

Методи пошуку інформації в глобальній мережі Інтернет

Однією з тенденцій розвитку світової спільноти і системи освіти є підключення зростаючої кількості користувачів до глобальної мережі Інтернет. У будь-якому місці земної кулі стає доступним небачений раніше обсяг інформації. При цьому нерідко поліпшення програмних і технічних характеристик засобів доступу не вирішує, а, швидше, загострює проблему ефективного відбору релевантної (відповідної) інформації, показуючи невідповідність і учнів, і вчителів до використання можливостей, що відкриваються. Виникає парадокс: *чим більше інформації доступно, тим важче її відібрати.*

Саме тому одним з завдань курсу сучасного інформатики є навчання учнів здійснювати результативний пошук інформації, що потребує від них знань про способи її представлення в електронному вигляді, технології збереження та передавання через комунікаційну мережу Інтернет.

Учні повинні навчитися орієнтуватися в інформаційному просторі мережі Інтернет, ознайомитись із механізмами пошуку інформації в тематичних каталогах та індексних базах даних, навчитися використовувати пошукові інструменти для організації пошуку в базах даних, будувати пошукові запити і правильно оцінювати знайдену інформацію, уточнювати пошуковий запит, вміти за результатами пошуку отримувати потрібну інформацію із першоджерел.

При навчанні учнів методам пошуку інформації в Інтернет спочатку необхідно пригадати основні інформаційні і комунікаційні послуги мережі Інтернет, які є джерелами для одержання інформації (Рис.1) та підкреслити, що кожний з них може стати середовищем пошуку відповідних даних, в кожному з яких існують спеціальні правила організації такого процесу.

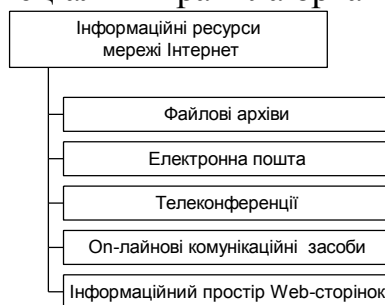


Рис 1.

Далі при повторенні особливостей роботи з інформаційними ресурсами доцільно разом з учнями скласти таблицю з відображенням специфіки їх інформаційного наповнення.

Таблиця 1.

Інформаційні ресурси	Інформаційний об'єкт	Інформація для пошуку
Файлові архіви	Файл	Ім'я файлу, папка розміщення
Електронна пошта	Електронний лист, зміст конфіденційний (тобто недоступна для індексації)	Електронна адреса людини

Телеконференції	Електронний лист	Електронні адреси людей, тема і зміст листа
Списки розсилки	Електронний лист	Тема і зміст листа
Комунікаційні засоби режиму On-line	Електронне повідомлення	Адреса для ідентифікації
Інформаційний простір Web-сторінок	Web-сторінка	Текстова і мультимедійна інформація

Далі вчителю доцільно звернути увагу на організацію пошуку інформаційних об'єктів в кожному виділеному ресурсі глобальної мережі Інтернет.

Сьогодні інформаційний простір Web-сторінок став універсальним по відношенню до всіх існуючих послуг мережі Інтернет, тому що є не тільки носієм інших інформаційних потоків але і містить в собі різні послуги мережі: поштові інтерфейси, систему телеконференцій, системи спілкування Chat і ICQ та інші. Крім того, кількість Web-сайтів з часом значно збільшується, тому виникає необхідність розглянути з учнями створення механізмів пошуку, структуризації, впорядкування, інформаційного простору Web-сторінок.

Учні повинні одержати спочатку загальне уявлення про пошукові служби, з'ясувати їх призначення та структуру. Важливо зауважити, що *пошукові служби* існують для зручності роботи в World Wide Web. При організації їх роботи задача користувача полягає в тому, щоб пояснити, що слід знайти в Інтернет, задача пошукової служби – надати готову Web-сторінку з посиланнями на адреси, по яких можна знайти потрібну інформацію.

Необхідно підкреслити, що на питання, яке сформульовано правильно, в Інтернет практично завжди можна одержати відповідь. Але при цьому не слід забувати, що питання задається не людині, а машині, яка аналізує інформацію в залежності від занесених до неї алгоритмів пошуку. Програми, які опрацьовують запити – це *інформаційно-пошукові системи (ІПС)*, що розробляються спеціалістами.

За принципом дії розрізняють кілька видів пошукових служб. Найбільш поширені три основних типи: *пошукові каталоги, рейтингові системи і індексні бази даних*. Учні повинні розуміти, що оскільки пошук інформації – це основна задача користувача в WWW-просторі, то в світі постійно з'являються нові типи і види пошукових служб, і вивчити всі їх неможливо, тому суттєвим є з'ясування принципів їх організації і роботи та правил використання.

З учнями доцільно скласти схему-карту розміщення інформаційного Web-простору, яка допоможе учням орієнтуватися серед Web-сайтів різноманітного тематичного призначення і здійснювати пошук в індексних базах даних (Рис.2).

Пошукові тематичні каталоги організовані за тим самим принципом, що і тематичні каталоги великих бібліотек. Звернувшись за адресою пошукового каталогу, користувач знаходить на його основній сторінці скорочений список великих тематичних категорій, наприклад таких, як *Освіта (Education)*.

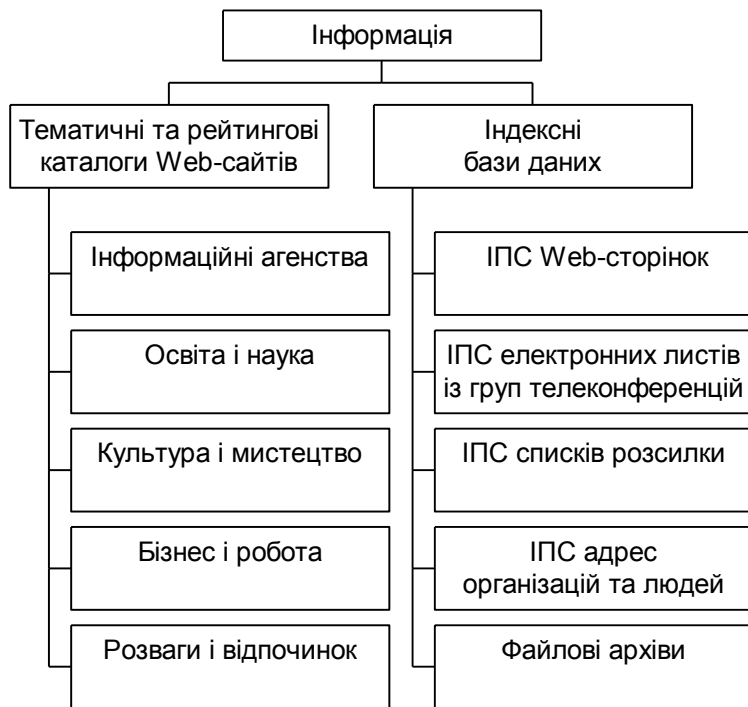


Рис. 2 .

Як правило, вони являють собою ієрархічні гіпертекстові меню з пунктами і підпунктами, які визначають тематику сайтів, адреси яких містяться в даному каталозі, з поступовим, від рівня до рівня, уточнюванням теми (Рис. 3).

Кожний запис в списку категорій – це *гіперссилка*. Натиснення на ній відкриває наступну сторінку пошукового каталогу, на якому дана тема представлена докладніше, наприклад по предметах: *Фізика, Хімія, Математика* і т. п. Вибір за допомогою мишки назви теми (наприклад, *Хімія*) відкриває сторінку із списком розділів (*Загальна хімія, Органічна хімія, Неорганічна хімія*). Продовжуючи занурення в тему, можна дійти до списку конкретних Web-сторінок і вибрати собі той ресурс, який краще підходить для розв'язування конкретної інформаційної задачі.

Працювати з пошуковими каталогами просто. У них пошук інформації відбувається на інтуїтивному рівні і практично завжди закінчується успіхом. Однак за цією простотою приховується найвища складність створення і ведення каталогу. Пошукові каталоги створюються, як правило, вручну. Висококваліфіковані редактори особисто переглядають інформаційний простір WWW, відбирають те, що на їх думку являє загальний інтерес, і заносять адреси до каталогу. Як правило, каталоги ресурсів складаються на основі експертних оцінок. Найбільшим пошуковим каталогом світу є пошукова служба Yahoo! (www.yahoo.com). На неї постійно працюють понад 150 редакторів, але і при цьому загальний обсяг каталогізованого простору складає біля 1 млн. Web-сторінок.

Найбільшим російським пошуковим каталогом є служба Атрус (www.atrus.ru), АУ! (www.au.ru), в Україні такими є: *Ukrainet Yellow Pages* (www.ukrainet.lviv.ua/yellow/pages.htm), *Nsearch* (www.search.kiev.ua) *Шерлок Холмс* (www.holms.ukrnet.net).

