

42. Поиск информации в Интернете



Для поиска информации в Интернете используются специальные поисковые системы, которые содержат в базах данных постоянно обновляемую информацию о веб-сайтах, файлах и других информационных ресурсах. Разные поисковые серверы могут использовать различные механизмы поиска, хранения и предоставления информации.

Интерфейс поисковых систем общего назначения содержит поле поиска и список разделов каталога. Заполнение баз поисковой системы осуществляется с помощью специальных программ-роботов, которые периодически «обходят» веб-серверы. Программы-роботы читают все встречающиеся документы, выделяют в них ключевые слова и заносят в базу данных, содержащую URL-адреса документов.

Так как информация в интернете постоянно меняется (создаются новые веб-сайта и страницы, удаляются старые, меняются их URL-адреса и т.д.) поисковые роботы не всегда успевают отследить все изменения. Информация, хранящаяся в базе данных поисковой системы, может отличаться от реального состояния и тогда пользователь в результате поиска может получить адрес уже не существующего или перемещенного документа.

Поиск по ключевым словам

При поиске по ключевым словам в поле поиска вводится одно или несколько ключевых слов, которые являются главными для искомого документа. Можно также использовать сложные запросы, использующие логические операции, шаблоны и т.д.

После отправки запроса поисковая система вернет аннотированный (с короткими комментариями содержания документа) список URL-адресов документов, в которых были найдены указанные вами ключевые слова. Если ключевые слова были выбраны неудачно, то список URL-адресов документов может быть слишком большим (содержать десятки и сотни тысяч ссылок). Чтобы уменьшить список, можно в поле поиска ввести дополнительные ключевые слова или воспользоваться каталогом поисковой системы.

Язык построения запросов

Поиск по одному ключевому слову всегда производится однозначно (ключевое слово либо имеется в рассматриваемом тексте, либо отсутствует). При поиске же по нескольким ключевым словам возможны различные способы их комбинирования:

- в искомом тексте должны присутствовать все заданные ключевые слова («И то, И другое»)
- достаточно, чтобы в искомом тексте имелось хотя бы одно из заданных ключевых слов («то ИЛИ другое»)
- в тексте должно иметься одно ключевое слово, но обязательно должно отсутствовать другое («то, но НЕ другое»)
- ключевые слова составляют фразу, которая обязательно должна присутствовать в тексте «как есть»
- ключевые слова обязательно должны быть взаимосвязаны (т.е. должны располагаться близко друг от друга, — например, в фразе могут быть разделены союзом или прилагательным).

Чтобы «объяснить» поисковой системе, как нужно понимать заданную последовательность ключевых слов, используется особый язык построения запросов. Принципы построения языка запросов, как правило, универсальны для большинства поисковых систем, но для иной системы язык запросов может иметь свои особенности.



Как ни спроси, **я** понимаю

www.yandex.ru

Пример языка построения запросов в поисковой системе Яндекс:

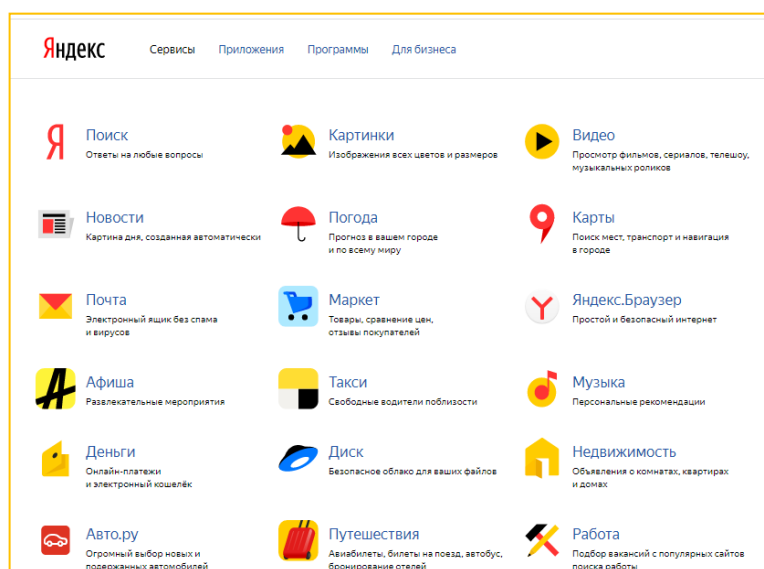
- заяц & кролик — поиск текстов, в которых есть оба заданных ключевых слова (и «заяц» и «кролик»);
- заяц | кролик — поиск текстов, в которых имеется хотя бы одно из заданных ключевых слов («заяц» или «кролик»);
- заяц -кролик — знак минуса перед словом «кролик» предписывает искать только такие тексты, в которых есть слово «заяц», но нет слова «кролик»;

