## Поиск информации в интернет

*Общие сведения.*

В настоящее время Интернет объединяет сотни миллионов серверов, на которых размещены миллиарды различных сайтов и отдельных файлов, содержащих различного рода информацию. Это гигантское хранилище информации. Существуют различные приемы поиска информации в Интернет.

Поиск по известному адресу. Необходимые адреса берутся из справочников. Зная адрес, достаточно ввести его в адресную строку Браузера.

Пример 1.

www.gov.ru - сервер органов государственной власти России.

Конструирование адреса пользователем. Зная систему формирования адреса в Интернет, можно при поискеWeb-сайтов конструировать адреса.

К ключевому слову (названию фирмы, предприятия, организации или простому английскому существительному) необходимо добавить домен тематический или географический, при этом необходимо подключать интуицию.

Пример 2.

Адреса коммерческих Web-страниц:

www.cnn.com (всемирные новости CNN),

www.sony.com (фирма SONY),

www.mtv.com (музыкальные новости MTV).

Пример 3.

Адреса учебных заведений:

www.ntu.edu (Национальный университет США).

Пример 4.

Адреса региональных серверов:

www.poland.net (Польша),

www.israil.net (Израиль).

## Поисковые системы Интернет

Для поиска информации в Интернет разработаны специальные информационно-поисковые системы. Поисковые системы имеют обычный адрес и отображаются в виде Web-страницы, содержащей специальные средства для организации поиска (строку для поиска, тематический каталог, ссылки). Для вызова поисковой системы достаточно ввести ее адрес в адресную строку Браузера.

По способу организации информации информационно-поисковые системы делятся на два вида: классификационные (рубрикаторы) и словарные.

Рубрикаторы (классификаторы) - поисковые системы, в которых используется иерархическая (древовидная) организация информации. При поиске информации пользователь просматривает тематические рубрики, постепенно сужая поле поиска (например, если необходимо найти значение какого-то слова, то сначала в классификаторе нужно найти словарь, а затем уже в нем найти нужное слово).

Словарные поисковые системы - это мощные автоматические программно-аппаратные комплексы. С их помощью просматривается (сканируется) информация в Интернет. В специальные справочники-индексы заносятся данные о местонахождении той или иной информации. В ответ на запрос осуществляется поиск в соответствии со строкой запроса. В результате пользователю предлагаются те адреса (URL), на которых в момент сканирования найдены искомые слово или группа слов. Выбрав любой из предложенных адресов-ссылок, можно перейти к найденному документу. Большинство современных поисковых систем являются смешанными.

Наиболее известные и популярные системы поиска:

www.aport.ru www.yahoo.com www.rambler.ru www.yandex.ru www.altavista.com www.google.com

Существуют системы, специализирующиеся на поиске информационных ресурсов по различным направлениям.

Поиск людей в Интернет:

www.whowhere.ru ww. bigfoot.com

Поиск по телеконференциям (Usenet):

www.dejanews.com

Предметные поисковые системы:

www.webring.org

Поиск программного обеспечения:

www.files.com

www.files.ru

Поиск по файловым архивам:

http://ftpseach. city.ru, http://ftpsearch. licos.com

Каталоги (тематические подборки ссылок с аннотациями):

http://www.atrus.ru

www.aup.ru

Часто эффективный поиск информации можно провести с помощью региональных каталогов - специализированных серверов, содержащих данные о предприятиях или Web-ресурсах какого-то города или региона. Например, для Санкт-Петербурга такой каталог располагается по адресу http://www.spb.ru.

Список ИПС можно найти на сайте www.monk. newmail.ru

Более подробный перечень поисковых систем и каталогов представлен в табл. 3.2.

## Правила выполнения запросов

В каждой поисковой системе в разделе Помощь (Help) можно получить сведения о том, как искать, как составить строку запроса. Ниже приведена информация о типовом, "усредненном" языке запросов.

*Простой запрос.*

Ввести одно слово, определяющее тему поиска. Например, в поисковой системе Rambler.ru достаточно ввести: автоматика.

