

Тема 4. Пошук інформації у мережі Internet

4.1. Веб-браузери. Їх призначення та функціональні можливості.

4.1.1. Призначення та функції браузерів

Браузер – це програма, що використовується для перегляду веб-сайтів. Вона встановлює зв'язок із веб-сервером, завантажує на комп'ютер сторінку, розміщену за вказаною користувачем адресою, та відтворює цю сторінку на екрані.

Вигляд веб-сторінки (гіпертекстового документа) у вікні браузера залежить, від використаних у ній тегів – команд мови розмітки HTML. Стиль оформлення сайтів визначає також інформація, записана мовою CSS (Cascade Style Sheets – каскадні таблиці стилів) у HTML-документах чи окремих файлах. Сучасні браузери здатні відображати ілюстрації, відтворювати анімацію, відеоролики і звук.

Одна з головних особливостей браузерів полягає у тому, що вони надають можливість переміщатися всім простором Всесвітньої павутини. Переглядаючи з їх допомогою веб-сторінки, на яких є гіперпосилання, користувач з легкістю, лише клацаючи відповідні посилання, переходить з однієї такої сторінки на іншу.

4.1.2. Можливості й інтерфейс популярних браузерів

Усі сучасні браузери мають приблизно однакові можливості: забезпечують високу швидкість виведення веб-сторінок, підтримують роботу як з багатьма вікнами (кожна веб-сторінка відкривається в окремому вікні браузера), так і в режимі одного вікна (кожна веб-сторінка відкривається на окремій вкладці одного вікна браузера), дозволяють з допомогою панелі інструментів вводити критерії пошуку і вибирати пошукову систему, пропонують вбудовані завантажувачі файлів і здатні забезпечити високий рівень безпеки під час роботи. Далі ми стисло опишемо переваги та недоліки найпопулярніших браузерів.

Internet Explorer. Найбільшу популярність у користувачів Інтернету традиційно виборює браузер Internet Explorer (рис.2). Він не лише входить до складу кожної нової версії операційної системи Windows, а й є доступним для безкоштовного завантаження з сервера корпорації Microsoft (<http://microsoft.com/>) і встановлення в операційній системі Windows

попередніх версій (наприклад, у Windows XP). Слід визнати, що перші версії браузерів Internet Explorer дуже рідко використовувалися як браузери на «кожний день», хоча й були встановлені на більшості комп'ютерів. Проте в Internet Explorer версії 9 було внесено суттєві зміни, а саме перероблено інтерфейс, збільшено швидкість завантаження сторінок і покращено безпеку програми. Завдяки цьому браузер Internet Explorer 9 має хорошу функціональність.

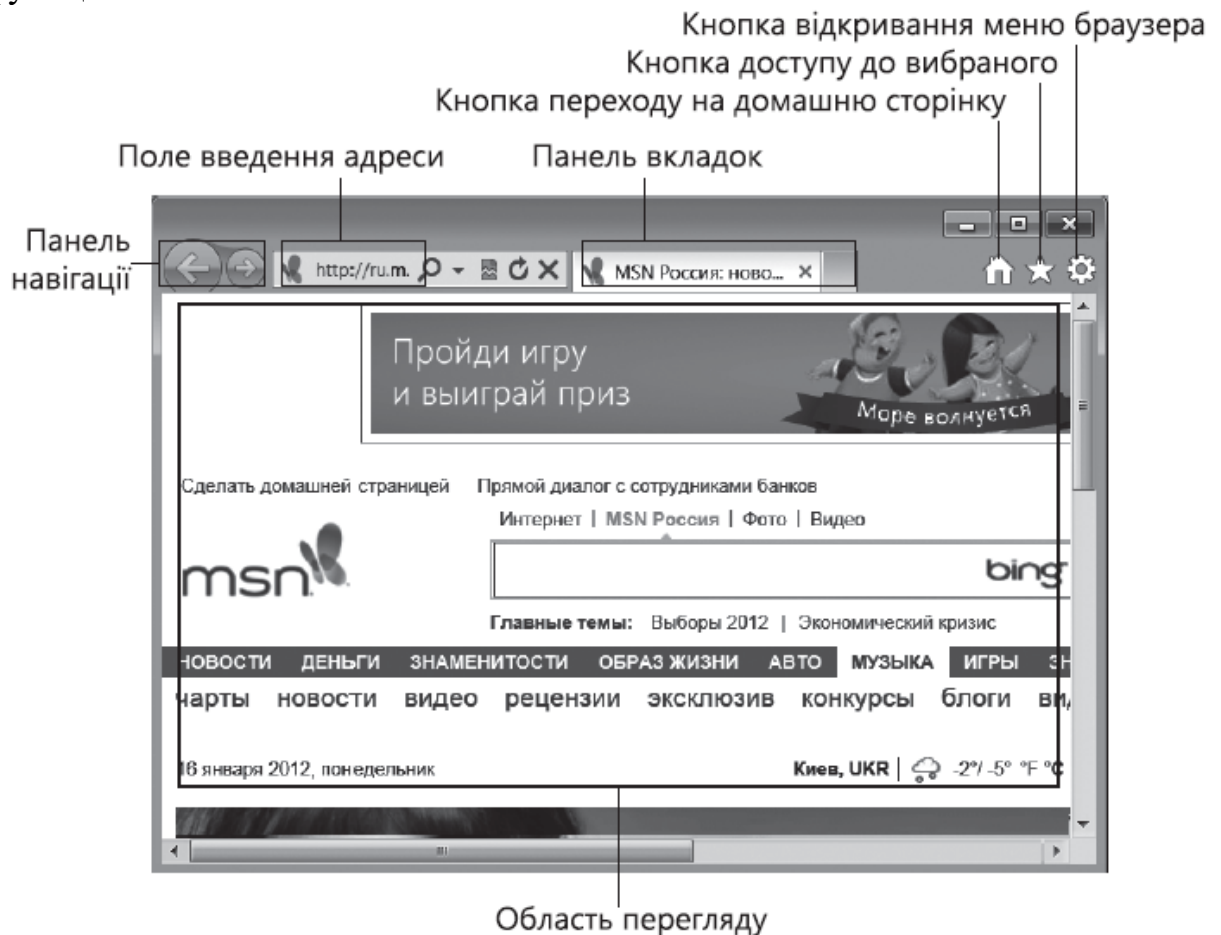


Рис. 2. Вікно браузера Internet Explorer 9

Головною перевагою цього продукту є те, що абсолютна більшість веб-сторінок відображається в ньому саме так, як задумав веб-дизайнер. Браузер має простий інтерфейс, містить засоби керування пошуком інформації в Інтернеті, надає легкий спосіб налаштування параметрів безпеки та інших корисних функцій, зокрема блокування спливаючих вікон, які під час відвідування деяких сайтів відкриваються автоматично і зазвичай містять інформацію рекламного характеру. Серед недоліків Internet Explorer слід відзначити неможливість змінювати інтерфейс і відсутність розширень від сторонніх розробників.

Mozilla Firefox. Головна перевага Mozilla Firefox (www.mozilla.com/ru/firefox) – браузера, створеного компанією Mozilla (рис.3), полягає у тому, що кожен користувач може вибрати доповнення (міні-програми), які розширюють вбудовані функціональні можливості браузера, а також дозволяють змінити його вигляд (теми). Велика кількість

доповнень міститься на сайті Firefox Add-ons за адресою <https://addons.mozilla.org>.

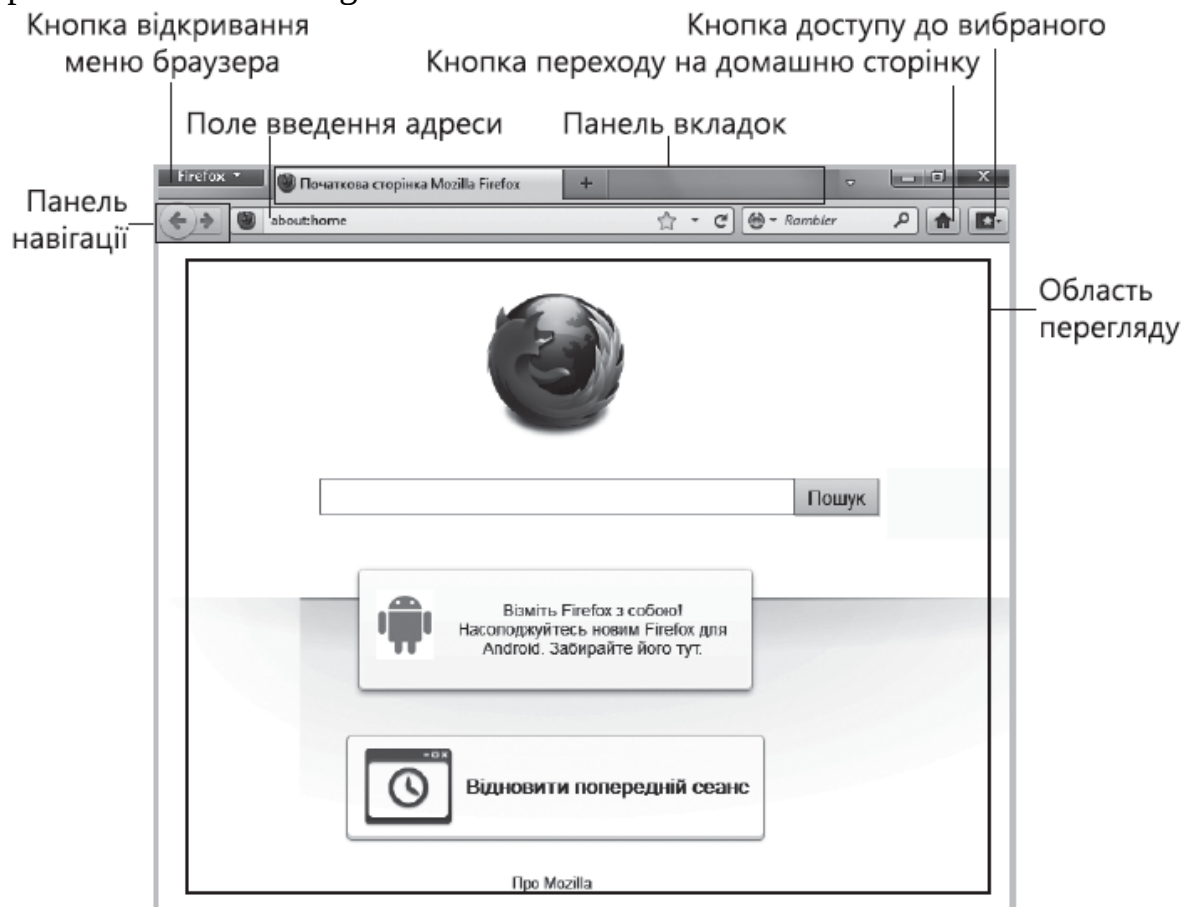


Рис. 3. Вікно браузера Mozilla Firefox

З допомогою доповнень можна розширити можливості пошуку, додати в браузер ігри, зробити більш зручним керування вкладками, підвищити безпеку браузера, відключити набридливу рекламу. Кожне доповнення має автора, назву і рейтинг (чим більше жовтих зірочок, тим він вищий). Встановити доповнення дуже просто – достатньо клацнути кнопку Add to Firefox, яка з'являється після наведення на нього вказівника миші (рис. 4). Одним з найпопулярніших вважається доповнення Easy YouTube Video Downloader, яке дозволяє зберігати відеоролики з сайту YouTube на жорсткий диск.

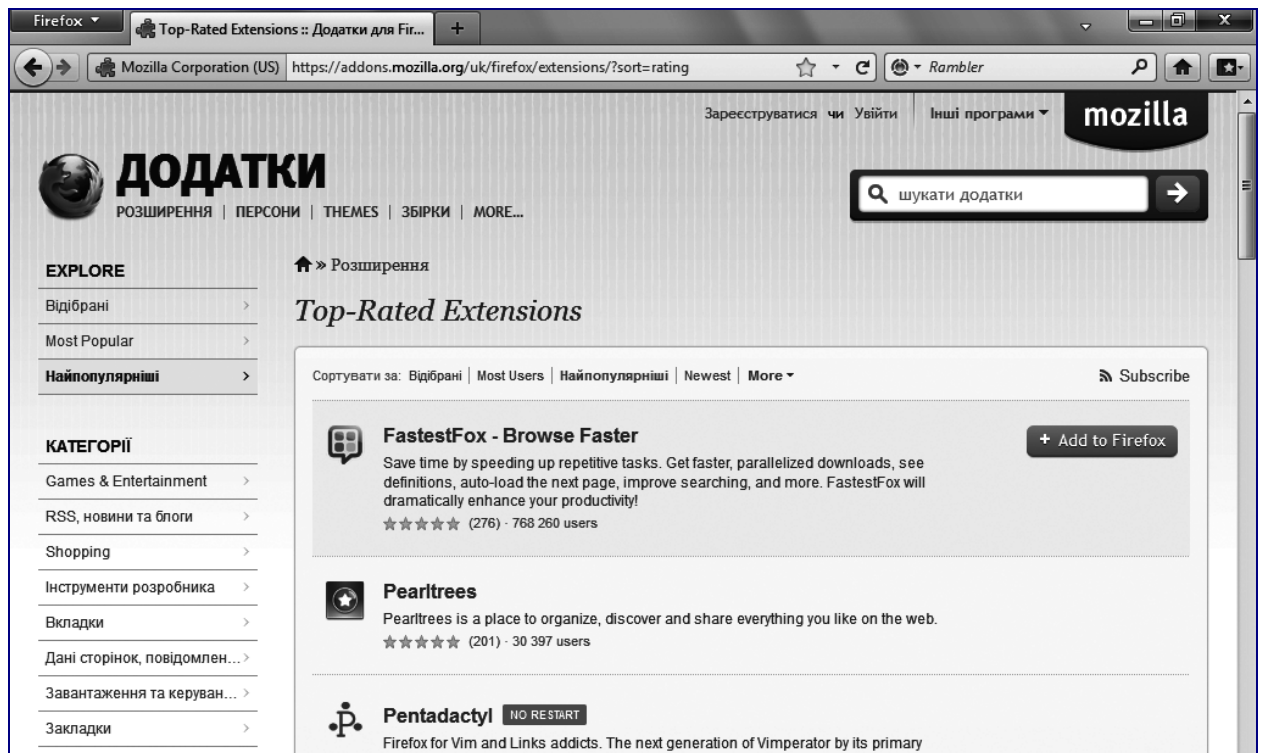


Рис. 4. Сторінка доповнень і кнопка Add to Firefox

Недоліком Mozilla Firefox є те, що цей браузер не вирізняється високою стабільністю роботи і потребує багато оперативної пам'яті, що інколи призводить до «гальмування» всієї системи.

Opera. Інтегрований пакет програм норвезької компанії Opera Software (<http://www.opera.com/>) є серйозним конкурентом продуктів Microsoft і Mozilla. Браузер Opera (рис. 5) невимогливий до ресурсів комп'ютера, характеризується високою швидкістю роботи, має зручний інтерфейс, панель інструментів, що налаштовується, а також чудовий засіб для масштабування відображуваних сторінок, що містять не лише текст, а й графіку. Стан браузера Opera може зберігатися після виходу з його вікна та відновлюватися під час його наступного відкриття. Працюючи в режимі Turbo, можна в декілька разів пришвидшити завантаження веб-сторінок і зекономити трафік. Така можливість буде доречною насамперед тим користувачам, що мають повільне або дороге інтернет-з'єднання, наприклад мобільний доступ до мережі. Режим Turbo вмикається клацанням кнопки Turbo, що має вигляд спідометра та розташована в лівому нижньому куті вікна.

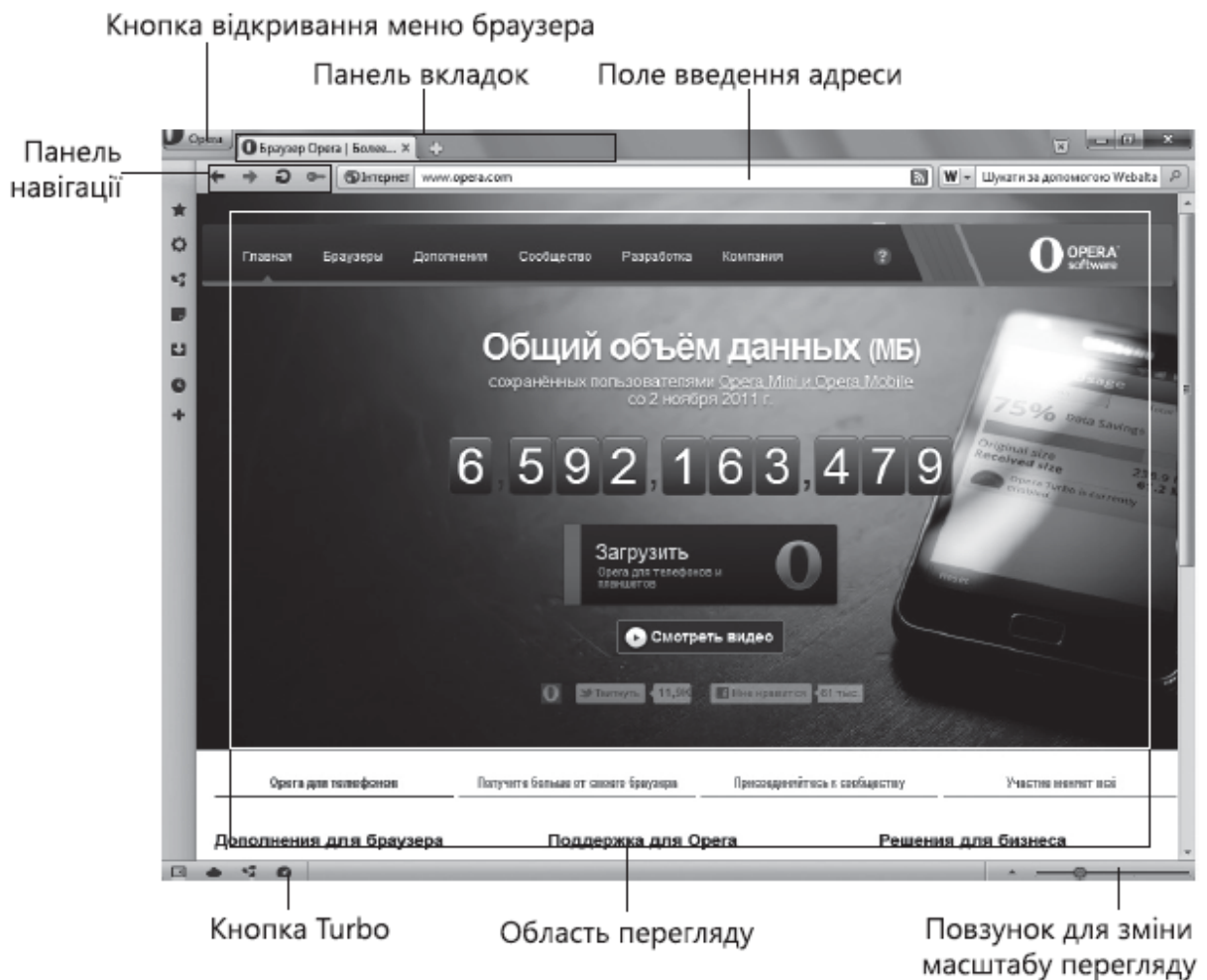


Рис. 5. Вікно браузера Opera

Саме у браузерах Opera були вперше розширені функції миші. Наприклад, для повернення на попередню сторінку достатньо, утримуючи праву кнопку миші, трохи перемістити курсор ліворуч.

