

Алексей Яковлев
Василий Ткачев

РАСКРУТКА САЙТОВ

Основы, секреты, трюки
5-е издание

Санкт-Петербург
«БХВ-Петербург»
2015

[>>>](http://kniga.biz.ua) Купить книгу на сайте kniga.biz.ua

УДК 004.738.5
ББК 32.973.202
Я47

Яковлев, А. А.

Я47 Раскрутка сайтов: основы, секреты, трюки. — 5-е изд., перераб. и доп. /
А. А. Яковлев, В. Г. Ткачев. — СПб.: БХВ-Петербург, 2015. — 352 с.: ил.
ISBN 978-5-9775-3568-7

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 и др., описаны основные элементы, влияющие на совместимость сайта с поисковыми машинами. Даны рекомендации по поиску и анализу ключевых слов, персонализации поиска, персонализации сайтов, написанию контента для сайта, технологии анализа сайтов конкурентов. Рассмотрены проблемы поискового спама и "черных" методов оптимизации и особенности продажи 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 пятом издании обновлены все главы, уточнена и исправлена информация, связанная со ссылочным ранжированием, внутренней оптимизацией и социальными факторами. Обновлена информация о сервисах для анализа сайтов. Добавлена новая глава "Ценообразование в SEO".

Для широкого круга читателей

УДК 004.738.5
ББК 32.973.202

Группа подготовки издания:

Главный редактор	<i>Екатерина Кондукова</i>
Зам. главного редактора	<i>Евгений Рыбаков</i>
Зав. редакцией	<i>Екатерина Капалыгина</i>
Редактор	<i>Владимир Красовский</i>
Компьютерная верстка	<i>Ольги Сергиенко</i>
Корректор	<i>Зинаида Дмитриева</i>
Оформление обложки	<i>Марины Дамбиевой</i>

Подписано в печать 01.12.14.
Формат 70×100^{1/16}. Печать офсетная. Усл. печ. л. 28,38.
Тираж 3000 экз. Заказ №
"БХВ-Петербург", 191036, Санкт-Петербург, Гончарная ул., 20.
Первая Академическая типография "Наука"
199034, Санкт-Петербург, 9 линия, 12/28

ISBN 978-5-9775-3568-7

© Яковлев А. А., Ткачев В. Г., 2015
© Оформление, издательство "БХВ-Петербург", 2015

Купить книгу на сайте kniga.biz.ua >>>

Оглавление

Введение	11
Глава 1. Общие сведения о поисковой оптимизации.....	13
1.1. Что такое поисковые машины и как они работают	13
1.2. Особенности поиска с точки зрения поисковой машины	17
1.3. "Анатомия" поисковой машины.....	19
1.4. Что такое SEO (Search Engine Optimization).....	23
1.5. Поисковая оптимизация: прошлое, настоящее и будущее	26
1.5.1. Прошлое поисковой оптимизации	26
Внестраничные факторы как новое решение	26
Издаваемые сообществами каталоги	27
Подъем Google	28
1.5.2. SEO и SEM (Search Engine Marketing) сегодня	28
Стабильная SEO против нестабильной	29
Популярные стратегии поисковой оптимизации	30
1.5.3. Будущее поискового маркетинга.....	32
1.6. Ценообразование в SEO	32
Глава 2. Основы поисковой оптимизации	39
2.1. Создание веб-страниц с учетом пожеланий поисковых систем	39
2.2. Написание грамотного HTML-кода	40
2.3. Верстка с помощью тегов <DIV> против табличной.....	42
2.4. Основные элементы, влияющие на совместимость с поисковыми машинами	43
2.5. Релевантность и факторы, влияющие на нее.....	44
Глава 3. Особенности некоторых поисковых систем.....	46
3.1. Особенности Google	46
3.1.1. Google PageRank	49
3.1.2. Управление индексацией Google.....	51
Атрибут <i>nofollow</i>	52
Метатег <i>nofollow</i>	54
Метатег <i>noindex</i>	54
Файл robots.txt.....	55

