

3. Mailvaganam H. Introduction to OLAP / H. Mailvaganam. – Slice, Dice and Drill / Hari Mailvaganam (2007). DWreview. Retrieved 2008-03-05.
4. Plattner H. Online Analytical Processing / Prof. Dr. Hasso Plattner; [Trends and Concepts Lectures]. – Potsdam : HPI, 2008.
5. Plattner H. OLAP : Slicing and Dicing / Prof. Dr. Hasso Plattner; [Trends and Concepts Lectures]. – Potsdam : HPI, 2008.
6. Plattner H. Compressed and Optimized OLAP : Storing Attributes in Columns / Prof. Dr. Hasso Plattner; [Trends and Concepts Lectures]. – Potsdam : HPI, 2008.
7. Adjoined Dimension Column Index (ADC Index) to Improve Star Schema Query Performance / [P. E. O'Neil, X. Chen, E. J. O'Neil] ; In ICDE, 2008.
8. The Star Schema Benchmark (SSB) / [P. E. O'Neil, E. J. O'Neil, X. Chen]. – Available at: <http://www.cs.umb.edu/poneil/StarSchemaB.PDF>.
9. «Column Oriented Database» / Harvard Research Group. – Available at: <http://www.hrgresearch.com/ColumnDB.html>
10. ColumnStores vs. RowStores: How Different Are They Really? / [Daniel J. Abadi, Samuel R. Madden, Nabil Natchem]. – Available at: <http://db.csail.mit.edu/projects/cstore/abadi-sigmod08.pdf>.

A. Apostol

ONLINE ANALYTICAL PROCESSING SYSTEMS

In this article, the review of the OLAP technology is presented. OLAP is the technology of the information processing that includes the composing and dynamic publication of the reports and documents. It is used by the analysts for the fast processing of the complex database queries. The comparison of OLAP and OLTP, and the methods that increase the performance of the processing of complex queries are also presented in this work. As one of these methods, the technology of the OLAP cube is thoroughly examined. Much attention is also paid to the special type of data stores, column-oriented databases. In comparison to the traditional row-oriented approach to the storing of data, column-based databases give plenty of advantages while processing the complex queries. This theory is proved by the example in which the performance of C-Store, one of the most popular column-oriented DBMSs, and the performance of the commercial row-oriented DBMS are compared.

УДК 681.3:658.5

Олецкий О. В.

ОРГАНІЗАЦІЯ ОНТОЛОГІЧНО-ОРІЄНТОВАНИХ ЗАСОБІВ АВТОМАТИЗОВАНОГО ЕКСПЕРТНОГО ДОБОРУ ІНФОРМАЦІЙНИХ РЕСУРСІВ НА ТЕМАТИЧНОМУ ПОРТАЛІ

Проблему експертного добору інформаційних ресурсів розглянуто в аспекті онтологічно-орієнтованого пошуку на базі системи математичних співвідношень над вузлами онтології в рамках моделі «онтологія–артефакт–користувач–проект». Висвітлено основні компоненти системи, що реалізують такий пошук.

Вступ

Проблема пошуку інформаційних ресурсів, які б найбільш точно відповідали цілям користувача, зважаючи на його індивідуальні особливості, є винятково важливою й актуальною. Протягом тривалого часу розвиваються різні погляди як її розв'язати [1–3 та ін.], але вони здебільшого мають евристичний характер і залучають до розгляду лише окремі аспекти пошуку.

Навіть такі ключові поняття, як «релевантність документа запитові», «схожість документів», формалізовано недостатньо.

Окрему увагу слід звернути на веб-орієнтовані інформаційні ресурси, для яких характерні висока інформаційна зв'язність, тематична однорідність, достатньо висока структурованість та якість інформаційного наповнення. До таких ресурсів належать, зокрема, тематичні портали,

в тому числі навчального спрямування. Для подібних систем особливого значення набуває максимальна орієнтованість на семантику, онтологію предметної області. Тоді можна говорити уже не про звичайний інформаційний пошук за «класичними» схемами, а про автоматизований експертний добір найбільш адекватних інформаційних ресурсів, і такі системи набувають експертно-консультаційних рис. Застосування такого підходу до порталів навчального спрямування дасть змогу повною мірою реалізувати сучасні методики *blended learning* та колаборативного навчання [4, 5 та ін.]. Не викликає сумнівів і перспективність застосування такого пошуку для більш традиційних веб-орієнтованих ресурсів (зокрема, для електронних крамниць).

