

Загуменов А. П.

*Как раскрутить
и разрекламировать*

Web-сайт

в сети Интернет

3

**причины
для покупки
этой книги:**

- Вы увеличите посещаемость своего сайта
- Вы увидите свой сайт в первых строках Яндекс и Rambler
- Вы сможете автоматически управлять сайтом

DMK
www.dmk.ru

Александр Петрович Загуменов

Как раскрутить и разрекламировать Web- сайт в сети Интернет

Текст предоставлен издательством

http://www.litres.ru/pages/biblio_book/?art=432502

Как раскрутить и разрекламировать Web-сайт в сети Интернет:

ДМК Пресс; Москва; 2005

ISBN 5-94074-027-8

Аннотация

Настоящая книга заинтересует всех, кто столкнулся с вопросами подготовки, размещения в Сети и популяризации Internet ресурсов различного уровня: от домашней странички до корпоративного сайта. В ней вы найдете все, что необходимо для оптимизации Web сайтов под поисковые системы: приемы написания Web-страниц, описание множества самых популярных специализированных программ, предназначенных для подготовки сайта и его раскрутки, создания удачного HTML-кода страниц с правильными метаданными. Книга является практическим руководством для разработчиков Web сайтов и всех, занимающихся их продвижением. Автор приводит множество советов, касающихся создания и анонсирования Web страниц. Рассмотрены средства автоматизации для

повышения эффективности разработки и маркетинга при создании и обслуживании сайта. Описание программных и сетевых средств, автоматизирующих процессы тестирования и отладки сайта, обеспечивающих проверку работоспособности и корректности гиперссылок, синтаксиса HTML кода и грамматики размещенного на странице текста, занимает центральное место в книге. Подробно излагаются возможности таких программ, как Linkbot Developer Edition, Domain NameChecker, Retrieve, CyberSpyder Link Test, HTML Link Validator, CSE HTML Validator, A Real Validator, MetaTag ToolKit, MetaMan, WebQA. Отдельная глава посвящена регистрации Web ресурсов в поисковых системах и каталогах. Описываются программы автоматической регистрации (WebPosition, Page Promoter, Web Регистратор), способы взаимодействия с индексирующими роботами поисковых машин, правила применения метаданных. Рассматриваются приемы и методы рекламы сайтов в Internet, указаны критерии ее эффективности. Издание рассчитано на широкий круг читателей и будет полезно как начинающим создателям Web сайтов, так и профессионалам, которые хотят научиться более качественно продвигать в Сети свой Web продукт.

Содержание

Введение	6
Глава 1	9
Собственное доменное имя	13
Разработка семантического ядра	17
Обновление информации	19
Заголовки	21
Метаданные	23
Метатэги – эквиваленты HTTP-заголовков	25
Свойство Expires	25
Свойство Pragma	28
Свойство Content-Type	28
Немного о кодировках	29
Свойство Content-language	32
Свойство Refresh	34
Свойство Window-target	35
Свойство Content-Script-Type	35
Свойство Content-Style-Type	36
Метатэги для поисковых машин	36
Ключевые слова и описания	39
Язык документа	49
Языковые варианты документа	49
Начало набора	50
Инструкции для роботов – файл	51

robots.txt	
Тэг МЕТА для роботов	56
Другие метатэги	59
Проверка ссылок	60
Конец ознакомительного фрагмента.	62

Александр Петрович Загуменов Как раскрутить и разрекламировать Web- сайт в сети Интернет

Введение

Настоящее издание заинтересует всех, кто разместил в Internet Web-страницу или целый сайт и стремится привлечь к ним посетителей или своих клиентов. В книге рассматриваются подготовка и проведение рекламной кампании Web-ресурсов, анализ ее результатов. Особое внимание уделено оценке популярности Internet-страниц, исправлению ошибок и поддержке посещаемости на высоком уровне. Указанные задачи вполне поддаются автоматизации при помощи программного обеспечения, а также средствами Internet, чему и посвящена большая часть страниц издания.

В главе 1 рассказывается обо всех аспектах раскрутки сайтов, выборе доменного имени, разработке семантического ядра запросов, о том, как управлять индексированием стра-

ниц в поисковых системах, о пользе регистрации в каталогах и обмена ссылками с партнерами, о почтовых рассылках, о влиянии баннерной и контекстной рекламы на посещаемость сайта. Большое внимание уделено описанию метаданных, применяемых для оптимизации Web-сайтов под поисковые системы Internet. Описаны различные методы обнаружения технических ошибок при размещении сайтов в Сети, – неверных ссылок, погрешностей HTML-кода, ошибок правописания. Предлагаются разнообразные способы всесторонней проверки страниц.

Глава 2 посвящена автоматизации всех этапов работы над сайтом по его продвижению в Internet. То, о чем написано в главе 1, здесь подкреплено описанием возможностей ресурсов Internet и таких программ, как Linkbot Developer Edition, Domain Name Checker, Retrieve, CyberSpyder Link Test, HTML Link Validator, CSE HTML Validator, A Real Validator, MetaTag ToolKit, MetaMan, WebQA.

В главе 3 рассмотрены методы и средства анонсирования сайта в поисковых системах, благодаря которому сайт при поиске по ключевым словам попадает в первую десятку найденных. Подробно описываются наиболее популярные отечественные и зарубежные поисковые машины. Поиск в каталогах часто обеспечивает лучшие результаты, нежели в автоматических поисковых системах, поэтому и они не обойдены вниманием, описаны способы добавления сайта в разнообразные каталоги, приведены методы работы с со-

ответствующим программным обеспечением: AI RoboForm, WebPosition, Page Promoter, Web-Регистратор.

Прочитав первую главу книги и воспользовавшись какими-либо средствами, представленными во второй и третьей ее главах, вы можете и не читать книгу дальше, так как у вас будет информация и инструменты для практической работы. Однако если вы хотите увидеть не только направление, в котором движетесь, но и результаты своей работы, вам потребуется ознакомиться с главой 4, которая повествует о том, какую важную и интересную информацию можно узнать из журнальных файлов сервера, и посвящена методам их обработки с помощью соответствующего программного обеспечения.

В приложении к книге приводится список полезных ссылок на сайты, рекомендуемые к посещению.

Книга предназначена для лиц, столкнувшихся с вопросами размещения и продвижения Web-ресурсов независимо от их уровня: будь то домашняя страничка или корпоративный сайт.

Вопросы к автору присылайте по электронному адресу: www@inotec.ru, посетите его сайт <http://promote.inotec.ru>.

Глава 1

Этапы большого пути

Хороший web-сайт – это не просто набор страниц, связанных гиперссылками, и далеко не только то, что видит пользователь на экране монитора. Его внутреннее устройство довольно сложно. Ведь требуется обеспечить максимум удобств, как для посетителей, так и для его владельцев, которые будут управлять сайтом. Поэтому при его создании должны быть учтены тысячи мелочей, что, собственно, и отличает хороший web-сайт от плохого.

Но самое главное – сайт должны найти люди, он не должен затеряться на бескрайних просторах Сети. А иначе, какой от него толк, если сайт никто не увидит? Пусть он даже фантастически красив.

Грамотно построенный web-сайт – довольно сложная конструкция, пронизанная многочисленными внутренними и внешними связями, с четкой и жесткой логикой. Уже в процессе разработки сайта должны быть заложены механизмы, работающие на его будущую популярность.

Internet – мощный и стремительно развивающийся канал для продвижения информации, товаров и услуг. Однако чтобы добиться успеха, недостаточно лишь разместить в Сети свой сайт. Его необходимо разрекламировать или, как гово-

рят, «раскрутить». Вот этим мы с вами и займемся.

Чтобы сайт работал, и с его помощью появлялись новые клиенты, его следует правильно спроектировать, учтя все детали, которые важны для поисковых систем, ведь основной поток целевых посетителей идет именно через них. Если считать основной задачей сайта привлечение новых клиентов или читателей, следует начинать работу с анализа ключевых слов, по которым будут искать сайт. На его основе строится так называемое семантическое ядро возможных запросов, с учетом которого проводится разработка страниц и подготовка исходных данных.

