

Поисковые системы Интернет



Кемь

2009 год

Составитель: Л. Проскурякова

зав. информационно-библиографическим отделом

Наш адрес: г. Кемь, пр. Пролетарский, д. 34

Межпоселенческая районная библиотека

Информационно-библиографический отдел

От составителя

Одно из заблуждений, бытующее у пользователей Интернета, гласит, что в Интернете найдётся всё. Это не так по двум причинам: во-первых, в Интернете может найтись лишь то, что туда кем-то было положено. В Интернете никогда не появится информация о том, почему вы вчера поздно пришли домой и где вы были, если конечно вы сами об этом не позаботитесь, оставив след о своих похождениях во Всемирной Паутине. Во-вторых, для того, чтобы найти в Интернете то, что там наверняка есть, нужно обладать определёнными навыками поиска. Сеть Интернет растёт очень быстрыми темпами, поэтому найти нужную информацию среди сотен миллиардов Web-страниц и сотен миллионов файлов становится все сложнее. Для поиска информации используются специальные поисковые системы, которые содержат постоянно обновляемую информацию о местонахождении Web-страниц и файлов на сотнях миллионов серверов Интернета.

Поисковые системы содержат тематически сгруппированную информацию об информационных ресурсах Всемирной паутины в базах данных. Специальные программы-роботы периодически "обходят" Web-серверы Интернета, читают все встречающиеся документы, выделяют в них ключевые слова и заносят в базу данных Интернет - адреса документов.

Что и как можно искать в Интернете? Чем при этом пользоваться, и какие ошибки следует избегать во время поиска? На все эти вопросы поможет ответить памятка «Поисковые системы Интернет», которая располагает информацией, позволяющей сориентироваться в многообразии поисковых систем Интернет и выбрать приемлемое для вас средство поиска необходимой информации. Практические советы, предложенные в памятке, помогут сделать ваш поиск наиболее результативным.

Поисковые системы Интернет

Что же такое поисковые системы? Это Интернет - порталы, которые предназначены для того, чтобы обеспечить их посетителю возможность поиска

сайтов, на которых встречаются заданные слова или целые фразы. Поисковые системы хороши в тех случаях, когда вы точно знаете, что вам нужно, когда вы можете быть уверены, что если на какой-то странице присутствуют те или иные слова или фразы, то это как раз то, что вы ищете.

Как это работает? Как мы уже говорили, порталы - это целый набор серверов, которые обеспечивают работу единой системы какого - либо Интернет-сервиса. В данном случае нас интересует служба поиска сайтов по заданным словам. Это значит, что поисковая машина должна знать, где и на каком сайте встречаются те или иные слова. Согласитесь, если бы поиск по сайтам начинался в момент, когда мы жмём на кнопку «Найти», то мы бы не дождались его завершения и через несколько дней. Сейчас интернет насчитывает более 1 терабайта информации. Это примерно $2 \cdot 10^8$ страниц машинописного текста. Если просматривать эти страницы по 10 в секунду, то на это уйдёт чуть меньше года! Как же получается, что поисковые системы предоставляют нам результат за несколько секунд?

Для этих целей существуют специальные поисковые роботы, или, как их ещё называют, пауки. Эти роботы представляют собой сервера, на которых запущена программа сканирования сайтов. После того, как владелец сайта регистрирует свой ресурс на поисковом сервере, робот запишет этот адрес в очередь, и через несколько дней зайдёт на этот сайт и проиндексирует его. Все слова, найденные на страницах сайта, будут занесены в специальную базу данных, отсортированы соответствующим образом, чтобы их поиск по этой базе занимал как можно меньше времени. И когда вы в строке поискового запроса набираете нужное вам слово, поиск на самом деле ведётся не по всему Интернету, а лишь по базе данных поискового сервера. Периодически «паук» повторно обходит известные ему страницы, чтобы обновить сведения о них. Таким образом, результаты поиска будут достаточно актуальны, но в пределах времени, необходимого роботу для повторной проверки ресурса.

Далее, чтобы предоставить вам информацию, поисковая машина должна её как-то классифицировать и упорядочить. Вот здесь и начинаются «секретные» технологии. Дело в том, что поисковые серверы пытаются предугадать, насколько

