

## Глава 1

# Поиск для нетерпеливых

*В этой главе...*

- ◆ Простой поиск с помощью браузера
- ◆ Каталоги и поисковые машины
- ◆ Рейтинги и порталы
- ◆ Анализ результатов простого поиска
- ◆ Недостатки простого поиска

## 1.1. Простой поиск с помощью браузера

Итак, вы подключились к Internet, потому что понимаете: эта глобальная Сеть — неиссякаемый источник информации любого рода (а также великолепное средство связи, но это — тема уже других книг). Нужно только уметь найти среди миллиардов файлов, хранящихся на миллионах компьютеров, те, в которых содержатся необходимые вам документы. Под словом “документ” здесь и далее понимается, конечно, не официальная бумага с подписью и печатью, а любой файл, содержащий полезную текстовую или графическую информацию. В принципе файл с понадобившейся вам записью музыкальной композиции тоже можно было бы назвать “документ”, но столь широкое толкование этого термина пока еще противоречит словарям. Тем не менее, дабы не упоминать каждый раз отдельно аудиофайлы, компьютерные программы, видеоклипы и другие “объекты”, которые вам могут понадобиться при работе в Internet, очень часто мы будем все их обозначать одним и тем же термином “документ”.

Конечно, как-нибудь на досуге вы пролистаете этот самоучитель и научитесь быстро и эффективно находить нужную вам информацию. Но вот беда: для чтения книги нужно время, а вам срочно (как обычно, еще “на вчера”) понадобилась информация. Что делать?

Открыть программу-браузер и провести *простой поиск*. Не исключено, что вы очень быстро найдете ту информацию, которая вам понадобилась.

Допустим, вам нужно получить какие-то сведения по теме “Автоматические выключатели концерна Moeller” (скажем, узнать технические характеристики этих выключателей). Для того чтобы быстро найти информацию по интересующей вас теме (или убедиться в том, что “с наскока” нужную информацию получить не удастся), выполните следующее.

1. Запустите программу Internet Explorer. Если вы до этого работали с ней в автономном режиме, снимите флажок Работать автономно в меню Файл.
2. Щелкните на кнопке Поиск в панели инструментов Internet Explorer. В левой части рабочего окна программы откроется панель Поиск (рис. 1.1). По умолчанию переключатель Вы-берете категорию поиска: должен быть установлен в положение Поиск веб-страницы.
3. В поле Поиск веб-страницы, содержащей: введите слова, которые обязательно должны быть в интересующем вас документе либо его заголовке. В нашем примере это

будут слова, достаточно полно характеризующие, по нашему мнению, предмет поиска: “Автоматические выключатели Moeller”. Щелкните на кнопке Поиск.

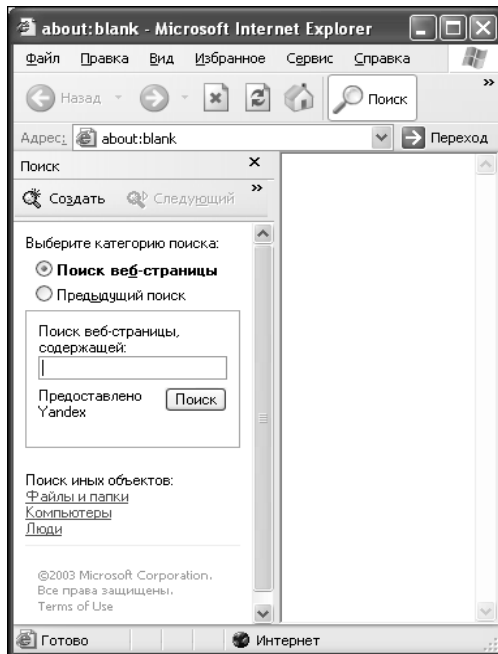


Рис. 1.1. Панель Поиск программы Internet Explorer

4. Если вы работали в автономном режиме, откроется окно Подключение удаленного доступа, и программа Internet Explorer начнет, как обычно, устанавливать соединение с сервером вашего провайдера. Если соединение было установлено ранее, сразу же начнется процедура поиска. В результате на панели Поиск появится перечень ссылок на Web-страницы, соответствующие сделанному вами *запросу на поиск*.
5. Щелкните на ссылке, наиболее соответствующей, по вашему мнению, предмету поиска. В результате в правой части окна Internet Explorer откроется Web-страница сайта, содержащая один из соответствующих вашему запросу документов (рис. 1.2).



Если вам срочно понадобилось найти в Сети какую-то информацию, воспользуйтесь инструментом Search (Поиск) вашего браузера. Введите слова, которые обязательно должны быть в искомым документах, в предназначенном для них поле ввода и щелкните на кнопке Search, Поиск, Go или другой, инициирующей процедуру поиска. Браузер обратится к одному из поисковых серверов, и вы быстро получите какие-то результаты.

Итак, введя в поле ввода поисковой машины слова, которые должны быть в искомым документах, вы тем самым сформулировали запрос на поиск. Поисковый сервер провел поиск информации в своей базе данных и выдал адреса Web-страниц или даже конкретных документов, наиболее точно, по его мнению, соответствующих сформулированному запросу.



Web-узел, специализирующийся на поиске информации, принято называть поисковым сервером, или информационно-поисковой системой (ИПС).

Первая же выбранная нами ссылка (но не первая в представленном списке!) вывела нас на страницу с подробной информацией об автоматических выключателях концерна Moeller. Осталось только открыть документ то ли с общими характеристиками, то ли с техническими параметрами выключателя конкретного типа. Все? Проблема поиска успешно решена?

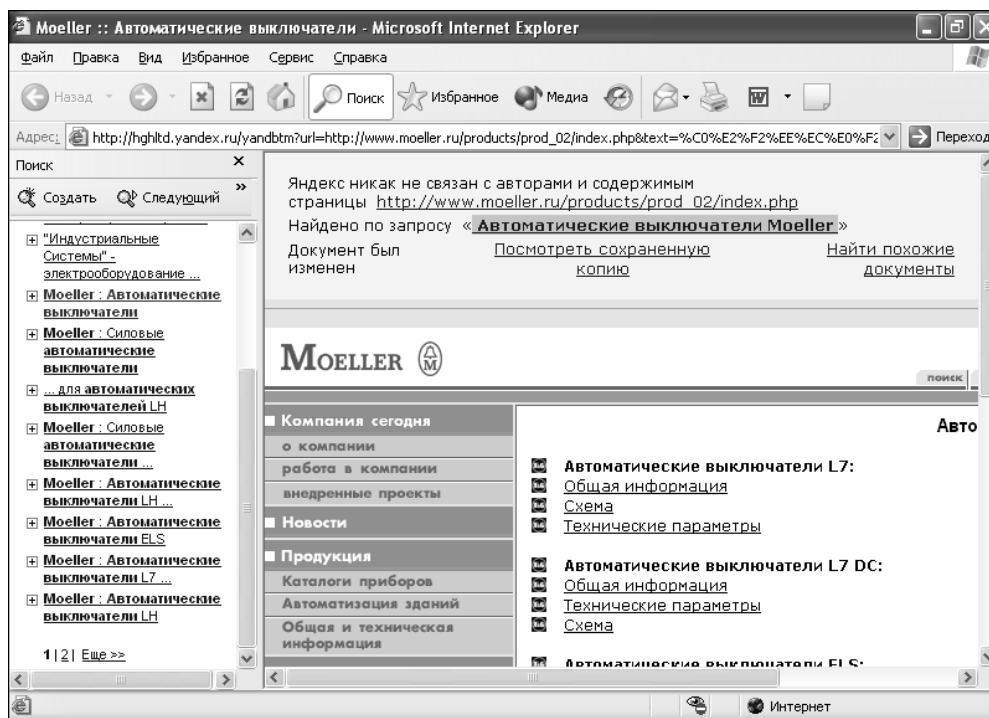


Рис. 1.2. Результаты поиска по запросу и одна из найденных Web-страниц

В данном конкретном случае — вполне возможно. Именно для таких случаев и предусмотрена кнопка Поиск в любом браузере. Но...

Не будем спешить с выводами. Мы специально составили запрос на такую тему, по которой можно быстро получить конкретный результат. Прочитав нашу книгу, вы научитесь составлять запросы именно такого типа по любой теме. Но давайте все же внимательно посмотрим на приведенные выше рисунки.

Во-первых, сразу становится ясно, что на самом деле поиск проводил не Internet Explorer, а один из самых популярных русскоязычных поисковых серверов Yandex. Кроме того, в нижней части панели Поиск вы видите цифры 1 | 2 и слово Еще с двумя символами “больше” (в дальнейшем мы будем называть их *две стрелки вправо*). Это означает, что все найденные ссылки не поместились на панели Поиск. Щелкнув на цифре 2, вы откроете вторую страницу со ссылками, на двух стрелках вправо — следующие. Сколько всего страниц со ссылками, Yandex умалчивает.

Если вы теперь щелкнете на кнопке Следующий панели Поиск, будет проведен еще один (следующий) поиск на заданную тему. Но теперь его будет проводить другой поисковый сервер — Rambler. Как видите, результаты поиска получились несколько разными, хотя нужный нам сайт также представлен в первой десятке (рис. 1.3). Это вызвано тем, что различные поисковые серверы используют разные алгоритмы и механизмы поиска, по-разному оценивают релевантность найденных ссылок (т.е. степень соответствия

найденных документов запросу пользователя). Более подробно о терминологии мы будем говорить в следующей главе. А пока отметим следующее: поисковый Web-узел Rambler сразу же указал, что всего им найдена 97 ссылка, показаны первые 15.



