



ПРАВОВА ІНФОРМАТИКА

№ 1 (9) / 2006

У номері:

- *Проблеми взаємовідношення правової інформатики та інформаційної безпеки*
- *Правове регулювання доступу до офіційної правової інформації в Україні*
- *Правові аспекти інформатизації*
- *Використання здобутків правової інформатики у боротьбі з корупцією*
- *Питання відповідальності за поширення недостовірної інформації,
отриманої в мережі Інтернет*
- *2GW – майбутнє Інтернету*
- *Термінологічні та організаційні аспекти створення
інформаційно-аналітичної системи ОВС України*
- *Про економічний аспект захисту персональних даних у контексті
права власності на інформацію*
- *До питання ідентифікації особи за допомогою біометричних даних*
- *До питання засобів індивідуалізації найманого працівника*
- *Інформатизація комплексної системи детінізації відносин
у сфері погашення податкового боргу платників податків*
- *Оцінка умов безпеки руху в зоні впливу автомобільної стоянки*

РЕКОМЕНДАЦІЇ РАДИ ЄВРОПИ № R(87)15 від 17.09.1987 р.

“Про регулювання використання персональних даних у секторі поліції”

**НАУКОВИЙ ЖУРНАЛ
З ПИТАНЬ ПРАВОВОЇ ІНФОРМАТИКИ,
ІНФОРМАЦІЙНОГО ПРАВА
ТА ІНФОРМАЦІЙНОЇ БЕЗПЕКИ**

НАУКОВО-ДОСЛІДНИЙ ЦЕНТР ПРАВОВОЇ ІНФОРМАТИКИ
АКАДЕМІЇ ПРАВОВИХ НАУК УКРАЇНИ
ІНСТИТУТ ЗАКОНОДАВСТВА ВЕРХОВНОЇ РАДИ УКРАЇНИ

ПРАВОВА ІНФОРМАТИКА

ЗАСНОВАНИЙ
У ГРУДНІ 2003 РОКУ

НАУКОВИЙ ЖУРНАЛ З ПИТАНЬ ПРАВОВОЇ
ІНФОРМАТИКИ, ІНФОРМАЦІЙНОГО ПРАВА ТА
ІНФОРМАЦІЙНОЇ БЕЗПЕКИ

ВИХОДИТЬ
ЩОКВАРТАЛЬНО

№ 1 (9)

січень – березень
2006

Редакційна колегія:

М.Я. ШВЕЦЬ (*голова редакційної колегії*), В.М. БРИЖКО (*заступник голови*),
М.І. КОВАЛЬ (*заступник голови*), В.В. БОНДАР, В.Д. ГАВЛОВСЬКИЙ, О.В. ГЛАДКІВСЬКА,
І.Б. ЖИЛЯЄВ, Л.М. ЗАДОРЖНЯ, А.П. ЗАКАЛЮК, І.О. ЗДЗЕБА, Р.А. КАЛЮЖНИЙ,
О.Л. КОПИЛЕНКО, О.Д. КРУПЧАН, О.П. ОРЛЮК, О.В. ПЕТРИШИН, В.М. ПОПОВИЧ,
Б.В. РОМАНЮК, М.Я. СЕГАЙ, В.М. СЕЛІВАНОВ, І.В. СЕРГІЄНКО,
В.П. ТИХИЙ, Ю.М. ТОДИКА, В.М. ФУРАШЕВ,
В.Г. ХАХАНОВСЬКИЙ, В.С. ЦИМБАЛЮК, В.К. ШКАРУПА

Засновники:

Науково-дослідний центр
правової інформатики
Академії правових наук України,
Інститут законодавства
Верховної Ради України

Редакція:

01032, м. Київ-32,
вул. Саксаганського, 110-В
Тел.: 234-94-56, 246-48-58
Факс: 234-55-60
e-mail: bib_rada@i.kiev.ua

Виготовлено:

Київська філія
державного підприємства
Науково-дослідного
економічного інституту
Міністерства економіки України