Находятся документы, в которых встречаются слова, указанные в запросе. Распознаются все формы слов русского языка, как правило, регистр букв игнорируется.

В запросе можно использовать символ "\*" или "?". Знаком "?" в ключевом слове заменяется один символ, на место которого может быть подставлена любая буква, а знаком "\*" - последовательность символов.

Например, запрос автомат\* позволит найти документы, включающие слова автоматический, автоматика и т.д.

*Сложный запрос.*

Часто возникает необходимость комбинирования ключевых слов для получения более определенной информации. В этом случае используются дополнительные слова-связки, функции, операторы, символы, комбинации операторов, разделенные скобками.

Например, запрос музыка & (beatles | битлз) означает, что пользователь ищет документы, содержащие слова музыка и beatles или музыка и битлз.

В табл.3.1 приведены правила формирования запросов, принятые в системе Апорт (http://www.aport.ru).

Таблица 3.1

Операторы для формирования запросов

|  |  |  |
| --- | --- | --- |
| Оператор | Синонимы | Комментарий |
| И | AND & | По запросу будут найдены документы, содержащие оба ключевых слова. Его можно и не писать. Например, запрос: информатика и учебник эквивалентен информатика учебник |
| ИЛИ | OR | | Производится поиск тех документов, в которых используется любое из указанных слов или оба слова одновременно |
| НЕ | NOT - ~ | Поиск ограничивается документами, не содержащими слово, указанное после оператора |
| " " | ' ' | Двойные или одинарные кавычки позволяют находить словосочетание |
| Дата= | дата:date= | Поиск ограничивается документами, попадающими в заданный интервал дат.Пример 1. валюта дата=01/02/2002-01/03/2002. По этому запросу будут выданы документы, содержащие слово "валюта" и имеющие дату от 1 февраля 2002 г. до 1 марта 2002 г.Пример 2. date=01/03/2002 валютаПример 3. дата: <02/03/2002 валюта |

Таблица 3.2

Список поисковых серверов и каталогов

|  |  |
| --- | --- |
| Адрес | Описание |
| www.excite.com | Поисковый сервер с обзорами узлов и путеводителями |
| www.alta-vista.com | Поисковый сервер, имеются возможности расширенного поиска |
| www.hotbot.com | Поисковый сервер |
| www.poland.net www.israil.net | Региональные поисковые серверы Польши, Израиля |
| www.ifoseek.com | Поисковый сервер (простой в использовании)  |
| www.ipl.org | Internet Publik library, публичная библиотека, функционирующая в рамках проекта "Всемирная деревня" |
| www.wisewire.com | WiseWire - организация поиска с применением искусственного интеллекта  |
| www.webcrawler.com | WebCrawler - поисковый сервер, прост в обращении |
| www.yahoo.com | КаталогWeb и интерфейс для обращения к полнотекстовому поиску на сервере AltaVista |
| www.aport.ru | Апорт - русскоязычный поисковый сервер |
| www.yandex.ru | Яндекс - русскоязычный поисковый сервер |
| www.rambler.ru | Рамблер - русскоязычный поисковый сервер |
| Справочные ресурсы Интернет |
| www.yellow.com | Желтые страницы Интернет |
| monk. newmail.ru | Поисковые системы различного профиля |
| www.top200.ru | 200 лучшихWeb-сайтов |
| www.allru.net | Каталог русских ресурсов Интернет |
| www.ru | Каталог русских ресурсов Интернет |
| www.allru.net/z09. htm | Образовательные ресурсы |
| www.students.ru | Сервер российского студенчества |
| www.cdo.ru/index\_new. asp | Центр дистанционного обучения |
| www.open. ac. uk | Открытый университет Великобритании |
| www.ntu.edu | Национальный университет США |
| www.translate.ru | Электронный переводчик текстов |
| www.pomorsu.ru/guide. library.html | Список ссылок на сетевые библиотеки |
| www.elibrary.ru | Научная электронная библиотека |
| www.citforum.ru | Электронная библиотека |
| www.infamed.com/psy | Психологические тесты |
| www.pokoleniye.ru | Web-сайт Федерации Интернет образования |
| www.metod. narod.ru | Образовательные ресурсы |
| www.spb. osi.ru/ic/distant | Дистанционное обучение в Интернет |
| www.examen.ru | Экзамены и тесты |
| www.kbsu.ru/~book/ | Учебник информатики |
| Mega. km.ru | Энциклопедии и словари |