Як і в Mozilla Firefox, в Opera можна встановлювати розширення (міні-програми, що збільшують кількість функціональних можливостей браузера) і теми. Щоб встановити розширення, потрібно знайти його на сайті <https://addons.opera.com/ru/addons/extensions/> і клацнути розташовану праворуч кнопку Установить.

Мабуть, єдиним недоліком Opera, який іноді дає про себе знати і призводить до зависання усієї системи (на щастя, не на всіх комп'ютерах), є раптове збільшення використання ресурсів процесора до 100 %.

Google Chrome. Google Chrome (www.google.com/chrome) – порівняно молодий браузер, випущений компанією Google (рис.6). Він позиціонується розробниками як швидкий, надійний та безпечний. І дійсно, стосовно надійності він найкращий. Справа в тому, що у разі збою закривається лише та веб-сторінка, яка стала причиною помилки, а всі інші веб-сторінки залишаються відкритими. Інші ж браузери в разі збою закривають свої вікна повністю, з усіма відкритими вкладками, тож їх доводиться завантажувати заново.

Google Chrome також підтримує розширення і теми, доступні за адресою <https://chrome.google.com/webstore/category/extensions>. Щоб встановити розширення, потрібно клацнути посилання Увійти в правому верхньому куті сайту і ввести дані свого облікового запису Google. Після цього слід знайти потрібне розширення, навести на нього вказівник миші та клацнути кнопку ДОДАТИ В CHROME.

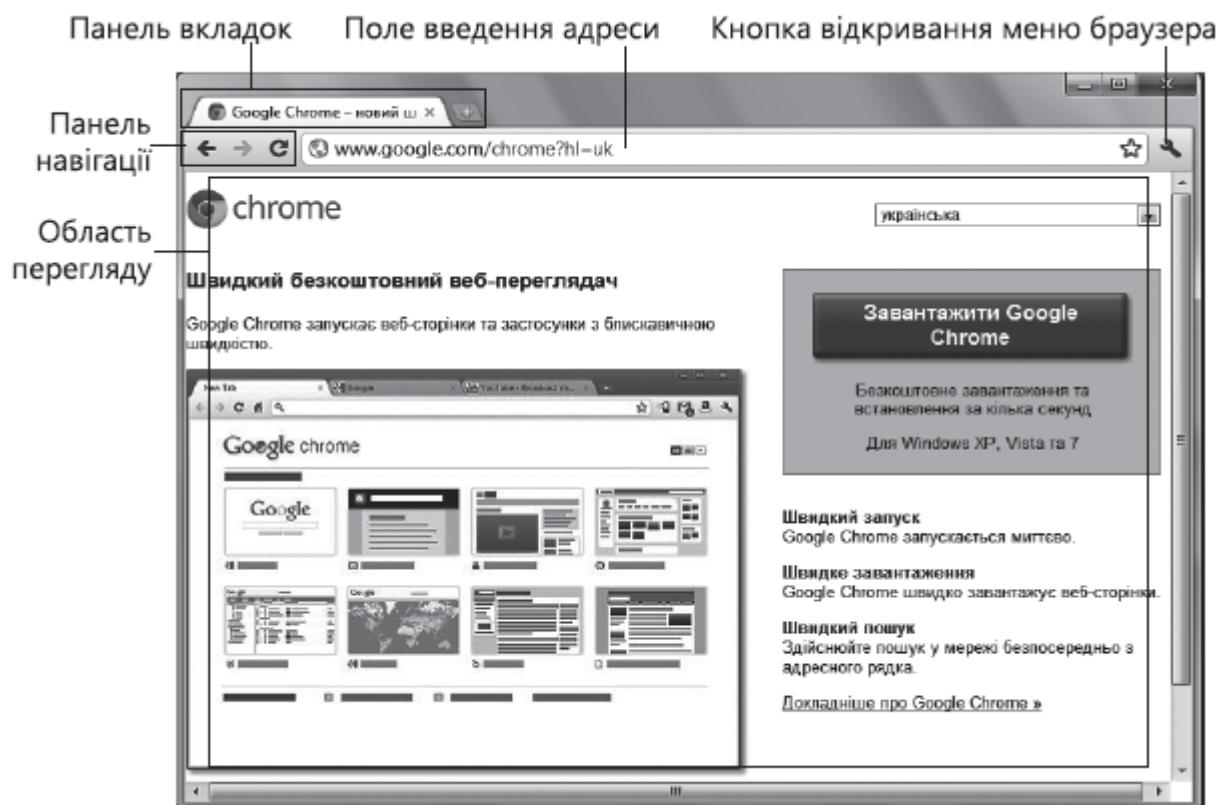


Рис. 6. Вікно браузера Google Chrome

4.1.3. Домашня сторінка браузера

Відразу після запуску браузер завантажує веб-сторінку, що задана в його настройках як домашня. Загалом кажучи, домашня сторінка браузера – це сторінка, яку він відображає за умовчанням. Як домашню зазвичай задають сторінку, з якої зручно розпочинати сеанс роботи з Інтернетом; це може бути, наприклад, веб-сторінка певної пошукової системи. У деяких браузерах, зокрема в Internet Explorer та Mozilla, перейти на домашню сторінку можна будь-коли під час роботи, клацнувши кнопку Додому (із зображенням будиночка).

4.1.4. Поняття cookie та кешу

Під час відвідування будь-якої веб-сторінки браузер зберігає файли, що її утворюють, на жорсткому диску комп'ютера у спеціальному місці – кеші, щоб у разі повторного відкриття цієї сторінки не завантажувати їх знову з мережі, а використовувати вже наявні. Інші тимчасові файли називаються

cookie. Це створені веб-сервером і збережені на комп'ютері текстові файли, в яких містяться індивідуальні настройки якогось сайту або особисті дані користувача.

4.2. Пошукові системи. Огляд популярних пошукових серверів.

4.2.1. Пошукові системи і каталоги

Пошукові системи – це складні технічні комплекси, що містять десятки швидкодіючих комп'ютерів, обслуговування яких ведуть сотні фахівців.

Якщо розумно використати пошукову систему, можна протягом достатньо короткого часу знайти інформацію, на пошук якої без використання Інтернет можна витратити місяці і навіть роки. Але, практика доводить, що ефективно і правильно використовувати пошукові системи вміють не більше 3% користувачів. Зазвичай, пересічний користувач на запит з 1-2 слів отримує інформацію, що охоплює значно більший простір ніж потрібно.

Пошукові системи інакше називають: пошуковими засобами або пошуковими машинами, жаргонною назвою є пошуковики. Еквівалентними іноземними термінами є: англійською - *Search Engines*, німецькою - *Suchmaschinen*, французькою - *Le systeme de prospection*.

Пошукові системи містять три основні компоненти:

- веб-сторінка з пошуковим механізмом, яку користувачі використовують як інтерфейс для взаємодії з базою даних;
- база даних, де міститься інформація, що зібрана спеціальними програмами пошукової системи. Власне наявність баз даних пояснюється високою швидкістю виведення результатів пошуку на сторінку пошукової системи;
- пошукові роботи (Robots), павуки (Spiders) або хробаки (Worms) - спеціальні програми, які автоматично періодично «відвідують» сайти, збирають відомості про вміст сторінок, тобто індексують їх і наповнюють бази даних пошукової системи.

Щоб скористатися пошуковою системою потрібно завантажити її веб-сторінку і сформулювати запит, за яким відбувається відбір документів з розподіленої бази даних, що зберігається на серверах в Інтернеті. Запит формується за допомогою ключових слів (одного або кількох). Результати пошуку видаються користувачу у вигляді переліку адрес сторінок (гіперпосилань) і короткої анотації до них.

Пошук - це процедура відбору потрібних документів, що зберігаються в Інтернеті. Пошук здійснюється або автоматично за допомогою ключових слів, або шляхом послідовного проходження користувачем по рубриках в каталогах.

Ключове слово - це будь-яка лексична одиниця природної мови, яка найбільшою мірою відображає зміст шуканого документа. При формуванні

запиту можуть використовуватися не лише окремі ключові слова, але і словосполучення, що складаються з кількох ключових слів.

Запит - це набір ключових слів, за допомогою яких пошукова система автоматично проводить пошук і відбір необхідних документів. Іншими словами, запит - це інструкція (команда) для пошукової системи на пошук потрібних документів.

Запити бувають двох типів: прості і складні (або розширені, *advanced*). Прості запити складаються з окремих ключових слів або словосполучень. Складні запити, окрім ключових слів, містять логічні чи інші оператори.

Індексація - це процедура автоматичного заповнення бази даних пошукової системи, в якій зберігаються короткий зміст та анотації до сторінок і доменні адреси, за якими розміщені сторінки.

Коли пошукова система отримує конкретний запит на пошук потрібної інформації, то її пошуковий механізм порівнює ключові слова, що введені користувачем, з ключовими словами, що отримані під час індексації і збережені в базі даних системи. При збіганні цих слів користувачу видається назва сторінки, короткий опис, або абзац, де зустрічаються пошукові слова і доменна адреса даної сторінки.

Сучасні пошукові системи постійно вдосконалюють свої алгоритми роботи, і прагнуть проводити індексацію всього документа, а не лише його назви чи перших абзаців тексту. Найбільш досконалі роботи при індексації сканують і головну сторінку і решту сторінок, заходячи за гіперпосиланнями вглиб сайту. Попередня індексація документів Інтернету дозволяє згодом за кілька секунд обслуговування запиту обробити (відсортувати) гігабайти різноманітної інформації. За допомогою механізму попередньої індексації можна отримати якісний результат пошуку у випадках, коли вдається точно сформулювати запит за допомогою невеликого числа ключових слів. В інших випадках користувач або отримує велике число посилань, які є лише "інформаційним шумом", або зовсім не знаходить необхідного документа.

Значно підвищують вибірковість пошуку фільтри, якими забезпечуються пошукові системи.

Фільтри дозволяють:

- обмежити список відібраних документів за допомогою логічних операторів (так званий складний пошук);
- обмежити простір пошуку типом протоколу, за допомогою якого було створено документ (пошук на сайтах або у телеконференціях);
- обмежити відібраний матеріал часовим відрізком, певними датами створення шуканого документа (наприклад, між 1 вересня 2005 р. і 20 вересня 2007 р.);
- відібрати документи, що складені однією мовою (українською, російською тощо);
- обмежити відбір документів територією розміщення серверів (наприклад, тільки з Європи, України чи навіть, лише з Львівської області);
- обмежити пошук певною частиною документа (заголовок, доменна адреса);

- відібрати документи, які містять фразу із заданим порядком розташування ключових слів.

При пошуку інформації пошукова система може робити дві помилки: не відбирати потрібні користувачеві документи і, навпаки, відбирати зайві документи, що містять інформаційний шум. Для уникнення цих помилок варто ознайомитися з поняттями "чутливість" і "вибірковість".

Чутливість - здатність пошукової системи відбирати документи, що відповідають запиту, не пропускаючи потрібних документів. Чутливість характеризується числом пропусків потрібних документів. Чим вищою є чутливість, тим менше вірогідність пропуску потрібного документа.

Вибірковість - здатність пошукової системи відбирати документи, що відповідають запиту і не виводити невідповідні документи. Якість вибіркості характеризується числом відібраних документів, що не відповідають запиту. Чим вищою є вибірковість, тим менше сторонніх документів потрапляє до користувача.

Вибірковість можна змінювати (регулювати) за допомогою логічних операторів (фільтрів).

Для уточнення запиту (фільтрації інформації) призначені логічні оператори OR, AND, NOT. Використання логічного оператора AND (І) приводить до відбору документів, які обов'язково містять всі перелічені в запиті ключові слова, що сполучені цим оператором.

Оператор NOT (НІ) дозволяє виключити документи, які містять ключове слово, вказане після цього оператора.

За допомогою оператора NEAR (поблизу, поряд) користувач може відбирати документи, в яких ключові слова, що сполучені цим оператором, будуть міститися поблизу один від одного, а не в різних кінцях документа.

Оператор FOLLOWED BY дозволяє відбирати документи, в яких ключові слова слідує один за одним в заданому порядку.

Оператор ADJ відбирає документи, в яких ключові слова є суміжними (слідують одне за іншим).

Сучасні пошукові системи стають інтелектуальними. Використовуючи принципи штучного інтелекту, вони формують список документів (посилань на них), відповідно до ступеня їх релевантності.

Релевантність – відповідність знайденого в процесі пошуку документа до зробленого запиту. При цьому пошукові системи аналізують розташування знайдених ключових слів в документі (заголовку або тексті), число повторень ключових слів, їх взаємне розташування в документі. Найбільш точно відібрані документи розташовуються на початку списку документів, що знайдені в процесі пошуку. Для ранжирування знайдених документів використовують наступні показники.

Розташування ключового слова на сторінці (*keyword prominence*) - показник, що визначає, як близько від початку документа знаходиться задане ключове слово. Як правило, чим ближче до початку сторінки розташовується ключове слово, тим точніше документ відповідає запиту.

Частота ключового слова (*keyword frequency*) - показник, що враховує абсолютну частоту використання ключових слів (скільки разів зустрічається дане ключове слово на сторінці). Найбільшу «вагу» при ранжируванні документів мають слова, що розташовані в заголовку сторінки (так званий титул, він міститься між тегами title). Саме цей заголовок відображається в рядку заголовка браузера при перегляді сторінки, і слова, що там вказані мають найбільшу цінність (вагу) для пошукової системи.

Деякі пошукові системи при ранжируванні враховують індекс цитування (посилальна популярність, авторитетність, тематичний індекс цитування, ІЦ, ТІЦ, *Page Rank*, *Link Popularity*) - кількість посилань в Інтернеті, що вказують на конкретний сайт. Чим їх є більше, тим вищим є індекс цитування.

Результати пошуку за допомогою різних пошукових систем будуть різнитися між собою. Це пояснюється тим, що в кожній пошуковій системі застосовуються різні алгоритми індексації, різна періодичність оновлення інформації в базі даних, охоплюється різний простір навколишніх серверів, і відповідно індексується різне число документів.

Спеціалізовані пошукові системи

Метапошукова система дозволяє вести автоматичний пошук за запитом з використанням відразу кількох пошукових машин.

Метапошукові системи:

<http://www.metacrawler.com>

<http://www.inctor.com>

Пошукові системи, що призначені для пошуку файлів:

<http://www.filesearch.ru>

<http://www.files.ru>

<http://www.freeware.ru>

Інші спеціалізовані пошукові системи:

www.midi.ru

Музичні файли.

www.graphscarch.com

Малюнки.

<http://bukinist.agava.ru>

Книги.

www.cooking.ru/search.html

Кулінарні рецепти.

В спеціалізованих системах пошук часто відбувається не за допомогою ключових слів, а за відомими іменами файлів.

Окрім пошукових систем, використовуються інші механізми пошуку і збереження інформації в глобальній мережі. Природним чином утворилися комбіновані (гібридні) пошукові системи, які називають порталами. Вони містять відомості про сторінки різних тематичних напрямів і дозволяють проводити пошук, як за допомогою ключових слів, так і за допомогою каталогів.

Каталоги – це є ієрархічні структури, що містять впорядкований перелік сайтів, які супроводжуються короткими коментарями (анотаціями). Ресурси каталогів можна сортувати за датою долучення, за популярністю, за алфавітом тощо.

Каталоги можуть бути глобальними, національними, загальними і спеціалізованими (галузевими, тематичними тощо).

Каталоги з'явилися самим природним чином: користувачі відбирали для себе посилання на улюблені місця в Інтернеті і складали впорядковані тематичні списки з деталізацією кожного розділу (рубрики). Саме так виник всесвітньо відомий каталог Yahoo!, що створено студентами Стендфордського університету Девідом Філо і Джері Янгом.

Рубрикація - розміщення документа у відповідний розділ (підрозділ, рубрику), яке, як правило, проводиться власноруч модераторами, розробниками чи власниками сайтів. Процедура рубрикації є достатньо суб'єктивною і здійснюється на підставі індивідуальних уявлень людей про дану наочну область. Автори при розміщенні своїх сторінок деколи керуються не точною відповідністю документа до назви рубрики, а іншими міркуваннями. Наприклад, з метою збільшення відвідуваності своїх сторінок їх поміщають в популярні рубрики. Сайт, що містить різноманітну інформацію інколи поміщають в кілька рубрик.