3.1.3. Сервис Google Webmaster Tools	56
3.1.4. Как оптимизировать сайт для Google	57
3.1.5. Проблема дублированного контента	59
3.1.6. Персонализированный поиск Google	61
3.1.7. Привлечение ссылок с блогов, не содержащих тег <i>nofollow</i>	61
3.2. Особенности Яндекса	63
3.2.1. Индекс цитирования	67
3.2.2. Быстрый робот	68
3.3. Как оптимизировать сайт для иностранных и иноязычных пользователей, которые находятся вне основного рынка	70
Глава 4. Поиск и анализ ключевых слов	73
4.1. Почему важно правильно подбирать ключевые слова	73
4.2. Основные интернет-сервисы для поиска и анализа ключевых слов	77
4.2.1. Яндекс	77
4.2.2. Rambler	80
4.2.3. Другие сервисы для подбора ключевых фраз	83
4.3. Пример подбора и анализа ключевых слов	86
4.4. Работа с трафиком	91
Глава 5. Оптимизация основного текста	94
5.1. Наполнение главной страницы	94
5.1.1. Нужен контент? Нет проблем!	94
5.2. Обогащение текста ключевыми словами	95
5.2.1. Наполнение страниц текстом	96
5.2.2. Оптимизация основного текста	97
5.3. Внедрение ключевых слов	99
5.4. Оптимизация PDF-документов	100
5.5. Создание идеальной веб-страницы	109
Глава 6. Описание основных элементов, влияющих на совместимость с ПМ	114
6.1. Заголовок веб-страницы	114
6.1.1. Общие сведения	114
6.1.2. Некоторые хитрости использования тегов <i><TITLE></i>	116
6.1.3. Практические советы при создании заголовков	117
6.2. Метатег описания	121
6.2.1. Назначение и особенности метаописания	121
6.2.2. Практические советы по написанию метаописания	123
6.2.3. Связанные теги <i><TITLE></i> и <i><META></i>	124
6.3. Метаключи	125
6.3.1. Общие сведения о метаключях	125
6.3.2. Практические советы по подбору метаключей	126
6.4. Основной текст	127
6.4.1. Плотность ключевых слов	127
6.4.2. Расположение текста на странице	128
6.4.3. Расположение ключевых слов (близость ключевых слов друг к другу — Adjacency & Proximity)	130
6.4.4. Заголовки <i>H1–H6</i>	131
6.4.5. Выделение текста полужирным	132

6.5. Доменное имя.....	132
6.6. Дополнительные возможности.....	134
6.6.1. Темы	134
6.6.2. Выделенный IP-адрес	134
6.6.3. Некоторые элементы веб-дизайна.....	135
6.6.4. Атрибут <i>alt</i> тега <i></i>	135
6.6.5. Текстовые ссылки.....	137
6.6.6. Размер файла.....	138
6.6.7. "Возраст" страницы.....	139
6.6.8. Скрытые теги форм	139
6.6.9. Внешний код	139
6.6.10. Каскадные таблицы стилей (CSS)	140
6.7. Динамически формируемый контент.....	141
6.7.1. Что такое динамически формируемое содержимое.....	141
6.7.2. Влияние динамического содержания на совместимость с ПМ	141
6.7.3. Будьте осторожны с ID-сессиями.....	142
6.7.4. Оптимизация динамических сайтов.....	143
6.7.5. Специальные приемы для сайтов с динамическим контентом.....	144
6.8. Структура каталогов.....	144
6.9. Элементы интерфейса	145
6.9.1. Графическая навигация (меню).....	145
6.9.2. Раскрывающиеся меню	145
6.9.3. Множественные точки входа.....	146
6.9.4. Карта сайта (sitemap).....	147
Карты сайта для больших сайтов	148
6.10. Flash- и splash-страницы.....	149
6.10.1. Общие замечания.....	149
6.10.2. Вводные, или splash-страницы	149
6.10.3. Как сделать Flash дружественным ПМ.....	150
6.10.4. Поддержка Flash поисковыми машинами	151
6.11. Спецэффекты	151
6.12. Тире, подчеркивание и другие символы в доменах и названиях файлов.....	152
6.13. Файл robots.txt и метатег <i>ROBOTS</i>	152
6.14. Поведенческие факторы и их влияние на поиск	154
Глава 7. Типы и функции ссылок.....	156
7.1. Динамические ссылки (JavaScript-ссылки).....	156
7.2. Графические ссылки.....	156
7.3. Текстовые ссылки.....	157
7.4. Входящие и исходящие ссылки.....	158
7.5. Взаимные ссылки.....	158
7.6. Треугольные ссылки.....	159
7.7. Ссылочная популярность	159
7.8. Узнайте разницу между естественной и искусственной структурами ссылок	161
Глава 8. Построение ссылок	164
8.1. Внешние факторы раскрутки сайта.....	164
8.2. Оценка качества ссылки.....	167

