

ІНТЕГРАЦІЯ ФІЛОЛОГІЇ І ТЕХНІЧНИХ НАУК

УДК 658.012

А.Ю. ДОРОШЕНКО, студент, НТУ«ХП», Харків
САЙЄД МОХАММАД ТАУХІД СІДДІКІ, канд. техн. наук.,
НТУ«ХП», Харків
Н.В. ШАРОНОВА, докт. техн. наук, проф., НТУ«ХП», Харків

ВИКОРИСТАННЯ ОНТОЛОГІЙ ДЛЯ СЕМАНТИЧНОГО ПОШУКУ ДОКУМЕНТІВ

У статті розглядається методологія побудови онтологій, визначається поняття семантичного пошуку у зв'язку з онтологіями, пропонується використання семантичної мережі та діаграм потоків даних при пошуку.

В статье рассматривается методология построения онтологий, определяется понятие семантического поиска в связи с онтологиями, предлагается использование семантической сети и диаграмм потоков данных при поиске.

The article deals with methodology of construction of ontologies, defines the semantic search in connection with ontologies and it is proposed to use the Semantic Web and data flow diagrams in the search.

Вступ. Комп'ютерне моделювання процесів обробки текстів різної складності дасть змогу автоматизувати різні види інтелектуальної діяльності людини, розширити його можливості щодо прийняття ефективних управлінських рішень. Ефективність автоматизованих інформаційних систем (АІС) визначається їхньою здібністю оброблювати інформацію неформалізовану або слабо формалізовану. Задача створення пакетів ділової документації є однією з таких актуальних галузей обробки інформації [1, 2].

У межах окресленої проблеми важливим є наукове завдання розробки методів, моделей, алгоритмів і програм, які здійснюють моделювання процесів інтелектуальної обробки даних при створенні пакетів ділової документації з метою визначення їх основних характеристик для побудови інформаційно-

го, математичного, лінгвістичного і програмного забезпечення систем електронного документообігу. Один із методів, які застосовуються у галузі інформаційного пошуку, може бути пов'язаний з онтологіями.

Онтологія сприяє встановленню коректних зв'язків між значеннями елементів області тексту, тим самим створюючи умови для їх спільного використання. Онтології можна застосовувати в якості будівельних блоків компонентів баз знань, схеми об'єктів в об'єктно-орієнтованих системах, концептуальної схеми баз даних, структурованого глосарію взаємодіючих частин, словника для зв'язку між агентами, визначення класів для програмних систем.

1. Методологія побудови онтологій. Вона припускає розгляд наступних важливих питань: Позначення цілей і сфери застосування створюваної онтології. Для цього необхідно визначити, для чого створюється онтологія, і як вона надалі використовуватиметься. Побудова онтології включає в себе фіксацію знань про предметну область, яка включає: визначення основних понять і їх взаємин у вибраній предметній області; створення точних несуперечливих визначень для кожного основного поняття і відношення; визначення термінів, які пов'язані з цими термінами і стосунками; остаточне узгодження усіх вище названих етапів. Кодування, яке має на увазі розподіл сукупності основних термінів, використовуваних в онтології, на окремі класи понять; вибір або розробку спеціальної мови для представлення онтології; безпосередньо завдання фіксованої концептуалізації на вибраній мові представлення знань. Нині існує велика кількість проєктів, в основу яких покладені онтології [1-9].

2. Семантичний пошук. «Семантичний пошук – вид автоматизованого повнотекстового інформаційного пошуку з урахуванням смислового змісту слів і словосполучень запиту користувача і пропозицій текстів проіндексованих інформаційних ресурсів» [3].

Як і у будь-якій системі перекладу, виникає завдання зняття омонімії, що правильно відображає сенс текстового висловлювання.

Основна ідея semantic web при вирішенні цієї проблеми полягає в тому, щоб зробити для машини доступну семантику анотацій. Це повинно бути досягнуто шляхом використання онтологій – концептуальних схем, для надання формального значення термінів, використовуваних в анотаціях, перетворюючи їх на семантичні анотації.

Повна реалізація такої семантичної мережі потребує досить складних рішень і піднімає проблеми дослідження в таких областях як представлення знань, баз даних, комп'ютерної лінгвістики, комп'ютерного зору і агентних систем. Однією з таких проблем є досягнення компромісу між суперечливими вимогами для виразної сили мови, використовуваних в анотаціях і масштабованістю системи, використовуваної для їх обробки. Інша проблема полягає в тому, що інтеграція різних онтологій може виявитися принаймні так же важко, як інтеграція ресурсів, які вони описують. Нові

проблеми включають створення відповідних анотацій і онтологій і джерел веб-інформації різної якості.

Незважаючи на вищезгадані проблеми, значний прогрес був досягнутий в розвитку інфраструктури, необхідної для підтримки Semantic Web. Зокрема, стався прогрес в розвитку мов і інструментів для управління вмістом анотацій, а так само для розробки і впровадження онтологій [6, 8].