Далее выполняются следующие работы по оптимизации сайта для поисковых систем:

- формируются заголовки HTML-страниц, которые видны в заголовочной части окна браузера;
- создаются метатэги с описаниями и ключевыми словами. Их не видят посетители сайта, но учитывают многие поисковые системы;
- изготавливается карта сайта, которая очень важна для поисковых машин, так как она позволяет быстрее и с большим весом проиндексировать страницы сайта. Иногда карта сайта не видна посетителям;
- вырабатываются альтернативные подписи ко всем иллюстрациям. Альтернативная подпись возникает на экране при наведении мыши на изображение. Так как эти подписи также индексируются поисковыми системами, их следует на-

сыщать ключевыми словами;

- создаются дублирующие переходы. Многие поисковые системы не понимают переходов по формам с выпадающими списками, поэтому нужно сделать дополнительные переходы в виде обычных ссылок;

- специальным образом организуется архив новостей. Появление новой информации должно приводить к обновлению всех страниц архива, что очень привлекательно для поисковых машин;

- формируется дублирующая навигация в виде цепочки гиперссылок, если дизайн сайта допускает такой подход, что не только улучшает удобство пользования сайтом, но и позволяет увеличить количество ключевых слов;

- навигационные элементы сайта по возможности делаются текстовыми. При этом используются слова из семантического ядра запросов;

- новости выводятся на главную страницу, в результате чего информация на главной странице постоянно обновляется, а это благоприятно сказывается на результатах поиска;

- на всех страницах, включая главную, организуются информационные блоки. Они не являются обязательными, но заполнение их информацией и периодическая ее смена очень полезна для повышения рейтинга в результатах поиска;

- организуются места для баннеров каталогов и счетчиков. При продвижении сайта обязательно придется пропи-

сывать его в различных каталогах. Чтобы не пришлось делать крупных доработок, заранее следует закладывать эту возможность;

- ссылки с сайта делаются такими, чтобы в них присутствовали ключевые слова.

Конечно, это не все приемы, которые используются при оптимизации сайта для того, чтобы он занимал первые места в результатах поиска. Поэтому, рассматривая вышеназванные приемы, мы поговорим и о других.

Собственное доменное имя

Доменное имя наряду с логотипом и торговой маркой является частью бизнес-идентичности компании. При правильном выборе оно ассоциируется у покупателей именно с вашей компанией, ее продуктами, услугами и т. д. Часто доменные имена используются для нескольких целей, включая имя web-сайта, адрес электронной почты, имя FTP- или другого сервера. Собственное доменное имя придаст вам дополнительную солидность и доверие. Доверие – это одна из основ вашего сайта.

Если доменное имя легко запоминается, вы сможете привлечь больше покупателей. Люди будут использовать сайт или адрес электронной почты для быстрого и удобного доступа к информации или услуге, которую вы предлагаете.

Доменные имена мобильны. Вы можете сменить адрес, номера телефонов, провайдера Internet-услуг, но доменное имя не изменится. Это дает возможность постоянно поддерживать контакты с клиентами и пользователями ваших услуг.

Полная длина доменного имени, как правило, составляет от 2 до 67 символов, включая точку и имя домена первого уровня. Имя может состоять из следующих символов:

- латинские буквы (от а до z);
- арабские цифры (от 0 до 9);

- дефис (-), если доменное имя не начинается и не оканчивается этим символом.

Доменные имена не чувствительны к регистру букв, так что часто удобно использовать в рекламе прописные буквы. Например, www.ShoppingWorld.com. В настоящее время можно регистрировать и многоязычные доменные имена.

При выборе имени ориентируйтесь на свою целевую аудиторию. В зависимости от этого имя может быть популярным, веселым, строгим, состояться из профессиональных терминов и т. д.

Доменное имя должно быть простым и коротким. Как правило, чем оно короче, тем легче его запомнить или записать, и тем меньше вероятность того, что его воспроизведут неправильно. При остальных равных условиях отдавайте предпочтение словам с меньшим количеством слогов и словам, которые легко произносить. Не используйте сложные для запоминания акронимы, такие как qmxf.

Чтобы получить более высокий рейтинг у некоторых поисковых машин, регистрируйте доменные имена, составленные из ключевых слов вашего web-сайта. В этом случае лучше разделять слова дефисами. Можно создать и зарегистрировать отдельную страницу с таким длинным именем и автоматически переводить посетителей на основной сайт.

Употребляйте подходящие слова во множественном числе. Например, searchengines.com (поисковые машины) лучше, чем searchengine.com (поисковая машина), если только

web-сайт не имеет отношения к какой-то отдельной поисковой машине.

Не забывайте о перспективе. Если в будущем вы планируете выпускать новые продукты, позаботьтесь, чтобы доменное имя соответствовало как новым, так и первоначальным продуктам и услугам. Эксперты по Internet-маркетингу утверждают, что web-сайт должен иметь только одну тематику. Сайт широкого профиля не очень привлекает посетителей, которые ищут конкретную информацию.

Регистрируйте доменное имя в правильном домене – глобальном или региональном. Для глобальных сайтов используйте домены .com, .org или .net. Для региональных – соответствующие домены страны, для России – .ru или .org.ru. Расширения .com и .net предпочтительны для бизнес-сайтов. Доменные имена .org традиционно используются некоммерческими организациями. Обратите внимание, что многие домены имеют ограничения по выбору имен. В некоторых доменах можно использовать только имя, производное от названия организации или зарегистрированной торговой марки. Можно зарегистрировать несколько доменных имен в разных доменах и направлять их на один и тот же сайт.

Проверяйте, не посягаете ли вы на торговую марку или имя, которое по праву принадлежит другой компании или отдельному человеку. Не всякое доступное доменное имя следует регистрировать и использовать.

По возможности не используйте доменные имена, похо-

жие на другие. Ведь нужно, чтобы посещали именно ваш сайт, а не конкурента! Исключением из этого правила будет ситуация, когда с целью привлечения дополнительных покупателей вы регистрируете доменное имя, состоящее из слов с распространенными орфографическими ошибками.

Чтобы проверить доступность конкретного доменного имени, можно воспользоваться сетевыми сервисами или специализированными программами, которые мы рассмотрим в следующей главе.

Разработка семантического ядра

Процедура проработки семантического ядра запросов не вполне технологична и зиждется, скорее, на интуитивных догадках исследователя, чем на строгих выкладках. Существуют способы получения данных о популярности того или иного слова, но нет четкого механизма определения, являются ли люди, воспользовавшиеся им, вашими целевыми посетителями. Кроме того, далеко не всегда ясно, а какие, собственно, слова нужно рассмотреть. То, что кажется очевидным продавцу, может совсем иначе представляться покупателю.

При анализе семантического ядра особое внимание следует уделить запросам с низкой частотой появления. Очень часто практически невозможно позиционировать сайт по запросам, имеющим высокую популярность, зато это достаточно просто делается по большому количеству запросов низкочастотных. А по общей сумме удачных запросов результат продвижения может оказаться и лучше.

В итоге работы над семантическим ядром должна быть составлена таблица, в которой отражается:

- запрос (слово или выражение);
- частота запроса (его популярность);
- количество сайтов, найденных по запросу;
- место в результатах поиска первого реального конкурента;

та.

Такие таблицы составляются на основании данных, по крайней мере, трех основных российских поисковых систем Яндекс, Рамблер и Апорт, а также Google и Yahoo!. После этого отбираются те ключевые слова, по которым имеет смысл бороться за место на первых трех страницах результатов поиска: многочисленные исследования показывают, что пользователи редко заглядывают дальше.

Старайтесь использовать эти слова, где только возможно (но не злоупотребляйте). Этот прием действенен и очень важен. Лучше уделить больше внимания продумыванию ключевых слов на начальной стадии разработки сайта, чем исправлять их позже, теряя гораздо больше драгоценного времени.