8.3. Регистрация сайта в каталогах.....	169
8.3.1. Общие сведения.....	169
8.3.2. Регистрация в Open Directory	172
8.3.3. Регистрация в Яндекс.Каталоге.....	172
8.4. Создание и ведение блогов	174
8.5. Написание и публикация статей.....	178
8.5.1. Общие рекомендации	178
8.5.2. Методы распространения статей.....	181
8.5.3. Каталоги статей.....	181
Финансовая сторона вопроса.....	182
Правила публикации статей.....	182
Безопасна ли реклама с помощью статей	182
Что делать с дублированием материалов	183
Рейтинги	183
Что нужно для публикации	184
Программное обеспечение для автоматической публикации статей	185
8.5.4. Соблазны грязных технологий	186
8.6. Википедия (Wikipedia)	187
8.7. Линкбэйттинг и социальные сети	188
8.7.1. Социальная сеть Digg.....	190
Разнообразие ценностей Digg.....	191
Диггеры не осуществляют диггинг для поисковой оптимизации	192
8.7.2. Twitter и его роль в построении ссылок	196
8.7.3. Facebook в роли ссылочного донора.....	196
8.8. Тематические форумы.....	196
8.9. Распространение пресс-релизов.....	197
8.10. Поиск и покупка заброшенных сайтов	198
8.11. Как покупать ссылки и не попасться	198
8.12. Покупка рекламы в электронных журналах и бюллетенях.....	199
8.13. Создание и дистрибуция программного обеспечения	200
8.14. Вирусный маркетинг	200
8.14.1. Принципы вирусного маркетинга	200
Определение вирусного маркетинга	201
Классический пример на Hotmail	201
Элементы стратегии вирусного маркетинга.....	201
Сравнение стратегий вирусного маркетинга по шести принципам	203
8.14.2. Современные приемы вирусного маркетинга	204
Поощряйте ссылки на ваш сайт.....	204
Предлагайте желаемые товары или услуги, чтобы распространить свое сообщение	205
8.14.3. Выводы	206
8.15. Конкурсы, тесты, спецопросы	207
8.16. Эффективный обмен ссылками	208
8.16.1. Доступность для поисковых машин.....	208
8.16.2. Осторожность при обмене ссылками.....	210
8.17. Другие советы по привлечению ссылок	211
8.18. Ссылки, которые не приносят пользы	212
8.18.1. Взаимные ссылки.....	212
8.18.2. Покупка ссылок	213

8.18.3. Регистрация в каталогах.....	213
8.18.4. Публикация статей и пресс-релизов.....	214
8.18.5. Правильный ссылочный баланс	214
Глава 9. Привлечение трафика и наращивание ссылочной базы с помощью Яндекс.Новости и Новости Google	216
9.1. Как воспользоваться сервисами Яндекс.Новости и Новости Google для раскрутки сайтов	216
9.2. Партнерская программа Новости Google	217
9.2.1. Какая организация страниц нужна для новостей Google	217
9.2.2. Как представить сайт на рассмотрение в Новости Google	220
9.2.3. Как создать карту сайта Новости Google.....	222
9.3. Партнерская программа Яндекс.Новости.....	222
9.3.1. Технология работы с Яндекс.Новостями	222
9.3.2. Как стать партнером Яндекс.Новостей.....	223
Глава 10. Как анализировать сайты конкурентов.....	226
10.1. Общие рекомендации	226
10.2. Анализ структуры ссылок	227
10.2.1. Сервис LinkPad (бывшее Solomono)	228
10.2.2. Сервис Ahrefs	228
10.3. Инструменты анализа, которые предоставляет Google.....	230
10.4. Анализ доменов ваших конкурентов	231
Глава 11. Как продавать через Интернет, или Секреты копирайтинга	234
11.1. Наиболее важные аспекты копирайтинга	234
11.2. Какой тип личности наиболее соответствует вашему потенциальному клиенту	239
11.2.1. Доминирование.....	239
11.2.2. Влияние	240
11.2.3. Стабильность.....	240
11.2.4. Исполнительность (добросовестность)	241
11.3. Различия в написании рекламных текстов для мужчин и женщин	243
11.3.1. Как покупают мужчины и женщины.....	243
11.3.2. Мужчины и женщины в Интернете.....	245
11.4. Писать, угождая каждому. Длинный текст против короткого.....	245
11.5. Заголовки, введение и заключение.....	247
11.5.1. Заголовки "Как..."	249
11.5.2. Заголовки, ориентированные на выгоды	250
11.5.3. Подзаголовки	250
11.6. Свойства против выгод. Вы действительно знаете различие?.....	252
Глава 12. Поисковый спам и "черные" методы оптимизации	259
12.1. Устаревшие спамовые технологии.....	263
12.1.1. Спаминг ключевых слов	264
12.1.2. Невидимый и "полувидимый" текст	265
12.1.3. Дорвей-страницы	267
12.1.4. Информационные страницы	270
12.1.5. Главное — содержание	270
12.1.6. Мета- и JavaScript-переадресация	272