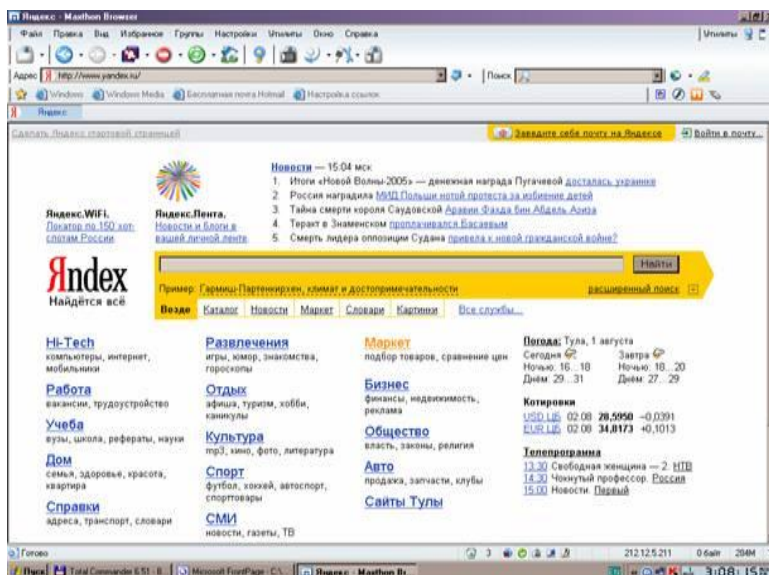
тот или иной сайт соответствуют введённому вами запросу. Это соответствие называется релевантность. У каждой поисковой системы свои собственные методы оценки релевантности. А держатся они в строжайшем секрете не только для обеспечения своего авторского права, но и для того, чтобы веб - мастера, создающие сайты не смогли чётко вычислять факторы, влияющие на то, чтобы сайты оказывались среди первых в списке найденных результатов. Тогда они будут искусственно фабриковать эти факторы, и страница не будет реально отвечать введённому слову или фразе.

Российские поисковые системы

Русскоязычным пользователям, которым нужно найти что-то в Рунете, лучше всего начинать поиск в Интернете с российских поисковых систем. И, в первую очередь, с поисковой машины Яндекс. Для следующих этапов поиска подойдут Rambler и Mail.ru. Во многих случаях уточнить результаты поиска в Яндекс можно с помощью глобальной поисковой системы Google.

Поисковая машина Яндекс

Русская буква «Я» среди латинских в названии - не ошибка, а волеизъявление авторов системы. В английском варианте название более читаемо - Yandex. Но в любом случае перед вами весьма перспективная поисковая машина (слово «машина» - это также изыск авторов), которая находится по адресу <http://www.yandex.ru>



Как указывается в документе на сервере, продукты Yandex предназначены для работы с большими объемами русских текстов всех типов – в виде файлов различных форматов, полей баз данных и страничек Интернет.

Поисковая машина Yandex была запущена в эксплуатацию в конце сентября 1997 г. и область поиска системы – русскоязычный Интернет. **Отличительные особенности системы, в соответствии с описанием на <http://yandex.ru>:**

- ✘ полнотекстовый поиск с полным учетом морфологии русского и английского языков;
- ✘ мощный язык запросов (в том числе возможность поиска с расстоянием);
- ✘ возможность запроса на естественном языке;
- ✘ корректная обработка сленговых слов;
- ✘ возможность поиска похожего документа;
- ✘ очень компактный индекс;
- ✘ подсветка слов из запроса в найденных документах;
- ✘ развитая релевантность, позволяющая найти искомую информацию, не перегружая пользователя тысячами ссылок на не нужные документы;
- ✘ высокая скорость поиска.

Как работает Yandex

Система просматривает все указанные ей тексты, преобразует каждое русское слово в нормальную форму (для существительных – это именительный падеж единственного числа, для глаголов – неопределенная форма и т.д.) и запоминает подробный адрес каждого слова. Алгоритмы морфологического разбора, которыми мы занимаемся, позволяют проанализировать слово, определить его характеристики и найти все формы, например: идти – идешь – шел; ребенок – дети; окно – окон; отзывать – отзывали.

Yandex обеспечивает индексацию одновременно с морфологическим разбором. Это позволяет создавать компактный индекс - около 30% исходных текстов, сохраняя достаточно высокую скорость индексации - 1-2 Мб/мин. Такая технология дает возможность почти полного снятия омонимии на этапе индексации.

Морфология базируется на словаре из 90 тыс. слов, к которому добавлены алгоритмы словообразования, а также на механизме построения гипотез для слов, отсутствующих в словаре. Создан и пополняется словарь имен собственных.

Поисковая машина Яндекс обладает самым быстрым механизмом обновления и поиска новой информации, можно примерно считать, что для этого требуется 1-2 недели. Но полностью весь Рунет она не просматривает, хотя количество выдаваемых ссылок максимально среди остальных поисковых систем.

Кроме поисковой машины на Яндекс вы найдете каталог интернет -ресурсов и другие интересные сервисы. Например, в ряде случаев может быть интересна система «Закладки», которая разрешает пользователям сохранять личные закладки в каталоге, размещенном в Интернете.

Информационно-поисковая система Rambler

Информационно-поисковая система Rambler («Праздношатающийся»), несмотря на англоязычное название, изначально была создана для поиска по серверам русского сектора Сети, ее адрес в Интернете - <http://www.rambler.ru>. Содержит миллионы документов, работает с 1996 года. До недавнего времени она была единственной из всех российских поисковых систем, которая признавалась фирмой Microsoft.



Рамблер - единственный в российском Интернете портал, объединивший поисковую систему, рейтинг-классификатор, а также ряд бесплатных сервисов и информационных проектов. Ресурсы портала регистрируют ежедневно более 3,5

млн. посещений, а ежемесячная аудитория Рамблера составляет 60-70% всех пользователей Рунета.

Только поисковая система содержит информацию о более чем 12 миллионах документов, расположенных на серверах России и стран СНГ. Рамблер обрабатывает ежедневно не менее 500 тысяч поисковых запросов (в среднем – 5 запросов в секунду), сканируя 48 тысяч web-серверов и используя несколько одновременно работающих программ-роботов. Все проекты Интернет - холдинга реализованы на базе собственных технологических разработок.