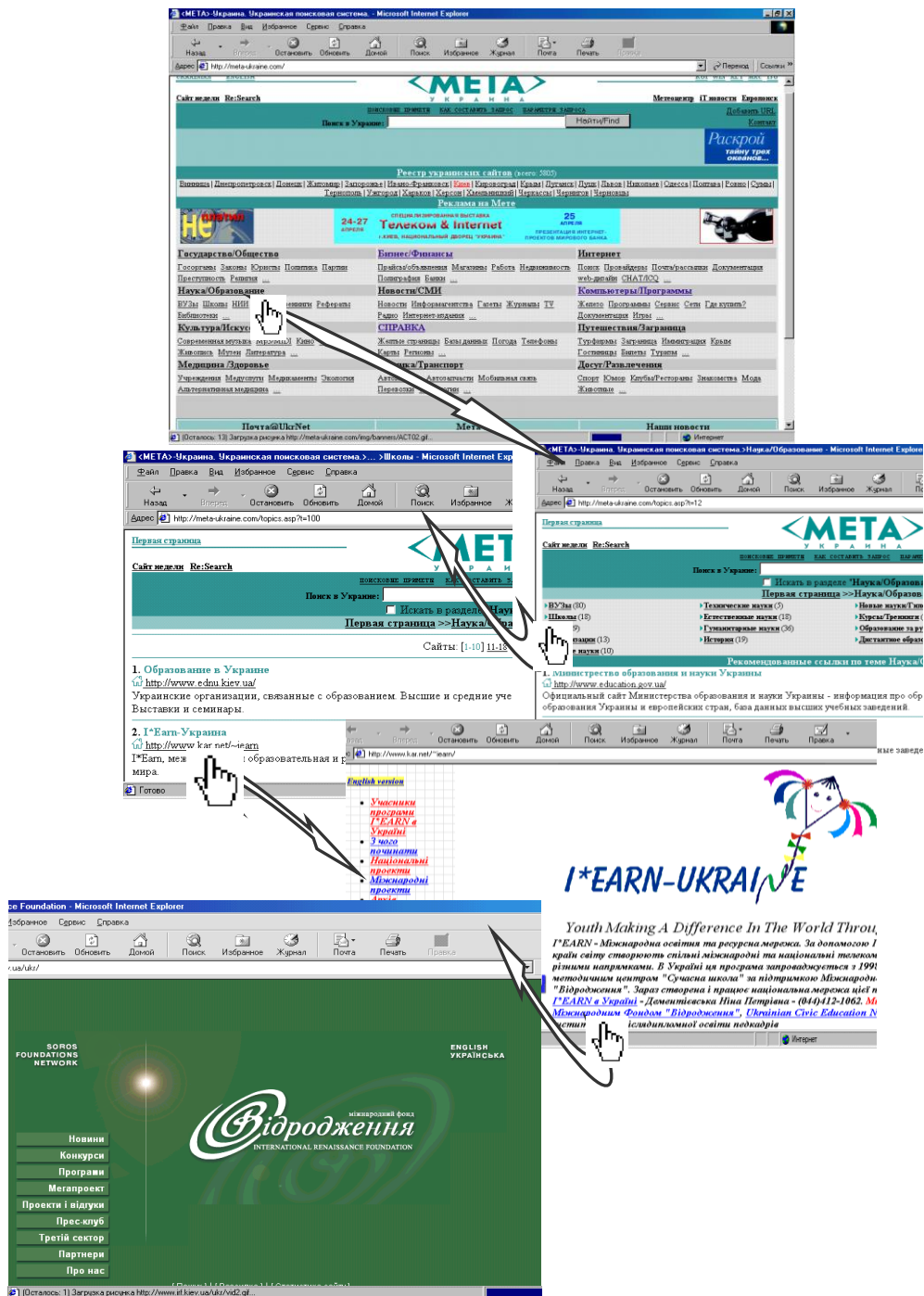


Рис.3

Далі з учнями необхідно розглянути основні характеристики Web-каталогів:

1. Спрямованість тематики і широта охоплення ресурсів: загальні, спеціалізовані, глобальні, регіональні.
2. Обсяг бази даних (загальна кількість силок у каталозі).
3. Структура рубрик, наявність кількох індексів.
4. Можливість пошуку по каталогу.

Учні повинні зрозуміти, що Web-каталоги є початковими засобами пошуку інформації, і дозволяють розв'язувати стандартні пошукові задачі:

пошук сайтів певного тематичного призначення. Але інколи в такий спосіб неможливо знайти потрібну інформацію або процедура пошуку затягується на невизначений час.

Основною перевагою такого засобу пошуку є чітка відповідність вмісту сайтів тематиці, яка оголошена для кожного розділу тематики (в окремих випадках – з додатковим впорядкуванням адрес в межах рубрики за результатами експертної оцінки або за рейтингом відвідування). Недоліком тематичних ПС є порівняно невелике охоплення існуючих ресурсів мережі, оскільки відслідкувати весь обсяг інформації, яка є в Інтернет, практично нереально навіть для значного колективу персоналу, якій підтримує сервер. Тому доцільно рекомендувати учням звертатися до каталогів ресурсів у випадку, коли необхідно швидко відшукати "типову" інформацію за тематикою, яка вимагається (якщо немає потреби відслідкувати будь-які її нюанси).

При формуванні практичних навичок учнів слід розглянути роботу з тематичним каталогом на прикладі. Можна запропонувати учням знайти інформацію про організований відпочинок взимку або інформацію про погоду на завтра, або інформацію про комп'ютерні курси тощо.

Сутність ідеї *рейтингової* системи заснована на тому, що на основній сторінці сервера розташовується список найбільш важливих тим, всередині кожної з яких організується список тематичних розділів. Вибравши потрібну тему, клієнт одержує обмежений список Web-вузлів, присвячених даній темі. Поруч з кожним Web-вузлом наводиться його оцінка популярності, яка вимірюється в кількості відвідувань за останню добу.

Використавши одну із посилань рейтингової системи, користувач переходить на рекомендований Web-вузол. При цьому збільшується показники лічильника кількості переходів. Ті, хто будуть користуватися рейтинговою системою пізніше, побачать зростання популярності даного Web-вузла.

Пошук джерел інформації за допомогою рейтингової системи можна розглядати як пошук *за рекомендаціями*. Далеко не завжди високий рейтинг об'єктивно відповідає якості Web-ресурсу, до якого веде популярна гіперссилка. Однак в тих випадках, коли мова йде про інформацію, що має суспільний інтерес, наприклад новини політики, економіки, культури, спорту і т. п., показнику популярності можна довіряти.

Вперше рейтингова система для пошуку Web-ресурсів була реалізована на сервері Top Hit (www.tophit.com). У Росії основним засобом рейтингового пошуку є Web-служба Ремблер (www.rambler.ru). В Україні це: Top Ping (www.topping.com.ua), Alpha Counter (www.a-counter.kiev.ua).

Більшість пошукових служб світу відноситься до *індексних баз даних*, які ще називають *пошуковими покажчиками*. Перед ними ставиться задача якнайкраще охопити інформаційний WWW-простір і подати його користувачам в зручному вигляді.

Принцип роботи з пошуковими покажчиками засновано на *ключових словах*. Розшуковуючи інформацію з деякої теми, клієнт повинен підібрати ключові слова, які описують цю тему, і задати їх пошуковій системі як

завдання. Відвідувачам такого пошукового сервера надається форма для введення ключового слова (слів) або фрази (Рис.4).

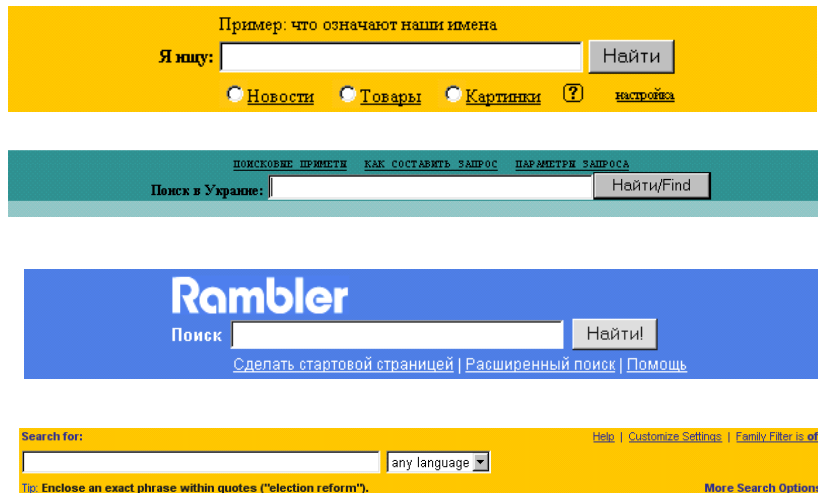


Рис.4

Після натиснення клавішею мишки на кнопці *Найти*, яка розташована поруч з полем введення, ключова фраза пересилається на сервер, опрацьовується, потім пошукова система знаходить в своїх базах даних (вони називаються *індексами* або *покажчиками*) адреси Web-ресурсів, в яких присутні ключові слова, і видає клієнту Web-сторінку з посиланнями на ці ресурси (Рис. 5).

