

Proceedings of the Sixth International Conference on Knowledge Discovery and Data Mining (ACM SIGKDD-2000), Boston, MA, 2000. ACM Press. 5. Flake G., Lawrence S., Giles C., Coetzee F. *Self-Organization of the Web and Identification of Communities. IEEE Computer*, 35(3), 66–71, 2002 <http://webselforganization.com/>. 6. Gillet S., Kapor M. *Self-govering Internet: Coordination by Design. Massachusetts Institute of Technology. Center for Coordination Science. Technical Report. 1997., 25p., http://ccs.mit.edu/CCSWP197/CCSWP197.html. 7. J. Srivastava, R. Cooley, M. Deshpande, P-T. Tan. *Web usage mining: discovery and applications of usage patterns from Web data. SIGKDD Explorations*, (1) 2, 2000. 8. Pennock D., Flake G., Lawrence, Glover E., Giles C. *Winners don't take all: Characterizing the competition for links on the web. Proceedings of the National Academy of Sciences, Volume 99, Issue 8, pp. 5207–5211, April, 2002.* 9. Пелецишин А.М. *Методи та алгоритми моделювання Web-систем // Вісник Держ. ун-ту "Львівська політехніка". – 2000. – №406. – С.199–211.* 10. Пелецишин А.М., Гулка Т.Б. *Інформаційна система аналізу діяльності Web-вузла // Вісник Нац. ун-ту "Львівська політехніка". – 2001. – №438. – С.115–120.* 11. Пелецишин А.М. *Методи визначення та оптимізації тематики сайту // Вісник Нац. ун-ту "Львівська політехніка", 2004. – №519. – С.254–267**

УДК 683.1

А.М. Пелецишин, О.Л. Березко,
Національний університет “Львівська політехніка”,
кафедра інформаційних систем та мереж

ІНТЕЛЕКТУАЛЬНА СИСТЕМА ІДЕНТИФІКАЦІЇ ОСОБИСТОСТІ У WWW

© Пелецишин А.М., Березко О.Л., 2006

Вводяться поняття “Веб-особистість” для позначення деякої множини інформаційних атрибутів, яка є проекцією реальної особи на WWW, та “коефіцієнт достовірності ідентифікації”, який є характеристикою деякого інформаційного об’єкта і відображає ступінь достовірності приналежності цього інформаційного об’єкта до деякої Веб-особистості. Пропонується структура та засоби інтелектуальної системи, яка дає змогу автоматизувати пошук невідомих значень атрибутів Веб-особистості.

Introduced are the notions “Web personality” in order to denote some set of information attributes which is a projection of a real person on the WWW and “authenticity identification coefficient” which is a characteristic of some information object and represents the level of authenticity of belonging of this information object to some Web personality. Proposed are the structure and means of an intellectual system which allows to automatize the search of unknown values of attributes of the Web personality.

Вступ

Сьогодні існує тенденція стрімкого зростання об’єму даних, доступних через Інтернет (Веб-контенту). Кількісно це означає подвоєння загального об’єму Веб-контенту приблизно кожні 8 місяців. Переважна більшість даних, доступних через цю глобальну мережу, має слабоструктурований характер, що викликало появу автоматичних програмних засобів пошуку інформації та отримання даних при використанні певних Веб-ресурсів. Було створено інтелектуальні системи, основним завданням яких є ефективне видобування знань з Інтернету (Web mining). [1,2,4,5,10,11]

Значна частина користувачів Інтернету мають власні Веб-сторінки, скриньки електронної пошти, є активними учасниками різноманітних форумів тощо, наслідком чого є накопичення у

мережі інформації про цих користувачів, яке може відбуватися без їх активної участі та навіть без їх відома (прикладом можуть бути сайти фан-клубів деякої особи). Отже, можна говорити про створення певної Веб-особистості (деякої множини інформаційних атрибутів) у мережі як проєкції реальної особистості на WWW.

До 90% користувачів Інтернету використовують пошукові системи (такі як www.google.com, www.altavista.com, www.yahoo.com та багато інших) для пошуку необхідної інформації, зокрема і про конкретних осіб (рис.1, 2). Проте, через слабоструктурованість, невпорядкованість та величезні розміри Веб-контенту результати використання цих систем не завжди можна вважати задовільними. Як правило, користувач змушений сам виконувати аналіз результатів пошуку, що дуже часто є неприйнятним у зв'язку з великими обсягами отриманої інформації [3,4,9,10,11].

Виходячи з вищенаведеного, актуальним є створення спеціалізованих інтелектуальних засобів, які б автоматизували пошук невідомих значень атрибутів Веб-особистості, ґрунтуючись на значеннях атрибутів, які задав користувач.