Поисковая система Rambler поддерживает различные логические операции между словами, а также усечения слов с помощью метасимволов. Пользователь может определить выходной формат документов, максимальное количество результатов поиска и т. д.

Другие поисковые системы России

Приведем перечень наиболее популярных адресов, не упоминавшихся ранее. Порядок следования поисковых систем и каталогов примерно соответствует их популярности.

Mail.ru

<http://mail.ru/>

WWW.ru

<http://www.ru>

Кирилл и Мефодий

<http://www.km.ru/>

Иван Сусанин

<http://www.susanin.net/>

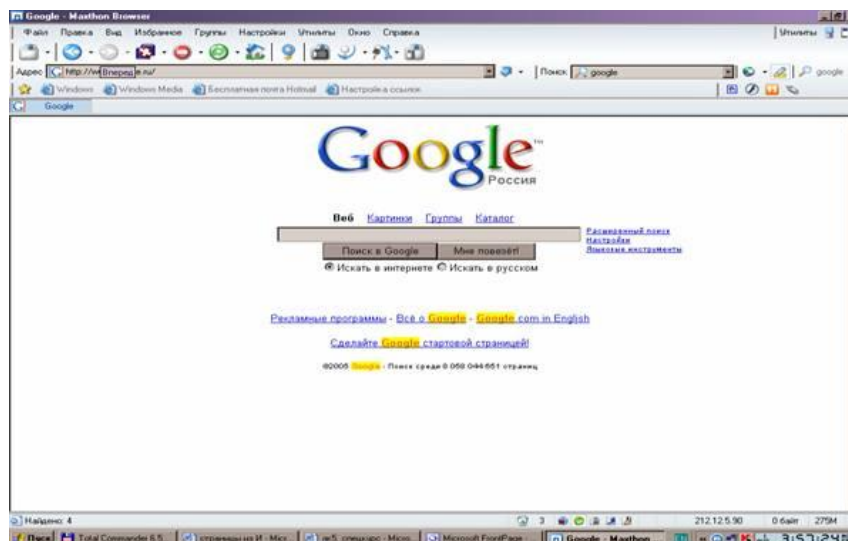
Особенно хочется отметить интернет-каталог Mail.RU (ранее List.RU). Его рейтинговая система наиболее популярна после Rambler в России, а баннеры-счетчики Mail.RU вы встретите не только на многих сайтах Рунета, но и на зарубежных публичных серверах.



Зарубежные поисковые системы

Google

В последнее время самой популярной глобальной поисковой системой стала Google (<http://www.google.com> и <http://www.google.com.ru>). Созданная русским программистом, переехавшим жить и работать в Кремниевую Долину, она отлично справляется не только с англоязычными документами, но и с мешаниной кодировок кириллицы. С 2003 года Google предлагает пользоваться своим сервисом чуть ли не на сотне различных языков.

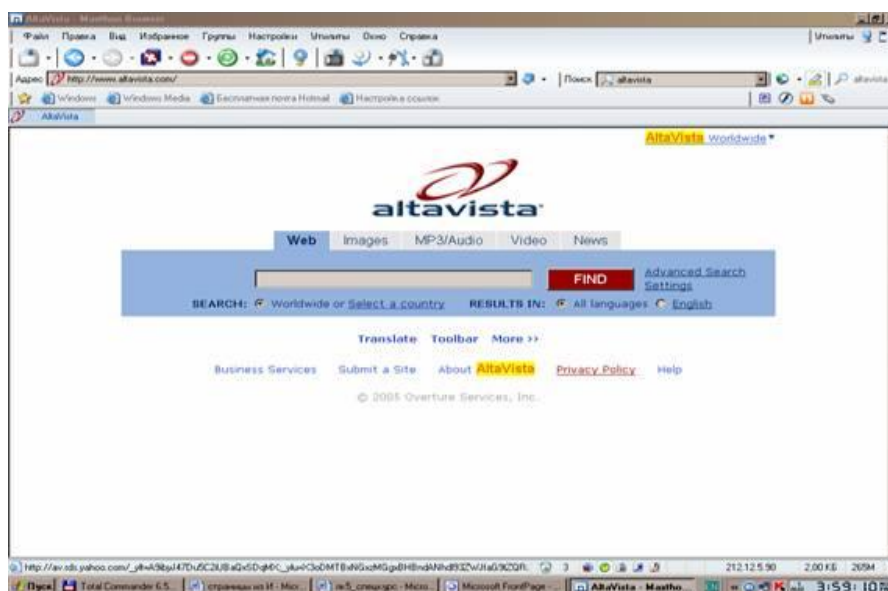


Google обладает уникальными возможностями, так как не хуже опытного хакера отыскивает в сети все доступные документы, даже те, к которым закрыт публичный доступ. Результаты поиска по Рунету у нее оказываются более

полными, чем у Яндекс. Можно даже рекомендовать после поиска в Яндекс щелкнуть внизу на странице результатов на ссылке Google, что автоматически передаст запрос из Яндекс в Google.