Релевантными принято называть документы, в наибольшей степени соответствующие потребностям пользователя, проводящего их поиск.

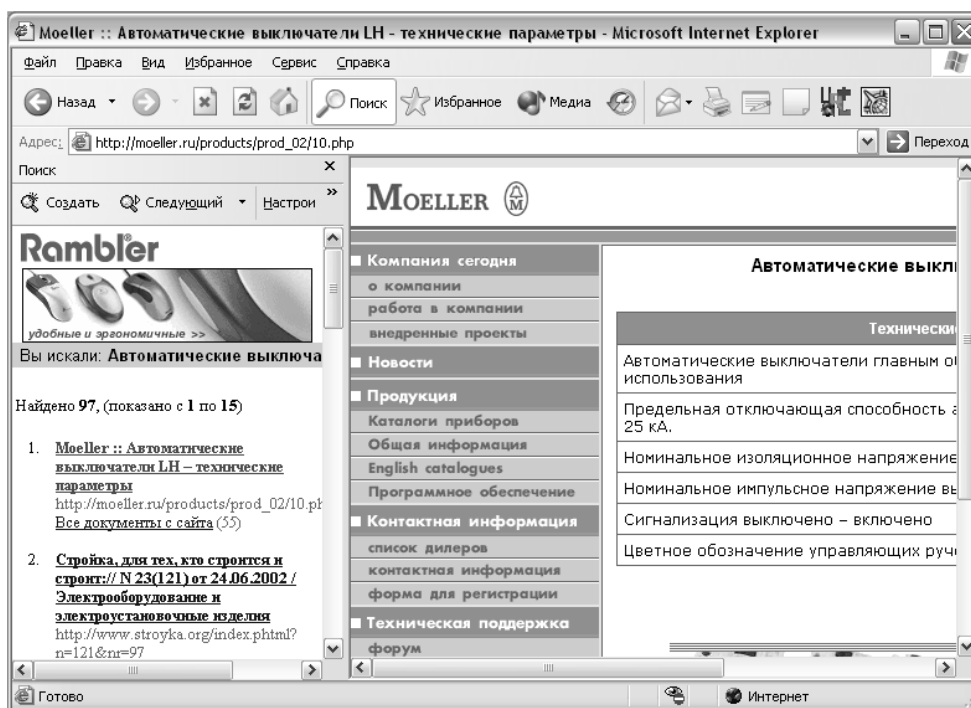


Рис. 1.3. А это часть результатов поиска на ту же тему, выданных поисковым Web-узлом Rambler

Внимательного читателя это должно насторожить. Если по достаточно четко сформулированному запросу по очень узкой теме найдено более полусотни ссылок, то что же будет, если нас будет интересовать не продукция концерна Moeller, а допустим история джаза? Заметьте, не музыки, а всего лишь одного из ее направлений, причем не самого популярного? В ответ на запрос История джаза Rambler выдал 9465 ссылок. Понятно, что никто не станет просматривать все найденные Web-страницы — в лучшем случае внимания удостоятся 5–8 из первой двадцатки. Если поисковый сервер способен располагать найденные страницы в таком порядке, когда наиболее релевантные предмету поиска Web-страницы и документы попадают в первую десятку–двадцатку, к его услугам чаще прибегают пользователи, если с этим у информационно-поисковой системы (ИПС), как иногда называют поисковые серверы, проблемы, — интерес к такому поисковику падает. Но о механизмах и алгоритмах поиска мы будем говорить в следующей главе. А пока давайте разберемся с тем, какой поисковый сервер должен искать необходимые вам документы первым. Если вы, предварительно установив соединение с Internet, щелкнете на кнопке с двумя стрелками, расположенной правее кнопки Следующий, откроется окно Настройка параметров поиска (рис. 1.4).

В версии 6 броузера Internet Explorer, которой мы пользовались при подготовке материалов этой книги, предлагается при проведении поиска пользоваться услугами четырех поисковых Web-узлов — трех наиболее популярных на начало тысячелетия

русскоязычных и, конечно же, поискового узла самой компании Microsoft. Перечень этих Web-узлов приведен в области Поиск веб-страницы справа. Вы можете, сняв флажок с названия оказавшегося малоэффективным поискового сервера, вовсе исключить его из числа тех, к услугам которого прибегает ваш браузер, а можете изменить порядок, в котором он будет обращаться к поисковым серверам, выделив один из них в левой части области Поиск веб-страницы и изменив его место в иерархии, щелкая на кнопках-треугольниках, расположенных ниже названной области.

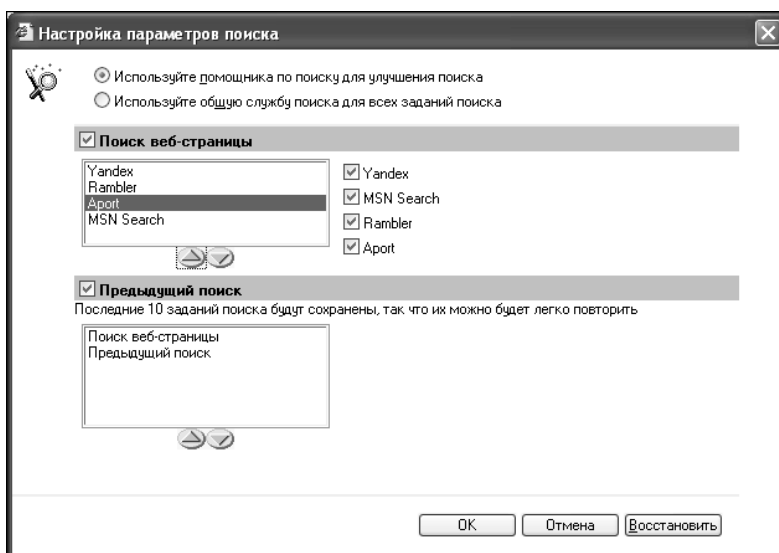


Рис. 1.4. Окно Настройка параметров поиска браузера Internet Explorer

В области Предыдущий поиск рассматриваемого окна вы можете указать, в каком режиме по умолчанию будет открываться панель Поиск после щелчка на одноименной кнопке в главном окне браузера — Поиск веб-страницы или Предыдущий поиск. Во втором случае удобнее возвращаться к результатам предыдущих поисков — их перечень отображается на панели Поиск сразу же после открытия главного окна браузера (если только накануне перед выходом из этой программы вы не закрывали панель Поиск щелчком на кнопке с крестиком в ее правом углу). Впрочем, переход из одного режима в другой легко выполнить, установив переключатель Выберите категорию поиска: в положение Предыдущий поиск или Поиск веб-страницы.

Если вы, проведя поиск по одной теме, решили изменить ключевые слова, по которым поисковый узел отыскивает нужные вам Web-страницы и документы, или вообще тему поиска, щелкните на кнопке Создать панели Поиск (а затем, при необходимости, еще и установите переключатель в положение Поиск веб-страницы). Вы можете провести новый поиск или очистить панель Поиск от результатов предыдущих поисков, щелкнув на кнопке Очистить.

С помощью вашего браузера вы можете не только проводить поиск информации в Internet, но и отыскивать “утраченные” папки и файлы на своем компьютере, компьютеры в локальной сети и людей в своей адресной книге, выбрав соответствующую ссылку в области Поиск иных объектов панели Поиск. Но наша цель — поиск информации в Internet, так что о поиске информации на собственном компьютере мы пока говорить не будем.

Для того чтобы завершить разговор о поиске информации с помощью браузера Internet Explorer (на самом деле — с помощью одного из специализированных поисковых Web-узлов), обратим ваше внимание на ссылку Найти похожие документы, по-

являющуюся в правой верхней части главного окна Internet Explorer при отображении в нем найденной Web-страницы или документа. Поисковый узел достаточно умен для того, чтобы понимать: десять тысяч ссылок — это слишком много. Если вы в первой двадцатке результатов оказавшегося слишком “урожайным” поиска обнаружите хотя бы один документ, в какой-то степени соответствующий вашим требованиям, щелкните после его открытия на названной ссылке. Поисковый сервер попытается сообразить, что именно вам было нужно, и найти среди оставшихся (практически тех же десяти тысяч) документы, похожие, по мнению поискового сервера, на найденный.

Как обычно, для ускорения просмотра найденных ИПС Web-страниц вы можете загружать их в отдельных окнах (для этого щелкните на очередной ссылке не левой, а правой кнопкой мыши и в открывшемся контекстном меню выберите пункт Открыть в новом окне). Если вы нашли то, что искали, и панель Поиск стала вам мешать, закройте ее щелчком на кнопке с крестиком в правом верхнем углу панели.

Итак, мы с вами очень быстро научились проводить простой поиск в Internet с помощью стандартного браузера. Но для успеха такого поиска мы должны хорошо знать, что хотим найти, предельно четко сформулировать запрос и уповать на то, что документов и страниц в Internet на эту тему не слишком много. Иначе мы или утонем в море обрушившейся на нас информации, или должны будем изменить стратегию и тактику поиска. Например, мы можем воспользоваться поисковым каталогом или поисковым сервером, не входящим в число тех четырех, которые предлагает Internet Explorer. Именно об этом мы и будем говорить в следующем разделе.

