Алексей Яковлев
Василий Ткачев

РАСКРУТКА САЙТОВ

Основы, секреты, трюки

3-е издание

Санкт-Петербург
«БХВ-Петербург»
2011

УДК 681.3.06
ББК 32.973.26-018.2
Я47

Яковлев, А. А.

Я47 Раскрутка сайтов: основы, секреты, трюки. — 3-е изд., перераб. и доп. / Яковлев А. А., Ткачев В. Г. — СПб.: БХВ-Петербург, 2011. — 320 с.: ил.

ISBN 978-5-9775-0675-5

Рассказано, как добиться успеха в привлечении на веб-сайт целевых посетителей. Книга содержит необходимый и достаточный материал для самостоятельного продвижения интернет-сайтов любой сложности. Даны основы поисковой оптимизации (SEO), изложены современные методы SEO, особенности продвижения в Yandex, Google, Rambler и др., описаны основные элементы, влияющие на совместимость сайта с поисковыми машинами. Даны рекомендации по поиску и анализу ключевых слов, персонализации поиска, написанию контента для сайта, технологии анализа сайтов-конкурентов. Рассмотрены проблемы поискового спама и "черных" методов оптимизации и особенности продаж товаров через Интернет. Книга основана на материалах курсов поисковой оптимизации проектов SEO-Study.ru и AllinTop.ru, практических примерах из жизни, а также ответах на часто задаваемые вопросы.

В третьем издании книги добавлены новые главы, касающиеся автоматизации процесса продвижения сайтов, анализа эффективности инвестиций, направляемых на продвижение; другие главы обновлены.

Для широкого круга читателей

УДК 681.3.06
ББК 32.973.26-018.2

Группа подготовки издания:

Главный редактор	*Екатерина Кондукова*
Зам. главного редактора	*Евгений Рыбаков*
Зав. редакцией	*Григорий Добин*
Редактор	*Анна Кузьмина*
Компьютерная верстка	*Натальи Караваевой*
Корректор	*Виктория Пиотровская*
Оформление обложки	*Елены Беляевой*
Зав. производством	*Николай Тверских*

Лицензия ИД № 02429 от 24.07.00. Подписано в печать 01.12.10.
Формат 70×100 1/16. Печать офсетная. Усл. печ. л. 25,8.
Тираж 3000 экз. Заказ №
"БХВ-Петербург", 190005, Санкт-Петербург, Измайловский пр., 29.

Санитарно-эпидемиологическое заключение на продукцию
№ 77.99.60.953.Д.005770.05.09 от 26.05.2009 г. выдано Федеральной службой
по надзору в сфере защиты прав потребителей и благополучия человека.

Отпечатано с готовых диапозитивов
в ГУП "Типография "Наука"
199034, Санкт-Петербург, 9 линия, 12.

ISBN 978-5-9775-0675-5

© Яковлев А. А., Ткачев В. Г.., 2010
© Оформление, издательство "БХВ-Петербург", 2010

Оглавление

Глава 1. Общие сведения о поисковой оптимизации ... 1

1.1. Что такое поисковые машины и как они работают? ... 1
1.2. Особенности поиска с точки зрения поисковой машины ... 5
1.3. "Анатомия" поисковой машины ... 7
1.4. Что такое SEO (Search Engine Optimization)? ... 11
1.5. Поисковая оптимизация: прошлое, настоящее и будущее ... 13
1.5.1. Прошлое поисковой оптимизации ... 13
1.5.2. SEO и SEM (Search Engine Marketing) сегодня ... 16
1.5.3. Будущее поискового маркетинга ... 20

Глава 2. Основы поисковой оптимизации ... 21

2.1. Создание веб-страниц с учетом пожеланий поисковых систем ... 21
2.2. Написание грамотного HTML-кода ... 22
2.3. Верстка с помощью тегов <*DIV*> против табличной ... 23
2.4. Основные элементы, влияющие на совместимость с поисковыми машинами ... 24
2.5. Релевантность и факторы, влияющие на нее ... 25

Глава 3. Особенности некоторых поисковых систем ... 27

3.1. Особенности Google ... 27
3.1.1. Google PageRank ... 29
3.1.2. Управление индексацией Google ... 31
3.1.3. Сервис Google Webmaster Tools ... 35
3.1.4. Как оптимизировать сайт для Google? ... 36
3.1.5. Проблема дублированного контента ... 38
3.1.6. Персонализированный поиск Google ... 39
3.1.7. Привлечение ссылок с блогов, не содержащих тег *nofollow* ... 40
3.2. Особенности Яндекса ... 41
3.2.1. Индекс цитирования ... 45
3.2.2. Быстрый робот ... 46
3.3. Особенности Rambler ... 49
3.3.1. Оптимизация страницы под Rambler ... 50
3.4. Как оптимизировать сайт для иностранных и иноязычных пользователей, которые находятся вне основного рынка ... 53

Глава 4. Поиск и анализ ключевых слов ... 55

4.1. Почему важно правильно подбирать ключевые слова? ... 55
4.2. Основные интернет-сервисы для поиска и анализа ключевых слов ... 59
4.3. Пример подбора и анализа ключевых слов ... 67
4.4. Работа с трафиком ... 71

Глава 5. Оптимизация основного текста ... 75

5.1. Наполнение главной страницы ... 75
5.1.1. Нужен контент? Нет проблем! ... 75
5.2. Обогащение текста ключевыми словами ... 76
5.2.1. Наполнение страниц текстом ... 77
5.2.2. Оптимизация основного текста ... 78
5.3. Внедрение ключевых слов ... 80
5.4. Оптимизация PDF-документов ... 81

Глава 6. Описание основных элементов, влияющих на совместимость с ПМ ... 89

6.1. Заголовок веб-страницы ... 89
6.1.1. Общие сведения ... 89
6.1.2. Некоторые хитрости использования тегов *<TITLE>* ... 91
6.1.3. Практические советы при создании заголовков ... 92
6.2. Метатег описания ... 95
6.2.1. Назначение и особенности метаописания ... 95
6.2.2. Практические советы по написанию метаописания ... 97
6.2.3. Связанные теги *<TITLE>* и *<META>* ... 98
6.3. Метаключи ... 99
6.3.1. Общие сведения о метаключах ... 99
6.3.2. Практические советы по подбору метаключей ... 99
6.4. Основной текст ... 101
6.4.1. Плотность ключевых слов ... 101
6.4.2. Расположение текста на странице ... 102
6.4.3. Расположение ключевых слов (близость ключевых слов друг к другу — Adjacency & Proximity) ... 104
6.4.4. Заголовки *H1—H6* ... 104
6.4.5. Выделение текста полужирным ... 105
6.5. Доменное имя ... 105
6.6. Дополнительные возможности ... 107
6.6.1. Темы ... 107
6.6.2. Выделенный IP-адрес ... 107
6.6.3. Некоторые элементы веб-дизайна ... 108
6.6.4. Атрибут *alt* тега *<IMG>* ... 108
6.6.5. Текстовые ссылки ... 110
6.6.6. Размер файла ... 111
6.6.7. "Возраст" страницы ... 111
6.6.8. Скрытые теги форм ... 112
6.6.9. Внешний код ... 112
6.6.10. Каскадные таблицы стилей (CSS) ... 113
6.7. Динамически формируемый контент ... 113
6.7.1. Что такое динамически формируемое содержимое? ... 113
6.7.2. Влияние динамического содержания на совместимость с ПМ ... 114
6.7.3. Будьте осторожны с ID-сессиями ... 115
6.7.4. Оптимизация динамических сайтов ... 115
6.7.5. Специальные приемы для сайтов с динамическим контентом ... 116
6.8. Структура каталогов ... 116
6.9. Элементы интерфейса ... 117
6.9.1. Графическая навигация (меню) ... 117

6.9.2. Раскрывающиеся меню ... 118
6.9.3. Множественные точки входа ... 118
6.9.4. Карта сайта (sitemap) ... 119
6.10. Flash- и splash-страницы ... 121
6.10.1. Общие замечания ... 121
6.10.2. Вводные или splash-страницы ... 121
6.10.3. Как сделать Flash дружественным ПМ ... 122
6.10.4. Поддержка Flash поисковыми машинами ... 123
6.11. Спецэффекты ... 123
6.12. Тире, подчеркивание и другие символы в доменах и названиях файлов ... 123
6.13. Файл robots.txt и метатег *ROBOTS* ... 124

Глава 7. Типы и функции ссылок ... 127

7.1. Динамические ссылки (JavaScript-ссылки) ... 127
7.2. Графические ссылки ... 127
7.3. Текстовые ссылки ... 127
7.4. Входящие и исходящие ссылки ... 129
7.5. Взаимные ссылки ... 129
7.6. Треугольные ссылки ... 130
7.7. Ссылочная популярность ... 130
7.8. Узнайте разницу между естественной и искусственной структурами ссылок ... 131

Глава 8. Построение ссылок. Решающий фактор высокого ранжирования ... 135

8.1. Внешние факторы раскрутки сайта ... 135
8.2. Оценка качества ссылки ... 139
8.3. Регистрация сайта в каталогах ... 140
8.3.1. Общие сведения ... 140
8.3.2. Регистрация в Open Directory ... 143
8.3.3. Регистрация в Яндекс.Каталоге ... 143
8.4. Создание и ведение блогов ... 145
8.5. Написание и публикация статей ... 149
8.5.1. Общие рекомендации ... 149
8.5.2. Методы распространения статей ... 151
8.5.3. Каталоги статей ... 151
8.5.4. Соблазны грязных технологий ... 156
8.6. Википедия (Wikipedia) ... 157
8.7. Линкбэйтинг и социальные сети ... 157
8.7.1. Социальная сеть Digg ... 159
8.8. Тематические форумы ... 164
8.9. Распространение пресс-релизов ... 165
8.10. Поиск и покупка заброшенных сайтов ... 165
8.11. Ссылочные агрегаторы ... 166
8.11.1. SEOpult ... 166
8.11.2. WebEffector ... 167
8.11.3. Rookee (ex-REACTOR) ... 168
8.12. Как покупать ссылки и не попасться? ... 169
8.12.1. Покупка рекламы в электронных журналах и бюллетенях ... 170
8.13. Создание и дистрибуция программного обеспечения ... 170