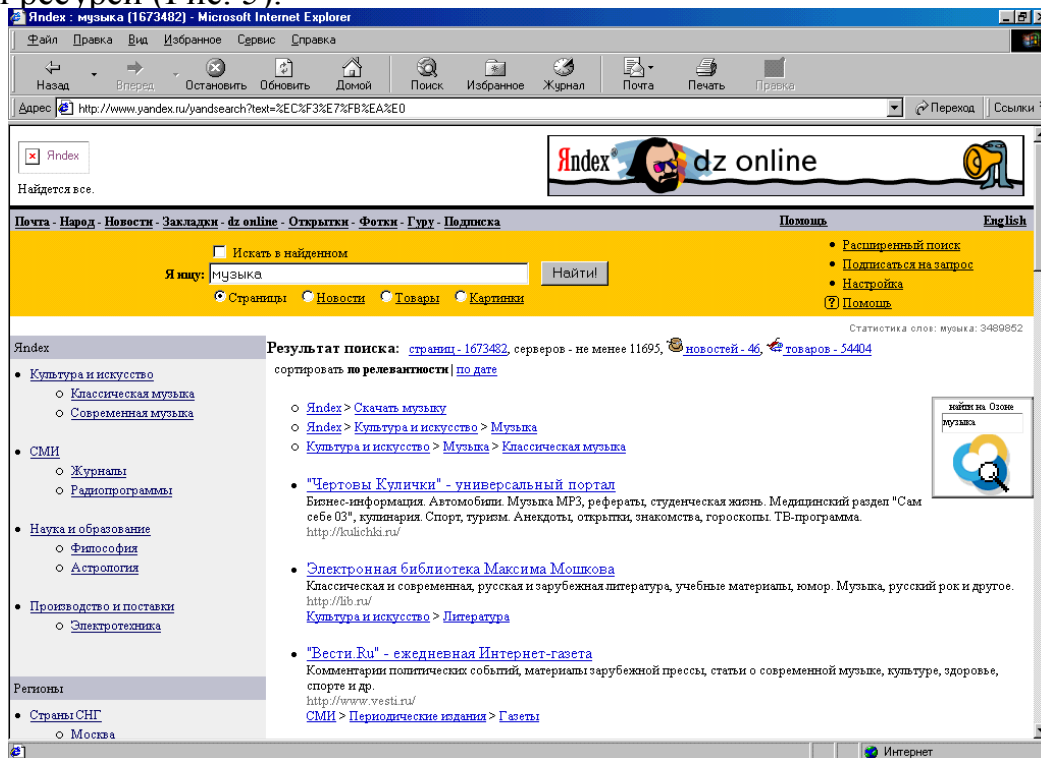


Рис.5

Самий великий покажчик, що наближається до 300 млн. Web-документів, має пошукова система Fast Search (www.alltheweb.com), а самий зручний для наукових досліджень покажчик має система Northern Light (www.northernlight.com).

У Росії пошукові покажчики розвинені дуже добре. Можна назвати три найбільших системи: Апорт 2000 (www.aport.ru), Яндекс (www.yandex.ru) і Ремблер (www.rambler.ru). В Україні – МЕТА (www.meta-ukraine.com).

Вчителю доцільно порівняти індексні бази даних і пошукові каталоги, акцентувати увагу учнів на позитивних характеристиках кожної із систем пошуку та відмітити недоліки тематичних каталогів і індексованих баз даних.

Пошукові покажчики створюються автоматичними засобами при мінімумі ручної праці. З одного боку це дає можливість порівняно швидко опрацьовувати великі масиви Web-документів і створювати індекси розміром в сотні мільйонів Web-сторінок. З іншого боку, ніхто не може гарантувати, що наявність в документі певних слів означає належність документа до конкретної теми. Тому точність пошуку інформації в базі даних, заповнених автоматичними засобами, залишає бажати кращого. Пошукові каталоги, як правило, позбавлені таких недоліків індексних баз даних, як:

- неадекватність посилань (посилання веде до документу, який не відповідає темі пошуку);
- неактуальність посилань (посилання вказує на документ, якого вже немає в наявності);
- дублювання посилань (кілька посилань ведуть до однакових документів, що зберігаються в різних місцях, або до одного документа, який проіндексовано в різні дати).

Крім того, важливо підкреслити, що для оцінювання якості пошуку використовують два параметри: *коефіцієнт обхвату* і *коефіцієнт попадання*. *Коефіцієнт обхвату* показує, яку частину загальносвітових ресурсів WWW відображає пошукова система в своїх базах даних. Для пошукових каталогів цей коефіцієнт надзвичайно низький (частки процента). Для пошукових покажчиків він порівняно високий (десятки процентів).

Коефіцієнт-попадання показує, яка частина посилань, що повертаються пошуковою системою, дійсно відповідає запитам клієнта. Для пошукових каталогів цей коефіцієнт дуже високий, оскільки ці каталоги складають люди. Для пошукових покажчиків цей коефіцієнт достатньо низький.

Однак, як свідчить практика, пошукові покажчики дозволяють провести найбільш глибокий пошук в рамках заданої теми. Слід пояснити учням, що робота пошукового покажчика проводиться в три етапи. На першому етапі сканується інформаційний простір і збираються копії Web-ресурсів. На другому етапі бази даних, складені за результатами сканування, перетворюються так, щоб по них можна було проводити прискорений пошук. На третьому етапі пошуковий покажчик приймає запит від клієнта, проводить пошук в своїх базах даних і видає Web-сторінку оформлених результатів пошуку.

Далі доцільно проаналізувати особливості кожного з етапів. Зрозуміло, що цей матеріал не є обов'язковим, але він має світоглядне значення, саме

тому слід зупинитися на поясненні роботи кожного з етапів організації пошукової роботи індексних баз даних.

Для збирання відомостей про ресурси WWW пошуковим покажчиком використовуються спеціальні програмні засоби, які називають пошуковими роботами або павуками, черв'яками, гусеницями, краулерами і т. п. Як правило, кожний пошуковий покажчик має свою власну програму, побудовану на патентованому алгоритмі. Загальний принцип дії пошукових роботів полягає в тому, що вони починають перегляд Мережі з деякої заданої адреси Web-сторінки, копіюють знайдений документ на сервер своєї системи, переглядають його, знаходять в ньому всі гіперсилки, переходять по них, знаходять нові документи, копіюють їх, виявляють в них гіперсилки, знову виконують перехід і т. ін. "Павуки" дуже подібні за принципами свого функціонування комп'ютерним вірусам, можуть "саморозмножуватися" та розсилати свої копії за силками, що вказані на сторінках, які аналізуються, а після виконання поставленого перед ними завдання самоліквідуються.

Проводити по зібраним копіям Web-сторінок пошук інформації, яка була замовлена клієнтом, дуже незручно, оскільки це займає достатньо велику кількість часу. Тому зібрані дані проходять попереднє опрацювання, яке називається *індексацією*. Метою індексації є одержання *індексного файлу*, за допомогою якого запит клієнта можна швидко опрацювати. Кожна пошукова система проводить таку індексацію за своїми алгоритмами, які складають комерційну таємницю системи.

Можна навести один із прикладів такого пошукового процесу, коли при індексації формується файл із словником, до якого входять всі слова, виявлені на Web-сторінках, скопійованих з WWW-простору. Поруч з кожним словом проставляються номери Web-документів, в яких це слово було виявлено. Часто також може проставлятися і *вага слова* – це число, яке показує, як часто дане слово зустрілося в даному Web-документі. Додатково додається ще адресна таблиця, в якій для кожного номера Web-документа вказана його URL-адреса .

На наступному етапі *ранжування* пошукова система намагається за своїми алгоритмами визначити цінність кожного із знайдених ресурсів і впорядкувати їх так, щоб найбільш цінні (з точки зору системи) розташовувалися на початку списку. Для цього кожному ресурсу дається деяка умовна оцінка. При її розрахунку окремим Web-ресурсам можуть нараховуватися "призові" бали, а деяким – "штрафні".

