**Технология поиска информации в Интернете: информационные поисковые системы Интернета**

28 Февраля 2006 года, 12:00

|  |  |  |
| --- | --- | --- |
| [Сохранить](http://vk.com/share.php?url=http%3A%2F%2Fwww.seonews.ru%2Fmasterclasses%2Ftehnologiya-poiska-informatsii-v-internet-vidyi-poiskovyih-instrumentov-informatsionnyie-poiskovyie-sistemyi-interneta%2F) |  | **[9](http://vk.com/share.php?url=http%3A%2F%2Fwww.seonews.ru%2Fmasterclasses%2Ftehnologiya-poiska-informatsii-v-internet-vidyi-poiskovyih-instrumentov-informatsionnyie-poiskovyie-sistemyi-interneta%2F)** |

Шрифт:

[0](http://www.seonews.ru/masterclasses/tehnologiya-poiska-informatsii-v-internet-vidyi-poiskovyih-instrumentov-informatsionnyie-poiskovyie-sistemyi-interneta/#comments) 24571

1. [Введение](http://www.seonews.ru/masterclasses/tehnologiya-poiska-informatsii-v-internet-vidyi-poiskovyih-instrumentov-informatsionnyie-poiskovyie-sistemyi-interneta/#1)   
2. [Технологии поиска](http://www.seonews.ru/masterclasses/tehnologiya-poiska-informatsii-v-internet-vidyi-poiskovyih-instrumentov-informatsionnyie-poiskovyie-sistemyi-interneta/#2)

2.1 [Поисковые инструменты](http://www.seonews.ru/masterclasses/tehnologiya-poiska-informatsii-v-internet-vidyi-poiskovyih-instrumentov-informatsionnyie-poiskovyie-sistemyi-interneta/#2.1)   
2.2 [Поисковые машины (search engines)](http://www.seonews.ru/masterclasses/tehnologiya-poiska-informatsii-v-internet-vidyi-poiskovyih-instrumentov-informatsionnyie-poiskovyie-sistemyi-interneta/#2.2)   
2.3 [Каталоги (directories)](http://www.seonews.ru/masterclasses/tehnologiya-poiska-informatsii-v-internet-vidyi-poiskovyih-instrumentov-informatsionnyie-poiskovyie-sistemyi-interneta/#2.3)   
2.4 [Подборки ссылок](http://www.seonews.ru/masterclasses/tehnologiya-poiska-informatsii-v-internet-vidyi-poiskovyih-instrumentov-informatsionnyie-poiskovyie-sistemyi-interneta/#2.4)   
2.5 [Базы данных адресов ( addresses database)](http://www.seonews.ru/masterclasses/tehnologiya-poiska-informatsii-v-internet-vidyi-poiskovyih-instrumentov-informatsionnyie-poiskovyie-sistemyi-interneta/#2.5)   
2.6 [Поиск в архивах Gopher (Gopher archives)](http://www.seonews.ru/masterclasses/tehnologiya-poiska-informatsii-v-internet-vidyi-poiskovyih-instrumentov-informatsionnyie-poiskovyie-sistemyi-interneta/#2.6)   
2.7 [Система поиска FTP файлов (FTP Search)](http://www.seonews.ru/masterclasses/tehnologiya-poiska-informatsii-v-internet-vidyi-poiskovyih-instrumentov-informatsionnyie-poiskovyie-sistemyi-interneta/#2.7)   
2.8 [Система поиска в конференциях Usenet News](http://www.seonews.ru/masterclasses/tehnologiya-poiska-informatsii-v-internet-vidyi-poiskovyih-instrumentov-informatsionnyie-poiskovyie-sistemyi-interneta/#2.8)   
2.9 [Системы мета-поиска](http://www.seonews.ru/masterclasses/tehnologiya-poiska-informatsii-v-internet-vidyi-poiskovyih-instrumentov-informatsionnyie-poiskovyie-sistemyi-interneta/#2.9)   
2.10 [Системы поиска людей](http://www.seonews.ru/masterclasses/tehnologiya-poiska-informatsii-v-internet-vidyi-poiskovyih-instrumentov-informatsionnyie-poiskovyie-sistemyi-interneta/#2.10)