8.14. Вирусный маркетинг ... 171
8.14.1. Принципы вирусного маркетинга ... 171
8.14.2. Современные приемы вирусного маркетинга ... 175
8.14.3. Выводы ... 176
8.15. Конкурсы, тесты, спецопросы ... 177
8.16. Эффективный обмен ссылками ... 178
8.16.1. Доступность для поисковых машин ... 179
8.16.2. Осторожность при обмене ссылками ... 181
8.17. Другие советы по привлечению ссылок ... 182
8.18. Ссылки, которые не приносят пользы ... 182
8.18.1. Взаимные ссылки ... 182
8.18.2. Покупка ссылок ... 183
8.18.3. Регистрация в каталогах ... 184
8.18.4. Публикация статей и пресс-релизов ... 184
8.18.5. Правильный ссылочный баланс ... 185

Глава 9. Привлечение трафика и наращивание ссылочной базы при помощи Яндекс.Новости и Google Новости ... 187

9.1. Как воспользоваться сервисами Яндекс.Новости и Google Новости для раскрутки сайтов ... 187
9.2. Партнерская программа Google Новости ... 188
9.2.1. Какая организация страниц нужна для новостей Google? ... 188
9.2.2. Как представить сайт на рассмотрение в Google Новости? ... 191
9.2.3. Как создать карту сайта новостей Google? ... 192
9.3. Партнерская программа Яндекс.Новости ... 192
9.3.1. Введение в Яндекс.Новости ... 192
9.3.2. Как стать партнером Яндекс.Новостей? ... 194

Глава 10. Как анализировать сайты конкурентов ... 197

10.1. Общие рекомендации ... 197
10.2. Анализ структуры ссылок ... 198
10.3. Инструменты анализа, которые предоставляет Google ... 200
10.4. Анализ доменов ваших конкурентов ... 201

Глава 11. Создание идеальной веб-страницы ... 203

Глава 12. Как продавать через Интернет, или секреты копирайтинга ... 207

12.1. Наиболее важные аспекты копирайтинга ... 207
12.2. Какой тип личности наиболее соответствует вашему потенциальному клиенту? ... 211
12.2.1. Доминирование ... 212
12.2.2. Влияние ... 212
12.2.3. Стабильность ... 213
12.2.4. Исполнительность (добросовестность) ... 213
12.3. Различия в написании рекламных текстов для вашего сайта для мужчин и женщин ... 215
12.3.1. Мужчины и женщины в Интернете ... 217
12.4. Писать, угождая каждому ... 217
12.4.1. Длинный текст против короткого ... 218
12.4.2. Ваше изделие (программа) или сервис ... 219

12.5. Заголовки, введение и заключение ... 219
12.5.1. Заголовки "Как..." ... 221
12.5.2. Заголовки, ориентированные на выгоды ... 221
12.5.3. Подзаголовки ... 222
12.6. Свойства против выгод. Вы действительно знаете различие? ... 223

Глава 13. Поисковый спам и "черные" методы оптимизации ... 231

13.1. Устаревшие спамовые технологии ... 235
13.1.1. Спаминг ключевых слов, стаффинг и спамдекс ... 236
13.1.2. "Невидимый" или "полувидимый" текст ... 237
13.1.3. Дорвей-страницы ... 238
13.1.4. Информационные страницы ... 241
13.1.5. Главное — содержание ... 242
13.1.6. Мета- и JavaScript-переадресация ... 243
13.1.7. Скрытие текста и ссылок на странице с помощью CSS ... 245
13.1.8. Тег комментария *<!-- ... -- >* ... 245
13.1.9. Порядок цифр и букв в тексте ... 245
13.1.10. Секрет "невидимого" пиксела ... 246
13.1.11. Техника "Приманка и переключение" ... 247
13.1.12. Перерегистрация в поисковиках ... 247
13.2. Еще несколько спамовых технологий ... 247
13.2.1. IP Delivery, или клоакинг ... 247
13.2.2. DHTML-спаминг ... 250
13.2.3. Автоматически сгенерированный контент ... 250
13.2.4. Линк-фермы ... 250
13.2.5. Спаминг на основе *<noframes>* и *<noscript>* ... 251
13.3. Продвинутые спамовые технологии ... 251
13.3.1. Публицистические Империи (мини-сети) ... 252
13.3.2. "Перелинковка" и создание мини-сетей ... 252

Глава 14. "Черный" PR в Интернете и методы защиты ... 259

14.1. Работа с негативными поисковыми результатами ... 260
14.2. 9 лучших стратегий по предотвращению негативных результатов ... 261
14.3. Небезопасное SEO ... 264
14.3.1. Отслеживайте вашу онлайн-репутацию ... 265
14.3.2. Онлайн-репутация — это все! ... 267

Глава 15. Эффективность интернет-маркетинга для B2B-отраслей ... 269

15.1. Отличия в раскрутке сайтов B2B- и B2C-компаний ... 269
15.2. Проблемы интернет-маркетинга, которые упоминают представители B2B-отраслей ... 270
15.3. Различие в целях поисковой оптимизации ... 272
15.4. Копирайтинг для B2B ... 273
15.4.1. Один пользователь, много поисковых запросов ... 273
15.4.2. Один договор, много пользователей ... 274
15.5. Стратегия ключевых фраз для B2B ... 275
15.6. Под какие поисковики лучше оптимизировать сайт? ... 275
15.7. Инструменты для повышения конверсий в B2B-отраслях ... 276
15.8. Электронные торговые площадки ... 277
15.9. Выводы ... 278

Глава 16. Отслеживание результатов и оценка успеха ... 279

16.1. Введение в веб-аналитику ... 279
16.1.1. Почему измерение успеха так важно для SEO-процесса ... 280
16.2. Измерение поискового трафика ... 283
16.2.1. Общий обзор систем веб-анализа ... 283
16.2.2. Ценные данные SEO в веб-аналитике ... 285
16.2.3. Реферальные сайты ... 287
16.3. Взаимосвязь между SEO, конверсиями и ROI ... 288
16.3.1. Стратегия отслеживания конверсий ... 289
16.4. Сегментирование SEO-кампаний согласно стоимости конверсий ... 289
16.5. Ключевые показатели эффективности для SEO по длинным целевым запросам ... 290
16.5.1. Соотношение брендовых и небрендовых запросов (использование бренда в ключевых словах) ... 290
16.5.2. Сканирование уникальных URL ... 290
16.5.3. Соотношение количества посетителей и страниц, приносящих пользу от поисковых систем ... 291
16.5.4. Соотношение числа ключевых слов и страниц ... 291
16.6. Вывод ... 291

Глава 17. Сводная информация о поисковой оптимизации (SEO Checklist) ... 293

17.1. Ключевые слова ... 293
17.2. Доменное имя и имена файлов ... 293
17.3. Расположение ключевых слов ... 294
17.4. Внутренние ссылки ... 295
17.5. Входящие ссылки ... 296
17.6. Наружные (исходящие) ссылки ... 296
17.7. Внутренние факторы ... 297
17.8. Внешние факторы ... 299
17.9. IP-адрес сервера ... 299
17.10. Накопление ссылок ... 300
17.11. Что можно и что нельзя на сайте ... 300

Заключение ... 302

Глоссарий ... 303

Приложение. Обучение и услуги в области интернет-маркетинга ... 305

П1. Обучение интернет-маркетингу ... 305
П2. Корпоративные семинары ... 306
П3. Услуги в области интернет-маркетинга ... 307
П4. Бесплатная презентация услуг по интернет-маркетингу в офисе вашей компании ... 307

Предметный указатель ... 310

Глава 1

Общие сведения о поисковой оптимизации

1.1. Что такое поисковые машины и как они работают?

В самом начале книги нам хотелось бы подробно рассказать, на чем базируется работа поисковых машин, а также дифференцировать категории поиска так, как это делают поисковые машины (ПМ). Мы считаем это важным, потому что как только вы поймете, что такое поисковые машины и как они работают, вам будет легче понять результаты, поскольку вы увидите, как они появляются на странице после поиска определенного ключевого слова. Это поможет вам рационализировать и затем оптимизировать способ создания веб-страниц, которые будут внесены в базу данных поисковой машины, и лучше понять, почему необходима поисковая оптимизация.

Поисковые системы состоят из пяти отдельных программных компонентов.

- Система выдачи результатов (search engine results engine) извлекает результаты поиска из базы данных.
- Паук (спайдер, spider) — программа, которая скачивает веб-страницы. Она работает точно так же, как ваш браузер, когда вы соединяетесь с веб-сайтом и загружаете страницу. То же действие (скачивание) вы можете наблюдать, когда просматриваете некоторую страницу, выбирая команду **Просмотр HTML-кода** в своем браузере.
- Краулер (Crawler, "путешествующий" паук) — программа, которая автоматически проходит по всем ссылкам, найденным на странице. Краулер выделяет все ссылки, присутствующие на странице. Его задача — определить, куда дальше должен идти паук, основываясь на ссылках или исходя из заранее заданного списка адресов. Краулер, следуя по найденным ссылкам, осуществляет поиск новых документов, еще неизвестных поисковой системе.
- Индексатор (Indexer) разбирает страницу на различные ее части и анализирует их. Элементы типа заголовков страниц, заголовков, ссылок, текста, структурных элементов, элементов `BOLD`, `ITALIC` и других стилевых частей страницы вычленяются и анализируются.
- База данных (Database) — хранилище всех данных, которые поисковая система скачивает и анализирует. Она часто требует огромных ресурсов.

