

ОЛЕЦЬКИЙ О.В.,

*факультет інформатики Національного Університету «Києво-Могилянська Академія»,
кандидат технічних наук, доцент
E-mail: oletsky@ukma.kiev.ua*

ПРИНЦИПИ ПОБУДОВИ ФОРМАЛІЗОВАНИХ ОНТОЛОГІЧНО-ОРІЄНТОВАНИХ МОДЕЛЕЙ ОРГАНІЗАЦІЇ ВІРТУАЛЬНИХ СПІЛЬНОТ У РАМКАХ ВЕБ-ПОРТАЛУ

Формування і розвиток віртуальних спільнот на базі корпоративних, навчальних, тематичних та інших веб-порталів є одним з найбільш магістральних напрямків розбудови сучасного Інтернету, і більш загально - інформаційного суспільства; особливо в рамках ідеології та концепції Web 2. З огляду на домінування в цій царині стихійно-евристичних підходів, все більш актуальною стає проблема розвитку більш-менш формалізованих теоретичних моделей, на основі яких можна було б описувати процеси, пов'язаних з організацією і функціонуванням віртуальної спільноти, в найпростішому випадку - на базі окремого порталу.

Оскільки віртуальна спільнота виникає і розвивається внаслідок взаємодії окремих користувачів, то очевидно, що в основі її організації мають лежати персональні робочі середовища окремих користувачів (які можна охарактеризувати як персональні робочі кабінети). Такі середовища мають бути максимально адаптованими до індивідуальних потреб; поєднувати в собі засоби керування особистими документами, планування завдань тощо. Відповідно до цього, формалізована модель віртуальної спільноти має описувати, як мінімум, наступні компоненти [1]:

- організацію окремого персонального середовища;
- зв'язки між окремими персональними середовищами та можливу взаємодію між ними;
- вплив загальної інфраструктури порталу на створення, наповнення, функціонування та підтримку персонального робочого середовища;
- вплив окремих персональних середовищ на формування віртуальної спільноти в цілому, і більш загально - на інфраструктуру порталу.

На основі подібних формалізацій може розглядатися ряд важливих і цікавих задач, зокрема:

- ефективний інформаційний пошук, який залучає до розгляду не просто міри релевантності документа запиту, а індивідуальні міри релевантності для кожного користувача з урахуванням завдань, які перед ним стоять, та проблем, які його дійсно цікавлять; деякі підходи до організації такого пошуку обговорювалися, зокрема, в [2];
- інтелектуальний аналіз даних (Data Mining), наприклад, динамічне користування груп користувачів, які мають спільні або схожі інтереси; автоматизований пошук можливих партнерів на основі аналізу мір схожості між персональними робочими середовищами;
- децентралізоване керування процесом формування спільноти та структурою порталу на основі принципів, що обговорювалися, зокрема, в [1]; у цьому контексті дуже перспективними видаються моделі ройового інтелекту; в ряді випадків можуть застосовуватися і більш прості моделі - наприклад, автомати з лінійною тактикою.

У цьому контексті необхідним стає підвищення інформаційної зв'язності порталу. Йдеться, зокрема, про те, що якщо в поточний момент часу на порталі відображається матеріал з певного питання, то користувачеві має бути надана можливість швидкого переходу до споріднених матеріалів, пов'язаних з ним за тематичною або іншими ознаками. Цьому повинен максимально сприяти механізм динамічного формування гіпертекстових посилань, реалізований на порталі.

Ключовим тут є максимальне врахування онтології предметної області. Важливо розвивати концептуальні моделі, які дозволяють пов'язати в єдине ціле множину документів, що складають основу інформаційного забезпечення веб-орієнтованих систем, з одного боку, та семантику предметної області, з іншого боку, на основі онтологічно-орієнтованого підходу, зокрема, на основі формальних моделей онтологій [3-7].

Тому розглядається підхід, в якому первинною стає формалізована модель "онтологія-артефакт", в яку "занурюються" інші категорії сутностей. Модель "онтологія-артефакт" описана в [7] наступним чином.

Модель інформаційного наповнення веб-орієнтованої системи розглядається як 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 . Відповідно до цього, якщо 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, то розширена онтологія описується 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^* . Іншими словами, документ d вважається релевантним відносно W^* , якщо існують хоча б один вузол w та функція інтерпретації f , такі що $d=f(w)$.

Наведена формалізація може бути легко розширена до четвірки "онтологія-артефакт-користувач-проект". Іншими словами, йдеться про розгляд формальних моделей, які дозволили б встановлювати відповідність між темами та пов'язаними з ними документами, з одного боку, та з користувачами та проектами і роботами, в яких вони беруть участь, з іншого. Подібні формалізації на базі четвірки "онтологія-артефакт-користувач-проект" можуть лягти в основу опису організації віртуальних спільнот на базі веб-порталу та окремих персональних робочих середовищ, які складають основу цих спільнот. Більш формалізовано, може бути запропонований наступний підхід. Якщо w є елементом розширеної онтології, а d - артефактом інформаційної системи, то функція інтерпретації f може формуватися на основі множин користувачів та проектів.

На цій основі можна визначати не просто міри релевантності документа запитові, а індивідуальні міри релевантності для кожного користувача, а також для окремих завдань і проектів. Зокрема, перспективним видається отримувати міри релевантності на основі процесу поширення активації між відповідними вузлами, застосування якого до задачі інформаційного пошуку було описано в [8]. Ініціюється певний хвильовий процес пошуку, під час якого активація вузлів може здійснюватися в чотирьох просторах:

- пошук в просторі документів, при цьому міри спорідненості можуть задаватися явним чином або визначатися динамічно; наприклад, на основі просторово- векторної моделі [9, 10];
- пошук на графі, який задає розширену онтологію;
- пошук на множині користувачів;

- пошук на множині окремих завдань і цілих проектів.