3. [Заключение](http://www.seonews.ru/masterclasses/tehnologiya-poiska-informatsii-v-internet-vidyi-poiskovyih-instrumentov-informatsionnyie-poiskovyie-sistemyi-interneta/#3)

**1. Введение**

С каждым годом объемы Интернета увеличиваются в разы, поэтому вероятность найти необходимую информацию резко возрастает. Интернет объединяет миллионы компьютеров, множество разных сетей, число пользователей увеличивается на 15-80% ежегодно. И, тем не менее, все чаще при обращении к Интернет основной проблемой оказывается не отсутствие искомой информации, а возможность ее найти. Как правило, обычный человек в силу разных обстоятельств не может или не хочет тратить на поиск нужного ему ответа больше 15-20 минут. Поэтому особенно актуально правильно и грамотно научиться, казалось бы, простой вещи – где и как искать, чтобы получать ЖЕЛАЕМЫЕ ответы.

Чтобы найти нужную информацию, необходимо найти её адрес. Для этого существуют специализированные поисковые сервера (роботы индексов (поисковые системы), тематические Интернет-каталоги, системы мета-поиска, службы поиска людей и т.д.). В данном мастер-классе раскрываются основные технологии поиска информации в Интернете, предоставляются общие черты поисковых инструментов, рассматриваются структуры поисковых запросов для наиболее популярных русскоязычных и англоязычных поисковых систем.

**2. Технологии поиска**

Web-технология World Wide Web (WWW) считается специальной технологией подготовки и размещения документов в сети Интернет. В состав WWW входят и web-страницы, и электронные библиотеки, каталоги, и даже виртуальные музеи! При таком обилии информации остро встает вопрос: «Как сориентироваться в столь огромном и масштабном информационном пространстве?»   
В решении данной проблемы на помощь приходят поисковые инструменты.

**2.1 Поисковые инструменты**

Поисковые инструменты - это особое программное обеспечение, основная цель которого – обеспечить наиболее оптимальный и качественный поиск информации для пользователей Интернета. Поисковые инструменты размещаются на специальных веб-серверах, каждый из которых выполняет определенную функцию:

1. Анализ веб-страниц и занесение результатов анализа на тот или иной уровень базы данных поискового сервера.
2. Поиск информации по запросу пользователя.
3. Обеспечение удобного интерфейса для поиска информации и просмотра результата поиска пользователем.

Приемы работы, используемые при работе с теми или другими поисковыми инструментами, практически одинаковы. Перед тем как перейти к их обсуждению, рассмотрим следующие понятия:

1. Интерфейс поискового инструмента представлен в виде страницы с гиперссылками, строкой подачи запроса (строкой поиска) и инструментами активизации запроса.
2. Индекс поисковой системы – это информационная база, содержащая результат анализа веб-страниц, составленная по определенным правилам.
3. Запрос – это ключевое слово или фраза, которую вводит пользователь в строку поиска. Для формирования различных запросов используются специальные символы ("", , ~), математические символы (\*, +, ?).

Схема поиска информации в сети Интернет проста. Пользователь набирает ключевую фразу и активизирует поиск, тем самым получает подборку документов по сформулированному (заданному) запросу. Этот список документов ранжируется по определенным критериям так, чтобы вверху списка оказались те документы, которые наиболее соответствуют запросу пользователя. Каждый из поисковых инструментов использует различные критерии ранжирования документов, как при анализе результатов поиска, так и при формировании индекса (наполнении индексной базы данных web-страниц).

Таким образом, если указать в строке поиска для каждого поискового инструмента одинаковой конструкции запрос, можно получить различные результаты поиска. Для пользователя имеет большое значение, какие документы окажутся в первых двух-трех десятках документов по результатам поиска и на сколько эти документы соответствуют ожиданиям пользователя.

Большинство поисковых инструментов предлагают два способа поиска – **simple search** (простой поиск) и **advanced search** (расширенный поиск) с использованием специальной формы запроса и без нее. Рассмотрим оба вида поиска на примере англоязычной поисковой машины.