## 1.2. Каталоги и поисковые машины

Условно все ИПС можно разделить на два типа — каталоги и поисковые машины. Мы говорим “условно”, потому что любой поисковый каталог сейчас обязательно предлагает и услуги поисковой машины. И наоборот, многие поисковые машины предлагают заранее сузить область поиска, выбрав или тему поиска, или хотя бы ее область (искать во всей сети Internet, только товары и услуги, только новости и статьи и т.д.). У поискового сервера каждого из названных типов есть свои преимущества и недостатки, и тот из них, который хорошо подходит для одних целей, значительно уступает конкурентам при решении иных задач.



Все ИПС можно условно разделить на два типа — поисковые каталоги и поисковые машины. У каждой из названных ИПС есть свои преимущества и недостатки. Почти каждый поисковый каталог оснащается еще и поисковой машиной (а вот поисковая машина не всегда предлагает пользователю достаточно подробный каталог).

Поисковые каталоги напоминают систематические каталоги обычных библиотек. Тот, кому приходилось бывать в зале каталогов крупной библиотеки, непременно вспомнит о нем, впервые попав на начальную страницу поискового каталога (рис. 1.5).

Профессиональный библиотекарь, увидев, что в тему “Культура и искусство” на равных входят темы “Книги” и “Литература”, а в тему “Образование/Наука”, опять-таки на равных, темы “Психология” и “Техника”, конечно, схватился бы за голову. Но рядовой пользователь, поблуждав между “стеллажами” каталога, достаточно быстро начинает ориентироваться в его темах. (Поскольку очень часто темы Internet-каталогов называют *категориями*, в дальнейшем мы будем пользоваться этими терминами как синонимами, хотя это и не совсем корректно.) То, что в первую очередь волнует пользователя — работа, дом, дача, машина, развлечения, новости, здоровье, спорт, — вынесено на самый верх информационного айсберга. А далее, путем последовательных уточнений, можно достаточно быстро дойти до самых нижних ступеней иерархии, которой является, по сути, всякий каталог. Например, выбрав категорию “Техника”, а затем следующую “Аудио- и видеотехника” (рядом с нею вы увидите цифру, указывающую количество Web-узлов на эту тему, упомянутых в каталоге), вы откроете страницу со ссылками на

сайты по выбранной теме. В верхней части страницы (рис. 1.6) вы обнаружите поле ввода для поиска в Internet, в данном каталоге либо в Internet-магазинах, если речь идет о каком-нибудь товаре. Ниже приведены темы, имеющие отношение к выбранной вами. Так, поскольку вы интересовались аудио- и видеотехникой, вам напомнят о том, что можно посетить сайты производителей соответствующей аппаратуры.

Среди включенных в каталог Web-узлов вы можете выбрать сайты определенного региона, персональные, корпоративные (в данном случае, скорее всего, это Web-узлы фирм, производящих аппаратуру интересующего вас типа) и т.д.

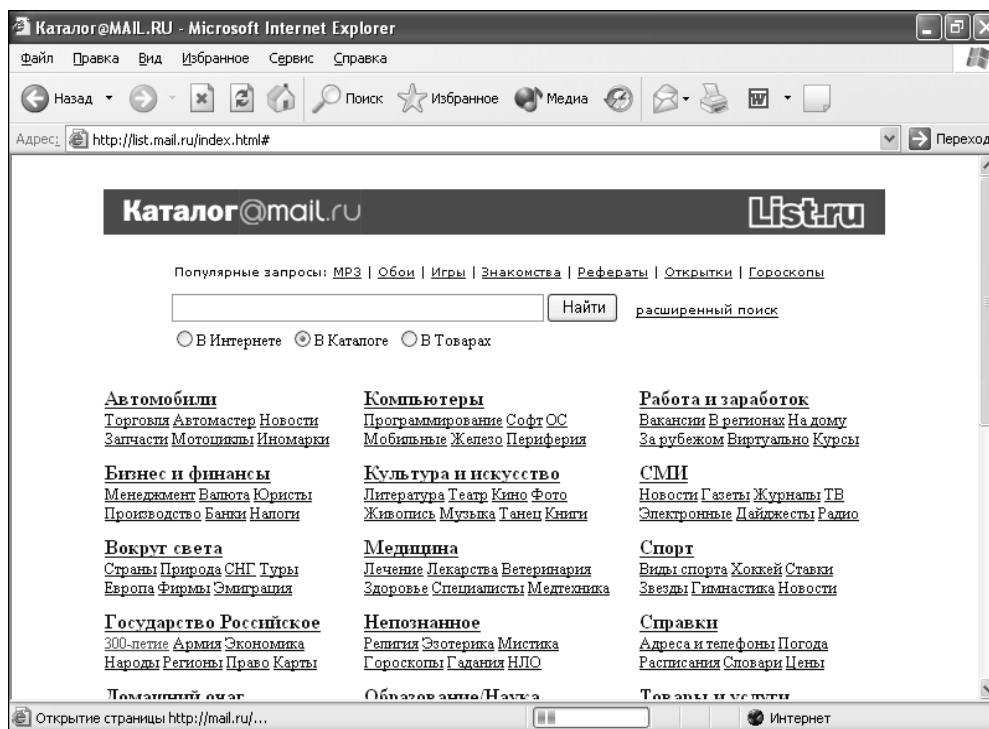


Рис. 1.5. Начальная страница поискового каталога @mail.ru-List.ru

Если ссылок на сайты по выбранной теме в каталоге слишком много, вы можете провести их сортировку. В частности, в каталоге @mail.ru-List.ru можно сортировать сайты по “оценке гидов”, алфавиту, популярности и дате. Чаще всего в поисковых каталогах ссылки на относящиеся к той или иной теме сайты сортируют по популярности. Оно и понятно: если какой-то сайт посещают многие пользователи, значит, они находят на нем полезную информацию. Скорее всего, именно на таком сайте и вы найдете то, что вас интересует. И наоборот, “не пользующиеся спросом” Web-узлы игнорируются посетителями потому, что собранная на них информация не представляет интереса для большинства людей. Но сделать сайт популярным можно искусственными (и, добавим, нечестными) методами. Одно из преимуществ поисковых каталогов как раз и состоит в том, что все включаемые в них сайты просматриваются редакторами. И если какой-либо сайт включен в каталог, значит, он заслуживает внимания. Более того, “гиды” каталога могут расставить ссылки на сайты так, что первыми будут упоминаться наиболее информативные из них — естественно, в соответствии с субъективным мнением редакторов.

Ниже вы увидите перечень ссылок на Web-узлы по выбранной вами теме, включенные в каталог (рис. 1.7). Приводится название сайта, его аннотация, URL (адрес в сети Internet), а также указывается, сколько в среднем за день делается переходов на данный сайт, когда он был открыт и на каком (или каких) языках ведется.

В самом конце страницы вы увидите ссылку **Добавить сайт**. Владелец любого Web-узла, ради его популяризации, может попытаться включить ссылку на него в каталог. Окончательное решение по этому вопросу принимает редактор каталога. Так, если, по его мнению, сайт содержит мало полезной информации, он может быть не включен в каталог.

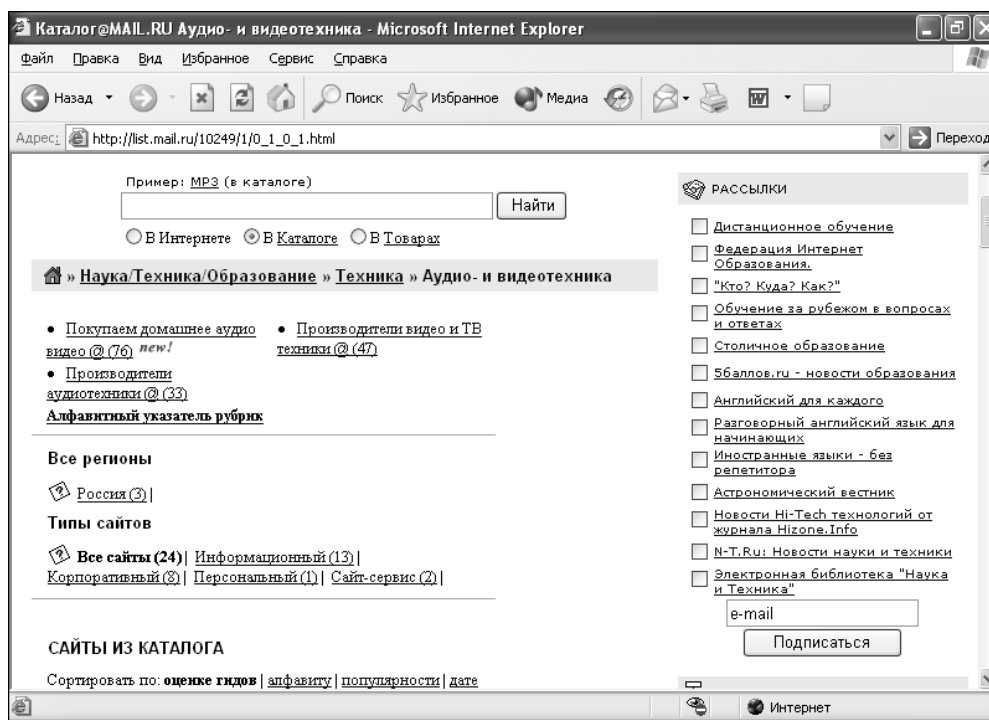


Рис. 1.6. Верхняя часть одной из страниц каталога

Как видите, с помощью каталога вы сможете достаточно быстро найти несколько сайтов, содержащих (возможно) информацию по интересующей вас теме. На самих сайтах нужные страницы придется отыскивать самостоятельно. Если вы только начинаете интересоваться какой-то темой, каталог может оказаться чрезвычайно полезным — в нем вы найдете самые “информированные” и авторитетные Web-узлы по данной теме. Но вот проводить с помощью каталога поиск информации по узконаправленной тематике может оказаться нецелесообразным. Так, среди рубрик на тему “Техника” в данном каталоге пока нет рубрики “Электротехника”. Поэтому наши попытки отыскать с помощью каталога информацию об автоматических выключателях Moeller были прекращены в силу их неэффективности. Это не значит, что с помощью какого-нибудь каталога эту информацию вообще нельзя найти. Но только очень терпеливый пользователь будет просматривать содержимое полудюжины тем каталога, если среди них нет той, которая с большой вероятностью соответствует теме поиска. Более того, если на странице представлено больше трех десятков рубрик, нужная, не исключено, будет пропущена при первом просмотре, а второго может и не быть. Ко-