В роботі [6] розглядається підхід, спрямований на формалізацію моделі інформаційного пошуку як процесу хвильового поширення активації на деякому графі. Цей граф як модель інформаційного наповнення системи описується в [6] у вигляді четвірки «онтологія–артефакт–користувач–проект». В загальних рисах, до розгляду залучаються різні типи зв'язків між поняттями предметної області, з одного боку, і документами, що зберігаються в системі, – з іншого, а міри важливості цих зв'язків залежать від цілей відвідувачів та від їхніх індивідуальних характеристик.

Безумовно, ці базові положення потребують уточнень і формалізацій через з'ясування формалізованих співвідношень і відповідних постановок задач. Не викликає сумнівів і те, що априорно визначити найбільш адекватні параметри цих співвідношень (наприклад, міри важливості зв'язків) часто буває неможливим, тому параметри мають підбиратися експериментальним шляхом. Тому виникає необхідність у проектуванні та розробці інструментальних засобів для проведення такого експерименту. Саме ці питання і досліджуються в цій роботі.

Як базова модель в [6] розглядається трійка $M = \langle W^*, D, L \rangle$, де W – онтологія предметної області, W^* – розширена онтологія, наповнення онтології W конкретними екземплярами класів (фактично – база знань), D – множина документів; L – множина зв'язків між W^* та D . Власне онтологія описується як трійка $\langle Q, R, F \rangle$, де Q – множина класів, які відповідають поняттям предметної області, R – множина зв'язків між ними, а F – множина функцій інтерпретації. Відповідно, розширена онтологія позначена як трійка $\langle Q^*, R^*, F^* \rangle$, де Q^* – множина класів разом з їх екземплярами, R^* – множина зв'язків між цими елементами, а F^* – множина функцій інтерпретації, визначених у найпростішому випадку на елементах з Q^* , R^* та $Q^* \times R^* \times F^*$. Тоді елементи D можуть бути значеннями функцій з F^* . По суті, така формалізація окреслює граф,

вузли якого відповідають поняттям предметної області та інформаційним ресурсам, а дуги – зв'язкам між ними, причому ці зв'язки можуть бути різних типів.

Зважаючи на цілі відвідувачів порталу та їхні характеристики, у рамках цієї моделі можна здійснювати на основі «занурення» в неї відповідної інформації. Ідея «занурення» інших категорій сутностей в загальних рисах формулюється так: якщо w є елементом розширеної онтології, а d – артефактом інформаційної системи, то функції інтерпретації f та відповідні вагові коефіцієнти можуть формуватися на основі цих категорій сутностей. Таким чином, здійснюється перехід до моделі «онтологія–артефакт–користувач–проект», де міри важливості зв'язків залежать від характеристик та цілей відвідувачів.

На основі описаних положень можна ввести ряд співвідношень, зокрема:

- співвідношення, спрямовані на опис мір подібності понять та інформаційних ресурсів, а також релевантності документів вузлам онтології;
- співвідношення, які визначають перерозподіл певного ресурсу (наприклад, мір важливості або авторитетності окремих вузлів).

Нехай W – множина понять предметної області, D – множина артефактів інформаційної системи, Q – задана множина можливих типів зв'язків, зокрема між поняттями предметної області, а також між поняттями предметної області та артефактами інформаційної системи. Позначимо через $r_q(w, d)$, де $q \in Q$, $w \in W$, $d \in D$, міру релевантності документа d поняттю w за зв'язком q .

Можна навести методики для розрахунку мір $r_q(w, d)$ як на основі певних узагальнень класичної векторно-матричної моделі, так і на основі теоретико-множинного підходу. Твердження «міра релевантності документа запиту q дорівнює x », по суті, еквівалентне твердженню «документ належить до множини документів, релевантних запиту q зі ступенем належності x ». Тому міри подібності та релевантності можуть мати нечітку інтерпретацію, і до побудови таких мір природно застосовувати розвинену теорію нечітких множин.

Природно залучити до розгляду деяку комбіновану міру релевантності документа d поняттю w , усереднену за всіма зв'язками з урахуванням їх вагових коефіцієнтів:

$$R(w, d) = \sum_{q \in Q} \alpha_q r_q(w, d), \quad (1)$$

де α_q – вага (змістовно – міра важливості) q -го типу зв'язків.