Обновление информации

Устаревшая информация – это очень серьезная проблема для владельца сайта. Ведь множество сайтов, сходных по тематике с вашим, продолжают постоянно обновляться. Сейчас же, с развитием электронной торговли, особенно важно не потерять доверия посетителей, а показать им устаревшую информацию – верный способ дать понять, чего стоят ваш сайт и ваши услуги. Обратите внимание на то, что архивы и информация о старом товаре или продукте – как раз достоинство сайта, а не недостаток, и они очень сильно отличаются от устаревшей информации.

Если вы описываете в документе некую сложную часть системы или чувствуете, что читатели рады иметь хоть какие-то сведения о данном предмете, нет смысла в детальной проверке такого документа. Если есть люди, действительно нуждающиеся в этой информации, они могут смириться с некоторыми недочетами и просмотреть весь материал, чтобы понять ваши идеи. Такой путь может быть весьма эффективным. Обратите на это внимание, потому что существует определенный слой информации, предназначенной для быстрого ознакомления или создания новых файлов на ее основе, и она важна для будущих читателей. Лучше, чтобы эта информация была доступна хотя бы в незаконченной форме, чем отсутствовала вовсе. До появления электронных

технологий усилия на публикацию «сырого» материала приводили к браку и могли быть приняты за оскорбление читателя, так как выглядели как издание низкого качества. В настоящее время публикации идут на всех уровнях, и имеют свою ценность, как документы высокого качества, так и недоработанные. Поэтому обратите внимание читателей на текущий уровень подготовки документа, дабы избежать разочарования.

Заголовки

Корректные заголовки, кратко отражающие общую тематику, облегчают поиск Internet-ресурсов. Использование каждого слова в названиях страниц следует тщательно обдумать. Очень важно грамотно составить начальные фразы; это требование особенно касается текста внутри тэгов `<HEAD><TITLE> . </TITLE > . </HEAD>`, так поисковые системы, отвечающие на запросы пользователей, работают в первую очередь с этой лексикой. У всех документов может быть только один заголовок – он должен идентифицировать содержание в довольно широком контексте. Обычно именно эти фразеологические единицы видны в результатах запросов, и именно по ним читатели будут определять, стоит ли щелкнуть по ссылке или продолжить поиск.

Заголовок не является частью текста документа. Он не может содержать ссылок, знаков параграфа или подчеркивания. Идеальная длина заголовка – не более 64 символов. Как и прочие смысловые элементы текста Web-страницы, заголовки должны быть написаны иначе, нежели для других средств массовой информации: ведь это часть пользовательского интерфейса и навигационный элемент сервера. Если вы учитываете данный фактор, то многие приложения могут показать заголовки ваших документов в соответствующем окне, в меню, закладках, – другими словами, там, где

место ограничено. Специальных ограничений на длину заголовков не существует, они легко генерируются автоматически, но слишком длинные могут быть обрезаны.

Заголовок страницы должен содержать как можно больше ключевых слов и в то же время не выглядеть тяжеловесным.

Часто заголовки вырезают из контекста, чтобы составить оглавление, краткий обзор сервера и т. п. Кроме того, именно эти элементы в первую очередь считываются роботами поисковых машин. В любом случае писать заголовки текстов для Internet нужно, ставя перед собой две задачи:

- дать пользователю точную информация о содержании документа, чтобы ему не приходилось строить догадки;
- не ввести пользователя в заблуждение привлекательным, но дезинформирующим заголовком. Сиюминутный успех обернется в дальнейшем резким снижением посещаемости вашего сайта из-за подрыва доверия.

Метаданные

Метаданными (от греч. *meta* – после, за; между, среди) называют информацию о документе, служащую дополнением к его содержанию.

Язык гипертекстовой разметки HTML позволяет авторам указывать подобные сведения множеством способов. С этой целью используют *метатэги* – данные, размещаемые за словом <МЕТА ... >. Если обычные тэги языка HTML применяются для разметки страниц, то есть для придания им определенного внешнего вида, то назначение метатэгов совсем другое. Они несут в себе справочную и управляющую информацию, которая предназначена для различных агентов (клиентских приложений, поисковых роботов). Включение метатэгов в документ, как правило, обуславливает следующие основные цели:

- управление процессом индексации документа или всего сайта в целом поисковыми роботами;
- описание содержимого документа (также для поисковых роботов);
- управление некоторыми функциями браузеров.

Многие из метатэгов довольно абстрактны, то есть не содержат никаких полезных сведений и не заслуживают внимания с точки зрения агентов. Сами метатэги и их содержание скрыты от посетителя сайта. Однако поисковые машины

их находят, а содержащиеся там данные играют не последнюю роль при определении рейтинга страницы. Однако не следует забывать, что некоторые поисковые машины не придают им значения.

Метатэги не несут никакой визуальной информации и располагаются в заголовочной части HTML-документа между тэгами <HEAD> и </HEAD>. В любом другом месте они будут проигнорированы. Особенно важны они для страниц, использующих фреймы.

Существует два вида метатэгов:

- метатэги – эквиваленты HTTP-заголовков, имеющие формат записи: <META HTTP-EQUIV="свойство" CONTENT="значение">
- информационно-указательные, записываемые следующим образом: <META NAME = "свойство" CONTENT="значение">

Элементом META задается свойство, которому присваивается значение. Метатэги первого вида эквивалентны HTTP-заголовкам и обычно управляют действиями браузеров. Они используются, чтобы оптимизировать информацию, которую содержат обычные заголовки, и особенно важны, если документы загружаются по протоколу HTTP (протоколу передачи гипертекста – <ftp://ftp.nsc.ru/pub/rfc/rfc2068.txt>). Серверы HTTP могут использовать имя свойства, указанное в атрибуте HTTP-EQUIV, для создания заголовка в ответе HTTP. Мета-тэги второго вида чрезвычай-

но важны для работы поисковых машин.

Элемент META лежит в основе механизма введения метаданных. Существуют, однако, некоторые тэги и атрибуты тэгов языка HTML, которые обрабатывают часть метаданных и могут использоваться авторами вместо элементов META. тэги TITLE, ADDRESS, INS и DEL, атрибуты тэгов TITLE и CITE.

Метатэги – эквиваленты HTTP-заголовков

Для метатэгов этого вида могут быть использованы следующие свойства:

- Expires;
- Pragma;
- Content-Type;
- Content-language;
- Refresh;
- Window-target;
- Content-Script-Type;
- Content-Style-Type.

Рассмотрим их применение.

Свойство Expires

В примере ниже показано, как выглядит формат метатэга

со свойством *Expires* (Дата устаревания):

```
<META HTTP-EQUIV="Expires" CONTENT="Mon, 29  
Nov 2004 08:21:57 GMT">
```

Клиенту будет возвращен такой заголовок HTTP:

Expires: Mon, 29 Nov 2004 08:21:57 GMT

Заголовок данного вида используется для управления кэшированием. Если указанный в метатэге срок устаревания истек, то при запросе документа он будет доставлен по сети, а не загружен из кэша.

Нулевое значение даты (CONTENT="0") приводит к тому, что браузер проверяет при каждом запросе, не изменился ли документ. Поисковые роботы могут либо совсем не индексировать его, либо постоянно отслеживать изменения в документе.

Формат даты описан в стандарте RFC850 (<ftp://ftp.nsc.ru/pub/rfc/rfc850.txt>).

Для решения ряда задач требуется, чтобы при каждом запросе документ загружался не из кэша браузера, а с сервера. Обычно это необходимо для сайтов, которые поставляют динамически изменяющееся содержание. В качестве примера можно привести результат какого-либо поиска (информация о курсе доллара в реальном режиме, электронный магазин или обычный чат). В таких случаях страница в кэше не обязательно соответствует реальной.

В подобной ситуации можно запретить браузеру кэшировать страницы при помощи свойства метатэга Expires, кото-

рое указывает дату устаревания содержимого страницы. Для того чтобы она не кэшировалась, достаточно установить прошедшую дату – например, вот так:

```
<META HTTP-EQUIV="Expires" CONTENT="Mon, 01 Jan 2000 00:00:00 GMT">
```

Если страница содержит постоянные данные и не требует регенерации при каждом запросе, то вам не следует отключать кэширование для браузера. Это позволит браузерам использовать сохраненную в кэше копию страницы в течение некоторого отрезка времени, которым вы можете управлять. Кэширование может значительно снизить нагрузку на вашем сервере.

