

# SEO Леонид Гроховский

Михаил Сливинский,  
Алексей Чекушин, Станислав Ставский

## руководство по внутренним факторам

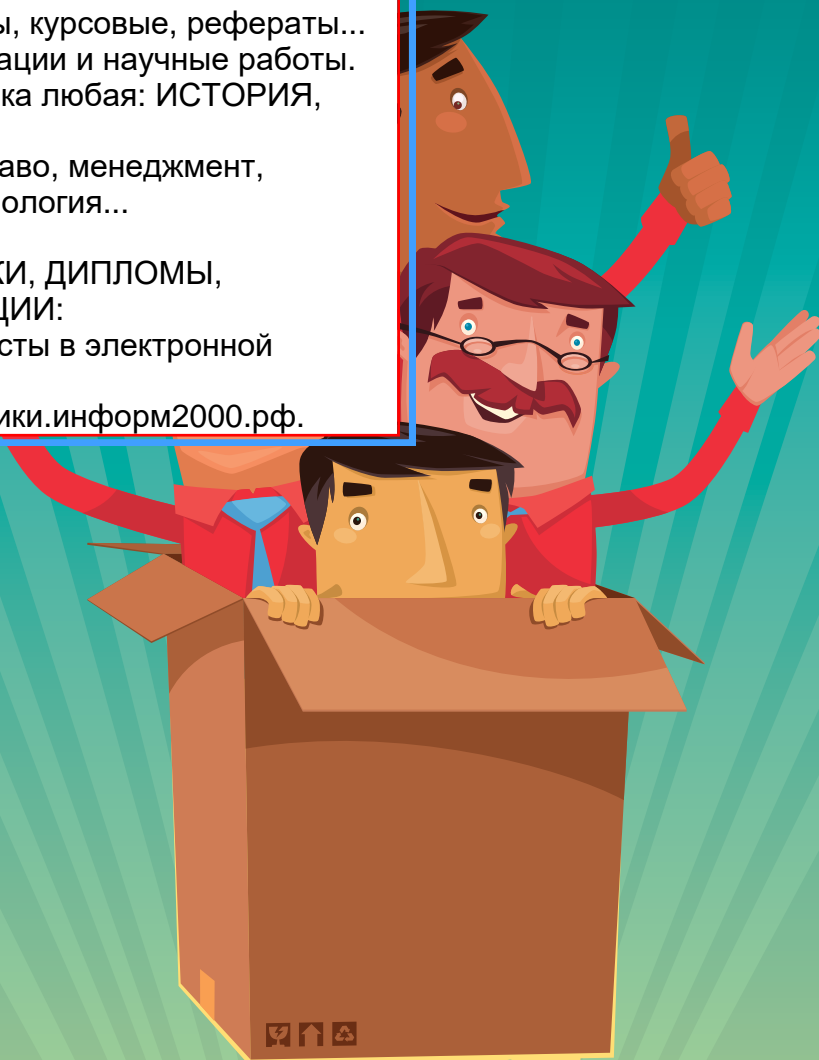
НАПИСАНИЕ на ЗАКАЗ:

1. Дипломы, курсовые, рефераты...
2. Диссертации и научные работы.

Тематика любая: ИСТОРИЯ,  
экономика,  
техника, право, менеджмент,  
финансы, биология...

УЧЕБНИКИ, ДИПЛОМЫ,  
ДИССЕРТАЦИИ:  
полные тексты в электронной  
библиотеке

[www.учебники.информ2000.пф](http://www.учебники.информ2000.пф).



Вернуться в библиотек учебников

Вернуться в каталог учебников  
<http://учебники.информ2000.пф/uchebniki.shtml>

Разработка и продвижение сайтов  
<http://сайт.информ2000.рф/index.htm>

Вернуться в каталог учебников  
<http://учебники.информ2000.рф/uchebniki.shtml>

Уникальная подборка информации по эффективности корпоративных сайтов:

- для самообразования топ-менеджеров;
- для повышения квалификации преподавателей;
- для рефератов и контрольных.

Рекламируйте бизнес через собственный сайт

Рейтинг дипломных и курсовых работ

Научу дистанционно создавать эффективные сайты

Уникальная информация по менеджменту и экономике

# SEO

## Леонид Гроховский

Михаил Сливинский,

Алексей Чекушин, Станислав Ставский

### руководство по внутренним факторам

Леонид Гроховский, Михаил Сливинский, Алексей Чекушин, Станислав Ставский  
SEO: руководство по внутренним факторам. – М.: Центр исследований и образования «ТопЭксперт.РФ», 2011. – 133 с.