Модератор – співробітник, що відповідає за вміст каталогу (його окремого розділу, підрозділу тощо), за його достовірність та актуальність.

Окрім ієрархічних каталогів, зустрічаються і інші типи каталогів - неієрархічні, в яких немає супідрядності, наприклад, каталоги, що використовують алфавітно-наочну класифікацію, наприклад, енциклопедії, музичні сайти, телефонні довідники тощо.

Особливості та відмінності пошукових систем та каталогів.

Отже, пошукові системи і тематичні каталоги мають свої особливості і взаємно доповнюють один одного.

Пошукові системи «оглядають» велике число сайтів, що знаходяться на різних серверах. Тут, нажаль, автоматичний відбір шуканих документів супроводжується чималим числом помилок.

База даних каталогу зберігається на одному сервері і містить менший об'єм інформації, ніж в базі даних пошукової системи. Проте, результати пошуку за допомогою каталогів мають більшу релевантність в порівнянні з пошуком за допомогою пошукової системи.

Пошукові машини.

- індексація сайтів відбувається в автоматичному режимі за допомогою спеціальних програм – роботів;
- пошукові системи переглядають кожен сторінку сайту окремо. Може бути проіндексована як одна, так і всі сторінки сайту;
- зазвичай, сучасні пошукові системи самостійно знаходять та індексують нові сайти. Але, іноді за певних умов, пошукові системи цього не роблять і сайт потрібно самостійно зареєструвати у пошуковій системі. Термін долучення сайту до бази даних (за умови відповідності сайту до вимог, що пред'являються) може складати від 1 до 2 тижнів;
- сайт може бути знайдений як результат обробки різноманітних запитів, що можуть відноситися до абсолютно різних галузей або видів діяльності;
- база даних в пошуковій системі постійно оновлюється. Алгоритм роботи пошукової системи є циклічним;

- при роботі з пошуковими системами кожна окрема сторінка сайту може бути оптимізованою під визначені ключові слова окремо від решти сторінок;
- алгоритм кожної пошукової системи є унікальним і може бути змінений у будь-який час без повідомлення власників тих сайтів, які розміщені в базі даних пошукової системи. Високий рейтинг сайту в результатах пошуку однієї системи, за певними ключовими словами, не гарантує високий рейтинг в результатах інших пошукових систем.

Каталоги.

- каталоги містять опис сайту як єдиного цілого. В каталогах немає опису окремих сторінок сайту. Власник сайту самостійно складає короткий опис сайту, що відображає його суть, призначення тощо. Після реєстрації сайту в каталозі, його зазвичай переглядає модератор, що відповідає за вміст каталогу або його окремого розділу. Змінити опис сайту в каталозі може модератор або власник сайту;
- складання каталогу (класифікація і анотація ресурсів) здебільшого є ручною роботою, і зазвичай в базі каталогу міститься значно менше сайтів, ніж в базі пошукової системи;
- у каталозі, як правило, зазначено доменну адресу головної сторінки сайту;
- при пошуку в каталозі вирішальну роль має короткий опис сайту, ніж ключові слова і метатеги, що містяться в тексті сторінок;
- сайти в каталогах можуть бути знайдені за датою розміщення, за алфавітом, за оцінкою модератора, за індексом цитування, за популярністю (числу відвідувачів) тощо;
- сортування сайтів за алфавітом і датою розміщення, пошук за назвами сайтів є абсолютно об'єктивними критеріями і не залежать від модератора;
- сортування сайтів за категоріями є об'єктивним критерієм. Якщо власник сайту при додаванні сайту до каталогу вибрав відповідну категорію, а модератор з цим погодився;
- сортування сайтів за іншими критеріями не може мати 100% об'єктивний характер. В певних каталогах рейтинг може мати комерційний характер;
- сайт може займати перші позиції в каталозі, якщо назва сайту починається з перших літер алфавіту чи якщо сайт було додано в каталог нещодавно, а сортування застосовано за датою додавання;
- виявлення посилань до неіснуючих сайтів в каталозі може бути виконане автоматично, але часто потрібно додатково перевіряти існування ресурсу модератором.

4.2.2. Огляд Пошукових систем

В даний час існує багато пошукових систем, які цікаві, якісні, але в силу ряду причин не дуже популярні. Більш того, багато хто з них надають інформацію в такому цікавому і незвичайному вигляді, що робота з ними

перетворюється з рутинного дії в захоплюючу гру. Аналіз тенденцій на цьому ринку дозволяє зробити висновок, що основний напрям розвитку - візуалізація отриманих результатів пошуку і більш вузька спеціалізація пошуку (пошуковики зображень, пошуковики ігор, пошуковики для жінок і т.д.). Більшість пошукачів дають прекрасну можливість отримання необхідної інформації. Всі пошуковики можна поділити на кілька основних груп. В даному огляді зроблена спроба такої систематизації, наведені деякі приклади членів кожної групи. Для кожного пошуковика дано його короткий опис і підкреслені основні відмінні риси. Приклади наведені таким чином, щоб дати максимально вичерпну інформацію про кожну групу пошуковиків, їх можливості, сфери застосування, відмінні особливості. Маючи уявлення про основні характеристики, користувач може провести самостійне додаткове дослідження з даної теми і знайти для себе ще багато цікавих, красивих і незвичайних пошукових систем.

Найбільш популярні пошукові системи

В даний час існує 3 основних міжнародних пошукових системи - Google, Yahoo і MSN, що мають власні бази даних і пошукові технології. Більшість інших пошукових систем використовує в тому чи іншому вигляді технології трьох перерахованих. Наприклад, пошук AOL (search.aol.com) використовує базу Google, а AltaVista, AllTheWeb і Lycos-бази Yahoo. Портал Mail.ru довгий час використовував пошукову технологію Google, а з 2006 року - Yandex. У Росії основною пошуковою системою є Yandex, за ним йдуть Mail.ru, що використовує технологію Yandex, замикає трійку лідерів - Rambler. Однак найбільша кількість пошукових запитів обробляє Google, російська версія якого (Google.ru) розпочала свою роботу в 2004 році. Кожен користувач Інтернету орієнтується на ту пошукову систему, до якої він звик або яку йому порадили його колеги. Нижче наведені короткі характеристики основних пошукових систем.

Google (www.google.com)



Це найшвидша і найбільша пошукова система. На даний час в ній проіндексовано понад 1,3 мільярда сторінок (з них повністю - трохи більше 700 мільйонів, про інші відомий тільки адресу і текст посилання). Google може знаходити інформацію на 117 мовах. Система добре працює з російськомовним ресурсом, є можливість вибрати мову інтерфейсу. На відміну від більшості пошукових систем, Google оцінює популярність ресурсу по кількості посилань, що ведуть до нього з інших сторінок. Google - це пошукова система, яка використовує кількість посилань на веб-сайт, як основний параметр популярності сайту. Це є особливо корисним у пошуку хороших сайтів при простих пошукових запитах. Google має дуже велику базу даних проіндексованих сайтів і надає частину своїх результатів Yahoo і Netscape Search. Найбільшим придбанням Google з'явилася компанія YouTube Пошукова система Google має дуже якісним пошуковим "движком".

Правильність видачі результатів пошуку в Google.ru часто перевищує якість видачі результатів пошуку в російських пошукових систем, наприклад, у Яндекс. У своїй системі Google використовує механізм PageRank, що змінює "важливість" сайту при видачі результатів пошуку. PageRank залежить від кількості і якості посилань на ресурс (тобто майже те ж саме, що й індекс цитування у Яндекс). Але на відміну від Яндекса, вплив PageRank у Google не настільки значно. Всі сторінки Google кешує (заносить в свою базу) і дозволяє людині, що проводить пошук, дивитися документ, не відкриваючи його в першоджерелі, а беручи з кешу Google (що часто набагато швидше). Google - одна з небагатьох пошукових систем, яка повністю індексує всі сторінки, а не тільки найголовніші.

Пошукова система Google має також різноманітними сервісами та додатковими можливостями пошуку, які постійно розширюються і вдосконалюються. Наприклад, рядок пошуку в Google можна використовувати як калькулятор. У рядку пошуку вводиться, припустимо, $(20 + 35) * 319$ і система видає правильну відповідь. Гарне опис можливостей пошуковика наведено на сайті компанії.

Yandex (www.yandex.ru)



Краща з пошукових систем вітчизняного виробництва. Вона індексує в основному російськомовні ресурси, при цьому по можливостях не поступається зарубіжним системам. Також в ній існує багато різних можливостей для удосконалення системи пошуку інформації. Це призводить кхорошім результатам і потрібні посилання, як правило, виявляються вже в першій десятці результатів. Має "полегшену" версію (з мінімумом елементів дизайну) на <http://www.ya.ru>. Офіційно пошукова машина Yandex.Ru була анонсована 23 вересня 1997. Уже тоді пошуковик мав деякими перевагами - можливістю перевірки документів на унікальність, обліком морфології російської мови, можливістю пошуку з урахуванням відстані (наприклад, при пошуку точного словосполучення). Основною відмінною рисою Yandex був ретельно розроблений алгоритм оцінки відповідності відповіді запиту (релевантності), що враховує не тільки кількість слів запиту, знайдених в тексті, а й "контрастність" слова (його відносну частоту для даного документа), відстань між словами і положення слова в документі.

AltaVista (www.altavista.com)



Система AltaVista є однією з найбільш великих пошукових систем (за кількістю проіндексованих сторінок). Даний факт, а також можливість вести пошук по ускладненим критеріям відбору привели до великої популярності AltaVista. Ця пошукова система також пропонує додаткові послуги у вигляді пошуку по каталогах (взятим з Open Directory and LookSmart), а також службу під назвою "Ask AltaVista" ("запитай AltaVista"), результати якої беруться з Ask Jeeves. AltaVista почала надавати свої послуги в грудні 1995

року. В даний час AltaVista володіє пошуковою системою Raging Search. До недавнього часу AV була великим порталом, але з причин фінансового (і не тільки) характеру значно скоротила кількість сервісів.

Yahoo! (www.yahoo.com)



Один з перших (був заснований в 1994) і найбільш популярних пошукових серверів в Інтернет. На Yahoo! складений великий структурований каталог категорій (categories). Спочатку пошук здійснюється в них, потім у власному архіві, потім - з використанням системи Google. Yahoo має базу даних в більш ніж 1 млн. проіндексованих сайтів. Yahoo є найстаршою пошуковою системою, яка почала надавати свої послуги в 1994 році. Секрет успіху Yahoo полягає в людях. Yahoo має близько 150 редакторів, для того, щоб складати і редагувати вміст своїх каталогів. Дивно, але ця неймовірно популярна система, яка обслуговує мільйони запитів щодня, зародилася як проста колекція закладок, яку поповнювали всього 2 людини - Девід Філо та Джеррі Янг.

Lycos (www.lycos.com)



Останнім часом - одна з найпопулярніших систем. У той же час ніяких особливих можливостей вона не надає - "AND" "OR", пошук фраз, обов'язкова присутність / відсутність слова; в розширених можливостях - пошук в назві, URL, імені хоста та / або назві домена; 25 мов, включаючи російську, - словом, весь "загальноприйнятий" набір. Можна вказати тип змісту ресурсу - авто, книги, ftp, download, новини і т.д. Очевидно, популярність Lycos - наслідок масштабу цього великого проекту.

Рамблер (www.rambler.ru)



До недавнього часу найвідоміша російська пошукова система. Але останнім часом його популярність різко впала. Зараз в цій системі використовується поліпшений механізм пошуку, змінився дизайн, але за якістю Rambler все одно не зрівнявся з Яндексом або Google.ru. На сайті присутня рейтинг-каталог ресурсів Rambler Top 100, один з визнаних джерел статистичної інформації про інтернет-проектах.

MSN (www.msn.com)



Пошуковик розроблений і запущений компанією Microsoft в 1997 році. На відміну від інших пошукових систем, раніше у MSN ніколи не було власного павука або каталога. З 1997 року для видачі результатів пошуку використовувалися різні бази даних, такі як: Yahoo!, LookSmart, Altavista, DirectHit, Inktomi і RealNames. Тільки з початку 2005 року MSN запустив бета-версію власного пошукового алгоритму. Користувачі MSN Search, як і раніше, зможуть здійснювати пошук по всій Мережі в цілому, а також за

окремими тематичними категоріями, в тому числі і по енциклопедії Microsoft Encarta. Новий движок включає можливість локалізованого пошуку (Near Me) - система здатна автоматично визначати місцезнаходження користувача по IP-Адрес його комп'ютера. Є російськомовна версія.

Метапошуковики (searchbots)

Так як різні пошуковики використовують різні алгоритми пошуку і приділяють "особливе" увагу різних ділянок мережі, до яких вони найбільш адаптовані, логічно в багатьох випадках шукати відразу декількома пошуковими машинами. Цю можливість і пропонують searchbots, Автономні пошукові агенти. В основу роботи searchbots закладений наступний принцип: із запиту користувача генеруються запити, відформатовані в синтаксисі і логічних конструкціях кожного конкретного пошукового ресурсу.

Таким чином, з одного запиту метапошукова машина робить безліч запитів, які потім розсилаються широкому колу пошукових машин (та / або каталогів). Зібравши результати, мета-пошукова система видаляє дубльовані посилання і, відповідно до свого алгоритмом, об'єднує / ранжирує результати в загальному списку.

Безумовно, автономні пошукові агенти стали значним проривом у розвитку технологій пошуку. Але використання цих коштів приховує в собі певні проблеми. Наприклад, іноді пошук із застосуванням мета-засобів може виявитися дуже повільним, адже цим системам доводиться координувати за часом надходження результатів обробки пошукового запиту від декількох серверів. Ще одним недоліком мета-засобів є те, що вони не дозволяють повною мірою використовувати можливості мови запиту кожного з застосовуваних пошукових засобів. Більшість таких систем погано розуміє (чи не розуміє зовсім) російську мову.

Метапошуковики не ведуть власної бази інтернет-сайтів, а лише обробляють результат, представлений іншими пошуковими машинами. Обсяг цієї обробки може бути різним. Найпростіші метапошуковики дозволяють видалити повторювані посилання та створити єдиний список сайтів, впорядкований за релевантністю.

Метапошуковики, що з'являються останнім часом, дозволяють проводити кластеризацію (об'єднання) отриманих адрес сайтів за різними критеріями. Такі системи дозволяють виявляти в списку отриманих сайтів загальні ключові фрази і групувати сайти відповідно до них. Автономні пошукові агенти продовжують розвиватися - головним плюсом метапошукових систем є їх можливість шукати в базах безлічі каталогів і пошукових машин, що дозволяє швидко переробляти великі обсяги інформації й істотно економити трафік. Автономні пошукові агенти бувають двох класів. Одні, як і звичайні пошуковики, розташовуються на публічному онлайн-ресурсі, інші встановлюються на персональний комп'ютер.

Vivisimo (www.vivisimo.com). Основною перевагою цього пошуковика є те, що він, крім видачі посилань на відповідні сторінки в Інтернеті, сортує інформацію за категоріями. За запитом користувача система будує дерево

тим, в якому розміщені не тільки ключові слова, а й типи документів (статті, новини).

Якщо в кластерах-темах знову зустрічаються повторювані комбінації, то створюються підтеми (підклстера). Система пропонує не тільки зручний для сприйняття результат пошуку, але і не менш зручні засоби роботи з ним. Поруч з назвою кожного сайту є посилання, за допомогою яких можна відкрити посилання в новому вікні, в фреймі поточного вікна і здійснити попередній перегляд знайденого сайту. Для кожного знайденого посилання зазначено також засіб, за допомогою якого вона знайдена. Крім того, користувач може провести додатковий пошук по вже відібраній інформації.

Browsys (www.browsys.com). При використанні даної системи проблему пошуку можна вирішити простим перебором варіантів (Google, Bing, YouTube, News, Blogs, Wikipedia і т.д). Browsys пропонує 17 різних варіантів пошуку, до яких відносяться великі пошукові системи, а також кілька сервісів Google, таких як Google Новини, популярні соціальні мережі, обмін знаннями та Bookmarking сайти, і, нарешті, Ebay. Як і в звичайній системі, запит вводиться в спеціальне віконечко, а ось кнопки «шукати» тут немає. Замість цього користувачеві пропонується вибрати один з пошуковиків, назви яких написані на кнопках. Після натискання однієї з них, на екрані з'являться посилання, які знайшла обрана пошукова система. Якщо щось не влаштовує, можна натиснути іншу кнопку і так, поки потрібний результат не буде знайдений, або не закінчиться весь список пошукових систем. Для того, щоб спеціалізувати ваш запит, потрібно вибрати відповідне посилання у верхньому меню. Відповідно до цього змінюється і список задіяних пошукових систем. Є об'єднаний пошук по Google і Bing. Шукати можна не тільки за словами, а й по зображеннях, звуку і відео. Є пошук новин по WorldNews, Topix, the Guardian, the NY Times, Google News, і ін. Іншою цікавою особливістю даного пошуковика є можливість створення власного списку сайтів по даній тематика. Цей список створюється в сервісі «віртуальний пошук» і зберігається в пошуковику під обраним для нього ім'ям.

Dogpile (www.dogpile.com). Це пошуковик від InfoSpace, який об'єднує результати пошуку найбільш популярних пошукових систем Google, Yahoo! Search, MSN, Ask.com, About, MIVA, LookSmart і ін. Досить цікавий метапошуковик з веселим мультяшним інтерфейсом. Він почав працювати в 2006 році. Dogpile в першу чергу виводить ті посилання, які дублюються відразу в декількох пошукових машинах (поруч з кожним посиланням вказується, де вона була знайдена). Чим нижче по списку ви опускаєтесь, тим більше унікальні посилання бачите, а це означає, що по ним міститься або інформація, що не індексується іншими пошуковими системами, які матеріали, які були додані відносно недавно.