"Призові" бали призначаються, наприклад, за наступні чинники:

- якщо ключові слова, які використовувались клієнтом в запиті, зустрічаються в заголовку Web-сторінки – це свідчить про те, що Web-сторінка дійсно присвячена тій темі, яку досліджує клієнт;
- якщо ключові слова, які виявлено в документі, мають шрифт збільшеного розміру – це говорить про те, що вони входять до заголовків розділів всередині документів;

- якщо ключові слова неодноразово повторюються всередині документа (особливо якщо повторення відбувається в перших п'яти-восьми абзацах тексту);
- якщо ключові слова входять до так званого *альтернативного тексту* (це текст, який підміняє ілюстрації, якщо їх відображення на екрані з деяких причин відключено); наявність ключових слів в альтернативному тексті свідчить про те, що автор документа надає їм особливе значення;
- якщо існують посилання з інших Web-сторінок тощо.
 - “Штрафні” бали можуть призначатися за те, що:
 - ключові слова повторюються підряд або регулярно з постійною періодичністю – це свідчить про те, що автор примусово намагається підвищити рейтинг своєї сторінки;
 - ключові слова зроблені невидимими або непомітними, наприклад набрані кольором, співпадаючим з кольором фону (чорним по чорному, білим по білому) або набрані шрифтом мінімального розміру. Це свідчить про те, що автор використав ці слова в документі не для читача, а для того, щоб впливати на пошукову роботу.

Далі необхідно звернути увагу учнів на основні характеристики індексних баз даних (Рис.6) та прокоментувати кожну з них.

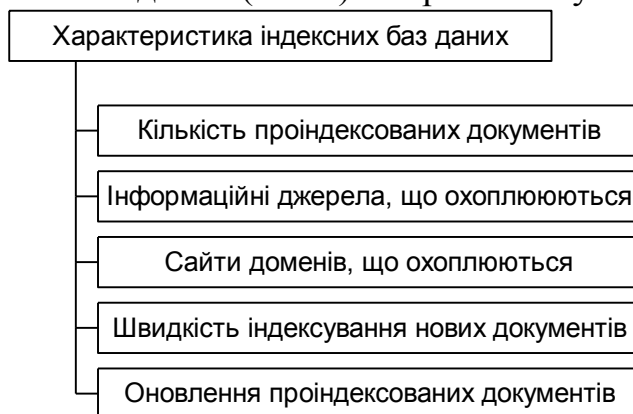


Рис. 6

Перше ніж працювати з конкретною пошуковою системою слід розібрати з учнями принципи та прийоми пошуку інформації. Учні повинні засвоїти, що для пошуку за одним ключовим словом необхідно ввести це слово до поля введення і клацнути на кнопці *Найти (Search)*. У різних пошукових системах ця кнопка може називатися по-різному, але суть її від цього не міняється.

Як правило, пошук за одним словом приводить до формування величезних списків Web-сторінок, на яких воно зустрічається. Знайти в цьому списку потрібні ресурси не просто, і тому пошук за одним словом малоефективний. Можна для прикладу звернутися із таким запитом до однієї з пошукових машин та проаналізувати одержані результати: визначити кількість знайдених Web-документів та Web-вузлів. Тоді учні самостійно приходять до висновку, що набагато ефективнішим є пошук за

кількома словами, але тут важливу роль грає правило, яке вказує пошуковій системі, як опрацювати групу слів. Можна навести конкретні приклади, коли користувача можуть цікавити документи, що містять І перше слово, І друге одночасно або слід знайти документи, в яких ці слова зустрічаються поруч або недалеко один від одного або можуть цікавити документи, в яких зустрічається АБО перше слово, АБО друге, АБО обидва слова разом.

Таким чином, учні повинні дійти до висновку, що для ефективного пошуку за кільком ключовим словами потрібні спеціальні команди, які дозволяють зв'язати окремі слова між собою. Ці команди в пошукових системах утворюють спеціальну, як правило, нескладну мову запитів, яка містить кілька команд.

Важливо підкреслити, що кожний пошуковий покажчик використовує свою власну мову запитів, тому при використанні різних пошукових систем треба знати особливості кожної. Але є загальний принцип, згідно з яким всі команди можна розділити на три групи: *простого пошуку*, *розширеного пошуку* і *спеціального пошуку*. В режимі простого пошуку запити створюються нескладними методами, але вони, як правило, призводять до широких результатів, з яких може бути важко вибрати конкретне.

Команди розширеного пошуку дозволяють досить точно описати необхідний документ.

Команди спеціального пошуку призначені не для пошуку документів за їх змістом, а, наприклад, для пошуку Web-вузлів за їх назвами, за фрагментами їх адреси, за адресами посилань, які зустрічаються на їх Web-сторінках і т. п.

Важливо, щоб учні зрозуміли, що при пошуку в Інтернет важливі дві складові: *повнота пошуку* (тобто намагання не пропустити деяку важливу інформацію – *нічого не загубити*) і *точність* (тобто відсутність в результатах пошуку зайвої інформації – *не знайти нічого зайвого*). Як правило, обидві ці складові називають спільним словом *релевантність*, яка і передбачає максимально змістовну відповідність результатів пошуку вказаному запиту. Іншими словами, релевантність – це *відповідність відповіді запиту*. Релевантність також можна розглядати і як спосіб впорядкування знайдених за запитом документів: чим більше документ відповідає запиту, тим ближче до початку списку повинна знаходитися його адреса. Для досягнення цього результату пошуковою системою враховуються такі параметри, як кількість знайдених в тексті ключових слів, "*контрастність*" слова (його відносна частота зустрічаємості для даного документу), відстань між словами, розташування слова в документі і його підрозділах і т.п.

При цьому слід враховувати, що *релевантність* – *якісна міра*. Відповідність відповіді ІПС інтересам користувача залежить не лише від властивостей даної пошукової системи, але і від того, наскільки вдало сформульовано запит, тобто наскільки правильно останній відображає інтереси користувача.

Для оптимальної і швидкої роботи з пошуковими системами індексних баз даних існують певні правила написання запитів. Ретельний перелік таких правил для конкретного пошукового серверу можна, як правило, знайти на самому сервері за силками "*Помощь*", "*Подказка*", "*Правила составления запроса*" тощо. Учнів можна познайомити з правилами, які частіше використовуються для ПС "Апорт", "Яндекс" або будь-якої іншої індексної бази даних.

1. Незалежно від того, яка форма слова стоїть у запиті, пошук враховує всі словоформи за правилами російської мови. Наприклад, якщо задано запит "идти", то в результаті пошука будуть знайдені силки на документи, які містять слова "идет", "шел", "шла" і т.д. На запит "окно" буде видана інформація, яка містить і слово "окон", а на запит "отзывали" – документи, які містять слово "отозвали". На жаль україномовної частини Інтернет цей процес не торкнувся, хоча на пошуковому сервері META проводяться роботи щодо коректного врахування при запиті особливостей української та російської мов та можливої омонімії мови (серед результатів за запитом російською мовою "лист каштана" крім російського омоніму отримуються ще й україномовні документи з цим словом в значенні "письмо").

2. Якщо в запиті слово набрано з великої літери, то всі ПС будуть враховувати реєстр, в якому набрані літери (маленька/велика.)

3. Якщо запит являє собою одно чи кілька слів, то на ньому будуть знайдені документи, в яких зустрічаються всі слова запиту, крім обмеженої кількості слів (сполучники, прийменники і т.д.), які в запиті ігноруються, тому що не несуть змістовного навантаження). Наприклад, по запиту "яблука на снігу" будуть знайдені всі документи, в яких зустрічаються одночасно два слова. Де само в межах документа розташовані слова, в якій граматичній формі вони знаходяться – не важливо. Слово "на", яке є прийменником, ігнорується. Тому вказаний запит можна було написати і так: "сніг на яблуках" – результат пошуку буде таким самим.

4. Зірочка в кінці слова символізує будь-яку кількість будь-яких літер. Наприклад, за запитом "вер*" будуть знайдені документи, які містять будь-яке із слів "по вере, верес, Вера, вересень", в будь-якій граматичній формі.

5. Лапки вказують пошуковій системі, що задані в запиті слова відносяться до ключовою фрази, яку слід шукати в тексті повністю. Наприклад, за запитом "експертна система" будуть знайдені всі документи, де зустрічаються вказані словосполучення або близькі до нього фрази.

6. При складанні запита можна комбінувати ключові слова за допомогою логічних операторів (І, АБО, НЕ).

Далі слід розглянути прийоми формування запитів на прикладі конкретної системи. При цьому після введення звичайного простого запиту

для пошуку одного слова необхідно ретельно розібрати, як працювати з результатами пошуку.