Обычно имеет смысл отключать кэширование лишь для динамических страниц, которые содержат информацию, изменяющуюся со временем. И то далеко не всегда. Какие же из динамических страниц могут быть кандидатами на кэширование? Например, страница с прогнозом погоды, где информация обновляется каждые 5 минут. Или начальная страница сайта, которая содержит список материалов на сервере или официальные сообщения для печати, которые модифицируются два раза в день. Другие подобные страницы, где обновления происходят раз в несколько часов.

Заметьте, что с кэшированием в браузере вы получите меньшее количество посетителей, зарегистрированных на вашем сервере. Поэтому, если вы хотите точно измерять количество просмотров страниц или показов баннеров, то, ско-

рее всего, вы откажетесь от кэширования.

Свойство Pragma

Свойство *Pragma* (Контроль кэширования) применяется для отказа от кэширования и задается следующим образом:
<META HTTP-EQUIV="Pragma" CONTENT="no-cache">

Отказ от кэширования означает, что при запросе документа он будет доставлен по сети, а не загружен из кэша. Обратите внимание, что неразумное запрещение кэширования усложняет навигацию по сайту и увеличивает время ожидания. Так, при нажатии в браузере кнопки **Back** (Назад) происходит новый запрос сервера и приходится ждать загрузки страницы, которая минутой ранее была отправлена клиенту и уже есть у него в компьютере.

Свойство Content-Type

Свойство *Content-type* (Тип содержимого) служит для указания типов содержимого (типов MIME), к которым относятся «text/html», «image/png», «image/gif», «video/mpeg», «audio/basic», «text/tcl», «text/javascript» и «text/vbscript» (с учетом регистра). Текущий список зарегистрированных типов MIME можно найти по адресу <ftp://ftp.isi.edu/in-notes/iana/assignments/media-types/>. Тип содержимого «text/css», хотя он и не зарегистрирован, используется для иерархиче-

ских, или каскадных, таблиц стилей (CSS).

Тип документа "text/html" дополнительно расширяется указанием кодировки страницы charset (набор символов) – тогда браузер выводит ее сразу в заданной кодировке. Однако обратите внимание: если текст документа написан в windows-1251, а значение указано charset=KOI8-r, то изменить кодировку в браузере невозможно и текст совершенно не читается.

Формат метатэга следующий (для типа содержимого "text/html" и набора символов windows-1251):

```
<META HTTP-EQUIV="Content-type" CONTENT="text/html; charset= windows-1251">
```

Немного о кодировках

Кодировка представляет собой таблицу, где каждому символу – буквам алфавита, цифрам и специальным знакам – соответствует свой уникальный номер, код символа.

Полностью стандартизирована лишь первая часть таблицы, так называемый ASCII-код – первые 128 символов, включающих цифры и буквы латинского алфавита. Поэтому с ними никогда не бывает проблем. Вторая же часть таблицы (всего в ней 256 символов – по числу состояний, которые может принять один байт) отведена под специальные символы. Она используется для кодировки символов национальных алфавитов, причем для каждого языка и в каждой

стране она различна. При этом для букв русского языка создано пять различных кодировок, то есть одному символу соответствуют разные цифровые коды. Таким образом, при неправильной кодировке мы получим совершенно нечитаемый текст.

На заре развития вычислительной техники очень широко использовалась кодировка с кириллицей KOI-8. Она появилась в семидесятые годы при адаптации операционной системы UNIX к русскому языку – персональных компьютеров тогда еще не было. KOI-8 – до сих пор основная кодировка в UNIX. Потом началось победное шествие персональных компьютеров IBM, а с ними – операционных систем MS DOS, имевших DOS-кодировку (или кодовую страницу 866). Параллельно развивались компьютеры Macintosh, и, разумеется, была придумана еще одна кодировка – MAC. Наконец, с появлением операционной системы Microsoft Windows появилась и новая windows-кодировка (кодовая страница 1251). Она-то и стала самой распространенной в России.

Еще один вариант связан с попытками стандартизации кодировок на уровне всей планеты. Разработчики из ISO, международной организации по стандартам, создали еще одну кодировку и назвали ее ISO-8859-5. В настоящее время она практически не применяется (по-видимому, ее используют лишь в базе данных Oracle), однако поддержка данной кодировки предусмотрена во всех браузерах.

В настоящее время создана универсальная кодировка *UNICODE*, в которую вошли все языки мира, поэтому на каждый символ в ней отведено по два байта. Таким образом, максимальное число знаков в таблице расширилось до 65 535. Сейчас *UNICODE* находит свое применение, но пока не слишком широкое.

Правильное отображение символов национальных алфавитов касается как Web-серверов, так и браузеров. Они должны осуществлять информационный обмен на одном языке и в одной кодировке, и только в этом случае сообщения будут доступны.

Сервер должен иметь эффективную систему предварительного оповещения о том, в какой кодировке будет прислана страница. Клиентской программе (браузеру) необходимо принять такое сообщение и, соответственно, настроиться на нужное отображение. Если все сделать правильно, то никаких трудностей не возникнет. Однако при некорректной настройке Web-сервера он сообщает об одной кодировке (например, windows-1251), а страницу пересылает в другой (например, в КОИ-8). Пользуясь указаниями сервера, браузер, естественно, ошибается и отображает страницу неверно.

Как показано выше, можно задать кодировку документа не на сервере, а непосредственно в HTML-коде – с помощью метатэга со свойством Content-type. Но не все так просто. В России распространен способ, при котором сервер автоматически определяет, в какой кодировке приходит запрос от

клиента, и пересылает страницу Web-браузеру уже перекодированной. Вот тут-то использование указанного метатэга и может сыграть с вами злую шутку. Дело в том, что указания на странице имеют приоритет над командами, присылаемыми Web-сервером, и, правильно перекодировав страницу, сервер не может изменить содержимое метатэга. Происходит несовпадение реальной кодировки, в которой пришел документ, и указаний в тэге META. Такую страницу нельзя будет нормально просмотреть и перекодировать средствами браузера. Выбор кодировки вручную в данном случае не поможет, так как метатэг имеет приоритет и над установками браузера. Единственный способ прочесть документ – это сохранить страницу и удалить злосчастный тэг.

Учитывая все вышесказанное, данный тэг можно вообще не применять. Тогда просмотр будет вестись в той кодировке, на которую настроен браузер, если сервер не пришлет уведомление о другой кодировке документа. В этом случае клиент сможет легко ее переключить. У подавляющего большинства ваших посетителей из России по умолчанию выставлена кодировка windows-1251, поэтому у них страница сразу же будет показана правильно.

Свойство Content-language

Язык документа, указанный в метатэге со свойством *Content-language* (Язык содержимого), может использоваться

поисковыми машинами при индексировании страниц.

Метатэг имеет следующий формат:

```
<META HTTP-EQUIV="Content-language"  
CONTENT="en-GB">
```

Язык задается комбинацией двух значений *язык-диалект* (при этом значение диалект может отсутствовать), в данном примере – английский-Великобритания. Язык представляет собой двухбуквенный код, зарезервированный для сокращений типа *fr* – французский, *de* – немецкий, *ru* – русский.

Спецификация HTML 4.0 допускает явное указание автором языка в самом документе:

```
<HTML LANG="en">
```

Подобная информация может пригодиться:

- поисковым машинам;
- синтезаторам речи;
- агентам пользователей (клиентским программам) при выборе вариантов глифов для типографской печати высокого качества;
- агенту пользователя при выборе набора кавычек;
- агенту пользователя при настройке переноса, лигатур и интервалов;
- программам проверки грамматики и орфографии.

Все коды языков, которые должны использоваться в документах на языке HTML, определены и описаны в стандарте RFC1766 (<ftp://ftp.nsc.ru/pub/rfc/rfc1766.txt>).