Другий тип співвідношень пов'язаний зі спробою опису моделі самоорганізації порталу

на основі певної системи рівнянь – алгебраїчних або диференціальних (подібно до того, як у вигляді таких рівнянь прийнято описувати поширення певної речовини в просторі). Змістовно це означає, що зміна оцінки деякого вузла тягне за собою зміну оцінок суміжних з ним вузлів. Якщо, як і раніше, розглядати суміжність за різними типами зв'язків, то можна записати таку систему співвідношень відносно мір важливостей вузлів:

$$m(u) = \sum_{v \in G(u)} \beta_{uv} m(v). \quad (2)$$

Тут $m(u)$ – міра важливості вузла u ; $G(u)$ – множина вузлів, суміжних з u , β_{uv} – коефіцієнти, які характеризують міру важливості відповідного зв'язку.

Можна розглядати і інші типи модельних співвідношень; у будь-якому випадку, одним із ключових завдань навчання і самоорганізації пошукової системи стає добір коефіцієнтів, аналогічних α_q формули (1) та β_{uv} формули (2).

Таким чином, важливими компонентами системи, яка реалізує пошук на онтологічно-орієнтованому тематичному порталі, мають стати:

- база знань, яка описує онтологію предметної області – важливі сутності та зв'язки між ними на основі вищенаведених формалізацій. Оскільки така інформація є недостатньо структурованою, видається доцільним будувати інформаційне наповнення системи на основі триад «об'єкт–атрибут–значення»;
- інформація про інформаційні ресурси та про їхні зв'язки з вузлами онтології; на найбільш загальному рівні можна виділити наступні типи таких ресурсів: власні матеріали, посилання на зовнішні URL, посилання на статті, посилання на книги;
- класи, які описують модельні співвідношення (1), (2) або подібні їм;
- класи, які реалізують ті чи інші методики оптимізації, навчання і самоорганізації. Сьогодні інтенсивно розвиваються нові інтелектуальні методики розв'язання перебірних пошукових задач, зокрема на основі інформаційного керування випадковим пошуком, а також механізмів, що імітують природні еволюційні процеси. В рамках цих методик перспективним видається застосування генетичних алгоритмів [7–9 та ін.] для знаходження оптимальних (або субоптимальних) наборів параметрів, а також методик, характерних для ройового інтелекту – зокрема, «алгоритму мурашки» [10, 11]. Слід зазначити, що використання алгоритму мурашки по суті тяжіє до імітаційного моделювання, оскільки в даному контексті він дає змогу зімітувати поведінку великої кількості відвідувачів;

- інструментальні засоби для експериментальних досліджень, зокрема для імітаційного моделювання;
- засоби Data Mining для інтелектуального аналізу даних, наприклад [6]: прийняття рішень за аналогією (зокрема, якщо користувач A для розв'язання задачі C вважає корисним документ W , то користувачеві X , характеристики якого схожі на характеристики користувача A , для розв'язання задачі K , схожої на C , можна порекомендувати список документів, схожих на W); автоматичне визначення відповідних функцій інтерпретації формальної моделі та пов'язаних з ними вагових коефіцієнтів на основі документів, відібраних користувачем тощо.

Власне онтологічно-орієнтований пошук при цьому здійснюється за такою схемою: аналіз запиту та його зіставлення з певною множиною вузлів онтології; пошук релевантних документів та упорядкування результатів; надання результатів користувачеві. Керування переміщенням відвідувача порталом здійснюється або на основі власне навігаційних засобів (гіпертекстові посилання при цьому формуються динамічно на основі онтології предметної області), або за запитом користувача. Варто зауважити, що особливої принципової різниці між цими типами керування немає; але якщо при навігаційному керуванні можна явно і точно встановити потрібний контекст, то запит відвідувача потрібно тим чи тим чином аналізувати. Доцільною видається також організація порталу взагалі і системи експертного підбору матеріалів зокрема на основі «класної дошки».

Такий портал може природним чином набути рис віртуальної спільноти. У цьому контексті важливого значення набувають персоналізація, підтримка персональних робочих середовищ та взаємодія між ними. Слід зазначити і те, що в рамках такої спільноти відвідувачі повинні мати можливість додавати власні інформаційні матеріали й оцінювати їх, а це має суттєве значення для динамічного перерахунку мір важливості вузлів та документів з урахуванням того, що ці міри для кожного відвідувача може бути індивідуальними, а також для перерахунку мір компетентності експертів. У [6] наголошено, що доцільним є зближення підходів, в рамках яких можна описати власне процес пошуку, з одного боку, та формування віртуальної спільноти, з іншого; обговорюються деякі підходи до цього.

