



ВИКОРИСТАННЯ ОНТОЛОГІЙ ДЛЯ СЕМАНТИЧНОГО ПОШУКУ ДОКУМЕНТІВ

Дорошенко А. Ю.

*Національний технічний університет
"Харківський політехнічний інститут",
м. Харків, вул. Пушкінська, 79/2, тел. 0932468666,
e-mail: marykate90@mail.ru*

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.

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.

Мета дослідження – запропонувати методіку семантичного пошуку, дозволяючи 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.

Пропонуємо методіку дослідження, яка полягає у наступних етапах.

1. Методологія побудови онтологій.

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 полягає у визначенні основних понять і їх взаємин у вибраній предметній області; створенні точних несуперечливих визначень для кожного основного поняття і відношення; визначенні термінів, які пов'язані з цими термінами і стосунками; остаточному узгодженні 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 безліч проектів, в основу яких покладені онтології.



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» [1].

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

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

3. Діаграма потоків даних при пошуку.

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 формує семантичні образи документів. Образи документів порівнюються з образами запиту, робиться висновок про релевантність кожного з документів і результати аналізу надаються користувачеві.

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 за допомогою семантичного пошуку на основі онтологій.

Список літератури

1. *Udo Han*, «Системи семантичного пошуку», 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.