

Поиск информации в сети Средства и способы поиска

Для решения стандартных, часто встречающихся поисковых задач сеть имеет стандартные же средства - мощные инструменты, разработанные именно для этой цели. Эти основные средства поиска работают автоматически, а значит, отвечают на запрос достаточно быстро и им не надоедает множество запросов от одного абонента.

Дополнительные способы поиска используют возможности, предоставляемые некоторыми другими службами сети, ее персоналом, а также ее абонентами, которые могут как-то способствовать поиску. Эти способы являются дополнительными, поскольку они:

- либо не предназначены для массового использования;
- либо не являются универсальными (накапливают адреса в недостаточном объеме или по узким направлениям);
- либо не являются стандартными или обязательными для того, кто их предоставляет (то есть, вообще говоря, нет гарантии, что на запрос хоть как-то ответят).

Характерным примером такого дополнительного способа поиска является публикация запроса ("крика о помощи") в соответствующей телеконференции - есть некоторая вероятность, что кто-нибудь отзовется по существу, но могут и промолчать.

Ниже в основном описываются средства и способы поиска, а круг решаемых ими задач ясен из описания. Это оказалось удобнее, чем сначала называть задачи, а затем перечислять пути их решения.

Если вы хотите, чтобы и вас могли найти при помощи поисковых служб Интернет, то не забывайте оставлять о себе информацию на поисковых серверах, которые предоставляют эту возможность. Большинство систем предлагают "зарегистрироваться", т.е. заполнить форму своими координатами. Обычно регистрация не обязательна, хотя бывают и исключения. Часто регистрация позволяет пользоваться дополнительными возможностями поиска.

White- и yellow- поиск

Для целей дальнейшего изложения следует различать *white-pouск* и *yellow-pouск*. Понимание того, что является исходными данными и что результатом, весьма полезно как само по себе, так и для ориентации среди многочисленных поисковых средств Интернет.

Под White-поиском понимается поиск адреса одного конкретного адресата по его достаточно определенному имени (человека - по фамилии, организации - по названию). "Достаточно определенное имя" означает, что объект поиска заранее известен: есть уверенность в его существовании и предлагаемое для поиска его имя вполне уникально.

Yellow-поиск предназначен для поиска информации не только по конкретному собственному имени адресата (человека, организации), но и по неким общим признакам, когда конкретное имя не определено, а есть

целая область поиска, например “фирмы, производящие легковые автомобили”, “что-нибудь о телескопах”.

Поиск по иерархическому классификатору и по ключевым словам

Информация в Интернет, в частности в WWW, часто ищется методом yellow-поиска с помощью специальных серверов. Эти поисковые серверы периодически “сканируют” содержимое WWW-документов в Интернет и составляют базы данных, позволяющие найти ссылки на документы с той или иной информацией.

Свяжитесь с поисковым сервером по одному из адресов, указанных ниже. Как правило, поисковый сервер поддерживает два метода поиска: *по ключевым словам* и *по иерархическому классификатору*.

Для осуществления поиска по ключевым словам сформулируйте не более трех слов, как можно более точно определяющих область вашего поиска. Введите эти слова в специальное окно ввода (вы увидите его на заглавной странице сервера) и нажмите кнопку *Go* (или *Поиск*, или *Submit*, и т.п.), располагающуюся тут же на странице. Сервер просканирует свою базу данных в поисках документов, содержащих указанные слова, и выдаст вам список ссылок на обнаруженные документы. Как правило, документы отсортированы в порядке убывания важности ключевых слов. Вместе со ссылкой выводится аннотация содержимого документа или его первые строки, что позволяет вам оценить, нужен вам этот документ или нет. Разумеется, вы можете воспользоваться любой из ссылок и получить заинтересовавший вас документ целиком.

Подбор ключевых слов - не совсем простая задача. Например, не стоит вводить очень распространенные или многозначные слова, т.к. они встречаются в большом числе документов, которые могут к вашей теме не иметь никакого отношения. Или, например, вас интересует конструкция телескопов, вы вводите ключевое слово “телескоп” и неожиданно получаете список ссылок на программу телевидения. Причина: слово “Телескоп” является также и названием ТВ-программы. Не вводите слова, значения которых являются подмножествами друг друга, например: “мухомор” и “гриб”, достаточно ввести просто “мухомор”.

Поиск по иерархическому классификатору представляет собой выбор из списка сначала большого раздела информации, потом его подраздела - и так далее, пока не будет достигнут желаемый результат. На каждом шаге вы будете видеть список ссылок на документы, тематика которых соответствует данному разделу, или список его подразделов. Например: “Наука” а “Астрономия” а “Приборы” а “Телескопы”.

Применение того или иного метода поиска определяется конкретной задачей.

Однако часто искомый адрес можно просто угадать - когда речь идет о white-поиске, т.е. вы ищете, например, домашнюю страницу какой-либо конкретной организации, группы и т.п. Например: фирма IBM - www.ibm.com, Белый дом США - www.whitehouse.gov и т.п. Интересно, что

такой метод иногда помогает и при yellow-поиске: “что-нибудь о группе Аквариум” - www.aquarium.ru, “что-нибудь о музыке” - www.music.ru, “что-нибудь об операционной системе BSD” - www.bsd.org, и т.п.

Язык запросов

Большинство поисковых серверов поддерживают формальный логический язык запросов в дополнение к естественному языку. Рассмотрим некоторые правила такого языка на примере сервера Яндекс. Ниже приводятся выдержки из документа

http://www2.yandex.ru:8081/ya_detail.html.