нечно, можно ввести ключевые слова в поле Запрос, но тем самым вы задействуете для поиска уже фактически не каталог, а поисковую машину. О поисковых машинах мы поговорим в этом же разделе, но чуточку позже, а пока проанализируем преимущества и недостатки поисковых каталогов. К преимуществам можно отнести следующие.

- Каталоги имеют удобную древовидную структуру, позволяющую быстро найти сайты на заданную тему.
- Web-узлы на заданную тему перечисляются в порядке их значимости (с точки зрения редакторов каталога). Возможна сортировка по посещаемости, дате открытия и другим критериям.
- В каталоге обязательно бывают представлены все наиболее популярные Web-узлы по той или иной теме (из числа представленных в каталоге).
- Искусственные методы повышения популярности в поисковых каталогах, как правило, не работают, поэтому вероятность быстро найти наиболее “информированный” по данному предмету сайт выше.

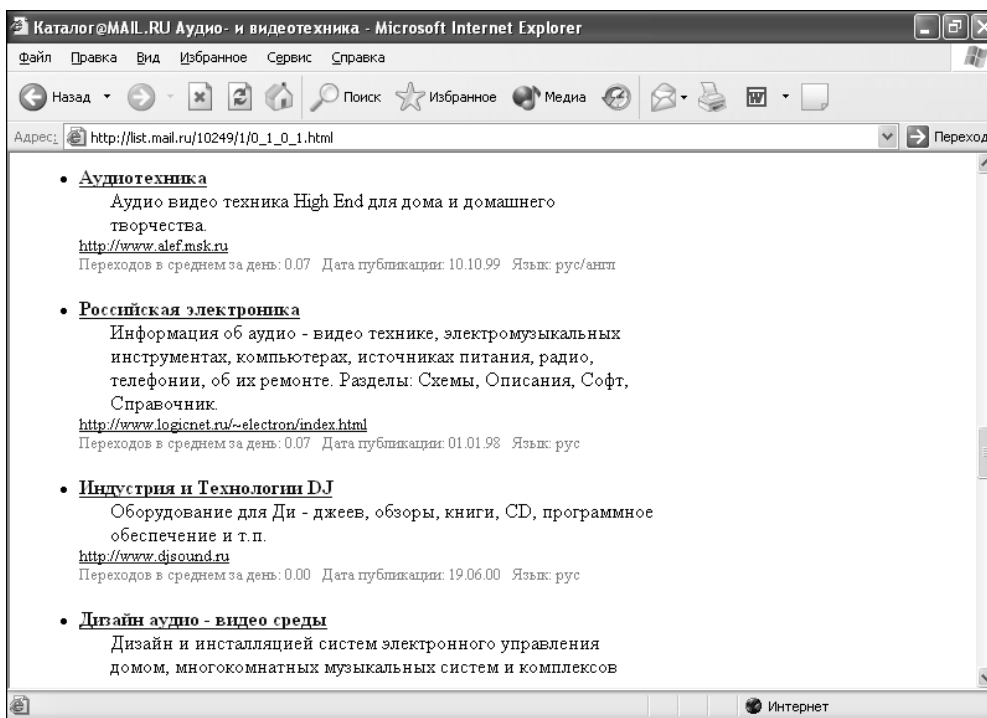


Рис. 1.7. Ссылки каталога сопровождаются аннотациями и другой информацией

Но есть у каталогов и недостатки. Наиболее серьезные из них перечислены ниже.

- Поскольку отбор ресурсов Internet, упоминаемых в каталоге, ведут его редакторы, это сказывается на объективности “ранжирования” сайтов.
- В каталоги попадают далеко не все сайты. Владельцы некоторых не афишируют свои детища, а редакторы каталога не в состоянии просмотреть сотни тысяч Web-узлов Сети, чтобы включить в свой каталог наиболее интересные.

- Некоторые сайты не упоминаются в каталогах из-за их “неприличности” (а ее степень субъективно определяет редактор каталога) или просто из-за различия политических либо религиозных убеждений редактора каталога и создателя сайта (не всякий приверженец правых сил станет рекламировать с помощью “своего” каталога сайты, пропагандирующие левые идеи).
- Обновление каталогов происходит достаточно медленно. Сайты, ссылки на которые вы нашли, могут за это время поменять адрес (обычно их все равно потом можно как-то найти), а то и вовсе исчезнуть из Сети (в этом случае тоже можно найти хотя бы часть из Web-страниц, но так бывает далеко не всегда).

От большинства из названных недостатков свободны поисковые серверы, основной инструмент которых — так называемая *поисковая машина*, или *автоматический индекс*. Это довольно сложный комплекс программ, обеспечивающих оперативное обновление информации о тысячах Web-узлов Internet. В периоды наименьшей загрузки линий связи и Web-узлов особые программы (их называют *поисковые роботы*, *HTTP-роботы* или просто *пауки*) проводят так называемую *индексацию* Web-страниц и документов, размещенных в Сети, т.е. заносят в базу данных поисковой машины информацию о них. Глубина индексации может быть разной. Если речь идет о Web-странице, то поисковый робот обследует следующие ее элементы: заголовок, перечень ключевых слов, составленный автором страницы, краткое описание страницы, также составленное ее разработчиками, и собственно содержимое страницы. Полные тексты документов, размещенных на странице, в базу данных копируют не все поисковые роботы — некоторые ограничиваются лишь их заголовками. Когда пользователь формулирует запрос на поиск, поисковая машина просматривает свою базу данных и выдает перечень Web-страниц, содержащих слова, введенные пользователем в поле ввода (их тоже часто называют *ключевые слова*). Как уже говорилось, таких страниц может быть очень много. Задача поисковой машины — отобрать те из них, которые в наибольшей степени отвечают запросу пользователя (т.е. *релевантны* ему) и указать ссылки на них в числе первых. Разные поисковые машины используют для этого различные алгоритмы. Скажем, если одно из указанных пользователем слов встречается и в заголовке страницы, и среди ключевых слов самой страницы, и в ее кратком описании, и в заголовках либо текстах документов страницы, она, возможно, будет упомянута в результатах поиска одной из первых. Мы говорим “возможно”, потому что некоторые роботы игнорируют и ключевые слова (наиболее полно характеризующие страницу по мнению ее разработчиков), и краткие описания страниц, составленные авторами; они ориентируются только на содержимое Web-страниц. Дело в том, что вероятность посещения страницы, упомянутой в результатах поиска среди первых, довольно высока, даже если ее заголовок не соответствует теме запроса. А чем больше посещений регистрируется на сайте, тем дороже стоит его “рекламная площадь”. Владельца такого сайта мало волнует, что вы впустую тратите время и деньги на просмотр его содержимого. Главное то, что он успел показать вам рекламные баннеры и окна. Поэтому ключевые слова и краткое описание некоторых Web-страниц (они остаются невидимыми для посетителей) иногда совершенно не соответствуют их содержанию.

Примеры русскоязычных поисковых машин — те же Yandex, Rambler и Апорт, к которым обращается браузер, когда вы щелкаете на кнопке Поиск в его панели инструментов. Существуют и многие другие. Запросы на русском языке принимают некоторые зарубежные поисковые машины, например Google, одна из самых мощных и популярных. При формировании результатов запроса она, среди многих прочих факторов, учитывает количество ссылок на Web-страницу, сделанных на других сайтах. Чаще цитируемые страницы выдвигаются в первые ряды. Этот критерий оказался столь эффективным, что степень релевантности результатов поиска системы Google почти всегда оказывается весьма высокой. (Мы говорим “почти”, потому что “раскрутки” сайтов все время ищут новые (нечистоплотные) способы проталкива-

ния своих подопечных в первые ряды, и иногда им это удается.) Поэтому в последнее время и другие поисковые машины начали учитывать “фактор цитируемости”.

К преимуществам поисковых машин следует отнести следующие.