Например, AltaVista удобно использовать для произвольных запросов, «Something about online degrees in information technology», тогда как поисковый инструмент Yahoo позволяет получать мировые новости, информацию о курсе валют или прогнозе погоды.

Освоение критериев уточнения запроса и приемов расширенного поиска, позволяет увеличивать эффективность поиска и достаточно быстро найти необходимую информацию. Прежде всего, увеличить эффективность поиска Вы можете за счет использования в запросах логических операторов (операций) Or, And, Near, Not, математических и специальных символов. С помощью операторов и/или символов пользователь связывает ключевые слова в нужной последовательности, чтобы получить наиболее адекватный запросу результат поиска. Формы запросов приведены в таблице 1.

Таблица 1

|  |
| --- |
| Простой запрос |
| account |
| merchant account |
| internet merchant account |
| "merchant account" |
| "internet merchant account" |
| Расширенный запрос |
| internet merchant account and online payments |
| internet merchant near gov\* |
| internet merchant near education |
| Расширенный запрос с использованием математических символов |
| +internet +merchant +account |
| internet ~merchant ~gov\* |
| internet ~merchant ~governor |
| Internet ~merchant ~(governor account) |

Простой запрос дает некоторое количество ссылок на документы, т.к. в список попадают документы, содержащие одно из слов, введенных при запросе, или простое словосочетание (см. таблицу 1). Оператор and позволяет указать на то, что в содержании документа должны быть включены все ключевые слова. Тем не менее, количество документов может быть все еще велико, и их просмотр займет достаточно времени. Поэтому в ряде случаев гораздо удобнее применить контекстный оператор near, указывающий, что слова должны располагаться в документе в достаточной близости. Использование near значительно уменьшает количество найденных документов. Наличие символа "\*" в строке запроса означает, что будет осуществляться поиск слова по его маске. Например, получим список документов, содержащих слова, начинающиеся на "gov", если в строке запроса запишем "gov\*". Это могут быть слова government, governor и т.д.

Наиболее развитый сервис поиска русскоязычной информации предоставляет поисковый сервер [Яndex](http://www.yandex.ru). В Яndex можно просто написать по-русски фразу, описывающую то, что Вы хотите найти, и система проанализирует и обработает Ваш запрос, а затем постарается найти все, что относится к заданной теме. Вы можете, используя специальные операторы, составить строку, поясняющую поисковой системе, каким Вашим требованиям должна отвечать интересующая Вас информация. Некоторые из операторов языка запросов Яndex можно посмотреть здесь: <http://help.yandex.ru/search/?id=481939>

Не менее популярная поисковая система Rambler ведет статистику посещаемости ссылок из собственной базы данных, поддерживаются те же логические операторы И, ИЛИ, НЕ, метасимвол \* (аналогично расширяющему диапазон запроса символу \* в AltaVista), коэффициентные символы + и -, для увеличения или уменьшения значимости вводимых в запрос слов.

Давайте рассмотрим наиболее популярные технологии поиска информации в Интернет.

**2.2 Поисковые машины (search engines)**

Машины веб-поиска - это сервера с огромной базой данных URL-адресов, которые автоматически обращаются к страницам WWW по всем этим адресам, изучают содержимое этих страниц, формируют и прописывают ключевые слова со страниц в свою базу данных (индексирует страницы).

Более того, роботы поисковых систем переходят по встречаемым на страницах ссылкам и переиндексируют их. Так как почти любая страница WWW имеет множество ссылок на другие страницы, то при подобной работе поисковая машина в конечном результате теоретически может обойти все сайты в Интернет.

Именно этот вид поисковых инструментов является наиболее известным и популярным среди всех пользователей сети Интернет. У каждого на слуху названия известных машин веб-поиска (поисковых систем) – Яndex, Rambler, Aport.

