

Памятка по поиску нужной информации в Интернете

В Интернете размещены миллионы сайтов, причем с актуальной информацией соседствует много устаревших ресурсов, мусора и недобросовестной рекламы. Интернет — это наиболее демократичный источник информации. Каждый может разместить в Сети собственный ресурс и высказать свое мнение. В этом одновременно сила и слабость Всемирной Сети. Плюсы наличия столь демократической площадки очевидны.

Однако отсутствие единоначалия и единых требований по оформлению информации, приводит к тому, что в Сети мало кто озабочен тем, чтобы избежать дублирования информации или следовать стандартам, принятым на сайте соседа и т.д.

Находить информацию в Интернете, вероятно, было бы очень трудно, если бы не были созданы мощные поисковые инструменты: поисковые машины (поисковики), каталоги (рубрикаторы), рейтинги, метапоисковые системы и тематические списки ссылок, онлайн-энциклопедии и справочники.

Индексированные каталоги

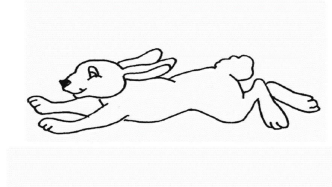
Каталог представляет собой данные, структурированные по темам в виде иерархических структур. Тематические разделы первого уровня определяют наиболее популярные, максимально широкие темы, такие как "спорт", "отдых", "наука", "магазины" и т.д. В каждом таком разделе есть подразделы. Таким образом, пользователь может уточнять интересующую его область, путешествуя по дереву каталога и постепенно сужая область поиска. Например, при поиске информации о ноутбуках цепочка поиска может выглядеть следующим образом:

Информационные технологии → Компьютеры → Ноутбуки.

Дойдя до нужного подкаталога, пользователь находит в нем набор ссылок. Обычно в каталоге все ссылки являются профильными, поскольку составлением каталогов занимаются не программы, а люди. Очевидно, что если ведется поиск общей информации по некоторой широкой теме, то целесообразно обратиться к каталогу. Если же необходимо найти конкретный документ, то каталог окажется малоэффективным поисковым средством.

Существует огромное количество каталогов. Один из наиболее популярных каталогов в России находится по адресу <http://list.mail.ru/>.

Помимо каталогов общего профиля в Сети достаточно много специализированных каталогов. Например, по адресу www.kinder.ru можно найти прекрасный каталог, посвященный детским ресурсам. В случае если внутри отдельной темы каталога



находится огромное количество ресурсов, возникает проблема выбора. В некоторых каталогах имеется сортировка по популярности, например в каталоге Яндекса сортировка идет по индексу цитирования – числу ссылок на сайт с других сайтов. Помимо каталогов в Сети существуют рейтинги. Одним из наиболее популярных рейтингов является [Rambler's Top100](http://top100.rambler.ru/top100/). (<http://top100.rambler.ru/top100/>).

Тематические коллекции ссылок

Тематические коллекции ссылок — это списки, составленные группой профессионалов или даже коллекционерами-одиночками. Очень часто узкоспециализированная тема может быть раскрыта одним специалистом лучше, чем группой сотрудников крупного каталога. Тематических коллекций в Сети так много, что давать конкретные адреса не имеет смысла.

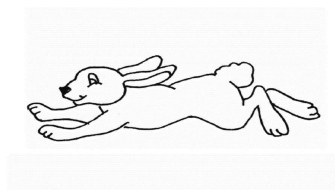
Подбор доменного имени

Каталог — удобная система поиска, однако для того чтобы попасть на сервер компании Microsoft или IBM, вряд ли имеет смысл обращаться к каталогу. Угадать название соответствующего сайта нетрудно: www.microsoft.com, www.ibm.com или www.microsoft.ru, www.ibm.ru — сайты российских представительств этих компаний.