Эта книга кардинально отличается от аналогичных изданий: она написана оптимизаторами для оптимизаторов и является первым изданием по теме внутренних факторов в SEO. Авторы соединили свои разноплановые познания в области внутренней оптимизации и смогли простым языком изложить сложный материал. Теперь вы можете воспользоваться актуальной информацией и применить новые знания в своей практике, наиболее полно задействовав потенциал собственного сайта и улучшив его позиции в выдаче поисковых систем. Книга написана в соавторстве с лучшими аналитиками нашей профессии – Станиславом Ставским, Михаилом Сливинским и Алексеем Чекушиным. Их участие сделало книгу полезной не только для начинающих, но и для продвинутых специалистов.

© Центр исследований и образования «ТопЭксперт.РФ»

© Леонид Гроховский, Михаил Сливинский,  
Алексей Чекушин, Станислав Ставский

Вернуться в каталог учебников  
<http://учебники.информ2000.рф/uchebniki.shtml>

## Об авторе и соавторах



**Л**еонид Гроховский, В SEO с 2005 года, в настоящий момент директор по технологии продвижения и автоматизации компании «Оптимизм.ру», входящей в топ-10 SEO-компаний. Опыт продвижения – более 300 сайтов. Область

деятельности: автоматизация, аналитика, проведение исследований. Руководитель центра образования и исследований «ТопЭксперт.РФ», автор ряда публикаций в профессиональных изданиях, постоянный докладчик на отраслевых конференциях. Руководитель разработки аналитической системы «СайтРепорт». Автор исследования «SEO-альманах 2010». Специализируется на продвижении порталов и крупных интернет-магазинов. Продвигал и консультировал такие проекты, как [championat.ru](http://championat.ru), [woman.ru](http://woman.ru), [rabota.ru](http://rabota.ru), [ucheba.ru](http://ucheba.ru), [pressfoto.ru](http://pressfoto.ru), [aforex.ru](http://aforex.ru), [best-realty.ru](http://best-realty.ru), [panasonic.ru](http://panasonic.ru). Преподает SEO в Академии интернет-рекламы и в учебном центре «Специалист» при МГТУ им. Н. Э. Баумана. Ведущий программы «ТопЭксперт.РФ» на интернет-телевидении MEGAINDEX.TV.

Автор и преподаватель курсов:

- ▶ «Основные ошибки в SEO от создания сайта до покупки ссылок»
- ▶ «Пошаговая реализация SEO-стратегии для маркетологов и руководителей»

- ▶ «Оптимизация сайтов: задействуем внутренние факторы»
- ▶ «Источники привлечения внешних ссылок»
- ▶ «Юзабилити, аналитика и Яндекс.Метрика для SEO-специалиста»
- ▶ «Комплексная стратегия продвижения»
- ▶ «Автоматизация в SEO»
- ▶ «Продвижение порталов и крупных интернет-магазинов»
- ▶ «Высший пилотаж в SEO»
- ▶ «Полный курс SEO – 130 часов»

### СОАВТОРЫ



**Михаил Сливинский**, руководитель отдела аналитики, [Wikimart.ru](http://Wikimart.ru); преподаватель учебного центра «ТопЭксперт.РФ»



**Алексей Чекушин**, руководитель департамента продвижения, [Wikimart.ru](http://Wikimart.ru); преподаватель учебного центра «ТопЭксперт.РФ»



**Станислав Ставский**, руководитель направления зарубежных проектов, [Sare.ru](http://Sare.ru); ex. аналитик веб-поиска Яндекса; преподаватель учебного центра «ТопЭксперт.РФ»