Чтобы воспользоваться данным видом поискового инструмента, необходимо зайти на него и набрать в строке поиска интересующее Вас ключевое слово. Далее Вы получите выдачу из ссылок, хранящихся в базе поисковой системы, которые наиболее близки Вашему запросу. Чтобы поиск был наиболее эффективен, заранее обратите внимание на следующие моменты:

* определитесь с темой запроса. Что именно в конечном итоге Вы хотите найти?
* обращайте внимание на язык, грамматику, использование различных небуквенных символов, морфологию.Важно также правильно сформулировать и вписать ключевые слова. Каждая поисковая система имеет свою форму составления запроса — принцип один, но могут различаться используемые символы или операторы. Требуемые формы запроса различаются также в зависимости от сложности программного обеспечения поисковых систем и предоставляемых ими услуг. Так или иначе, каждая поисковая система имеет раздел "Help" ("Помощь"), где все синтаксические правила, а также рекомендации и советы по поиску, доступно объясняются (скриншот страничек поисковиков).
* используйте возможности разных поисковых систем. Если не нашли на Яndex, попробуйте на Google. Пользуйтесь услугами расширенного поиска.
* чтобы исключить документы, содержащие определенные термины, используйте знак "-" перед каждым таким словом. Например, если Вам нужна информация о работах Шекспира, за исключением "Гамлета", то введите запрос в виде: "Шекспир-Гамлет". И для того, чтобы, наоборот, в результаты поиска обязательно включались определенные ссылки, используйте символ "+". Так, чтобы найти ссылки о продаже именно автомобилей, Вам нужен запрос "продажа+автомобиль". Для увеличения эффективности и точности поиска, используйте комбинации этих символов.
* каждая ссылка в списке результатов поиска содержит [сниппет](http://www.seonews.ru/tags/snippety) – несколько строчек из найденного документа, среди которых встречаются Ваши ключевые слова. Прежде чем переходить по ссылке, оцените соответсвие сниппета теме запроса. Перейдя по ссылке на определенный сайт, внимательно окиньте взглядом главную страничку. Как правило, первой страницы достаточно, чтобы понять – по адресу Вы пришли или нет. Если да, то дальнейшие поиски нужной информации ведите на выбранном сайте (в разделах сайта), если нет – возвращайтесь к результатам поиска и пробуйте очередную ссылку.
* помните, что поисковые системы не производят самостоятельную информацию (за исключением разъяснений о самих себе). Поисковая система – это лишь посредник между обладателем информации (сайтом) и Вами. Базы данных постоянно обновляются, в них вносятся новые адреса, но отставание от реально существующей в мире информации все равно остается. Просто потому, что поисковые системы не работают со скоростью света.

К наиболее известным машинам веб-поиска относятся Google, Yahoo, Alta Vista, Excite, Hot Bot, Lycos. Среди русскоязычных можно выделить Яndex, Rambler, Апорт.

Поисковые системы являются самыми масштабными и ценными, но далеко не единственными источниками информации в Сети, ведь помимо них существуют и другие способы поиска в Интернете.

**2.3 Каталоги (directories)**

Каталог Интернет-ресурсов – это постоянно обновляющийся и пополняющийся иерархический каталог, содержащий множество категорий и отдельных web-серверов с кратким описанием их содержимого.Способ поиска по каталогу подразумевает «движение вниз по ступенькам», то есть движение от более общих категорий к более конкретным. Одним из преимуществ тематических каталогов является то, что пояснения к ссылкам дают создатели каталога и полностью отражают его содержание, то есть дает Вам возможность точнее определить, насколько соответствует содержание сервера цели Вашего поиска.

Примером тематического русскоязычного каталога можно назвать ресурс <http://www.ulitka.ru/>.

На главной странице данного сайта расположен тематический рубрикатор,

с помощью которого пользователь попадает в рубрику со ссылками на интересующую его продукцию.

Кроме того, некоторые тематические каталоги позволяют искать по ключевым словам. Пользователь вводит необходимое ключевое слово в строку поиска

и получает список ссылок с описаниями сайтов, которые наиболее полно соответствуют его запросу. Стоит отметить, что этот поиск происходит не в содержимом WWW-серверов, а в их кратком описании, хранящихся в каталоге.

В нашем примере в каталоге также имеется возможность сортировки сайтов по количеству посещений, по алфавиту, по дате занесения.