Реализовывать рассмотренную возможность следует

крайне осторожно, поскольку могут возникнуть проблемы, из-за которых текст на странице станет нечитаемым.

Свойство Refresh

Свойством *Refresh* (Период обновления) определяется время задержки, после которой браузер автоматически обновляет документ. Метатэг часто используется для автоматической загрузки другого документа (или, как говорят, *переедресации*, перенаправления).

Формат метатэга таков:

```
<META HTTP-EQUIV="Refresh" CONTENT="3, URL=http://www.name.ru/page.html">
```

CONTENT – число, указывающее время задержки в секундах; далее следует адрес, который нужно загрузить по прошествии этого времени. Подобный способ широко используется для создания кратковременных заставок. Но, поскольку некоторые браузеры его не поддерживают, в заставке, чтобы она не «зависала», необходимо предусмотреть возможность перехода на следующую страницу. Кроме того, применение данного способа затрудняет возврат на предыдущие страницы стандартными средствами. При переедресации (то есть при значении 0) пользователь не может вернуться на предыдущую страницу с помощью кнопки **Назад** (Back), так как его в ту же секунду пересылают на страницу, с которой он пытается уйти.

Как известно, изменение адреса сайта влечет за собой потерю его посетителей. Поэтому обычно по старому адресу размещают информацию, которая поможет найти сайт. Но можно сделать и так, чтобы в дополнение к этому браузер сам перешел на использование нового адреса – благодаря метатэгу со свойством Refresh.

Свойство Window-target

Свойство *Window-target* (Окно текущей страницы) служит для определения окна текущей страницы. Его можно использовать, чтобы исключить появление новых окон во время применения фреймовых структур.

Вид метатэга может быть, например, таким:

```
<META HTTP-EQUIV="Window-target"  
CONTENT="_top">
```

Свойство Content-Script-Type

Поскольку в документах HTML не предполагается использовать определенный язык сценариев, авторы должны указывать его непосредственно в каждом фрагменте кода. Это можно сделать с помощью объявления скрипта по умолчанию для всего документа или с помощью локального объявления.

Чтобы объявить скрипт по умолчанию, в текст HTML-документа необходимо включить следующий метатэг:

```
<META HTTP-EQUIV="Content-Script-Type"  
CONTENT="type">
```

Здесь «type» – тип содержимого, указывающий язык скрипта. Примеры значений – "text/tcl", "text/javascript", "text/vbscript".

Свойство Content-Style-Type

Чтобы установить для документа язык таблицы стилей по умолчанию, надо использовать следующую строку:

```
<META HTTP-EQUIV="Content-Style-Type"  
CONTENT="text/css">
```

Документы, в которых применен тэг STYLE, но не определен язык таблиц стилей по умолчанию, являются некорректными.

Метатэги для поисковых машин

Метатэги, которые помогают поисковым машинам индексировать сайт, делают документы для них более доступными.

Представители большинства поисковых систем уже сошлись во мнении, что применение метатэгов способствует повышению релевантности отклика при обработке запросов. Но есть и прямо противоположная точка зрения, которой придерживаются, например, эксперты российской поисковой машины Rambler.

Если ваша компания широко известна и связана с производством уникальной продукции или оказанием услуг повышенного спроса, то проблем с локализацией ваших узлов в Internet у пользователя, скорее всего, не возникнет.

Другое дело, если вы пытаетесь предоставить клиентам или читателям сервис или материал, не отличающийся оригинальностью, например, связанный с разработкой Web-страниц. В этом случае даже в первую сотню ссылок из списка отклика в глобальной поисковой системе попасть не просто. Способов повышения доступности узла из поисковых машин – и оправданных, и сомнительных с точки зрения этики – немало. Важно помнить о том, что универсальных рецептов пока не существует: слишком многое зависит от специфики работы отдельного поискового сервиса. Проблема заключается еще и в том, что фактически существует два подхода к решению рассматриваемого вопроса. Один предлагается в рекомендациях по созданию документов экспертами самой поисковой системы, другой используется разработчиками, чтобы достичь высокого рейтинга.

По-видимому, самый полезный совет заключается в следующем: анализируйте HTML-код тех документов, которые добились в интересующей вас сфере деятельности и поисковой системе наивысших рейтинговых результатов. Это относится и к метатэгам, и к остальному содержимому страниц. Естественно, такой анализ требует специальных навыков и отличается трудоемкостью, что может послужить веским до-

водом для обращения к профессионалу.

Оптимальный результат обеспечила бы компания-посредник, имеющая прямой контакт с разработчиками поисковых систем. Если нечто подобное и существует, информация об этом по понятным причинам вряд ли когда-либо будет передана огласке. Есть, однако, косвенные, хотя и медленно действующие приемы, которые позволяют провести анализ работы поисковых систем и доступны каждому: речь идет о тестировании. Известны примеры создания целых тестовых Web-узлов, единственная задача которых – определить, насколько чувствительна работа алгоритма отдельной поисковой машины к тому, как размещена информация на Web-страницах. Широкое распространение в Internet получила и разработка так называемых страниц-мостиков, которые оптимально нацелены на конкретную поисковую систему. Добраться с них до основной страницы узла адресатам информации позволяют гиперссылки.

В целом ситуация, когда материалы сайта в первую очередь ориентированы на поискового робота, а не на конечного пользователя, не может не вызывать опасений. Так, например, одно время AltaVista особенно высоко оценивала содержимое заголовка страницы, помещаемое в контейнер TITLE. В результате список отклика этого индекса возглавили сотни документов, имеющих вместо связного заголовка набор ключевых слов и фраз. Именно этот набор становился по умолчанию именем закладки на документ при работе с большин-

ством браузеров.

Как оказалось, даже такая «прецизионная» настройка на систему не дает долговременных гарантий: после того как AltaVista изменила характер ранжирования документов, рейтинг многих фаворитов катастрофически снизился.

Ключевые слова и описания

Чтобы получить высокий рейтинг в поисковых системах, требуется выбрать для передачи им основные ключевые слова и фразы сайта. В общей сложности их должно быть не меньше десяти (желательно – несколько десятков), и они должны соответствовать содержанию сайта. Кроме того, требуется составить краткое описание сайта – абзац длиной не более 1 000 знаков (для некоторых поисковых машин – не более 700).

Ваш сайт для поисковых систем – один из многих, поэтому текст описания, который обычно появляется на страницах с результатами поиска, должен быть интересным, привлекающим внимание, чтобы у читателя возникло сразу желание посетить сайт. Ключевые слова в нем необходимо использовать как можно чаще – от этого зависит высота рейтинга.

Вообще, к подбору ключевых слов следует подойти очень серьезно. Правильный их выбор позволяет сильно повысить позицию сайта при запросах.

Какие же слова следует выбирать? Разумеется, те, которые наиболее точно отражают специфику именно вашего проекта. При этом следует избегать случайных и общих фраз. Представьте себя на месте человека, который пытается найти ваш сайт. С каких слов вы начнете поиск? Что наберете в поле запроса поисковой системы?

Каждый вариант следует проверить в какой-либо поисковой системе (лучше – в нескольких). Оцените, насколько список найденных сайтов связан с тематикой вашего. Если такая связь не вызывает сомнений, то выбранные вами слова подходят на роль ключевых; в противном случае стоит поискать другие. Для наглядности не мешает посмотреть, какие ключевые слова используются на первых страницах результата поиска.

Многие индексирующие машины проводят поиск элементов тэга META, в которых задан разделенный запятыми список ключевых слов и фраз или дается краткое описание. Поисковые машины могут представлять их как итог поиска или использовать для улучшения его результатов.

Формат указанных метатэгов иллюстрируют следующие примеры:

```
<META NAME="keywords" CONTENT="отпуск, Греция, солнце">
```

```
<META NAME="description" CONTENT="Идиллический отпуск в Европе">
```

В поле CONTENT не должно быть знаков конца строки,

кавычек и других специальных символов; регистр символов значения не имеет.