Є зручна функція для контролю вмісту. Він передбачає три варіанти фільтрації: відсутність фільтрації, відкидання найбільш нескромних сторінок і, нарешті, максимальна фільтрація, після якої, по ідеї, повинні залишитися тільки бездоганно благопристойні посилання. Ви можете налаштувати спосіб

сортування результатів пошуку зображень, новин і медіафайлів. Крім того, відповідно до своєї назви на сайті пошуковика організовуються реальні акції на допомогу тваринам, викладені різні картинки з собаками і т.д.

GahooYoougle (www.polycola.com). Це сервіс з назвою GahooYoougle. На ньому проводиться пошук в двох пошукових системах - Google і Yahoo!. Запит відправляється як Google, так і Yahoo!, відкривається сторінка результатів розділена вертикально навпіл: зліва розташовуються результати Yahoo!, а праворуч - Google. Це повнофункціональні сторінки даних пошуковиків з працюючими посиланнями. В результаті знаходиться в два рази більше посилань на сторінці. Перехресні посилання не фільтруються. Для користувачів, чий IP визначається як російські цей сервіс має особливе значення. Локальні версії в Росії мають як Google, так і Yahoo !, але останній, працюючи з запитом з нашої країни, видає, в першу чергу знахідки на російськомовних ресурсах. В результаті, на одній сторінці будуть результати пошуку, як в рунеті, так і в Інтернеті.

Rollyo (www.rollyo.com). Користувачеві пропонують створити свій власний пошуковик. Для цього потрібно зареєструватися, придумати назву вашої пошукової машини і вказати ті ресурси, за якими вона буде шукати, не більше 25 штук. При цьому треба мати на увазі, що приймаються тільки домени. Звичайно ж, ніякого власного пошуковика при цьому не створюється. Це хороший спосіб звуження кола пошуку. Створений користувачем мікропошуковик з'являється в одному з розділів, і будь-хто може ним скористатися: при цьому він буде бачити фотографію і контактну інформацію творця пошуковика (якщо, звичайно, він цю інформацію вкаже). Зручний варіант, коли часто доводиться користуватися певним набором сайтів, наприклад для моніторингу з якоїсь певної тематики.

Copernic Agent (copernic.com). Метапошукові машини можуть встановлюватися на комп'ютери користувачів. Вони зручні і можуть налаштовуватися на коло запитів і інтересів конкретної людини. Зараз можна знайти кілька десятків систем цього класу. Одна з найпопулярніших таких систем - Copernic Agent, програма компанії Copernic Technologies. Результати пошуку зберігаються в своїй основі програми, їх можна зберігати в файлах стандартних форматів разом з гіпертекстовими посиланнями. Розміри анотацій (описів) можна змінювати, програма сама оновлює бази пошукових ресурсів при кожному підключенні до Інтернету. Є безкоштовна і дві платні версії програми. Платні версії працюють не тільки з індексами пошукових машин. Вони вміють працювати з контентом багатьох великих видавництв і агентств. Посилають запити до CNN, MSNBC, New York Times, Forbes і т.д. Досить серйозний мінус, присутній і в ToolBar програми і в її безкоштовної версії - відсутність пошуку по базі Google.

Turbo10 (turbo10.com). Даний пошуковик дозволяє налаштовувати список використовуваних пошукових машин. Система за замовчуванням пропонує набір з 10 пошукових машин, званий turbo10 collections. Перейшовши за посиланням Edit my collections, можна сформувати нову або змінити існуючу колекцію пошукових систем. Процес редагування полягає в

відборі пошукових засобів зі стандартного списку turbo10 в свій список. Якщо стандартний список недостатній (а це так хоча б тому, що в ньому немає жодного російськомовного пошукового засобу), список можна доповнити. Список знайдених сайтів містить не тільки назва, посилання і пошукове засіб, за допомогою якого сайт вже знайдений, але і зменшений скріншот сторінки.

Метапошуковий засіб turbo10 допускає кластеризацію отриманих посилань, причому список посилань можна згрупувати одночасно за двома критеріями - традиційним способом, за ключовими словами і по використовуваних пошуковим системам. В обох випадках вказано кількість сайтів, що належать одній категорії.

У тому, що Google і Yandex - це незамінні інструменти для пошуку в мережі, сумніватися не доводиться - пошук в цих системах зручний, гнучкий і дуже точний. Але це не означає, що альтернативні пошукові системи не мають права на існування. У більшості випадків вони індексують меншу кількість сторінок, ніж стандартні пошукові системи. Але у таких пошукових систем є одна незаперечна перевага - вони пропонують щось нове, відмінне від прийнятих стандартів. Оскільки альтернативні пошукові сервіси використовують інший підхід до підбору ресурсів, відповідних запиту, результат пошуку буде інший, ніж у випадку з звичними пошуковими системами. Корисно використовувати метапошуковики або вузькоспеціалізовані пошукові системи. Знайдені результати фільтруються, відбираються найкращі і часто для кращого сприйняття візуалізуються за допомогою діаграми, карти сайтів, хмари тегів і т.д. Розробники альтернативних пошукових систем часом так далеко заходять у своїх пошуках нового універсального інтерфейсу, що, іноді важко визнати в web-сторінці пошукову машину. Головне, щоб візуальні ефектів не підмінили собою основну задачу, для вирішення якої і були створені пошукові системи (збір та систематизація максимальної кількості інформації та її подальша обробка).

Інтелектуальні пошукові системи

Як відзначають дослідники Інтернету, результати пошуку в великій мірі залежать від правильно сформованого запиту. Багато людей, намагаючись отримати будь-яку інформацію, формулюють свої питання в досить розмовної формі у вигляді питань типу, «яка завтра буде погода». Звичайний пошуковик видає посилання у відповідь на таке питання відповідно до принципу максимально пересічних слів в питанні і в змісті тексту на сайті. Тому, щоб отримати відповідь на своє питання, користувач повинен сам переглянути видану інформацію і знайти в ній шукане рішення задачі. У створенні інтелектуальних пошукових систем реалізується ідея про комп'ютер, який зможе спілкуватися з людиною на зрозумілій йому мові, володіючи при цьому набагато більшими знаннями, ніж звичайна людина. Передбачається, що така система буде відповідати на найрізноманітніші запитання користувачів, задані на звичайній розмовній мові. І відповідати не

посиланнями на сторінки або цитатами з них, а фактами. В останні кілька років такі пошукові системи з'явилися. Це Nakia, Powerset, True Knowledge і WolframAlpha.

True Knowledge ([www.True Knowledge.com](http://www.TrueKnowledge.com)). Нова пошукова машина True Knowledge вміє працювати зі звичайними запитам. Однак головна особливість цього пошуковика - планується, що він буде відповідати на запити користувачів не зовсім звичайно - в пошуковій видачі будуть розміщені не тільки посилання на сторінки або цитати, а факти. В даний момент за запрошеннями доступна бета-версія системи, яка комбінує ці дві видачі - задавши запит, можна отримати і список посилань, і, в окремій області вікна, список фактів.

Наприклад, запит виду - «who is George Bush» підносить у верхній частині списку фактів, які представляють собою своєрідні кнопки переходу на більш детальну інформацію.

Користувачам сервісу надається можливість поповнення бази даних фактами, подіями, особистостями; в наявності кілька рівнів, а для досвідчених користувачів доступна можливість додавання нових класів, об'єктів і зв'язків. У деяких випадках пошукова видача є досить релевантні посилання і факти, однак іноді система не справляється навіть з найпростішими запитам. А в цілому видача досить сильно залежить від заданого запиту.

Wolfram Alpha (www.wolframalpha.com). Wolfram | Alpha - база знань з елементами штучного інтелекту, в якій закладено величезну кількість математичних моделей та інформації.

Її творець Стивен Вольфрам після навчання в Ітоні і Оксфорді отримав ступінь доктора наук в галузі теоретичної фізики в Каліфорнійському технологічному інституті в 20 років. У 1981 році йому був вручений "грант геніїв", стипендія фонду Макартур. Його книга "Нова наука" (A New Kind of Science), в якій він заявляє, що в основі всіх наук можуть лежати прості алгоритми, а не складні структури і правила, була зустрінута захопленнями з боку одних і несхваленням інших.

В основі Wolfram Alpha лежить "движок" Mathematica. У нього вбудований потужний логічний механізм, який може робити висновки, маючи дані і математичну модель. Цей підхід поширений на області, слабо пов'язані з математикою, зокрема, географію, бізнес і кулінарні рецепти. Wolfram Alpha може, наприклад, не тільки відповідати на запитання на кшталт "Де знаходиться Єгипет?" (Прості пошуковики з таким навчилися справлятися), а й "Де завтра опівдні буде знаходитися МКС?". Для того щоб відповісти на останнє запитання, потрібно зрозуміти, коли настане завтра, а також мати математичну модель руху МКС або розрахувати її, виходячи з наявних про станції даних. Звичайні пошуковики цим не займаються. Wolfram Alpha для цього створено.

Пошуковиком Wolfram Alpha можна назвати з великою натяжкою. Він у вкрай рідкісних випадках звертається до зовнішніх джерел даних. Команда Стивена Вольфрама витратила кілька років на створення бази даних фактів, а

також великої кількості математичних моделей. Пошуковик займається пошуком в бібліотеці, заснованій на програмних продуктах Стефана Вольфрама - Mathematica і A New Kind of Science. Розробляються автором алгоритми дозволяють надати цим даним зрозумілу для користувача форму. Коли користувач задає питання, дані для відповіді будуть братися саме з цієї бази. Мова не йде про запити на природній мові, хоча сама мова запитів дуже схожий на англійський. У той же час система розуміє короткі запити, схожі з тими, до яких звикли користувачі звичайних пошуковиків. Відповідь виводиться у вигляді діаграм, таблиць і графіків.

Запуск першої публічної версії системи відбувся 16 травня 2009 року.

Спеціалізовані пошуковики

Пошук зображень

TagGalaxy (www.TagGalaxy.com). Цей пошуковий сервіс являє собою засіб для пошуку зображень на Flickr.com, з попереднім переглядом. Кількість картинок, що завантажуються на сервери цього сервісу настільки велике, що для того щоб знайти якесь конкретне зображення в цьому океані знімків і картин, необхідна окрема пошукова система. А незвичайним його робить інтерфейс пошуку, який повністю зроблений тривимірним. Процес пошуку за ключовим словом нагадує якусь комп'ютерну гру. TagGalaxy - ця незвичайна пошукова система має тривимірний інтерфейс у вигляді планетної системи, тим самим він виглядає справді красиво і захоплююче.

Picsearch (www.picsearch.com)

Сервіс працює дуже швидко, виводить результати пошуку в зручній формі і майже ніколи не показує битих посилань. Під час пошуку не забувайте користуватися додатковими настройками (кнопка Advanced Search), де можна задати розмір шуканого зображення, його кольоровість і наявність на картинках анімації.

Retrievr (labs.systemone.at/retrievr). Пошуковик шукає саме картинки (але не по всьому інтернету, а тільки по Flickr). Однак в пошуку потрібно ставити не слово, а, як не дивно, теж картинку. У маленькому віконці малюєте щось (графічний редактор найпростіший - 4 розміру кисті, так палітра, як в «Фотошопі»). Намалювали, відсунули курсор в сторону - і retrievr тут же видає безліч зображень, які хоч чимось схожі на ваше «творчість». Збіги, правда, трапляються нечасто: система орієнтується тільки на основні геометричні форми, а значить, намалювавши чорну морду кота, ви в результатах пошуку цілком можете отримати кілька портретів людей, якусь чорнильну пляму і, припустимо, камінь. Але весело. Можна приємно провести час і знайти чимало несподіваних зображень.

Picollator (picollator.ru). Пошук ведеться по фотографіях. На сайт завантажуються портрет. В результаті пошуку система видає фотографії всіх людей, хоч чимось схожих на запит.

Corbis (pro.corbis.com). Це один з найбільших фотобанків (з можливістю пошуку картинок в інших галереях), а головне в ньому - найбільший відсоток безкоштовних зображень.

Пошук музики та звуків

FindSounds (www.findsounds.com). Пошуковик створений для пошуку звуків (не плутати з музикою), різних звукових ефектів і т.д. Ресурс дозволяє шукати звукові файли різних форматів - wav, mp3, aiff, au. У базі даних ресурсу є найрізноманітніші звуки - крики тварин, скрегіт машин, дзвін, стукіт, сирени, дзижчання комах, гуркіт вибухів і стрілянини, сплеск води і т.д. У результатах пошуку ресурс показує не тільки посилання на знайдені файли, але і їх основні характеристики, а також показує графік амплітуди звуку, по якій можна судити про характер звучання даного прикладу. Шукає по аудіо форматів - AIFF, AU, MP3 і WAVE, за кількістю каналів - mono і stereo, і ще за кількома параметрами, включаючи максимальний розмір файлу. Шукає швидко і багато, що в зв'язці з зручним і інтуїтивно-зрозумілим інтерфейсом дозволяє використовувати його в потребах професійних і не тільки. Він схожий на Google, але сфокусований саме на звуках. Даний пошуковик надає потужні можливості, але в той же час є простим, зрозумілим і зручним в користуванні. Є дуже зручна функція для батьків: в цьому пошуковику фільтрується музика, яка містить в собі ненормативну лексику. Отримавши запит в командному рядку - наприклад, Rocket, - пошуковик видає кілька сотень посилань на файли, в яких хтось його вимовляє, або лунає звук ракети, або звук старту космічного човника, а то і зовсім звук пострілу з ракетниці в Quake або DOOM. База звукових ефектів FindSounds може знайти застосування в самих різних областях - від розробки комп'ютерних ігор та інших додатків, до створення презентацій і всіляких кліпів.

Musicplasma (www.musicplasma.com). Вам подобається певна музика і ви хочете знайти що-небудь схоже? Для цього існує пошуковик Musicplasma.com. На відміну від інших пошуковиків він шукає не музику як таку, а виконавців. Якщо задати сервера чиєсь ім'я або назва групи, то засобами флеш-графіки в браузері відобразиться хитромудре сузір'я імен: в центрі - шуканий виконавець, а десятки інших розмістяться ближче або далі від нього, у міру схожості їх музики. Картинка дуже наочна і дозволяє відразу побачити, хто на кого і як сильно схожий. Кулька з пошуковим запитом знаходиться в самому центрі карти. Найбільш схожі на нього виконавці розташовуються поруч. Якщо вам подобається той, що в центрі, то майже напевно вам сподобаються і сусідні, хоча ви могли ніколи раніше не чути навіть їх назв. Розмір кожної кульки відповідає популярності групи. Імена з'єднуються лініями, і, простежуючи ланцюжка, що відходять від знаменитих команд, робиш чимало відкриттів, натикаючись на невідомі гурти, які грають, на думку Musicplasma, схожу музику. Наприклад, з її допомогою можна виявити неймовірні зв'язку між класичними композиторами - такими як Бетховен, Брамс, Ліст, Вагнер - і сучасними

музикантами. Деякі "карти зірок" такі великі, що не поміщаються на екрані, але по ним легко переміщатися, просто клацаючи мишею. Клацання по будь-якому імені запускає новий пошук асоціативних зв'язків, і через пару секунд можна спостерігати картину, створену вже навколо іншого центру. Musicplasma - це, в першу чергу, чудовий інструмент для досліджень. Якого б найвідомішого виконавця НЕ ввести, обов'язково поряд з ним виявиться пара-трійка «родинних», але абсолютно невідомих музикантів. Для багатьох музикантів в спеціальному віконці ліворуч виводиться їх повна дискографія. Кожен з альбомів можна купити (посилання ведуть в Amazon).

Allmusic (www.allmusic.com). Велика онлайнова музична база даних, що належить організації All Media Guide, поряд з All Movie Guide і All Game Guide. Заснована в 1991 році Майклом Ерлевайном і математиком Володимиром Богдановим. Інтернет-портал працює з 1995 року. На ньому надана інформація про жанрах музики, музикантів і групи, а також інформація про вихід альбомів і професійні рецензії. У цій основі містяться 1,695,140 альбомів, 317,552 класичних композицій, 1,207,634 виконавців, 282,732 композиторів, 98,384 біографій, 930 музичних стилів, 77,241 фотографій виконавців і т.д.

Новинні пошуковики

Wikio (wikio.com). Wikio - новини та інформація. Пошукова система є як би величезним інформаційним порталом, суть полягає в пошуку новинних сайтів і блогів зі свіжою пресою.

Addictomatic (www.addictomatic.com). Пошуковик пропонує відмінну заміну ранковій газеті. Вводите в поле пошуку будь-яке слово, пов'язане з подіями зараз подіями, відомими людьми або подіями, і отримуєте результат - огляд найпопулярніших сервісів зі згадуванням цікавить вас інформації на них. Серед сайтів, за якими проводиться пошук і збір інформації: новини live.com, повідомлення з Twitter, відео з YouTube, Google Blogs, картинки з Flickr і багато іншого.

Наукові пошукові системи

В даному розділі зібрані посилання на спеціалізовані наукові пошукові системи, електронні архіви, засоби пошуку статей та посилань.

Scirus (Scirus.com). Універсальна наукова пошукова система. Багаторазово визнавався найкращим спеціалізованої пошуковою системою, що включає в себе реферати понад 28 млн. Статей з більш 14,000 журналів (52% з них європейські) 4,000 видавництв, патентів, матеріалів конференцій і т.д. з усього світу, включаючи 260 російських журналів. Глибина охоплення - 40 років (з 1966 року), включені всі спеціалізовані бази Elsevier, основні бази інших видавництв (напр., Inspec, Medline і т.д.), дані з платформ всіх наукових видавництв (напр., Springer), дана інформація по індексації в інших базах по кожного запису, а також інформація по цитованості (зокрема для російських авторів).