AltaVista

Самая именитая поисковая система - AltaVista, обладающая наиболее быстрым механизмом поиска, а также самой большой базой данных о ресурсах современного Интернета. Но, к сожалению, у AltaVista медленный темп обновления базы данных, но это беда всех систем, претендующих на глобальный охват Интернета. Первая страница AltaVista находится по адресу: <http://www.altavista.com>



Заметим, что эта наиболее авторитетная поисковая система начала работать в декабре 1995 года.

В AltaVista существуют два варианта запроса: простой (Simple Search) и расширенный (Advanced). Можно задать страну, серверы которой надо просмотреть для составления ответа. Кроме индексирования веб-страниц.

Yahoo!

Yahoo! - один из самых популярных тематических каталогов. Информация в нем организована по темам и подтемам, что в сочетании с отличной поисковой

системой дает возможность успешно использовать Yahoo! В качестве отправной точки для поиска нужного документа в Интернете. Каталог Yahoo! имеет национальные разделы ряда стран, ссылки на которые можно найти внизу главной страницы.

Первая страница Yahoo! расположена по адресу: <http://www.yahoo.com>, где вы получаете доступ к двум основным методам работы с каталогом - поиску по ключевым словам и иерархическому дереву разделов (категорий).



При путешествии по каталогу вы увидите путеводные значки, например:

- ✘ значки с надписями New приглашают познакомиться с новыми, недавно появившимися в каталоге гиперссылками;
- ✘ число в скобках информирует о количестве гиперссылок в подразделе;
- ✘ символ @ означает, что данный подраздел относится к другому разделу каталога.

Как улучшить результаты поиска



Очень часто поисковый сервер возвращает более скромный, чем хотелось бы, список документов. Или наоборот, список полученных ссылок огромен, но они совершенно не относятся к искомой теме. Прежде чем переходить к другому поисковому серверу или впадать в уныние, попробуйте улучшить результаты поиска. Если вы работаете через интерфейс Internet

Explorer, то выйдите из этого режима и загрузите заглавную страницу одного из поисковых серверов в окне просмотра. Затем постарайтесь внимательно изучить те страницы поискового сервера, где содержатся советы по организации поиска и различного рода подсказки. Они содержат множество примеров, которые помогут вам правильно применить то или иное правило. К сожалению, практически на каждом сервере применена своя система правил (синтаксис команд), определяющая поиск фраз из нескольких слов, применение булевых операций и т. д. Поэтому конкретные правила, действующие на том или ином сервере, вам лучше изучить самостоятельно.

Сформулируйте ваш запрос по возможности точнее. Хотя ряд лидирующих поисковых машин позволяет вам задать вопрос на естественном языке, этой возможностью следует пользоваться с осторожностью. Дело в том, что в таком случае поисковый сервер «отбрасывает» все незначимые по его мнению слова, кроме тех, которые можно интерпретировать как ключевые слова. Однако существует вероятность, что при этом будет отброшено и то слово, которое важно для результатов поиска. Поэтому по возможности старайтесь пользоваться четкими и ясными ключевыми словами, наиболее полно характеризующими предмет вашего поиска.

Проверьте правильность написания ключевых слов. Если результаты поиска по какой-либо теме полностью отрицательны, проверьте орфографию введенных вами терминов. Очень удобно, если на вашем компьютере установлен MS Word из комплекта Office 2003 (или теперь Office 2007). В этом случае можно не рыться в словарях, а ввести вашу фразу в Word, который подчеркнет неправильно набранные слова. Если теперь щелкнуть правой клавишей мыши на подчеркнутом слове, то Word предложит вам его правильное написание.

Используйте синонимы. Во многих предметных областях можно улучшить результаты поиска, если задать поиск по синонимам. Это особенно касается областей знания, где в силу новизны предмета еще не полностью устоялась терминология (в том числе в области компьютерных технологий).

Используйте опцию «расширенного поиска» (Advanced Search). На многих серверах имеется режим «расширенного» или «сложного» поиска, где действуют специальные синтаксические правила.

Постарайтесь локализовать поиск. Во многих случаях, когда предмет поиска можно локализовать географически, полезно выполнить поиск на серверах местных организаций, например, на серверах университетов. Там находится огромное количество ссылок на местную прессу, компании и общественные организации, работающие в этом регионе и т. д. По разным причинам поисковые серверы не индексируют значительную часть таких документов, и они могут быть найдены только в результате кропотливого локального поиска информации.

Старайтесь сами определить наиболее вероятный узел, где хранится искомый документ или файл. При поиске файлов с драйверами, программными продуктами и др. имеет смысл, в первую очередь, обратиться на Web-сервер фирмы-производителя. На многих серверах существуют локальные поисковые системы, обеспечивающие поиск документов по этому серверу, и корпоративные распределенные системы поиска, работающие с несколькими тематически связанными серверами (такие, к примеру, как поиск по серверам издательской корпорации Ziff-Davis). Поэтому, если вы, например, хотите найти информацию по операционной системе Windows XP, то, в первую очередь, имеет смысл провести поиск по ключевым словам на сервере Microsoft (<http://www.microsoft.com>), драйверы к принтерам и сканерам Hewlett-Packard – на <http://www.hp.com>, а для поиска сведений, относящихся к OS/2 Warp Connect, логично обратиться к поисковой системе WWW-сервера IBM (<http://www.ibm.com>). Аналогичный подход можно использовать при нахождении информации по иным товарам и услугам - расписаниям авиакомпаний, маршрутам турагентств, и т.д.