Неверно считать, что поисковые машины сразу ищут сайт, как только вы ввели ключевое слово в запросе. Это — чистой воды миф, что поисковые машины просматривают весь Интернет в поисках нужного запроса. Поисковая машина может искать только в пределах собственной базы данных (индекса). И конечно, количество и новизна данных определяется тем, как часто эта база данных обновляется. Часто такое обновление базы данных той или иной поисковой машины называют апдейтом (от англ. *update* — модернизация, корректировка, обновление информации, данных) — пополнение базы данных поисковой машины новой информацией.

Крупные поисковые машины индексируют информацию подобно каталогу в библиотеке. Храня большое количество информации, они должны уметь быстро находить нужные документы по заданным ключевым словам или фразам. Но самих веб-страниц еще больше, чем непосредственно данных о них. Интернет в целом не имеет никакой четкой структуры, и сайты обладают огромным количеством вариантов авторской разработки и содержания по сравнению со стандартными текстами. Это делает почти невозможным для поисковой машины применение стандартных методов, используемых в управлении базой данных и в обычном информационном поиске.

Поисковые алгоритмы (математические методы программирования, сортирующие найденные результаты) каждого из поисковых сервисов являются уникальными. Вы можете проверить это сами: введите ключевое слово или фразу в поисковую систему **www.yandex.ru** и запомните результаты. Теперь зайдите на **www.rambler.ru** или **www.google.ru** и повторите то же самое. У вас всегда будут разные результаты в различных поисковых машинах. Учитывая этот факт, можно отметить, что необходимо иметь индивидуальный подход к различным поисковым службам. Стоит отметить, что некоторые поисковые системы используют базу и алгоритмы более авторитетных собратьев, например, **www.mail.ru**, использует алгоритм Яндекса, многие зарубежные поисковые системы используют общие алгоритмы.

Google, являющийся самой крупной поисковой машиной в мире и по популярности среди тех, кто ищет, и по базе данных из нескольких миллиардов файлов, имеет только малую долю данных, содержащихся во всей глобальной сети. Сеть растет быстрыми темпами. Исследование, проведенное в 2000 г., обнаружило приблизительно 7,5 млн страниц, добавляемых каждый день. Таким образом, невозможно представить, что какая-либо поисковая машина будет когда-нибудь иметь ежедневно обновляемые данные обо всех сайтах сети.

Здесь речь идет о практических ограничениях, заключающихся в размере веб-страницы. В среднем, каждая страница содержит около 5—10 Кбайт текста, поэтому, даже если учитывать, что поисковая машина хранит информацию только о текстовых страницах, это уже выливается в десятки терабайт, которые находятся в базе данных поисковых машин. Есть так называемая "невидимая сеть", которая представляет собой более 550 млрд документов. С этими документами поисковые машины либо не знакомы (не связаны с другими сайтами), либо не могут получить доступ (некоторые участки защищены паролем), либо их технология просто не позволяет им "захватить" эти страницы (например, те, которые включают только сложные типы файлов: визуальные, аудио, мультипликация, сжатые файлы, файлы под паролем и т. д.). Постоянно "ползать" по сети и индексировать веб-страницы с множеством документов, как это делают поисковики, — задача не из дешевых.

Вы сами убедитесь в этом, когда мы изучим "анатомию" поисковой машины. Поддержка базы данных поисковой машины требует много инвестиций для обеспечения работы, необходимых технических ресурсов и продолжения научных исследований.

Мы должны понимать, что базы данных поисковых машин постоянно изменяются. Google может иметь больше страниц, внесенных в свой индекс, чем, скажем, Яндекс (**www.yandex.ru**). Однако, например, если Яндекс обновит свои данные быстрее, чем Google, то даже при его относительно небольшом количестве страниц это может дать пользователю более свежие и всесторонние результаты. Кроме технического фактора, есть также много других, которые должны быть приняты во внимание.

Мы должны упомянуть, что зачастую поисковые машины указывают большой объем страниц, содержащихся в их базе данных, как признак своей исключительности. Это своего рода игра или соревнование между количеством и качеством имеющихся ресурсов.

Хотя размер — это важный индикатор, другие факторы, касающиеся качества базы данных, могут обеспечить лучшие результаты в плане того, что они гораздо больше соответствуют ключевому запросу пользователя (релевантны, от англ. *relevant* — относящийся к делу). Обнаружение релевантных страниц в сети для индексации — приоритет поисковых машин. Но как машина может определить, насколько важна та или иная страница? Некоторые поисковые машины, например, Яндекс, даже используют ручную проверку на релевантность, так называемые *асессоры* (от англ. *assessor*). Асессоры работают по заданной методике. Существуют определенные критерии, по которым должно измеряться качество страницы. Асессор вводит поисковый запрос и ставит определенную оценку, насколько по этим критериям релевантны сайты, находящиеся в поисковой выдаче. Позже мы еще поговорим о методах, используемых поисковыми машинами для определения того, что делает одни веб-страницы важнее других.

Поскольку поисковые машины часто возвращают несоответствующие запросам результаты, мы должны также остановиться подробнее на том, что информация в базах данных машин должна постоянно обновляться. Кроме новых страниц, которые каждый день появляются в сети, непрерывно обновляются и старые. Рассмотрим пример. В одном из исследований говорится о том, что в результате четырехмесячного изучения полумиллиона страниц было выяснено, что более 23% всех веб-страниц обновляются ежедневно. Около половины страниц обновлялась каждые 10 дней, а некоторые документы и вовсе перемещались на новый доменный адрес.

Спайдеры поисковых машин находят миллионы страниц в день, которые заносятся в базу данных и индексируются. Но, как вы можете понять из вышеописанного, поисковикам очень трудно определить частоту изменений страниц. Спайдер поисковой машины может "заползти" на страницу один раз, затем вернуться к ней, чтобы обновить, и, возможно, обнаружит, что были сделаны какие-то изменения. Но он не может обнаружить, сколько раз страница изменялась, начиная с последнего его посещения.

Некоторые веб-сайты изменяются очень часто. Пример — новостные сайты, где информация должна постоянно обновляться, или сайты интернет-магазинов, где регулярно происходит изменение цен, номенклатуры продукции и т. д. Сегодня проводится много как научных, так и коммерческих исследований для развития методов оперативного обнаружения свежей информации. Даже если "важная"

страница будет проверяться спайдером каждые 48 часов, веб-мастера могут обновлять эти страницы гораздо чаще.

Критически важно, чтобы ваш сайт работал, когда его посещает робот поисковой машины. Если в этот момент он не работает, то вы можете исчезнуть из индекса до следующего обновления! Поисковая машина думает, что ваш сайт не существует, поэтому удаляет его из списков после нескольких заходов.

Если веб-мастер загружает страницу на сервер, а затем делает страницу доступной для поиска через опцию **Добавить URL** в поисковой машине, или если страница просто найдена поисковой машиной по ссылке с другого сайта, то содержание страницы в индексе будет именно таким, каким оно было при индексировании поисковой машиной во время обхода.

Так, если в день индексации страница имеет определенное количество слов, которые содержатся в определенном числе параграфов и в определенной степени относятся к ключевому слову, — все это будет записано в индексе поисковой машины до следующей индексации. Если автор страницы решит дополнить ее (добавить изображения, заголовки, выполнить редактирование текста), то поисковая машина не будет знать об этом до следующего посещения страницы. Если пользователь делает запрос на определенную тему в тот день, когда поисковая машина как раз обновила эту страницу, то он получит обновленную информацию, которая уже попала в базу данных поисковика. Однако если пользователь выполняет поиск после того, как автор изменил страницу, то поисковая машина приведет его по ключевой фразе к той же странице, даже если автор, возможно, изменил контекст или убрал важные ссылки на тему, не ставя об этом в известность поисковые машины. Конечно, такая ситуация расстраивает пользователя, который хотел найти релевантную его запросам веб-страницу. Это, как вы понимаете, является главной проблемой поисковых машин. Они просто не могут постоянно следить за изменениями веб-страниц.

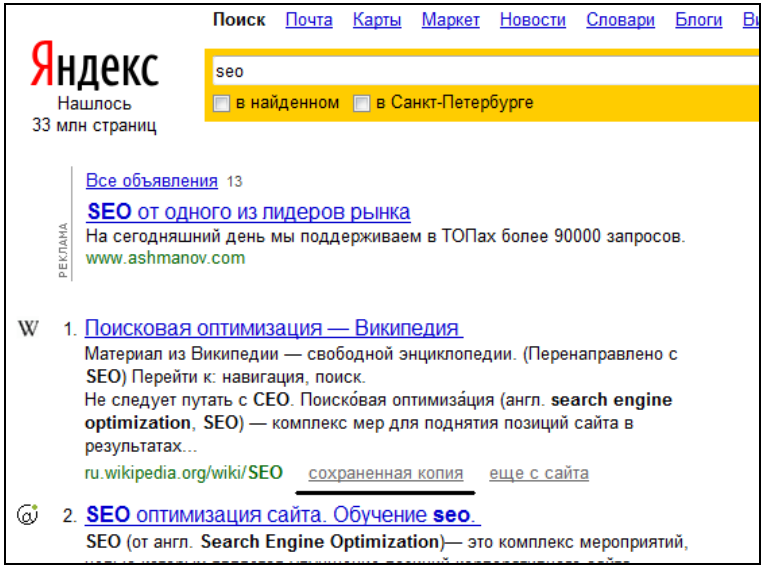

Рис. 1.1. Сохраненная копия страницы в поисковой системе Яндекс

Посмотреть, какая версия страницы в настоящее время находится в индексе поисковой системы, можно, щелкнув в результатах поиска по ссылке **сохраненная копия** (рис. 1.1).