З недавніх пір з'явилася можливість встановити браузерну панель Scirus. Це дозволяє здійснювати пошук в Scirus'е прямо з браузера, швидко переміщатися між результатами пошуку по різних запитах, підсвічувати на знайдених сторінках слова з пошукових запитів, додавати цікаві наукові сайти в індекс цієї пошукової системи.

ScholarGoogle (scholar.google.com). Пошукова система по науковій літературі. Включає статті великих наукових видавництв, архіви препринтів, публікації на сайтах університетів, наукових товариств та інших наукових організацій. Шукає статті в тому числі і російською мовою. Що не маловажно, розраховує індекс цитування публікацій і дозволяє знаходити статті, які містять посилання на ті, що вже знайдені.

Scholar (www.scholar.ru). Пошуковик був створений для спрощення пошуку документів наукової тематики російською мовою. В першу чергу, проект розрахований на електронні публікації, виконані в Росії. Проект індексує також дисертації, монографії та інші наукові матеріали. індексуються і перекладені з російської матеріали. Основна мета проекту - збір інформації щодо вільно скачуваних публікацій. Проект не розрахований на зберігання повних текстів публікацій в тому чи іншому вигляді. Замість цього, використовується база посилань на тексти документів з інформацією про самих публікаціях (анотація, автори і т.д.). Проект Scholar.ru знадобиться насамперед студентам і тим, хто цікавиться науковими публікаціями. Як і у випадку з іншими «пошуковими системами», сам сайт не містить будь-яких публікацій, а тільки надає посилання на них. Можна додавати посилання на свої статті в базу даних. Є цікавий форум по різних наукових тематиках. Також на сайті наводиться програма наукових конференцій.

Science Research Portal (www.rad.pfu.edu.ru). Наукова російськомовна пошукова система, підтримувана компанією Deep Web Technology (DWT), що здійснює повнотекстовий пошук в журналах багатьох великих наукових видавництв, таких як Elsevier, Highwire, IEEE, Nature, Taylor & Francis та ін. Шукає статті та документи у відкритих наунчних базах даних: Directory of Open Access Journals, Library of Congress Online Catalog, Science.gov і Scientific News.

MedPoisk (www.medpoisk.ru). Пошук інформації відбувається виключно на медичних сайтах. Дана пошукова система використовує движок пошуку від Google. Medpoisk.ru - це універсальний пошуковик, який призначений для пошуку виключно на медичних сайтах. Цей сайт - відмінний інструмент для кожного медика і всіх, хто бажає отримати відповідь на будь-яке питання з області медицини. Як лікувати ту чи іншу хворобу, які протипоказання у тих або інших ліків, до якого лікаря звернутися - все це і багато іншого можна дізнатися, задавши запит пошуковій системі. Пошуковик включає в себе біржу праці і може використовуватися для пошуку роботи серед медичних працівників. Ресурс також містить каталог медичних установ, розсортовані по регіонах. Серед цих установ адреси клінік, медичних центрів різної спрямованості, пологові будинки, діагностичні центри, косметологічні салони та ін.

PubMed (pubmed.com). Це пошукова система розроблена в Національному Центрі Біотехнологічної Інформації (National Center for Biotechnology Information - NCBI), який є підрозділом Національної Медичної бібліотеки США (National Library of Medicine - NLM), яка є частиною Національного Інституту Здоров'я США (National Institutes of Health - NIH). MEDLINE - база даних медичної інформації, що включає бібліографічні описи з більш ніж 4800 медичних періодичних видань з усього світу, починаючи з 1949р. В даний час MEDLINE доступна безкоштовно для пошуку через Інтернет як для фахівців, так і для широкого загалу. Плюси цієї системи очевидні: величезна база даних і велику кількість функцій для управління пошуком. Зараз це універсальний пошуковик літератури з медичної тематики. Справа не тільки в тому, що бази даних охоплюють велику кількість публікацій, але і в зручності пошуку і обробки отриманих результатів.

HighWire Press (highwire.stanford.edu). Це електронне підрозділ бібліотека Стенфордського Університету. Ресурс містить повнотекстові статті, частина з яких доступна безкоштовно, частину можна купити. Це найбільше у світі сховище безкоштовних повнотекстових статей з галузі life science. Станом на 13 січня 2009р. у них є 1831270 повнотекстових статей у вільному доступі із загальної кількості в 5095262 статті. Пошук здійснюється як по своїй базі даних, так і по PubMed.

Інтернет бібліотеки

Це, звичайно, не зовсім пошукові системи. Але при їх правильному використанні допоможуть зекономити багато часу і привести до бажаного результату набагато швидше, ніж за допомогою власне пошукової машини. Тим більше, що серед посилань будь-якого пошукача якась із таких бібліотек обов'язково буде вказана. Електронних бібліотек існує багато як загального, так і вузькоспеціалізованого призначення. На деяких сайтах є додаткові сервери, на зразок «розгадувача кросвордів» або on-line переклад заходів одиниць в різні системи. У деяких випадках разом з самими енциклопедіями дається досить повний каталог on-line словників. Крім того, існують спеціалізовані пошукові машини по електронним бібліотекам. Їх дуже багато і вони всі схожі. Тому, на них зупинятися не має сенсу. В даному огляді наведені приклади найбільш цікавих і повних з бібліотек.

Wikipedia (www.wikipedia.org). Вікіпедія (англ. Wikipedia) - вільна загальнодоступна багатомовна універсальна енциклопедія, підтримувана некомерційною організацією «Фонд Вікімедіа». У ній є інформація більш ніж на 50 мовах світу. Природно, є і російськомовний розділ Wikipedia. Запущена в січні 2001 року Вікіпедія зараз є найбільшим і найбільш популярним довідником в Інтернеті. За обсягом відомостей і тематичним охопленням вважається найповнішою енциклопедією з коли-небудь створювалися за всю історію людства. Одним з основних достоїнств Вікіпедії як універсальної енциклопедії є можливість представити інформацію рідною мовою, зберігаючи її цінність в аспекті культурної приналежності. Крім виконання

функції енциклопедичного довідника, Вікіпедія стала головним об'єктом уваги ЗМІ як мережевий джерело останніх новин з тієї причини, що вона постійно оновлюється. Власне, про Вікіпедію знають практично всі відвідувачі Інтернету.

Світ енциклопедій (www.encyclopedia.ru). Величезна добірка енциклопедій як по загальних знань, так і по конкретних наук. До більшості сайтів зі змістом статей доступ безкоштовний. Існує каталог друкованих енциклопедій. Є внутрішня пошукова система.

Didib (www.diclib.com). Словники і енциклопедії on-line на трьох мовах. На головній сторінці наведено систематичний каталог з посиланнями на більшість електронних версій найбільш відомих енциклопедій, таких як Велика радянська енциклопедія, Енциклопедія Брокгауза і Ефрона, словник Мюллера (англо-російський) і т.д. Існує гарна пошукова система як автоматична, так і ручна. Для більшості словників є функція їх скачування в запакованому вигляді.

4.3. Використання веб-браузерів для пошуку ресурсів.

У мережі Internet знаходяться мільярди Web-сторінок, терабайти інформації в різних файлах, але вся проблема в тому, що для одержання потрібної інформації необхідно вказати її точну адресу – URL. Саме цю адресу нам і потрібно знайти в першу чергу. Як Ви пам'ятаєте, адреса URL складається з доменного імені (або IP-адреси), шляху до файлу й імені файлу.

Якщо ми знайдемо такий URL, то без проблем введемо його в адресний рядок програми Internet Explorer і побачимо потрібну Web-сторінку, або скачаємо файл.

У мінімальному варіанті нам необхідно знати хоча б доменне ім'я ресурсу, а потім, переходячи по гіперпосиланнях, ми зможемо добратися до потрібної нам інформації.

Як знайти адресу (URL)?

Способів декілька. Найпростіший спосіб – це побачити доменні імена в різних рекламних матеріалах. В даний час будь-яка поважаючи себе фірма, у тому числі будь-яке видавництво книг, журналів, редакції телеканалів мають свою адресу в Internet і звичайно вказують її на упаковках, у книгах, журналах, в ефірі телепрограм. Як правило, вказується коротка адреса, що складається лише з доменного імені, щось на зразок: <http://www.1tv.ru> або ще коротше: 1tv.ru . Введіть таку адресу в адресний рядок і "ходіть" по сайту.

Для виклику пошукової системи необхідно в адресному рядку програми Internet Explorer вказати її адресу. Пошукових систем у мережі Internet зараз досить багато. У таблиці нижче перераховані лише найбільш популярні пошукові системи. Звичайно, у російськомовних системах міститься більше посилань на російськомовні ресурси, а в англomовних – на англomовні, тому пошук російськомовних ресурсів краще здійснювати в російськомовних пошукових системах.

Як працює пошукова система? Перший етап роботи пошукової системи – це індексування інформації, що знаходиться в Internet. *Мережевий робот* пошукової системи переглядає величезну кількість сторінок і заносить адреси і короткий зміст цих сторінок до себе в базу даних, точніше в пошуковий індекс. Власники сайтів будуть дуже раді, якщо їх сайт перегляне мережевий робот. Для того, щоб робот обов'язково переглянув сайт і вніс його в пошуковий індекс, власники самі реєструють свої сайти в пошуковій системі (як правило, у декількох пошукових системах). Процес занесення адреси й описи сторінки в пошукову систему називається індексація.

Таким чином, коли Ви запитаете пошукову систему про те, де знаходиться розклад потягів далекого прямування на літо, вона не буде шукати цю інформацію у величезній мережі Internet, а всього лиш шукатиме у своєму заздалегідь підготовленому пошуковому індексі, причому зробить це дуже швидко.

Як Ви вже зрозуміли, другий етап роботи пошукової системи – це видача накопиченої і відібраної, відсортованої і класифікованої інформації згідно запитів користувачів. У наступному пункті буде докладно розглянута робота з найбільш популярною пошуковою системою Яндекс.

Примітка: пошукові системи надають посилання на ресурси безкоштовно, а заробляють вони на рекламі, на платному занесенні інформації в каталоги, на сортуванні результатів пошуку на користь рекламодавця і так далі.

4.4. Принципи організації та використання пошукових машин. Формулювання пошукових запитів. Стратегії пошуку інформації.

4.4.1. Пошук по ключових словах

Найбільш простим і результативним пошуком є пошук по ключових словах. Давайте підключимося до Internet, відкриємо програму Internet Explorer і введемо в адресний рядок адресу пошукової системи: <http://www.yandex.ru>. Ми докладно розберемося з правилами пошуку по ключових словах у пошуковій системі Яндекс (рис.7). В інших системах ці правила можуть відрізнятися, але основні положення будуть ідентичні.

Ми вибрали Яндекс, оскільки Яндекс – це досить велика і популярна пошукова система. Протягом дня Яндекс відвідують, у середньому, не менш 75% користувачів російськомовної зони Internet. Щоденна аудиторія Яндекса складає більш 1 мільйона чоловік, а щомісячна – близько 8 мільйонів. Обсяг проіндексованої інформації, включеної в пошуковий індекс складає 5610 Гб – це 180 мільйонів унікальних документів, розташованих більш ніж на мільйоні серверів, і, звичайно, ці цифри постійно ростуть.

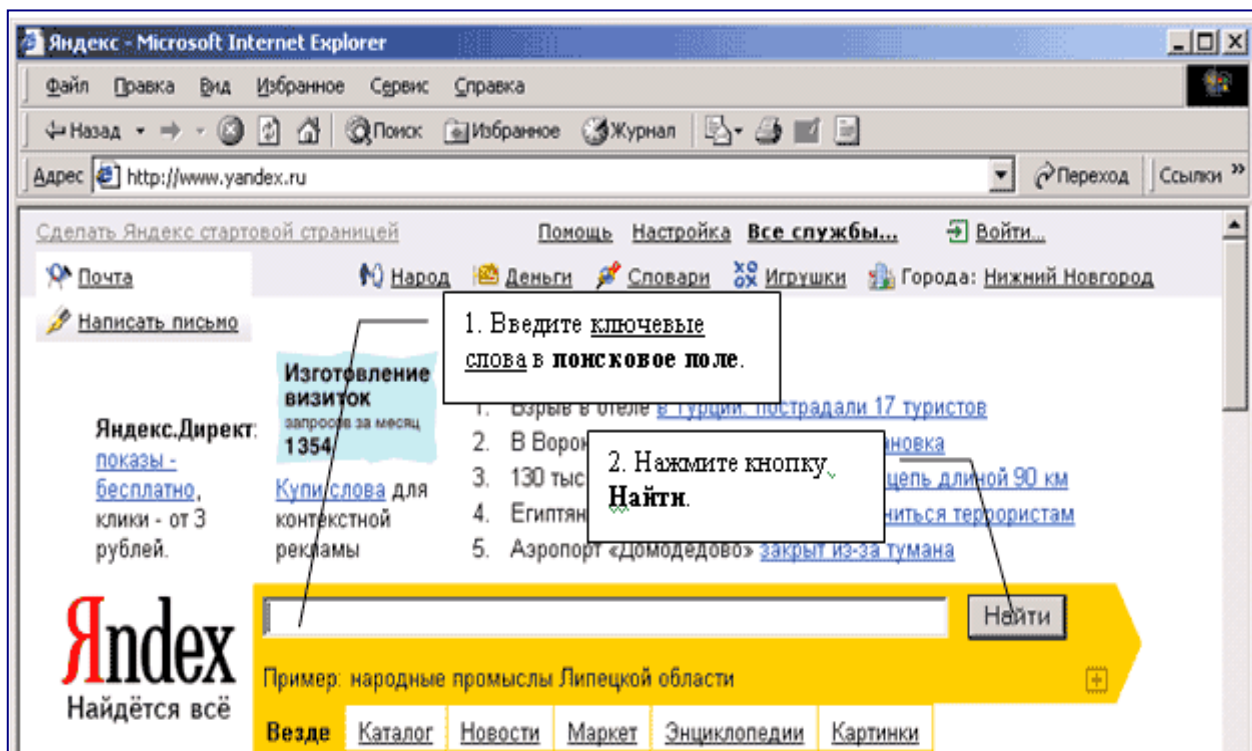


Рис.7. Головна сторінка пошукової системи Яндекс.

Основний алгоритм пошуку наступний:

- введіть ключові слова в поле Запиту (наприклад, *розклад потягів*),
- натисніть кнопку Знайти.

Після натискання кнопки Знайти, Яндекс виведе список посилань на документи, що найбільше точно відповідають запиту, і Ви побачите наступне:

Давайте розберемося, яка інформація міститься на даній сторінці. Тут є деякі спеціальні позначення і посилання, призначені для полегшення перегляду і сортування знайдених сторінок.

- Пошукова форма. Головний її елемент – поле Запиту.
- Статистика пошуку: кількість знайдених документів і частота заданих у запиті слів.
- Рубрики каталогу. При натисненні на посилання відбувається перехід у відповідний розділ каталогу (виводиться лише при точній відповідності запиту). Пошук по каталозі буде розглянутий нижче.
- Результати пошуку – список знайдених посилань. Для кожного документа видається наступна інформація: заголовок - посилання на ресурс,

коротка анотації ресурсу (сторінки), знайдені слова, схожі документи, і, якщо знайдений сайт описаний у каталозі, перехід у відповідну рубрику каталогу. Крім цього, інформація про розмір, дату документа, URL документа.

- Перехід на наступні сторінки результату (потрібно клацнути по номерах сторінок). Сортування по спаданні дат, спочатку список сортується по релевантності, тобто по ступені відповідності результату запитові).

- «Перетворювач запитів». Тут даються ради і пропозиції по уточненню вашого запиту. Можна звузити пошук по рубриці каталогу або по регіоні.

- Популярні знахідки наших користувачів - посилання на документи, обрані користувачами по цьому запиту (виводиться лише при точній відповідності).

- Можливість пошуку в інших російськомовних пошукових машинах.

- Результат пошуку в стрічках новин інформаційних агентств (виводиться лише при точній відповідності запитові). При високій відповідності й актуальності ця секція може виявитися над результатами пошуку.

- Посилання на інші служби Яндекс.

- Результат пошуку в енциклопедичних статтях (виводиться лише при точній відповідності запитові).

- Результат пошуку в базі товарних пропозицій магазинів (виводиться лише при точній відповідності запитові).

- Мітками R1-R5 на малюнку позначені різні рекламні блоки.

Оскільки в системі Яндекс дуже багато різної інформації, то результат пошуку по фразі «розклад потягів» дуже великий, для полегшення перегляду потрібної інформації її слід уточнити, для цього давайте розглянемо, що для цього передбачено в пошуковій формі (рис.8).

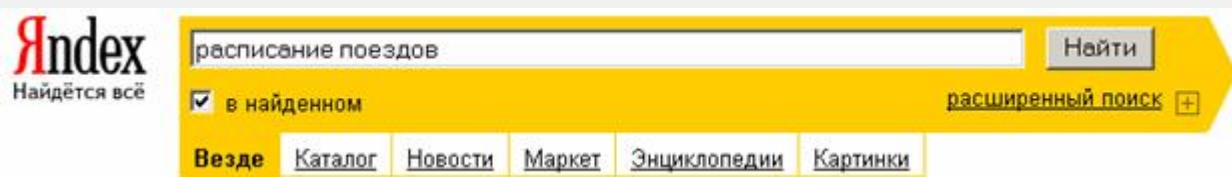


Рис.8. Пошукова форма системи Яндекс.

Два основних елементи в пошуковій формі ми вже знаємо, це пошукове поле і кнопка Знайти. Прапорець «У знайденому» дозволяє шукати в результатах попереднього запиту. Наприклад, по запиті розклад потягів система видала 300 тис. сторінок, ми вводимо наступний запит далекого прямування і встановлюємо прапорець «У знайденому» – тепер пошук буде вестися серед цих 300 тис. сторінок.

