

## Лекция «Поиск информации с использованием компьютера»

В настоящее время существует множество справочных служб Интернет, помогающих пользователям найти нужную информацию. В таких службах используется обычный принцип поиска в неструктурированных документах – по ключевым словам.

**Поисковая система** – это комплекс программ и мощных компьютеров, способные принимать, анализировать и обслуживать запросы пользователей по поиску информации в Интернет.

### ПС бывают 2 видов:

#### Тематические каталоги Автоматические индексы

Тематические каталоги: создаются специалистами, которые занимаются классификацией и анализом информации, создавая базы данных.

Автоматические индексы: составляются специальной программой, работа которой выбирает на Web страницах гиперссылки и помещает их в базу данных с указанием URL адреса.

**Служба World Wide Web (WWW)** – это единое информационное пространство, состоящее из сотен миллионов взаимосвязанных электронных документов.

Отдельные документы, составляющие пространство Web, называют **Web-страницами**.

Группы тематически объединенных Web-страниц называют **Web-узлами** (сайтами).

Программы для просмотра Web-страниц называют **браузерами** (обозревателями).

### **К средствам поисковых систем относится язык запросов.**

Используя различные приёмы можно добиться желаемого результата поиска.

1. **!** – запрет перебора всех словоформ.
2. **+** – обязательное присутствие слов в найденных документах.
3. **-** – исключение слова из результатов поиска.
4. **&** – обязательное вхождение слов в одно предложение.
5. **~** – требование присутствия первого слова в предложении без присутствия второго.
6. **|** – поиск любого из данных слов.
7. **«»** – поиск устойчивых словосочетаний.
8. **\$title** (в переводе название) – поиск информации по названиям заголовков.
9. **\$anchor** – поиск информации по названию ссылок.

## **Поиск информации: основные понятия, виды и формы организации**

Поиск информации или информационный поиск представляет один из основных информационных процессов, цели, возможности и характер которого всегда зависели от наличия, информации, её важности и доступности, а также средств организации поиска.

**Поиск** - процесс, в ходе которого в той или иной последовательности производится соотнесение отыскиваемого с каждым объектом, хранящимся в массиве.

Цель любого поиска заключается в потребности, необходимости или желании находить различные виды информации, способствующие получению нужных сведений, знаний и т.д.

**"Поисковые системы"** осуществляют поиск среди документов базы или иных массивов машиночитаемых данных, содержащих заданные слова.

### **Информационно-поисковые системы**

ПС с большим набором функций и возможностей которые входят в состав СУБД называются информационно-поисковыми системами. Они также создаются и используются для эффективного нахождения пользователями необходимых им данных, в том числе в Интернете.

**"Информационно-поисковая система"** представляет систему, предназначенную для поиска и хранения информации; пакет программного обеспечения, реализующий процессы создания, актуализации, хранения и поиска в информационных базах и банках данных.

Информационный поиск подразумевает использование определённых стратегий, методов, механизмов и средств.

**Стратегия поиска** - общий план (концепция, предпочтение) который определяет достижение компромисса между практическими потребностями и возможностями имеющихся средств.

**Метод поиска** - совокупность моделей и алгоритмов реализации отдельных технологических этапов.

**Поисковый образ запроса** - записанный на ИПЯ(информационно-поисковых языков) текст, выражающий смысловое содержание, необходимые для наиболее эффективного осуществления информационного поиска.

**Механизмы поиска** - совокупность реализованных в системе моделей и алгоритмов процесса формирования выдачи документов в ответ на поисковый запрос.

Процесс поиска информации представляет последовательность шагов, приводящих при посредстве системы к некоторому результату, и позволяющих оценить его полноту

Процесс поиска можно представить в виде следующих основных компонент:

- 1) формулирование запроса на естественном языке, выбор поисковых системы и сервисов, формализация запроса на соответствующем ИПЯ;
- 2) проведение поиска в одной или нескольких поисковых системах;
- 3) обзор полученных результатов (ссылок);
- 4) предварительная обработка полученных результатов: просмотр содержания ссылок, извлечение и сохранение релевантных данных;
- 5) при необходимости, модификация запроса и проведение повторного (уточняющего) поиска с последующей обработкой полученных результатов.

### **Комбинация условия поиска.**

Этапы организации поиска:

1. Определение области знаний;
2. Выбор типа и источников данных;
3. Сбор материалов необходимых для наполнения информационной модели;
4. Отбор наиболее полезной информации;
5. Выбор метода обработки информации (классификация, кластеризация, регрессионный анализ и т.д.);
6. Выбор алгоритма поиска;
7. Поиск закономерностей, формальных правил и связей в собранной информации;
8. Творческая интерпретация полученных результатов;
9. Интеграция извлеченных "знаний".

Для проведения поиска первоначально на компьютере пользователя загружается интерфейс работы с соответствующей БД. Это может быть локальная или удалённая БД.

### ***Методы обработки результатов поиска***

По характеру преобразований методы обработки результатов поиска можно условно разделить на две группы:

1. Структурно-форматные преобразования;
2. Структурно-семантические преобразования (информационно-аналитические, логико-семантические).

### **Поисковые системы**

Эффективный доступ к информации в Интернете обеспечивают такие **зарубежные поисковые системы** (машины), как Альта-Виста (AltaVista), "Yahoo", "Google", "OpenText", "WebCrawler" и др. Их адреса в

Интернете: [www.altavista.com](http://www.altavista.com), [www.yahoo.com](http://www.yahoo.com), [www.gogle.com](http://www.gogle.com),  
[www.opentext.com](http://www.opentext.com),

К *отечественным поисковым* машинам относятся: Rambler (фирма Stack Ltd.), Яндекс ("Yandex" фирма CompTek Int), "Русская машина поиска", "Новый русский поиск", и др. Их адреса в Интернете: [www.aport.ru](http://www.aport.ru), [www.rambler.ru](http://www.rambler.ru), [www.yandex.ru](http://www.yandex.ru), [search.interrussia.com](http://search.interrussia.com), [www.openweb.ru](http://www.openweb.ru) соответственно) и др.