

Технология поиска информации в сети Интернет.



Инструменты информационного поиска

Поиск:

- **фактографический поиск:** в энциклопедиях, справочниках, словарях;
- **библиографический поиск:** библиотеки, каталоги, программы;
- **документальный поиск:** электронные документы, электронные библиотеки, электронные журналы.

ПОИСКОВЫЕ СЛУЖБЫ ИЛИ СЕРВИСЫ:

- *справочники или каталоги*
- *поисковые системы*

Поведение пользователя, осуществляющего управление процессом поиска, определяется не только информационной потребностью, но и инструментальным разнообразием системы - технологиями и средствами, предоставляемыми системой.

Поисковые технологии -

инициализированные (оптимизированные в рамках конкретной информационно-поисковой системы) последовательности эффективного использования отдельных средств поиска в процессе взаимодействия пользователя с системой.

информационные системы:

- *тематические каталоги и специализированные каталоги (онлайновые справочники);*
- *поисковые машины (полнотекстовый поиск);*
- *средства мета-поиска.*

- *Тематические каталоги* предусматривают обработку документов и отнесение их к одной из нескольких категорий, перечень которых заранее задан.
- *Специализированные каталоги* или *справочники* создаются по отдельным отраслям и темам.

- *Поисковые машины* реализуют технологию полнотекстового поиска. Индексируются тексты, расположенные на запрашиваемых серверах.
- При использовании средств *мета-поиска* запрос осуществляется одновременно несколькими поисковыми системами. Результат поиска объединяется в общий, упорядоченный по степени релевантности, список.

- **Средства поиска** - взаимозависимый комплекс информационно-поисковых языков и языков определения/управления данными, обеспечивающий структурные и семантические преобразования объектов обработки (документов, словарей, совокупностей результатов поиска).

1. Справочники

- имеют четкую иерархическую систематическую или логико-тематическую структуру.
- Работа со справочниками позволяет ориентироваться в Интернет-ресурсах в пределах отдельных отраслей знания, углубляясь от общего к частному, меняя иерархические ветви, возвращаясь на несколько шагов назад и т.д.

- Среди российских разработок в этой области значатся:
- Апорт (адрес: www.aport.ru),
- List.ru (адрес: list.mail.ru),
- Weblist (адрес: www.weblist.ru),
- Иван Сусанин (адрес: www.susanin.net)
- Улитка (адрес: www.ulitka.ru).

- Главной отличительной особенностью справочников является то, что они сделаны вручную.

Параметры, характеризующие достоинства справочников, являются:

- *объем;*
- *оперативность отражения новых или изменившихся ресурсов;*
- *логичность и последовательность иерархической схемы классификации;*
- *перекрестность структуры.*

- Объем - степень его надежности.
- Логичность - степень простоты, с которой пользователи находят требуемые сведения.
- Система **перекрестных ссылок** позволяет выявить информацию, используя разные подходы.

Справочники предназначены для решения трех типов задач:

- ориентация в незнакомой отрасли знания;
- разыскание крупных объектов, каковыми являются, к примеру, серверы организаций или значительных проектов;
- получение готового перечня ресурсов, имеющих размытый поисковый образ (библиотек определенного типа, транспортных расписаний или сайтов политических партий и т.д.)

2. Поисковые системы

- Задача поисковых машин - обеспечить детальное разыскание информации, что может быть достигнуто только за счет учета (**индексирования**) содержания максимально возможного числа веб-страниц. В отличие от справочников, поисковые машины функционируют в автоматизированном режиме и имеют единообразный принцип действия.

Поисковая система

```
graph TD; A[Поисковая система] --> B[Программа-робот]; A --> C[Компонент индексации, модуль ранжирования];
```

Программа-робот

Компонент
индексации,
модуль ранжирования

Вывод результатов осуществляется с помощью специального модуля, который производит **интеллектуальное ранжирование результатов**. При этом в расчет берется:

- *местоположение термина в документе (название, заголовок, основной текст), частота его повторения,*
- *процентное соотношение искомого термина к тексту страницы,*
- *число и авторитетность внешних ссылок на данную страницу с других сайтов.*

К основным параметрам поисковых систем относятся:

- *число проиндексированных серверов и отдельных документов (объем индексных файлов);*
- *степень оперативности обновления базы данных за счет включения сведений о новых материалах и удаления устаревших;*
- *возможности для составления запроса;*
- *интеллектуальность системы ранжирования результатов поиска;*
- *наличие дополнительных сервисных функций, облегчающих работу пользователя.*

Примеры поисковых систем

- **Google** (адрес: www.google.com),
- **AlltheWeb**(адрес: www.alltheweb.com),
- **Alta Vista** (адрес: www.altavista.com).