Можливості RDF були розширені шляхом підставлення додаткового значення до певних ресурсів. OWL-мова опису онтологій для семантичної павутини. Мова OWL дозволяє описувати класи і стосунки між ними, властиві для веб-документів і додатків.

Сучасні Web-технології пошуку засновані на повнотекстовому пошуку. Усі пошукові запити обслуговуються на основі індексу, що містить деякі описи входжень слів з відомих цій пошуковій системі документів. При цьому виникають різні проблеми – вибору того, що слід індексувати, забезпечення рівноправного індексування усього інформаційного простору, а також рішення, в контексті яких пошукових запитів слід видавати ту або іншу інформацію. Таким чином, семантична мережа (Semantic Web) – це концепція мережі, в якій кожен ресурс природною мовою був би забезпечений описом, зрозумілим комп'ютеру.

У семантичній мережі можна знаходити і об'єднувати дані з самих різних джерел, а також використовувати правила логічного висновку для оцінки цінності і якості знайдених джерел, перетворювати результати в придатну для аналізу форму.

3. Фундамент семантичної мережі. Існує три основні технології:

– специфікація XML (eXtensible Markup Language), що дозволяє визначити синтаксис і структуру документів;

– система онтологій, що дозволяє визначати терміни і стосунки між ними;

– механізм опису ресурсів RDF (Resource Description Framework), що забезпечує модель кодування для значень, визначених в онтології.

У семантичній мережі використовуються також інші технології і концепції, зокрема, універсальні ідентифікатори ресурсів URI (Uniform Resource Identifier), цифрові підписи, системи логічного висновку [4]. Онтології визначають терміни, використані для опису і представлення знань тій або іншій предметній області, зокрема – базових понять цієї області і зв'язків між ними. Можна сказати, що основна мета онтологій полягає в представленні сенсу понять, використовуваних в конструкціях RDF (фактично, RDF – це мова спілкування програмних систем, працюючих в середовищі Інтернет, а онтології складають його словник-тезаурус). Онтологія потрібна для людей, для додатків систем баз даних і різних інших інформаційних систем, які спільно використовують специфічну інформацію в певній предметній області. Онтології включають доступні для комп'ютерної обробки визначення основних понять предметної області і

зв'язку між ними. Вони забезпечують можливості повторного використання знань, можуть бути застосовані для пошуку інформаційних ресурсів в Інтернет і управління знаннями в цьому середовищі [3, 7].

Онтологія визначає загальну угоду про семантику конкретної області і сприяє встановленню коректних зв'язків між значеннями елементів області, тим самим створюючи умови для їх спільного використання. Онтології використовуються для підтримки автоматизованого обміну даними і для інтеграції додатків, механізми пошуку також застосовують онтології для вибірки сторінок із синтаксично різними, але семантично однаковими словами. Базова складова онтології визначає класи об'єктів і взаємодію між цими класами. Ключовими поняттями тут є поняття підкласу, суперкласу і спадкоємства. Наприклад, клас Студент є підкласом класу Людина (а Людина - суперклас класу Студент), оскільки для будь-якого об'єкту, якщо цей об'єкт є студентом, то він є і людиною. Засобами семантичного програмування класи визначаються як підмножини спадково-кінцевої надбудови, що мають спеціальні якості.

Онтології використовуються як механізм вираження і розподілу знань для визначення загального словарного запасу і підтримки інтелектуальних запитів в різноманітних базах даних. Онтології і метадані описують організацію і зміст ресурсів.

Основні завдання, які успішно вирішуються на базі онтологій, включають надання знань для виведення інформації, що відповідає запиту користувача, фільтрацію і класифікацію інформації, індексування зібраної інформації, організацію загальної термінології, якою можуть користуватися для комунікації програмні агенти і користувачі. Завдання семантичного пошуку в електронній бібліотеці є спрощеним аналогом пошуку інформації в Інтернет, таким чином передбачається, що пошук здійснюватиметься по запиту користувача природною мовою в аналогічному рядку пошуку.

4. Діаграма потоків даних при пошуку. На рис. 1 представлена схема семантичного пошуку інформації. Користувач вводить запит, який підлягає лінгвістичному аналізу, розширюється за рахунок використання синонімів, потім перетворюється в ключові слова і прямує до пошукової машини. Пошукова машина повертає знайдені документи, вони також піддаються лінгвістичному розбору і формуються семантичні образи документів. Образи документів порівнюються з образом запиту, робиться висновок про релевантність кожного з документів і результати аналізу надаються користувачеві. Схема лінгвістичного аналізу приведена на рисунку [5].

Висновок. Запропонована методика семантичного пошуку дозволяє відбирати близькі за загальним контекстом документи, навіть якщо вони належать до різних предметних областей. Тим самим можна збирати і узагальнювати знання, розосереджені в різних областях з допомогою семантичного пошуку на основі онтологій.