Интернет - Intranet. Публично доступные документы, содержащиеся на различных серверах Интернета, составляют лишь небольшую часть того информационного богатства, которое находится на файл-серверах в локальных сетях корпораций, правительственных учреждений, исследовательских центров, университетов и т. д. Эта «подводная» часть сети, носящая название Intranet,

закрыта для посторонних пользователей. Однако ссылки на имена и адреса электронной почты конкретных людей - профессоров, исследователей, журналистов, а также на области их деятельности, как правило, можно найти на WWW-серверах. Завяжите с этими людьми электронную переписку, попросите прислать копии их статей или URL известных им документов по интересующей вас теме – в общем, налажьте информационный обмен. Это очень эффективный путь поиска информации, поскольку знания эксперта – это высшая степень владения информацией.

Несколько рекомендаций по составлению поисковых запросов:

- ✘ Не следует в запросе писать только одно слово, если конечно это не какое-то определённое название. Лучше составить запрос из нескольких слов, каждое из которых может находиться в искомом документе. Причём слова эти будут располагаться недалеко друг от друга (в пределах нескольких предложений).
- ✘ Если вы хотите, чтобы слова стояли друг за другом, напишите их в кавычках.
- ✘ Можно использовать комбинированные запросы, например "старый дом" покрашен. Это значит, что слова «старый» и «дом» должны стоять рядом, а слово «покрашен» будет где-то недалеко.
- ✘ Поиск производится с учётом морфологии. Описанному выше запросу будут соответствовать фразы «старые дома», «Старый дома» и т.п. Если вы хотите произвести поиск без учёта морфологии, поставьте перед словом, которое не должно изменяться восклицательный знак.
- ✘ Не пишите слова заглавными буквами. Если вы вводите строчные буквы, то в результате окажутся слова, встречающиеся как заглавными, так и строчными буквами. Если же вы, например, написали слово с заглавной буквы, то найдены будут только те страницы, где заглавные буквы вашего слова совпадают с такими же в тексте. С заглавной буквы необходимо начинать запрос, если это слово - имя собственное.

✘ На каждом поисковом сайте есть раздел помощи, зайдя в который можно почитать об особенностях синтаксиса составления запросов. При помощи специальных символов можно составлять достаточно сложные запросы, в которых указывать не только нужные слова, их формы, но и слова, которые не должны присутствовать на странице. Можно указывать, например, что искомые слова должны встречаться только в заголовках, или только в тексте ссылок.

Краткий словарь терминов по Интернет

ADSL (Asymmetric Digital Subscriber Line) - Ассиметричная абонентская цифровая линия. Технология ADSL предполагает передачу данных с использованием цифровых модемов по обычным телефонным проводам. Данные передаются по кабелю уже в цифровом виде. При этом в момент соединения по Интернет сохраняется возможность пользоваться этим же кабелем для телефонных переговоров.

CMS (Content Management System) - система управления содержанием сайта. Представляет собой программные средства для подготовки, редактирования и публикации информации на сайте, а также средства для управления функциональностью сайта.

FTP (File Transfer Protocol) - протокол передачи файлов из семейства TCP/IP, обеспечивающий возможность найти, получить или отправить файлы через Интернет с одного компьютера на другой. Как правило, используется при передачи файлов большого объема.

E-mail – см. Электронная почта.

GIF (Graphic Interchange Format) - формат представления графических изображений. Получил наибольшее распространение в Интернет, за счет возможности хранения изображений, имеющих до 256 цветов, поддерживания прозрачности, анимации и способности сохранения в одном файле нескольких изображений. GIF имеет хороший алгоритм сжатия, что крайне важно для создания компактных графических файлов.

Home page - см. Главная страница.

Hosting – см. Хостинг.

HTML (Hyper Text Markup Language) - язык разметки гипертекста, позволяющий с помощью управляющих меток (тэгов) определять структуру и внешний вид HTML-документа (web-страницы) при отображении в браузере, а также создавать ссылки на другие файлы.

HTTP (Hyper Text Transfer Protocol) - протокол, обеспечивающий взаимодействие пользователя, запрашивающего доступ к web-документам, с сервером, предоставляющим возможность такого доступа.

ICQ (I Seek You) - приложение Интернет, используемое для прямого интерактивного общения между пользователями. Посредством ICQ возможен обмен текстовыми сообщениями, пересылка файлов, участие в коллективных играх и др.

Intranet - сеть в масштабах организации (группы организаций), построенная по Интернет - технологии.

IP (Internet Protocol) - протокол, обеспечивающий доставку данных в виде пакетов, снабженных IP-адресом.

IP-адрес - числовой идентификатор, который присваивается каждому компьютеру (хосту), подключенному к Интернету. IP-адрес состоит из адреса сети и адреса данного хоста в этой сети и представляет собой четыре десятичных числа (от 0 до 255), разделенных точкой. Например: 217.174.97.59.

