

Магістрант Бездорожев Д.В., к.т.н., доцент Волкогон В.П.,

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

ІНФОРМАЦІЙНО-ПОШУКОВА СИСТЕМА З ВИКОРИСТАННЯМ ОНТОЛОГІЧНОГО АНАЛІЗУ ПРЕДМЕТНИХ ОБЛАСТЕЙ

Вступ

Діяльність окремих людей і організацій зараз усе більше залежить від наявної в них інформації й здатності її ефективно використовувати. Сьогодні у великих компаніях та на підприємствах, де є велика кількість робочих місць з комп'ютерами, які об'єднані у єдину мережу, поширена практика створення мережових репозиторіїв даних, призначених для загального використання співробітниками компанії. Та й самі співробітники теж не рідко обмінюються даними між собою шляхом безпосереднього відкриття мережевого доступу до своїх ресурсів. Всі ці ресурси можна представити як корпоративну базу знань. Вміння вчасно і досконало користуватися цими знаннями може забезпечити компанії перевагу над конкурентами, а швидкість і точність їх використання керівництвом компанії дозволить вижити та адаптуватися в економічному середовищі, яке постійно змінюється.

Але корпоративна база знань найчастіше знаходиться у розрізненому вигляді, у вигляді різних баз даних, електронних документів, листів тощо. Такий вигляд породжує декілька проблем. По-перше, важко організувати доступ до цих даних у зручному для користувача вигляді. По-друге, різні групи користувачів використовують різну термінологію для одних й тих самих понять, і навпаки, одне й те саме поняття має різні спеціальні терміни у різних груп користувачів.

Постановка задачі

Зважаючи на перелічені вище проблеми, які виникли при обробці великих масивів інформації, постає нагальна потреба у розробці системи з використанням онтологічного аналізатора, яка б виконувала наступні дії:

- пошук та вилучення інформації з корпоративної бази знань;
- структурування отриманої інформації;

- формалізацію інформації шляхом її опису мовами, зручними для машинної обробки;
- обслуговування формалізованих даних та надання користувачам можливості ці данні використовувати (додавання й оновлення бази знань, видалення застарілої інформації, фільтрування та й, власне, пошук інформації).

Опис методів побудови онтологій над предметною областю, визначеною конкретним користувачем

Для моделювання таких складних систем, як великі підприємства, корпорації та інші, була розроблена ланка методологій IDEF (Integrated DEFinition), яка містить у собі 14 державних стандартів США з комп'ютеризації промисловості. Для підтримки онтологічного аналізу призначена методологія IDEF5. Згідно з IDEF5, у загальному вигляді процес побудови онтологій має наступну послідовність:

- вивчення початкових умов та встановлення цілей розробки онтології;
- збір і накопичення початкових даних, які необхідні для побудови онтології;
- аналіз і групування даних;
- формування попередньої версії онтології;
- формування заключної версії онтології.

Кожна предметна область складається з об'єктів, які утворюють систему, та взаємозв'язків між цими об'єктами, що характеризують стан системи. Отже, при побудові онтології були вирішені наступні питання:

- створення і документування словника термінів;
- опис правил і обмежень, згідно з котрими на базі введених термінів формуються достовірні ствердження, які описують стан системи;
- побудова моделі, котра на базі існуючих стверджень дозволить сформулювати необхідні додаткові твердження.

Таким чином, щоб створити онтологію, користувач повинен задати кінцеву множину термінів предметної області, кінцеву множину відносин між цими термінами й кінцеву множину функцій їх інтерпретації, а потім указати, між якими саме термінами існують певні відношення. Онтологія предметної області може бути візуалізована у вигляді лісу орієнтованих графів з навантаженими дугами, у яких вершини відповідають термінам предметної області, а дуги - відношенням між ними.

Для побудови онтологій планується використати вже існуючі програмні рішення, а саме безкоштовну Java-програму Protege, яка спеціально призначена для побудови, редагування і перегляду заданої предметної області. Вона включає редактор онтологій, який дозволяє

проекувати онтології, розгортати ієрархічну структуру класів і понять. Цей інструмент підтримує використання мови OWL і дозволяє генерувати html-документи, які відображають структуру онтології.