## Поиск информации в Интернете: подводные камни

Проблемы, не лежащие на поверхности, нередко дают о себе знать лишь "задним числом", после того как определенный этап поисковых работ завершен и, возможно, исходя из его результатов уже принято какое-либо решение. Что же мешает сделать ситуацию прозрачной с самого начала эксплуатации той или иной информационно-поисковой системы (ИПС)? Ответ довольно прост: отсутствие исчерпывающей информации подобного рода со стороны разработчика. Прямым следствием этого становятся недостоверность получаемых данных и их неконтролируемая потеря. Редко удается встретить в Сети поисковую систему, которая не обладала бы некоторыми "недокументированными" особенностями. Казалось бы - пользователю необходимо не так уж много сведений, а именно:

как происходит наполнение базы данных ИПС и каков ее объем;

полный спектр возможностей поискового языка системы;

основные особенности представления результатов поиска, прежде всего алгоритма ранжирования записей из списка отклика на поисковый запрос.

Увы, источником подобной информации обычно является не документ, доступный с головной страницы поискового сервера, а разбросанные по Сети, книгам и компьютерным журналам публикации отдельных авторов. К причинам такого положения дел, по-видимому, можно отнести не только небрежность разработчика, но и фактор, именуемый маркетинговой политикой. Проще говоря, предоставление поисковой системой наиболее полной информации о самой себе не всегда положительно сказывается на ее рейтинге. Тем не менее, взять ситуацию под контроль в ряде случаев пользователю оказывается вполне по силам. Выяснить особенности работы избранного поискового сервиса часто удается с помощью тестирования. Построение специальных тестовых запросов, быстро проясняющих именно тот аспект работы системы, который наиболее важен для текущей задачи, во многих случаях оказывается нетривиальным. Тому, как избежать некоторых неприятностей при работе с ИПС, мы и посвятим наше обсуждение. В качестве примеров, иллюстрирующих изложение, будут рассмотрены широко известные поисковые системы Интернета.

Любая поисковая машина или каталог регламентирует свою работу по сбору данных из Сети. Очевидно, что формирование поискового образа информационного объекта, или, другими словами, его "отражения" в "зеркале" поисковой системы, неизбежно связано с некоторыми искажениями. По сути, главным при этом становится вопрос о том алгоритме, на основе которого создается поисковый образ. Объектом-оригиналом при этом может стать как Web-страница, так и файл "закрытого" формата, который не доступен для проникновения сканирующих программ ИПС, например видео - или аудиозапись. Определенный шаблон обычно используется и при построении поискового образа для физического или юридического лица в момент его регистрации в поисковой службе. Отсечение, фильтрация информации от оригинала свойственны всем без исключения ИПС, в том числе и полнотекстовым системам глобального охвата и самого общего назначения.

Фильтрация может регламентироваться как на техническом, так и на лингвистическом уровне, однако задача у нее одна - при минимальных материальных затратах добиться реальной эффективности поиска.

В связи с этим на практике часто возникает вопрос - что становится причиной неудачного поиска: высокая ли вероятность отсутствия в Сети на данный момент времени информации, релевантной запросу, или то, что эта информация потенциально не доступна для рассматриваемой поисковой системы. "Подводным камнем" этот аспект становится, когда получен ненулевой отклик на поисковый запрос, а доля недополученных данных оказывается неконтролируемой. Некоторый свет на особенности работы глобальных ИПС проливает сравнительный анализ их возможностей, который был приведен в прошлой публикации. Однако, если детали алгоритма фильтрации не известны, наиболее чувствительные потери данных возникают именно при использовании специализированных поисковых служб.