Учні повинні з'ясувати, що після введення з клавіатури запиту, на екран виводиться список посилань на документи, які точно відповідають запиту. На такій сторінці користувач може побачити спеціальні позначення і силки, які призначені полегшати перегляд знайдених сторінок. Слід звернути увагу учнів на такі елементи сторінки з результатами пошуку (Рис.7 – для спрощення будемо розглядати пошукову систему Яндекс):

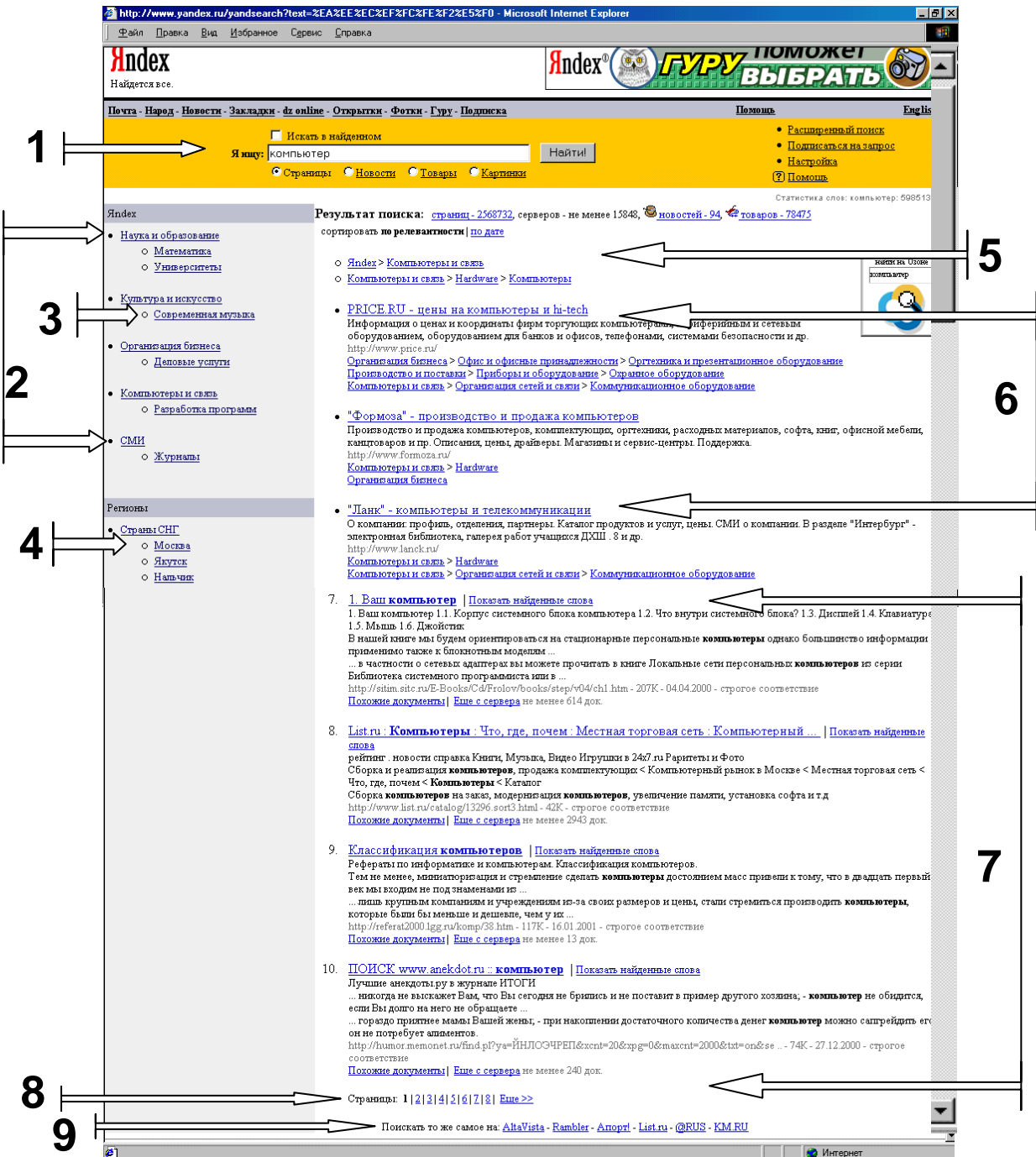


Рис. 7

1. *Пошукова форма*, головним елементом якої є рядок запиту, до якого з клавіатури користувач вводить простий або складений запит. При цьому додатково можна уточнити, що саме слід шукати: сторінки, новини, товари чи картинки. Здійснити подальший пошук в результатах попереднього дозволяє параметр “Искать в найденном”.
2. *Тематичний спектр* результатів пошуку. Показуються тільки ті теми каталогу, в яких є знайдені документи. Список тем впорядкований за релевантністю. Натиснення на посилання призводить до обмеження пошуку по сторінках вибраної теми.
3. *Теми більш глибокого рівня*, які найкращим чином відповідають запиту. Натиснення на посилання призводить до обмеження пошуку за сторінками вибраної теми.
4. *Регіони пошуку* – можна обмежити пошук сайтами вказаного регіону.
5. *Пошук по каталогу* – опис ресурсів, що містять слова запиту. Натиснення на посилання веде до переходу у відповідний розділ каталогу (выводиться на екран тільки при точній відповідності запиту).
6. *Пошук по каталогу* – опис ресурсів, що містять слова запиту. Натиснення на посилання веде до переходу на ресурс (выводиться тільки при точній відповідності запиту).
7. *Результати пошуку по Web-простору*. Заголовок документу – це є посилання на ресурс. За допомогою силки *Показать найденные слова* можна здійснити перехід на "підсвічений" документ, в якому відповідність запиту (слова, які входять до запиту) виділена стрілочками та кольором (Рис. 8).

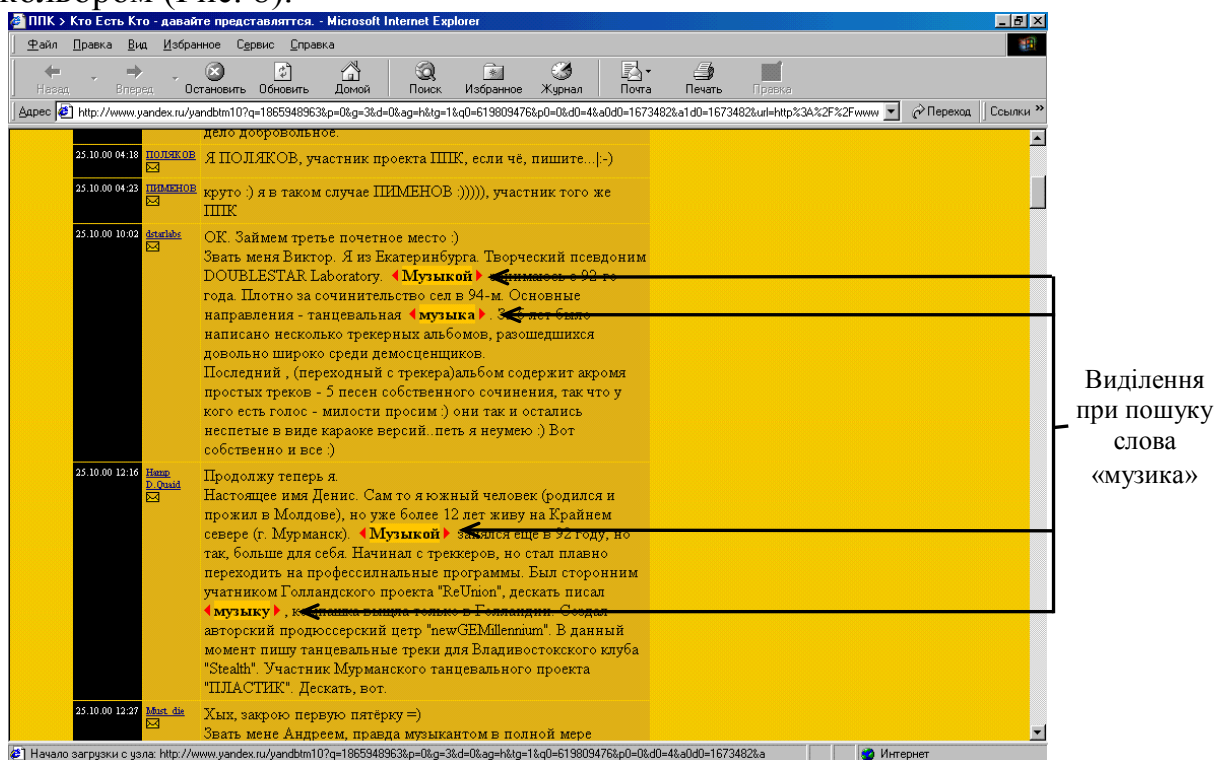


Рис.8

8. Посилання на інші служби пошукової системи або на інші пошукові системи.

9. Перелік сторінок, які містять список знайдених за запитом силок.