Другие примеры русскоязычных каталогов:   
[Каталог@Mail.ru](http://list.mail.ru/)   
[Weblist](http://www.weblist.ru/)   
[Vsego.ru](http://www.vsego.ru/)   
Cреди англоязычных каталогов можно выделить:   
<http://www.DMOS.org>   
<http://www.yahoo.com/>   
<http://www.looksmart.com>

**2.4 Подборки ссылок**

Подборки ссылок – это отсортированные по темам ссылки. Они достаточно сильно отличаются друг от друга по наполнению, поэтому чтобы найти подборку, наиболее полно отвечающую Вашим интересам, необходимо ходить по ним самостоятельно, дабы составить собственное мнение.

В качестве примера приведем Подборку ссылок ["Сокровища Интернет" АО "Релком"](http://www.relcom.ru/Internet/Treasures/)

Пользователь, нажимая на любую из заинтересовавших его рубрик

 СОДЕРЖАНИЕ

Автомобилистам

* Астрономия и астрология
* Ваш дом
* Ваши питомцы
* Дети - цветы жизни
* Досуг
* Города в Сети Internet
* Здоровье и медицина
* Информационные агентства и службы
* Краеведческий музей и т.д.,

попадает на подборку со ссылками на полезные Интернет-ресурсы   
Автомобилистам

* Автомобильная электроника.
* Музей автомото старины.
* Коллегия Правовой Защиты Автовладельцев.
* Sportdrive.

Преимуществом такого вида поисковых инструментов является их целенаправленность, обычно подборка включает в себя редкие интернет ресурсы, подобранные конкретным веб-мастером или хозяином интернет странички.

**2.5 Базы данных адресов ( addresses database)**

Базы данных адресов – это специальные поисковые сервера, которые обычно используют классификации по роду деятельности, по выпускаемой продукции и оказываемым услугам, по географическому признаку. Иногда они дополнены поиском по алфавиту. В записях базы данных хранится информация о сайтах, которые предоставляют информацию об электронном адресе, организации и почтовом адресе за определенную плату.

Крупнейшей англоязычной базой данных адресов можно назвать: <http://www.lookup.com/> -

представляет собой рубрикатор со множеством поддиректорий, таких как: Find Person, Phone Numbers, Name Search и другие.

Попадая в данные поддиректории, пользователь обнаруживает ссылки на сайты, которые и предлагают интересующую его информацию.

Широко доступных и официальных баз данных адресов в РФ нам неизвестно.

**2.6 Поиск в архивах Gopher (Gopher archives)**

Gopher – это взаимосвязанная система серверов (Gopher-пространство), распределенная по Интернет.

В пространстве Gopher собрана богатейшая литературная библиотека, однако материалы недоступны для просмотра в удаленном режиме: пользователь может только просматривать иерархически организованное оглавление и выбирать файл по названию. С помощью специальной программы (Veronica) такой поиск можно сделать и автоматически, используя запросы, построенные на ключевых словах.

До 1995 года Gopher являлся самой динамичной технологией Интернет: темпы роста числа соответствующих серверов опережали темпы роста серверов всех других типов Интернет. В сети EUnet/Relcom активного развития серверы Gopher не получили, и сегодня о них практически никто не вспоминает.

**2.7 Система поиска FTP файлов (FTP Search)**

Система поиска FTP-файлов – это особый тип средств поиска в Internet, который позволяет находить файлы, доступные на «анонимных» FTP-серверах. Протокол FTP предназначен для передачи по сети файлов, и в этом смысле он функционально является своеобразным аналогом Gopher.

Основным критерием поиска является название файла, задаваемое разными способами (точное соответствие, подстрока, регулярное выражение и т.д.). Данный тип поиска, конечно же, не может соперничать по возможностям с поисковыми машинами, так как содержимое файлов никак не учитывается при поиске, а файлам, как известно, можно давать произвольные имена. Тем не менее, если Вам требуется найти какую-нибудь известную программу или описание стандарта, то с большой долей вероятности файл, его содержащий, будет иметь соответствующее имя, и Вы сможете найти его при помощи одного из серверов [FTP Search](http://www.filesearch.ru/):

FileSearch ищет файлы на FTP-серверах по именам самих файлов и каталогов. Если Вы ищете какую-либо программу или еще что-то, то на WWW-серверах Вы скорее найдете их описание, а с FTP-серверов Вы сможете перекачать их к себе.