При цьому розрахунок мір релевантності здійснюється на основі коефіцієнтів, що пов'язуються з окремими типами зв'язків між вузлами розширеної онтології, а також з окремими типами функцій інтерпретацій. В результаті будуть сформовані множини понять і документів, в тій чи іншій мірі релевантних запиту, з урахуванням формування функцій інтерпретації за вищенаведеним принципом. В типовому випадку кожний елемент цих множин буде поданий у вигляді $(u, t_1(u), \dots, t_r(u))$, де u - знайдений вузол, $t_i(u)$ - міра релевантності цього вузла, обчислена за i -м критерієм, можливо, недостовірна або нечітка. Припускається навіть динамічне породження нових критеріїв, якщо задати процедуру такого породження.

Основна проблема, яка виникає при цьому, пов'язана зі значними обсягами інформації, і відповідно - зі значною часовою складністю. Тому ключовим є наступне питання: як спрямувати процес поширення активації в потрібному напрямку, і яким повинен бути критерій зупинки цього процесу? Важливою є також проблема комбінування критеріїв, яка полягає в тому, щоб перейти від кількох мір релевантності документа за різними критеріями до однієї комбінованої міри релевантності.

Для розв'язання перелічених проблем видається доцільним застосовувати методики випадкового керування інформаційним пошуком [11]; зокрема генетичні алгоритми, які добре зарекомендували себе для розв'язання ряду перебірних задач [13, 14]. В контексті, який розглядається, можна виділити як мінімум два аспекти застосування цих алгоритмів:

- власне для вибору найбільш перспективної підмножини документів, серед яких документи, потрібні користувачеві для вирішення його конкретної задачі, будуть міститися з максимальною ймовірністю;
- для експериментального підбору параметрів хвильового процесу поширення активації.

Важливим є і те, що описаний хвильовий процес поширення активації з урахуванням ефективних процедур його припинення, а також імовірнісних та нечітких мір, може розглядатися не тільки як інструмент пошуку, а й як концептуальна основа власне для самоорганізації та формування самих віртуальних спільнот, і відповідно - персональних робочих середовищ в рамках цих спільнот.

ЛІТЕРАТУРА

1. Олецкий О.В. До проблеми організації онтологічно-орієнтованих персональних робочих середовищ в рамках віртуального співтовариства. // Системи підтримки прийняття рішень. Теорія і практика. Матеріали 4-ї дистанційної науково-практичної конференції з міжнародною участю СППР'2008. Київ, червень 2008. - С.81-83. - WWW: http://conf.atsukr.org.ua/nles/conf_dir_8/oletskey_sppr2008.pdf.
2. Олецкий О.В. До проблеми онтологічно-орієнтованого пошуку в інформаційних системах. // Теоретичні та прикладні аспекти побудови програмних систем. Матеріали міжнародної конференції ТАAPSD'2007. Бердянськ, 4-9 вересня 2007 р. - С.73-77.
3. Гаврилова ТА., Хорошевский В.Ф. Базы знаний интеллектуальных систем. - СПб: Питер, 2000. - 384 с.
4. Плскач В.Л., Рогушина Ю.В. Агентні технології. - К.: Київ. нац. торг.-екон. ун-т, 2005. -338 с.
5. Проскудіна Г.Ю., Овдій О.М. Онтології в інформаційних системах // Теоретичні та прикладні аспекти побудови програмних систем. Спеціальний випуск Вісника Київського національного університету ім.Т.Г.Шевченка, 2004. -С. 164-169.
6. Олецкий О.В. Застосування формальних моделей онтологій для формалізації інформаційних потоків у системах управління контентом. // Теоретичні та прикладні аспекти побудови програмних систем. Матеріали міжнародної конференції ТАAPSD'2005, Київ, 7-9 грудня 2005 р. - С. 26-29.
7. Діренко І.С., Олецкий О.В. Система управління вмістом веб-ресурсів на основі онтологічно-документного моделювання //Теоретичні та прикладні аспекти побудови програмних систем. Матеріали міжнародної конференції ТАAPSD'2006. Київ, грудень 2006 р. – С. 171-176.
8. Глибовець М.М. Моделі та методи створення і супроводу високопродуктивного розподіленого навчального середовища. Автореферат дисертації на здобуття наукового ступеня доктора фізико-математичних наук. - Національний університет «Києво-Могилянська Академія», Київ, 2006.
9. Сэлтон Дж. Автоматическая обработка, хранение и поиск информации. - М.: Сов.радио, 1973. - 560 с.
10. Ландэ Д.В. Поиск знаний в Интернет. - М.: Изд. дом «Вильяме», 2005. - 272 с.
11. Глибовець М.М., Олецкий О.В. Про деякі підходи до проблеми інформаційного керування випадковим пошуком //Dynamical System Modelling and Stability Investigation. Thesis of Conference Reports, May 22-25, 2007. -С.370.

12. Глибовец Н.Н., Медведь С.А. Генетические алгоритмы и их использование для решения задач составления расписания//Кибернетика и системный анализ, 2003, №1. - С.95-108.
13. Рутковская Д., Пилиньский М., Рутковский Л. Нейронные сети, генетические алгоритмы и нечеткая логика. - М.: Горячая линия - Телеком, 2004. - 452 с.