1.2. Особенности поиска с точки зрения поисковой машины

Одна вещь, которая выходит на первый план в области автоматического поиска, выполняемого поисковыми машинами, — это тот факт, что обычные методы поиска информации (например, как в библиотеке) просто не подходят для поиска сетевого. Со времен одной из первых работ в области автоматического поиска по тексту, написанной американцем Джерардом Салтоном с помощью векторной пространственной модели, до последних экспериментов и открытий анализа ссылок и технологий автоматического изучения, а также классификации текста, остается вопрос: как же мы можем получать столь же эффективные и точные результаты поиска, как в библиотечном каталоге?

Чем же обладает библиотечный каталог, чего нет у поисковых машин? Давайте посмотрим на характеристики поиска, проводимого поисковыми машинами, а затем вернемся к библиотечному каталогу.

Основные особенности поиска делятся на три категории. Первая вещь, которую следует отметить, — это различие между традиционным поиском информации и теми проблемами, с которыми такой поиск сталкивается в сетевом применении. Хотя алгоритмы были приспособлены для традиционного поиска, чтобы направлять системы гипертекста, сеть нуждается в более четкой структуре и четком различии этих родственных систем. Например, небольшие, тщательно контролируемые частные собрания текстов вроде научных работ или новостей легче привести в соответствие с критериями отбора. Конференция по поиску текста (TREC) утвердила точку отсчета для такого большого объема (собрания писем, тексты и т. д.) информации, как 100 Гбайт информации. (Поисковые системы уже имеют десятки терабайт хранимой информации: для того чтобы дать представление о размере информации, скажу, что одна большая энциклопедия заняла бы порядка 1 Гбайт, а общественная библиотека более чем из 300 000 книг равнялась бы приблизительно 3 Тбайт информации.)

Сеть, как известно, является огромным собранием гетерогенных страниц, которые создаются и развиваются любым желающим, не подвергаясь какому-либо контролю.

Эта нехватка структуры управления или стандартизации в сети ведет к "взрыву" количества доступной информации, но это также создает огромные проблемы для поисковых машин при поиске необходимой информации в сети. Основной вопрос, являющийся актуальным в данном случае: насколько продуктивны результаты, которые мы получаем в процессе поиска?

Существуют три обобщенных класса поиска в сети.

- **Информационный.** Такой поиск осуществляют пользователи, которые действительно ищут необходимую им информацию в сети. Они формулируют запрос так же, как в речи, например, "низкий гемоглобин". Это медицинский термин. Пользователь ищет определенную информацию об этом термине. Это очень близко к классическому информационному поиску.

- **Навигационный.** Применяется, когда пользователь хочет попасть на определенный веб-сайт. При этом формулируется такой запрос, как, например, "РосБизнесКонсалтинг". В этом случае то, что он действительно хочет найти, должно находиться непосредственно на сайте компании "РосБизнесКонсалтинг" (**www.rbc.ru**). Если бы кто-то напечатал "РИА Новости" в строке поисковика, то, вероятно, он хотел бы попасть на сайт "РИА Новости", а не найти историю РИА и т. п. Все мы формулируем большое количество подобных запросов, и они, фактически, составляют порядка 20% от всех запросов.
- **Деловой.** Означает, что, в конечном счете, пользователь хочет сделать что-либо через сеть, с помощью сети. Хороший пример — посещение интернет-магазина. Вы действительно хотите купить что-либо. Или вам необходимо скачать файл, или найти услугу, например, Желтые страницы. То, что вы действительно хотите сделать, вовлекает вас в сделку. Если обратиться к вопросу посещения магазина, здесь люди хотят купить товар и т. д., таким образом, они желают найти ответ на свой запрос, который может удовлетворить их потребность.

Таким образом, когда мы говорим о точности и актуальности ответа на запрос, важно различать эти три класса. Поскольку, например, классический деловой запрос для жителя Санкт-Петербурга будет нечто иным, чем для кого-либо, живущего в Москве.

В случае с некоторыми деловыми запросами трудно решить, какой из результатов лучший. Контекст играет большую роль. Что касается деловых запросов, иногда более эффективный результат мы получаем из других источников, а не из того, что собираем сами. Это как поход за покупками: чтобы что-то купить, вам нужно идти в торговый центр, а не в библиотеку.

Вернемся к библиотеке. Поисковой машине довольно трудно постигать природу запроса пользователя. Машина может найти сайт на подходящую по ключевому слову тему, связующую ссылку в том или ином тексте, даже выбрать цитату, но она не может интуитивно понять цель запроса. Если же вы приходите в библиотеку маленького городка и просите библиотекаря помочь найти ту или иную информацию, то он, наверняка, поймет, для чего вам нужна эта информация, и предложит конкретное место поиска.

Как уже говорилось, многое из того, что пытаются достигнуть поисковые машины, основано на обычной системе информационного поиска. Предположим, я пришел к библиотекарю и спрашиваю, есть ли в библиотеке определенная популярная книга. Если библиотекарь понимает, что есть эта книга, то он или найдет книгу, или закажет ее. Когда библиотекарь получит книгу, то сделает определенные записи о книге в библиотечном каталоге. Запись о книге будет включать: название книги, имя автора, некоторые ключевые слова, описывающие содержание, порядковый номер идентификации (ISBN), заголовок, относящий книгу к той или иной категории, и номер индекса для последующего поиска книги. В этом случае, книга была бы помещена на полки библиотеки в алфавитном порядке, а карточки с индексом книги занесены в каталог.

Качественная библиотечная система каталогов позволяет вам искать не только по названию, но и по автору или по категории. Получая большое количество запросов о некой особенной книге или теме, библиотекарь может иногда интуитивно

сказать, где находится данная книга, или, по крайней мере, указать секцию. Регистрируя выбор и возврат книг, она может также заметить, сколько времени та или иная книга отсутствовала и сколько читателей брали ее за какое-либо время, другими словами, какие книги пользуются наибольшей популярностью, а какие просто пылятся на полках. Все эти наблюдения помогут в усовершенствовании библиотечной системы, позволяя перемещать "устаревшие" книги, которые больше не пользуются популярностью, на более отдаленные полки, освобождая место для более популярной литературы.

Если вы подумаете над этими двумя только что прочитанными абзацами, то поймете, что фактически, хотя и в довольно странном виде, это описание принципа работы поисковых машин. На первый взгляд кажется, что это простой процесс, но в реальности проблемы, с которыми сталкиваются крупнейшие разработчики в этой сфере, пытаясь подражать этому принципу, весьма обширны и серьезны.

Можно привести пример школьника, который приходит в библиотеку и спрашивает книгу об Италии. Здесь библиотекарь, обладая достаточно скромной информацией, может предположить, что школьнику необходимо написать реферат, и поэтому ему нужны книги об истории и культуре Италии. Если подросток приходит в библиотеку летом и спрашивает книги об Италии, библиотекарь может решить, что он собирается в Италию на каникулы и ему нужны путеводители по Италии. Поэтому и укажет ему на подобные книги.

Иными словами, библиотекарь может помочь вам более точно сформулировать ваш вопрос, чтобы подвести вас к наиболее подходящим для вас темам, понимая смысл и цель вашего поиска. Это доказывает, что те базы данных, которые отсортированы по категориям, имеют меньше проблем, и именно поэтому некоторые называют редакторов интернет-каталогов "библиотекарями" Интернета. Поисковые машины не всегда способны точно определить сферу вашего запроса, но они могут, по крайней мере, попытаться вывести те страницы, которые они посчитают специфичными и связанными с вашим запросом при помощи "топологии сети".

Кроме того, независимо от тематики запроса, большинство поисковых машин все еще выводит несколько тысяч или несколько миллионов вариантов, из которых только некоторая часть может быть действительно релевантной. Так как же быстро снижается показатель релевантности? Ответ таков: после первых двух страниц результатов этот показатель начинает резко снижаться.

1.3. "Анатомия" поисковой машины

Поисковые оптимизаторы регулярно пытаются разобраться в алгоритмах поисковых машин, чтобы было легче объяснить, что на самом деле представляет процесс от индексации до нахождения нужного результата. Хорошая поисковая машина, работающая со всей своей оптимальной производительностью, должна обеспечивать эффективное местонахождение веб-страниц, полный охват сети, новейшую информацию, непредвзятый равноправный доступ ко всей информации, удобный интерфейс для пользователей, кроме того, выдачу наиболее подходящих результатов на момент выполнения запроса.

Обеспечение содержательного доступа к большим объемам информации — это трудная задача. Наиболее успешные методы и подходы к информационному поиску, распределение документов по категориям очень сильно полагаются на статистические технологии.

Существует пять наиболее важных модулей, благодаря которым работают поисковые машины. Для специфических компонентов используются различные термины, но я взял стандартные и надеюсь, что эти описания и пояснения легче воспринять, чем те, которые приводятся в технических документах.