Далі доцільно звернутися до однієї з пошукових систем, наприклад, Yandex, та запропонувати учням ввести з клавіатури до пошукової форми простий запит, який складається з одного слова, після чого слід запропонувати учням частково-пошукову роботу по вивченню додаткових посилань, які знаходяться на сторінці, що є результатом пошуку. Робота полягає в тому, щоб шляхом методу випробувань з'ясувати, для чого на сторінках з результатами пошуку за запитом використовуються силки: *"результаты поиска (страницы, сервера), сортировка (по релевантности, по дате), показать найденные слова, похожие документы, еще с сервера"*. Для організації такої роботи вчитель може задавати учням додаткові питання, підказувати, на що в кожному конкретному випадку із перерахованих вище слід звертати увагу: порядок відображення сторінок, зміст документів, додаткові символи, які висвітлюються на екрані тощо.

Учні повинні під керівництвом вчителя дійти до висновків:

1. Якщо в результаті пошуку було знайдено занадто багато документів, і користувач має бажання скоротити такий список, то для цього використовується прапорець *"Искать в найденном"*.

2. При одержанні результатів пошуку на запит, що складається з кількох слів, користувачеві необхідно звертати увагу на статистику результатів пошуку, яка відображається праворуч від рядком пошуку, де виводиться статистика слів запиту і знайдених документів.

Статистика слова видається завжди по всій базі і не залежить від області пошуку. Якщо слово запиту жодного разу не зустрічалось в базі, воно виділяється червоним кольором. В цьому користувачеві рекомендується перевірити його правопис.

Статистика документів складається з двох чисел – знайдені документи (сторінки) і сайти (сервери). Якщо вибрати силку про сторінки, то на екрані послідовно сторінка за сторінкою (по 10 документів на кожній екранній сторінці) відображаються анотації та адреси кожної з них (Рис.7). Якщо вибрати силку про сервери, то на екрані відображається список, що складається із одного знайденого документу з кожного сервера.

3. Всі знайдені за запитом документи впорядковуються при відображенні на сторінці з результатами пошуку. За замовчуванням результати пошуку впорядковуються за релевантністю і згруповані по серверам. Знайдені документи можуть мати помітки *"совпадение фразы"*, *"строгое соответствие"* або *"нестрогое соответствие"* (по степеню зменшення релевантності) або *"нестрогие"* документи відділені від інших горизонтальною лінією. Активний порядок впорядкування і групування виділяється жирним шрифтом, альтернатива – силкою. При групуванні по серверам у кожного знайденого документа існують силки *"Все с сервера"* (*Еще*

с сервера) – запит на одержання всіх сторінок цього сервера, які відповідають запиту.

4. При впорядкуванні по серверам кожний знайдений документ має такий вигляд:

[Заголовок документа \(ссылка на нього\)](#)

Опис документа (Meta description, якщо існує, інакше - перші 256 кілобайт тексту)

URL-адреса документа - його розмір - ступень відповідності запиту

[Ссылка на розмічений документ](#) | [Ссылка на запит схожих документів](#) | [Ссылка на інші документи з того ж самого сервера, які відповідають даному запиту](#)

Наприклад,

[Тенета-Ринет'2000: Стихотворения](#) | [Показать найденные Слова](#)

Тенета-Ринет'2000: Стихотворения КОНКУРСНАЯ РАБОТА:

<http://teneta.rinet.ru/2000/stihi/gb955460406453260.html> - 24К - совпадение фразы

[Похожие документы](#) | [Еще с сервера](#) около 16 док.

При впорядкуванні по сторінкам знайдений документ має інший вигляд: в останньому рядку відсутня силка "Еще с сервера...", наприклад,

[Поэтическая обсуждалка](#) | [Показать найденные Слова](#)

Участники Курилка Прозаическая обсуждалка Обсуждаемые работы Архивы

<http://www.lito.spb.ru/archive/poet/lukin.html> - 35К - совпадение фразы

[Похожие документы](#)

5. Якщо один із знайдених документів ближче до потрібної теми, ніж всі останні, то слід використати силку "*Найти похожие документы*". При цьому за спеціальним алгоритмом пошукової системи сторінка проаналізується і будуть знайдені документи, що за змістом схожі на вказані користувачем.

6. В кінці кожної сторінки з результатами пошуку відображається силка "*Поискать тоже самое на* : [AltaVista](#) - [Rambler](#) - [Апорт!](#) - [List.ru](#) - [@RUS](#) - [K&M.](#), яка може стати корисною, якщо пошуковий запит не призвів до бажаних результатів. Тоді запит користувача передається до інших пошукових систем і каталогів.

Далі можна запропонувати учням звернутися до іншої пошукової системи з індексною базою даних та здійснивши аналогічний пошук за простим запитом, проаналізувати одержану результуючу сторінку, знайти відповідні її елементи. Наприклад, за запитом "компьютер" на пошуковиз системах META та Rambler будуть одержані відповідно такі результати (Рис.9, Рис. 10)