З М І С Т

- 5 **М. ШВЕЦЬ, О.ГЛАДКІВСЬКА, В. ЦИМБАЛЮК.** Проблеми взаємовідношення правової інформатики та інформаційної безпеки
- 12 **О. ЯРЕМЕНКО.** Правове регулювання доступу до офіційної правової інформації в Україні
- 18 **В. МАЦЮК.** Правові аспекти інформатизації
- 22 **В. ЛОЖКІН, В. ЦИМБАЛЮК.** Використання здобутків правової інформатики у боротьбі з корупцією
- 29 **М. КРАСНОСТУП, Г. КРАСНОСТУП.** Питання відповідальності за поширення недостовірної інформації, отриманої в мережі Інтернет
- 36 **Д. ЛАНДЕ.** 2GW – майбутнє Інтернету
- 44 **В. ХАХАНОВСЬКИЙ, В. СМАГЛЮК.** Термінологічні та організаційні аспекти створення інформаційно-аналітичної системи ОВС України
- 47 **В. БРИЖКО, М. ШВЕЦЬ.** Про економічний аспект захисту персональних даних у контексті права власності на інформацію
- 57 **М. ГУЦАЛЮК.** До питання ідентифікації особи за допомогою біометричних даних
- 61 **Д. СОПЛЬНЯК.** До питання засобів індивідуалізації найманого працівника
- 65 **С. ПОЗНЯКОВ.** Інформатизація комплексної системи детінізації відносин у сфері погашення податкового боргу платників податків
- 72 **О. ЗАГОРУЙ, Б. РАЦЬБОРИНСЬКИЙ.** Оцінка умов безпеки руху в зоні впливу автомобільної стоянки
- 78 **ВІД РЕДАКЦІЙНОЇ КОЛЕГІЇ**
РЕКОМЕНДАЦІЇ РАДИ ЄВРОПИ № R(87)15 від 17.09.1987 р.
“Про регулювання використання персональних даних у секторі поліції”
- 93 **До відома читачів**

Рекомендовано до друку Вченою радою НДЦПІ АПрН України, протокол № 12 від 19 грудня 2005 р.

Створення оригінал-макета – В. Брижко. Редактор – А. Москаленко.
Формат 70 x 108/16. Папір офсетний. Гарнітура Times.
Офсетний друк. Ум. друк арк. 8,6. Обл. вид. арк. 8,6.
Тираж: 100 прим. – паперовий варіант, 1000 прим. – електронний варіант на CD-ROM.

УДК 681.3

Д. ЛАНДЕ, кандидат технічних наук

2GW – МАЙБУТНЄ ІНТЕРНЕТУ

***Анотація.** Розглядаються проблеми сучасного інформаційного простору Інтернет, показано, що основні складності його ефективного використання пов'язані сьогодні зі специфікою традиційного подання інформації. Представлено новий підхід, який одержав назву “веб другого покоління” або 2GW, що базується на семантичних методах роботи з інформацією. Дано аналіз основних складових 2GW, особливостей їх реалізації на сучасному етапі, у тому числі й адаптивний інтерфейс уточнення запитів, реалізований за участю автора.*

За 15 років свого існування Інтернет перетворився на найбільший у світі розподілений інформаційний ресурс завдяки декільком закладеним у його основу принципам. До цих принципів відноситься реалізація гіпертексту, що дозволяє інтегрувати неоднорідні інформаційні ресурси, використання простої, доступної розумінню користувачів мови розмітки HTML (що обумовило легкість публікації документів у мережі), природну, адаптовану до людської логіки систему навігації в гіпертекстовому середовищі.

Разом з тим, можливості подання й доступу до інформації в Інтернеті обмежувалися статичністю мови HTML, що обумовлювала тільки навігаційний доступ до ресурсів, практичну відсутність підтримки метаінформації, недосконалість ідентифікації інформаційних ресурсів, і, найголовніше, той факт, що розмітка HTML відносилася тільки до зовнішнього подання документів, не стосуючись їх семантики.

На початку існування Інтернету невелика кількість веб-сайтів публікувала інформацію окремих авторів для відносно великої кількості відвідувачів. Сьогодні ситуація різко змінилася. Самі відвідувачі веб-сайтів стають авторами контенту, широко розвинулися форуми, “живі журнали” тощо. Це веде до різкого росту обсягів інформації: тільки у відкритій частині Інтернету міститься понад 20 млрд. документів.

У міру розвитку Інтернету першого покоління його можливості розширювалися, еволюційно були додані динамічні компоненти, можливість керувати стильовими рішеннями, були розроблені й деякі принципи подання контенту, зафіксовані як рекомендації. У процесі цієї еволюції з'явилися Java-аплети та Java-скріпти, численні мета-теги, мова каскадних таблиць стилів CSS та інше.

