

Наук. співробітник Михайлюк О.С., магістрант Сінкевич Ю. С.

**Національний технічний університету України
«Київський політехнічний інститут»**

ОНТОЛОГІЇ ЯК ЗАСІБ ПІДВИЩЕННЯ ЕФЕКТИВНОСТІ ПОШУКУ ІНФОРМАЦІЇ

Вступ

На сьогоднішній день глобальна інформаційна мережа Internet включає декілька десятків мільйонів серверів, на кожному з яких доступно від декількох мегабайт до десятків гігабайт інформації. Маючи потенційний доступ до такої кількості інформації, користувачеві хотілося б одержувати тільки потрібні йому об'єкти, у той час як сучасні пошукові системи працюють часто незадовільно, оскільки реалізують в основному механізми пошуку за ключовими словами, не враховуючи при цьому контекст. У підсумку результатом роботи таких пошукових систем стають сотні тисяч посилань, більшість з яких вказують на документи, що не становлять інтересу для користувача, тим самим ускладнюючи сприйняття й вибір корисної інформації.

Розв'язанням зазначеної проблеми є застосування таких методів пошуку, які допускають формулювання запитів природною мовою, мають здатність до аналізу семантики документів і запитів з наступним обчисленням їх семантичної близькості. Однак розробка подібних пошукових систем пов'язана як зі складнощами забезпечення розуміння пошуковими машинами природномовних запитів, так і зі складнощами реалізації універсальних алгоритмів семантичного аналізу пошукового відгуку. Тому дана проблема пов'язана з дослідженнями в галузі штучного інтелекту і є актуальною.

Постановка задачі

Ефективним засобом реалізації семантично-орієнтованого пошуку є онтології [1], котрі описують семантику наявних у мережі ресурсів. Онтологія може бути базовим інструментом семантичного аналізу, тим полем, у межах якого можна обчислювати смислову близькість семантичних інтерпретацій лексем тексту щодо найближчого оточення, тобто контексту. На формальному ж рівні, онтологія – це система, що складається з набору понять і набору тверджень про ці поняття.

Побудова онтологій базується на фіксації логічних конструкцій, які містять у собі словники термінів тематичної галузі та на описі визначень цих термінів, їх теоретично можливих і неможливих змістовних зв'язків. Основою для побудови онтології можуть бути запити користувача. З кожного запиту користувача, точніше, з його образу, виділяється деяка інформація, котра конвертується до бази знань та використовується для побудови онтології. Створення онтології – процес поступовий і залежить від насиченості інформації, що надходить впродовж тривалого проміжку часу діяльності користувачів. Завданням даної роботи є огляд деяких відомих шляхів реалізації пошуку за змістом із застосуванням апарату онтологій.

Застосування онтологій для реалізації пошуку природномовних текстів за змістом

Всі онтології зберігаються у базі знань, відсортовані по конкретних предметних галузях (рис. 1).



Рис. 1. Схема організації пошуку на основі онтологій [2]

При подібній організації пошуку можна виділити такі основні етапи [3]:

- 1) формування та розширення запиту з використанням онтологій;
- 2) виконання запиту пошуковою системою;

3) семантична фільтрація результатів даних з використанням бази онтологій.

На першому етапі відбувається побудова запиту до пошукової системи, шляхом його перетворення в сполучений логічними зв'язками набір термінів і понять, які безпосередньо використовуватимуться пошуковою системою. Також запит розширюється шляхом додавання синонімів та асоціацій. Так, наприклад, якщо запит користувача містить в собі поняття “семантична мережа”, то він доповнюється синонімічним поняттям “асоціативна мережа”, а також такими поняттями, які з тієї або іншої причини можуть викликати у користувача відповідні асоціації. Наприклад, деякі дослідники під поняттям “Система подання знань” мають на увазі “Експертну систему”.

Приклад запиту: Семантичний пошук інформації у мережі Internet.