- Малое количество в результатах поиска устаревших ссылок (почтовые роботы проверяют уже занесенные в базу данных сайты гораздо чаще, чем это делают редакторы каталогов).
- Намного большее количество Web-узлов, по которым производится поиск. Если наиболее крупные поисковые каталоги содержат ссылки на десятки тысяч сайтов, то поисковые машины в своих базах данных — на сотни тысяч и миллионы.
- Более высокую скорость поиска. При поиске в каталоге вам, уточняя тему, придется открывать несколько страниц, а потом еще “перелистывать” страницы найденных сайтов; поисковая машина сразу выдает адреса конкретных документов.
- Высокую релевантность поиска. Правда, это относится не ко всем поисковым машинам. Но каждая из них развивается, и повышение релевантности — одно из главных направлений совершенствования.

Существуют у поисковых машин и недостатки. Среди главных отметим следующие.

- Ограниченная область поиска. Если какой-либо сайт не был внесен в базу данных поисковой машины, он для нее “не существует”, и его документы в результаты поиска попасть не могут.
- Относительная сложность использования. Для того чтобы составленный вами запрос на поиск точно соответствовал тому, что именно вы хотите найти, нужно хотя бы немного представлять, как работает поисковая машина, и уметь использовать простейшие логические операторы. Поисковые каталоги в этом смысле проще и привычнее.
- Менее наглядная форма представления результатов запроса. Каталог выдает вам название сайта с его краткой аннотацией и другой полезной информацией (количество посещений, дата создания и т.д.). Результаты работы поисковой машины менее наглядны (см. рис. 1.2).
- Поскольку базу данных поисковой машины пополняют программы-роботы, нечестные владельцы рекламных сайтов могут их “обмануть”, из-за чего релевантность поиска может быть значительно снижена.

Помимо каталогов и поисковых машин, отыскать нужную информацию в Сети вам помогут рейтинги и порталы. О них пойдет речь в следующем разделе.

### 1.3. Рейтинги и порталы

Как уже говорилось, в любом поисковом каталоге ссылки на Web-узлы, относящиеся к той или иной теме, можно отсортировать в порядке их посещаемости. Это довольно объективный критерий полноты и качества материалов, выложенных на сайте. Если другие пользователи регулярно посещают определенный Web-узел (а за счет случайных посетителей сделать этот показатель высоким невозможно), значит, имеющаяся на нем информация, скорее всего, окажется интересной и для вас.

Поэтому неудивительно, что в Internet появились Web-узлы, специализирующиеся на ведении *рейтингов*. На таком узле множество, допустим, русскоязычных сайтов (на Web-узле SruLog их зарегистрировано около 130 тысяч) делится на категории (рис. 1.8), и в каждой из них ссылки на сайты отображаются в строгом соответствии со степенью популярности того или иного Web-узла. Популярность пропорциональна числу посещений сайта за день (и это будет один рейтинг), за неделю (этот рейтинг

также может быть выведен после щелчка мыши), за месяц или в среднем. Подсчитывается и общий рейтинг, не зависящий от “специализации” сайта (на рис. 1.8 он представлен в левой части окна). Естественно, первые места в “общем” хит-параде занимают поисковые серверы, Web-узлы, оперативно сообщающие о событиях в мире и о погоде, и разного рода развлекательные сайты. На “обычных” поисковых каталогах вы также можете “выстроить” сайты по степени популярности, но предлагаются и другие механизмы классификации, а также обязательно —поисковая машина. На “рейтинговых” узлах главное — посещаемость сайта; категории, на которые создателями такого узла разделены все сайты, — единственное, что облегчает вам поиск необходимой информации. Но зато по любой теме (ну, почти по любой) вы сразу можете найти самые популярные сайты, за которые граждане Internet “проголосовали” своими посещениями. А это значит, что и для себя вы найдете на них что-то интересное. Те же из пользователей, которые привыкли всегда и во всем быть оригинальными, могут смело “идти” в конец рейтинга. Ведь многие замечательные Web-узлы не пользуются широкой популярностью лишь потому, что рассчитаны на относительно узкие группы людей, интересующихся определенными вопросами, или “чересчур заумны” для рядового гражданина страны Internet. Но зато именно там вы можете найти тех немногих людей, которые разделяют ваши интересы, завязать интересные знакомства и даже принять участие в совместных проектах.

Еще один пример Web-узла, активно занимающегося рейтингами, — сайт Национальной почтовой службы России (<http://top.mail.ru/rating>).

К преимуществам рейтингов можно отнести, пожалуй, лишь одно: с их помощью вы очень быстро можете найти адреса наиболее посещаемых сайтов по интересующей вас теме и при этом будете уверены: ни один из популярных ресурсов не пропущен. К недостаткам поиска информации с помощью рейтингов можно отнести следующие.

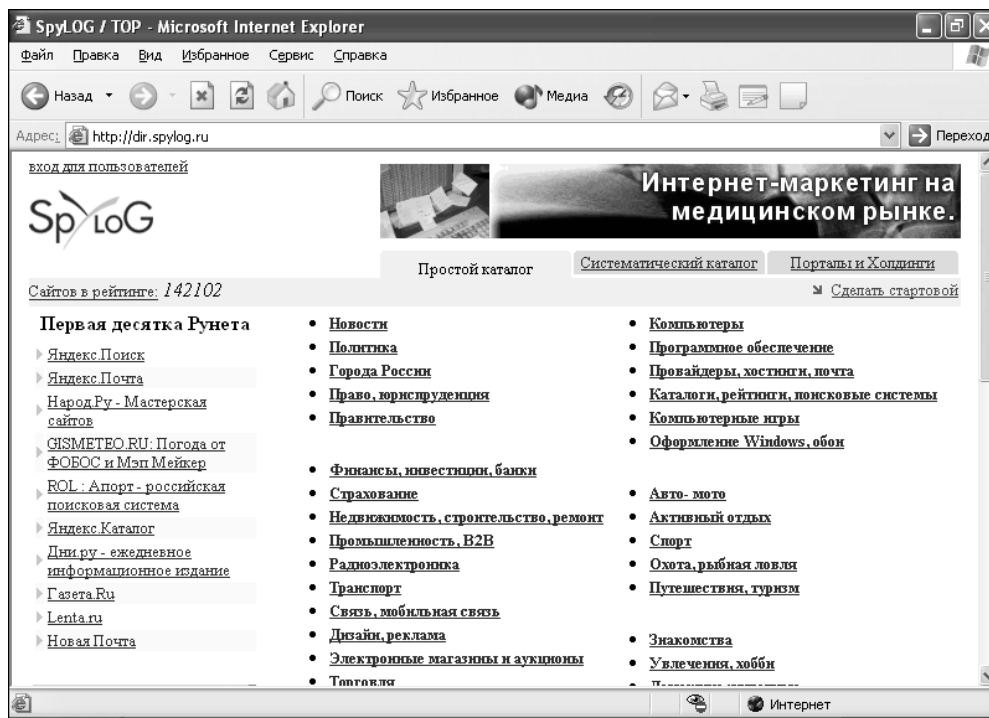


Рис. 1.8. Рейтинговый Web-узел

- Искать информацию на малопосещаемых сайтах, относящихся к интересующей вас теме, — занятие и малопродуктивное, и утомительное.
- Даже наиболее крупные рейтинги содержат ссылки лишь на несколько десятков (очень редко — сотен) тысяч Web-узлов; значительная часть имеющейся в Сети информации находится вне поля их внимания.
- Если интересующая вас тема не относится к числу популярных, рейтинги, фактически, окажутся вообще бесполезными.

Как сказано выше, наиболее высокие позиции в рейтингах обычно занимают поисковые Web-узлы, а также сайты, предоставляющие последние новости и всевозможные развлечения (начиная с анекдотов и кончая сетевыми компьютерными играми со многими участниками). Естественно, в Internet появились сайты, предоставляющие все это и многое другое (например, бесплатные почтовые ящики) по принципу “все в одном”. Называются такие сайты *порталы*, самые крупные из них именуют себя *мультипорталы*.

Типичный “джентльменский набор” сервисов и сайтов мультипортала обычно включает поисковую систему, каталог Web-страниц, ленту новостей, электронную почту (бесплатные почтовые ящики), набор словарей и энциклопедий, Internet-магазин, доску объявлений, чат, электронную библиотеку, гороскоп, подборку анекдотов и т.д. Причем лент новостей может быть несколько — здесь новости и политические, и экономические, и из области культуры, и из мира Internet... Прочитав одну новость, вы где-то в конце сообщения найдете ссылки на другие новости по данной теме. Благодаря этому можно проследить, как развивалось то или иное событие, как оно комментировалось, какие прогнозы относительно его финала оказались верными, какие комментарии ошибочными, а какие — просто лживыми.

### 1.3. Рейтинги и порталы

На страницах каждого солидного портала бывает представлено так много информации, что в его поисковой системе обычно бывает опция типа “Искать в пределах сайта”. И если вам не нужно что-то весьма специфическое, скорее всего, непосредственно на страницах портала вы найдете всю необходимую в данный момент информацию. Конечно, портал не заменит ни Web-узел какого-нибудь физического общества, на котором вы можете по ключевым словам сделать подборку статей по интересующей вас теме, ни сайт, посвященный, например, русской фантастике. Но зато вы очень быстро узнаете прогноз погоды на завтра, курс валют, самые главные (по мнению редакторов портала) новости, сможете “пролистать” программу телепередач, узнать, что сегодня в театрах и какие другие интересные мероприятия можно посетить. Многие поисковые системы представляют собой, по сути, порталы (а любой портал считается “ущербным”, если не предлагает своим посетителям удобную систему поиска).