Разом з цим, традиційному веб все ж таки властиві такі недоліки, як високий рівень інформаційного шуму, неможливість гарантування цілісності документів, відсутність можливості змістовного пошуку, обмеженість доступу до “прихованого” веб.

Над вирішенням названих проблемам працюють численні колективи вчених і фахівців в усім світі, зокрема, консорціум W3C, де під керівництвом засновника Інтернет Тіма Бернерса-Лі реалізується концепція семантичного веб [1]. Поряд із цією концепцією, революційний прорив обіцяє дати більш загальний підхід, а саме – веб-2, або, як його називають, “веб другого покоління” (2GW) [2], що містить у собі реалізацію концепції Семантичного веб, багаторівневу підтримку мета-даних, нові підходи до дизайну й відповідного інструментарію, технологію глибинного аналізу текстів (Text Mining), а також ідеологію веб-сервісів, базуючись на інформаційних ресурсах, накопичених у веб першого покоління. Таким чином, 2GW передбачає перегляд усього комплексу стандартів й архітектурних принципів Інтернету.

Сьогодні очевидно, що центральною ланкою інструментарію подання й обміну даними буде Розширювана Мова Розмітки (XML) [3], що лежить в основі Семантичного веб. Передбачається також використання нового принципу ідентифікації інформаційних ресурсів, формування нової архітектури веб-простору на основі багаторівневого подання інформаційних ресурсів і стандартизованих веб-сервісів.

Передбачається, що 2GW на початку буде базуватися на ресурсах (базах даних, сайтах, Інтернет-співтовариствах) таких популярних Інтернет-компаній, як Google, Amazon, eBay тощо.

Семантичний веб

Однією з основних частин 2GW, яку її творці вважають абсолютно самодостатньою, є Семантичний Веб (Semantic Web). Концепцію Семантичного веб висунув Тім Бернерс-Лі, один з основоположників World-Wide-Web і голова веб-консорціуму (W3C) на міжнародній конференції XML-2000, що пройшла у 2000 році у Вашингтоні.

Основна ідея цього підходу полягає в організації такого подання даних у мережі, щоб допускалася не тільки їх візуалізація, але й ефективна автоматична обробка програмами різних виробників. Шляхом таких радикальних змін концепції традиційного веб- передбачається перетворення його на систему семантичного рівня. Семантичний веб повинен забезпечити “розуміння” інформації комп'ютерами, виділення ними даних, що найбільше підходять за тими або іншими критеріями, і вже після цього – надання інформації користувачам [4].

Семантичний веб можна представити як симбіоз двох напрямів, перший з яких охоплює мови подання даних. На сьогодні основними такими мовами є Розширювана Мова Розмітки XML (eXtensible Markup Language) і Засіб Опису Ресурсів RDF (Resource Description Framework). Існує також ряд інших форматів, однак XML й RDF надають більше можливостей, тому вони мають статус рекомендацій W3C.

Другий, концептуальний напрям несе в собі теоретичну уяву щодо моделей предметних областей, які в термінології Семантичного веб називаються онтологіями. 10 лютого 2004 року консорціумом W3C була затверджена й опублікована специфікація мови мережних онтологій OWL (Ontology Web Language).

У результаті дві гілки Семантичного веб спираються на три ключові мови (відповідно, технології) [5]:

- специфікація XML, що дозволяє визначити синтаксис і структуру документів;
- механізм опису ресурсів RDF, що забезпечує модель кодування для значень, певних в онтології;
- мова онтологій OWL, що дозволяє визначати поняття й відносини між ними.

Семантичний веб використовує також й інші мови, технології й концепції, зокрема, універсальні ідентифікатори ресурсів, цифрові підписи, системи логічного висновку тощо.

Практична реалізація Семантичного веб критично залежить від існування веб-сторінок, що містять мета-дані, формування яких не входить у стандартний процес веб-розробки. Навряд чи вдасться змусити авторів веб-сторінок вручну індексувати свої ресурси за допомогою термінологічних словників, онтологій Семантичного веб. Очевидно, що інтегрувати існуючі ресурси Інтернету в 2GW (що передбачено базовою концепцією) можна тільки автоматично. Дане завдання є дуже складним, вимагає глибокого аналізу текстів (Text Mining), якій, у свою чергу, сьогодні бурхливо розвиваються.