Висновки

Тематичні портали, для яких характерні висока інформаційна зв'язність, тематична однорідність, достатньо висока структурованість та якість інформаційного наповнення, надають ши-

роки можливості для організації експертного добору інформаційних ресурсів, релевантних запитові відвідувача з урахуванням його мети та індивідуальних особливостей. Такий пошук має суттєво спиратися на онтологію предметної області та потребує відповідних формалізацій. У роботі наведено певні математичні співвідношення, які можна використовувати для пошуку і самоорганізації порталу в рамках формальної моделі «онтологія–артефакт–користувач–проект».

Коефіцієнти цих співвідношень мають налаштуватися, зокрема на основі навчання та експериментальним шляхом, і пошукова система має передбачати для цього розвинені інструментальні засоби. В роботі в загальних рисах обговорюється архітектура і основні компоненти програмної системи, яка реалізує ці можливості. Подальшим кроком має стати розвиток відповідних моделей в напрямку побудови віртуальної спільноти.

1. Ландэ Д. В. Поиск знаний в Интернет / Д. В. Ландэ. – М. : Изд. дом «Вильямс», 2005. – 272 с.
2. Гаврилова Т. А., Хорошевский В. Ф. Базы знаний интеллектуальных систем / Т. А. Гаврилова, В. Ф. Хорошевский. – СПб: Питер, 2000. – 384 с.
3. Сэлтон Дж. Автоматическая обработка, хранение и поиск информации / Дж. Сэлтон. – М. : Сов. радио, 1973. – 560 с.
4. Boublik V., Hesser W., Schmidt-Braul I. E. Towards cooperative e-teaching and e-learning / V. Boublik, W. Hesser, I. E. Schmidt-Braul // Теоретичні та прикладні аспекти побудови програмних систем. Матеріали міжнародної конференції ТАAPSD'2006, грудень 2006 р. – К. – С. 231–235.
5. Бублик М. М., Глибовець М. М., Олецький О. В. Моделі трансформації інформаційної освіти в контексті руху до інформаційного суспільства : досвід факультету інформатики НаУКМА / М. М. Бублик, М. М. Глибовець, О. В. Олецький // Наукові праці. Науково-методичний журнал. Педагогічні науки. – Миколаїв : Вид-во МДГУ ім. П.Могили. – Т. 71. – Вип. 58. – С. 60–64.
6. Олецький О. В. Онтологічно-орієнтований інформаційний пошук на основі хвильового процесу поширення активації / О. В. Олецький // Наукові записки НаУКМА. Комп'ютерні науки. – 2008. – Т. 86. – С. 50–52.
7. Рутковская Д., Пилинский М., Рутковский Л. Нейронные сети, генетические алгоритмы и нечеткая логика / Д. Рутковская, М. Пилинский, Л. Рутковский. – М. : Горячая линия – Телеком, 2004. – 452 с.
8. Глибовец Н. Н., Медведь С. А. Генетические алгоритмы и их использование для решения задач составления расписания / Н. Н. Глибовец, С. А. Медведь // Кибернетика и системный анализ. – 2003. – №1. – С. 95–108.
9. Глибовець М. М., Олецький О. В. Про деякі підходи до проблеми інформаційного керування випадковим пошуком / М. М. Глибовець, О. В. Олецький // Dynamical System Modelling and Stability Investigation. Thesis of Conference Reports, May 22–25. – 2007. – С. 370.
10. Тарасов В. Б. От многоагентных систем к интеллектуальным организациям : философия, психология, информатика / В. Б. Тарасов. – М. : Эдиториал УРСС, 2002. – 352 с.
11. Джонс М. Т. Программирование искусственного интеллекта в приложениях / М. Т. Джонс. – М. : ДМК Пресс, 2004. – 312 с.

O. Oletsky

BUILDING ONTOLOGIC-ORIENTED TOOLS FOR SEARCHING INFORMATION ON THE THEMATIC PORTAL

The problem of expert selecting relevant information resources is regarded as the problem of ontological search on the base of certain mathematical relations in the framework of the model «ontology–artifact–user–project». Main components of such system are discussed.