Одним из наиболее крупных порталов Рунета (так часто именуют русскоязычную часть Internet) является мультипортал Кирилл и Мефодий (рис. 1.9), где самое интересное — набор словарей и энциклопедий. Большое внимание уделено также образованию. Здесь вы найдете и портал *для студентов* (по-английски *for students*, следуя моде записывать сходно звучащие английские слова цифрами, его адрес выбрали таким: <http://4students.ru>), и виртуальную школу для учеников старших классов (<http://vschool.ru>), и образовательный портал для детей младшего возраста, и страницу для преподавателей. Есть на мультипортале и торговый отдел (прайс-листы более чем 3 тыс. фирм и Internet-магазинов), и новости на любую тему, и огромный “парк культуры и отдыха” (афиши, всевозможные развлечения, программы телепередач и т.д.).

Поэтому, если вас в данный момент интересует информация “общего назначения”, имеет смысл сразу же отправляться на какой-нибудь портал и искать ее там. Заодно вы узнаете последние новости, погоду на завтра, курс валют и последние сплетни о наиболее известных (или наиболее одиозных) персонах.

К преимуществам порталов, используемых в качестве поисковых систем, можно отнести следующие.

- На них можно быстро найти самую свежую информацию по наиболее животрепещущей теме, а также ранее опубликованные на портале сообщения и документы, имеющие к ней отношение.
- С помощью поисковой машины, имеющейся на каждом портале, можно быстро найти в его пределах информацию разного типа на заданную тему. Это могут быть и новости, и аналитические статьи, и статьи словарей и энциклопедий, и иллюстрации, и реплики из форумов и чатов, и многое другое (вплоть до майки с портретом поп-звезды, продаваемой в Internet-магазине). Как правило, можно осуществлять поиск и за пределами портала, т.е. портал предоставляет в распоряжение пользователей еще и “универсальную” поисковую машину.

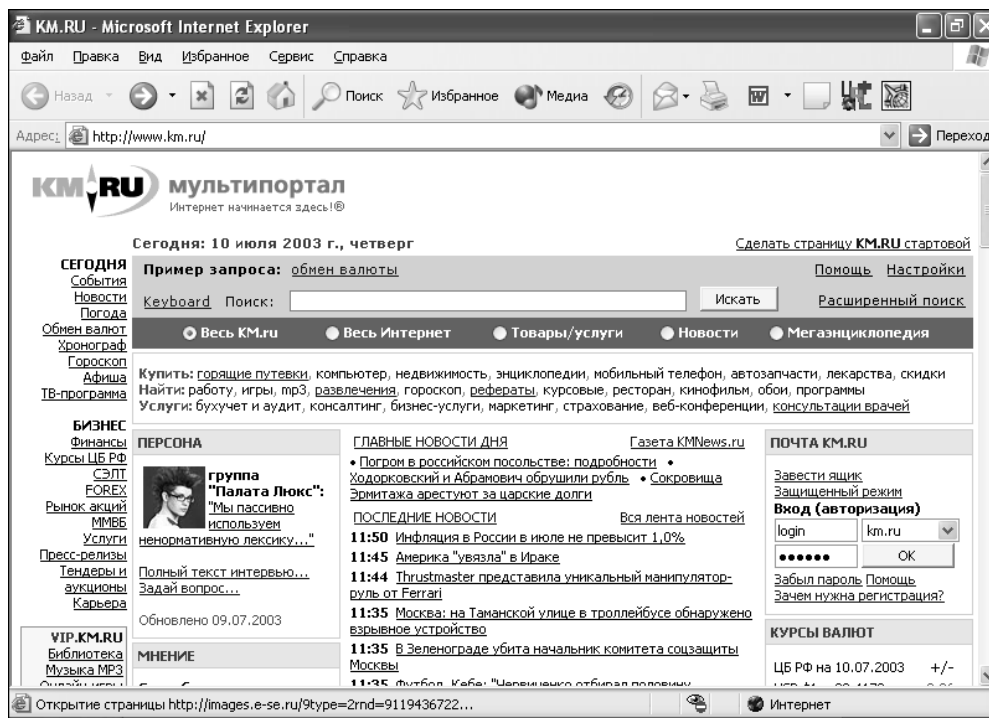


Рис. 1.9. Мультипортал Кирилл и Мефодий

Недостатки поиска информации на портале очевидны.

- Ограниченность охватываемых тем.
- Предоставляемая информация рассчитана на массового посетителя и не отличается особенной глубиной.
- Как бы ни велик был портал, он не заменит тысячи других Web-узлов, содержащих информацию по интересующей вас теме.

## 1.4. Анализ результатов простого поиска

Итак, мы с вами убедились, что, даже не обладая какими-либо существенными знаниями о том, как работают поисковые машины, вы можете найти в Internet информацию по выбранной вами теме. Точнее, вы найдете страницы и даже конкретные документы, содержащие, допустим, ключевые слова, введенные вами в соответствующем поле ввода поисковой машины, или ссылки на Web-узлы, имеющие отношение к интересующему вас предмету. Большинству пользователей для поиска информации вполне хватает умения составлять простые запросы для поисковых машин и навыка обращения с каталогами, рейтингами и порталами. Но в любом случае полученные даже в результате простого поиска результаты нужно уметь анализировать. Именно об этом мы и будем говорить в данном разделе.

Как уже говорилось, каталоги предоставляют аннотированные ссылки на сайты, имеющие отношение к той или иной теме. Некоторые каталоги позволяют сортировать ссылки по популярности, дате создания и другим критериям (см. описание каталога @mail.ru-List.ru в разделе 1.2). Наиболее популярный в США и других англо-

зycznych странах каталог Yahoo! сам отбирает наиболее популярные сайты и перечисляет ссылки на них первыми (рис. 1.10), остальные дает в алфавитном порядке. Кроме того, перед списком ссылок на сайты он дает ссылки на другие их категории, в которых также может содержаться необходимая вам информация.

Как видим, основную часть анализа при простом поиске информации с помощью каталогов пользователь должен выполнить при выборе категории, к которой, по его мнению, должна относиться интересующая его тема, и при выборе подлежащих более подробному изучению сайтов по их аннотациям.

Практически те же возможности вы получаете, используя список ссылок рейтинга или портала (рис. 1.11).

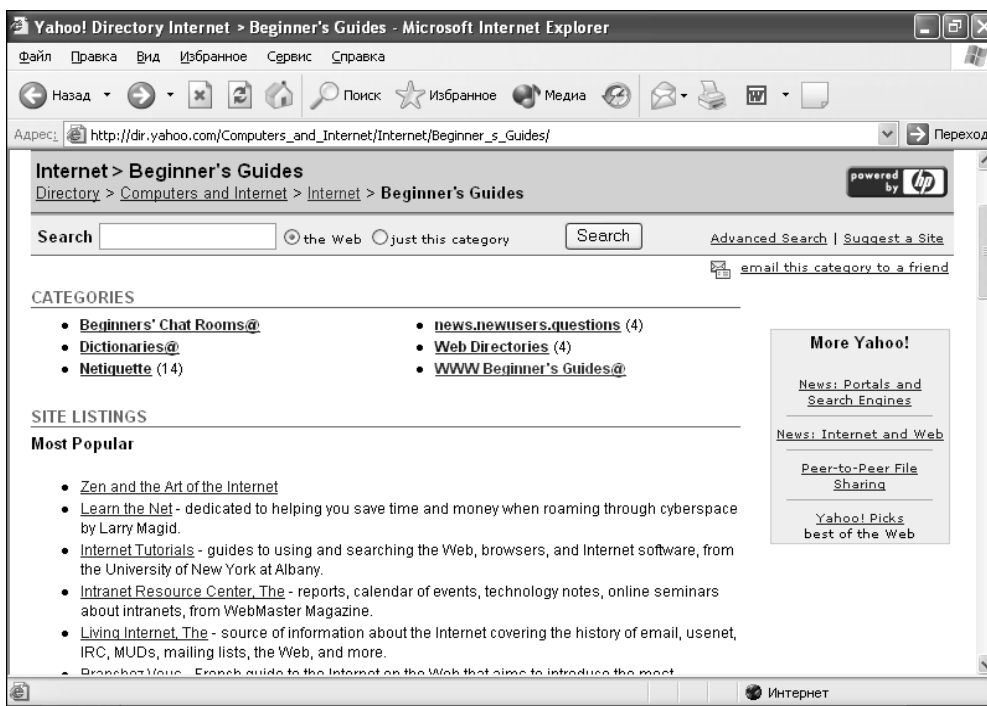


Рис. 1.10. Популярный каталог Yahoo! самостоятельно сортирует сайты

Как мы видели (см. раздел 1.1), если поисковая машина вызывается браузером, вы получаете еще меньше “сопутствующей” информации, чем при проведении поиска в каталоге. Но если вы воспользуетесь услугами информационно-поисковой системы Yandex “без посредников”, игнорируя инструменты поиска браузера Internet Explorer, результат будет иным (рис. 1.12). Как видите, в этом случае поисковая машина дополнительно предоставила нам довольно много информации, которая поможет быстрее найти именно то, что нас интересует, или изменить тактику поиска. Так, она (машина) сообщает, что всего нашла 3194 страницы на 755 серверах. Слово “автоматические” встретилось более трех миллионов (!) раз, “выключатель” — почти 665 тысяч раз, Moeller — всего лишь 6675 раз.



К ИПС лучше обращаться напрямую, не используя в качестве посредника браузер. В этом случае каждая найденная ссылка сопровождается “сопутствующей” информацией, благодаря которой можно быстрее отыскать нужный сайт.