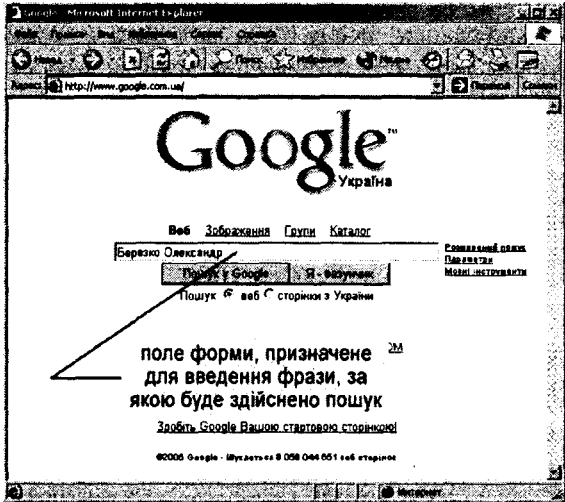


Рис. 1. Стартова Веб-сторінка пошукової системи Google

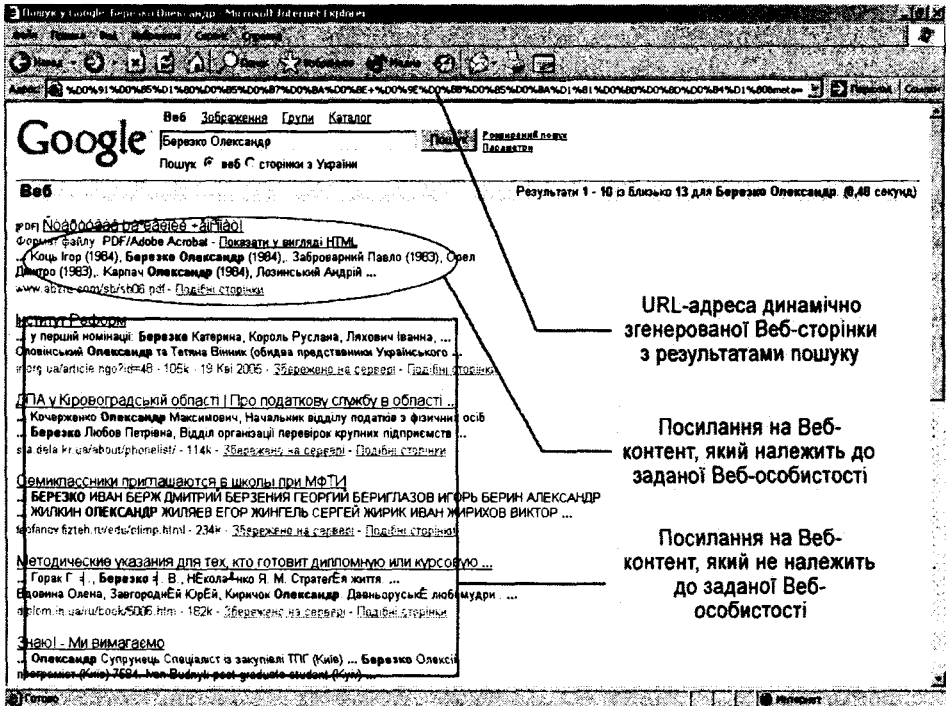


Рис. 2. Веб-сторінка пошукової системи Google з результатами пошуку

Веб-особистість

Всі дані, доступні в Інтернеті, можна віднести до однієї з трьох категорій:

- Web content (Веб-контент: тексти, зображення, аудіо- та відеофайли тощо);
- Web structure (Веб-структура: гіперпосилання, теги тощо);
- Web usage (дані про використання Інтернету: http logs, app server logs тощо) [2].

Вводимо поняття “Веб-особистість” для позначення множини інформаційних атрибутів – проекції реальної особистості на WWW. Веб-особистість – це множина даних, доступних через Інтернет, які стосуються конкретної особи і можуть належати до будь-якої категорії або до будь-якої комбінації вищенаведених категорій даних. Отже, якщо розглядати всі дані, доступні через Інтернет як деяку множину, конкретна Веб-особистість є підмножиною цієї множини.

Треба одразу зазначити, що найцікавішим з точки зору пошуку невідомих значень атрибутів Веб-особистості є саме Веб-контент, а інші вищенаведені категорії даних можуть виконувати допоміжну роль у пошуку, слугуючи “ключем” до місцезнаходження корисної інформації.

