

A. Afonin, A. Glybovets, O. Bobko

COORDINATING DISTRIBUTED SOFTWARE DEVELOPMENT

This article describes a deep research of problems and characteristics of distributed software development project management systems, the levels of cooperation between members of distributed teams investigated the role of organizational memory. It was made a comparative analysis of known software products of distributed UPC support. It proposes criteria to compare these products.

Keywords: software, distributed development, team development, comparative analysis, methodology of software development, project management.

Матеріал надійшов 12.05.2014

УДК 004.891.2

Глибовець М. М., Конюшенко О. В.

ВИКОРИСТАННЯ ДІЯЛЬНОСТІ ВІРТУАЛЬНИХ СПІЛЬНОТ ДЛЯ ПОКРАЩЕННЯ ІНФОРМАЦІЙНОГО ПОШУКУ

У статті обґрунтовано задачу розробки системи персоналізованого пошуку, яка здійснювала б пошук документів тематичного порталу з урахуванням результатів кооперативної діяльності віртуального співтовариства порталу за оцінкою документів.

Ключові слова: персоналізований пошук, віртуальні спільноти, генетичні алгоритми, тематичні портали.

Вступ

У зв'язку зі збільшенням об'єму даних, розміщених в Інтернет і корпоративних базах даних, актуальною стає проблема підвищення якості пошукових і навігаційних систем [2]. Наразі більшість пошукових систем можуть дізнаватися про інформаційну потребу користувача тільки за допомогою ключових слів, заданих у запиті. Причому, система не має в своєму розпорядженні відомостей про контекст, в якому користувач вживає ці слова. Та й самому користувачеві часом важко підібрати ключові слова, які точно описують тематику, що його цікавить. Пошукова система надає користувачеві результати, відсортовані за релевантністю до запиту. При цьому враховується текст документів або структура посилань між ними, але не враховується ряд інших факторів, які могли б покращити якість пошуку [1; 5], зокрема:

- коло інтересів і індивідуальні особливості користувача, який веде пошук;
- результати кооперативної діяльності віртуального співтовариства за оцінкою якості ресурсів.

Перший з цих факторів можна охарактеризувати як проблему впровадження персоналізованого пошуку, другий – поширення впливу результатів діяльності одного користувача (зокрема, оцінювання якості ресурсів) на результати пошуку інших, що в перспективі може призвести до підвищення ефективності пошуку.

Актуальність досліджень у цій галузі підтверджується як діяльністю багатьох наукових спільнот у цьому напрямі, так і впровадженням систем, які реалізують ці ідеї, великими корпораціями, що представлені на ринку пошуку інформації. Зокрема, не так давно компанія Google представила свій сервіс персоналізованого пошуку «Вікіпошук»

(SearchWiki) [14]. Але сервіс має низку обмежень, спричинених природою Інтернет. Тому прийнято рішення про проведення досліджень у цій галузі на базі окремого веб-порталу.

Мета роботи полягає у розробці основних принципів функціонування та впровадженні системи персоналізованого пошуку, яка дозволяла б здійснювати пошук документів тематичного порталу з урахуванням результатів кооперативної діяльності віртуального співтовариства порталу за оцінкою якості документів.

Персоналізований пошук

Щодня пошукові системи отримують запити від користувачів, які вибирають результати, що їх зацікавили. Пошукові системи збирають, систематизують і видають інформацію у вигляді цифр статистики: популярності запитів, інтересів, рейтингів тощо. Зрозуміло, цифри загальні і не стосуються конкретної людини. Але при впровадженні засобів однозначної ідентифікації користувача дістаємо можливість збору інформації про конкретного користувача. Спеціалізована обробка цієї інформації дозволяє дізнатися багато цікавого про користувача, що може стати привабливим при формуванні індивідуальних результатів видачі на запит. Сама процедура пошуку значно спрощується, оскільки пошукова машина вже «знає» багато чого про його потреби та інтереси. Це і називають персоналізацією пошуку [8].