Не рекомендуется повторять несколько раз одни и те же ключевые слова, так как это может быть воспринято как спам, в результате чего страница может быть удалена из индекса поисковой машины. Обычно *спамом* называют несанкционированную рассылку электронной почты. К ней зачастую прибегают недобросовестные рекламодатели, распространители вирусов и т. д. Однако здесь и далее по тексту под спамом будет подразумеваться намеренное злоупотребление ключевыми словами в заголовках, комментариях и тексте самого документа.

Не стоит использовать одинаковые списки ключевых слов для разных страниц сайта. Конечно, так проще, но содержимое документов различное, да и поиск ведется по-разному. Если вам хочется автоматизировать эту работу (действительно, довольно трудоемкую, ведь общий объем ключевых слов одного документа может достигать 50 % от его объема), напишите программу, которая выбирала бы текст из определенных блоков документа, например, из контейнеров, заключенных в тэгах H, I и B. Задача не кажется сложной, да и можно найти подобную программу в архивах программного обеспечения в Internet.

Если строка в поле CONTENT получается слишком длинной, не возбраняется разбить ее на несколько конструкций этого вида.

Указание ключевых слов – это основной способ использования тэга META. Информация о документе, составленном с использованием различных языков, может быть представлена в нескольких таких тэгах. Тогда поисковые машины могут фильтровать атрибут LANG и отображать результаты поиска с применением выбранного пользователем языка – например, так:

<- Для говорящих на английском в американской версии->

```
<META      NAME="keywords"      LANG="en-us"  
CONTENT="vacation, Greece, sunshine">
```

<- Для говорящих на британском английском ->

```
<META      NAME="keywords"      LANG="en"  
CONTENT="holiday, Greece, sunshine">
```

<- для русскоязычных пользователей ->

```
<META NAME="keywords" LANG="ru" CONTENT="от-  
пуск, Греция, солнце">
```

Эффективность обращения к поисковым машинам можно повысить также за счет использования тэга LINK для создания ссылок на разноязычные переводы страницы, на версии документа для другого носителя (например, PDF) и, если документ является частью набора, ссылок на соответствующую начальную точку для просмотра набора.

Метатэг NAME="keywords" позволяет автору документа

самостоятельно создать адекватный содержанию набор ключевых слов и фраз. Допустимая для восприятия роботом длина перечня варьируется от 874 до 1 000 символов. При отсутствии указанного метатэга робот формирует набор автоматически на основе своего алгоритма. Если индексируется все содержимое документа, то он будет участвовать в отклике и по тем терминам, которые входят в текст страницы, но отсутствуют в метатэге. Автоматический индекс при создании поискового образа документа может комбинировать содержимое метатэгов и текста, должным образом взвешивая термины из разных полей Web-страницы. При этом далеко не все системы, которые поддерживают метатэги, отдают явное предпочтение входящим в них терминам.

Отметим также, что метатэг ключевых слов стоит размещать в одну строку, поскольку некоторые роботы не умеют переходить к новой строке.

Ключевые слова лучше не повторять более двух раз, иначе поисковая система может расценить это как спам. Нежелательно, например, задавать последовательности вида «free, free, free, free, free», но вполне допустимы такие, как «free, free web, free stuff, internet for free».

Большого успеха можно добиться, если использовать не ключевые слова, а ключевые фразы. Если вы подберете их удачно, то, возможно, при запросе ваша ссылка окажется одной из первых. Есть еще один, не вполне серьезный прием, основанный на известной ошибке. Многие забывают пере-

ключать языковую раскладку клавиатуры, и тогда, допустим, вместо запроса «free» возникает запрос: «акуу», то есть английское слово, введенное буквами кириллицы, расположенными на соответствующих клавишах. Как ни странно, результат бывает положительным: поисковая система выдает ссылку на ваш сайт!

Для хорошего индексирования документа поисковыми системами рекомендуется в первом абзаце текста на каждой странице использовать как можно больше отобранных для нее ключевых слов.

Как известно, основную смысловую нагрузку несут в языке имена существительные. Поэтому именно они составляют большую часть ключевых поисковых слов. Значительно реже используются прилагательные, совсем редко – глаголы, наречия, предлоги, союзы. Имена прилагательные просто незаменимы, если требуется индексировать «голландский сыр» с «баварским пивом» в «ночном клубе».

Очень эффективно использование редких слов. К такому можно отнести специальные термины, названия местностей, организаций, имена людей и т. п. Применение подобной лексики позволяет максимально конкретизировать тематику.

Существует целый ряд слов, которые поисковая система будет игнорировать при запросе; их поиск невозможен. Это так называемые стоп-слова, например: *на, что, это, для* и др. Они настолько часто встречаются в текстах, что искать

по ним что-либо крайне затруднительно. Интересно, что в списки стоп-слов для некоторых поисковых машин уже входят такие, как *Internet, компьютер, Сеть*. Они стали настолько распространенными, что утратили свое значение (с точки зрения поиска, конечно). Очевидно, что использовать стоп-слова в качестве ключевых нет смысла.

Метатэг NAME="description" позволяет включить в поле CONTENT краткое описание документа. В зависимости от работа воспринимаемая длина текста составляет от 150 до 250 символов. После индексирования поисковой машиной описание должно появиться рядом со ссылкой на документ при попадании его в список отклика.

Если ресурс ориентирован не только на русскоязычную публику, целесообразно сделать отдельные описания на разных языках. Следует создавать лаконичные и привлекающие внимание тексты; суть должна быть ясна с первого взгляда, поскольку на страницах с результатами поиска в поисковой системе информацию о каждом найденном сайте приводят в одном небольшом абзаце.

При составлении описания не нужно указывать в нем название сайта – оно и так будет проиндексировано. С точки зрения индексирования сайта не имеют смысла общие фразы типа: «Здесь есть все», «Вам будет интересно», «Заходи, не пожалеешь!» и т. п. Обратите внимание посетителя именно общее содержание, а он сам разберется, нужна ли ему ваша страница или нет. Стоит ли заманивать читателей, если они

сразу же покинут сайт, убедившись, что он не соответствует описанию?

Если в документе нет описания, то в качестве такового поисковые машины приводят несколько начальных строк документа. Поэтому случается, что если в самом начале документа есть фрагмент, написанный на языке JavaScript, то вместо нормального описания выдается «абракадабра» в виде фрагмента скрипта.

Словом, необходимо, чтобы описание имело вид аннотации, состоящей из нескольких ясных предложений, и тогда пользователь поисковой машины сможет легко понять смысл документа.

При подборе ключевых слов для метатэга документа необходимо учитывать различные нюансы. Многие алгоритмы придают больший вес тому термину или фразе, которые расположены ближе к началу перечня. Число повторений ключевых слов не должно превышать определенного количества (как правило, двух-трех), во избежание применения санкций против спама: иначе страница не будет проиндексирована и включена в базу данных поисковой системы. Неплохую помощь могут оказать системы, которые отслеживают запросы, поступающие от пользователей на поисковые машины. Важно найти не просто адекватные содержанию ключевые слова и фразы, а именно те, которые часто применяются пользователями на практике.

Итак, поисковые роботы берут ключевые слова и описа-

ние сайта непосредственно из кода страницы и содержимого метатэгов NAME = «keywords» и NAME="description". Однако не все системы придерживаются такой методики. Многие проверяют соответствие ключевых слов основному тексту страницы, а полнотекстовые поисковые машины анализируют все ее содержимое, хотя ключевые слова в ней при поиске являются приоритетными.

Для более широкого охвата потенциальной аудитории ключевые слова и описание сайта, которые вы стремитесь вставить в большинство документов, рекомендуется корректировать в соответствии со смыслом конкретной страницы. Это повысит вероятность попадания на ваш сайт. Общее же его описание пригодится для страниц с «нейтральным» содержанием.

Большинство поисковых машин придает большое значение титулу страницы и метатэгам с ее описанием и ключевыми словами. Однако стоит подумать не только о поисковых системах, но и о тех, кто увидит эти ссылки, – ведь в конечном итоге ссылки будут выведены на экран, а решать, переходить по ним или нет, будет пользователь.