З точки зору змісту атрибути Веб-особистості можна класифікувати так:

- Персональні дані про реальну особу: ім’я, прізвище, дата народження, освіта тощо;
- “Паспортні”, з точки зору мережі Інтернет, дані Веб-особистості: псевдонім (nickname); адреси персональних Веб-ресурсів: Веб-сторінки, блогу тощо; адреси електронних поштових скриньок;
- Сфера інтересів (сфера діяльності) – може бути представлена як множина термінів, що позначають поняття, пов’язані з діяльністю особи;
- Посилання на статті, книги, мистецькі твори, дописи на форумах – довільні об’єкти авторського права особи, доступні в Інтернеті, – можуть бути представлені множиною URL-адрес цих об’єктів;
- Посилання на Веб-ресурси, де є інформація про особу (чи принаймні згадки про неї) – можуть бути представлені множиною URL-адрес цих ресурсів.

Деякі з перерахованих атрибутів можуть мати більш ніж одне значення. Наприклад, людина може застосовувати для своєї ідентифікації в Інтернеті декілька псевдонімів, мати декілька електронних поштових скриньок тощо.

Вищенаведену класифікацію можна доповнити, розділивши атрибути за ознакою авторства:

- Створені особою, інформацію про яку шукаємо;
- Створені будь-ким іншим.

Комплексний підхід до пошуку

Видобування знань можна визначити як знаходження і аналіз корисної інформації. В контексті Інтернету цей напрямок діяльності прийнято поділяти на три частини: автоматичний пошук якісної інформації у Веб-контенті – Web content mining, виявлення і обробка інформації, що стосується роботи користувачів із сервером – Web usage mining, виявлення і обробка інформації, що стосується структури Інтернету – Web structure mining. Специфіка Web mining полягає у застосуванні традиційних технологій data mining для аналізу слабоструктурованої і значної за об’ємом інформації, що міститься на Веб-вузлах. [1,2,10,11]

Пошук невідомих значень атрибутів Веб-особистості, ґрунтуючись на значеннях атрибутів, які задав користувач, належить до задач Web content mining. Завданням пропонованої системи є саме такий пошук.

Як зазначалося вище, результати використання традиційних пошукових систем не завжди можна вважати задовільними. Пропонується застосування комплексного підходу до пошуку з подальшою обробкою отриманих результатів.

Комплексний підхід до пошуку передбачає застосування таких засобів:

1. Одночасне використання цілої низки пошукових систем: www.google.com, www.altavista.com, www.lycos.com, www.msn.com, www.yahoo.com, www.teoma.com, www.alltheweb.com, www.overture.com, www.yandex.ru, www.rambler.ru, тобто пропонована система є мета-пошуковою.

Мета-пошукова система (metacrawler) – це пошуковий інструмент, який надсилає запит користувача одразу на декілька пошукових систем та/або каталогів, а іноді і в так звану невидиму (приховану) частину Вебу – сукупність on-line інформації, не проіндексовану традиційними пошуковими системами. Після збирання результатів мета-пошукова система видаляє посилання, які дублюються, і об'єднує та ранжує результати в загальному списку або виконує інші дії згідно зі своїм алгоритмом.[8,9] На рис.3 зображено узагальнену структуру мета-пошукової системи.

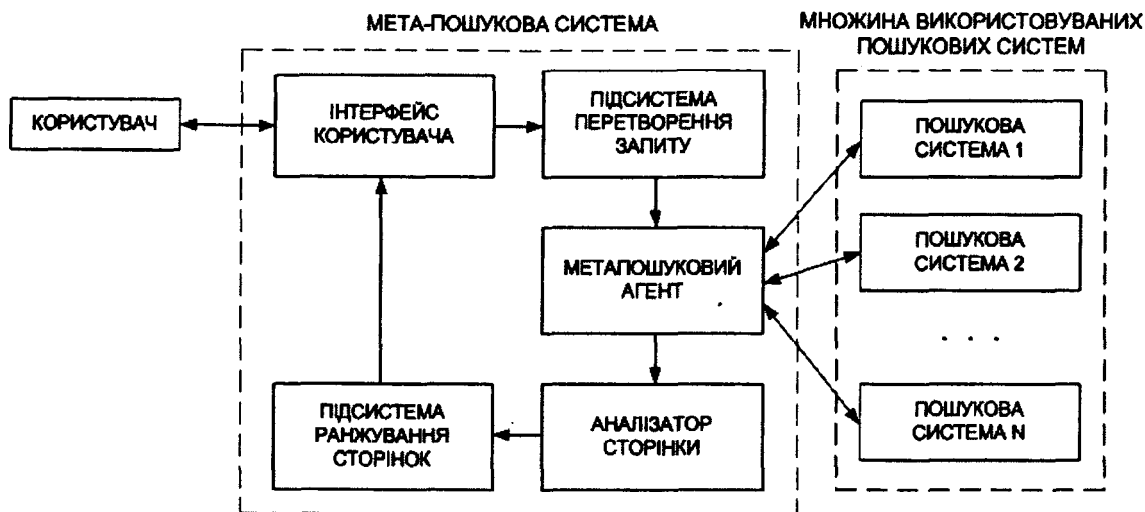


Рис. 3. Узагальнена структура мета-пошукової системи

Незважаючи на постійне зростання індексів пошукових систем, збільшення Веб-контенту загалом лише погіршило картину доступності інформації, оскільки відсоток Веб-контенту, охоплений окремими індексами, суттєво зменшився. Звідси випливає, що лише застосування сукупності пошукових серверів здатне забезпечити повноцінну інформаційну картину для пошукових задач, для вирішенні яких є важливою повнота пошуку [3,4,5,6].