Как трактуются слова

Независимо от того, в какой форме вы употребили слово в запросе, поиск учитывает все его формы по правилам русского языка. Например, если задан запрос *идти*, то в результате поиска будут найдены ссылки на документы, содержащие слова *идти, идет, шел, шла* и т.д. На запрос *окно* будет выдана информация, содержащая и слово *окон*, а на запрос *отзывали* - документы, содержащие слово *отозвали*.

Если вы набрали в запросе слово с большой буквы, будут найдены только слова с большой буквы, в противном случае будут найдены как слова с большой, так и с маленькой буквы.

Например, запрос *лебедь* найдет и птицу, и генерала. Запрос *Лебедь* - генерала и те случаи упоминания птицы, когда она написана с большой буквы.

По умолчанию поиск учитывает все формы заданного слова согласно правилам русского языка. Однако существует возможность поиска по точной словоформе, для этого перед словоформой надо поставить восклицательный знак '!'. Так, по запросу *!Лужкову* будут найдены все документы, содержащие словоформу 'Лужкову', а по запросу

Лужков ~ !Лужкову

документы, в которых упоминается Лужков, кроме тех, которые были найдены по первому запросу.

Основные операторы

Несколько набранных в запросе слов, разделенных пробелами, означают, что каждое из них должно входить в один абзац искомого документа. Тот же самый эффект произведет употребление символа &.

Например, при запросе *холодная вода* (или *холодная & вода*) результатом поиска будет список документов, в которых в одном абзаце содержатся и слово 'холодный', и слово 'вода'.

Между словами можно поставить знак | (или запятую ','), чтобы найти документы, содержащие любое из этих слов. Запрос вида *вода | огонь* или *вода,огонь* задает поиск документов, содержащих в одном абзаце хотя бы одно из слов 'вода' или 'огонь'.

Еще один знак, тильда ~, позволит найти документы с абзацем, содержащим первое слово, но не содержащим второе.

По запросу *холодный ~ горячий* будут найдены все документы, содержащие слово 'холодный', рядом с которым (в пределах абзаца) нет слова 'горячий'.

Чтобы подняться на ступеньку выше, от уровня абзаца до уровня документа, просто удвойте соответствующий знак. Одинарный оператор (&, ~) ищет в пределах абзаца, двойной (&&, ~~) - в пределах документа.

Например, по запросу *холодная && вода* будут найдены документы, содержащие, где бы то ни было, оба эти слова. А запрос *квас ~~ теплый* выдаст все документы со словом 'квас', но без слова 'теплый'.

Поиск с расстоянием

Как Yandex-Web адресует слова? Если все слова в тексте перенумеровать по порядку их следования, то расстояние между словами а и b - это разница между номерами слов а и b. Таким образом, расстояние между соседними словами равно 1 (а не 0), а расстояние между соседними словами, стоящими "не в том порядке", равно -1. То же самое относится и к абзацам.

Если между двумя словами поставлен знак /, за которым сразу напечатано число, значит, требуется, чтобы расстояние между ними не превышало этого числа слов.

Например, задав фразу *холодная /2 вода*, вы требуете найти документы, в которых содержатся и слово 'холодный', и слово 'вода', причем расстояние между ними должно быть не более двух слов и они должны находиться в одном абзаце.

Если порядок слов и расстояние точно известны, можно воспользоваться пунктуацией */+n*. Так, например, задается поиск слов, стоящих подряд.

Запрос *холодная /+1 вода* означает, что слово 'вода' должно следовать непосредственно за словом 'холодный'. (Кстати, к тому же результату приведет запрос *холодная вода*).

В общем виде ограничение по расстоянию задается при помощи пунктуации вида */(n m)*, где n - минимальное, а m - максимально допустимое расстояние. Отсюда следует, что запись */n* эквивалентна */(-n +n)*, а запись */+n* эквивалентна */(+n +n)*.

Запрос *холодная /(-2 4) вода* означает, что 'вода' должна находиться от 'холодный' в интервале расстояний от 2 слов слева до 4 слов справа.

Практически все знаки можно комбинировать с ограничением расстояния.

Например, результатом поиска по запросу *холодный ~ /+1 вода* будут документы, содержащие слово 'холодный', причем в этих документах слово 'вода' не следует непосредственно за словом 'холодный'.

Когда знаки ограничения по расстоянию стоят после двойных операторов, то употребленные там числа - это расстояние не в словах, а в абзацах. Расстояние в абзацах определяется аналогично расстоянию в словах.

Запрос *холодная && /I вода* означает, что слово 'вода' должно находиться в том же самом либо в соседнем со словом 'холодный' абзаце.

Скобки

Вместо одного слова в запросе можно подставить целое выражение. Для этого его надо взять в скобки.

Например, запрос *(холодный,горячий) /+I (вода|огонь)* задает поиск документов, которые содержат любую из фраз 'холодная вода', 'холодный огонь', 'горячая вода', 'горячий огонь'.

Поисковые серверы в России

Rambler <http://www.rambler.ru/>

Яндекс <http://yandex.ru:8081/>

Созвездие Интернет <http://www.stars.ru/>

Ермак <http://www.medialingua.ru/>

Желтые страницы Интернет -

<http://proxy.menatop.elcom.ru/>

Апорт! <http://russia.agama.com/aport/>

Русский Интернет <http://www.rocit.ru/au/>

Russian Internet Search <http://www.search.ru/>

Russian FTP Search <http://ftpsearch.city.ru:8000/ftpsearch/>

Русскоязычные поисковые серверы

Altavista <http://www.altavista.com/>

EuroSeek <http://euroseek.net/>