- заяц +кролик — знак плюса перед словом «кролик» предписывает искать только такие тексты, в которых есть слово «заяц» и обязательно есть слово «кролик»;

Большинство современных поисковых систем являются «интеллектуальными» — при поиске учитываются все возможные формы (падежи, склонения, спряжения, единственное и множественное число) заданных ключевых слов. Например, при задании ключевого слова заяц поисковая система будет искать все варианты ключевых слов — «заяц», «зайца», «зайцами» и т.п.

Поиск в иерархической системе каталогов

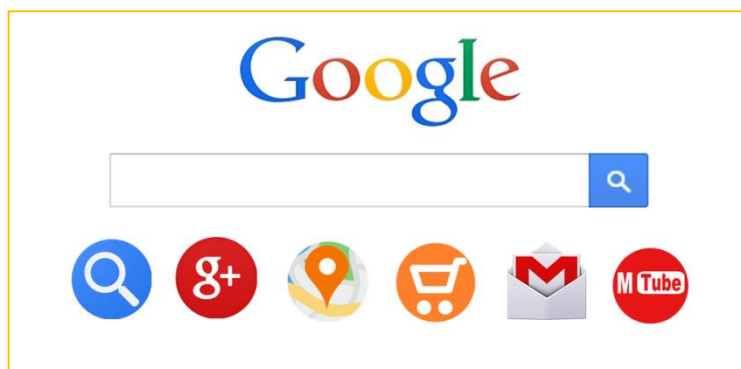
Каталоги составляются редакторами, просматривающими каждый новый сайт до его включения в иерархическую систему каталогов. Каталоги обычно организованы в соответствии с предметной классификацией.



Поиск информации в каталоге сводится к выбору определенного каталога, после чего пользователю будет представлен список ссылок на URL-адреса наиболее посещаемых и важных веб-сайтов. Каждая ссылка обычно аннотирована.

Современные поисковые системы

Одной из наиболее полных и мощных поисковых систем является Google (www.google.ru). В Рунете (русской части Интернета) обширные базы данных имеет поисковая система Яндекс (www.yandex.ru).



Сайты в базе данных ранжируются по количеству их посещений в день, неделю или месяц. Посещаемость сайтов определяется с помощью специальных счетчиков, которые могут быть установлены на сайте. Счетчики фиксируют каждое посещение сайта и передают информацию о количестве посещений на сервер поисковой системы.

Современные поисковые системы часто являются информационными порталами, которые представляют пользователям не только возможности поиска документов в Интернете, но и доступ к другим информационным ресурсам (к новостям, к информации о погоде, о валютном курсе, к интерактивным географическим картам и т.д.)

ЗАДАНИЕ № 1

Заполните таблицу. Файл WORD присылайте на адрес informatika-emk64@mail.ru

Положительные стороны Интернета	Отрицательные стороны Интернета
1.	1.
2.	2.
3.	3.
4.	4.
5.	5.
6.	6.
7.	7.
8.	8.
9.	9.
10.	10.

ЗАДАНИЕ № 2

Определите, какое из указанных имен файлов удовлетворяет маске: **pre??.***

- | | |
|---------------|---------------|
| 1) spread.gif | 3) pre12.docx |
| 2) prelab.avi | 4) pred.xls |

ЗАДАНИЕ № 3

В таблице приведены запросы к поисковому серверу. Расположите номера запросов в порядке возрастания количества страниц, которые найдет поисковый сервер по каждому запросу.

1	Лекарства
2	Лекарства & Витамины & БАДы
3	Лекарства & Витамины
4	Лекарства БАДы

Выберите правильный ответ: 1) 1432 2) 2341 3) 2314 4) 4132

ЗАДАНИЕ № 4

Повторить пройденные темы можно по видеороликам из интернет:

1. «Web-страницы и Web-сайты. Структура Web-страницы» :
<https://youtu.be/vcrE0UAnwLI>
2. «4 апреля – День веб-мастера»
<https://youtu.be/ipv77M-yfaw>
3. «Браузер и основы работы с информацией в Интернет»
<https://youtu.be/t57rjPzMVKM>