Головна ідея цього підходу полягає в тому, що різні користувачі не повинні отримувати однакові результати при однакових запитах. Незважаючи на початкове незвичне звучання, ідея про те, що при кожному запиті мусить враховуватися контекст (*хто запитує, що користувач запитував до цього*), видається дуже слушною. Способи реалізації ідеї різні.

Так, наприклад, пошукова система *collarity.com* веде облік усіх пошукових запитів користувача, дозволяючи підібрати найвідповідніші результати, спираючись як на історію запитів, так і на досвід інших учасників системи. У ній зручно реалізовано можливість управління персоналізацією пошуку – можна вибрати, на що спиратимуться результати: на власну історію запитів, на сукупну історію запитів всього співтовариства користувачів або результати видаватимуться взагалі без прив'язки до чийогось досвіду.

Представником іншого способу персоналізації пошуку є система Rollyo. Вона дозволяє користувачам створювати і публікувати власні пошукові машини, включаючи в пошуковий індекс лише ті сайти, які користувачеві цікаві. Також

система дозволяє легко знаходити і користуватися персональними пошуковими системами, створеними іншими користувачами. Ідея персоналізованого пошуку такого типу здається дуже перспективною і вже активно реалізовується і застосовується [6].

Персоналізований пошук має багато переваг. Головна з них полягає в тому, що такий пошук практично знімає проблему вибору, оскільки завжди допомагає вибрати краще. А це вже економить час, гроші тощо. Персоналізований пошук знає про інформаційні потреби і переваги користувача, а значить, завжди запропонує цікаву, актуальну інформацію. Наприклад, персоналізована стрічка новин міститиме лише те, що цікаво конкретному користувачеві, персоналізована пошта, як би не дивно це звучало, здатна відбрати цікавий конкретному користувачеві спам.

Актуальність технології персоналізованого пошуку підтверджується і зацікавленістю та певними розробками відомих компаній, які присутні на ринку надання сервісів інформаційного пошуку. Зокрема в своєму інтерв'ю керівник групи мережного спаму компанії Google Метт Каттс заявив: «Я вважаю, що персоналізація має широкі можливості щодо покращення результатів пошуку для середньостатистичного користувача. Найцікавіше в цьому те, що вам насправді непотрібно здійснювати багато дій. Як тільки ви вкажете, що цікавитесь тією чи іншою інформацією, Google подбає про багато інших деталей, які пов'язані з цією тематикою... Кажучи про пошук наступного покоління, я вважаю, що персоналізація пошуку є дуже актуальною темою» [13].

Як вже зазначалося, компанія Google розробила та впровадила власну модель персоналізованого пошуку під назвою «Вікіпошук». Використовуючи цей сервіс, кожен користувач сайту Google може налаштувати свою пошукову видачу завдяки можливості пересувати результати, видаляти їх, додавати свої варіанти і залишати коментарі. Окрім цього, можна ознайомлюватися з налаштуваннями пошукової видачі за тим або іншим запитом інших користувачів. Інструмент «Вікіпошук» розроблено спеціально для того, щоб дати користувачам можливість самим вибирати, які результати пошуку їх задовольняють. Наступного разу, коли користувач авторизується в Google і введе той самий пошуковий запит, він бачитиме сторінку з результатами пошуку, де вже будуть враховані його коментарі. У будь-який момент він може відмінити внесені зміни. Зміни, які вносяться до пошукової видачі, впливатимуть лише на результати пошуку для цього користувача. Зміни в результатах пошуку

завжди відображатимуться при пошуку за тим самим запитом, поки користувач не вийде з аккаунту Google або не відмінить їх. Окрім правки, вони відображатимуться на спеціальній загальнодоступній сторінці, там само можна буде побачити, як налаштували пошукову видачу за цим запитом інші користувачі [12].

Однак, як видно з короткої характеристики «Вікіпошуку», цей сервіс повністю не використовує весь можливий потенціал персоналізованого пошуку. Головними його обмеженнями є:

- налаштування видачі одного користувача не впливає автоматично на результати видачі іншого, хоча їхні інформаційні потреби можуть співпадати;
- налаштування видачі йде за конкретним запитом і мало впливає на запити суміжної тематики.