Як приклад реалізації такого підходу можна навести техніку й методологію австрійсько-швейцарської групи розробників, призначену для створення семантично анотованих веб-сторінок. Технологія WEESA (Web Engineering for Semantic Web Applications)

дозволяє здійснювати автоматичну генерацію мета-даних у форматі RDF для структурованого контенту веб-сторінок. Для генерації мета-даних використовується Java-програма, що бере контент одного або декількох атрибутів у якості вихідних даних і повертає стандартну тріаду RDF (“об’єкт – атрибут – значення”). За твердженням авторів технології, вони вже успішно застосували техніку WEESA для обробки веб-застосувань на сайті Міжнародного віденського фестивалю. Там були магазин квитків, більше 60-ти описів різних заходів, а також архів за останні 52 роки. Експеримент показав, що WEESA добре підходить для розробки веб-застосувань Семантичної Мережі.

Пошукові системи

Оскільки кількість веб-сайтів продовжує стрімко збільшуватися, користувачі 2GW мають потребу у більш ефективних пошукових системах. Пошукові машини наступних поколінь повинні будуть краще класифікувати інформацію й наочніше її представляти. У майбутньому пошук не буде обмежуватися лише обробкою уведених ключових слів. Наприклад, до уваги буде братися місце розташування користувача. Системи стануть відслідковувати інтереси користувачів, роблячи пошук більше цілеспрямованим. Нове програмне забезпечення буде працювати з мультимедійною інформацією так само легко, як з текстом. Нові пошукові машини будуть “бачити” опубліковані в мережі тексти, аудіо- і відеоматеріали, які в цей час недоступні.

Останнім часом одержали поширення адаптивні інтерфейси уточнення запитів [5], найчастіше реалізовані шляхом кластеризації результатів первинного пошуку. З’явилося таке поняття, як метод “папок пошуку” (Custom Search Folders), що не зв’язується з певним алгоритмом кластеризації, а являє собою безліч підходів, загальне в яких – спроба згрупувати результати пошуку й представити кластери у зручному для користувачів вигляді.

До подібних механізмів можна віднести, наприклад, австралійський пошуковий сервер Mooter (<http://www.mooter.com>), на якому застосовується візуальний підхід до надання результатів пошуку по оброблюваних запитах шляхом згрупування результатів первинного пошуку за категоріями. Інший пошуковий сервер iBoogie (<http://www.iboogie.com>) також групує результати пошуку, але відображає їх у вигляді, близькому до екрана провідника Windows. Слова й словосполучення в інформаційних портретах, застосовуваних, наприклад, у системі Галактика Зум, також дозволяють адаптивно уточнювати первинні запити.

В інформаційному центрі “Електронні вісті” за участю автора була розроблена система InfoStream [7], що застосовується для вирішення завдань автоматизованого збору новітньої інформації з веб-сайтів, її обробки й забезпечення доступу до неї в пошукових режимах. Ця система охоплює понад 1200 веб-джерел – більше 30000 унікальних новітніх повідомлень на добу, при цьому в ретроспективних базах даних зберігається понад 20 млн. повідомлень. Для ефективної роботи з такими обсягами інформації найпростішого інформаційного портрета виявилось замало – знадобився “інформаційний альбом” – багатоаспектна добірка параметрів вибірки за заздалегідь складеним запитом. І така можливість була реалізована. При цьому, на відміну від більшості подібних систем, в InfoStream уточнюючі параметри пошуку задаються не заповненням складної форми розширеного пошуку, а вказуються шляхом вибору з інформаційного альбому, одержуваного в результаті пошуку за первинним запитом. Сьогодні в системі InfoStream інформаційний альбом, що відповідає первинному запиту, містить такі параметри, як ключові слова, рубрики, мови, країни. Зокрема, в адаптивному інтерфейсі системи істотно полегшений множинний вибір джерел інформації, що відповідають заданому запиту. Крім того, користувачу надано можливість визначення характеристик розмірів документів,

що шукаються. Передбачено й такий “екзотичний” параметр, як рівень насиченості документів цифровою інформацією, що є корисним, наприклад, при пошуку аналітичних документів, цінкових таблиць, рейтингів тощо.

Експериментальною реалізацією ідеї колективної роботи в Інтернеті, що входить у концепцію 2GW, стала пошукова система Snap (<http://www.snap.com>), що забезпечує не тільки пошук веб-сторінок за ключовими словами, але й надає додаткову інформацію, близьку інтересам користувачів. Наприклад, до результату пошуку щодо виробників цифрових камер додається порівняльна таблиця моделей, які раніше були викликані іншими користувачами системи. Розроблювач системи Білл Гросс вважає, що Snap стала першою системою з “хвостовими даними” (data trail), але незабаром таких систем стане більше. Ця пошукова система є провісником такого етапу розвитку Інтернет, на якому в ній будуть активно використовуватися результати роботи всього співтовариства користувачів.