JPEG (Joint Photographic Experts Group) - файлы растровых изображений с использованием до 16,7 миллионов цветов (24-битовый цвет), сжатие сопровождается потерей информации (отбрасываются мелкие детали). Популярный в Интернет графический формат. Для хранения изображения JPEG использует алгоритм сильного сжатия с потерей данных, не позволяет использовать анимацию и прозрачность. Как правило, применяется для хранения высококачественных фотографий.

JPG - см. JPEG

Portal - см. Портал.

PHP (Personal Home Page) - средство создания интерактивных web-страниц, главным достоинством которого является возможность формирования страниц посредством интерактивного взаимодействия «клиент-сервер». Включает CGI-интерфейс, интерпретатор языка и набор функций для доступа к базам данных и различным объектам Всемирной паутины.

Search engine - см. Машина поисковая.

Site - см. Сайт.

Spam – незапрашиваемые и нежелательные сообщения, приходящие по электронной почте. Как правило, рассылка производится одновременно по множеству адресов с целью рекламы товаров и услуг.

URL (Uniform Resource Locator) - Интернет-адрес, присвоенный каждой web-странице. Каждый URL в Интернет уникален.

Web-браузер – см. Браузер.

Web-сайт - см. Сайт.

Web-сервер - компьютер со специальным программным обеспечением, обеспечивающий доступ многих пользователей к расположенной на нем информации.

Web-страница (HTML-документ) - логическая единица Интернет (точнее, Всемирной паутины), однозначно определяемая адресом (URL). Физически представляет собой HTML-файл. Может содержать текст, изображения, аудио- и видеофрагменты, Java-апплеты и другие элементы. Web-страница может быть статической или динамически сгенерированной (примеры динамических страниц - перечни результатов, выдаваемые поисковыми машинами). В случае использования фреймов, каждый фрейм рассматривается в качестве отдельной

страницы. Страницы загружаются и просматриваются пользователем на свой компьютер с помощью браузера. Логически связанная совокупность web-страниц образует сайт.

Web-узел - см. Сайт.

Wi-Fi - (синоним: 802.11b) - стандарт высокоскоростной передачи данных по беспроводным сетям (от 11 до 20 Мегабит в секунду).

WYSIWYNG (What You See Is What You Get) - "что вы видите, то и получаете". Обозначение типа HTML-редакторов, которые позволяют пользователю создавать страницы без знания HTML. Примеры: Front Page, Netscape Composer.

Автоматизация сайта - создание программных решений, позволяющих владельцу сайта самостоятельно управлять содержимым и параметрами работы этого сайта.

Базы данных - специальное программное обеспечение, предназначенное для организации хранения и доступа к данным (информации). Используются при создании программных решений для автоматизации сайта.

Баннер - статичное или анимированное изображение, размещаемое на страницах сайта с целью рекламы чего-либо. Существуют также баннеры, использующие технологию Flash, HTML.

Баннерная реклама - размещение баннеров на сайтах Интернет. Возможно размещение путем взаимного обмена баннерами, путем участия в баннерообменных службах или путем покупки баннерных показов. Отличается возможностью выбора времени показа и определения целевой аудитории по интересам и территориальному признаку. Оптимизация графики - максимальное уменьшение размера файлов с изображениями при сохранении приемлемого их качества. Используется в любой серьезной системе автоматизации сайта для обеспечения высокой скорости загрузки его страниц.

Браузер (Browser) - клиентская программа для работы во Всемирной Паутине (WWW). Позволяет пользователю просматривать содержание web-страниц. Браузер обращается к web-серверу (сайту), запрашивает HTML-документ, интерпретирует полученную информацию и отображает документ на экране компьютера. Браузеры делятся на графические и текстовые. Последний вариант браузеров, примером которого является Lynx, в настоящее время практически полностью вышли из употребления. Примеры браузеров: Mosaic, Netscape Navigator, Internet Explorer, Opera, Mozilla.

Главная страница (Home page) - начальная (титульная) страница web-сайта. На главной странице размещаются общие сведения о сайте с указанием того, что представлено во всех его разделах. Внешние ссылки на ресурс, как правило, делаются именно на главную страницу, поэтому число ее посещений намного больше, чем любых других страниц сайта.

Домен (Domain) - см. Доменное имя.

Доменное имя (доменный адрес) - уникальный текстовый идентификатор компьютера (хоста), подключенного к Интернет. Состоит из слов, написанных латинскими буквами и разделённых точками. Пробелов и других знаков препинания в доменных именах нет. Каждому доменному имени соответствует определенный IP-адрес или несколько IP-адресов. Например, доменному имени www.rbc.ru соответствует IP-адрес 194.186.36.150. Доменные имена являются составляющей частью URL, указывающих на конкретные web-страницы. Доменные имена преобразовываются в IP-адреса службой DNS. Система доменных имен создана для удобства пользователей, которым легче запомнить доменный адрес (например, www.harvard.edu, www.fbi.gov, www.louvre.fr или www.ddt.ru), чем числовые значения IP-адресов. Регистрацией доменных имен занимается InterNIC (представитель в России - РОСНИИРОС). Регистрация доменного адреса означает внесение его и соответствующего ему IP-адреса в базу данных DNS-сервера.