Проте ці обмеження спричинено об'єктивними причинами, що пов'язано з природою Інтернет. По-перше, для підтримки розвинених профілів користувачі опису всіх можливих інформаційних потреб і, що найголовніше, для автоматичної обробки та застосування для кожного користувача інформації про ресурси мережі потрібні величезні апаратні ресурси. По-друге, з уведенням за відсутності названих вище обмежень можна припустити появу фактів накручування тим чи іншим способом рейтингу певних ресурсів за допомогою цього інструменту. Також варто зазначити, що впроваджуючи системи персоналізованого пошуку, деякі експерти застерігають про ризики, пов'язані з можливостями маніпулювання думками при використанні даних систем, та вказують на необхідність мати декілька альтернативних потужних персоналізованих пошукових систем [7].

Більшість зазначених проблем, які обмежують використання всього потенціалу персоналізованого пошуку в Інтернет, відсутні або не суттєві для систем локального пошуку всередині порталу або домену.

Можливості використання діяльності віртуальних спільнот для покращення результатів пошуку

Наразі поширення набувають розвинені тематичні, навчальні, корпоративні портали, на базі яких часто формуються віртуальні співтовариства.

У рамках цих ресурсів, характерних для Web 2.0, необхідним є створення для окремих користувачів персональних робочих середовищ (робочих кабінетів), максимально адаптованих до їх індивідуальних потреб.

З огляду на домінування евристичних підходів, актуальною стає проблема побудови достатньо формалізованих моделей, які описували б взаємозв'язки між окремими робочими кабінетами, з одного боку, та порталом або співтовариством у цілому, з іншого. Таким чином, подібні моделі повинні описувати такі аспекти проблеми [9]:

- вплив існуючої інфраструктури порталу на створення, наповнення, функціонування та підтримку персонального робочого середовища;
- зворотний вплив окремих середовищ на формування віртуальних співтовариств в цілому; по суті це означає врахування сучасних тенденцій до децентралізованого формування контентом веб-ресурсів та управління ним;
- зв'язки між індивідуальними робочими кабінетами та моделі можливої взаємодії між ними.

Враховуючи перелічені вище напрями, організація порталу на основі взаємодії окремих персональних робочих середовищ може мати суттєве значення для сумісної роботи груп користувачів, а також для підтримки колективного та/або децентралізованого прийняття рішень. Зокрема, в рамках окремого віртуального співтовариства може бути організована «класна дошка» для сумісного обговорення проблеми, закріплення завдань між окремими виконавцями, електронного голосування тощо. При цьому адекватне врахування інтересів окремих користувачів і тем, які їх цікавлять, дозволяють говорити про динамічне формування співтовариств і, відповідно, засобів для управління ними.

У цьому контексті дуже важливою та актуальною видається проблема підвищення онтологічної орієнтованості порталу; зокрема, максимального врахування онтології предметної області. Серед магістральних напрямів розвитку таких онтологічно-орієнтованих порталів, інтегрованих із віртуальними спільнотами, слід відзначити [11]:

- підвищення інформаційної зв'язності порталу для швидкого переходу від поточного ресурсу до матеріалів, пов'язаних із ним за певним критерієм. Критично важливим тут є те, що перелік таких зв'язків не повинен бути статичним, а формуватися динамічно – залежно від поточної ситуації і на основі семантики предметної області та відповідних онтологій;
- формування віртуальної спільноти на основі персональних робочих середовищ окремих користувачів (які можна охарактеризувати як персональні робочі кабінети). Такі

середовища повинні 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 такі компоненти: організацію окремого робочого кабінету; зв'язки між окремими персональними середовищами та можливу взаємодію між ними; вплив загальної інфраструктури порталу на створення, наповнення, функціонування та підтримку персонального робочого середовища;
- ефективний інформаційний пошук мусить залучати 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 справді цікавлять.

У цьому контексті видається доцільним зближення підходів, у рамках яких можна описати власне процес пошуку, з одного боку, та формування віртуальної спільноти, з іншого. Важливими є не тільки пошук на уже сформованій моделі, а й самоорганізація, формування окремих персональних середовищ та віртуальних спільнот. Принципи, які можуть лягти в основу формування коефіцієнтів зв'язків і мір активації в ході такої самоорганізації, багато в чому аналогічні загальним принципам ройового інтелекту, зокрема – алгоритму мурашки [3].