Рассмотрим несколько примеров. Немало специализированных систем имеет собственный интерфейс для ввода поисковых запросов. Тем не менее можно считать веянием времени ситуацию, когда многие подобные сервисы интегрируются в шаблоны глобальных ИПС в виде фильтров. Такими возможностями всегда отличался HotBot; недавно соответствующие элементы были внедрены на AltaVista; есть они и на Еxcite. Постоянно расширяется набор фильтров поисковой системы Lycos (см. рис.1), на которой мы остановимся подробнее.

Представьте себя на месте пользователя, впервые посетившего такую известную глобальную поисковую систему, как Lycos, с целью найти в Сети сведения о некоем книжном издании. Введя соответствующие ключевые слова и выбрав фильтр Books, он получает отклик, который, при отсутствии дополнительной информации, нельзя расценить иначе, как получение данных о книгах, собранных по всему Интернету. Интересно было бы задать вопрос, а может ли в масштабе Сети автоматически вестись отбор подобных сведений? Если говорить только о пространстве WWW, то в большинстве случаев программы-пауки, сканирующие Сеть, используют для распознавания типа данных специальные элементы языка HTML, с помощью которых в Web-страницу внедряются определенные информационные блоки. Название элемента может нести смысловую нагрузку и отождествляться с типом информации. Так, если бы гипотетически существовал элемент HTML book, заключающий в себе сведения о книге и ее авторе, он мог бы размещаться на странице и в простейшем случае иметь следующий вид:

<book>Название книги и автор</book>

(сами элементы <book> в окне браузера не должны отображаться) При этом вся информация о книгах, публикуемая в WWW подобным образом, могла бы благополучно и без участия человека накапливаться в базе данных ИПС. Но элемента book в стандарте HTML пока не существует. Следовательно, приходится прибегать либо к "ручному" отбору, либо к автоматическому просмотру некоторых, заданных наперед каталогов отдельных узлов, возможно, имеющих отношение к продаже книжной продукции или к библиотекам.

В случае Lycos все гораздо проще. Поиск происходит всего-навсего по одному-единственному узлу компании (http://www.barnesandnoble.com), заинтересованной в реализации своего товара. К чести разработчика следует сказать, что после нескольких лет молчания по поводу фильтра "books" в недрах предлагаемой документации сегодня можно найти скромное упоминание об арендаторе фильтра. Ранее его владельца просто нельзя было идентифицировать, и только спустя некоторое время стало понятно, что система работает с довольно незначительной по объему и специфически пополняемой базой данных.

Не менее серьезно звучат опасения в случае, когда поиск связан с информацией, привязанной к определенному формату ее хранения, например к звуковым файлам. В течение нескольких месяцев поиск "звуков в Интернете" на Lycos оставался чем-то таинственным, напоминающим работу с небольшой, но со вкусом собранной коллекцией. Тестирование системы с помощью простых запросов показывало, что в основном в ней представлены форматы WAV и AU. Недавно стало известно, что теперь поддерживаются также и MP3, MID, RA, RAM и AIF. При этом объем накопленных записей, доступных через большинство фильтров, продолжает сохраняться в тайне.

Ясно, что, если интересующий вас формат не входит в поддерживаемый на данный момент системой перечень, вы получите нулевой отклик, причину которого следовало бы четко представлять с самого начала.

Происхождение сопроводительных записей к звуковым файлам на Lycos, которые отображаются в результатах поиска, по-прежнему не регламентировано разработчиком.

Аналогичные проблемы существуют и на других ИПС. Хотелось бы отметить типичный в этом отношении прием: использование шаблона глобальной ИПС как для поиска информации, относящейся ко всему Интернет-пространству, так и для поиска по некоторым избранным базам данных или коллекциям. К сожалению, реальное поле поиска оговаривается далеко не всегда, и часто его приходится выяснять самостоятельно во избежание неверных выводов в дальнейшем

Ситуация может осложниться тем, что на поисковом сервере вы не найдете исчерпывающего описания того, как работают операторы языка запросов.

C этим можно столкнуться даже на "зрелых", не первый год работающих ИПС. Рассмотрим на примере AltaVista, каким образом это может стать источником определенных проблем.