Закладки (Bookmarks) - сервисная функция браузеров, позволяющая пользователю создавать перечень примечательных Интернет-ресурсов. Щелчок “мышью” по закладке вызывает загрузку данной страницы на компьютере пользователя. В браузере Internet Explorer закладки именуются “Избранное” (Favorites).

Индексация - лексический анализ и разбор текстовых материалов сайта с целью составления списка используемых слов и выражений. Составленный таким образом список используется при поиске информации на сайте по запросу пользователя. Индексация производится в автоматическом режиме специальными программами и используется всеми поисковыми системами.

Клик - (Click) - Щелчок кнопкой “мыши” по какому-либо элементу электронного документа, приводящий к загрузке определенной информации.

Кодировка (Encoding) - соответствие между символами (язык людей) и числами (язык компьютеров). Одни и те же числа можно превратить в символы различных алфавитов. Чтобы изменить правило превращения цифры в буквы, выберите в меню Вид Internet Explorer пункт Вид кодировки. В Internet наиболее используемы следующие кодировки русского алфавита КОИ8-Р и Win1251.

Контент - (Content) - содержание. Под данным термином чаще всего понимается содержательное наполнение электронных ресурсов, например, web-сайтов.

Куки (Cookies) - элемент данных, которыми web-сервер помечает конкретный браузер при его посещении. При следующем визите сервер уже “узнает” пользователя и может предложить ему информацию с учетом заявленных прежде пристрастий или, наоборот, не показывать клиенту те данные (например, рекламный баннер), которые он уже видел. Cookies не способны читать диск компьютера пользователя. Некоторые их значения хранятся только в течение одного сеанса работы с сервером и удаляются после закрытия браузера. Другие записываются в файл и хранятся на жестком диске в специальных директориях.

Кэш – (Cache - "тайный запас") - системная папка, в которую компьютер записывает все документы, полученные пользователем из Сети. При запросе документа вторично, показывается содержимое кэша. Наиболее эффективно кэширование, производимое прокси-сервером, который хранит документы, полученные из Интернет всеми сотрудниками организации. Обращение к кэшу в случае повторного запроса одного и того же документа позволяет не только снизить трафик, но и увеличивает скорость предоставления данных клиенту. Единственным недостатком кэширования является возможность получения старой версии документа в случае, если документ на удаленном сервере изменился, а кэш еще содержит старую версию. Прокси-сервер использует весьма сложный алгоритм определения степени устаревания документов, поэтому в большинстве случаев пользователь все же получает самую свежую версию документа. См. также Прокси-сервер.

Машина поисковая - (Search engine, синонимы: поисковая система, поисковый сервер) - инструмент для поиска информации в Интернет. Поисковые машины состоят из двух основных частей: программы-робота и индексного файла (базы данных). Робот сканирует сайты Интернет, выгружая web-страницы на базовый компьютер, где производится их индексирование. Учитываются не только тексты, но и иллюстрации, а также аудио- и видеофайлы. Пользователь, вводя состоящий из ключевых слов запрос, обращается к индексному файлу и получает в ответ перечень страниц, содержащих искомые термины. По территориальному охвату поисковые машины различаются на глобальные и локальные (национальные или отраслевые).

Портал - (Portal) – Интернет-сайт, предоставляющий максимально широкий спектр услуг, соответствующих потребностям среднестатистического пользователя Сети. В перечень предоставляемых сервисов, как правило, входят поиск информации, бесплатная электронная почта, новостная лента, прогноз погоды, сведения о курсах валют, перечень ссылок на сетевые ресурсы и др. В большинстве случаев у порталов есть также значительное собственное содержательное наполнение. Цель порталов – в создании места в Сети, с которого большое число пользователей регулярно начинает свою работу в Интернет и которые поэтому служат привлекательным местом для размещения рекламы. Порталами могут считаться сайты, имеющие стабильно высокую посещаемость. Большинство порталов сформировано вокруг поисковых машин и справочников Интернет-ресурсов. Помимо универсальных порталов, рассчитанных на всех пользователей без ограничения, существуют отраслевые (тематические) порталы, четко ориентированные на потребности определенной группы пользователей. Примеры: Yahoo!, Яндекс, KM.RU, AUTO.RU.

Посещаемость - количество посетителей web-сайта за определенную единицу времени. Является ключевой характеристикой при определении рейтинга сайта для участия в баннерных сетях и иных проектах по продвижению сайта. До настоящего времени единых подходов измерения посещаемости еще не выработано. Для оценки числа "уникальных посетителей" обычно используется анализ IP-адресов или функция Cookie. См. также Трафик.

Почтовый адрес - идентификатор почтового ящика пользователя. Образуется из имени пользователя и доменного имени почтового сервера, разделенных символом @ (коммерческая “эт”). Например: stepanov@shpl.ru, ivanov@inbox.ru, info@inion.ru. В интранет-сетях организаций регистрация почтового адреса выполняется системным администратором, на бесплатных почтовых серверах - самими пользователями.