Типовим результатом пошуку, який здійснюється на основі наведених вище принципів, мусить стати формування множини понять і документів, тією чи іншою мірою релевантних запиту, з урахуванням формування вагових коефіцієнтів, пов'язаних із функціями інтерпретації. Кожний елемент цієї множини буде подано у вигляді $(u, m1(u), m2(u), \dots)$, де u – знайдений вузол, $m_i(u)$ – міра релевантності цього вузла, обчислена за i -м критерієм, можливо, недостовірна або нечітка [11]. Припускається навіть динамічне породження нових критеріїв, якщо задати процедуру такого породження. Основна проблема, яка виникає у процесі такого пошуку, пов'язана зі значними обсягами інформації і, відповідно, – зі значною часовою складністю. Тому ключовим є питання: як спрямувати процес поширення активації в потрібному напрямі та яким повинен бути критерій зупинки цього процесу? Важливою проблемою є комбінування критеріїв, яка полягає в переході 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; зокрема, генетичні алгоритми, які добре зарекомендували себе для розв'язання низки перебірних задач [3]. У контексті, що розглядається, можна виділити як мінімум два аспекти застосування цих алгоритмів:

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

Корисна інформація, що накопичується у процесі пошуку, мусить максимально враховуватися, тобто стратегія пошуку – динамічно коригуватися. Одна з можливих стратегій такого пошуку полягає в такому [4]: множина можливих рішень Ω розбивається на підмножини $\Omega^*, \dots, \Omega_n$. Задача полягає у виборі найперспективнішої підмножини Ω^* , на якій здійснюватиметься подальший пошук. До уваги повинні братися:

- коефіцієнт звуження області пошуку $h(\Omega^*) = |\Omega^*| / |\Omega|$;
- оцінка $p(\Omega^*, I)$ перспективності підмножини Ω^* , яка залежить, крім самої підмножини Ω^* , від наявної інформації I ; остання складається з апіорної оцінки перед розбиттям та результатів апостеріорної перевірки, в якій з підмножин знаходиться потрібне рішення.

Тоді проблема вибору найперспективнішої підмножини Ω^* може бути математично сформульована як задача мінімізації критерію

$$E(h(\Omega) p(\Omega) + (1-h(\Omega))(1-p(\Omega))),$$

де E – символ математичного сподівання.

Слід зазначити, що ідеологія роботи генетичних алгоритмів тісно пов'язана з даною формулою. Справді, схеми можливих рішень генетичного алгоритму фактично задають розбиття на підмножини, а накопичення інформації здійснюється на основі того, що, згідно з теоремою про схеми [3], відсоток найбільше пристосованих схем (тобто критеріїв вибору найперспективніших підмножин) у наступних поколіннях зростає.

У рамках цього підходу можна розв'язувати також задачі, характерні для інтелектуального аналізу даних (Data Mining) [10]:

- пошук потенційних партнерів на основі аналізу мір схожості між персональними робочими середовищами;

- динамічне формування груп користувачів, які мають спільні або схожі інтереси;
- прийняття рішень за аналогією (наприклад, якщо користувач А для розв'язання задачі С вважає корисним документ W, то користувачеві Х, характеристики якого схожі на характеристики користувача А, для розв'язання задачі К, схожої на С, можна порекомендувати список документів, схожих на W);
- автоматичне визначення відповідних функцій інтерпретації формальної моделі та пов'язаних із ними вагових коефіцієнтів на основі документів, відібраних користувачем.

Висновки

У роботі обґрунтовано актуальну наукову задачу розробки основних принципів функціонування та впровадження системи персоналізованого пошуку, яка дозволяла б здійснювати пошук документів тематичного порталу з урахуванням результатів кооперативної діяльності віртуального співтовариства порталу за оцінкою якості документів.