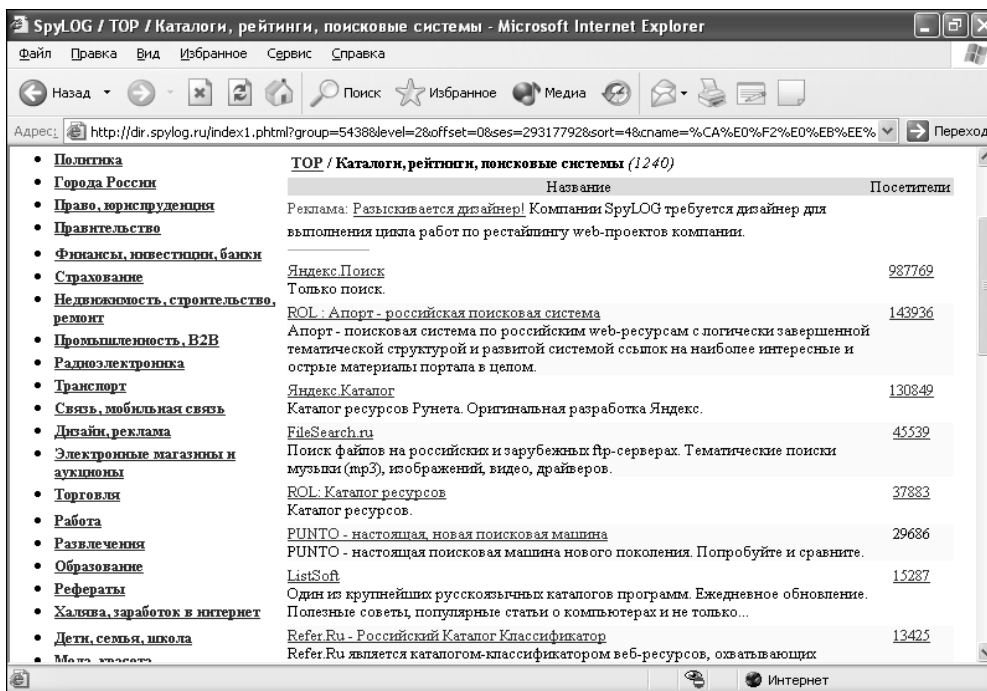


Рис. 1.11. Рейтинг также предоставляет аннотацию каждого сайта

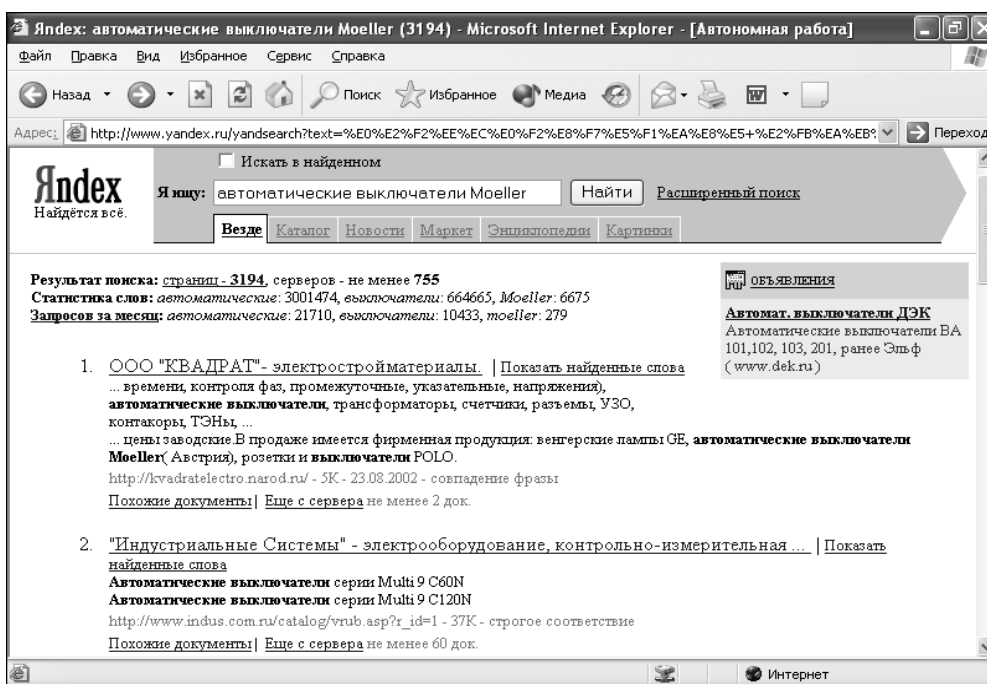


Рис. 1.12. Результаты поиска, представленные поисковой машиной при обращении к ней "без посредников"

Итак, нам уже есть над чем задуматься. Несмотря на достаточно конкретный вопрос, ссылок получено чересчур много. На первой странице результатов запроса приведены десять из них, среди которых, по мнению поисковой машины, должны быть именно те, которые нам нужны. Хорошо, если так. А если нет? Просматривать три с лишним тысячи найденных страниц не имеет смысла. Видимо, запрос нужно было сформулировать несколько иначе. Как именно? Об этом мы поговорим в главе 3. А пока попробуем все же разобраться в том, что “принесла нам на хвосте” поисковая машина.

Первым в результатах поиска представлен, скорее всего, какой-то Internet-магазин, предлагающий среди прочей продукции автоматические выключатели, в том числе и автоматические выключатели концерна Moeller. Поисковая машина привела два фрагмента страницы, на которой встречаются слова, введенные нами в качестве ключевых, и выделила их полужирным шрифтом. Поскольку это фрагменты, каждый абзац начинается с маленькой буквы и предваряется многоточием. Далее дается адрес страницы, на которой были найдены абзацы, содержащие ключевые слова, и объем страницы в килобайтах. Может быть также указана дата создания либо обновления страницы и критерий, которым руководствовалась поисковая машина, определяя место ссылки в выводимом списке. В частности, первой ссылка в рассматриваемом случае была приведена потому, что найденная в документе фраза полностью совпадает с введенными ключевыми словами.

Завершают своеобразную аннотацию к ссылке предложения просмотреть похожие документы в Internet (на случай, если вам мало тех трех тысяч страниц, которые были найдены в результате поиска) и загрузить с “цитируемого” сервера другие документы на ту же тему (с указанием их числа).

Если вас интересовало приобретение автоматических выключателей в Internet-магазине, уже первая ссылка окажется полезной. Но если вас интересовали типы и параметры выключателей Moeller, полезной будет третья ссылка (см. рис. 1.2).

В конце первой страницы с результатами поиска (рис. 1.13) вы увидите, что по умолчанию ссылки были отсортированы по релевантности. Альтернативная возможность — отсортировать их по дате размещения документа в Internet, ограничить регион, ограничить поиск предлагаемым поисковой машиной каталогом и, наконец, обратиться к помощи других поисковых машин.

Как видите, поисковая машина, во-первых, может выдать ссылки не на сайты, а на конкретные документы и даже процитировать слова и фразы, расположенные в найденном документе рядом с теми, которые вы задали в качестве ключевых. Во-вторых, она позволяет быстро модифицировать запрос на поиск, ограничив отображаемые ссылки по тому или иному критерию, и тем самым уменьшить количество страниц, которые придется открывать и просматривать.

Но какой бы поисковой машиной вы ни пользовались, в случае простого поиска результат будет далеким от оптимального. Согласитесь, три с лишним тысячи ссылок, многие из которых, возможно, придется просматривать “вручную” — это слишком много. А ведь тема поиска была очень узкой, очень конкретной. Будь она несколько шире, шансов найти искомое в первом десятке выданных ИПС ссылок были бы весьма невелики. Но более подробно о недостатках простого поиска мы поговорим в следующем разделе.

## 1.5. Недостатки простого поиска

О недостатках простого поиска, проводимого с помощью каталогов, рейтингов и порталов, мы говорили в соответствующих разделах и здесь повторяться не будем. Как следует из предыдущего раздела, поиск с помощью поисковой машины зачастую (хотя и не всегда) оказывается более эффективным. Но у него также есть существенные недостатки. Один из главных — непомерно большое количество выдаваемых ссылок.

Продолжим анализ результатов проведенного нами поиска. Почему ссылок получилось слишком много? Даже беглый взгляд на первую же найденную ссылку

(см. рис. 1.12) дает ответ: нас интересовали автоматические выключатели концерна Moeller, но поисковая машина выбирала и страницы со словом “автоматические”, и страницы со словом “выключатели”, и страницы со словом Moeller отдельно. При этом, как нетрудно заметить (см. рис. 1.13), заданные нами слова в найденных документах могут стоять в любом порядке, и десятая ссылка указывает на страницу, посвященную шинопроводам Moeller, которые нас совершенно не интересуют. Потом поисковая машина отсортировала найденные документы и первой выдала все же ссылку на ту страницу, в документе которой нашла совпадающую фразу, т.е. все три слова, указанных нами в качестве ключевых. Но уже вторая ссылка (и несколько других в первом десятке) бьет мимо цели — нас интересуют не автоматические выключатели вообще, а только конкретной фирмы.