У процесі розробки системи список використовуваних систем може змінюватися.

2. Надсилання комбінації запитів замість одного запиту. Інформація про особу (наприклад, ім'я та прізвище) попри всю свою однозначність може бути присутньою у Веб-контенті формально у різних формах. Наприклад, шукаючи інформацію за словосполученням “Петренко Іван Петрович”, до пошукових серверів доцільно надіслати такі запити: “Петренко Іван Петрович”, “Петренко Іван”, “Іван Петренко”, “Петренко І.П.”, “Петренко І.”, “Петренко”.

Доцільно також сформувані множину понять, які можуть стосуватися сфери діяльності особи та розбити їх за категоріями. Наприклад, поняття “комп'ютер” та “база даних” можуть стосуватися категорії “програміст”. За наявності хоча б одного з цих понять серед множини слів, які користувач вводить під час заповнення полів форми, до фрази, за якою буде здійснюватися пошук, необхідно додати слово “програміст”.

Важливо також сформувані множину продукційних правил, за якими можна додавати до фрази, за якою буде здійснюватися пошук, слова, які позначають поняття, потенційно дотичні до Веб-особистості, значення атрибутів якої шукаються. Наприклад:

ЯКЩО

(рік народження особи $\geq Y-22$) і (рік народження особи $\leq Y-17$),

де Y – поточний рік

ТО

(додати до фрази, за якою буде здійснюватися пошук слово “студент”).

У процесі розвитку системи планується експериментально визначити, які саме комбінації запитів мають найбільшу вагу для пошуку та скоректувати підсистему перетворення запитів за результатами експериментів.

3. Транслітерація фрази запиту англійською мовою і надсилання транслітерованої фрази як запиту до пошукових систем. Така дія доцільна, якщо: інформація, яка стосується Веб-особистості,

може належати до англomовного Веб-контенту, який сьогодні у Всесвітній павутині абсолютно переважає; якщо у Вебі існує персональна Веб-сторінка чи інші персональні Веб-ресурси конкретної особи, пошук інформації про яку здійснюється, є велика ймовірність того, що прізвище цієї особи буде присутнім в URL-адресах цих ресурсів. В обох випадках прізвище особи буде написано буквами латинського алфавіту з використанням транслітерації.

Транслітерація – це точне відтворення букв або сполучень букв однієї мови засобами алфавіту іншої мови. Важко досягнути взаємної однозначності такого відтворення в прямому і зворотному напрямках без розробки жорстких стандартів. Такі стандарти існують і широко використовуються, наприклад, в алфавітних каталогах іноземної літератури. Зокрема стандарт ISO-9-1986-(E)/ISO/TC 46 регламентує транслітерацію символів слов'янської кирилиці символами латинського алфавіту. Проте на практиці спостерігаються суттєві відхилення від будь-яких стандартів. Тенденції таких відхилень частково визначає англomовне походження Інтернету [7].

У проєктованій системі буде використано систему транслітерації, що тяжіє до українсько-англійської системи. За зразок було взято російсько-латинську систему транслітерації, наведену в [7], що створювалась на основі аналізу великої кількості імен російської частини Інтернету і кількох поширених у Веб схем транслітерації.

Структура пропонованої системи

Застосуємо методологію IDEF0 (Integration Definition for Function Modeling) для побудови контекстної діаграми, яка відображає функціонування пропонованої системи загалом (рис.4).

Дані та об'єкти, що стосуються пропонованої системи, можна класифікувати так:

- дані та об'єкти, змінювані в процесі функціонування системи: запит користувача, поточні коректування, база даних системи;
- правила і обмеження, за якими функціонує система: мережні протоколи (TCP/IP, HTTP), швидкість каналів передавання даних, швидкодія пошукових серверів, швидкодія програмного забезпечення, швидкодія апаратного забезпечення;
- дані та об'єкти, що з'являються в результаті функціонування системи: нові знання про об'єкт пошуку, нові правила та закономірності, нові записи в базі даних системи;
- ресурси, які необхідні для функціонування системи, але не змінюються під час її функціонування: апаратне забезпечення, програмне забезпечення, пошукові системи, доступ до Інтернету.



Рис. 4. Контекстна діаграма пропонованої системи

Використаємо діаграму потоків даних (Data Flow Diagram) для документування механізмів передавання та обробки інформації в проєктованій системі (рис.5).