Обмежити область пошуку можна, клацнувши по посиланнях Каталог, Новини, Маркет, Енциклопедії, Картинки. Щиглик по потрібному посиланню

заміняє натискання кнопки Знайти. Зручно звужувати область пошуку для перебування картинок, наприклад, введіть у пошукове поле *Ющенко* і клацніть по посиланню Картинки. Для пошуку картинки можна також вказати бажаний розмір картинки від значення «Дріб'язок» до «Величезні».

Обмежуючи область пошуку, ми тим самим відкидаємо частину свідомо непотрібної нам інформації. Але часто буває необхідність вести пошук у всій базі, але для відкидання непотрібної інформації доводиться вводити різні додаткові відомості. Для цього служить Розширений пошук (рис.9). Він вмикається щигликом по посиланню «розширений пошук», або по значку «+».

Искомые слова
расположены относительно друг друга:

- ☐ подряд
- ☐ в одном предложении
- ☒ не очень далеко [?]
- ☐ на одной странице

расположение на странице:

- ☒ где угодно
- ☐ в заголовке
- ☐ в тексте ссылки

на сайт: (URL сайта)

употреблены в тексте:

- ☒ в любой форме
- ☐ точно так, как в запросе

Страницы
Найденные страницы должны обладать следующими свойствами:

язык:
русский
английский
французский
немецкий

дата: [?]
последние 2 недели
последний месяц

формат:
HTML
PDF
DOC
RTF

содержат ссылку: (URL ссылки)

находятся на сайте: (URL сайта)

похожи на страницу: (URL страницы-образца)

В результатах поиска показывать ссылок на странице

Итого: искать «», все слова расположены не очень далеко, в любой части страницы.

Рис. 9. Сторінка введення умов розширеного пошуку.

Ця сторінка дозволяє Вам більш тонко вказати умови пошуку. Важливим є те, що знайдені в результаті документи будуть відповідати відразу всім умовам, поставленим Вами. Давайте розберемося з особливостями введення умов розширеного пошуку.

- Поле запиту – введіть слова, що обов'язково повинні бути в знайдених документах, перед словами, яких не повинно бути в документах

поставте знак мінус (пробіл ставте до знака, але не після), наприклад, – електричка.

- Розташування слів відносно один одного – підряд, в одному словосполученні, не дуже далеко (у межах декількох рядків), на одній сторінці.

- Розташування слів на сторінці – де завгодно (у будь-якому місці сторінки), у заголовку (заголовок сторінки відображається в заголовку вікна програми Internet Explorer, тобто в синій смузі), у тексті посилання (слово є гіперпосиланням), також можна вказати адресу, куди веде дане посилання – таким чином, зручно шукати ті сторінки, що посилаються на дані (адресу яку Ви вказали).

- Форма вживання слів. Система Яндекс вміє ставити у відмінки введене слово і за замовчуванням знаходить документи, у яких зустрічаються усі форми слова. Якщо Ви виберете «точно так, як у запиті», то Яндекс буде шукати лише введену Вами форму слова. Це зручно при пошуку цитат.

- Мова сторінки – вкажіть мову, на якій повинна бути виведена сторінка. Мова міститься в невидимому користувачеві заголовку сторінки. В базу Яндекса вносяться документи російськомовного Інтернету, що знаходяться в доменах *su, ru, am, az, by, ge, kg, kz, md, tj, ua, uz*, а також закордонні сайти, що представляють інтерес для російськомовного пошуку.

- Дата сторінки – вкажіть кожен із запропонованих варіантів, або введіть період (дати початку і кінця), у межах якого повинна знаходитися дата документа.

- Формат документа. Усі Web-сторінки мають формат HTML, тобто виконані з використанням мови розмітки гіпертексту, однак, у мережі Internet присутні документи й інших форматів, що також можна переглядати за допомогою програми Internet Explorer: PDF – файл публікації, DOC – документ, створений у програмі Microsoft Word, RTF – відформатований текст (універсальний формат).

- Сторінки містять посилання – вкажіть URL цього посилання.
- Сторінки знаходяться на сайті – вкажіть доменне ім'я сайта.
- Сторінки схожі на сторінку – вкажіть URL сторінки-зразка.

У нижній жовтій панелі «Разом:» буде цілком сформульований Ваш запит. Оскільки ми шукаємо «свіжий» розклад руху потягів, то можна скористатися умовою Дата сторінки.

4.4.2. Поради при пошуку в системі Яндекс

1. Перевіряйте орфографію

Якщо пошук не знайшов жодного документа, то Ви, можливо, допустили орфографічну помилку в написанні слова. Перевірте правильність написання. Якщо ви використовували для пошуку кілька слів, то подивіться на кількість кожного зі слів у знайдених документах (перед їх списком після фрази «Результат пошуку»). Якесь зі слів не зустрічається жодного разу? Швидше за все, його ви і написали невірно.

2. Використовуйте синоніми

Якщо список знайдених сторінок занадто малий або не містить корисних сторінок, спробуйте змінити слово. Наприклад, замість «реферати» як найбільше підійдуть «курсові роботи» або «твору». Спробуйте задати для пошуку три-чотири слова-синоніми відразу. Для цього перелічіть їх через вертикальну риску (|). Тоді будуть знайдені сторінки, де зустрічається хоча б одне з них. Наприклад, замість «фотографії» спробуйте «фотографії, фотознімки».

3. Шукайте більше, ніж по одному слову

Слово «психологія» або «продукти» дадуть при пошуку поодинокі велику кількість безглузвих посилань. Додайте одне або два ключових слів, пов'язаних із шуканою темою. Наприклад, «психологія Юнга» або «продаж і покупка продовольчих товарів». Рекомендуємо також звужувати область вашого питання. Якщо ви цікавитесь автомобілями ГАЗ, то запити «автомобіль Волга» або «автомобіль ГАЗ» видадуть більш потрібні документи, а ніж «легкові автомобілі».

4. Не пишіть великими буквами

Починаючи слово з великої літери, ви не знайдете слів, написаних з маленької літери, якщо це слово не перше в пропозиції. Тому не набирайте звичайні слова з Великої Літери, навіть якщо з них починається ваше питання Яндекс. Великі літери в запиті рекомендується використовувати лише в власних іменах. Наприклад, «група Чорний кофе», «телепередача Здоров'я».

5. Знайти схожі документи

Якщо один зі знайдених документів ближче до шуканої теми, ніж інший, натисніть на посилання «схожі документи». Посилання розташоване під короткими описами знайдених документів. Яндекс проаналізує сторінку і знайде документи, схожі на той, що ви вказали.

6. Використовуйте знаки "+" і "-"

Щоб виключити документи, де зустрічається визначене слово, поставте перед ним знак мінус. І навпаки, щоб визначене слово обов'язково було присутнє в документі, поставте перед ним плюс. Зверніть увагу, що між словом і знаком плюс-мінус не повинно бути пробілу. Наприклад, якщо вам потрібен опис Парижа, а не пропозиції численних турагентств, має сенс задати такий запит «путівник по Парижу -агентство -тур». Плюс варто використовувати в тому випадку, коли потрібно знайти так звані стоп-слова (найбільш частотні слова російської мови, в основному це займенники, прийменники, частки). Щоб знайти цитату з Гамлета, треба задати запит «+бути або +не бути».

7. Спробуйте використовувати мову запитів

За допомогою спеціальних знаків ви зможете зробити запит більш точним. Наприклад, вкажіть, яких слів не повинно бути в документі, або що два слова повинні йти підряд, а не просто обидва зустрічатися в документі.

8. Шукати без морфології

Ви можете змусити Яндекс не враховувати форми слів із запиту при пошуку. Наприклад, запит Іваново знайде лише сторінки зі згадуванням цього прізвища, а не міста "Іваново".

9. Пошук картинок і фотографій

Яндекс вміє шукати не лише в тексті документа, але і відшуковувати картинки за назвою файлу або підпису, назві посилання на картинку.



Рис. 4.10. Пошук картинок.

При пошуку можна задати бажаний розмір картинки. Картинки групуються по площі:

- «Дрібні» до 1000 квадратних пікселів
- «Маленькі» від 1000 до 10 000 квадратних пікселів
- «Середні» від 10 000 до 100 000 квадратних пікселів
- «Великі» від 100 000 до 1 000 000 квадратних пікселів
- «Величезні» понад 1 000 000 квадратних пікселів

Для пошуку введіть пошуковий вираз, вкажіть бажаний розмір картинки і клацніть по посиланню Картинки.

При пошуку по можливості виключаються рекламні банери і копії картинок (однакові зображення). У результатах пошуку видається не більш однієї картини від одного сайту, інші доступні по посиланню «ще із сайту». Також доступна розширена інформація про картинку.

Синтаксис мови запитів.

У системі Яндекс як і в будь-якій іншій пошуковій системі існує спеціальна мова запитів, використовувати яку більш складно, ніж форму розширеного пошуку, але при її використанні можна одержати більш кращий результат.

Пошуковий запит вводиться в пошукове поле, він може містити ключові слова і спеціальні символи, що дозволяють встановити взаємозв'язок між цими словами і ввести додаткові параметри. Більшість цих символів представлено в наступній таблиці.

Синтаксис мови запитів системи Яндекс.		
Символ	Призначення	Приклад
" "	пошук фрази	"червона шапочка" (еквівалентно червона /+1 шапочка)
+	обов'язкова наявність слова в знайденому документі	+бути або +не бути
~~ або -	не повинно бути слова в межах документа (I НЕ)	путівник по парижеві ~~ (агентство тур)
~	не повинно бути слова в межах	банки ~ закон

	одного словосполучення (I НЕ)	
!	шукати лише зазначену форму слова	!Ющенко
пробіл або &	логічне I (у межах пропозиції)	фабрика зірок
&&	логічне I (у межах документа)	музика && (фабрика зірок)
	логічне АБО	малюнок картинка фото колаж
/(n m)	відстань між словами (-назад +вперед)	постачальники /2 кава музичне /(-2 4) утворення вакансії ~ /+1 студентів
&&/(n m)	відстань у кроках (-назад +вперед)	банк && /1 податки
()	групування слів	(технологія виготовлення) (сиру сиру)

Основні положення мови запитів:

- Якщо ключові слова є стійким словосполученням або єдиною фразою, то вкладіть їх у лапки.
- Якщо слова не об'єднані лапками, то кожне слово буде саме по собі і перед кожним з них можна поставити знак плюс «+», якщо слово обов'язково повинно бути в знайдених документах, мінус «-», якщо слово не повинно бути в знайдених документах (пробіл ставиться перед знаком, але не після). Якщо перед словом поставити знак ~ (тильда), то цього слова не повинно бути в межах одного словосполучення в сукупності з поруч розташованим в запиті словом. *Примітка:* за замовчуванням будуть знайдені і ті документи, що задовольняють хоча б одному з ключових слів. Такі посилання будуть мати низьку релевантність і будуть знаходитися наприкінці результатів запиту.
- Незалежно від того, у якій формі ви вжили слово в запиті, Яндекс враховує усі форми цього слова за правилами російської мови. Щоб цього не відбувалося, поставте знак оклику перед незмінним словом.
- Усі слова, написані через пробіл або знак & (логічне I) повинні одночасно знаходитися в знайдених документах у межах одного словосполучення. Усі слова, написані через && повинні одночасно знаходитися в знайдених документах, але відстань між ними не враховується.
- Слова, написані через символ | (логічне АБО) є замінюючими одне одного (синонімами), і будуть знайдені документи, що задовольняють хоча б одному з цих слів.
- Можна вказати відстань між словами. Якщо пронумерувати слова в словосполученнях, то відстань між словами – це різниця номерів слів. Наприклад, якщо між двома словами може знаходитися лише одне слово, то відстань між ними дорівнює 2 (3 мінус 1). Число вказується після знаку /, наприклад регіональний /2 центр. У цьому випадку будуть знайдені документи, у яких ці слова знаходяться або разом, або між ними є ще одне

слово. Запис /2 еквівалентна запису /(-2 +2), у такій формі можна вказати максимальну і мінімальну кількість слів, наприклад, від 3 до 5 записується /(3 5). Мінус і плюс вказують на порядок слів: мінус – зворотний порядок. Якщо перед символом / указати &&, то відстань буде обчислюватися в кроках.

- Для групування окремих частин запиту використовуйте круглі дужки.

4.4.3. Пошук у зонах і елементах web-сторінки.

Web-сторінка складається з визначених зон і елементів. Відповідно можна здійснювати пошук у зонах і в елементах. Наприклад, для пошуку в заголовку сторінки (заголовок відображається в заголовку вікна оглядача) вказують: *\$title (вираз)*, пошук у тексті посилань аналогічний (див. наступну таблицю), а загальний синтаксис такий: *\$ім'я_зони (вираз)*.

Примітка: вираз може бути представлений як одним ключовим словом, так і декількома словами, об'єднаними зазначеними вище знаками логічних операцій.

Для пошуку в елементах використовується синтаксис: *#ім'я_елемента=(вираз)*.

Елементи відрізняються від зон тим, що їх здебільшого не видно користувачеві, що переглядає сторінку. Так, наприклад, ключові слова вказуються в невидимому заголовку Web-сторінки і не виводяться в браузері. Їх можна побачити лише виконавши Вид - В виде HTML. Більшість документів мають опис (abstract), що також не виводиться на сторінку. Підпис малюнків видно лише в тому випадку, якщо завантаження малюнків відключене, або якщо навести мишку на малюнок, то підпис з'явиться у вигляді підказки.

Синтаксис пошуку в елементах і зонах		
Синтаксис	Призначення	Приклад
\$title (вираз)	пошук у заголовку	\$title (Яндекс)
\$anchor (вираз)	пошук у тексті посилань	\$anchor (Яндекс Апорт)
#keywords=(вираз)	пошук у ключових словах	#keywords=(пошукова система)
#abstract=(вираз)	пошук в описі	#abstract=(искалка пошук)
#image="значення"	пошук файлу зображення	#image="tort*"
#hint=(вираз)	пошук у підписах до зображень	#hint=(lenin лєнін)
#url="значення"	пошук на заданому сайті (сторінці)	#url="www.comptek.ru*"
#link="значення"	пошук посилань на заданий	#link="www.yandex.ru*"

	URL	
#mime="значення"	пошук у документах даного (pdf або rtf) типу	#mime="pdf"

4.4.4. Сортування результатів запиту.

Після того, як пошукова система вибере сторінки, що задовольняють запиту, вона сортує посилання на ці сторінки в порядку спадання їхньої релевантності.

Релевантність – це ступінь відповідності змісту документа пошуковому запиту. Релевантність документа залежить від ряду факторів, у тому числі від частотних характеристик шуканих слів, ваги слова або виразу, близькості шуканих слів у тексті документа один від одного і т.д.

Користувач може вплинути на порядок сортування, використовуючи оператори ваги й уточнення запиту. Вага вказується для того, щоб збільшити релевантність документів, що містять слово або вираз, вага якого зазначена.

Синтаксис: слово:число або (пошуковий_вираз):число.

Чим більша вага зазначена для слова (або виразу), тим вище релевантність документів, що його містять.

Наприклад, по запиту батьківщина Ющенка:5 у результатах пошуку нагорі списку будуть знаходитися документи, в яких частіше зустрічається саме слово Ющенко.

Уточнює слово або вираз застосовується для того, щоб збільшити релевантність документів, що їх містять.

Синтаксис: <- слово або <- (уточнюючий_вираз).

Наприклад, по запиту телефон <- автовідповідач будуть знайдені всі документи, що містять слово телефон, але першими будуть видані сторінки, що містять слово автовідповідач.

Примітка: крім сортування по релевантності Ви можете вибрати сортування по даті документів, клацнувши по відповідному посиланню.

4.4.5. Пошук по каталогах

Коли мережевий робот пошукової системи сканує Internet і автоматично вносить описи й адреси сторінок у свій пошуковий індекс, то разом з корисними й адекватними посиланнями в базі даних виявляється багато пошукового сміття або пошукового спаму, тому що деякі несумлінні власники сайтів, намагаючись збільшити відвідуваність своїх ресурсів, додають у свої сторінки найбільш популярні пошукові слова, що зовсім не відносяться до теми сторінки. Це і деякі інші обставини змушують творців пошукових систем крім автоматичного індексу вести також каталоги ресурсів, складені вручну. Такий каталог є й у пошуковій системі Яндекс.

Каталог дозволяє класифікувати ресурси і згрупувати їх по тематиці. Тематичне дерево каталогу Яндекс зроблено максимально простим. На першому рівні в ньому всього вісім тем:

- Бізнес і економіка,
- Суспільство і політика,
- Наука й освіта,
- Комп'ютери і зв'язок,
- Довідки,
- Будинок і родина,
- Розваги і відпочинок,
- Культура і мистецтво.

Якщо клацнути за назвою теми, то відкриваються підтеми (рис.10). При подальшому русі всередину будуть відкриватися наступні рівні (підрубрики), і в підсумку відкриється список ресурсів (сайтів) обраної тематики.

Примітка: кількість рівнів у глибину не перевищує 4-х.