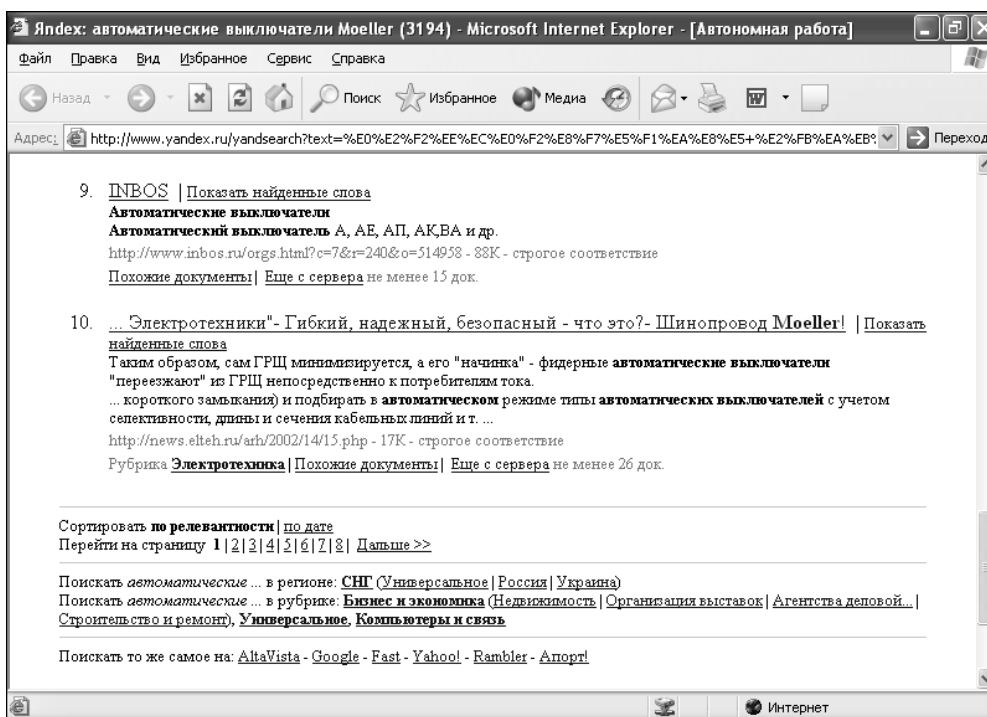


Рис. 1.13. Найденные ссылки можно отсортировать наиболее удобным для пользователя способом

По-видимому, поиск оказался бы более эффективным, если бы мы каким-то образом дали знать ИПС, что в нужных нам документах обязательно должны находиться все три слова. При этом они должны находиться неподалеку одно от другого, в одном предложении или хотя бы абзаце. Порядок слов в данном случае не имеет значения. Обратите внимание: первой дана ссылка на документ, в котором обнаружилось совпадение фразы. Между тем наиболее полезна в первой десятке для нас третья ссылка, в которой ключевые слова следуют в другом порядке. Но возможны случаи, когда порядок слов как раз имеет значение: если вы, например, по цитате из какого-то рассказа или стихотворения пытаетесь отыскать его текст в Internet.

Третья ссылка указывает на страницу, содержащую нужные нам слова не в текстах документов, а в ее заголовке. В рассматриваемом случае ценность такой страницы для нас намного выше — мы сразу получаем много сведений о предмете поиска. Если бы мы начали поиск со страниц, содержащих ключевые слова в заголовках, наверное, эффектив-

ность поиска значительно повысилась бы. Но стихи по фрагментам строк, наоборот, в заголовках искать не стоит. Если нас интересует Web-узел какой-то фирмы (допустим, “Рога и копыта”), правильной тактикой будет искать не документы и страницы, а сайты, в адресе которых есть слово *rog*, *roga*, *kozya*, *kozya*, *horn*, *horns*, *antler*, *antlers* *hoof* или *hoofs*. При этом хорошо бы не делать запрос по каждому слову отдельно, а вводить их хотя бы по два-три. То есть машина должна знать, что нас интересуют сайты, в которых есть *или* слово *rog*, *или* слово *horn*, *или* слово *antler*, причем во множественном числе.

Итак, мы могли бы гораздо точнее провести поиск, если бы могли “объяснить” поисковой машине, как именно должны быть расставлены ключевые слова в нужных нам документах, в каких элементах сайта, страницы или документа они должны находиться, а также многое другое (например, сколько “посторонних” слов могут разделять в найденном документе ключевые слова). И поисковые машины, конечно же, позволяют делать все это. Но для этого нужно проводить уже не простой, а *сложный* или *расширенный* поиск. О том, как можно значительно сэкономить время и средства, заменяя простой поиск сложным или расширенным, мы поговорим в главе 3. Но вначале имеет смысл рассказать о точности и полноте поиска, а также дать определения некоторых понятий. Именно этим мы и займемся в следующей главе.

## Резюме

Простой поиск можно провести с помощью браузера, например, такую возможность предоставляет Internet Explorer. Однако он проводит поиск не сам, а “нанимает” для этого одну из информационно-поисковых систем; для России это могут быть поисковые серверы Yandex, Rambler, Апорт и, конечно же, ИПС MSN Research компании Microsoft. К этим поисковым ресурсам браузер обращается последовательно. Но на экран при этом выводится ограниченное количество информации о найденных сайтах и документах. Кроме того, при посредничестве браузера можно провести поиск только по ключевым словам, в то время как в некоторых случаях бывает удобнее воспользоваться поисковым каталогом.

Для поисковых каталогов характерно то, что сайты, включаемые в них, просматриваются редакторами каталогов. Это позволяет отсеять малоинформативные и не заслуживающие доверия Web-узлы. Представленные в каталоге сайты на определенную тему зачастую перечисляются в порядке, соответствующем количеству их посетителей (хотя пользователь может избрать и другой способ сортировки). Таким образом, при поиске в каталоге вы можете опираться на опыт других пользователей: если они часто посещают какой-то сайт, значит и вы, скорее всего, найдете для себя на его страницах что-то интересное.

Существуют и каталоги особого рода, так называемые рейтинги, которые представляют ссылки только на наиболее посещаемые сайты (например, на первую сотню наиболее посещаемых по каждой представленной в рейтинге теме). С их помощью пользователь может не только отыскать самые популярные сайты, но и получить интересующую его статистическую информацию. Особенно полезны рейтинги для владельцев сайтов.

Если тема вашего поиска не совпадает с одной из тем, представленных в каталоге, вам придется просматривать сайты нескольких категорий, имеющих отношение к вашей теме, и все равно поиск нужных документов может оказаться безуспешным. Кроме того, каталоги недостаточно часто обновляются, поэтому в них могут оказаться устаревшие, уже не работающие ссылки. Количество сайтов, включаемых в каталог, относительно невелико.

Поисковые машины более оперативны, так как их программы-роботы чаще посещают уже занесенные в базу данных Web-узлы с целью обновления информации о них. Кроме того, они позволяют провести поиск на большем количестве Web-страниц (наиболее известные — уже на миллиардах). Для того чтобы поисковая машина нашла документы на интересующую пользователя тему, он должен ввести в ее поле ввода ключевые слова, которые обязательно должны присутствовать в найденных документах. Ко-

личество ссылок, которые можно получить в результате простого поиска, может составлять от нескольких сотен до нескольких сотен тысяч или даже миллионов. Для того чтобы успешно находить информацию с помощью поисковых машин, желательно научиться составлять запросы, которые уменьшают количество выдаваемых ссылок, т.е. повышают точность поиска. Поиск такого рода называется сложным или расширенным.

Многие порталы также снабжены поисковыми машинами, позволяющими проводить поиск как в пределах портала, так и в Internet. Если пользователя интересует информация общего характера — прогноз погоды, последние новости, курс обмена валют и т.п., — имеет смысл обратиться к какому-либо местному portalу (например, городскому, если таковой существует) и получить на нем всю необходимую информацию.

## Тесты

1. Броузер Internet Explorer способен провести поиск информации в Internet с помощью:
  - а) информационно-поисковой системы MSN Search компании Microsoft;
  - б) российской поисковой машины Yandex;
  - в) встроенной собственной поисковой машины;
  - г) поисковых машин MSN Search, Yandex, Rambler и Апорт.
2. Основное преимущество поиска информации с помощью поисковых каталогов состоит в том, что:
  - а) в каталоге можно найти информацию по любой интересующей вас теме;
  - б) в каталоге обязательно бывают представлены все наиболее значимые Web-узлы по интересующей пользователя теме (если, конечно, она представлена в каталоге);
  - в) база данных поискового каталога столь обширна, что в ней обязательно найдется нужный документ;
  - г) каталоги оперативно обновляются, и в них не бывает “мертвых” ссылок.
3. Среди основных преимуществ поиска с использованием поисковой машины следует особо отметить следующее:
  - а) все включаемые в результаты поиска сайты тщательно проверены редакторами поисковой машины и содержат только достоверную информацию;
  - б) она выдает большое количество ссылок на нужные пользователю документы;
  - в) база данных поисковой машины содержит огромное количество документов (иногда — миллиарды), поэтому вероятность найти среди них нужные достаточно высока;
  - г) поисковая машина имеет привычную древовидную структуру хранения файлов с документами, поэтому найти среди них нужный не составляет труда.
4. Рейтинг позволяет:
  - а) быстро найти самые популярные сайты по определенной теме (если, конечно, она представлена в рейтинге);
  - б) составить достоверный прогноз относительно того, какая команда выигрывает очередной чемпионат мира по футболу;
  - в) не пропустить наиболее интересные передачи по телевидению.