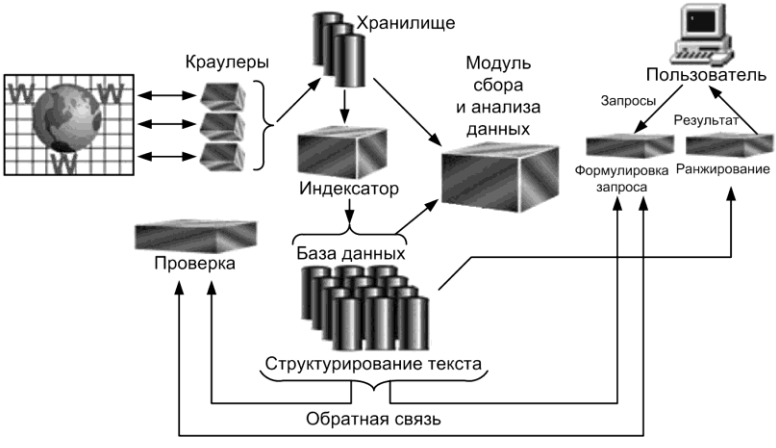

Рис. 1.2. Схема работы поисковых систем

Выделим пять наиболее важных компонентов поисковых систем (рис. 1.2):

- модуль краулера/спайдера;
- модуль склада/базы данных;
- модуль индексатора/анализа ссылок;
- модуль поиска/ранжирования;
- пользовательский интерфейс запроса.

Поисковые машины тщательно хранят свои методы индексации и ранжирования веб-страниц как коммерческие тайны. Каждая поисковая машина имеет свою уникальную систему. И хотя алгоритмы, которые они используют, могут различаться, на практике между ними много схожего в том, каким образом они ведут построение системы индексов.

Поисковые машины находят веб-страницы тремя способами:

- при помощи использования стартового сбора URL-страниц (по-другому веб-страниц) и извлечения из них ссылок, чтобы следовать по ним (например, выбирая их из каталогов);
- из списка URL, полученного при прошлом исследовании сети (используя первые результаты проведенных поисков);
- URL, добавленных искусственно веб-мастерами непосредственно в поисковую машину (при помощи опции **Добавить URL**).

Существует большое количество проблем, с которыми сталкиваются спайдеры поисковых машин из-за размеров сети, ее постоянного непрерывного роста и изменений. Как вы теперь знаете, в отличие от традиционного поиска информации, где все данные собраны в одном месте и готовы к проверке, информация в сети распределена между миллионами различных интернет-серверов. Это означает, что информация должна быть сначала собрана и систематически распределена по большим "складам", прежде чем она будет доступна для обработки и индексации.

Кроме того, хорошая поисковая машина должна иметь хорошие фильтры, позволяющие избежать множества проблем, которые могут быть, сознательно или нет, созданы владельцами веб-сайтов. Эти фильтры автоматически избавляются от миллионов ненужных страниц.

Современные поисковые машины умеют сами себя подстраивать, т. е. они сами определяют частоту сканирования того или иного сайта в зависимости от множества факторов, например, частоты обновления веб-ресурса, его рейтинга и т. п.

Существует много различных типов краулеров в сети. Есть такие, которые применяются для личного использования непосредственно с рабочего стола вашего компьютера, и такие, которые собирают адреса электронной почты, или разного рода коммерческие краулеры, которые выполняют исследования, измеряют сеть и выявляют работу шпионских программ.

Описанные краулеры, спайдеры и роботы — это автоматизированные программы, которыми обычно управляют поисковые машины, ползающие по ссылкам сети и собирающие первичные текстовые данные и другую информацию для индексации.

Ранние краулеры были запрограммированы для довольно-таки общих целей. Они обращали мало внимания на качество или содержание страниц, а больше внимания акцентировали на количестве. Их целью было собрать как можно больше страниц. Однако сеть тогда обладала гораздо меньшими объемами, поэтому они были достаточно результативными с точки зрения обнаружения и индексации новых веб-страниц.

С ростом сети краулеры столкнулись со многими проблемами: масштабируемостью, отказоустойчивостью и ограничением пропускной способности. Быстрый рост сети опередил возможности систем, которые не были достаточно подготовленными, чтобы досконально исследовать загружаемую информацию, с которой они сталкивались. Попытка управлять набором программ одновременно на таких уровнях, не повреждая при этом систему, стала невозможной.

Сегодняшние краулеры, которые появились в течение последних нескольких лет в результате роста сети, полностью изменились со времен ранних роботов. Однако, хотя они все еще используют ту же самую базовую технологию, теперь они запрограммированы на более индивидуальные собственные разноуровневые системы.

Хотя "ползание" по сети — очень быстрый процесс, фактически, краулер делает те же самые действия, что и рядовой серфер-пользователь.

Краулер начинает или с одного URL, или с набора страниц. Например, страниц, внесенных в определенный каталог, которые он загружает, извлекает гиперссылки и затем "ползет" к страницам, на которые указывают эти ссылки. Как только краулер натыкается на страницу без ссылок, по которым он может последовать, он возвращается на уровень назад и переходит к ссылкам, которые он, возможно, пропустил

раньше, или к тем ссылкам, которые были запрограммированы заранее в очереди для следующего раза. Процесс повторяется от сервера к серверу, пока на страницах больше нечего будет загрузить, или пока некоторые ресурсы (время, ограничение пропускной способности) будут достигнуты или исчерпаны.

Слово "краулер" почти всегда используется в единственном числе, однако большинство поисковых машин имеет множество краулеров с целым "флотом" агентов, выполняющих масштабную работу. Например, Google, как поисковая машина нового поколения, начавшаяся с четырех краулеров, каждый раз открывает приблизительно триста ссылок. На пиковых скоростях они загружали информацию более чем из 100 страниц за секунду. Google теперь работает на трех тысячах компьютеров с системой Linux, общий размер винчестеров которых более 90 Тбайт. Google добавляет 30 новых машин в день к их серверу, только чтобы не отставать от роста.

Краулеры используют традиционные схематичные алгоритмы, чтобы обследовать сеть. График составлен из того, что называется "узлами" и "краями" (nodes and edges). Узлы — это URL, а края — ссылки, вложенные в страницы. "Края" — это передовые ссылки ваших веб-страниц, которые указывают на другие страницы, и ссылки возврата, которые указывают путь назад на ваши страницы с любого другого места сайта. График сети может быть представлен математически для поисковых целей с помощью использования алгоритмов, где будет определено пересечение или "начальной широты", или "начальной глубины".

Поиск, исходя из "начальной широты", означает, что краулер восстанавливает все страницы вокруг отправной точки "ползания" прежде ссылок, ведущих еще дальше от начала. Это самый общий путь, по которому спайдеры следуют за ссылками. Поиск, исходя из "начальной глубины", может быть использован для того, чтобы следовать за всеми ссылками, начиная с первой ссылки первой страницы, затем с первой ссылки на второй странице и т. д. Как только первая ссылка на каждой странице будет посещена, спайдер перемещается на вторую ссылку, а затем — на каждую последующую. Метод поиска, исходя из "начальной широты", позволяет уменьшить нагрузку на серверы, которая быстро распределяется и помогает избежать того, чтобы какому-то одному серверу придется быстро отвечать на тот или иной запрос постоянно. Метод "начальной глубины" программируется легче метода "начальной широты", но ввиду своих возможностей может привести к добавлению менее важных страниц и отсутствию свежих результатов поиска.

Существует вопрос: насколько глубоко краулер может проникать в веб-сайт? Многое зависит от содержания веб-сайтов, с которыми сталкиваются краулеры, а также от того, какие страницы поисковая машина уже имеет в своей базе данных. Во многих случаях более важная информация содержится в начале страницы, и чем дальше идти от начала страницы, тем менее важная информация там содержится. Логика в том, что более важную информацию для пользователя стараются всегда расположить в начале чего бы то ни было. Стоит вам только зайти на сайт, как вы обнаружите, что у него, как и у многих сайтов, нет четкой структуры, правил и стандартов, однако часто именно ссылки, содержащие важную для пользователя информацию, располагаются в начале страницы.

Поисковые машины вообще предпочитают идти по более коротким URL на каждом посещенном сервере, предполагая, что URL с более короткими компонентами, вероятно, будет содержать более полезную информацию.

Это означает, что

http://www.mysite.ru/blue-cars.html

вероятно, считается более важным, чем

http://www.mysite.ru/products/webcatalog/cars/blue/spec~series17.html

или еще более длинный адрес, который ведет к более глубокой информации в иерархии веб-сайта. Спайдеры могут быть ограничены определенным количеством подразделов (слешей) сайта, по которым они будут искать информацию. Десять слешей — это максимальная глубина; средней же глубиной, как правило, считаются три слеша.

Важные страницы, которые находятся более глубоко на сайте, вероятно, придется напрямую регистрировать в поисковой машине владельцу сайта, а также обязательно создавать карту сайта. С постоянным развитием сети и связанных с ней технологий, таких как ASP, PHP и ColdFusion, часто многие важные страницы "скрыты" в глубине сетевых баз данных, но это уже не относится к алгоритмам поисковых систем. Все это и многое другое мы подробно рассмотрим в последующих главах, в особенности в *главе 6*.

А сейчас, считаю необходимым сказать пару слов о поисковой оптимизации. Что такое SEO?

1.4. Что такое SEO (Search Engine Optimization)?

Поисковая оптимизация (Search Engine Optimization, SEO) — это процесс корректировки HTML-кода, текстового наполнения (контента), структуры сайта, контроль внешних факторов для соответствия требованиям алгоритма поисковых систем, с целью поднятия позиции сайта в результатах поиска в поисковых системах по определенным запросам пользователей. Чем выше позиция сайта в результатах поиска, тем больше заинтересованных посетителей перейдет на него с поисковых систем.