Розглянемо докладніше об'єкти 3-го рівня, присутні на DFD-діаграмі пропонованої системи, оскільки деталізація опису системи на цьому рівні достатня для формування повноцінного уявлення про принципи її функціонування (рис.5).

1.1.1 Класифікація даних. Вхідними даними та параметрами є дані та параметри, які користувач вводить під час заповнення форми, працюючи з інтерфейсом системи. Отримані так дані класифікуються системою згідно з полями форми, звідки вони були отримані.

1.1.2 Заповнення полів відношень бази даних. Система здійснює заповнення полів бази даних, які стосуються атрибутів Веб-особистості, введених користувачем (рис.6).

1.1.3 Формування запитів до пошукових систем. На основі даних, введених користувачем, за принципами комплексного підходу до пошуку, описаного вище, формуються запити до пошукових систем. Формуючи запити, необхідно враховувати те, що кожна пошукова система використовує власне кодування запитів. Щоб дізнатися про формат запиту до конкретної пошукової системи, необхідно зайти на відповідний Веб-сайт і дати запит на пошук за довільною фразою. Формат запиту можна буде отримати у полі для адрес браузера – це буде URL-адреса динамічно згенерованої Веб-сторінки з результатами пошуку (рис.2). Для наступних запитів треба лише замінити використану на нову фразу, за якою має здійснюватися пошук.

Наприклад, для пошукової системи Google формат звичайного запиту є таким:

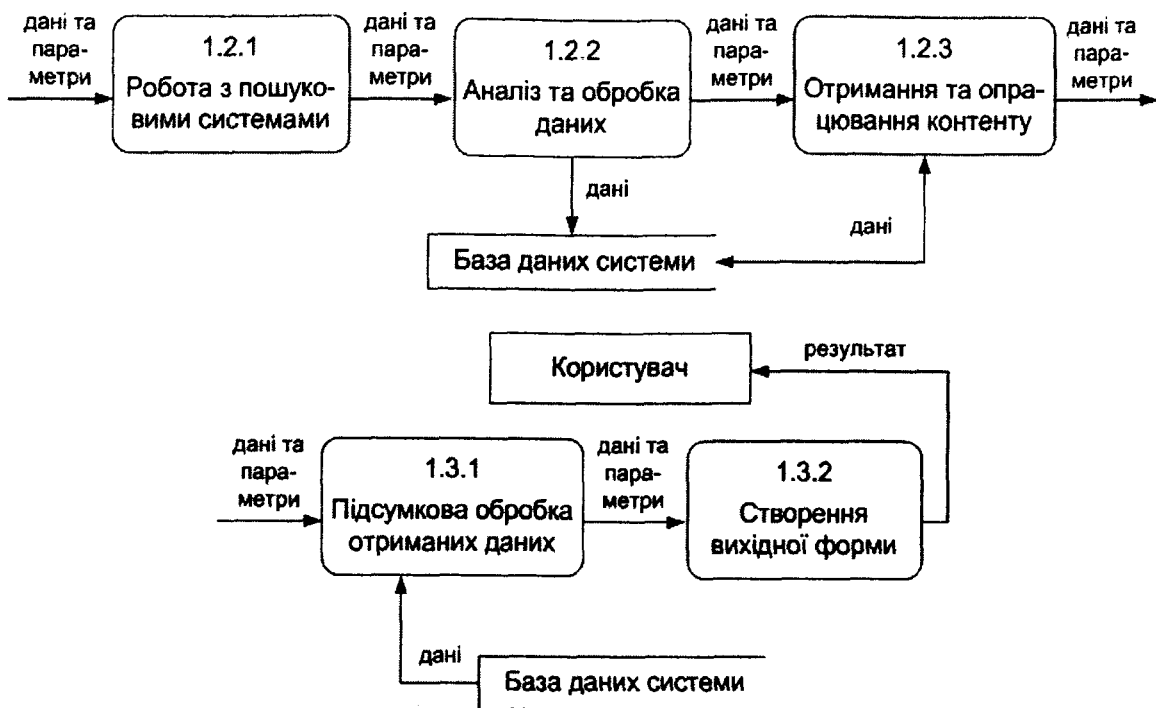
```
http://www.google.com/search?q=***** ,
```

де ***** – фраза, за якою буде здійснюватися пошук.

1.2.1 Робота з пошуковими системами. Надсилання запитів до пошукових систем та отримання відповіді у вигляді Веб-сторінок з результатами роботи відповідних пошукових систем.



Рис. 5. DFD-діаграма пропонованої системи



Продовження рис. 5. DFD-діаграма пропонованої системи

1.2.3 Отримання та опрацювання контенту. Послідовне отримання з мережі Інтернет тих Веб-сторінок, яким відповідають URL-адреси, занесені в базу даних системи на попередньому кроці, та їх аналіз.

