

УДК 004.02; 004.6

Золотар І. – ст. гр. СКмз-61

Тернопільський національний технічний університет імені Івана Пулюя

ВЕБ-ПРОСТІР

Науковий керівник: ст. викладач Дуда О.М.

Zolotar I.

Ternopil Ivan Pul'uy National Technical University

WEB SPACE

Supervisor: Duda O.

Ключові слова: Веб-простір, інформація

Keywords: Web space, information

Веб-простір, заснований на фізичній інфраструктурі мережі Інтернет та протоколі передачі даних HTTP, об'єднує сотні мільйонів веб-серверів, підключених до мережі Інтернет. На початку існування веб-простору на невеликій кількості веб-сайтів публікувалася інформація окремих авторів для відносно великої кількості відвідувачів. Сьогодні ситуація різко змінилася, відбувся перехід до веб другого покоління. Самі відвідувачі веб-сайтів активно беруть участь в створенні контенту, що призвело до різкого зростання об'ємів інформації та динаміки веб.

Сьогодні у веб вже існує вільно доступна для користувачів інформаційна база такого об'єму, який раніше важко було уявити. Більш того, об'єми цієї бази перевищують на порядки все те, що було доступно десятиліття назад. В серпні 2010 року компанія Yahoo! оголосила про те, що проіндексувала близько 40 млрд. документів. Досягнення компанії Google в 2009 році становило менше 50 млрд. документів. Сьогодні Google заіндексувала понад трильйон веб-документів. За даними служби Netcraft Web Server Survey [1], в даний час кількість веб-серверів перевищує 670 млн.

У відкритих джерелах і спеціалізованих базах даних, доступних у веб-просторі, міститься велика частина інформації, необхідної для проведення аналітичних досліджень, однак залишаються відкритими питання її знаходження і ефективного використання. При використанні веб-простору як найпотужнішого джерела інформації найсуттєвішими є проблеми об'єму, навігації, наявності інформаційного шуму і динамічного характеру інформації в Інтернет.

Можливості доступу до Інтернет-ресурсів, що привертає своєю відкритістю, об'ємами і змістовною багатогранністю, на перший погляд здаються безмежними. Однак важливі події в різних областях свідчать про зворотне. Саме в кризових ситуаціях Інтернет досить часто підводить. Існує безліч проблем – від перевантаженості мережевої інфраструктури – до вірусних атак, вразливостей і відмов в обслуговуванні окремих веб-серверів. Цілий ряд проблем породжено також об'ємами, різноманітністю уявлення і динамікою контентного сегменту інформаційного простору.

Незважаючи на такі якості, як відкритість і доступність, існуючу інфраструктуру веб-простору можна визнати надійною та достовірною. Наведемо ряд проблем, притаманних веб-простору:

– не вирішена задача доступу користувачів до різномірних веб-ресурсів з «одного вікна» для отримання узагальненого представлення потоків інформації з необхідної тематики;

– не забезпечена можливість своєчасного «нагадування» і «проштовхування» профільної для користувача інформації, що публікується на великій кількості веб-сайтів;

– досить велика ймовірність відмови в обслуговуванні критично важливих веб-ресурсів в найнеобхідніший час.

Відомо, що сьогодні існують технології інтеграції контенту, які дозволяють частково вирішувати названі проблеми, забезпечуючи ефективний пошук та навігацію у веб-просторі, моніторинг та агрегацію відкритих веб-ресурсів.

Для професійного пошуку у веб-просторі і моніторингу інформації використовується спеціалізоване програмне забезпечення, інформаційно-пошукові системи і сервіси. Наведемо деякі приклади програмних продуктів:

Copernic Agent [2] – програмний засіб, що дозволяє проводити мета-пошук, використовуючи 1000 пошукових систем, об'єднувати результати, усувати дублікати, блокувати неробочі посилання, показувати найбільш релевантні результати.

Newprosoft Web Content Extractor [3] – програмний засіб сканування і отримання даних з веб-сайтів.

Portable Offline Browser від MetaProducts Corporation [4] – програмний засіб, що дозволяє скачувати необхідні веб-сайти і мультимедіа-інформацію, в тому числі Flash-анімацію, скрипти і активний вміст сторінок.

WebSite-Watcher [5] – програмний засіб, що дозволяє проводити моніторинг веб-сайтів, форумів, локальних файлів, що забезпечує фільтрацію інформації, а також зручну візуалізацію результатів моніторингу.

В якості сервісних рішень приведемо:

WatchThatPage [6] – безкоштовний сервіс, що дозволяє автоматично збирати нову інформацію з веб-ресурсів, поставлених на моніторинг.

Diphur Monitor EveryThing [7] – безкоштовний сервіс моніторингу будь-яких веб-сайтів, що повідомляє про їх оновлення і доставляє користувачам оновлення.

Література.

1. Most Reliable Hosting Company Sites in March 2016 // Internet Research, Anti-Phishing and PCI Security Services | Netcraft. – URL: <http://news.netcraft.com/>. – Last access: april 2016 y. – Title from screen.

2. Desktop Search 5 // copernic. – URL: <http://www.copernic.com/en/products/desktop-search/>. – Last access: april 2016 y. – Title from screen.

3. NewProsoft. Web data extractions software. – URL: <http://www.newprosoft.com/>. – Last access: april 2016 y. – Title from screen.

4. Learning About Hiring Pest Control Experts // Portable Offline Browser. – URL: <http://portableofflinebrowser.com/>. – Last access: april 2016 y. – Title from screen.

5. Automatically check web pages for updates and changes // WebSite-Watcher. Save Time, Stay Informed! – URL: <http://www.aignes.com/>. – Last access: april 2016 y. – Title from screen.

6. Monitor pages, extract new information // WatchThatPage – Monitor web pages extract new information. – URL: <http://watchthatpage.com/>. – Last access: april 2016 y. – Title from screen.

7. You bookmark it, we notify you what changes. It's really that simple! // Diphur Monitor EveryThing. – URL: <https://diphur.com/intro>. – Last access: april 2016 y. – Title from screen.