Нові пошукові системи поліпшують якість результатів, усе глибше зариваючись у доступні сховища інформації, сортуючи її, і представляючи результати з обліком персональних користувальницьких переваг. Так недавно портали Amazon, Ask Jeeves й Google оголосили про впровадження механізму поліпшення результатів пошуку, який базується на персоналізації. Пошукові машини www.A9.com (проект Amazon) і www.MyJeeves.ask.com (проект Ask Jeeves) не тільки відслідковують запити й знайдені веб-сторінки, але й дозволяють зберігати їх у вигляді закладок. Користувач MyJeeves може багаторазово переглядати накопичені результати, які являють собою персонально організовану область Інтернету. Подібні функції підтримує й портал www.A9.com, на якому пропонується набір сторінок, сформований при аналізі особистої пошукової історії. Історії пошукових запитів на сайтах A9 й MyJeeves зберігаються на серверах пошукових систем. У системі Google користувач може вибрати з ієрархічного списку найбільш важливі для нього теми й указати ступінь свого інтересу до тієї або іншої області знань. Всі ці дані враховуються системою при оцінці результатів пошуку.

“Прихований” веб

2GW припускає відкрити доступ до “прихованого” веб. Більша частина змісту сайтів Інтернету першого покоління залишається недоступною для пошукових машин, тому що багато веб-серверів зберігають і переробляють інформацію не у тому вигляді, в якому вона надається відвідувачеві. При цьому багато веб-сторінок генеруються тільки тоді, коли користувачі звертаються до них. Традиційні мережні агенти не вміють працювати з подібними ресурсами й не в змозі визначити їх зміст. “Прихований” веб охоплює в першу чергу вміст он-лайнних баз даних [8]. Прихованою є й швидко обновлювана інформація – новини, конференції, он-лайнні журнали.

У 2000 році американська компанія BrightPlanet (<http://www.brightplanet.com>) опублікувала сенсаційну доповідь [9], в якій стверджувалося, що в Інтернеті в сотні разів більше сторінок, ніж їх удалося проіндексувати найпопулярнішими пошуковими системами.

На сьогодні розроблений цілий клас програм, що одержали назву пакувальників (wrappers). У цих програмах, щоб одержати доступ до “прихованого” змісту веб-сторінок, використовується звичний синтаксис пошукових запитів і стандартний формат он-лайн ресурсів. В інших системах реалізуються переваги програмного інтерфейсу, що дозволяє використати стандартний набір команд й операцій.

Для пошуку в “прихованій” мережі, а саме – в тому її сегменті, що становлять бази даних, сьогодні вже існують деякі спеціалізовані ресурси. Серед них, наприклад, системи BigHub (<http://www.bighub.com>) і InvisibleWeb (<http://www.invisible-Web.net>) компанії

IntelliSeek. Сайт Invisible Web містить у собі каталог баз даних, більшість із яких не проіндексовані відомими пошуковими машинами. При введенні запиту цей сайт видає посилання на ресурси, за допомогою яких пошук необхідної інформації стане найбільш оптимальним. На цьому сайті Криса Шермана (Chris Sherman) і Гарі Прайса (Gary Price) зібрані колекції посилань на різні бази даних, серед яких є чимало унікальних ресурсів, наприклад, збірник спічів політиків і бізнесменів. Програмний пакет BullsEye компанії IntelliSeek здійснює пошук більш ніж у 800 мережних ресурсах.

У 2005 році компанія Yahoo також запустила тестову версію пошукового сервісу, орієнтованого на роботу з базами даних сайтів. Він може проводити пошук не тільки у загальнодоступних сайтах, але й у ресурсах, що надають платну інформацію, – таких, як он-лайн версія Wall Street Journal, що стягує з відвідувачів певну плату. Сервіс одержав назву DeepWeb і доступний поки що тільки для мешканців США та Великобританії.