Для розв'язку цієї задачі проаналізовано низку суміжних досліджень та проведено їх адаптацію. На основі дослідження можна сформулювати основні подальші напрями досліджень та план реалізації системи персоналізованого пошуку на основі аналізу діяльності членів віртуальної спільноти.

Список літератури

1. Глибовець М. М. Алгоритм та онлайн-застосування пошуку осередків зацікавленості за обраною предметною областю / М. М. Глибовець, М. О. Сидоренко. – К. : Вісник Київського нац. ун-ту ім. Тараса Шевченка. Фізико-математичні науки. – 2012. – № 3. – С. 133–140.
2. Глибовець М. М. Веб-сервіси оброблення документів / М. М. Глибовець, А. А. Жигмановський, Р. І. Заболотний, П. О. Захоженко. – К. : Нац. ун-т «Києво-Могилянська академія», 2012. – 212 с.
3. Глибовець М. М. Еволюційні алгоритми. / М. М. Глибовець, Н. М. Гуласва. – К. : НАУКМА, 2013. – 828 с.
4. Глибовець М. М. Про деякі підходи до проблеми інформаційного керування випадковим пошуком / М. М. Глибовець, О. В. Олецкий // Dynamical System Modelling and Stability Investigation. Thesis of Conference Reports, May 22–25, 2007. – 370 p.
5. Глибовець Н. Н. Создание динамической системы распространения контента с использованием протокола BitTorrent / Н. Н. Глибовець, В. Е. Мельник, М. О. Сидоренко // Компьютерная математика. – 2012. – № 2. – С. 76–85.
6. Калуженков Н. Будущее интернет-поиска [Електронний ресурс] / Н. Калуженков. – Режим доступу: <http://hostinfo.ru/articles/internet/services/1055>. – Назва з екрана.
7. Кравцов А. Персонализация – две стороны одной медали [Електронний ресурс] / А. Кравцов. – Режим доступу: <http://www.seonews.ru/columns/detail/9126.php>. – Назва з екрана.
8. Ландэ Д. В. Поиск знаний в Internet. Профессиональная работа / Д. В. Ландэ. – М. : Вильямс, 2005. – 272 с.
9. Олецкий О. В. До проблеми організації онтологічно-орієнтованих персональних робочих середовищ у рамках віртуального співтовариства [Електронний ресурс] / О. В. Олецкий // 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. – Режим доступу: http://conf.atsukr.org.ua/files/conf_dir_8/oletskey_sppr2008.pdf. – Назва з екрана.
10. Олецкий О. В. Онтологічно орієнтований інформаційний пошук на основі хвильового процесу поширення активації / О. В. Олецкий // Наукові записки НАУКМА. – 2008. – Т. 86: Комп'ютерні науки. – С. 50–52.
11. Олецкий О. В. Принципи побудови формалізованих онтологічно-орієнтованих моделей організації віртуальних спільнот у рамках веб-порталу / О. В. Олецкий // Теоретичні та прикладні аспекти побудови програмних систем. Матеріали міжнародної конференції ТААПСД'2008. Київ–Чернігів, 22–26 вересня 2008 р. – С. 209–214.
12. Справка Google. Функции : Википоиск [Електронний ресурс]. – Режим доступу: <http://www.google.ru/support/websearch/bin/answer.py?hl=ru&answer=115764>. – Назва з екрана.
13. MacManus R. Interview with Google's Matt Cutts about Next-Generation Search [Electronic resource] / R. MacManus. – Mode of access: http://www.readwriteweb.com/archives/interview_with_matt_cutts_next_generation_search.php. – Title from the screen.
14. SearchWiki – персоналізований пошук от Google [Електронний ресурс]. – Режим доступу: <http://internetno.net/category/anonsi/searchwiki/>. – Назва з екрана.

M. Glybovets, O. Konyushenko

USING THE ACTIVITIES OF VIRTUAL COMMUNITIES FOR IMPROVING SEARCH OF INFORMATION

The article substantiates the task of development of personalized search that would be search for documents of thematic portal based on the results of the cooperative activity of an virtual community portal for the evaluation of documents.

Keywords: personalized search, virtual communities, genetic algorithms, thematic portals.

Матеріал надійшов 24.04.2014