Несмотря на недавнее появление графического фильтра (рис.2), многие пользователи системы продолжают эксплуатировать прозрачный по своей природе оператор *image*, позволяющий находить в индексе графические файлы. На этот счет справка AltaVista исчерпывается тем, что рекомендует ввести в шаблон запрос, в котором вслед за указанным оператором должно следовать имя или часть имени искомого файла. Таким образом, для поиска файла с изображением акрополя следует задать запрос в виде *image: acropolis*.

Увеличит ли наши шансы на успех знание того, как реально отрабатывает оператор *image?* Если посмотреть на откликнувшиеся документы, а затем на их HTML-источник, то легко убедиться, что в каждом из них в месте вставки графического образа присутствует элемент <IMG>. Внутри него в качестве обязательного атрибута стоит URL, с которого, собственно, и извлекается сам файл:

<IMG SRC="http://citforum.ru/buildings/acropolis. gif">

Фактически же Web-страница дает отклик, если ключевое слово входит не только в имя файла, но и в название любого каталога и в доменное имя сервера, содержащихся в URL элемента <IMG>, то есть документ, включающий в себя приведенную выше строку, откликнулся бы и на запрос *image: buildings*. Следовательно, поиск по имени каталога, которое так же, как и имя файла, несет смысловую нагрузку, позволяет получить графические данные, которые нельзя извлечь в первом случае. Предположим, что Web-мастер неосторожно назвал искомый файл ACR1. GIF, но разумно положил его в каталог buildings. Тогда по запросу *image: buildings* могут откликнуться релевантные документы с изображением акрополя, вставленным в Web-страницу с помощью строки:

<IMG SRC="http://www.citforum.ru/buildings/acr1. gif">

В расширенном поиске AltaVista используются логические операторы и скобки. Однако на сервере ничего не говорится о том, допустимо ли применять их внутри специальных полей поиска, таких как поле *image*. Уже заведомо зарегистрированный в индексе графический файл, найденный ранее, можно использовать для проверки работоспособности отдельных поисковых запросов. Так, если предположить, что файл с URL из последнего примера существует, то тестовый запрос в виде *image: (buildings AND acr1)* должен дать корректный ненулевой отклик и таким образом подтвердить, что комбинирование операторов допустимо. На практике это действительно возможно.

Хотелось бы еще раз подчеркнуть, что речь здесь идет не о несовершенстве отдельных поисковых систем, а о конструктивном подходе к разрешению возникающих вопросов. При этом нередки и ситуации, предугадать которые крайне сложно.

Если, скажем, на той же AltaVista организовать поиск по ключевому слову "президент" (оно специально выбрано в качестве тестового как довольно распространенное), легко убедиться, что отклик зависит от двух факторов: какой язык выбран в меню шаблона (см. рис.2, справа вверху) - русский (Russian) или любой (any language), а также какая русская кодировка установлена в меню браузера. Результаты поиска приведены в табл.1. Анализ списка отклика показывает, что, во-первых, при вводе запроса только в одной кодировке неминуемо теряются данные. Во-вторых, становится ясно, как система идентифицирует тот или иной язык документа. Оказывается, если некоторая начальная часть документа написана на языке, отличном от русского, то этот документ уже не описывается ИПС как русскоязычный. Результат этой недокументированной особенности - максимальный отклик индекса при поиске по русскоязычному термину достигается при установке пункта меню "any language", а не "Russian".

В шаблоне расширенного поиска популярной бизнес-ориентированной системы Open Text Livelink Pinstripe (OTLP) (рис.3) также скрыты некоторые проблемы, никак не освещенные в справочном материале ИПС.