Это техника достижения более высокого положения в поисковых системах путем изменения сайта в сторону максимальной совместимости с поисковыми машинами. Иногда под SEO подразумевают интернет-маркетинг в целом, но на самом деле, SEO — лишь одна из его сторон.

SEO — это часть интернет-маркетинга, цель которого — получить приток качественно важной аудитории на сайт с натуральных результатов поиска.

Поисковая оптимизация важна еще и потому, что помогает привлечь посетителей на сайт бесплатно. То есть вам не придется платить за каждого посетителя (за каждый щелчок по вашему объявлению контекстной рекламы). Многие серферы не смотрят на проплаченные позиции в результатах поисковиков, поэтому если вы на бесплатных первых позициях, то, скорее всего, этот серфер превратится в вашего посетителя.

Все это применимо как к малому, так и к среднему и крупному бизнесу. Некоторые компании не в полной мере понимают процесс формирования рейтинга,

поэтому могут игнорировать какие-то моменты маркетинговой стратегии, которые могут стать для них критичными.

Пока вы не станете продавать что-то уникальное, у вас всегда будет сотня сайтов-конкурентов, предлагающих тот же товар или услугу. В зависимости от вида предложения количество конкурентов может возрастать до тысяч и миллионов. Вы когда-нибудь задавались вопросом, насколько легко будет обычному посетителю найти ваш сайт через Яндекс по ключевым словам для вашего ресурса? Если вы купили данную книгу, то, наверное, задавались.

Если вы сотрудник крупной компании и ваш сайт не выводится в десятке лидеров (ТОП 10) поисковых систем по вашему ключевому слову, то ваша репутация повреждена. Если вы дальше 30-й позиции в рейтинге, считайте, что вас вообще не видно в сети. Исследования показывают, что посетители редко идут дальше 30-й позиции в результатах (согласно Forrester Research Inc.). Считается, что 30 топовых результатов дают приток 90% посетителей в результате поиска.

Вот как посетители Интернета находят сайты согласно Forrester Research Inc.:

- поисковые машины — 81%;
- ссылка с другого сайта — 59%;
- посредством маркетинговых коммуникаций — 56%;
- телевидение — 48%;
- посредством ввода и угадывания адреса — 41%;
- онлайн-реклама — 20%;
- радио — 19%;
- директ-мэйл — 10%.

Таким образом, SEO — это эффективный способ донесения ваших посланий для широкой и нужной аудитории. SEO поможет вам не затеряться во множестве конкурентов.

А вот другие причины:

- большинство поисковиков привлекает больше уникальных посетителей, чем любой другой сайт;
- 85% всех использований Интернета включает загрузку поисковиков. Поэтому поисковик играет важную роль в привлечении посетителей на ваш сайт;
- потребители используют поисковики для нахождения и покупки товаров или для решения некоторых задач (выбор места отдыха, медицинского лечения или покупке автомобиля);
- люди, которые нашли ваш сайт с помощью поисковиков, — это наиболее подходящая для вас целевая аудитория, т. к. они вводили в строке поиска ваше ключевое слово.

Потребность в специалистах SEO, или "поисковых оптимизаторах", значительно возросла за последние пять—десять лет. Обычно они работают или с администратором сайта, или с дизайнером, изменяя код и содержание сайта так, чтобы сайт занял достойное положение в поисковых системах. Ключ к успешной оптимизации — найти специалиста SEO с хорошими рекомендациями среди огромного количества неопытных и неэтичных исполнителей.

Путь к вершинам строчек выдачи поисковой машины для популярных ключевых слов — это постепенный процесс. Возможно, некоторые из ваших конкурентов уже могут похвастаться топовыми позициями. Они также могут иметь преимущество в том, что находятся в сети в течение долгого времени, а время — либо самый большой друг поисковой оптимизации, либо худший враг, — все зависит от вашей ситуации.

Если поисковая машина уже решила, что ваш конкурент — самый релевантный сайт по определенному ключевому слову, то вы должны понимать: необходимы время и усилия, чтобы эта поисковая машина поместила ваш сайт на первые строки выдачи.

Работа над достижением первых строк выдачи для определенного ключевого слова может занимать месяцы. За это время вы можете постепенно оптимизировать ваш сайт, собрать более релевантные ссылки, увеличить качество вашего сайта и не успеть понять, как вы оказались на первых строчках поисковых машин.

Для конкурентных ключевых слов описанное — это общий способ, который может сделать трудной жизнь SEO-специалистов, т. к. многие клиенты нереалистично ожидают немедленных результатов.

Если же вам необходим мгновенный результат, вы можете получить его через контекстную рекламу (или Pay-Per-Click (PPC)), а также через планирование широкого диапазона менее конкурентных ключевых слов. В *главе 4* вы узнаете, как найти правильные ключевые слова, которые прибавят вам клиентов из ниш, которые ваши конкуренты не успели занять. Но если вы будете конкурировать только в популярных ключевых словах, где предложений уже полно, вам придется позаботиться о большем количестве действительно хороших ссылок, чтобы добраться до верхних позиций в выдаче.

1.5. Поисковая оптимизация: прошлое, настоящее и будущее

1.5.1. Прошлое поисковой оптимизации

1997—1999 гг. — период становления поисковых машин. SEO-специалисты занимались тем, что добавляли сайты в поисковики. Отголосок прошлого слышен и поныне, когда работающие по старинке SEO-компании предлагают зарегистрировать ваш сайт в тысячах возможных поисковых системах. Программы индексации, используемые поисковиками (роботы, спайдеры), просматриваются через HTML-код страницы и используют некоторые алгоритмы Page Rank, Trast Rank и другие, которые держатся в секрете.

Те дни были блаженством для спамеров, и было гораздо проще добиться высоких рейтингов сайтов. Вы могли использовать ключевые слова сколько угодно раз в теле контента, в `META`-тегах, комментариях и т. д. Можно было также спрятать все это от глаз посетителей, сделав текст невидимым с помощью специальных HTML-уловок. Поисковая машина не обладала достаточными технологиями, чтобы распо-

знать данный спам, поэтому такие сайты легко становились на верхние позиции рейтингов. Сегодня тоже встречается такая примитивная оптимизация (однако сегодня подобные проекты чаще всего банятся (удаляются из индекса)).

Исключением всегда был Yahoo, где человек проводил индексацию и исключал спамерские страницы.

Со временем поисковые машины стали способны отличать спамовые страницы и штрафовать за использование спамовых методов. Тем не менее, оптимизаторы поисковиков всегда были на один шаг позади поисковиков в обнаружении новых способов обмана алгоритмов индексации. Именно поэтому поисковики доставляют нам только релевантные результаты, избавленные от различных спамерских трюков. Некоторые машины начинают применять различные способы индексирования.

Внестраничные факторы как новое решение

В середине 1999 г. поисковые порталы пытались использовать логику обычного интернет-серфера, чтобы улучшить поисковые результаты. Один из таких не зависящих от контента страницы (а значит, в какой-то мере и от спама) способов — это отслеживание щелчков пользователей. Поисковая система DirectHit представила технологию, которая отслеживала, какие сайты выбирают пользователи Интернета. Если посетители часто выбирают один и тот же сайт по определенному ключевому слову, то, скорее всего, этот сайт поднимется в рейтингах (данный эффект можно было увидеть в поисковой системе Рамблер, первые позиции в которой занимали сайты, находящиеся на первых местах в рейтинге Топ100).

Еще один способ сделать алгоритмы индексирования более эффективными — это отслеживать количество сайтов, ссылающихся на данную страницу. Это правило пришло из системы библиотек и архивов и до сих пор играет важную роль в притоке посетителей на сайт. Этот способ был назван "ссылочной популярностью" (link popularity) и остается одним из основных факторов в ранжировании.

Оба описанных параметра известны как внешние факторы ("off-the-page"), т. к. это факторы, не зависящие от контента вашей страницы. Такие страницы свободны от спамерского влияния.

Индустрия SEO нашла выход в создании так называемых ссылочных ферм (link farm). Идея заключается в следующем: если поисковые машины учитывают количество входящих ссылок на сайт, то можно создать специальную внешнюю ссылочную страницу, которая будет полезна спайдерам. На этой странице можно размещать много ссылок на нашу страницу и на другие подобные страницы, и эти ссылки сделают свое дело, когда робот начнет их индексировать.

Подобные ссылочные страницы имеют ссылки друг на друга. Официальной целью подобных ссылочных ферм было создание сообщества, объединенного ссылками. Не столь явная, но понятная цель — достижение высоких позиций в рейтингах для своих сайтов.

Пока ссылочные фермы наслаждались своей популярностью и разрабатывалось программное обеспечение для обмена ссылками между фермами, спайдеры поисковиков уже знали обо всем этом. Сегодня поисковики уже не дают большого преимущества от использования этих методов.

Издаваемые сообществами каталоги

Yahoo была первой поисковой машиной и до сих пор остается одной из самых популярных. Поскольку каталоги Yahoo создаются людьми, поисковик столкнулся с нехваткой человеческих ресурсов. Редакторы Yahoo все равно стремятся достигнуть максимального количества охваченных страниц.

Издаваемые сообществами каталоги впервые появились в 1999 г. Эта система позволяет тысячам редакторов, объединенных Интернетом, постоянно обновлять каталоги. Первая из таких сетей — это каталог Netscape Open, а каталог **www.go.com** стал еще одним лидером.

