*
*

А что, если иголку в стоге сена искать правильно?

ПОИСКОВЫЕ СИСТЕМЫ ИНТЕРНЕТ

Поиск в Интернете

По статистическим данным суммарное число Web-страниц в конце 2001 г. составляло 7,5 миллиардов, а к концу 2005 г. это число возрастет до 25 миллиардов, причем количество пользо-вателей Интернет к указан-ному году превысит миллиард человек.

Инструменты информационного поиска

Поисковые машины (search engines)

Поисковая машина состоит из баз данных и программного обеспечения для их обработки, которое делится на две части: анализатор содержимого Web-страниц (Spider - спайдер, или «паук») и классификатор-определитель рейтинга, осуществляющий сортировку ссылок, найденных по запросу пользователя.

Каталоги (directories)

Тщательно подобранные (чаще вручную) ссылки на ресурсы, сопровождаемые описаниями. Некоторые каталоги содержат внутренние поисковые машины.

Порталы

Поисковые машины с интегрированным предложением собственного информационного содержания, которые могут также предоставлять дополнительные услуги: бесплатные адреса электронной почты, места для размещения домашних Web-страниц и другие.

Метапоисковые системы

Надстройки над поисковыми машинами, которые не имеют собственной базы данных и при поиске по запросу пользователя формируют, в свою очередь, запросы для нескольких внешних поисковых машин, а затем анализируют полученные результаты и выдают список ссылок в порядке, определяемом соотношением рейтингов ответа сразу по нескольким поисковым машинам. Проще говоря, такая машина ведет опрос нескольких поисковых систем, а затем отбирает ссылки, следуя собственному алгоритму.

Yahoo! (http://www.yahoo.com)

- ✔ Один из самых первых, надежных и авторитетных справочников Всемирной паутины.
- ✔ Транснациональный проект.
- ✔ Внушительный объем (2 000 000 сайтов, 25 000 категорий).
- ✔ Научность и логичность используемой схемы классификации.
- ✔ 14 категорий: Бизнес и Экономика, Новости и СМИ, Образование, …
- ✔ Перекрестная структура.
- ✔ Встроенная поисковая система.

Open Directory Project (http://dmoz.org)

- ✔ Один из наиболее полных справочников в мире (6 000 000 сайтов, 540 000 категорий).
- ✔ Некоммерческая направленность.
- ✔ Научная и логическая иерархическая структура.
- ✔ Русскоязычный вариант (22 000 сайтов).
- ✔ Динамически развивающийся проект (среднее увеличение справочника более чем на 1000 сайтов в день).

dmoz open directory project

About (http://www.about.com/)

- ✔ Справочник, поддерживаемый экспертами различных областей знания.
- ✔ 24 категории: Автомобили, Хобби, Подростки, ...
- ✔ Огромное количество рекламы.

WHAT YOU NEED TO KNOW About

Апорт (http://aport.ru)

- ✔ Единственный профессионально поддерживаемый отечественный справочник.
- ✔ Является порталом.
- ✔ Многоуровневая иерархическая структура.
- ✔ Сортировка ссылок (по дате, по алфавиту, …).
- ✔ Низкая скорость актуализации сведений.

АПОРТ
ищет что надо

Российские справочники ресурсов

Пингвин

http://www.pingwin.ru

Иван Сусанин

http://www.susanin.net

Улитка

http://www.ulitka.ru

Глобальные поисковые системы

Google (http://www.google.com)

- ✔ Система запущена в 1998 году.
- ✔ Единоличный лидер среди глобальных поисковых систем (3 3000 000 000 web-страниц).
- ✔ Лучшие на сегодня возможности поиска иллюстраций (425 000 иллюстраций).
- ✔ Специальный модуль ранжирования результатов.
- ✔ Наличие русскоязычного интерфейса (http://www.google.com.ru).
- ✔ Высокая степень комфорта для пользователя.
- ✔ Простая методика поиска.

Глобальные поисковые системы

Lycos

http://www.lycos.com

HotBot

http://www.hotbot.com

Excite

http://www.excite.com

Яндекс (http://www.yandex.ru/)

- ✔ Запущен в сентябре 1997 года.
- ✔ Признанный лидер российского поискового сервиса.
- ✔ Еженедельная актуализация базы.
- ✔ Простая форма запроса.
- ✔ На начало февраля 2004 года Яндексом проиндексировано свыше 970 000 российских и зарубежных русскоязычных серверов, а также серверов на территории СНГ (всего учтено более 140 000 000 оригинальных документов).