Обробка результатів виконання інформаційних запитів з урахуванням контексту

Для того, щоб користувач мав можливість розпочати інформаційний пошук, йому необхідно надати непусту множину інформаційних ресурсів Q_1, \dots, Q_n , до яких він може звернутися. Потім користувач формує інформаційний запит. Спосіб виконання пошуку залежить від специфіки конкретного IP. У результаті виконання пошуку формується множина документів I , які ПС визначила релевантними запиту I_j , де I_j – результат пошуку в інформаційному ресурсі Q_j . На жаль, більшість ПС, що здійснюють пошук за ключовими словами, включають в I дуже багато непотрібної інформації – повтори, нерелевантні й застарілі посилання, а також посилання на документи, які вже відомі користувачеві. Щоб звільнити користувача від необхідності переглядати вручну всі ці документи, пропонується здійснити їх фільтрацію, використовуючи відомості про попередні запити цього користувача й сфери його інформаційних інтересів.

Обробка результатів виконання запитів складається з 6-и етапів:

Етап 1. У результаті виконання інформаційного запиту користувача до Q за ключовими словами формується множина I .

Етап 2. Якщо множина I не порожня, виконується впорядкування цієї множини за адресами інформаційних ресурсів. Інакше - завершення роботи.

Етап 3. Якщо отримана на етапі 2 множина I не порожня, відфільтровуються повторювані посилання. Інакше - завершення роботи.

Етап 4. Відфільтровуються застарілі посилання.

Етап 5. Якщо отримана на етапі 3 множина I не порожня, здійснюється перевірка за БД користувача, чи одержував він раніше кожне з посилань, що залишилися (якщо одержував, то рішення про те, чи залишати це посилання, залежить від того, як у минулому користувач повівся з цим посиланням, а також від інших його інструкцій). Інакше - завершення роботи.

Етап 6. Якщо сформована на етапі 5 множина I не порожня, виконується оцінка відповідності документів з цієї множини контексту пошуку. Інакше - завершення роботи.

Саме на шостому етапі використовується онтологія предметної області, яка була створена раніше користувачем. Тут контекст пошуку - це

непуста неупорядкована множина термінів і словосполучень, характерних, на думку користувача, для того ІР, що він хоче знайти. Так, наприклад, наявність в ІР термінів «монографія», «список літератури» і «анотація» підвищують ймовірність того, що розглянутий ІР - наукова праця.

Застосування користувацьких онтологій для задання контексту пошуку в першу чергу орієнтоване на користувачів, що мають постійні інформаційні інтереси в мережі й потребують постійного надходження відповідної інформації. Запити таких користувачів можуть повторюватися від сеансу до сеансу або змінюватися, але предметні області, у яких користувачі є експертами, практично не змінюються і є досить обмеженими. Опис цих предметних областей задається самими користувачами у вигляді онтологій. Один користувач може створювати кілька онтологій, якщо він має кілька прикладних областей, що не перетинаються, які його цікавлять.

Висновки

Сьогодні дані найчастіше є основою для прийняття важливих рішень, що впливають на роботу або навіть на виживання компаній. Інтенсивна інформатизація державних органів та комерційних структур, зростаюча доступність інструментарію для збору та моніторингу даних ведуть до появи надмірної кількості інформації, в якій важко знайти щось необхідне.

Ефективний пошук вчасно пропонує потрібну інформацію, позбуваючись інформаційного шуму, і часто виявляється вирішальним фактором для підвищення конкурентоздатності. Підвищення ефективності такого пошуку є сьогодні актуальним завданням. Цього можна домогтися шляхом обробки контексту запиту й відомостей про конкретного користувача, що послав запит, а також передісторії його звертання до різних ІПС.

Для формалізованого опису предметної області пошуку, до якої ставляться інформаційні запити користувача, доцільно використати онтологічний підхід. При цьому необхідно створити адекватні платформонезалежні інструментальні засоби для створення, модифікації й обробки онтологічних систем, які може застосовувати користувач, що не є фахівцем у галузі інформаційних технологій.

За прогнозами аналітичної компанії IDC, попит на подібні програми істотно зросте протягом найближчих 3-4 років. Так, до 2012 року очікується підвищення прибутків від такого ПЗ до \$ 1,5 млрд.. А в 2012 році такі системи будуть домінувати при проведенні аналізу інформації від клієнтів в компаніях будь-якого рівня.

Література

1. *Картышева Е.* Интеллектуальные поисковые системы Excalibur. – [Открытые Системы](#).
2. *Гаврилова Т.А., Хорошевский В.Ф.* Базы знаний интеллектуальных систем. – Спб.: Питер, 2001. – 382 с.
3. A Model-Theoretic Semantics for DAML+OIL.– [W3C Consortium](#).
4. W3C Web Ontology. – [W3C Consortium](#).
5. *Тузовский А.Ф., Козлов С.В.* Построение модели знаний организации с использованием системы онтологий. – [Конференция по компьютерной лингвистике «Диалог-2006»](#).