Каталог Netscape Open, помимо того, что издается сообществом редакторов, является также каталогом открытого доступа. Любой желающий, имеющий намерение улучшить работу своего поискового портала, может получить доступ к этому каталогу. Показателен один только факт, что к 2000 г. списки из каталога Netscape Open стали появляться во многих поисковиках.

Когда издаваемые сообществами редакторов каталоги имеют не только количественную, но и качественную сеть ссылок, они становятся весовыми игроками на поисковом пространстве. Созданные человеком каталоги играют роль своего рода поисковых адвокатов, т. к. обеспечивают весьма релевантные результаты поиска. С увеличением значения таких каталогов к 2000 г. маркетологи стали концентрироваться на сфокусированной, целевой и качественной оптимизации сайтов.

Подъем Google

Поисковая система Google начала свой путь к званию Короля поисковиков в 2000 г., а в 2002 г. уже с полным правом могла себя так называть, т. к. 70% поиска практически во всех странах (кроме России, Китая и Эстонии) происходило через нее. В России на первом месте по популярности находится поисковая система Яндекс, в Китае — Baidu.com и в Эстонии — Neti.ee. Google в этих странах занимает твердое второе место. Пока другие поисковики стараются быть как можно более универсальными, Google остается простой и уникальной поисковой системой — с простым интерфейсом и релевантными результатами.

Google также единственная поисковая система (ПС), которая в более полном объеме умеет искать и индексировать файлы формата SWF (shockwave flash). Кроме того, благодаря внестраничным технологиям, Google устойчива к спаму. Преобладание Google закрепилось, когда в 2000 г. Yahoo переключился на него с Inktomi в качестве второго поисковика. Сейчас Yahoo использует комбинацию результатов Overture и собственного поискового софта, т. е. здесь нет доминирования Google.

Теперь Google предлагает множество инструментов и сервисов, о которых стоит рассказать. Вот те из них, которым я решил уделить внимание: файлы Sitemap, Google Analytics, Оптимизатор веб-сайтов Google Website Optimizer, iGoogle, Page Creator и многие другие.

- **Файлы Sitemap.** Google представил новую возможность для ускорения процесса "ползания" по большим сайтам. Технология Sitemap — еще один способ проинформировать Google о вашем сайте и рассказать его спайдерам об основной информации на сайте (количество страниц, частота обновлений, страницы для

индексации). Технология также обеспечивает статистические данные о сайте. Все, что вам нужно сделать, — это создать XML-файл карты сайта, поместить его в корневой каталог и дать об этом знать по адресу **http://google.com/webmasters/sitemaps/siteoverview**. Всю информацию об этой технологии можно найти здесь: **http://www.google.com/webmasters/**.

- **Google Analytics** (**analytics.google.com**) — это относительно новый инструмент, который может дать важную для вашего бизнеса информацию: вы можете определить, откуда к вам приходят посетители, какие ссылки наиболее значимы в плане привлечения посетителей, сколько времени проводят посетители на страницах вашего сайта, на какие ссылки они нажимают, узнать ключевые слова поиска по сайту и др. Google Analytics может быть использован прямо из интерфейса AdWords. Тогда вы можете даже получать соответствующие данные и отчеты. Чтобы подписаться на эту услугу, посетите сайт **http://www.google.com/analytics/sign_up.html**.
- **Оптимизатор веб-сайтов Google Website Optimizer** — бесплатная платформа для тестирования, которая с легкостью позволяет создавать тесты и анализировать полученные результаты.

  В процессе тестирования происходит сравнивание элементов на сайте или в рекламной кампании с целью увидеть, какое изменение лучше всего убедит посетителей сделать то, что от них требуется. Примером таких действий может быть покупка, проявление определенной инициативы, подписка на рассылку новостей, или даже просто щелчок как следующий шаг в процессе продажи (или в процессе покупки).

  Сайт инструмента — **http://www.google.com/websiteoptimizer**.
- **iGoogle** — это сервис индивидуализированной стартовой страницы (Personalized Homepage). Цель этого сервиса — создать для каждого пользователя собственную стартовую страницу Google, на которой была бы только та информация, которая ему интересна: погода в выбранном регионе, RSS-новости с любимых сайтов, курсы валют, текущая дата и время и т. д.

  Сайт сервиса — **http://www.google.ru/ig**.
- **Page Creator** — инструмент, который дает возможность публиковать различные варианты своих страниц в сети (**http://pages.google.com**) и постоянно обновлять их. Конечно, это не сайты с профессиональным дизайном, а просто странички, например, для тех, кто хочет выложить свои фото с каникул. Для использования этого сервиса вам не надо устанавливать никакого дополнительного оборудования, вам нужен только браузер с разрешением использования cookies и JavaScript, а также аккаунт Gmail (если у вас его нет, зайдите на **www.gmail.com**). Google будет бесплатно обслуживать ваш сайт, а его спайдеры — индексировать его каждые несколько часов (однако у сайта не будет преимуществ в области SEO). Вся информация о сервисе доступна на **http://pages.google.com/-/about.html**.

1.5.2. SEO и SEM (Search Engine Marketing) сегодня

Если вы до сих пор думаете, что продвижение сайта может выполняться только путем автоматического программного обеспечения, то вы ошибаетесь. SEO требует усовершенствования качества и популярности сайта. Чтобы сайт мог

достигнуть топовых позиций в рейтингах, вы должны опираться на анализ целевой аудитории, анализ работы конкурентов и копирайтинг. Поскольку тенденции постоянно меняются, SEO-специалисты должны отслеживать все направления в этих вопросах.

Сегодня только неопытные оптимизаторы будут использовать спам для раскрутки коммерческого сайта. В большинстве случаев спам и так называемые "черные" методы поисковой оптимизации распознаются автоматическими спайдерами, которые становятся все умнее. В конце книги мы вернемся к так называемым "черным" методам оптимизации и рассмотрим спамовые методы, но только для того, чтобы предупредить вас о них. Мы не рекомендуем их использовать, т. к. они не только не гарантируют помощь, но и портят репутацию вашего ресурса.

Существует одно древнее высказывание: "Есть только одна вещь, которая тяжелее зарабатывания денег — сохранение этих денег". Эта жемчужина мудрости распространяется и на поисковые системы. Единственная вещь, которая тяжелее получения рейтинга — это его сохранение.

Профессионалы, которые "делают" деньги, знают, как их хранить, а мудрые поисковые оптимизаторы — как зарабатывать и поддерживать рейтинг. Таким образом, они развивают стабильные системы и стратегии, которые поддерживают их успех.

"Быстро разбогатеть" и "быстро подняться в рейтинге" — это глупые сказки. Ветер фортуны — переменчивая штука, поэтому надо думать о долгосрочных планах. Те, кто лидирует благодаря быстрым и коротким стратегиям, обречены быть обогнанными мудрыми старыми "черепахами", которые просчитали долгосрочный путь. Звучит скептически? Хорошо, давайте исследуем некоторые факты.

Стабильная SEO против нестабильной

Не секрет, что мы, поисковые оптимизаторы, всегда имели много способов манипулировать результатами поиска. Некоторые из "манипулирующих" стратегий ненавидимы поисковыми системами, а некоторые — любимы. Примеров слишком много, достаточно сказать, что такие стратегии, как уместные входящие ссылки, ключевые слова в `Title`-тегах, изобилие уникального контента — это стратегии, которые работают сегодня и будут работать всегда, т. к. они не только помогают рейтингам, но также поощряются поисковыми системами, потому что помогают им выдавать наиболее релевантные результаты.

Стратегии, которые используют искусственные входящие ссылки (нетематический обмен ссылками, покупные ссылки и т. д.), переполнение ключевыми словами и т. д., работали раньше и, возможно, все еще работают, если ими воспользовались грамотно и втайне от поисковых систем. Но все это относится к игре с огнем, т. к. в любой момент все ваши старания и достигнутые результаты могут рухнуть, и вы останетесь ни с чем.

Различие между этими примерами в том, что первые содержат стратегии, которые любят поисковые машины, потому что они помогают им поддерживать релевантность в пределах индекса. Последние примеры находятся в числе тех стратегий, которые поисковые машины ненавидят, потому что они специально предназначены

лишь для того, чтобы управлять результатами поиска, не добавляя ценный или уместный контент.

А теперь откроем вам секрет. Нам все равно, любят или ненавидят поисковые системы ту или иную специфическую тактику. Необходимо заботиться о стабильных результатах, потому что, если стратегия или система не в состоянии давать стабильный результат, тогда поисковики рано или поздно уничтожат сайт, который ее использует. Это весьма досадный факт. Почему? Потому что сайт закроется, возможно, даже навсегда.

В прошлом, если сайт был замечен в плохих делах, можно было вернуть незапятнанную репутацию. Но сегодня у поисковых машин нет причин делать поблажки "плохим парням". Поисковые машины, конечно, могут вернуть сайт в индекс, но на таком сайте будет стоять клеймо былых "заслуг", поэтому высокого места они или не займут, или пройдет довольно долгое время реабилитации.

Давайте разберем на примере Google, поскольку именно эта поисковая система имеет максимальные возможности для анализа. Сегодня Google имеет доступ к данным о PageRank, AdSense/AdWords, множеству персональных сервисов (Google Desktop, Gmail), а также информации из других источников типа Alexa, DMOZ, Yahoo Directory, Веб-архив (archive.org) и др.

Таким образом, они могут знать:

- полную историю вашего сайта;
- о любой странице, когда-либо индексированной их сервисом;
- ваш контент;
- любые "фишки", которыми вы пользовались на странице;
- любую страницу, на которую вы ссылались;
- любую страницу, которая ссылалась на вас;
- точную структуру ваших входящих и внутренних связей.