Під час пошуку невідомих значень атрибутів Веб-особистості виникає проблема відфільтрування Веб-контенту, який не належить до цієї Веб-особистості, але перетинається з її інформаційними атрибутами. Наявність посилань на такий Веб-контент у базі даних системи є небажаним, оскільки може спонукати систему до неправильних висновків щодо приналежності такого Веб-контенту до заданої Веб-особистості. Наприклад, існує більш ніж одна Веб-особистість, для якої атрибут “прізвище” є “Березко”. Пошукові системи не розрізняють такі деталі навіть при вказанні імені чи по батькові особи, оскільки ці слова можуть бути присутніми на тій самій Веб-сторінці у іншому контексті (рис.2).

Для відфільтрування такого Веб-контенту доцільно сформувати множину реальних імен людей. Як правило, у довільному тексті ім’я, по батькові чи ініціали містяться справа, рідше – зліва від прізвища. Отже, аналізуючи слова, що знаходяться зліва і справа від прізвища, можна з великою достовірністю виявити, чи належить цей Веб-контент до Веб-особистості, значення атрибутів якої шукаються.

Ще важче відфільтрувати Веб-контент, який перетинається з більш ніж одним значенням атрибутів Веб-особистості, але до неї не належить. Здійснити це можна, виявивши деякі значення атрибутів Веб-особистості на основі аналізу тієї частини Веб-контенту, яка характеризується високим коефіцієнтом достовірності ідентифікації, і здійснити пошук цих значень атрибутів у Веб-контенті, який викликає сумнів щодо приналежності до заданої Веб-особистості.

Доцільно також аналізувати мета-дані, що містяться на Веб-сторінках для точнішої класифікації Веб-контенту, який належить до заданої Веб-особистості з метою окреслення сфери інтересів особи.

1.3.1 Підсумкова обробка отриманих даних. Використання даних, накопичених у базі даних системи для виявлення атрибутів заданої Веб-особистості, що не були виявлені на попередніх кроках, зокрема формування множини слів, які позначають поняття, дотичні до заданої Веб-особистості.

1.3.2 *Створення вихідної форми.* Впорядкування знайдених значень атрибутів Веб-особистості. Динамічна генерація Веб-сторінки з результатами роботи системи, яка би передбачала можливість інтерактивної оцінки користувачем як роботи системи загалом, так і приналежність окремих значень атрибутів заданій Веб-особистості. Необхідно передбачити можливість запуску нового пошуку як “з нуля”, так і на основі попереднього.

Пропонована система в процесі своєї роботи оперує різними за змістом і форматом даними. Для зручного накопичення даних та швидкого оперування ними пропонується використання бази даних реляційного типу. На рис.6 зображено ERD-модель пропонуваної бази даних.

Розглянемо докладніше відношення пропонуваної бази даних:

Person – призначене для зберігання персональних даних про реальну особу: ім'я, прізвище, по батькові, дата народження, громадянство. Первинним ключем є поле `person_id`. Зазначимо, що це відношення має зв'язки типу “один до багатьох” з іншими відношеннями. Джерелом інформаційного наповнення може бути користувач або результати роботи системи.

Links – призначене для зберігання URL-адрес частини Веб-контенту, яка потенційно належить до заданої Веб-особистості, коефіцієнтів достовірності ідентифікації цих URL-адрес, а також мета-даних відповідного Веб-контенту. Первинним ключем є поле `id`. Зовнішнім ключем є поле `person_id`. Це відношення є з'єднаним з відношенням **Person** зв'язком типу “багато до одного”. Такий зв'язок відображає те, що до однієї Веб-особистості може належати більш ніж одна Веб-сторінка, а, отже, і більш ніж одне посилання на Веб-сторінку. Джерелом інформаційного наповнення є результат роботи системи.

Categories – призначене для зберігання слів, які позначають категорії, до яких можна віднести атрибут “сфера діяльності” заданої Веб-особистості. Наприклад: програміст, музикант, спортсмен, студент тощо. Первинним ключем є поле `id`. Зовнішнім ключем є поле `person_id`. Це відношення є з'єднаним з відношенням **Person** зв'язком типу “багато до одного”. Такий зв'язок відображає те, що задану Веб-особистість можна віднести більш ніж до однієї категорії за критерієм “сфера діяльності”. Наприклад: реальна особа може бути одночасно студентом і спортсменом. Джерелом інформаційного наповнення є розробник системи.