Яndex
Найдётся всё

Rambler (http://www.rambler.ru)

- ✔ Запущен в октябре 1996 года.
- ✔ Проведенная в декабре 2002 года коренная модернизация программно-аппаратной части позволил поисковой системе вновь приобрести былой авторитет (занимает второе место после Яндекса по величине базы данных).
- ✔ Производительность робота – 6 900 000 страниц в сутки.
- ✔ Простая и расширенная форма ввода запроса.
- ✔ Rambler Top 100.

Rambler®

Turtle (http://www.turtle.ru)

- ✔ Запущена в июне 2002 года.
- ✔ Около 100 000 000 оригинальных документов.
- ✔ Быстрый результат запроса.
- ✔ Перспективная развивающаяся система.

Мета-поисковые системы

MetaCrawler (http://www.metacrawler.com)

Мета-поисковые системы не имеют собственных поисковых роботов и баз данных (индексированных файлов). Их достоинство заключается в умении рассылать запросы по другим системам, а затем суммировать результаты.

Структура информационно-поисковых запросов

ОПЕРАТОР	RAMBLER	YANDEX	APORT
Логическое И	**And** **&** **Пробел**	**& , Пробел**(в пределах предложения) **&&** (в пределах документа)	**AND, &, +, И, пробел**
Логическое ИЛИ	**Or, |**	**|**	**Or, |, ИЛИ**
Логическое НЕ	**NOT, !**	**~** (оператор «И НЕ» в пределах предложения) **~~** (оператор «И НЕ» в пределах документа)	**NOT, -**
Группировка	**( )**	**( )**	**( )**

Стратегия поиска информации в Интернете

1. Выберите основные понятия, описывающие предмет вашего поиска.
2. Выберите ключевые слова, подходящие к данному понятию.
3. Подыщите как можно больше синонимов к вашим ключевым словам.
4. Определите, какой тип операторов поиска (OR, AND, NOT) лучше подойдет в вашем случае.
5. Выберите подходящую поисковую систему.
6. Изучите особенности поисковой системы, воспользовавшись разделом Помощь.
7. Подготовьте заранее выражения для поиска, проверив орфографию.
8. Проведите запросы несколько раз, слегка изменяя выражение.
9. Модифицируйте свои запросы в зависимости от результатов.
10. Попробуйте выполнить тот же запрос на других поисковых системах.

Язык поисковых запросов

РЕГИСТР

В общем случае, регистр написания поисковых слов и операторов значения не имеет, то есть **дом** и **ДОМ**, **Not** и **nOt** воспринимаются одинаково.

МОРФОЛОГИЯ

По каждому слову запроса поиск ведется с учетом правил словоизменения соответствующего языка. Например, при поиске по слову **'человек'** будут также найдены документы, содержащие слова **'человеку'**, **'человеком'**, **'человека'** и даже **'люди'**. Чтобы провести поиск только по одной определенной форме слова, нужно взять его в двойные кавычки или воспользоваться поиском точной фразы в расширенном поиске.

ОГРАНИЧЕНИЕ НА РАССТОЯНИЕ

Если запрос составлен из одного или нескольких слов без применения операторов и конструкций языка запросов, то будут найдены документы, в которых встречаются все слова запроса. При этом для каждого запроса всегда существует так называемое ограничение контекста. Например, по запросу 'красная армия' будут найдены те документы, в которых слова 'красная' и 'армия' хотя бы один раз встретятся менее чем в 40 словах друг от друга. Значение ограничения контекста можно изменять конструкцией '(число, запрос)‘. Например, '(2, красная армия)‘.

НЕНАЙДЕННЫЕ СЛОВА

Если запрос состоит из нескольких слов, и при этом некоторые из них вообще не удалось найти в Интернете, то выдаются результаты поиска по частичному запросу, из которого отсутствующие в Интернете слова исключены.

Как скачивать информацию

Копирование рисунков

1. Щелкнуть на рисунке правой кнопкой мыши.
2. Выбрать из контекстного меню строку **Сохранить как**.
3. В диалоговом окне **Сохранение рисунка** выбрать каталог, куда сохраняем.
4. Щелкнуть на кнопке **Сохранить**.

Копирование текста

1. Выделить копируемый текст.
2. Щелкнуть на тексте правой кнопкой мыши.
3. Выбрать из контекстного меню строку **Копировать**.
4. Открыть текстовый редактор.
5. В главном меню редактора щелкнуть команду **Правка | Вставить**.
6. После вставки текста из буфера обмена в главном меню редактора щелкнуть команду **Файл | Сохранить как**.
7. Сохранить текстовый файл как обычно.