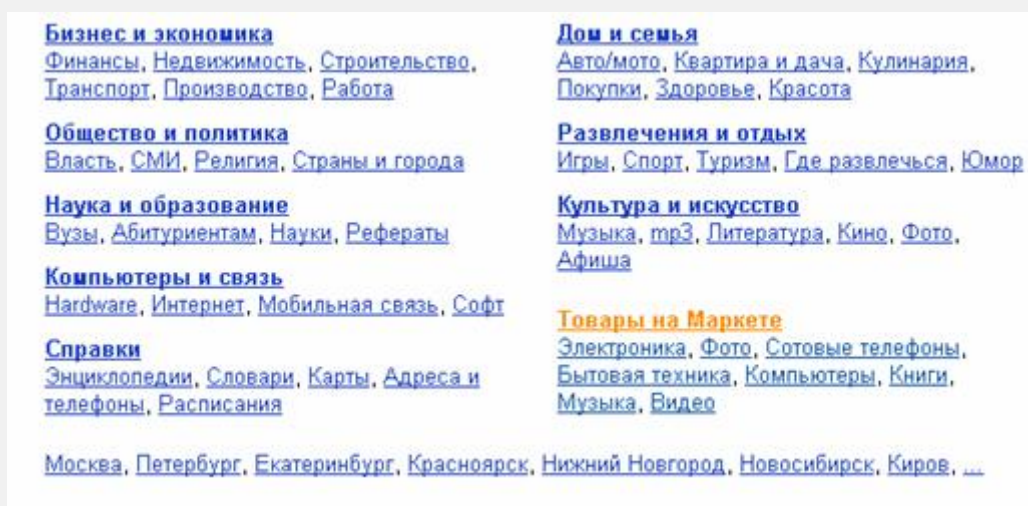


Рис. 10. Тематика каталогу системы Яндекс.

Ресурси в системі чітко структуровані по тематиці, однак, крім теми, у каталозі існує ряд додаткових ознак (фасет), що дозволяють уточнити характер ресурсів, що користувач хоче побачити в тематичних категоріях. Ці нетематичні ознаки характеризують ресурси по регіону, секторові економіки, ступені вірогідності (джерела) інформації, її потенційної аудиторії (адресатові інформації), жанрові (художня література, науково-технічна література, і т.д.), мети (пропозиція товарів і послуг, інтернет-представництво) і т.д.

Сайти в рубриках розташовані по спаданню їхнього тематичного індексу цитування. Тематичний індекс цитування – це показник, що показує кількість посилань на даний ресурс з інших ресурсів. При розрахунку цього індексу посиланням надається різна «вага» (значимість) у залежності від авторитетності сайту, на який посилається. Таким чином, на перших позиціях каталогу будуть саме найбільш визнані ресурси, ресурси – джерела інформації, ресурси, що цитують, на які посилаються.

У каталозі можна також здійснювати пошук по ключових словах, як і у всьому пошуковому індексі, для цього необхідно після введення ключових слів у пошуковій формі клацнути по посиланню Каталог.

Пошук у каталозі ведеться по текстах заголовків і описів, складених редакторами пошукової системи Яндекс. Оскільки заголовки відбивають основний профіль сайта, сайти, знайдені по заголовку, варують у результатах пошуку більше, ніж сайти, знайдені по опису.

На закінчення необхідно відзначити, що якщо пошук у системі Яндекс як у всій базі, так і в каталозі не дав необхідних результатів, то спробуйте пошукати те саме в іншій пошуковій системі. Для цього на сторінці результатів пошуку є відповідні посилання.

Тема 5. Сучасні ІКТ у металургійній промисловості

5.1. Металургійні довідкові системи

Металургійні довідкові системи можна знайти на сайті державного підприємства «Український Інститут Інтелектуальної власності (Укрпатент) <http://www.uipv.org/ua/bases2.html>, де вони групуються за розділами.

Винаходи та корисні моделі:

- ІДС "Стан діловодства за заявками на винаходи і корисні моделі";
- ІДС "Відомості про патенти України на винаходи (корисні моделі), які визнані недійсними, та патенти, які втратили чинність";
- ІДС "Міжнародна патентна класифікація. Повний текст (2016.01)" (чинна з 01.04.2016)

Знаки для товарів і послуг:

- ІДС "Відомості про стан діловодства за заявками на знаки для товарів і послуг;
- ІДС "Відомості про свідоцтва України на знаки для товарів і послуг, дія яких припинена, або які визнано недійсними";
- ІДС «Міжнародна класифікація товарів і послуг для реєстрації знаків (Ніццька класифікація). Десята редакція, версія 2016 року» (чинна з 01.01.2016);

ІДС "Міжнародна класифікація зображувальних елементів знаків (Віденська класифікація). Сьома редакція" (чинна з 01.01.2013)

Промислові зразки:

- ІДС "Стан діловодства за заявками на промислові зразки в Україні";
- ІДС "Відомості про патенти України на промислові зразки, які визнані недійсними, та патенти, які втратили чинність";

- ІДС "Міжнародна класифікація промислових зразків (Локарнська класифікація). Десята редакція" (чинна з 01.01.2014);
- ІДС "Міжнародна класифікація промислових зразків (Локарнська класифікація). Дев'ята редакція" (втратила чинність, доступна для інформаційних цілей).

5.2. Бази патентування та ліцензування

Державне підприємство «Український Інститут Інтелектуальної власності (Укрпатент)» має цифрову патентну бібліотеку, що поділена за розділами:

- патенти України на винаходи;
- патенти України на корисні моделі;
- патенти України на промислові зразки;
- патенти на винаходи зарубіжних країн;
- заявки на винаходи;
- міжнародні класифікації;
- нормативно-правові акти.

Діючи електронні бази даних наведені нижче.

Офіційний бюлетень:

- БД "Електронна версія акумулятивного офіційного бюлетеня "Промислова власність".

Винаходи та корисні моделі:

- БД "Перспективні винаходи України";
- спеціалізована БД "Винаходи (корисні моделі) в Україні";
- БД колективного користування "Відомості про заявки на винаходи, які прийняті до розгляду".

Знаки для товарів і послуг:

- БД "Зареєстровані в Україні знаки для товарів і послуг";
- БД "Відомості про добре відомі знаки в Україні";
- інтерактивна БД "Заявки на знаки для товарів і послуг, прийняті до розгляду".

Промислові зразки:

- інтерактивна БД "Промислові зразки, зареєстровані в Україні".

Зазначення походження товару:

- відомості про зареєстровані в Україні КЗПТ

5.3. Бібліотечні ресурси в металургії

Центральна державна науково-технічна бібліотека гірничо-металургійного комплексу України (ДЗ «ІДНТБ ГМК України») – державна спеціалізована бібліотека з галузей металургії, гірничої справи та суміжних

галузей. Єдина у Східному регіоні. ДЗ "ЦДНТБ ГМК України" - регіональний центр міжбібліотечного обслуговування та єдиний в країні державний депозитарій з металургії.

Завдання бібліотеки. Основними завданнями бібліотеки з самого початку були і залишаються:

- інформаційний супровід та документальне забезпечення галузевих програм розвитку металургії України;
- забезпечення фахівців галузі та суміжних галузей промисловості достовірною інформацією;
- координаційна та методична допомога НТБ підприємств галузі;
- збереження фондів літератури з металургії.

У розпорядженні: добре укомплектованої і раціонально організованою документальною базою даних (як на паперових носіях , так і в електронному вигляді) по чорній металургії, гірничорудної промисловості, технології металів, трубопрокатному виробництву, коксохімії, вогнетривами, металовироби, промислового будівництва, внутрішньозаводського транспорту та ін галузям промисловості. У фонді «ЦДНТБ ГМК України» повно представлені всі види і типи літератури , включаючи описи винаходів до авторських свідоцтв, каталоги на промислове обладнання, стандарти України , міждержавні стандарти, норми , методичні інструкції та вказівки, правила та іншу науково - технічну документацію.

Видає: більш 10 найменувань інформаційно-бібліографічних видань, по всіх переділах металургійної промисловості , включаючи весь спектр науково - технічної літератури, патентної, каталожної та нормативної документації, що забезпечує інформаційний супровід і документальне забезпечення програми розвитку ГМК України.

Складає: бібліографічні списки до курсових, дипломних, дисертаційних робіт; аналітичні, фактографічні довідки та огляди; тематичні добірки, сигнальну інформацію про переклади статей з поточних зарубіжних журналів; тематичний пошук літератури; поточне інформування по заданій тематиці.

Виконує: переклади з англійської , французької, української мов;

Проводить: Тематичні виставки та перегляди літератури, супроводжувані оглядами актуальних джерел та консультаціями перекладача і бібліографа; дні інформації; дні фахівців . На основі електронних баз даних, що включають довідково-пошуковий апарат « ЦДНТБ ГМК України ».

Надає послуги: за тематичним пошуку джерел; з комп'ютерної верстки документів; З метою підвищення кваліфікації бібліотечних працівників надає методичну допомогу науково - технічним бібліотекам галузі.

Організує: практикуми, семінари, конференції для НТБ галузі.

5.4. Системи супроводу виробничих процесів

Основними завданнями систем супроводу:

- забезпечення стійкості роботи системи і збереження інформації – адміністрування;
- своєчасна модернізація і ремонт окремих елементів – технічна підтримка;
- адаптація можливостей системи, що експлуатується, до поточних потреб бізнесу підприємства – розвиток системи.

Ці роботи необхідно включати в оперативний план інформатизації підприємства, який повинен формуватися обов'язково з дотриманням усіх умов стратегічного плану. В іншому випадку в межах існуючої системи можуть з'явитися фрагменти, які в майбутньому зроблять ефективну експлуатацію системи неможливою. Зараз за рубежом стало загальноприйнятим передавати функції технічної підтримки і частково адміністрування постачальникам системи або системним інтеграторам. Ця практика одержала назву "аутсорсинг". Часто в межах аутсорсингу стороннім підприємствам передаються й такі функції, як створення і підтримка резервних сховищ даних і центрів виконання критичних бізнес-додатків, які задіюються у разі стихійного лиха або інших особливих умов.

Особливу увагу на етапі експлуатації і супроводу потрібно приділити питанням навчання персоналу і, відповідно, плануванню інвестицій у цей процес.

Життєвий цикл (ЖЦ) утворюється відповідно до принципу низхідного проектування і зазвичай має ітераційний характер: реалізовані етапи, починаючи з найперших, циклічно повторюються відповідно до змін вимог і зовнішніх умов, введення обмежень тощо. На кожному етапі ЖЦ породжується певний набір документів і технічних рішень, при цьому для кожного етапу початковими є документи і рішення, отримані на попередньому етапі. Кожний етап завершується верифікацією породжених документів і рішень з метою перевірки відповідності їх вихідним.

Існуючі моделі ЖЦ визначають порядок виконання етапів у ході розробки, а також критерії переходу від етапу до етапу. Відповідно до цього найбільше поширення отримали такі три моделі:

1. Каскадна модель (70 – 80-ті роки) передбачає перехід до наступного етапу після повного завершення робіт на попередньому етапі і характеризується чітким поділом даних і процесів їх опрацювання.

2. Поетапна модель з проміжним контролем (80 – 85-ті роки) – ітераційна модель розробки з циклами зворотного зв'язку між етапами. Перевага такої моделі в тому, що між етапні коригування забезпечують меншу трудомісткість порівняно з каскадною моделлю; з іншого боку, час життя кожного з етапів розтягується на весь період розробки.

3. Спіральна модель (86 – 90-ті роки) – загострює увагу на початкових етапах ЖЦ: аналізі вимог, проектуванні специфікацій, попередньому й детальному проектуванню. На цих етапах перевіряється і обґрунтовується реалізованість технічних рішень створенням прототипів. Кожний виток спіралі відповідає поетапній моделі створення фрагмента або версії системи, на ньому уточнюються цілі й характеристики проекту, визначається його якість, плануються роботи наступного витка спіралі. Таким чином

поглиблюються і послідовно конкретизуються деталі проекту і в результаті обирається обґрунтований варіант, який доводиться до реалізації.

Фахівці відзначають такі переваги спіральної моделі:

- накопичення і повторне використання програмних засобів, моделей і прототипів;
- орієнтація на розвиток і модифікацію системи в ході її проектування;
- аналіз ризику і витрат у процесі проектування.

При використанні спіральної моделі відбувається накопичення і повторне використання проектних рішень, засобів проектування, моделей і прототипів інформаційної системи та інформаційної технології; здійснюється орієнтація на розвиток і модифікацію системи і технологи в процесі їх проектування; проводиться аналіз ризику та витрат у процесі проектування систем і технологій.

5.5. Металургійні системи типу B2B, B2C, C2C, C2B

B2B (англ. Business to Business) - термін, що визначає вид інформаційної та економічної взаємодії, класифікований за типом взаємодіючих суб'єктів, в даному випадку - це юридичні особи.

B2B - скорочення від англійських слів «business to business», в буквальному перекладі - бізнес для бізнесу. Це сектор ринку, який працює не на кінцевого, рядового споживача, а на такі ж компанії, тобто на інший бізнес. Прикладом B2B- діяльності може послужити виробництво барних стійок або надання рекламних послуг: фізичним особам реклама ні до чого, а от іншим організаціям вона необхідна.

Під поняттям B2B також мають на увазі системи електронної комерції, або системи електронної торгівлі - програмні-апаратні комплекси, які є інструментами для здійснення торгово-закупівельної діяльності в мережі Інтернет.

У західних країнах, під терміном B2B часто розуміється забезпечення яких-небудь виробничих фірм супровідними послугами, додатковим обладнанням і т. д. Однак в загальному сенсі визначення B2B відповідає будь-яка діяльність, спрямована на клієнтів, які є юридичними особами.

B2C. Business-To-Consumer (укр. Бізнес для Споживача) - форма електронної торгівлі, метою якої є прямі продажі для споживача.

B2C ефективний для усунення відмінностей між великими містами та віддаленими регіонами в сенсі доступності товарів і послуг для споживача. B2C дозволяє вести прямі продажі з мінімальною кількістю посередників. Усунення посередників дає можливість встановлювати конкурентні ціни на місцях і навіть збільшувати їх (виключаючи винагороду посередників), що, природно, призведе до зростання прибутку.

Система є однією з ланок в ланцюжку сучасних бізнес-процесів, і ця ланка будує бізнес-відносини наступного плану взаємодій: "Бізнес-клієнт".

B2C - це концепція побудови бізнес-процесів підприємства та комплекс Інтернет-технологій та інструментів, що забезпечують підвищення прозорості підприємства і полегшують його взаємодію з клієнтами. Один з найбільш популярних інструментів B2C - Інтернет-магазин.

B2C (Business-to-Consumer) - термін, що позначає комерційні взаємини між організацією (Business) і приватним, так званим, "кінцевим" споживачем (Consumer) /D.Chaffey, "E-Business and E-Commerce Management", 2007. Часто використовується для опису діяльності, яку веде підприємство, тобто, в даному випадку - продаж товарів і послуг, безпосередньо призначених для кінцевого використання.

Поряд з цим терміном існують і інші, також позначають різні види відносин, як комерційного, так і некомерційного характеру, наприклад: B2B (Business-to-Business) - "взаємини між комерційними організаціями", C2C (Customer-to-Customer) - "взаємини між споживачами", B2G (Business-to-Government) - "взаємини між організацією і урядом" і т.д. Ці терміни широко застосовуються в англomовному середовищі, як в теоретичних працях на тему ділового адміністрування, так і в повсякденному діловому спілкуванні.

B2G (business-to-government) - відносини між бізнесом і державою. Зазвичай термін використовується для класифікації систем електронної комерції. Прикладом B2G-систем можуть служити системи електронних держзакупівель.

G2B (англ. Government to Business) - набір програмних і апаратних засобів для здійснення он-лайн взаємодії виконавчої влади і комерційних структур з метою підтримки і розвитку бізнесу. Зокрема, до Класу G2B можна віднести інформаційні веб-сайти органів влади, системи електронних закупівель та ін.

Тема 6. Розподілені системи у сучасних ІКТ середовищах Internet

6.1. Системи розподіленого зберігання

Уявімо, що у вас є переносний комп'ютер з невеликим вінчестером. Також у вас є стаціонарний комп'ютер, на який і з якого ви можете пересувати через мережу файли. Ви працюєте, і в певний момент вам бракує місця на переносному комп'ютері – й ви пересуваєте частину файлів на стаціонарний. А ще частину – на зовнішній вінчестер. І ви опиняєтесь в становищі, коли ваші дані розпорошені між трьома різними пристроями.

Чи не чудово було б, якби в кожен такий момент ви могли продовжувати бачити перелік усіх ваших файлів, а також мали можливість легко отримати будь-який з них, не шукаючи перед тим півдня, на якому пристрої ви його розмістили? І якщо вам треба «розчистити» місце для отримання інших файлів, бути певними, що ви не вилучите файли, яких

більше ніде немає? І ще: якщо ви змінюєте, переміщуєте чи перейменовуєте файли на одному з ваших пристроїв, щоб ці самі зміни автоматично відтворювались і на решті ваших пристроїв, на які ви скопіювали цю частину своїх файлів? Ви, мабуть, довго мріяли про таку можливість, чи не так?

6.2. Системи розподілених розрахунків

Розподілені обчислення (розподілена обробка даних) – спосіб розв'язання трудомістких обчислювальних завдань з використанням двох і більше комп'ютерів, об'єднаних в мережу.

Розподілені обчислення є окремим випадком паралельних обчислень, тобто одночасного розв'язання різних частин одного обчислювального завдання декількома процесорами одного або кількох комп'ютерів. Тому необхідно, щоб завдання, що розв'язується було сегментоване – розділене на підзадачі, що можуть обчислюватися паралельно. При цьому для розподілених обчислень доводиться також враховувати можливу відмінність в обчислювальних ресурсах, які будуть доступні для розрахунку різних підзадач. Проте, не кожне завдання можна «розпаралелити» і прискорити його розв'язання за допомогою розподілених обчислень.

Список українських проектів розподілених обчислень:

- Magnetism@Home;
- SLinCA@Home – дослідження в галузі матеріалознавства;
- Physics@home – фізика твердих тіл, матеріалів, оптика та хімія.

