

Поисковые системы



Поисковые системы

- Поисковые системы сейчас – это огромные и сложные механизмы, представляющие собой не только инструмент поиска информации, но и заманчивые сферы для бизнеса.



Поисковые системы

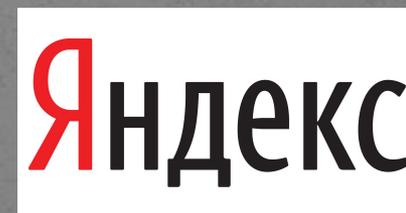
Наиболее крупные международные поисковые системы:

- [«Google»](#);
- [«Yahoo»](#);
- [«MSN»](#).



В русском Интернете:

- [«Яндекс»](#);
- [«Рамблер»](#);
- [«Апорт»](#).



Поисковые системы

- Первоочередная задача любой поисковой системы – доставлять людям именно ту информацию, которую они ищут.
- Поисковая система должна «думать» так же, как думает пользователь при поиске информации.



Поисковые системы

- Нашел ли он то, что искал?
- Если не нашел, то сколько раз ему пришлось перефразировать запрос, чтобы найти искомое?
- Насколько актуальную информацию он смог найти?
- Насколько быстро обрабатывала запрос поисковая машина?
- Насколько удобно были представлены результаты поиска?
- Был ли искомый результат первым или же сотым?
- Как много ненужного мусора было найдено наравне с полезной информацией?
- Найдется ли нужная информация, при обращении к поисковой системе, скажем, через неделю, или через месяц?

Поисковые системы

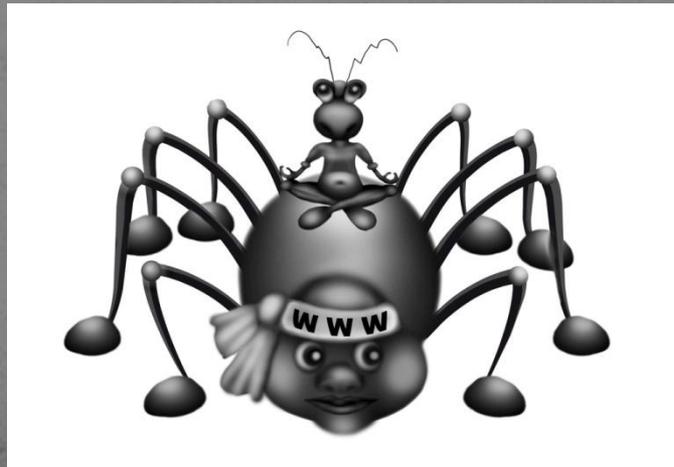
Основные характеристики поисковой системы:

- Полнота;
- Точность;
- Актуальность;
- Устойчивость к нагрузкам;
- Наглядность представления результатов.



Поисковые системы

- Spider (паук) – программа, предназначенная для скачивания веб-страниц.
- «Паук» обеспечивает скачивание страницы и извлекает все внутренние ссылки с этой страницы.
- Скачивается html-код каждой страницы.
- Для скачивания страниц роботы используют протоколы НТТР.



Поисковые системы

- Crawler («путешествующий» паук) – программа, которая автоматически проходит по всем ссылкам, найденным на странице.
- Его задача - определить, куда дальше должен идти паук, основываясь на ссылках или исходя из заранее заданного списка адресов.
- Crawler, следуя по найденным ссылкам, осуществляет поиск новых документов, еще неизвестных поисковой системе.

Поисковые системы

- Indexer (робот-индексатор) - программа, которая анализирует веб-страницы, скаченные пауками.
- Индексатор разбирает страницу на составные части и анализирует их, применяя собственные лексические и морфологические алгоритмы.



Поисковые системы

Поисковый сервер работает следующим образом:

- Полученный от пользователя запрос подвергается морфологическому анализу.
- Полученные данные передаются в качестве входных параметров специальному модулю ранжирования.
- В зависимости от выбора пользователя этот рейтинг может быть скорректирован.
- Генерируется сниппет.
- Полученные результаты поиска передаются пользователю в виде SERP (Search Engine Result Page) – страницы выдачи поисковых результатов.

**Спасибо за
внимание!**