лидирующая группа российских поисковых средств

- **Яндекс** (адрес: www.yandex.ru),
- **Rambler** (адрес: www.rambler.ru),
- **Апорт** (адрес: www.aport.ru)

В российском сегменте Интернета представлены и мета-поисковые системы, рассылающие запросы как в российские, так и глобальные поисковые системы.

- **Punto** (адрес: <http://punto.ru>)
- **MetaBot** (адрес: <http://www.metabot.ru>)

Три способа поиска информации в Интернет

Поиск информации – одна из самых востребованных на практике задач, которую приходится решать любому пользователю **Интернета**.

Существуют три основных способа поиска информации в Интернет:

1. Указание адреса страницы.
2. Передвижение по гиперссылкам.
3. Обращение к поисковой системе (поисковому серверу).

Поисковая система

— веб-сайт, предоставляющий возможность поиска информации в Интернете.

Большинство поисковых систем ищут информацию на сайтах Всемирной паутины, но существуют также системы, способные искать файлы на ftp-серверах, товары в интернет-магазинах, а также информацию в группах новостей Usenet.

Типы поисковых систем

По принципу действия поисковые системы делятся на два типа:



Поисковые каталоги

Поисковые каталоги служат для тематического поиска.

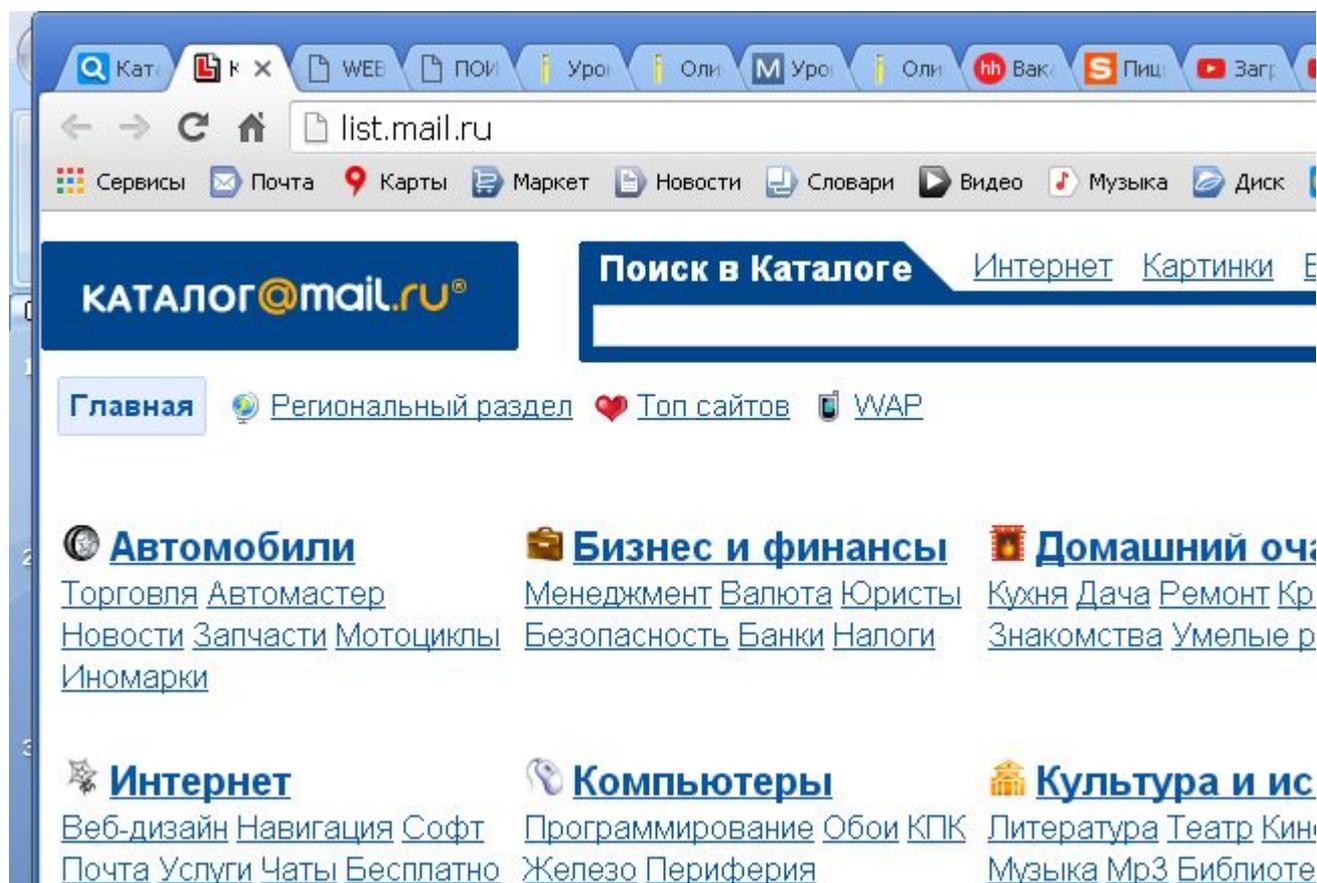
Информация на этих серверах структурирована по темам и подтемам.