Біологія та медицина:

- Rosetta@home – вирахування структури білка із самою низькою енергією;
- Folding@Home – проект з вирахування третинної структури білків;
- World Community Grid – дослідження, пов'язані із генетикою людини, а також із різноманітними важкими захворюваннями. Вивчення білків, та їх каталогізація;
- grid.org – проект з пошуку ліків від Карциноми (закінчений 27 квітня 2007 року (не знайдено));
- Find-a-Drug – проект з пошуку ліків від різноманітних хвороб шляхом обрахунку докінгу білків із різними молекулами (закінчено 2005 року), приєднався до WCG;
- DrugDiscovery@Home – російський проект, згортання білків, скрінінг баз даних біологічно активних сполук;
- Proteins@home;
- GPUGrid.net – молекулярне моделювання білків;
- POEM@HOME – передбачення біологічно активної структури білків;
- FightMalaria@Home – пошук ліків проти малярії;
- Docking@Home – дослідження взаємодії білків-ліганд;
- Malariacontrol.net – моделювання динаміки розповсюдження і впливу малярії на здоров'я;
- RNA World – дослідження пов'язані з ідентифікацією, аналізом і передбаченням структури РНК;

- SIMAP – пошук подібності між білками.

Математика:

- Seventeen or Bust – проект, який займається підтвердженням задачі Серпінського;

- GIMPS – проект з пошуку простих чисел Мерсенна;
- ZetaGrid – перевірка гіпотези Рімана (закінчено 2005 року);
- ABC@home – проект з пошуку ABC-трійок;
- Collatz Conjecture – дослідження гіпотези Колатца;
- PrimeGrid – проект підтримує декілька підпроектів з пошуку великих простих чисел різних форм;
- NFS@home – факторизація великих цілих чисел;
- NumberFields@home – дослідження у галузі теорії чисел;
- primaboinca – пошук контрприкладів до двох гіпотез із галузі теорії чисел;
- SubsetSum@Home – проект у галузі теорії чисел;
- SZTAKE Desktop Grid – пошук всіх систем узагальнених двійкових чисел.

Криптографія:

- distributed.net – проекти зі злому RC5-72 повним перебором, пошук оптимальних лінійок Голомба - OGR-27;

- RainbowCrack;
- AQUA@home – моделювання роботи адіабатичного квантового надпровідного комп'ютера (завершено 2011 року);
- Enigma@Home – спроба розшифрувати 3 повідомлення, зашифрованих німецькою шифрувальною машиною Енігма. Сигнали були перехоплені в Північноатлантичному океані в 1942 році і вважалося, що їх неможливо зламати;

- DistrRTgen – генерація гігантських "веселкових" таблиць, що дозволяють зламати довгі паролі.

Когнітивна наука та штучний інтелект:

- MindModeling@Home – дослідження механізмів з покращення процесів навчання та людського мислення.

Астрономія та астрофізика:

- SETI@Home – проект з обробки сигналів радіотелескопу, для пошуку радіосигналів позаземних цивілізацій;

- Einstein@Home – проект з перевірки гіпотези Ейнштейна про гравітаційні хвилі за допомогою аналізу гравітаційних полів пульсарів чи нейтронних зірок;

- MilkyWay@home – створення трьохмірної моделі галактики Чумацький шлях;

- Cosmology@home – пошук моделі, яка найкращим чином описує наш Всесвіт, а також визначення діапазону моделей, які узгоджуються із сучасними астрономічними та фізичними даними;

- Asteroids@home – метою проекту є визначення форми і характеру обертання значної частини астероїдів.

Фізика та хімія:

- Spinhenge@home – проект в області нанотехнологій, вивчення магнітних молекул;
- LHC@home – розрахунки магнітної підсистеми прискорювача заряджених частинок – Великого адронного колайдера (LHC);
- Muon1 DPAD – розрахунки з проектування ще потужнішого прискорювача – Neutrino Factory/Muon Collider;
- EDGeS@Home – проект у галузі фізики плазми, моделювання поведінки заряджених частинок в магнітному полі термоядерного реактора ITER;
- eOn;
- Leiden Classical – дослідження в галузі фізики поверхонь на основі Classical Dynamics.

Наука про Землю:

- Climateprediction.net – проект з моделювання впливу відходів вуглекислого газу на клімат Землі;
- Quake Catcher Network – створення найбільшої сейсмічної карти світу;
- Radioactive@Home – створення найбільшої карти радіаційного рівня.

ПЗ для організації розподілених обчислень:

- Apache Hadoop;
- BOINC – відкрита інфраструктура для розподілених обрахунків Університету Берклі (Berkeley Open Infrastructure for Network Computing), яка поширюється за ліцензією LGPL;
- Condor(англ.);
- Globus Toolkit(англ.) – набір програм, який значно спрощує створення та керування розподіленими обрахунками.

6.3. Сучасні термінальні системи

Термінальний доступ – доступ до інформаційної системи (ІС), організований так, що локальна машина-термінал не виконує обчислювальної роботи, а лише здійснює перенаправлення вводу інформації (від миші і клавіатури) на центральну машину (термінальний сервер) і відображає графічну інформацію на монітор. Причому вся обчислювальна робота в термінальній системі виконується на центральній машині.

Історично термінальний доступ вперше був організований на комп'ютерах, здатних одночасно обслуговувати декілька обчислювальних процесів. Це дозволило більш раціонально розподіляти обчислювальні ресурси між користувачами перших дуже дорогих обчислювальних машин. З появою дешевих персональних комп'ютерів (ПК) роль термінального доступу стала трохи знижуватися, оскільки склалася думка, що достатню продуктивність ІС можна отримати на робочому столі кожного користувача ПК.

Проте надалі стало очевидним, що дешевизна ПК не в змозі компенсувати щоденні витрати на супровід великої кількості робочих місць користувачів, що володіють нібито перевагами через можливість персоналізації налаштувань операційних систем (ОС) і ПЗ. Реально (у великих організаціях), наявність великої кількості «різношерстого» обладнання замість достоїнств створює додаткові складності користувачам і системним адміністраторам. Питання забезпечення безпеки ІС, також зажадали перегляду поглядів і повернення до термінального доступу, як більш уніфікованому і економічно виправданого.

Термінальний сервер (англ. terminal server) – сервер, що надає клієнтам обчислювальні ресурси (процесорний час, пам'ять, дисковий простір) для вирішення завдань. Технічно термінальний сервер – надпотужний комп'ютер (або кластер), підключений до мережі з термінальними клієнтами – у котрих є, як правило, малопотужні або застарілі робочі станції або спеціалізовані для доступу до термінального сервера. Термінальний сервер служить для віддаленого обслуговування користувача з наданням робочого столу.

Тема 7. Структура та принципи створення хмарних сховищ даних

Хмарне сховище даних це онлайнове сховище для файлів. Завантаживши файли в таке сховище, ви зможете отримати до них доступ з будь-якої точки світу, де є Інтернет. По суті це онлайн-флешка, яку не треба носити з собою і яку ви ніколи не забудете удома.

З технічного боку хмарне сховище даних відрізняється від звичайних серверів для зберігання файлів тим, що для зберігання даних використовується величезна кількість серверів. І ваші файли можуть бути розкидані після різних серверів. Але користувача не хвилює внутрішня структура сервісу. Із його точки зору хмара це один великий сервер з яким він і працює.

Зараз хмарні сховища даних набирають величезну популярність. Їх використовують не лише комерційні структури, але і приватні користувачі.

Dropbox – найбільш популярний сервіс такого роду. Сервіс працює вже багато років і за цей час встиг обзавестися великою кількістю прихильників. При реєстрації користувач отримує 2 Гб безкоштовного вільного місця в хмарному сховищі. Може здатися, що це дуже мало, але сервіс пропонує ряд способів, за допомогою яких ви можете безкоштовно розширити простір під свої файли до 22 Гб.

Окрім цього є можливість купити вільне місце. Dropbox пропонує два тарифні плани 50 Гб за 10 доларів в місяць і 100 Гб за 20 доларів.

Працювати з файлами в Dropbox можна як за допомогою звичайного браузера так і використовуючи спеціальне додатки клієнт, версії якого існують під усі популярні платформи.

Google Drive – хмарне сховище даних від компанії Google. Сервіс з'явився не давно і на даний момент пропонують користувачам 5

безкоштовних гігабайт вільного місця під зберігання файлів. Окрім безкоштовного вільного місця користувачам доступні тарифні плани на 25 і 100 Гб по 2,5 і 5 доларів в місяць. Що істотно дешевше за пропозицію від Dropbox.

Google Drive був побудований на базі старого сервісу Google Docs. По цьому Google Drive надає значно більше функцій чим звичайний сервіс для зберігання даних. Користувачі Google Drive можуть створювати текстові документи, презентації, електронні таблиці і інші документи прямо в інтерфейсі браузера. При цьому усі створені таким чином файли не витрачають вільне місце в сховищі.

В цілому Google Drive зручніший чим Dropbox. Інтерфейс сервісу простіший і інтуїтивно зрозуміліший.

SkyDrive –сервіс хмарного зберігання файлів від компанії Microsoft. При реєстрації користувачам доступно 7 Гб вільного простору, які можна розширити за допомогою платних тарифів. У SkyDrive існує 3 платні тарифи: на 20, 50 і 100 Гб за ціною в 310, 780 і 1570 рублів в рік, що навіть дешевше чим в Google Drive.

Також як і в Google Drive користувачі можуть створювати офісні документи прямо у браузері. Інтерфейс офісних онлайн додатків один в один повторює інтерфейс офісних програм з пакету Office 2010. Тому розібратися з ними не складе труднощів.

SkyDrive можна було б назвати кращим хмарним сховищем, але є один досить значний мінус. Максимальний розмір файлу, який можна завантажити в це сховище, обмежений 50 мегабайтами.

ЯндексДиск – хмарне сховище даних від пошукової системи Яндекс. Цей сервіс знаходиться на стадії тестування і доступ до нього можливий тільки після запрошень. Інтерфейс сервісу ЯндексДиск є частиною інтерфейсу ЯндексПошти. Також є присутніми додатки-клієнти для роботи з усіма популярними платформами.

При реєстрації користувачеві доступно 3 Гб вільного місця, яке можна безкоштовно розширити до 10 Гб. Інформації про платне розширення відведеного місця доки немає.

7.1. Особливості використання структури хмарних сховищ даних

Залежно від моделі розгортання вони бувають: private cloud - приватна хмара, призначене для використання однією організацією, може знаходитися як у власності цієї організації, так і якийсь інший. public cloud - публічне хмара, призначене для вільного використання різними користувачами різних компаній. hybrid cloud - гібридне хмара - комбінація з двох або більше різних хмарних інфраструктур (приватних, публічних або суспільних), що залишаються унікальними об'єктами, але пов'язаних між собою стандартизованими або приватними технологіями передачі даних і додатків. community cloud - суспільне або комунальне хмара, призначене для

використання конкретним спільнотою споживачів з організацій, що мають спільні завдання (наприклад, вимог безпеки). Громадське хмара може перебувати в кооперативній (спільної) власності, управлінні і експлуатації однієї або більше з організацій спільноти або третьої сторони (або будь-якої їх комбінації).

Способи реалізації хмари:

1. Приватне (для себе).
2. Кооператив (для групи).
3. Суспільне (публічне, для всіх).
4. Гібридне (комбінація всіх перерахованих).

Моделі надання хмарних послуг:

1. SaaS (Cloud Software as a Service) - додатки як послуга, «оренда».
2. PaaS (Cloud Platform as a Service) - програмна платформа як послуга, «розробка».
3. IaaS (Cloud Infrastructure as a Service) - інфраструктура як послуга, «міграція».

7.2. Програмні засоби для доступу к розподіленим системам

Програмні засоби складаються з набору послуг, що дозволяють декільком процесам, запущених на одній або декількох машинах взаємодіяти через мережу. Ця технологія забезпечує взаємодію при переході до розподілених архітектур, які використовуються найчастіше для підтримки та спрощення складних, розподілених прикладних програм. До проміжних програмних засобів відносять веб-сервера, сервера програм та інші аналогічні інструменти, які підтримують розробку та впровадження програм. Проміжні програмні засоби є невід'ємною частиною сучасних інформаційних технологій, заснованих на XML, SOAP, Web Service та сервісно-орієнтованій архітектури.

Такі програмні засоби розташовані між прикладними програмами, що працюють на різних операційних системах. Вони схожі на середній шар трирівневої архітектури, що знаходиться на одній системі, за винятком того, що вона розтягується на кілька систем або програм. Прикладами можуть бути бази даних, телекомунікаційні програмні засоби, монітори транзакцій, а також програмні засоби повідомлень і черг.

Відмінність між системними і проміжними програмними засобами, не дуже чітка. Якщо основна функціональність ядра надається самою операційною системою, то деякі функції, що раніше надавалися лише окремим проміжними програмними засобами тепер інтегровані в операційні системи. Типовим прикладом є стек TCP/IP в телекомунікаціях, який сьогодні включено практично у всі операційні системи.

У технологіях моделювання проміжні програмні засоби за звичай використовуються в контексті HLA, що застосовується в багатьох розподілених моделювання. Це – прошарок програмних засобів, що

знаходиться між прикладними програмами і інфраструктурою часу виконання.

7.3. Мобільні пристрої та засоби для доступу до розподілених систем

Мобільний агент – агент, не прив'язаний до системи, де розпочалось його виконання. Він має здатність переміщувати себе від однієї системи в мережі до іншої.

Здатність рухатись дозволяє мобільному агенту пересуватися до тієї системи, яка містить об'єкт, з яким цей агент хоче взаємодіяти, а також, працювати на тому ж комп'ютері або в тій же мережі, де знаходиться цей об'єкт.

Хоча технологія мобільних агентів викликає нечуване захоплення в сучасному світі, інтерес до них мотивується не технологією по суті, а скоріше перевагами, які з'являються при створенні розподілених систем на основі такої технології. Нижче наведено сім добрих причин для того, щоб використовувати мобільні агенти.

Вони зменшують завантаження в мережі. Розподілені системи часто залежать від протоколів зв'язку, які передбачають багатократну взаємодію для опрацювання одного завдання. Це особливо стосується тих систем, в яких велику увагу приділяється конфіденційності і безпеці роботи. Результатом є різке зростання мережевого трафіку. Мобільні агенти дозволяють пакувати можливий діалог (при віддаленому виклику процедури) і послати його до комп'ютера адресата, де взаємодія відбудеться локально. Мобільні агенти також корисні для зменшення трафіку при передачі великих обсягів даних в мережі. Дійсно, коли великі томи даних зберігаються в окремих комп'ютерах, наприклад, серверах баз даних, певну їх обробку було б зручніше виконувати там же, ніж передавати їх по мережі до машини, на якій встановлено програму обробки. Принцип тут простий: пересунути обчислення до даних скоріше, ніж дані до обчислень.

Вони долають мережеву латентність. Системам реального часу, як, наприклад, роботам в процесах виробництва, потрібно реагувати на зміни в їхніх середовищах максимально швидко. Контролювання таких систем через фабричну мережу великого розміру призводить до значної латентності, яка абсолютна неприйнятна для критичних систем реального часу. Використання в такій мережі мобільних агентів є чудовим рішенням, бо вони можуть бути послані від центрального контролера для дій у певному місці і безпосередньо виконувати його інструкції.

Вони інкапсулюють протоколи. Коли в розподіленій системі відбувається обмін даними, кожен комп'ютер має програмне забезпечення, яке реалізує протоколи, потрібні для належного кодування вихідних даних і інтерпретації вхідних. Однак, протоколи мають властивість розвиватися, а

перехід на нову, більш ефективну або захищену версію протоколу є важким, іноді цілком неможливим завданням по модернізації встановленого програмного забезпечення. Як результат, часто виникає проблема з підтримкою старих версій, або проблема несумісності версій протоколів. Мобільні агенти можуть пересуватися до віддалених комп'ютерів "каналами", які базуються на власних протоколах. До речі, перехід на нову версію протоколу в мережі можна виконати за допомогою мобільних агентів, які у визначеному порядку встановлюють на кожному комп'ютері нове програмне забезпечення.

Вони працюють асинхронно і автономно. Часто зовнішні мобільні пристрої (ноутбуки, мобільні телефони, автомобільні комп'ютери) залежать від дорогих або просто неякісних зв'язків з мережею. З цього випливає, що завдання, які вимагають безперервного зв'язку між мобільним зовнішнім пристроєм і деякою корпоративною мережею, певно будуть неекономними або технічно складними в реалізації. При використанні агентів такі завдання можуть бути вміщені в мобільні агенти, які після цього можуть бути послані в мережу. При встановленні зв'язку з мобільним пристроєм мобільні агенти переміщуються в його систему і функціонують незалежно від системи, що їх створила, працюють асинхронно і автономно. Мобільний зовнішній пристрій може повторно з'єднатися пізніше, щоб забрати агента і дізнатися про результат його дій.

Вони адаптуються динамічно. Мобільні агенти здатні аналізувати своє програмне середовище виконання і реагувати відповідно до змін, які в ньому відбуваються. Багато мобільних агентів здатні розподілятися між комп'ютерами в мережі таким чином, щоб підтримати оптимальну конфігурацію для розв'язання специфічної проблеми.

Вони природно різномірні (гетерогенні). Розподілені обчислення в своїй більшості різномірні, як з точки зору апаратних засобів, так і програмного забезпечення. Оскільки мобільні агенти є, в принципі, комп'ютер- і транспорт-незалежні, і залежать тільки від їх середовища виконання, то вони створюють оптимальні умови для безшовного інтегрування розподілених систем.

Вони стійкі до несправностей в мережі. Здатність мобільних агентів динамічно реагувати на несприятливі ситуації і події робить легким побудову стійких до критичних ситуацій розподілених систем. Якщо комп'ютер припиняється свою роботу, то виконання агентів на цій машині зупиниться, і їм буде наданий час, щоб переміститися і продовжити роботу на іншому комп'ютері в мережі.

Які види систем мобільних агентів доступні зараз? Після того, як з'явилась Java, на її платформі було згенеровано багато експериментальних систем мобільних агентів. Численні системи розробляються зараз, і більшість з них доступна для оцінки в Інтернеті.

Кількість систем зростає так швидко, що будь-яка спроба перелічити та класифікувати їх приречена на неповноту й застарілість, тому в цьому дослідженні здається доцільним згадати тільки найцікавіші базові системи