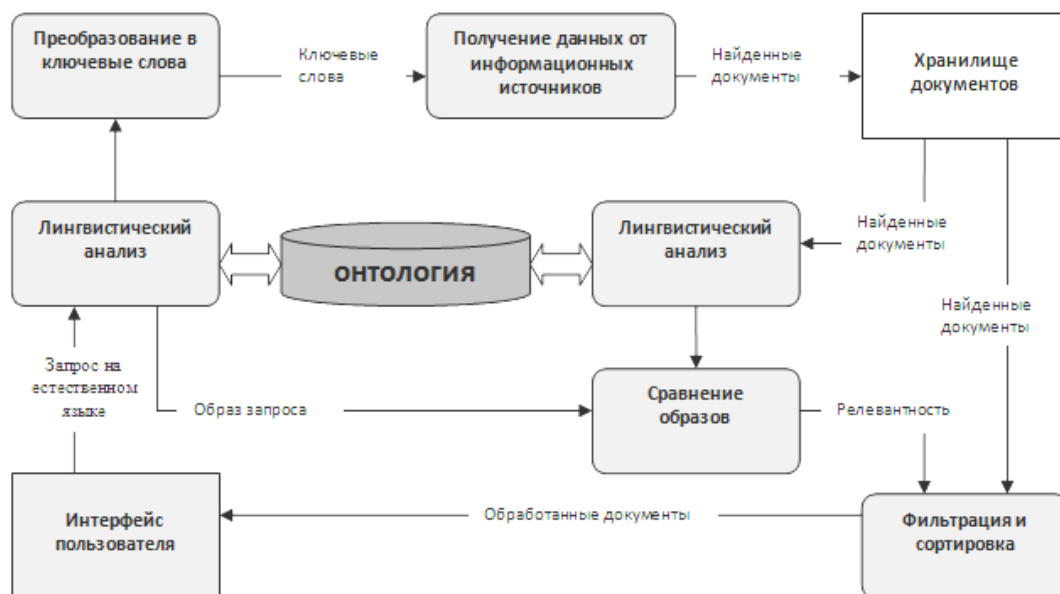


Рис. – Діаграма потоків даних при пошуку

- Список літератури:**
1. Udo Han. «Системы семантического поиска», 2009. – С. 19-25.
 2. Ланин В.В. Интеллектуальное управление документами как основа технологии создания адаптируемых информационных систем // Труды международной научно-технической конференций «Интеллектуальные системы» (AIS'07). Т. 2 / М.: Физматлит, 2007. – С. 334-339.
 3. Рябова Н., Козоплянська Г., Дяденко К. Застосування онтологічної семантики у зіставленні документів // Матеріали III Міжн. конф. молодих вчених «Комп'ютерні науки та інженерія» CSE-2009, 14-16 травня, 2009, Львів. – Львів: НУ «Львівська політехніка», 2009. – С. 107-108.
 4. Рябова Н.В. Методы и модели интеллектуальной обработки текстов в задачах онтологического инжиниринга // Математическое и программное обеспечение интеллектуальных систем (MPZIS-2008). Тез. докл. / VI междунар. науч.-практ. конф., Днепропетровск, 2008. – С. 269-270.
 5. Мальків М. В. Онтології в учбовому процесі [Електронний ресурс]: Освітні технології – Електрон.журн. – Освітні технології, 2007. – Режим доступу : http://www.naukarpro.ru/ot_2007/3_003.htm, вільний.
 6. Келеберда И.Н., Лісова Н.С., Рінка В. Би. Використання мультиагентного онтологічного підходу до створення розподілених систем дистанційного навчання. [Електронний ресурс]: Освітні технології і суспільство – Електрон.журн. – Освітні технології і суспільство, квітень 2004 - Режим доступу : http://ifets.ieee.org/russian/depository/v7_i2/html/3.html, вільний.
 7. Kravčev M. & Gašević D. Leveraging the Semantic Web for Adaptive Education. [Електронний ресурс]: Journal of Interactive Media in Education (Adaptation and IMS Learning Design. Special Issue, ed. Daniel Burgos), 2007/06. – Режим доступу: http://dspace.ou.nl/bitstream/1820/1080/1/kravcik_2007-06.pdf, вільний.
 8. Michal Laclavik1, Martin Seleng1, Emil Gatia11, Zoltan Balogh1, Ladislav Hluchy: Ontology based Text Search, 2004. – С. 189-211.
 9. Feldman R. and Hirsh H., «Mining associations in text in the presence of background knowledge», in Proc. 2nd Int. Conf. on Knowledge Discovery and Data Mining, Portland, USA, 1996. – С. 67-89.
 10. Канищева О.В. Идентификация документов в системах электронного документооборота / О. В. Канищева, Сайед Мохаммад Таухид Сиддики // Вестн. Херсонского нац. техн. ун-та. – Херсон: ХНТУ, 2007. – № 4(27). – С. 273-276.

Надійшла до редколегії 31.10.2011.