Google владеет историей, которую изменить, увы, нельзя. Так что, учтите это, когда будете использовать нелюбимые Google стратегии. "Прячьте" их тщательнее.

В последнее время выходят на арену и становятся популярными две стратегии поисковой оптимизации, о которых нам бы хотелось сейчас рассказать. Быть может, вы уже слышали о них: это линкбэйтинг и социальные сети.

"Линкбэйтинг" и "социальные сети" стали двумя наиболее широко распространенными специальными терминами в поисковой оптимизации и по вполне понятным причинам. Помимо удобства, перекрестных связей между участниками, они обеспечивают стратегию построения ссылок, которая одобрена поисковыми системами, и могут создать сотни настоящих ссылок за меньшее время, чем этот процесс отнимал ранее при использовании старых, плодящих бесполезные взаимные ссылки методов.

Начнем с линкбэйтинга. Что же это такое? Linkbaiting — "приманка" для ссылок, т. е. создание контента для сайта, который будет заставлять людей ссылаться на него. Данный метод раскрутки позволяет как бы управлять шумихой, которая создается вокруг вашего сайта. Это одна из самых быстрых и сильных стратегий из доступных на сегодняшний день, с помощью которой очень быстро можно созда-

вать множество ссылок, в то время как типы контента, которые хорошо работают для линкбэйтинга, меняются в зависимости от области его применения, например, использование юмора, страха, противоречий и т. п.

Создание хорошего контента, который способствует получению ссылок, — не новая идея. Существуют две основные причины, почему это может очень быстро повысить рейтинг сайта в поисковых системах. Во-первых, в свете возрастающей популярности социальных информационных сайтов их содержание "путешествует" намного быстрее и проще, чем прежде. Новый контент может принести более тысячи ссылок за день. Во-вторых, ссылки, принесенные линкбэйтингом, обычно являются естественными. Они следуют от реальных людей, связывающихся с вами по доброй воле — вирусный маркетинг в самой чистой форме!

Примерами социальных сетей могут быть:

- **http://www.digg.com**;
- **http://www.news2.ru**;
- **http://www.habrahabr.ru**;
- **http://www.toodoo.ru**.

Важный момент в процессе внесения в списки социальных сетей — продемонстрировать свой контент столь большому количеству людей, как это возможно, надеясь, что некоторые из них согласятся поставить на вас ссылку. Например, появление на главной странице такого гиганта социальных сетей, как сервис **www.digg.com**, даст вам значительное увеличение трафика, вплоть до 5—10 тыс. человек за первые несколько часов нахождения вашего сайта на первой странице Digg. Однако этот прирост трафика будет крайне недолговечным, и лишь немногие из этих людей вернутся снова на ваш сайт.

Так что забудьте о том, чтобы что-нибудь им продать. Вместо этого вам следует основываться на принципе, по которому чем больше людей увидят ваш сайт, чем больше он им понравится, тем больше вероятность, что кто-нибудь из них захочет провести ссылку к вам на сайт. Если 10 тыс. человек посетит ваш сайт и найдет ваш контент интересным, то маловероятно, но возможно, что хотя бы сто из них захотят поставить на вас ссылку.

Это означает для вас сто настоящих ссылок, о которых вам даже не придется просить. Все, что вы делаете в данном случае, — это создаете контент и демонстрируете его огромной группе людей, и некоторым из них он нравится настолько, что они ставят ссылку на него. Это как раз тот самый тип ссылок, которому поисковые системы отдают предпочтение, ранжируя страницы в результатах поиска.

Сегодня поисковая оптимизация и поисковый маркетинг похожи на приключенческий сериал (с элементами мыльной оперы). Как и в любой продолжительной саге, нужно разбираться в событиях, чтобы понимать, что происходит. Если что-то упустить, то можно безнадежно отстать.

Огромная проблема SEO/SEM в том, что слишком много противоречивой и устарелой информации может сильно навредить. Информацию нужно фильтровать. Наша работа как раз в том, чтобы применять такой фильтр и предоставлять вам только хорошую и полезную информацию. Для начала мы поможем вам завоевать высшие позиции, а затем — удержать их!

1.5.3. Будущее поискового маркетинга

Поисковые машины постоянно развиваются. В перспективе развития ПМ — возможность распознавания и индексирования ими не только HTML, но и графики, аудиоконтента, организация поисковых результатов в логические категории и т. д. Чтобы изнутри взглянуть на будущее поисковиков, зайдите на Google Labs (**http://labs.google.com**). Это зона, где Google представляет свои новые технологии.

Таким образом, поисковый маркетинг становится все более похожим на традиционный маркетинг. В целом, маркетинговый успех поисковиков в перспективе будет зависеть от трех составляющих: рекламы и аналитики, оптимизации сайта и контента и внестраничных факторов продвижения. Когда все сказано и сделано, вы получите ожидаемый приток посетителей, пусть даже ваши рейтинги низки. Ваш рейтинг может быть ниже, чем у конкурента, но процент посетителей вашего сайта, совершающих покупки у вас, может быть больше.

Глава 2

Основы поисковой оптимизации

В этой главе мы хотим поговорить об основных элементах, которые влияют на совместимость с поисковыми машинами. Технологии, описанные здесь, не вознесут вас на вершину рейтинга поисковых машин (вам нужны входящие ссылки для того, чтобы этого добиться), но они помогут вам избежать тех ошибок, которые вредят вашему рейтингу и затрудняют для спайдеров поисковых машин поиск и краулинг вашего сайта.

2.1. Создание веб-страниц с учетом пожеланий поисковых систем

Существуют два очень важных момента, которые следует знать до начала создания вашего сайта.

Во-первых, это качество оценки посещаемости вашего сайта. Поисковые машины зарабатывают деньги с помощью контекстной рекламы. Показывать рекламные объявления пользователям — это их модель коммерческой деятельности, и чем больше у них количество пользователей, тем больше денег они получают. Основной способ привлечения новых пользователей для поисковых машин — предоставление ими высокого качества поиска. Это означает, что если ваш сайт является наиболее полезным для клиентов в вашей сфере, то поисковые машины должны выводить его в лучших результатах поиска. И действительно, ведь средства к их существованию зависят от этого.

Во-вторых, следует запомнить, что поисковые машины — это компьютерные программы. Точнее, поисковые машины запускают программу, называемую спайдером, которая загружает ваши интернет-страницы, просматривает текст и ссылки на этих страницах и затем решает, что делать с вашей страницей, основываясь на этой информации. Этот процесс известен как краулинг. Поисковые машины — это компьютерные программы, которые занимаются краулингом веб-страниц.

И поскольку вы пользуетесь компьютером, то знаете, что компьютерные программы иногда выходят из строя, особенно если их перенапрягать. Вы, вероятно, замечали, что компьютер начинает работать медленнее, открывает слишком много окон или приложений и, в конце концов, может просто перестать работать. То же самое происходит и со спайдером поисковой машины. Если ваш сайт имеет сложное или запутанное местонахождение или если ссылки между вашими страницами

трудны для нахождения спайдером, то ваш сайт не будет качественно изучен поисковой машиной.

Это означает, что некоторые из ваших страниц не будут появляться в результатах поиска или будут появляться редко, что доставит вам много неприятностей, особенно если вы хотите, чтобы эти страницы были высоко оценены в поисковых машинах. Не забывайте, что в Интернете существуют миллиарды страниц, и поисковой машине необходимо проводить максимально эффективный поиск среди всех этих страниц при ограниченных ресурсах и времени, которыми она располагает. Ваша задача — сделать краулинг страницы настолько быстрым, насколько это возможно. Иначе вы рискуете быть незамеченными поисковыми машинами.

2.2. Написание грамотного HTML-кода

Для начала удостоверьтесь, что ваш HTML-код верен и хорошо построен. Качественный HTML-код не будет способствовать повышению вашего рейтинга, но нарушенный код может затруднить для поисковых машин краулинг вашего сайта. Убедитесь, что вы закрыли все соответствующие теги, и избегайте использования исходного кода, который может сбить с толку спайдера поисковой машины.

Когда вы собираетесь создавать веб-страницу, очень важно помнить, что спайдеры поисковых машин — это всего лишь компьютерные программы. Это значит, что они имеют ограничения, основанные на определенном наборе правил, и страницы, которые не соответствуют этим правилам, сами создают себе неудобства.

Фундаментальное правило для создания страниц, дружественных поисковым машинам, гласит: чем легче вы сделаете для спайдера поисковой машины чтение вашего кода, тем больше ваши шансы на успех.

В целом, это означает, что следует избегать двух видов HTML-кода:

- HTML, который устарел или работает только в определенном браузере (например, старый тег `<blink>`);
- HTML, который настолько нов, что еще не распознан большинством спайдеров (некоторые новые XML-теги попадают в эту категорию).

Лучший способ убедиться в том, что ваш HTML-код легок для обработки спайдером, — это использовать код, выведенный World Wide Web Consortium (W3C). Это может показаться сложным для выполнения, но на самом деле для этого требуется всего лишь проверить ваши страницы HTML-кода валидатором (validator), обратившись по адресу **http://validator.w3.org/** (рис. 2.1).

Если у вас не выявлено никаких ошибок, то вы можете быть уверены, что спайдеры смогут успешно обрабатывать ваши страницы без проблем (рис. 2.2).

Но опять-таки повторим: тестирование не означает, что страница будет занимать более высокое место в рейтинге поиска, это просто подтверждает, что ваш код не содержит ошибок, которые могли бы помешать поисковым машинам в обработке вашей страницы.