12.1.7. Скрытие текста и ссылок на странице с помощью CSS	274
12.1.8. Тег комментария <code><!-- ... --></code>	274
12.1.9. Порядок цифр и букв в тексте	274
12.1.10. Секрет "невидимого" пиксела	275
12.1.11. Техника "Приманка и переключение"	276
12.1.12. Перерегистрация в поисковиках	276
12.2. Еще несколько спамовых технологий	277
12.2.1. IP Delivery, или клоакинг	277
12.2.2. DHTML-спаминг	279
12.2.3. Автоматически сгенерированный контент	280
12.2.4. Линк-фермы	280
12.2.5. Спаминг на основе <code><noframes></code> и <code><noscript></code>	280
12.3. Продвинутое спамовые технологии	281
12.3.1. Публицистические империи (мини-сети)	281
12.3.2. "Перелинковка" и создание мини-сетей	282
Глава 13. Черный пиар в Интернете и методы защиты	289
13.1. Работа с негативными поисковыми результатами	290
13.2. Девять лучших стратегий по предотвращению негативных результатов	291
13.3. Небезопасное SEO	295
13.3.1. Отслеживайте вашу онлайн-репутацию	296
13.3.2. Онлайн-репутация — это все!	297
Глава 14. Эффективность интернет-маркетинга для B2B-отраслей	298
14.1. Отличия в раскрутке сайтов B2B- и B2C-компаний	298
14.2. Проблемы интернет-маркетинга, которые упоминают представители B2B-отраслей	299
14.3. Различия в целях поисковой оптимизации	301
14.4. Копирайтинг для B2B	302
14.4.1. Один пользователь, много поисковых запросов	303
14.4.2. Один договор, много пользователей	304
14.5. Стратегия ключевых фраз для B2B	304
14.6. Под какие поисковики лучше оптимизировать сайт	305
14.7. Инструменты для повышения конверсий в B2B-отраслях	305
14.8. Электронные торговые площадки	307
14.9. Выводы	308
Глава 15. Отслеживание результатов и оценка успеха	309
15.1. Введение в веб-аналитику	309
15.2. Современные возможности и тенденции в веб-аналитике	311
15.3. Измерение поискового трафика	314
15.3.1. Общий обзор систем веб-анализа	314
15.3.2. Ценные данные SEO в веб-аналитике	316
Трафик от поисковых машин	316
Трафик от ключевых слов	317
15.3.3. Реферальные сайты	318
15.4. Взаимосвязь между SEO, конверсиями и ROI	319
15.4.1. Стратегия отслеживания конверсий	320
15.4.2. Стратегия повышения конверсий	320

15.5. Сегментирование SEO-кампаний согласно стоимости конверсий.....	322
15.6. Ключевые показатели эффективности для SEO по длинным целевым запросам	322
15.6.1. Соотношение брендовых и небрендовых запросов (использование бренда в ключевых словах).....	322
15.6.2. Сканирование уникальных URL.....	323
15.6.3. Соотношение количества посетителей и страниц, приносящих пользу от поисковых систем	323
15.6.4. Соотношение числа ключевых слов и страниц.....	323
15.7. Вывод.....	324