Допустим, вы сделали запрос *program*. Появились результаты, в числе которых, например, такие:

Free program, program, free, program free, free program

Free program, program, free, program free, free program,
free

program, program, free, program free, free program, free program,

program, free, program free, free program, free program, program,

free, program free, free program

Воспользоваться такими ссылками уже не тянет – возникает предубеждение к сайту и его хитрому создателю. Современные поисковые роботы определяют спам и не индексируют страницу, но машина – это всего лишь машина.

Можно просто вписывать в титул метатэги, ключевые выражения, результат будет тот же: бессмысленные ссылки! Пожалуй, не зря считается, что оформление страниц для поисковых систем – это наука и искусство. Ваша задача – не только правильно подобрать ключевые фразы и слова, но и составить из них логичные предложения для титула и описания, которые и выводятся в результатах поиска. При этом еще надо уложиться в отведенное количество символов.

Подбор ключевых выражений начинает походить на некое таинство, особенно это касается иностранных поисковых систем: некоторые расхожие слова уже «проданы», и вы, как ни бейтесь, на первое место по ним не попадете.

Появилась даже платная услуга по подбору ключевых слов и оформлению документов под них, осуществляемая профессионалами. Если вы полагаете, что вам это не нужно, – готовьтесь постоянно следить за своим рейтингом в поиско-

вых машинах и определять положение ваших ссылок по заданным ключевым словам, ведь поисковики действительно могут обеспечить вам существенный приток посетителей.

Язык документа

В глобальном контексте Internet важно знать, на каком языке создается страница. Определите язык документа, как описано выше, в разделе «Свойство Content-language».

Языковые варианты документа

Если вы подготовили переводы документа на другие языки, используйте тэг LINK для ссылки на эти тексты. Такой подход повысит эффективность поисковых машин и позволит им предлагать результаты поиска на языке, предпочитаемом пользователем, независимо от построения запроса. Например, следующие ссылки предлагают поисковой машине французскую и немецкую версии:

```
<LINK rel="alternate"  
type="text/html"  
href="mydoc-fr.html" hreflang="fr"  
lang="fr" title="La vie souterraine">  
<LINK rel="alternate"  
type="text/html"  
href="mydoc-de.html" hreflang="de"
```

lang="de" title="Das Leben im Untergrund">

Указанным тэгом определяется связь. Хотя он может быть представлен неограниченное число раз, его присутствие допускается только в разделе HEAD документа. Несмотря на то, что тэг LINK не имеет содержимого, он является носителем информации об отношениях, которая может представляться агентами пользователей (браузерами) различными способами, например, в виде панели с выпадающим списком ссылок.

Таким же способом с помощью тэга LINK задаются ссылки на версии документа для другого носителя (например, PDF) и, если документ является частью набора, ссылки на соответствующую начальную точку для просмотра набора.

Начало набора

Наборы документов, или представлений систем обработки текстов, часто переводятся в наборы документов HTML. Для поисковых машин полезно указать ссылку на начало набора в дополнение к тому, что страница попала в результаты поиска. Вы можете помочь поисковым системам, применив тэг LINK, имеющий атрибуты REL="begin" и TITLE=, как показано в следующем примере:

```
<LINK REL="begin"
```

TYPE="text/html"

HREF="page1.html"

TITLE="Общая теория относительности">

Инструкции для роботов – файл robots.txt

Алгоритмы работы многих поисковых машин предусматривают возможности ограничения действий роботов. Это достигается с помощью файла robots.txt и тэга META в документах HTML.

Файл *robots.txt* объясняет роботу поисковой машины, что надо индексировать, а что не стоит. Зачем, например, индексировать служебные файлы, скажем, статистические отчеты?

Файл содержит набор команд, которые позволяют закрыть от индексирования отдельные каталоги узла. Обычно закрываются каталоги, содержащие скрипты, служебную информацию и т. п. Отчасти это повышает контрастность значимых документов узла в поисковой системе. К тому же поисковые машины нередко вводят ограничение на число ресурсов, регистрируемых для одного сервера. Некоторые – так, например, действовал робот поисковой системы Lycos, – вообще не проводят индексирования, если указанный файл отсутствует.

Итак, если вы поддерживаете работу сервера с доменным именем www.mysite.ru, то содержимое файла robots.txt должно быть доступно по адресу: <http://www.mysite.ru/>

robots.txt, то есть данный файл следует загрузить в корневой каталог вашего сайта. Когда робот просматривает Web-сайт, сначала он проверяет наличие документа <http://www.mysite.ru/robots.txt>. Если таковой обнаружен, анализируется его содержимое и выясняется, позволено ли загрузить документ. Вы можете настроить файл robots.txt для конкретных роботов и запретить доступ к определенным каталогам или файлам.

На одном сайте может быть только один файл robots.txt. Не следует помещать файлы с данным именем в каталоги, поскольку робот их не найдет. Обратите внимание, что при адресации в Internet учитывается регистр, поэтому название robots.txt должно быть задано в нижнем регистре.

Файл robots.txt – это обычный текстовый документ, содержащий одну или несколько записей, разделенных пустыми строками, как показано в следующем примере:

```
# robots.txt for http://www.mysite.ru
User-agent: *
Disallow: /cgi-bin/maillist/
Disallow: /tmp/
Disallow: /product1.html
User-agent: aport
User-agent: scooter
Disallow:
```

Вот пример файла robots.txt, полностью запрещающего доступ к сайту всем роботам:

User-agent: * # Применяется ко всем роботам.

Disallow: / # Запретить индексацию всех страниц.

Каждая запись должна начинаться с указания переменной User-agent отдельной строкой. Таких строк может быть и несколько. Значением переменной User-agent задается оригинальное имя программы-робота соответствующей поисковой системы, для которой предназначена информация.

Согласно спецификации на файл robots.txt, текст в нем чувствителен к регистру, так что следует записать именно User-agent, а не User-Agent. Трудно сказать, вызывает ли неверное написание проблемы у роботов поисковых систем, но для надежности лучше соблюсти требования спецификации.

Имя робота может быть похоже на название соответствующей поисковой системы, а может и сильно отличаться от него. Далеко не всегда оно соответствует названию механизма поиска. Поэтому нельзя просто написать AltaVista в Useragent и ожидать, что AltaVista исключит указанные вами страницы. Имя робота указывается без учета регистра и не включает информацию о версии. Можно перечислить несколько имен через пробел. Символ * является маской и означает: «для всех роботов». В этом случае запись описыва-

ет режим доступа по умолчанию для любого робота, неупомянутого явно в других записях. В файле robots.txt не может быть нескольких таких записей.

В табл. 1 приведены имена роботов некоторых поисковых систем.

Таблица 1. Имена роботов поисковых систем

Поисковая система или сервис проверки ссылок	Имя робота
AltaVista (http://www.altavista.com/)	Scouter
AOL (http://search.aol.com/)	Slurp
EuroSeek (http://www.euroseek.com/)	Arachnoidea
Excite (http://www.excite.com/)	ArchitextSpider
EZResult (http://www.ezresult.com/)	Mozilla
Fast (http://www.fastsearch.com/)	Fast
Google (http://www.google.com/)	Googlebot
Hotbot (http://www.hotbot.com/)	Slurp
Infoseek (http://www.infoseek.com/) – GO.com (http://go.com/)	Infoseek Sidewinder; Ultraseek
iWon (http://www.iwon.com/)	Slurp
LookSmart (http://www.looksmart.com/)	Slurp
Lycos (http://www.lycos.com/)	Lycos
MSN (http://www.msn.com/)	Slurp
Netscape Search (http://search.netscape.com/)	Googlebot
Northern Light (http://www.northernlight.com/)	Gulliver
Teoma (http://www.directhit.com/)	Grabber
W3C Link Checker (http://validator.w3.org/checklink/)	W3C-checklink
WebCrawler (http://www.webcrawler.com/)	ArchitextSpider
Yahoo! (http://www.yahoo.com/)	Googlebot
Rambler (http://www.rambler.ru/)	StackRambler
Апорт (http://www.aport.ru/)	Aport
Яндекс (http://www.yandex.ru/)	Yandex

Начиная со следующей строки после указания переменной User-agent, определяются собственно правила поведения робота в виде задания значений переменным Allow (Разрешить) и Disallow (Запретить).