Как видно из рисунка, шаблон позволяет задать свое поле поиска для каждого термина, а затем связать термины с помощью логических операторов. Однако как только терминов становится больше двух - возникает вопрос: в какой последовательности будут отрабатывать операторы и, соответственно, что будет представлять собой результат. Даже для такого простого запроса, как *term1 AND term2 OR term3*, разумно предположить двоякую интерпретацию, которую можно проиллюстрировать с помощью выделения в скобки логических единиц (в самом шаблоне скобки не применяются). И вариант (*term1 AND term2) OR term3*, и вариант *term1 AND (term2 OR term3)* кажутся приемлемыми, давая при этом совершенно разный отклик. Тестовый запрос и последующий анализ откликнувшихся документов показывают справедливость первого варианта, то есть то, что операторы выполняются по мере их появления в шаблоне и в документе будут присутствовать либо *term1* и *term2* одновременно, либо только *term3*. Как в таком шаблоне вводить запросы с участием фраз (а это возможно) - автор предлагает выяснить читателям самостоятельно. В данном случае приходится констатировать очевидную небрежность разработчика по отношению к пользователям системы.

Подавляющее большинство ИПС Интернета сегодня активно работает с так называемыми стоп-словами (stop-words). К последним относят служебные части речи, которые не несут смысловую нагрузку, а также некоторые наиболее общеупотребительные в Сети слова, такие как information, Internet, Web, business, и другие. Известно, что AltaVista, Excite, HotBot и Lycos применяют в работе технику стоп-слов, а Infoseek и NorthernLight ее не практикуют.

При появлении стоп-слов в поисковом запросе, не содержащем специальных ухищрений, ИПС может не учитывать их при поиске и ранжировании результатов, при этом иногда информируя об этом пользователя, а иногда - нет. В целом неучет стоп-слов при обработке запроса сокращает время поиска и повышает релевантность отклика. Однако стоит вам попробовать отыскать что-нибудь вроде классической фразы Шекспира "to be or not to be", состоящей только из стоп-слов, - и вы уже не владеете ситуацией.

Хотя стоп-слова и могут игнорироваться в простых запросах, в индексе полнотекстовой ИПС они присутствуют наряду с остальными. Такой системой является, например, AltaVista (индексируются все слова документа). HotBot, напротив - индексирует все, кроме стоп-слов.

Тем не менее и HotBot выполняет полнотекстовое индексирование отдельных значимых полей документа, так что запросы со стоп-словами, оформленные в виде фразы, дают и на этой ИПС результативный отклик.

Перечень стоп-слов не стандартизован, так что он может быть оригинальным для каждого сервиса. Разработчики редко приводят сведения об этом аспекте работы ИПС, однако при необходимости поиск по ключевым словам *stop, words* плюс название интересующей вас поисковой машины позволяет обнаружить в Сети версии соответствующих перечней.

Наиболее общие принципы выхода из проблемной ситуации следующие: по возможности избегать употребления стоп-слов в запросах, исключить применение логических операторов типа and, or, not и других в тех шаблонах, в которых они не поддерживаются и будут восприняты как стоп-слова.

Если же без стоп-слов в запросе обойтись нельзя, то следует включить их во фразу, что во многих системах означает заключение в кавычки. В отдельных случаях полезно протестировать работу шаблонов простого и расширенного поиска ИПС, в которых техника поддержки стоп-слов может быть различной

Самая захватывающая интрига Сети, которую порождают ИПС, связана с особенностями работы алгоритма, ранжирующего результаты в списке отклика. Эти сведения обычно не предаются широкой огласке, но они крайне необходимы Web-мастерам, продвигающим в суровой конкурентной борьбе свои узлы через поисковые системы Интернета. Попасть в первые несколько десятков записей из списка отклика на ИПС по часто повторяющимся в Сети запросам - значит обеспечить свою доступность для потенциальных клиентов (см. КомпьютерПресс №5’99, с.114).

Тем не менее, и при решении поисковых задач во время работы со списком отклика из-за недостатка информации также могут возникать некоторые проблемы.

В предыдущем выпуске мы говорили о том, что простые тестовые запросы позволяют с самого начала работы с ИПС понять, насколько широко в индексе представлена искомая информация. Однако не всякая ИПС дает полное число документов, содержащихся в отклике на запрос (например, Lycos, не дает). В какой-то мере это позволяет системе сохранить свое лицо, избежав сравнения с гигантами - Northern Light, AltaVista или HotBot. При решении профессиональных поисковых задач к таким сервисам следует обращаться в последнюю очередь.

