі Інтернет. [9]

Недавно в рамках ідеології Семантичного Web була розроблена в School Electronics & Computer Science (ECS) Університету Саутгемптона була розроблена система mSpace. Програмне забезпечення цієї системи є набором могутніх інструментів, що дозволяють збирати дані з різних джерел і організувати інформацію по категоріях і даючих можливість користувачу вільно орієнтуватися в ній.

Розробники наводять наступний приклад. Наприклад, якщо в Google набрати "класична музика", то пошукач покаже посилання на сайти, так чи інакше що стосуються класичної музики. Якщо ж шукати "класичну музику" на mSpace, то буде виданий список композицій, які можна тут же викачати. Інший приклад - по запиту "Гарі Поттер" користувач отримає не просто набір посилань, а відсортований звіт, в якому частина посилань лежатиме в графі "фільми", інша частина - в колонці "книги", а третя - в колонці "рецензії". Семантичний Web надасть користувачу можливість вибирати, в якому напрямі досліджувати інформацію, а не просто видавати саме релевантне по загальному алгоритму. [10]

Семантичний Web дозволяє створення системи з елементами "штучного інтелекту", яка б дозволила спеціальним додаткам якісно шукати в Internet необхідну інформацію, а також обмінюватися інформацією один з одним. При цьому саме мова онтологій OWL виступила вирішальною компонентою інтелектуалізації, базисом для побудови семантичних мереж. Необхідно відзначити, що вперше теорія семантичних мереж народилася в середині роках минулого століття і була орієнтована на задачі штучного інтелекту, зокрема, машинного перекладу. Знання в теорії семантичних мереж представлялися у вигляді вузлів, сполучених дугами, кожна з яких визначала тип відношення. Семантичний Web по суті є реалізацією ідеї штучного інтелекту, проте даний термін не дуже популярний зважаючи на велику кількість невдалих проєктів в цій області, тому поняття "семантична мережа" сьогодні викликає настороженість. Разом з тим, Web онтології по суті є справжньою базою знань, одну з концептуальних основ штучного інтелекту. [11]

Висновки Представленням знань в Семантичному Web властиві універсальні виразні можливості, синтаксична і семантична інтероперабельність. Отже, мови Семантичного Web дозволяють представляти будь-який вид даних, створювати численні синтаксичні аналізатори і інтерфейси прикладних програм необхідні для маніпулювання даними. Семантична інтероперабельність реалізується в онтологіях шляхом встановлення відповідності між термінами, що використовуються.

1. Основи системного аналізу. Підручник для вищих навчальних закладів. / [М..З.Згуровський, Н.Д. Панкратова]. Київ: Видавнича група BNV, 2007. - 543с.
2. Моделювання систем. Підручник для вищих навчальних закладів. / [Томашевський В.М.] Київ: Видавнича група BNV, 2005. - 350 с.
3. Международный Форум "Образовательные Технологии и Общество" – Восточно-европейская подгруппа International Forum of Educational Technology & Society <http://ifets.ieee.org/russian/>
4. Classic: A structural data model for objects. [Borgida A., Brachman R., McGuinness D., Resnick L.] / ACM SIGMOID Int. Conf. on Management of Data, Portland, Oregon, USA, — 1989.
5. An Overview of the KL-ONE Knowledge Representation System. [Brachman R., Schmolze J.] Cognitive Science, — Vol. 9, — No.2, — 1985.
6. Extensible Markup Language (XML) 1.0. W3C Recommendation. [Bray T., Paoli J., Sperberg C.] — Feb 2005.
7. The Description Logic Handbook. Theory, Implementation and Applications. / [F. Baader, D. Calvanese, D. McGuinness, D. Nardi, Peter Patel-Schneider], Cambridge, 2007. — 574 pages.
8. OIL in a Nutshell. / [Fensel D., Horrocks I., Harmelen F., Decker S., Erdmann M., Klein M.]. 12th Intern. Conf. , Juanles-Pins, France. — 2000.
9. Knowledge Interchange Format. Technical Report. Computer Science Department. / [Fikes M.]. Stanford University. Logic-92-1. — 1992.
10. Inside the LOOM classifier. / [MacGregor R.], SIGART bulletin. — 1991. — Vol.3, No.2, — P.70-76.
11. The PROMPT Suite: Interactive Tools For Ontology Merging And Mapping Stanford Medical Informatics [Noy N., Musen M.], Stanford University. — August 2003. — SMI-2003-0973.pdf.