Аналогично, если пользователю необходим сайт, посвященный погоде в мире, его логично искать на сервере www.weather.com. В большинстве случаев поиск сайта с ключевым словом в названии эффективнее, чем поиск документа, в тексте которого это слово используется. Если Западная коммерческая компания (или проект) имеет односложное название и реализует в Сети свой сервер, то его имя с высокой вероятностью укладывается в формат www.name.com, а для Рунета (русской части Сети) — www.name.ru, где name — имя компании или проекта. Подбор адреса может успешно конкурировать с другими приемами поиска, поскольку при подобной системе поиска можно установить соединение с сервером, который не зарегистрирован ни в одной поисковой системе. Однако если подобрать искомое имя не удастся, придется обратиться к поисковой машине.

Поисковые машины

Принцип работы поисковой машины сводится к тому, что она опрашивает свой внутренний каталог (базу данных) по ключевым словам, которые пользователь указывает в поле запроса, и выдает список ссылок, ранжированный по релевантности. Следует отметить, что обрабатывая конкретный запрос пользователя, поисковая система оперирует именно внутренними ресурсами (а не пускается в путешествие по Сети, как часто полагают неискушенные пользователи), а внутренние ресурсы, естественно, ограничены. Несмотря на то, что база данных поисковой машины постоянно обновляется, поисковая машина не может проиндексировать все Веб-документы: их число слишком велико.



Поэтому всегда существует вероятность, что искомый ресурс просто неизвестен конкретной поисковой системе.

Google и его поисковые сервисы

Поисковая система Google декларирует самый большой объем базы данных — более 8 млрд. документов. Google первой из поисковых машин начала индексацию документов в форматах PDF, PS, DOC, XLS, PPT, RTF и WP5. В последнее время Google расширила число не-HTML-файлов до 12 форматов, включив основные форматы Microsoft Office. Сейчас Google обеспечивает интерфейс на многих языках мира (более 100 языков), в том числе и на русском.

Фирменная кнопка Google — «I'm Feeling Lucky», или, в русской версии, «Мне повезет», отправляет пользователя на единственную страничку, которую Google выбрала на ваш запрос. Однако в этом случае других результатов поиска он не увидит. Кнопка «Мне повезет» помогает сэкономить время и действительно полезна, если вы уверены в том, что система легко может дать однозначный прогнозируемый вариант.

Google позволяет осуществлять поиск в различных источниках информации: во Всемирной паутине, в каталоге, в дискуссионных группах, а также искать картинки. Возможен также расширенный поиск, который позволяет разыскивать документ на разных языках и в различных форматах (PDF, DOC, XLS и т.д.), показывать страницы, измененные в течение определенного времени, и учитывать ряд других параметров.

Работая с Google, полезно знать несколько операторов, которые упрощают поиск. Например, для того чтобы получить толкование слова или сочетания слов, достаточно перед неизвестным термином (через пробел) напечатать «define». Так, задав запрос «define Интернет», вы получите специальную ссылку, предоставляющую всевозможные толкования термина в Сети (рис. 6).

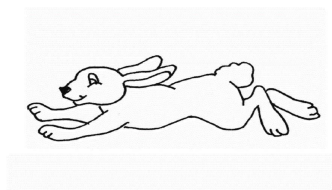
Слово «site», сопровождаемое двоеточием, позволяет ограничить поиск конкретным сайтом. Например, запрос «сеть site:www.hownetworks.boom.ru» позволяет найти документ со словом «сеть» на конкретном сайте.

Пример работы оператора define

По запросу типа link:siteURL вы получите страницы, которые ссылаются на данный URL. Например, запрос «link:www.compress.ru» выдаст список ссылок страниц, которые ссылаются на сайт www.compress.ru, а запрос «link:www.google.com» даст список страниц, которые ссылаются на стартовую страничку Google.

Если вы включите оператор «intitle:» в ваш запрос, то Google ограничит результаты документами, содержащими данные слова в заголовке. Например, «intitle:Гугл search» вернет документы, которые имеют слово «google» в заголовках, и слово «search» в любом месте документа.