**2.8 Система поиска в конференциях Usenet News**

USENET NEWS – это система телеконференций сообщества сетей Интернет. На Западе этот сервис принято называть новостями. Близким аналогом телеконференций являются и так называемые "эхи" в сети FIDO.

С точки зрения абонента телеконференции, USENET представляют из себя доску объявлений, в которой есть разделы, где можно найти статьи на любую тему - от политики до садоводства. Эта доска объявлений доступна через компьютер, подобно электронной почте. Не отходя от компьютера, можно читать или помещать статьи в ту или иную конференцию, найти полезный совет или вступать в дискуссии. Естественно, статьи занимают место на компьютерах, поэтому не хранятся вечно, а периодически уничтожаются, освобождая место для новых. Во всем мире лучшим сервисом для поиска информации в конференциях Usenet является сервер Google Groups (Google Inc.).

Группы Google – это бесплатное интерактивное сообщество и служба групп обсуждений, которая предлагает самый обширный в Интернете архив сообщений сети Usenet (более миллиарда сообщений).Подробнее ознакомиться с правилами пользования сервисом можно на странице <http://groups.google.com/intl/ru/googlegroups/tour/index.html>

Среди русскоязычных выделяется сервер Всемирная система USENET и телеконференции Relcom. Точно также как и в других поисковых службах, пользователь набирает строку запроса, а сервер формирует список конференций, содержащих ключевые слова. Далее надо подписаться на отобранные конференции в программе работы с новостями. Также имеет место аналогичный российский сервер FidoNet Online: конференции Fido на WWW.

**2.9 Системы мета-поиска**

Для быстрого поиска в базах сразу нескольких поисковых систем лучше обратиться к системам мета-поиска.

Системы мета-поиска – это поисковые машины, которые посылают Ваш запрос на огромное количество разных поисковых систем, затем обрабатывают полученные результаты, удаляют повторяющиеся адреса ресурсов и представляют более широкий спектр того, что представлено в сети Интернет.

Наиболее популярная в мире система мета-поиска [Search.com](http://www.search.com).

Объединенный поисковый сервер Search.com компании CNET, Inc. включает в себя почти два десятка поисковых систем, ссылками на которые пестрит весь Интернет.

С помощью данного вида поисковых инструментов пользователь может искать информацию во множестве поисковых систем, однако отрицательной стороной данных систем можно назвать их нестабильность.

**2.10 Системы поиска людей**

Системы поиска людей – это специальные сервера, которые позволяют осуществлять поиск людей в Интернет, пользователь может указать Ф.И.О. человека и получить его адрес электронной почты и URL-адрес. Однако, следует отметить, что системы поиска людей, в основном, берут информацию об электронных адресах из открытых источников, таких как конференции Usenet. Среди самых известных систем поиска людей можно выделить:

[WhoWhere?](http://www.whowhere.com/) - поиск адресов e-mail

в специальные графы поиска контактные данные (First Name. City, Last Name, Phone number), Вы можете найти интересующую Вас информацию.

Системы поиска людей - это действительно большие сервера, их базы данных содержат порядка 6 000 000 адресов.

**3. Заключение**

Мы рассмотрели основные технологии поиска информации в Интернет и представили в общих чертах поисковые инструменты, которые существуют на данный момент в Интернете, а также структуру поисковых запросов для наиболее популярных русскоязычных и англоязычных поисковых систем и, подводя итог вышесказанному, хотим отметить, что единой оптимальной схемы поиска информации в Интернет не существует. В зависимости от специфики нужной Вам информации, Вы можете использовать соответствующие поисковые инструменты и службы. А от того, как грамотно будут подобраны поисковые службы, зависит качество результатов поиска.

Теги: [Поисковые системы](http://www.seonews.ru/tags/poiskovye-sistemy/) [Новичкам](http://www.seonews.ru/tags/novichkam/) [Яндекс](http://www.seonews.ru/tags/yandeks/) [Рамблер](http://www.seonews.ru/tags/rambler/)