Запит після обробки: (Семантичний | Асоціативний) & пошук & (інформації | даних) & у & (мережі Internet | глобальній інформаційній мережі | WWW).

Після виконання запиту над усіма знайденими пошуковою системою документами виконується семантична фільтрація з використанням онтологій і лінгвістичний аналіз (морфологічний, синтаксичний і семантичний) [4], після чого виявляються найбільш релевантні фрагменти тексту документа (релевантність - міра відповідності результатів пошуку запиту користувача [5]).

В даний час представлено багато цікавих і амбітних проектів, які в своїй основі використовують онтології. Серед них: TOVE (Toronto Virtual Enterprise), KACTUS (Європейський проект в рамках проекту ESPRIT), Сус (проект Корпорації Мікроелектроніки і Комп'ютерних технологій (MCC)) [2]. Мета кожного з проектів – створити величезну базу знань, що включає всі факти, доступні людській увазі.

Розглянемо докладніше один зі згаданих проектів, а саме проект Сус. Даний проект шляхом створення онтологій для широкого кола додатків, залежних від конкретної предметної галузі, формує лексико-семантичну базу для систем, які імітують механізми людської логіки. Сус включає величезні мультитекстні бази знань, ефективні механізми виведення і ряд спеціалізованих прикладних модулів. База знань, основу якої складає ядро (Сус Inference Engine) з 1.000.000 простих тверджень, що фіксують знання про світ, містить також правила виведення і правила керування. Знання в Сус подаються засобами спеціально розробленої декларативної мови СусL (Сус Representation Language) у вигляді тверджень логіки першого порядку. Онтології, які складають основу Сус, організовані у набори модулів, котрі називаються мікротеоріями. Кожна така мікротеорія фіксує знання і твердження деякої окремої галузі знань, наприклад, про простір, час тощо. Причому, для однієї галузі може існувати безліч мікротеорій, в яких представлені різні погляди і припущення, зроблені

людьми, які моделюють цю галузь. У цьому сенсі Сус не є монолітною онтологією, а є мережею мікротеорій для безлічі галузей.

Висновки

Використання методів пошуку, котрі передбачають застосування онтологій, є вельми ресурсоємним і потребує постійного нарощення бази знань. Тому одним з головних завдань цього науково-технічного напрямку є створення для пошукових систем можливостей по самостійному вдосконаленню, для чого в першу чергу необхідний зворотний зв'язок, котрий надає системі матеріали для аналізу результатів роботи. Розробка методів самонавчання для пошукової машини може призвести до створення принципово нового покоління пошукових систем, онтологічні описи ресурсів яких удосконалюються протягом життєвого циклу такої системи без прямої участі людини. Серед перспективних напрямків розвитку даної технології найбільше виділяється вдосконалення методів інтерактивного розширення природномовних запитів синонімічною й іншою семантично еквівалентною інформацією з метою підвищення повноти пошуку.

Література

1. *Майкевич М.В.* От информационного пространства к пространству знаний. Онтологии в Интернет. //Праці конференції по штучному інтелекту, 1998. - С. 152-158.
2. *Загорулько Ю.А.* Организация эффективного поиска на основе онтологий// Праці міжнародної конференції "Діалог" 2001, 2001. - С. 333-342.
3. *Жигалов В.О.* Составляющие интеллектуального поиска. //Праці міжнародної конференції "Діалог" 2003. –М.: Наука, 2003.
4. *Осипов Г.С., Тихомиров І.А., Смирнов І.В.* Интеллектуальный поиск в глобальных и локальный вычислительных сетях и базах данных. // Праці міжнародної конференції "Програмні системи: теорія та пропозиції". - Переславль-Залесский, 2004. - С. 21-34.
5. *Козлович О.О.* Поиск релевантной Информации в среде интернет. //Перспективні інформаційні технології і інтелектуальні системи, 2004. - С. 21-30.