Если вы начнете запрос с оператора «allinurl:», Google ограничит результаты ссылками, которые будут содержать слова запроса в url. Например, «allinurl: Гугл search» выдаст только те документы, которые имеют слова «google» и «search» в url.



Поиск Google в блогах доступен по адресу: www.google.com/blogsearch. Сервис имеет режим «Advanced Search», позволяющий указать язык, на котором будет произведен поиск, время, когда написан документ, и ряд других параметров.

Google ищет в блогах, написанных на разных языках. Его возможности поиска в русскоязычных блогах пока уступают возможностям Яндекса.

Google не ограничивает сферу поиска только сетью Интернет. Он разрабатывает сервисы, которые позволяют вести поиск на отдельно стоящем компьютере.

Google Desktop предоставляет простой доступ к информации на Вашем компьютере и в Интернете. Это приложение, предназначенное для поиска в компьютере, обеспечивает полнотекстовый поиск в электронной почте, файлах, музыке, фотографиях, чатах, Gmail и просмотренных веб-страницах, а также другие возможности. Эта программа делает поиск на компьютере таким же легким, как и в Интернете на сайте Google.

Скачать программу можно на сайте по адресу <http://desktop.google.com/ru/>.

Страница результатов поиска Google Desktop включает все типы элементов из своего индекса: файлы, сообщения электронной почты, веб-страницы, чаты и "другие" элементы, в число которых входят контакты, встречи, задачи, заметки и журнальные записи. На Вашей странице результатов также показано, сколько найдено результатов каждого типа.

Если Вам нужны результаты только одного типа, нажмите на соответствующую ссылку. Например, нажав ссылку 3 файла, Вы перейдете на страницу, на которой отображены только 3 файла, соответствующие запросу. Находясь на странице результатов одного типа, просто нажмите ссылку Все, чтобы вернуться на страницу всех результатов.

Результаты можно сузить еще больше. Нажав ссылку Файлы, Вы увидите параметры фильтрации файлов по определенным типам: форматы Word, Powerpoint, Excel, TXT, HTML, PDF, графические форматы (GIF, JPG, PNG, BMP), звуковые форматы (WAV, MP3, AAC) и видеоформаты (WMV, MPG, AVI). Нажав ссылку Сообщения электронной почты, Вы увидите параметры фильтрации сообщений по отправителю или получателю.

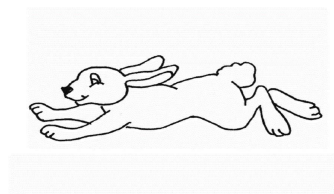
Настольный поиск Google можно использовать для обучения основам поиска и освоению операторов, которые используются при поиске различными поисковыми системами.

При выполнении поиска в компьютере можно использовать следующие операторы расширенного поиска.

Поиск по фразе. Если взять весь запрос или его часть в кавычки, то Desktop выведет только те результаты, которые содержат точное соответствие словосочетанию в кавычках.

Поиск "-" Когда оператор - стоит прямо перед словом, то из результатов поиска будут исключены элементы, содержащие это слово. Например, в результаты поиска по запросу [ворота -футбол] будут включены все элементы из кэша Desktop, которые содержат слово "ворота", кроме тех, в которых также содержится слово "футбол".

Поиск "site:" Если Вы включите в свой запрос оператор site:, то Desktop покажет только результаты, полученные с указанного веб-сайта. Например, поиск Desktop по запросу [справка site:www.google.com] вернет только просмотренные Вами страницы www.google.com, в которых содержится слово "справка". Примечание. Между оператором "site:" и следующим за ним веб-сайтом не должно быть пробела.



Поиск "filetype:" Типы файлов, отображаемые в результатах поиска Desktop, можно ограничить с помощью оператора filetype: и аргумента в виде расширения файла или полного названия приложения Office. Например, если Вы ищете [налог filetype:xls] или [налог filetype:excel], в результат будут включены только файлы Excel, содержащие слово "налог". Если Вы выполняете поиск [filetype:abc], в результат будут включены только файлы с расширением .abc. Как и в случае с оператором site: , между оператором filetype: и расширением или типом не должно быть пробела.