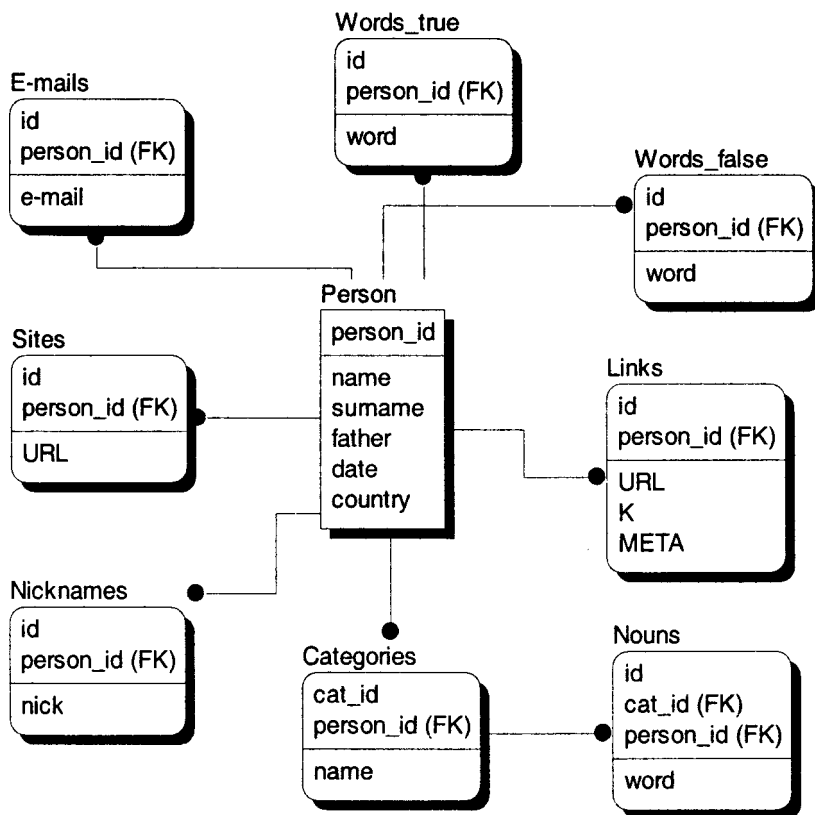


Рис. 6. ERD-модель бази даних пропонуваної системи

Nouns – призначене для зберігання слів, які позначають поняття, дотичні до певної категорії, що зберігається у відношенні Categories. Наприклад, поняття “гітара” та “концерт” можуть стосуватися категорії “музикант”. Джерелом інформаційного наповнення може бути розробник системи або результати роботи системи.

Відношення E-mails, Sites, Nicknames, Words_true, Words_false призначені для зберігання адрес електронних поштових скриньок, які належать реальній особі, відображенням якої є задана Веб-особистість; адрес Веб-сайтів, потенційно дотичних до заданої Веб-особистості; псевдонімів реальної особи, які використовуються під час роботи у Веб; слів, що потенційно належать до значень атрибутів заданої Веб-особистості; слів, що завідома не належать до значень атрибутів заданої Веб-особистості відповідно. Всі ці відношення з’єднані з відношенням “Person” зв’язком типу “багато до одного”. Такий тип зв’язку відображає те, що, наприклад, до деякої Веб-особистості можуть належати декілька адрес електронних поштових скриньок.

Коефіцієнт достовірності ідентифікації

Коефіцієнт достовірності ідентифікації – це деяке дійсне число, яке належить діапазону $[0,1]$. Це число є характеристикою певного інформаційного об’єкта і відображає ступінь достовірності того, що цей інформаційний об’єкт належить до деякої Веб-особистості. Значення “1” означає найвищу достовірність ідентифікації, значення “0” – найнижчу:

$$0 \leq \mu \leq 1, \quad (1)$$

де μ – коефіцієнт достовірності ідентифікації інформаційного об’єкта.

Значення коефіцієнта достовірності ідентифікації (1), який присвоюється URL-адресам, залежить від:

1. Фрази запиту, у відповідь на яку була отримана ця URL-адреса. Наприклад: посилання, отримане як відповідь на запит “Олександр Березко 1984 студент” матиме вищий коефіцієнт достовірності ідентифікації, ніж посилання, отримане як результат відповіді на запит “Олександр Березко”.

2. Позиції посилання серед інших посилань на Веб-сторінці, динамічно згенерованій пошуковою системою як відповідь на надісланий запит. Пошукові системи здійснюють ранжування результатів пошуку за критеріями наявності в тексті документа, на який вказує посилання, слів запиту, і за кількістю посилань на цю сторінку серед інших Веб-сайтів, враховуючи популярність цих Веб-сайтів. Це ранжування доцільно брати до уваги під час присвоєння коефіцієнтів достовірності ідентифікації. Перші посилання на Веб-сторінці є, з точки зору пошукової системи, релевантнішими до об’єкта пошуку, ніж наступні [9].