Почтовый сервер - сервер, обеспечивающий прием-передачу и маршрутизацию персональных электронных писем пользователей. Организация почтового сервера требует установки на компьютер соответствующего программного обеспечения, например, Mdaemon. Электронная почта является основным средством общения в Интернет.

Прокси-сервер (от англ. Proxy - право действия от имени) - промежуточный web-сервер, используемый как посредник между браузером и конечным web-сервером. Прокси-сервер выполняет две основные функции: экономия объема трафика и увеличение скорости доступа к данным за счет кэширования данных на своем локальном диске и обеспечение защиты локальной сети корпорации, например, библиотеки. Использование прокси-сервера способно также привести к экономии IP-адресов, поскольку в этом случае необходим всего лишь один публичный IP-адрес. См. также Кэш.

Промоушн - от английского слова promotion (продвижение). Комплекс мер для рекламы ресурса (баненная реклама, offline-реклама, регистрация в каталогах и поисковых системах, поиск рекламных партнёров и т.п.).

Регистрация доменного имени - закрепление определенного доменного имени за физическим или юридическим лицом путем внесения соответствующей информации в регистрационную базу данных организации, координирующей распределение доменных имен. Для зоны RU такой организацией является Российский НИИ развития общественных сетей (RIPN). Адрес сайта этой организации - www.ripn.net. Для зоны COM имена распределяют несколько авторизованных для этого частных компаний.

Релевантность сайта - степень соответствия текста и тематики сайта слову или выражению, заданному в качестве ключа при поиске информации. Используется поисковыми системами для определения порядка выдачи результатов поиска. Например, при поиске с помощью поисковой системы Rambler страниц, содержащих выражение "Создание сайтов", страницы, которые имеют большую релевантность (больше соответствуют ключу "Создание сайтов") будут находиться ближе к началу списка результатов поиска.

Рунет - российский сегмент Сети. Включает сайты, физически хостирующиеся на территории России, а также русскоязычные и содержательно относящиеся к России зарубежные сайты.

Сайт (Site) - совокупность логически связанных web-страниц, размещенных, как правило, на одном компьютере.

Сервер – компьютер, который предоставляет свои ресурсы другим компьютерам сети, либо программа, которая обслуживает запросы на доступ к ресурсам своего компьютера.

Система управления сайтом - см. CMS.

Спам – см. Spam.

Статистика сайта - статистические данные, полученные путем учета информации о посещении страниц сайта. В простейшем случае включает информацию о количестве и времени посещений. Пример такой статистики - статистика рейтинговой службы Rambler Top100. Для применения в качестве исходных данных бизнес анализа требуется более серьезная система сбора и обработки статистики, позволяющая получать отчеты о посещаемости сайта и предпочтениях его посетителей. Система управления сайтом "Сайт Менеджер" позволяет получать такие статистические отчеты и является одной из лучших систем в своем классе.

Трафик - (Traffic) - поток (объем) информации, прошедший через канал связи или объем переданной/полученной информации. Измеряется в мегабайтах. Иногда в понятие трафик вкладывается число посещений web-сайта (или его конкретной страницы) за определенный промежуток времени. См. также Посещаемость.

Тэг - (Tag, синоним: метка) - команда языка разметки гипертекста. Большинство тэгов заключено в угловые скобки и употребляется в парах, то есть имеют открывающую и закрывающую метку, которые обозначают, соответственно, начало и конец области действия.

Хостинг - (Hosting, Web-hosting, Collocation, синоним - web-хостинг) - размещение web-сайта на внешнем сервере, в отличие от размещения на собственном компьютере, физически расположенном внутри организации (например, библиотеки). Существуют специальные хостинговые площадки для размещения сайтов. Как правило, это мощные серверы, подключенные к Интернет по высокоскоростным каналам связи. Наряду с коммерческим хостингом, существует бесплатный хостинг для размещения сайтов, как частных граждан, так и организаций. В последнем случае, платой за возможность быть представленным в Интернет, является размещение рекламных банеров на страницах сайта. Кроме того, бесплатные хостинги могут содержать ограничения по объему размещаемой информации и использованию различных приложений. В ряде случаев на хостинговой площадке размещается не только сайт, но и непосредственно сервер (компьютер) организации.

Хостер (хост-провайдер) - организация, профессионально занимающаяся предоставлением услуг хостинга. Рекомендуется выбирать услуги профессиональных хостеров, т.е. компании, основной деятельностью является хостинг.

Электронная коммерция - проведение бизнес транзакций с помощью Интернет. Включает в себя любые формы деловой активности в сети.

Электронная почта (E-mail) – приложение Интернет, позволяющее обмениваться сообщениями в электронной форме с помощью компьютерных сетей. Наиболее распространенный способ общения в Интернет. В настоящее время по электронной почте можно пересылать не только текст, но и все другие виды данных приложениями к письмам. Сами письма ныне в состоянии поддерживать все операции со шрифтами, включать фон, таблицы и иллюстрации.

МУ «Кемская межпоселенческая районная библиотека»
186615, Республика Карелия, г. Кемь,
Пролетарский пр., д. 34
тел.: 2-15-70, e-mail: kemlib@library.karelia.ru