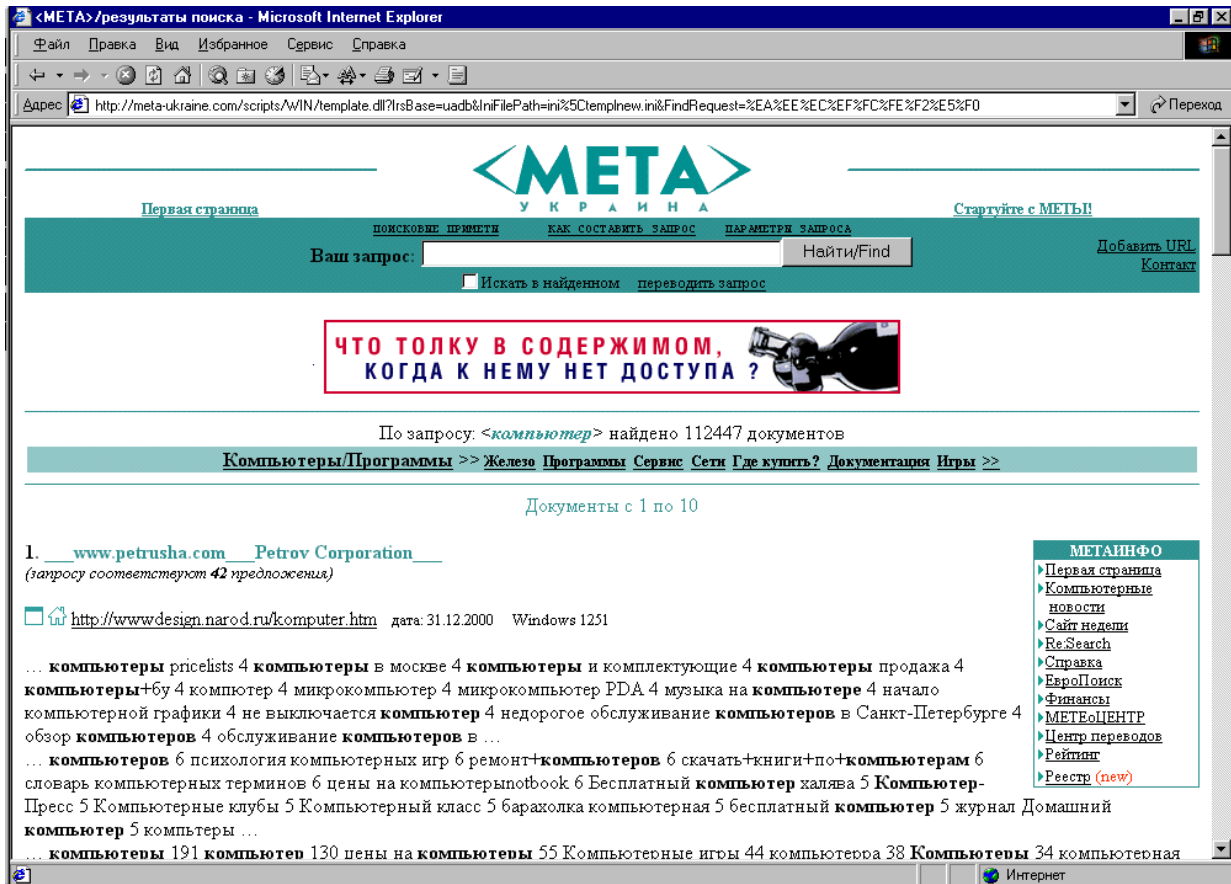


Рис.9

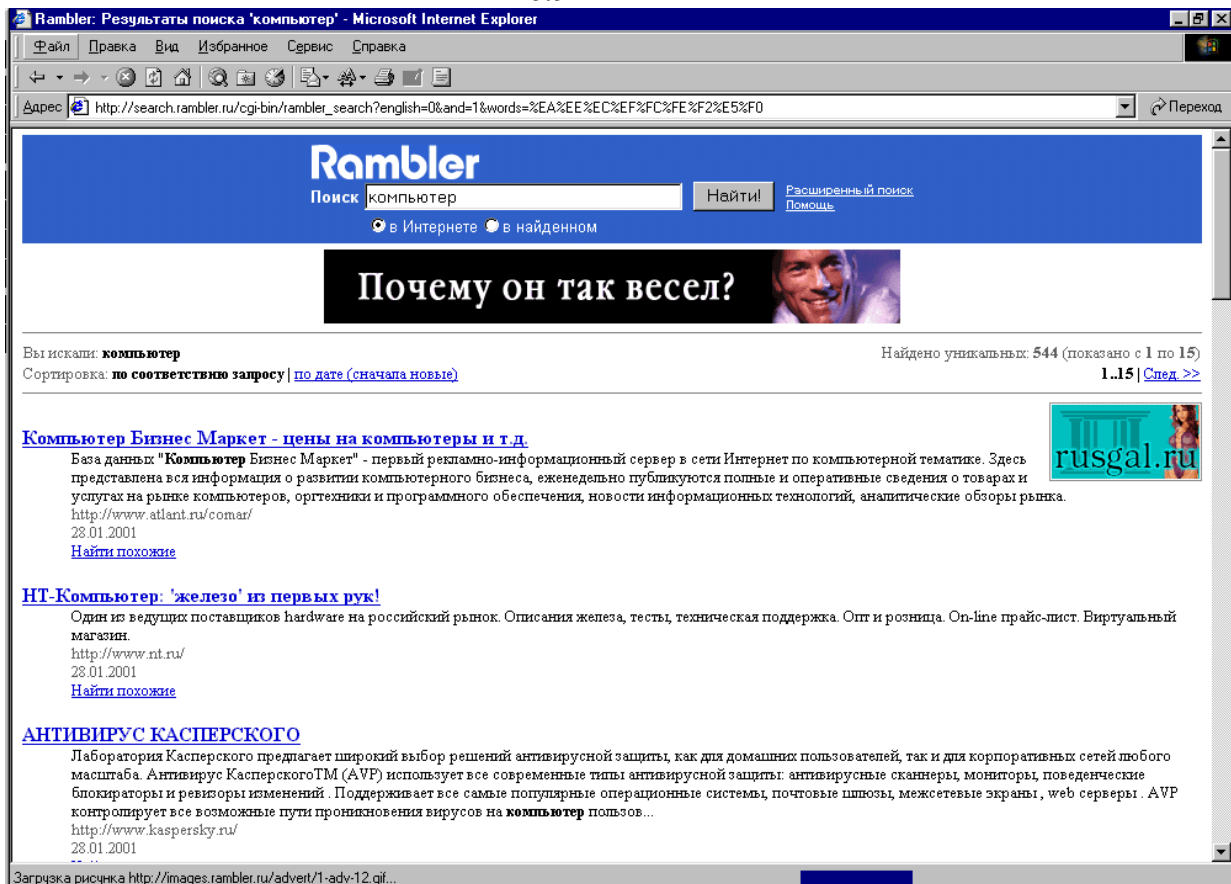


Рис.10

Практика свідчить, що вчителю доцільно розглянути з учнями фактори, які впливають на релевантність знайдених документів:

- Різні частини Web-сторінки, як і будь-якого документу, не є рівноцінними за своїм інформаційним навантаженням. Ті самі слова, але вживані в різних частинах документу, мають різний ступінь значимості. Так, слово з заголовка сторінки “важить” більше, ніж слово із основного тексту.
- При використанні в запиті кількох ключових слів, у більшості випадків буде справедливим залежність релевантності документу від ступеня близькості цих слів у тексті.
- Слово, що зустрічається кілька разів на сторінці більш значимо, ніж таке, яке зустрічається тільки раз. У більшості випадків саме цей фактор є вирішальним у визначенні рейтингу документу.
- За допомогою використання спеціальних операторів мови запитів у деяких пошукових машинах можна збільшити “вагу” слова, і документи, які його містять, будуть вважатися більш релевантними.
- При введенні списку результатів пошуковими машинами враховується також кількість зовнішніх сілок на сторінку: чим більше сілок на дану сторінку, тим вище її розташування в списку результатів пошуку; також враховується присутність знайденої сторінки в тематичному Web-каталозі, її рейтинг, що підтримується даною системою.
- При пошуку по датах створення документів, результати виводяться в довільному порядку або упорядковуються саме по датах.

Далі для вивчення мови запиту конкретної індексної бази даних можна запропонувати виконати такі завдання:

Знайти Web-сторінки, що містять певну інформацію, за запитами, які записані в таблиці у відповідному стовпчику. Визначити кількість посилань на знайдені документи і кількість сайтів, на яких знаходяться знайдені Web-сторінки, переглянути знайдені документи з виділеними словами, що входять до запитів, заповнити таблицю та за результатами пошуку зробити відповідні висновки.

Таблиця

№	Запит (ключове слово або вираз для пошуку)	Регіон пошуку	Кількість знайдених документів	Кількість сайтів, на яких знаходяться знайдені web-сторінки	Сортування (за релевантністю/датою)	Виведення вмісту по (серверам /сторінкам)	Виділені слова в знайденому документі
1.	music	Рунет			релев.	стр.	
2.	Музыка	Рунет			релев.	стр.	
3.	МУЗЫКА	Рунет			релев.	стр.	

4.	музика	Украина			релев.	стр.	
5.	!музика	Украина			релев.	стр.	
6.	!музыка	Рунет			релев.	стр.	
7.	"музыка"	Рунет			релев.	стр.	
8.	современная музыка	Рунет			релев.	стр.	
9.	современная электронная музыка	Рунет			релев.	стр.	
10.	"современная музыка"	Рунет			релев.	стр.	
11.	"сучасна музика"	Украина			релев.	стр.	

Після виконання завдань необхідно зробити висновки про побудову простих запитів, врахування системою морфологічного аналізу слів та побудувати схему проведення такого аналізу ключових слів запиту (Рис.11).

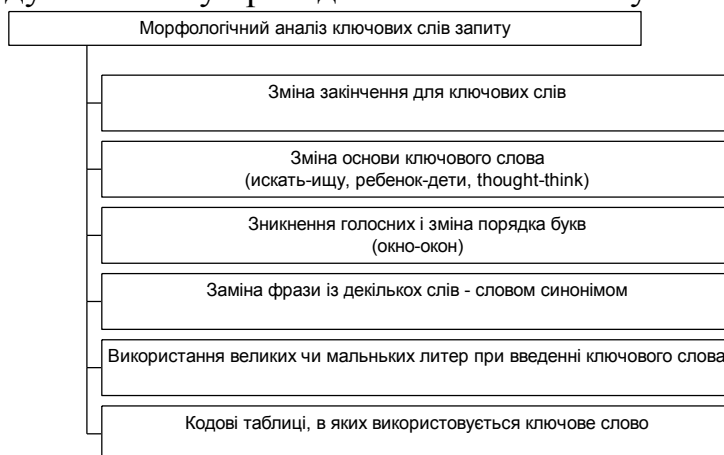


Рис. 11.

Далі доцільно ретельно розібрати з учнями принципи побудови в будь-якій пошуковій системі з індексною базою даних складеного запиту. Для цього можна запропонувати учням ознаки для введенні спеціальних операторів мови запитів:

- 1.Наявність кількох ключових слів в реченні.
- 2.Наявність кількох ключових слів в документі.
- 3.Обов'язкове входження кількох ключових слів до речення.
- 4.Обов'язкове входження кількох ключових слів до документу.
- 5.Входження одного слова із набору слів (для пошуку синонімів).
- 6.Відсутність ключового слова в реченні.
- 7.Пошук слів, що розташовані в реченні на певній відстані одне від одного.
- 8.Входження фрази до документу.
- 9.Входження цитати до документу.