В поле Allow задается перечень разрешенных для посещения роботами поисковых машин каталогов и файлов, а в поле Disallow – перечень закрываемых каталогов и файлов. Это может быть полный или частичный путь. Любой URL-адрес, начинающийся с этого значения, нельзя будет загрузить. Например, запись Disallow: /help запрещает доступ к /help.html и /help/index.html, в то время как запись Disallow: /help/ запретит доступ к /help/index.html, но разрешит доступ к /help.html.

Если файл находится в корневом каталоге, обязательно включите перед его именем символ «слеш» – косую черту вправо.

Пустое значение параметра Disallow означает, что все URL сайта могут быть проиндексированы. Частая ошибка при написании robots.txt состоит в том, автор ставит *, чтобы указать, что хочет исключить все файлы в каталоге. Однако правильным будет не включать никаких звездочек в поле Disallow.

В файле robots.txt должно быть, по крайней мере, одно поле Disallow. Символ # предваряет комментарии, необработываемые роботами.

В примере, рассмотренном выше, первая запись запрещает индексирование двух каталогов и файла product1.html. Роботу Scooter поисковой системы AltaVista и роботу Aport поисковой системы Апорт для доступа открываются все каталоги (поле Disallow пусто). При необходимости закрыть все

каталоги следовало бы написать: `Disallow: /`. Пустая строка между записями необходима.

Очень важно не допускать ошибок в файле `robots.txt`. В противном случае конструкции могут не работать, а вы даже не будете знать об этом. Возможно, некоторые поисковые системы без затруднений разрешат проблемы, возникающие из-за различий в синтаксисе, однако абсолютной гарантии нет. Поэтому, если вы обнаружите ошибку в синтаксисе, исправьте ее и повторно зарегистрируйте ресурс в поисковых системах. Роботы поисковых машин снова проиндексируют ваш сайт и выполнят все указания, имеющиеся в файле `robots.txt`.

Обратите внимание, что недопустимы строки вида:

```
Disallow: /tmp/*
```

или

```
Disallow: *.gif
```

Подробное описание стандарта исключений и синтаксиса команд файла `robots.txt` вместе с другой полезной информацией о роботах (на английском языке) можно найти по адресу <http://www.robotstxt.org/wc/robots.html>.

Тэг META для роботов

Файл `robots.txt` поддерживается практически всеми роботами, однако корневой каталог сервера может быть недоступен вам. В этом случае для аналогичных целей, но лишь в

пределах одного документа можно использовать специальные метатэги. Они не только решают проблему запрета, но и предоставляют позитивные возможности для управления индексированием. Это более гибкое средство управления индексацией, чем robots.txt. В частности, в тэге можно дать роботу поисковой машины предписание не уходить по ссылкам на чужие серверы, например, в документах со списками ссылок:

```
<META NAME="robots" CONTENT="index, follow">
```

Из данного примера видно, что все управление в метатэге сводится к указанию двух переменных, а именно NAME и CONTENT. Для переменной CONTENT в контексте NAME="robots" допустимо использовать следующие значения (они могут быть записаны как строчными, так и прописными буквами):

- index – разрешено индексировать документ;
- follow – разрешено следовать по ссылкам;
- all – эквивалентно употреблению index и follow одновременно (записывается через запятую), то есть разрешено индексировать данную страницу и все ссылки, исходящие из нее. Когда обнаружено слово all, все другие слова, если они указаны, игнорируются;
- noindex – не индексировать сам документ, но следовать по его ссылкам;
- nofollow – индексировать, но не идти по ссылкам;
- none – эквивалентно одновременному употреблению

`noindex` и `nofollow` через запятую, то есть всем роботам предложено игнорировать данную страницу при индексации.

Если указанный метатэг пропущен или не задано значение переменной `CONTENT`, то по умолчанию поисковый робот действует, как при значении `CONTENT="index, follow"` (или `CONTENT="all"`, что равноценно).

Если в переменной `CONTENT` содержатся противоположные по смыслу ключевые слова (например, `follow` и `nofollow`), то робот поступает по своему усмотрению; в данном случае выбрано `follow`).

В приведенном выше примере свойство `NAME="robots"` дает роботам предписание индексировать и саму страницу (`CONTENT="index"`), и документы, ссылки на которые она содержит (`CONTENT="follow"`). Вместо двух этих значений, приведенных через запятую, можно было бы написать одно – `CONTENT="all"` – с тем же результатом.

Указанный метатэг позволяет авторам HTML-документов сообщать роботам о том, может ли документ быть проиндексирован или его следует использовать, чтобы получить дополнительные ссылки. Для этого не требуется вмешательства администратора сервера.

В следующем примере роботу не предписано ни индексировать сайт, ни анализировать ссылки:

```
<META NAME="robots" content="noindex, nofollow">
```

Другие метатэги

Другие метатэги также могут использоваться роботами при сканировании. Например, тэг с параметром `NAME="author"` позволяет ввести имя автора документа и учитывается некоторыми поисковыми системами, что иногда может существенно помочь при поиске.

Похожий метатэг с параметром `NAME="generator"` часто автоматически проставляется различными HTML-редакторами и содержит информацию о программе, в которой был сделан документ. Некоторые авторы в его описании дублируют информацию метатэга "author".

Метатэг с параметром `NAME="copyright"` служит для указания принадлежности авторских прав.

Справедливости ради следует отметить, что поисковые системы все же редко используют указанные метатэги в своей работе.

Проверка ссылок

До начала рекламной кампании стоит убедиться в том, что в текстах страниц, размещенных на сайте, нет технических ошибок. К таковым относятся неправильно расставленные ссылки, отсутствие каких-либо файлов и собственно погрешности в HTML-коде, а также грамматические и орфографические ошибки в текстах документов. Напомним, кстати, что сайтом называют набор документов, которые в установленном порядке размещены на определенной машине и доступ к которым через сеть обеспечивается HTTP-сервером.

Реакция пользователя на ошибки такого рода может иметь самые печальные последствия для автора. Замечая их, посетитель испытывает досаду, раздражение и быстро уходит с сайта. Уговорить «обиженных» вернуться будет крайне трудно.

Формальную проверку сайта можно провести как сетевыми средствами, так и при помощи различных программ. Кроме того, необходимо обратить особое внимание на смысловые элементы, играющие немаловажную роль при «раскрутке» Web-ресурса:

- заголовки;
- использование фреймов;
- метатэги;

- навигацию по Internet-страничкам.

В качестве дополнительной проверки понадобится исследование времени загрузки ваших страниц. Кроме того, очень важно иметь доступ к файлам посещений, что позволит вести постоянный анализ жизнедеятельности сайта в самых разнообразных аспектах.

В каком-то смысле созданный вами гипертекст похож одновременно и на книгу, которую вы должны откорректировать, и на программу, которую необходимо отладить. По крайней мере, следует дать прочитать его кому-нибудь из той группы людей, для которой он написан, и проанализировать замечания. Но одного этого недостаточно, требуется протестировать документ. Просмотрите его в нескольких различных обозревателях (браузерах), чтобы убедиться, что его вид не зависит от средства вывода.

Проверка занимает время. Сколько времени отводить на тест? Все зависит от того, какое качество документа вас устраивает.

Есть сведения, что в Internet около 5 % имеющихся ссылок недоступны. Необходимо предотвратить подобные проблемы.

Конец ознакомительного фрагмента.

Текст предоставлен ООО «ЛитРес».

Прочитайте эту книгу целиком, [купив полную легальную версию](#) на ЛитРес.

Безопасно оплатить книгу можно банковской картой Visa, MasterCard, Maestro, со счета мобильного телефона, с платежного терминала, в салоне МТС или Связной, через PayPal, WebMoney, Яндекс.Деньги, QIWI Кошелек, бонусными картами или другим удобным Вам способом.