Але все ж таки лідером серед навігаторів у “прихованому” веб є сайт CompletePlanet (<http://www.completeplanet.com>) компанії BrightPlanet. Цей сайт – найбільший каталог, що нараховує понад 100 тисяч посилань. Компанія BrightPlanet також створила персональну утиліту для пошуку в он-лайн баз даних – LexiBot, яка може забезпечувати пошук у декількох тисячах пошукових систем “прихованого” веб. Метапошуковий пакет DeepQueryManager (DQM) цієї ж компанії забезпечує пошук у 55 тисячах “прихованих” веб-ресурсах.

Пошук і глибинний аналіз текстів

Пошукові технології 2GW повинні стати більш ефективними за рахунок потужних технологій, що поєднують пошук і глибинний аналіз текстів (Text Mining), знаходження аномалій і трендів у текстах. Одночасно ці технології будуть невидимими завдяки операціям інтелектуального пошуку, що вбудовані в інтерфейси “за замовчуванням”. В остаточному підсумку пошук інформації в 2GW стане нерозривно пов'язаним з її осмисленням.

Існує чотири основних види застосувань технологій Text Mining, які повинні знайти своє втілення в веб другого покоління [10]:

- класифікація тексту, у якій використовуються статистичні кореляції для побудови правил розміщення документів у визначені категорії;
- кластеризація, що базується на ознаках документів та використовує лінгвістичні й математичні методи без використання визначених категорій. Результат – таксономія або візуальна карта, що забезпечує ефективне охоплення великих обсягів даних;
- семантичні мережі, або аналіз зв'язків, які визначаються дескрипторами (ключовими фразами) у документах для забезпечення навігації;
- витяг фактів, призначений для одержання деяких фактів з тексту з метою поліпшення класифікації, пошуку й кластеризації.

Нещодавно компанія Google представила свої нарощені й плани щодо кластеризації знайдених документів у рамках технології Text Mining. Демо-версія цієї системи дозволяє виділяти з документів назви компаній, які є основними критеріями кластеризації.

Можна назвати ще кілька завдань технології Text Mining, наприклад, прогнозування, що полягає в тому, щоб прогнозувати за значеннями одних ознак об'єкта значення інших. Ще одне завдання – знаходження аномалій, тобто пошук об'єктів, які своїми характеристиками сильно виділяються із загальної маси. Для цього спочатку з'ясовуються середні параметри об'єктів, а потім досліджуються ті об'єкти, параметри яких найсильніше відрізняються від середніх значень. Подібний аналіз часто проводиться після класифікації для того, щоб з'ясувати, наскільки остання була точною.

Трохи окремо стоїть завдання пошуку пов'язаних ознак (феноменів, понять) в окремих документах. Від прогнозування це завдання відрізняється тим, що заздалегідь невідомо, за якими саме ознаками реалізується взаємозв'язок; ціль саме в тому й полягає, щоб знайти зв'язки ознак. Це завдання подібне до кластеризації, але не за безліччю документів, а за безліччю властивих їм ознак.

І нарешті, для обробки та інтерпретації результатів Text Mining велике значення має візуалізація. Візуалізація в 2GW на основі систем Text Mining передбачається як засіб надання контенту всього масиву документів, а також для реалізації навігаційного механізму, що може застосовуватися при дослідженні документів та їх класів.

Веб-сервіси

Одним із ключових елементів 2GW є веб-сервіси – автономні, модульні додатки, призначені для реалізації інформаційних процесів у мережі (зокрема, бізнес-процесів [11]). Веб-сервіси спираються на ряд галузевих стандартів: WSDL (для опису), UDDI (для інформування й публікації) і SOAP (для обміну повідомленнями).

У серпні 2002 року, усвідомивши складність звернення до веб-сервісів у синхронному й асинхронному середовищах, корпорації BEA, IBM, Microsoft, SAP і Siebel у результаті спільних зусиль розробили мову реалізації бізнес-процесів для веб-сервісів (Business Process Execution Language for Web Services, BPEL4WS або просто BPEL). Мова BPEL дозволяє описувати бізнес-процеси й те, як вони пов'язані з веб-сервісами, а також як бізнес-процеси використовують веб-сервіси для досягнення поставлених завдань. BPEL можна розглядати як декларативно-процедурну мову програмування. BPEL являє собою діалект мови XML. Як й у будь-якій мові програмування, в BPEL визначені зарезервовані слова (теги XML):

- Виклик операції за допомогою веб-сервісу (<invoke>).
- Очікування зовнішнього повідомлення (<receive>).
- Генерація відповіді для вхідних/вихідних даних (<reply>).
- Очікування протягом деякого часу (<wait>).
- Копіювання даних між позиціями (<assign>).
- Індикація помилки або збійної ситуації (<throw>).
- Зупинка реалізації всього сервісу (<terminate>).
- Відсутність дій (<empty>).
- Визначення послідовності виконання дій (<sequence>).
- Розгалуження за допомогою оператора вибору (<switch>).
- Визначення циклу (<while>).
- Виконання одного з декількох альтернативних маршрутів (<pick>).
- Індикація того, що крок повинен бути виконаний паралельно (<flow>).
- Індикація обробки помилкової логіки за допомогою <throw> й <catch>.