Глава 16. Сводная информация о поисковой оптимизации

(SEO Checklist)	325
16.1. Ключевые слова	325
16.2. Доменное имя и имена файлов.....	326
16.3. Расположение ключевых слов	327
16.4. Внутренние ссылки.....	328
16.5. Входящие ссылки	328
16.6. Исходящие ссылки	329
16.7. Внутренние факторы	330
16.8. Внешние факторы.....	332
16.9. IP-адрес сервера.....	333
16.10. Накопление ссылок.....	333
16.11. Что можно и что нельзя на сайте.....	334

Заключение.....	337
------------------------	------------

Приложение. Услуги и обучение в области интернет-маркетинга	339
П1. Услуги в области интернет-маркетинга	339
П2. Презентация услуг компании SEM Complex в офисе вашей компании	340
П3. Обучение интернет-маркетингу.....	341
П4. Корпоративные семинары.....	342
П5. Обучение и продвижение от профессионалов.....	343

Глоссарий.....	344
-----------------------	------------

Предметный указатель	347
-----------------------------------	------------

Введение

Это уже пятое издание книги, серьезно переработанное и дополненное. Рынок интернет-продвижения развивается очень динамично: алгоритмы работы поисковых систем постоянно совершенствуются, конкуренция в Интернете ежедневно усиливается, пожелания пользователей Интернета регулярно меняются.

Давно закончилось то время, когда владельцы сайтов гнались за поднятием тИЦ Яндекса, Google PageRank, CTR и других терминов, понятных лишь специалистам. Сейчас настали другие времена, во всяком случае, мы это видим по нашим заказчикам. В текущей ситуации пришло время думать стратегическими категориями, особенно в интернет-маркетинге. Все хотят реальных результатов. Все чаще приходится слышать фразы: "Какую долю поискового трафика я получу, если инвестирую в Интернет 1 000 000 руб.?", "Сколько у меня будет целевых посетителей на сайте через X месяцев работы?", "Сколько мне нужно заплатить, чтобы у меня было Y продаж?" и т. п.

Для ответа на эти вопросы понадобится стратегическое мышление, необходимое для прогнозирования результатов. Как спрогнозировать, сколько трафика придет на наш сайт через 6–12 месяцев, какую сумму необходимо инвестировать с целью захвата X доли поискового трафика. Ответы на эти вопросы лежат в плоскости комплексного интернет-маркетинга.

В современных условиях нужно понимать, что раскрутка сайта — это лишь одно из направлений. Если в 2006 г., когда выходило первое издание этой книги, раскрутка была полностью самодостаточной, то сейчас это лишь капля в море. Технологии интернет-маркетинга сильно изменились за эти годы и продолжают меняться галопирующими темпами. В настоящее время у нас вышло уже три книги по интернет-маркетингу: "Раскрутка сайтов: основы, секреты, трюки", "Контекстная реклама: основы, секреты, трюки" и "Веб-аналитика: основы, секреты, трюки". Сейчас мы готовим обобщающую книгу для менеджеров и руководителей, в которой будет разобрана вся цепочка задач, необходимых для успешного ведения бизнеса в Интернете, — от построения стратегий и прогнозирования до анализа эффективности конкретных используемых инструментов интернет-маркетинга с целью дальнейшей оптимизации стратегии интернет-присутствия сайта в Сети.

В пятом издании данной книги мы постарались изложить все нюансы раскрутки веб-сайта — от "белых" методов, которые уже зарекомендовали себя как надежные

[>>>](http://kniga.biz.ua)

и будут работать всегда, до "серых" и откровенно "черных" методов продвижения — чтобы вы могли четко понимать, что делать можно и нужно, что не стоит и какие штрафные санкции можно за это получить.

Книга будет полезна тем, кто хочет заняться продвижением своего сайта либо зарабатывать на продвижении чужих веб-ресурсов, а также тем, кто отдал эту работу на аутсорсинг в SEO-компанию.