Каталог ресурсов в Интернете —

структурированный набор ссылок на сайты с кратким их описанием.

Каталог в котором ссылки на сайты внутри категорий сортируются по популярности сайтов называется рейтинг (или топ).

Пример поискового каталога



Поисковые индексы

Поисковые индексы работают как алфавитные указатели. Клиент задает слово или группу слов, характеризующих его область поиска, — и получает список ссылок на web-страницы, содержащие указанные термины.

Первой поисковой системой для Всемирной паутины был «Wandex», уже не существующий индекс, разработанный Мэтью Грэйем из Массачусетского технологического института в 1993.

Язык запросов поисковой системы

- Это группа ключевых слов, сформированная по определенным правилам — с помощью языка запросов, называется запросом к поисковому серверу.
- Языки запросов к разным поисковым серверам очень похожи. Подробнее об этом можно узнать, посетив раздел «Помощь» нужного поискового сервера.

Правила формирования запросов на примере поисковой системы Yandex.

синтаксис оператора	Что означает оператор	Пример запроса
пробел или &	Логическое И (в пределах предложения)	лечебная физкультура
&&	Логическое И (в пределах документа)	рецепты && (плавленый сыр)
	Логическое ИЛИ	фото фотография снимок фотоизображение
+	Обязательное наличие слова в найденном документе	+быть или +не быть
()	Группирование слов	(технология изготовление) (сыра творога)

Правила формирования запросов на примере поисковой системы Yandex. продолжение

синтаксис оператора	Что означает оператор	Пример запроса
~	Бинарный оператор И НЕ (в пределах предложения)	банки ~ закон
~~ ИЛИ -	Бинарный оператор И НЕ (в пределах документа)	путеводитель по Парижу ~~ (агентство тур)
/(n m)	Расстояние в словах (минус (-) — назад, плюс (+) — вперед)	поставщики /2 кофе музыкальное /(-2 4) образование вакансии - /+1 студентов
“ ”	Поиск фразы	“красная шапочка” Эквивалентно: красная /+1 шапочка
&&/(n m)	Расстояние в предложениях (минус (-) — назад, плюс (+) — вперед)	банк && /1 налоги

Несколько простых правил поиска

1. Не искать информацию только по одному ключевому слову.
2. Лучше не вводить ключевые слова с прописной буквы, так как это может привести к тому, что не будут найдены те же слова, написанные со строчной буквы.
3. Если в итоге поиска вы не получили никаких результатов, проверьте, нет ли в ключевых словах орфографических ошибок.

Задания ЕГЭ по этой теме (№17)

Каталог заданий . Расположение запросов в порядке убывания/возрастания

[Пройти тестирование по этим заданиям](#)

[Вернуться к каталогу заданий](#)

[Версия для печати и копирования в MS Word](#)

№	Условие
1	<p>Задание 17 № 2705. Ниже приведены запросы к поисковому серверу. Расположите номера запросов в порядке возрастания количества страниц, которые найдет поисковый сервер по каждому запросу. Для обозначения логической операции «ИЛИ» в запросе используется символ , а для логической операции «И» – &.</p> <ol style="list-style-type: none">1) принтеры & сканеры & продажа2) принтеры & продажа3) принтеры продажа4) принтеры сканеры продажа <p>Показать пояснение</p> <p>Помощь по заданию Сообщить об ошибке</p>
№	Условие
2	<p>Задание 17 № 3151. В таблице приведены запросы к поисковому серверу. Расположите номера запросов в порядке возрастания количества страниц, которые найдет поисковый сервер по каждому запросу.</p> <p>Для обозначения логической операции «ИЛИ» в запросе используется символ , а для логической операции «И» – &.</p> <p>№ Запрос</p> <ol style="list-style-type: none">1 канарейки щеглы содержание2 канарейки & содержание3 канарейки & щеглы & содержание4 разведение & содержание & канарейки & щеглы <p>Показать пояснение</p> <p>Помощь по заданию Сообщить об ошибке</p>