На цей час уже існує безліч веб-сервісів, однак іншим програмам немає можливості розшукати в мережі веб-сервіс, що виконує ту або іншу функцію. Необхідний для підвищення ефективності роботи веб другого покоління процес, який називають виявленням сервісів, стане можливим лише після того, як пошириться наведена вище або подібна їй уніфікована мова, яка дозволяє описувати сервіси, для того щоб агенти могли “розуміти”, що дозволяє робити даний сервіс та яким чином ним користуватися. Наприклад, у рамках Семантичного веб-агенти виробника сервісу й агенти його користувачів можуть досягти розуміння один одного шляхом обміну онтологіями, що містять необхідні для спілкування термінологічні словники. Більше того, агенти зможуть навіть самі, знаходячи нові онтології, удосконалювати свої алгоритми.

Семантика мови опису сервісів (наприклад, BPEL) дозволяє агентів описувати, які саме функції він може виконувати і які вхідні дані йому потрібні. Технологія виявлення веб-сервісів відразу ж знайде своїх користувачів. Наприклад, у сфері малого бізнесу стане набагато простіше налагоджувати проведення трансакцій в області електронної комерції, що мають більший ступінь захисту й автоматизації.

Формат RSS

У веб другого покоління інформація неіто “відчужується” від джерела. Відповідно, передбачається широке застосування формату RSS (Really Simple Syndication, Rich Site Summary, RDF Site Summary), спеціально призначеного для легкого й швидкого обміну змістом веб-сайтів [12]. RSS забезпечує погоджений засіб резюмувати зміст веб-сайтів, а крім того, його застосування дозволяє адміністраторам сайтів новин, онлайнових щоденників, форумів та інших часто оновлюваних веб-ресурсів одержувати простий уніфікований метод подачі інформації про події, що відбуваються.

Сьогодні RSS прийнято розглядати у першу чергу як формат, призначений для публікації й забезпечення експорту новин на новинних сайтах. Після того як інформація перетворена у формат RSS, програма, орієнтована на цей формат, може завантажувати відомості про відновлення веб-сайтів і залежно від результату виконувати певні дії, наприклад, автоматично оновлювати список актуальних інформаційних повідомлень.

Користувачі можуть одержати доступ до даних у форматі RSS за допомогою спеціальних програм, які називають RSS-агрегаторами. Програма-агрегатор дозволяє групувати публікації з різних джерел, забезпечуючи можливість одночасно стежити за появою новин на всіх сайтах, не вимагаючи відвідування кожного сайту окремо. При цьому, звичайно ж, не потрібно завантажувати з мережі зайвої інформації, що стосується, наприклад, оформлення веб-сторінок.

Програми-агрегатори (або парсери) виконують синтаксичний розбір даних, наведених у форматі RSS, після чого можуть реалізовувати будь-які дії стосовно цих даних, приміром, відображати їх на обраному веб-сайті.

Перспективність і популярність RSS як стандарту обумовлена насамперед його доступністю й простотою. Сьогодні практично всі провідні інформаційні сайти у світі використовують RSS як інструмент оперативного подання своїх оновлень.

Про перспективність RSS уже сьогодні свідчать і спроби використання її в рекламному бізнесі. На конференції веб 2.0, що проходила у Сан-Франциско, один з керівників компанії Yahoo Ден Розенвейг (Dan Rosensweig) заявив, що їхня система контекстної реклами Overture буде експортувати посилання в RSS-канали.

Дизайн

2GW розглядається у першу чергу як ефективне середовище для роботи з контентом. Природно, нова концепція висуває нові вимоги до засобів візуалізації інформації, дизайну. Сьогодні створення інструментарію дизайнерів для роботи з веб другого покоління є передовим фронтом впровадження технологій 2GW. Уже сьогодні створюються інтерфейси, які агрегують інформацію з тисяч джерел. Так Amazon.com (<http://www.amazon.com>) дає доступ до своєї бази даних через відкритий API. Кожен бажаючий може створити персоналізований, більш дружній, на його думку, інтерфейс користувача, що володіє функціональністю сайта-першоджерела (наприклад, Amazon Light, <http://www.kokogiak.com/amazon>).