Нехай

$$Obj(Wp) = \{Obj_i^{(Wp)}\},$$

де Obj_i – інформаційний об’єкт, що має стосунок до заданої Веб-особистості.

$Obj(Wp)$ – це множина інформаційних об’єктів WWW, які можуть вважатися релевантними до Веб-особистості Wp .

Тоді нечітка множина $RelObj(Wp)$ відображає інформаційний профіль Веб-особистості:

$$RelObj(Wp) = \{(Obj_i^{(Wp)}, \mu)\} \quad (2)$$

де μ_i – міра достовірності ідентифікації інформаційного об’єкта $Obj_i^{(Wp)}$ до Веб-особистості Wp .

Висновки

З метою вдосконалення засобів сучасних пошукових систем та приділення більшої уваги інтелектуальному пошуку в слабоструктурованому середовищі, для позначення деякої множини інформаційних атрибутів, яка є проекцією реальної особи на WWW, введено поняття “Веб-особистість”. Аналізом теперішнього стану WWW та можливостей наявних пошукових систем доведено доцільність створення спеціалізованих інтелектуальних засобів для автоматизації пошуку невідомих значень атрибутів Веб-особистості, ґрунтуючись на значеннях атрибутів, які задає користувач. Запропоновано структуру та засоби інтелектуальної системи для реалізації такого

пошуку. Введено поняття “коефіцієнт достовірності ідентифікації”, що є характеристикою деякого інформаційного об’єкта і відображає ступінь достовірності того, що цей інформаційний об’єкт належить до деякої Веб-особистості. Запропоновано шляхи визначення значень цього коефіцієнта.

1. Щербина А. Основы извлечения знаний из Internet // Открытые системы. – №04/2003 – <http://www.osp.ru/os/2003/04/049.htm>. 2. Srivastava J. Web Mining : Accomplishments & Future Directions - www.ieee.org.ar/downloads/Srivastava-tut-paper.pdf. 3. Bayeza-Yates R., Castillo C. Relating Web Structure and User Search Behavior. Center for Web Research, Department of Computer Science, University of Chile, 2002. 24p. 4. Kosala R., Blockeel H. Web Mining Research: A Survey. Department of Computer Science, Katholieke Universiteit Leuven SIGKDD: SIGKDD Explorations: Newsletter of the Special Interest Group (SIG) on Knowledge Discovery & Data Mining, ACM, 2000, 15p. 5. Hu W.-C., Chen Y., Smalz M., Ritter G. An Overview of World Wide Web Search Technologies. Department of Computer Science. Auburn University, 2000, 6p. 6. Baeza-Yates R., Riebero-Nero B. Modern Information Retrieval. ACM Press. New York, 1999, 501p. 7. Талантов М. Поиск в Интернете: использование имён. // Компьютер пресс №2/2000 – <http://www.citforum.ru/internet/search/namesearch.shtml>. 8. Brin S., Page L. The Anatomy of a Large-Scale Hypertextual Web Search Engine. Computer Science Department, Stanford University, Stanford, 1998, 20p. 9. Brin S., Page L. The Page Rank Citation Ranking: Bringing Order to Web. Computer Science Department, Stanford University, Stanford, 1998, 17p. 10. Huang L. A Survey On Web Information Retrieval Technologies. Computer Science Department, State University of New York. 2000, 33p. 11. Kobayashi M., Takeda K. Information Retrieval on the Web. IBM Research, IBM Tokyo Research Laboratory. IBM Japan. 2000, 47p.

УДК 681.325

І.О. Процько

Національний університет “Львівська політехніка”,
кафедра систем автоматизованого проектування

РОЗРОБКА СХЕМИ УЗАГАЛЬНЕНОГО ЕФЕКТИВНОГО АЛГОРИТМУ ГАРМОНІЧНОГО ПЕРЕТВОРЕННЯ ДАНИХ

© Процько І.О., 2005

Для прийняття рішень з організації процесу обчислення гармонічних перетворень даних змінних обсягів використовується сформований циклічний розклад підстановки, який визначається на основі базисної матриці перетворень. Розглянуто узагальнену схему ефективного алгоритму на основі формування блочно-матричних структур аргументів базису гармонічного дискретного перетворення.

The passing decisions for synthesis efficient algorithm calculation of the harmonic discrete transforms the arbitrary size of data are used the formed cycle decomposition of substitution what determine for the discrete basis matrix of transform. The general structure of efficient algorithm on base of the constructing block-matrix structures of arguments the harmonic discrete transforms is considered.

Вступ

Теоретико-числові перетворення в частотну область, поряд з часовою, застосовуються в системах обробки і прийняття рішень та передавання інформації. Широке застосування дискретного перетворення вхідної інформації для подальшого аналізу, оброблення та синтезу вихідних даних ставить нові задачі для алгоритмічних, програмно-апаратних засобів. Спектральний аналіз