В данной книге вы ознакомитесь со всеми базовыми моментами, с которыми сталкивается специалист по поисковой оптимизации и которые необходимы для успешной раскрутки как русскоязычных, так и англоязычных веб-ресурсов.

Некоторые примеры мы будем приводить из англоязычной части Интернета, т. к. российских аналогов пока не существует (но они очень скоро появятся). В книге приводится огромное количество реальных примеров из жизни оптимизаторов, которые помогут вам лучше разобраться в этом нелегком деле.

Вы обнаружите, что информация, представленная в книге, даст вам огромное тактическое превосходство в соревновании за топовые позиции. Воспользуйтесь этой информацией для того, чтобы сфокусировать свои усилия там, где вы получите наибольший отклик. Если сказать коротко: вам следует работать разумно, а не тяжело.

Мы старались приводить наиболее свежую и эффективную информацию, которая только возможна. Однако Интернет — это динамичная и постоянно изменяющаяся среда, и то, что работает сегодня, может не работать уже завтра.

Тем не менее мы уверены, что методы, которые здесь описаны, работают сейчас и будут работать и в будущем. Однако очень важно, чтобы во время изучения материала вы научились самостоятельно анализировать используемые методики.

До начала чтения этой книги очень рекомендуем вам ознакомиться со статьей "Как изучить SEO", опубликованной на портале SEONews.ru (<http://www.seonews.ru/columns/kak-izuchit-seo-obuchenie-seo-optimizatsii/>).

Если во время изучения книги у вас возникнут какие-либо вопросы, вы всегда можете обратиться в службу поддержки учебного центра **www.seo-study.ru** по e-mail **customer@seo-study.ru** или по телефонам: в Санкт-Петербурге — +7(812) 332-53-06, в Москве — +7(499) 500-38-22, из регионов РФ — +7(800) 555-10-98. Либо можете задать их нам лично на основных тематических конференциях, например ConversionConf (www.conversionconf.ru) и AllinTopConf (www.allintopconf.ru).



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.

В поисковые системы входят пять отдельных программных компонентов.

- ❑ *Страница выдачи результатов* (Search Engine Results Page, SERP) извлекает результаты поиска из базы данных.
- ❑ *Паук* (спайдер, англ. *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 в своем браузере команду **Просмотр HTML-кода**.
- ❑ *Краулер* (от англ. *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* (от англ. *indexer*) разбирает страницу на 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* (англ. *database*) — хранилище всех данных, которые поисковая система скачивает и анализирует. 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 определяются 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) — пополнением 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, сортирующие найденные результаты) каждого из поисковых сервисов являются уникальными. Вы можете проверить это сами: введите ключевое слово или фразу в поисковую систему Яндекс (www.yandex.ru) и запомните результаты. Теперь зайдите на Google (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 Поиск@Mail.Ru (go.mail.ru) до середины 2013 г. использовал алгоритм Google, подмешивая свои алгоритмы; многие зарубежные поисковые системы используют общие алгоритмы.

Google является самой популярной и самой крупной по базе данных поисковой машиной в мире. 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, то 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. 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 *ассесоры* (от англ. *assessor*). Ассесоры работают по заданной методике. Существуют определенные критерии, по которым должно измеряться качество страницы. Ассесор вводит поисковый запрос и ставит определенную оценку, насколько по этим критериям релевантными будут 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 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 описанного выше, поисковикам очень трудно определить частоту изменений страниц. Робот 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 роботом 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 predetermined 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.

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



Рис. 1.1. Сохраненная копия страницы в поисковой системе Яндекс

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, Text REtrieval Conference) утвердила точку отсчета для такого большого объема (собрания писем, тексты и т. д.) информации, как 100 Гбайт. (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 300000 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. 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 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, чего 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 эта книга есть, то он или найдет ее, или закажет. 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 полку библиотеки в алфавитном порядке, а карточка с индексом книги занесена 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 некоей особенной книге или теме, библиотекарь может иногда интуитивно сказать, где находится данная книга, или, по крайней мере, указать секцию. Регистрируя выбор и возврат книг, библиотекарь 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, пытаются подражать этому принципу, весьма обширны и серьезны.

Можно привести пример школьника, который приходит в библиотеку и спрашивает книгу об Италии. Библиотекарь, обладая довольно 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 с помощью "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