Як основну мову розмітки веб-сторінок передбачається використати XML. Колишні мови розмітки – HTML й XHTML – вирішували переважно завдання відображення інформації, у той час як XML призначений для її опису.

Передбачається, що у 2GW буде реалізовані персоналізована, незалежна навігація й керування веб-сайтом. Тобто користувач сам зможе контролювати візуальний інтерфейс. Разом з цим, передбачається, що дизайнер ресурсів 2GW буде у більшій мірі програмістом, якій за допомогою інструментальних засобів буде визначати елементи структури, навігації й дизайну веб-сайта [2]. Технологічні, інтуїтивні інтерфейси – це те, до чого повинні прагнути дизайнери сайтів 2GW. Як найбільше технологічні сайти нового покоління вже сьогодні можна назвати картографічний сервіс Google Maps (<http://www.maps.google.com>), фотосервіс Flickr (<http://www.flickr.com>), а також Інтернет-співтовариство Del.icio.us (<http://www.del.icio.us>).

Перспективи

Передбачається, що наступний щабель розвитку Інтернету буде визначатися технологіями роботи з величезним обсягом інформації, що накопився у мережі. Зокрема, веб другого покоління повинен характеризуватися переходом від мережі документів до мережі даних, які при необхідності агрегують у семантично зв'язані документи за допомогою веб-сервісів нового покоління. Передбачається існування єдиного інформаційного простору у вигляді безлічі одиниць даних, які можуть розміщатися на численних сайтах. Користувач буде одержувати документ шляхом агрегування у себе на комп'ютері інформаційних одиниць, розподілених в Інтернеті.

Перспективи 2GW будуть багато у чому залежати від інфраструктури, у рамках якої будуть працювати програмні продукти з боку веб-серверів і користувачів. На думку багатьох учених й учасників Інтернет-ринку, веб другого покоління буде більшою мірою, ніж сьогодні пристосований для автоматизованої обробки, використання комп'ютерами. Завдяки цьому споживачі будуть мати справу з інформацією, зібраною провідними інформаційними компаніями, і створювати нові сервіси.

Використана література

1. Tim Berners-Lee, James Hendler, Ora Lassila, The Semantic Web, Scientific American, May 2001 (<http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21>).
2. Chris Preimesberger. Web 2.0: Possibly the best IT business conference of 2004 // NewsForge, 2004 (<http://www.newsforge.com/article.pl?sid=04/10/08/0849201>).
3. Ландэ Д.В. На границе стихий // ЧИП-Украина. – 2003. – № 5. – С. 72-77.
4. Семантический Вэб: воплощение идеи Телеком. – 2005. – № 6. – С. 60-65.
5. Фурашев В.М., Ландэ Д.В., Григор'ев О.М., Фурашев О.В. Електронне інформаційне суспільство України: погляд у сьогодення і майбутнє. – К.: Інжиніринг. – 2005. – 164 с.
6. Григорьев А.Н., Ландэ Д.В. Адаптивный интерфейс уточнения запросов к системе контент-мониторинга InfoStream // Труды Международного семинара “Диалог’2005”. – 2005. – С. 109-111.
7. Григорьев А.Н., Ландэ Д.В. Система мониторинга новостей InfoStream – информационное пространство из одних рук. Построение информационного общества: ресурсы и технологии // тезисы докладов и информационные материалы XI международной научно-практической конференции. – К.: УкрИНТЭИ. – 2005. – С. 17-20.
8. Ландэ Д.В. Затерянный вэб // “Телеком”. – 2005. – № 1. – С. 46-51.
9. Danny Sullivan. Invisible Web Gets Deeper // The Search Engine Report. – 2002. (<http://searchenginewatch.com/sereport/article.php/2162871>).
10. Ландэ Д.В. Поиск знаний в Internet. Профессиональная работа // – М.: Издательский дом “Вильямс”. – 2005. – 272 с.
11. Шапошников И. Веб-сервисы Microsoft.NET // – СПб.: БХВ-Петербург. – 2002. – 334 с.
12. Ландэ Д.В, Морозов А.Ю. Новостной Интернет // “Телеком”, – 2005. – № 1-2. – С. 58-62.