Можна для прикладу розглянути приклади використання таких операторів в мовах запитів різних пошукових систем:

<i>Логічні оператори</i>	<i>Яндекс</i>	<i>МЕТА</i>
И (синоніми: AND, &, +)	За замовчуванням: за будь-яким із запитів “метод проектв” або “метод И проектв” будуть знайдені документи, які містять обидва слова в одному реченні.	За замовчуванням: за будь-яким із запитів “нужная” або “метод И проектв” будуть знайдені документи, які містять обидва слова в одному реченні.
ИЛИ (синоніми: OR, , „кома”)	Дозволяє шукати документи, які містять хоча б одне із вказаних слів. Запит “книга підручник” або “книга, підручник” задає пошук документів, які містять в одному абзаці хоча б одне із слів.	Дозволяє шукати документи, які містять хоча б одне із вказаних слів або обидва слова одночасно. Запит “книга или підручник” задає пошук документів, які містять в одному абзаці хоча б одне із слів.
НЕ (синоніми: NOT, ~, „мінус”)	Знак тільди “~” дозволяє знайти документи, які містять перше слово, але які не містять друге. За запитом “спорт ~теніс” будуть знайдені документи, які містять слово “спорт”, але не містять слово “теніс”.	За запитом “напиток не кофе” будуть знайдені документи, які містять слово “напиток”, але не містять слово “кофе”.
Круглі дужки	Круглі дужки задають порядок дії логічних операторів. За запитом “(быстрый качественный) поиск” будуть знайдені документи, де зустрічаються одночасно слова “быстрый” і «поиск», або «качественный» і «поиск».	Аналогічно
Обмеження відстані між словами	Відстань між сусідніми словами дорівнює 1, а відстань між сусідніми словами, які стоять в зворотньому порядку, дорівнює - 1. Якщо між двома словами вставлено знак “/”, за яким слідує деяке число, то це означає вимогу, щоб відстань між ними не перевищувала вказане.	За запитом сл5(солнце ветер вода) будуть знайдені документи, в яких між словами “солнце”, “ветер”, “вода” стоїть не більше двох інших слів (тобто загальна кількість слів у фрагменті не більше 5).

Потім вчитель разом з учнями узагальнює міркування учнів та вводить основні команди мови запитів конкретної пошукової системи та пропонує проаналізувати призначення прикладів запитів, записаних мовою конкретної пошукової системи. Наприклад, для пошукової системи Яндекс визначити результати пошуку за запитом, спочатку теоретично, а потім підтвердити свою запропоновану версію практично:

- *!проект*
- *Проект*
- *"проект"*
- *телекомунікаційний проект*
- *проект ~бізнес*
- *(телекомунікаційний проект) ~бізнес*
- *інформація | технологія*

- навчання гітара +Київ ~(приватне | недержавне)
- віруси ~-бактерії
- фото| Фотографія| Знімок| Фотозображення
- (технологія | виготовлення) (іграшки +м'які)
- права людини ~(США |Америка)
- путеводитель по праже ~-(агенство | тур)
- поставщики /2 модем
- музыкальное /(-2 4) образование

Далі можна учням практичну роботу по пошуку інформації, коли запиті сформовані та записані вчителем, а учням необхідно заповнити таблицю за результатами пошуку та зробити висновки:

Таблиця 2

№	Запит (ключове слово або вираз для пошуку)	Регіон пошуку	Кількість знайдених документів	Кількість сайтів, на яких знаходяться знайдені web-сторінки	Сортування (за релевантністю/ датою)	Виведення вмісту по (серверам /сторінкам)	Виділені слова в знайденому документі
1	сучасна /+1 музика	Україна					
2	!сучасна класична музика	Україна			дате	стр.	
3	+сучасна ~рок музика	Україна			релев.	серверам	
4	доход прибыль	Рунет			релев.	стр.	
5	(доход прибыль) (бизнес "предпринимательской деятельности")	Рунет			релев.	стр.	
6	бизнес ~(торговый игровой)	Рунет			дате.	серверам	

Далі необхідно навчити учнів планувати пошук, що передбачає вибір пошукових систем і інструментів, послідовність їх використання в залежності від результативності пошуку.

Можна запропонувати учнями алгоритм побудови запиту:

1. Проведення лексичного і морфологічного аналізу інформації, яку потрібно знайти:

a) Познайомитись із інформаційними поняттями в довідниках, енциклопедіях. (Наприклад, можна використати сайт Кирила і Мефодія – www.km.ru).

b) Виділити ключові слова, що несуть інформаційне навантаження.

c) Знайти синоніми, антоніми, фрази із кількох слів.

d) Перевести слова на інші мови.

2. Технологічний етап. Вибір інформаційного поля, регіону пошуку, пошукової системи.

3. Побудова запиту на мові пошукової системи.

4. Аналіз результату пошуку.

5. Вдосконалення запиту :

1) Використання операторів з дією на максимальне звуження кількості знайдених документів,

2) Використання операторів з напрямленою дією на релевантність отриманих документів.

6. При невдалому пошуку повернутися до другого кроку алгоритму.

Після ознайомлення учнів з алгоритмом пошуку на основі роботи з конкретною пошуковою системою необхідно розглянути з учнями проблеми, які виникають при його організації:

повнота охоплення інформаційного простору;

достовірність знайденої інформації;

планування проведення пошуку серед інформаційних сервісів.

Необхідно пояснити учням, що пошук вважається *повним*, якщо він охопить повний доменний Web-простір. Тому доцільно використовувати при розв'язанні пошукових задач кілька пошукових систем.

Пошукові системи надають результати в залежності від поставленого запиту. Такі результати бувають позитивні або негативні, коли не вдається знайти жодного документу на поставлене запитання. Але учням слід пояснити, що завжди потрібно перевіряти отримані результати.

Доцільно розглянути з учнями методику оцінювання достовірності знайденої інформації:

1) Отримання однакової інформації одночасно з кількох першоджерел.

2) Перевірка знайденого матеріалу, встановлення частоти його використання іншими користувачами. При цьому використовуються рейтингові механізми перевірки.

3) З'ясування статусу документа і рейтингу Web-сайта, на якому він знаходиться, засобами пошукових систем.

4) Пошук інформації про компетентність і статус автора матеріалу засобами пошукових систем.

По закінченню вивчення теоретичного матеріалу про побудову простих та складених запитів доцільно провести лабораторну роботу, яка може містити такі завдання:

1. Знайти інформацію одночасно на пошукових системах *yandex i meta-ukraine* та порівняти одержані результати:

- 1) Історичний опис Києва.
- 2) Інформацію про комп'ютерні віруси.
- 3) Інформацію про мультимедійні програми.
- 4) Реферати чи курсові роботи.
- 5) Правила участі в дистанційній олімпіаді з права.
- 6) Список готелів у м. Києві.
- 7) Адреси підприємств Київської області.
- 8) Інформацію про посольства США, Великобританії, Франції в Києві.
- 9) Список вищих навчальних закладів, що готують фахівців з прикладної інформатики.
- 10) Рекламні агентства в Україні.
- 11) Адреси Web-сторінок пошукових машин.
- 12) Інформацію про автомобіль Ford на сайтах України та Росії окремо.
- 13) Картинки, які в імені файла містять слово *tar*.

2. Використовуючи пошукові системи, визначити:

- які спектаклі будуть йти сьогодні ввечері у вашому місті;
- яка погода буде завтра у вашому місті;
- які нові компакт-диски випустила фірма TeachPro;
- хто такий Томас Джефферсон;
- хто такий модератор;
- що таке Юкка;
- коли буде сонячне затемнення;
- як приготувати піцу;
- офіційний сайт своєї улюбленої групи чи виконавця;
- у кого із відомих музикантів були гастролі в Києві в цьому році;
- яка виставка проходить в Ермітажу в цьому місяці;
- в якому році було побудовано Андріївську церкву в Києві;
- де зараз працює А. Шевченко;
- як доїхати із Києва до Львова.

На закінчення можна ознайомити учнів з третьою групою запитів – *спеціальними*, для побудови яких використовуються оператори області пошуку. Учням слід пояснити, що за замовчуванням пошук проводиться по всіх полях гіпертекстового документу, які враховуються при його індексації програмою-павуком. Використання ж операторів області пошуку дозволяє обмежити пошук тільки основним текстом документу, або його заголовком, полями коментарів, мета-дескрипторами, alt-тегами графічних файлів,

ссылками з Web-сторінками, і тому подібне. Можна ознайомити учнів з прикладами таких операторів:

Текст (text) – все те, що безпосередньо видно при перегляді сторінки в браузері.

Заголовок сторінки (title) – заголовок Web-сторінки, якій відображається в заголовку вікна браузера.

Заголовки (header) – заголовки, які виділяються в основному тексті сторінки за допомогою тегів <H>.

ALT-тегі – коментарі до вбудованих зображень.

META-тегі – поля, які стисло описують вміст сторінки.

Коментарі (comment) – коментарі Web-майстра до HTML-коду сторінки.

Текст силки (anchor) – текст, який описує гіперсилку (власне, той текст, що бачить користувач).

Силка (link) – “підводна” частина гіперсилки, яка вказує на іншу Web-сторінку або файл.

Адреса сторінки (URL-адреса) – повний шлях до сторінки, включаючи ім’я серверу, каталогу і ім’я файлу.

Вбудовані програми – пошук сторінок з вбудованими JAVA-аплетами або VB- і JAVA-скриптами.

Вивчення спеціальних операторів мови запиту пошукових систем доцільно закріпити за допомогою виконання конкретних вправ.

Звершуючи розгляд пошукових служб, доцільно згадати і про інші типи ПС: пошукові сервери спеціального призначення та метапошукові сервери.