Поиск "under:" Оператор under: позволяет ограничить поиск только той папкой, результаты поиска файлов из которой Вас интересуют.

Поиск "machine:" Если функция "Поиск на других компьютерах" охватывает несколько компьютеров, оператор machine: позволяет указать компьютеры, с которых поступили результаты поиска. При поиске [отчеты machine:СЛУЖЕБНЫЙ] результаты будут содержать только элементы, найденные на компьютере с именем "СЛУЖЕБНЫЙ".

Расширенный поиск в сообщениях электронной почты. Можно ограничить свой поиск только заголовками сообщений электронной почты. Поддерживаются следующие операторы поиска в электронной почте: Subject: (тема); To: (кому); From: (от); Cc: (копия); Bcc: (скрытая).

В одном запросе можно комбинировать несколько ограничений для заголовка сообщения. Например, результаты по запросу [subject:дух from:денис] будут включать только сообщения, содержащие слово "дух" в поле "Тема:" и "денис" в поле "От:". Как и с другими операторами, между оператором заголовка сообщения и соответствующим запросом поиска не должно быть пробела.

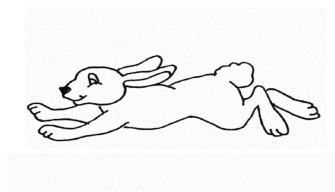
Другой интересный поиск на отдельной машине от Google - сервис Picasa <http://picasa.google.com/intl/ru/download/thanks.html> Picasa – это программное обеспечение, с помощью которого можно легко просматривать, упорядочивать, редактировать цифровые фотографии на персональном компьютере и обмениваться ими.

Яндекс и его поисковые сервисы

Яндекс — популярнейшая в России поисковая система, функционирующая с осени 1997 года. Поиск Яндекса позволяет искать по Рунету документы на русском, украинском, белорусском, английском, немецком и французском языках с учетом морфологии (Если вам необходимо искать без морфологии, можно заставить Яндекс не учитывать формы слов из запроса при поиске. Например, запрос !иванов найдет только страницы с упоминанием этой фамилии, а не города Иваново) русского и английского языков и близости слов в предложении.

Помимо Веб-страниц в формате HTML, Яндекс индексирует документы в форматах PDF (Adobe Acrobat), RTF (Rich Text Format), DOC (Microsoft Word), XLS (Microsoft Excel), PPT (Microsoft Power Point), SWF (Macromedia Flash), а также многие блоги и форумы.

Яндекс позволяет задать запрос по любому из источников информации («Каталог», «Новости», «Товары», «Адреса», «Словари», «Картинки»



Для каждого документа выдается ссылка на ресурс и ссылка [Показать найденные слова] — переход на подсвеченный документ, где контекст запроса выделен стрелочками, которые обеспечивают переход по выделенным словам.

При точном соответствии запросу система выводит результат поиска в новостных лентах информационных агентств, результат поиска в энциклопедических статьях или в базе товарных предложений магазинов.

Как и в Google, в Яндексe есть служба расширенного поиска. Она позволяет задать словарный фильтр, то есть указать, какие слова обязательно должны встретиться в документе, каких быть не должно, а какие — желательны. Поле «Все формы» или «Точная форма» указывает Яндексe, надо ли учитывать при запросе все словоформы. «Точная форма» обычно требуется только для поиска цитат. Зоной поиска слова может быть как текст документа (слова находятся в одном предложении или во всем документе), так и его заголовок, аннотация, ссылка (подпись URL) или адрес (сам URL). Вариант «Во фразе» означает необходимость искать слова в том порядке, в котором они введены. Вы можете задать несколько слов через запятую. Здесь же можно обозначить ограничение выдачи документов по дате, задать запрос только по страницам указанного сайта или поддиректории, найти ресурсы, которые ссылаются на некоторый Веб-сайт, найти документы, содержащие картинку с определенным названием или подписью.