Обычно в списке отклика появляется информация, которая включает в себя заголовок страницы, адрес и аннотацию. Аннотация берется либо из специального META-элемента, задаваемого автором документа, либо в этом качестве выступают несколько первых нередактируемых строк текста, взятых со страницы. В некоторых случаях указывается язык документа. Выше мы уже обращали внимание на проколы алгоритма AltaVista, связанные с идентификацией языка, и подобные случаи - не редкость и на других ИПС.

Другая обескураживающая неприятность - это возможное отсутствие в найденных документах тех самых ключевых слов, по которым проводился поиск. Причиной подобного явления, если не считать незарегистрированного обновления страницы без изменения адреса, становится тот факт, что ключевые слова были заданы автором в специальном поле - элементе META. Оно доступно для сканирования роботом ИПС, но не отображается на странице. В этом случае путем просмотра метаэлементов HTML-источника у вас есть возможность убедиться в недобросовестности автора: несоответствие ключевых слов содержанию документа - это прямая дезинформация.

Еще одна проблема вообще не очевидна для единичного пользователя. Речь идет о том, как поисковый сервер обрабатывает запросы в случае, когда их поступает слишком много, то есть в режиме переполнения. Так, автору статьи не раз приходилось сталкиваться с тем, что, например, на AltaVista при одинаковом и практически одновременном тестовом запросе с 10-15 компьютеров количество результатов, появляющихся в отклике для каждого пользователя системы, иногда может различаться на десятки тысяч. В действительности, попадая в режим перегрузки, поисковый сервер не имеет большого выбора, а именно: он либо отклоняет запрос, либо обслуживает его по "сокращенному" варианту. Последний вполне может предполагать предоставление лишь части удовлетворяющих запросу данных. Выход очевиден: проверять достоверность отклика ИПС многократно и в разное время суток.

Нам хотелось бы остановиться на некоторых более чем реальных опасностях, которые подстерегают пользователя, доверившегося малоизвестному поисковому серверу. Написать об этом автора заставил такой случай. Человеку была срочно необходима информация о наличии прямых электропоездов между двумя городами СНГ. Воспользовавшись каталогом Rambler, он быстро сумел локализовать сервер, предлагающий необходимые сведения (рис.4).

http://pavel. physics. sunysb.edu: 8080/

После введения станций отправления и назначения система ответила отрицательно (см. рис.4, строка внизу). Такой категоричный ответ сервера заставил человека прекратить дальнейшие поиски и принять решение, о котором ему скоро пришлось пожалеть. Предъявить претензии к разработчику системы также оказалось невозможным. Дело в том, что чуть ниже под результатом поиска пользователем не была замечена одна важная деталь, а именно надпись "Расписание рекламное, возможны изменения, за которые не несут ответственности ни распространитель, ни МПС". При этом если бы фраза об отказе была сформулирована чуть мягче, пользователь, вероятно, смог бы продолжить поиск в Сети и достичь положительного результата.

В некоторых случаях маркетинговая агрессивность разработчика начинает носить вызывающий характер. Вот уже не один месяц на серверах HotBot и AltaVista находится рекламное объявление крупнейшей книготорговой компании Amazon (http://www.amazon.com), а также ряда других. При этом на любой запрос в ИПС рядом с результатами поиска появляется баннер, намекающий на то, что как раз по тематике выполненного поиска и можно найти информацию на Amazon, даже если в запросе фигурировал мистический "господин Иванов" (см. рис.5).

Подстановка терминов из поискового шаблона в баннер производится путем их механического переноса и безо всякого контроля на предмет действительного наличия книг по данной тематике на сервере компании. К тому же найти "Иванова" на Amazon нельзя в принципе, поскольку вплоть до последнего времени русскоязычная литература там не продавалась. В данном случае плата за доверчивость - это несколько минут напрасно потраченного времени.

Таким образом, от привычного уважения к печатному слову в Сети лучше отказаться, особенно если сервер генерирует реплики автоматически.