# Содержание

---

<b>Предисловие</b> .....	<b>6</b>
<b>Введение в продвижение сайтов</b> .....	<b>11</b>
0.1. Понятие о внутренних факторах .....	12
0.2. Понятие о внешних факторах .....	14
0.3. Понятие о поведенческих факторах .....	15
0.4. Комплексный подход к продвижению сайтов .....	16
<b>Глава 1. Перелинковка страниц</b> .....	<b>23</b>
1.1. Как вычисляется PageRank и почему это так важно знать .....	24
• По материалам статьи в переводе Александра Садовского «Растолкованный PageRank» .....	25
1.2. Базовые схемы перелинковки .....	25
• Комплексное продвижение всего сайта .....	26
• Продвижение главной страницы под ВЧ .....	26
• Продвижение разделов под СЧ .....	27
• Продвижение внутренних страниц под НЧ .....	29
1.3. Локальные схемы перелинковки .....	30
• Кольцо .....	30
• Куб .....	30
• Звезда .....	31
1.4. Правила перелинковки .....	31
1.5. Способы перелинковки .....	34
1.6. Составление текста ссылки для перелинковки .....	38
1.7. Создание модуля для перелинковки .....	40
• Портал .....	40
• Интернет-магазин .....	41
• Корпоративный сайт .....	42
• Каталог .....	42
<b>Глава 2. Устранение дубликатов</b> .....	<b>43</b>
2.1. Чем вредны дубликаты .....	43
2.2. Классификация дубликатов .....	45
2.3. Кейсы .....	49
• Компания «БЕСТ-Недвижимость», <a href="http://www.best-realty.ru">www.best-realty.ru</a> .....	49
• ММЦ «ОнКлиник», <a href="http://www.onclinic.ru">www.onclinic.ru</a> .....	51
• Магазин итальянской мебели, <a href="http://www.eurointerier.ru">www.eurointerier.ru</a> .....	52
2.4. Способы устранения дубликатов .....	54

<b>Глава 3. Инструкции для поисковых систем</b> .....	<b>57</b>
3.1. Зачем нужен robots.txt. ....	57
3.2. Robots.txt – универсальные директивы .....	58
3.3. Robots.txt – директивы для Яндекса .....	59
3.4. Правила составления robots.txt .....	60
3.5. Разбор robots.txt на примерах .....	61
3.6. Зачем нужен sitemap.xml .....	63
3.7. Директивы sitemap.xml .....	63
3.8. Правила составления sitemap.xml .....	66
3.9. Пример sitemap.xml .....	66
3.10. Принцип взаимодействия sitemap.xml и robots.txt .....	69
3.11. Автоматизация добавления страниц в sitemap.xml .....	69
<b>Глава 4. Работа с авторским контентом</b> .....	<b>71</b>
4.1. Уникальность контента .....	71
4.2. Как защитить контент от воровства? .....	74
4.3. Как воровать контент? .....	76
4.4. Отсутствие и недостаток контента .....	76
• Разбор решения на примере <a href="http://www.mobiguru.ru">www.mobiguru.ru</a> .....	76
4.5. Автоматизация наполнения Description, Keywords, Title .....	77
4.6. Автоматизация генерации контента .....	79
• Автоматизация генерации описаний товаров .....	79
• Автоматизация генерации технических характеристик (уровень 1) .....	79
• Автоматизация генерации технических характеристик (уровень 2) .....	80
4.7. Обнаружение дубликатов .....	80
• Инструкция по обнаружению дубликатов вручную .....	81
• Способы автоматизации обнаружения дубликатов .....	82
<b>Глава 5. Коды ответов сервера</b> .....	<b>85</b>
5.1. Код ответа 200 .....	85
5.2. Коды ответов 301/302/303/307/3** .....	86
5.3. Код ответа 404 .....	87
<b>Глава 6. Стандарты W3C</b> .....	<b>89</b>
6.1. Правила HTML-разметки для SEO .....	91
6.2. Проверка верстки на соответствие стандартам W3C .....	95
<b>Глава 7. SEO-аналитика</b> .....	<b>97</b>
<b>Глава 8. Исследование «SEO-альманах 2010»</b> .....	<b>107</b>
<b>Семинары и курсы</b> .....	<b>117</b>
<b>DVD-семинары и курсы</b> .....	<b>125</b>

## Предисловие

---

**П**риветствую! Меня зовут Леонид Гроховский. В настоящий момент я работаю директором по технологии продвижения и автоматизации в компании «Оптимизм.ру». Моя область деятельности – исследование поисковых алгоритмов и создание технологии продвижения. Также я занимаюсь разработкой стратегии продвижения ключевых клиентов компании, т. е. порталов, крупных интернет-магазинов и сайтов – лидеров тематики. В SEO с 2005 года.

В этой книге я стремился собрать воедино свои познания в области внутренней оптимизации и доступным языком изложить этот материал, чтобы вы смогли применить почерпнутую здесь информацию в своей практике. Мне хочется, чтобы в результате вы смогли задействовать потенциал собственного сайта и улучшили его позиции в выдаче поисковых систем. Кроме того, мне удалось привлечь лучших аналитиков нашей профессии – Станислава Ставского, Михаила Сливинского и Алексея Чекушина. Они стали соавторами этой книги и дали комментарии по вопросам PRO-уровня, сделав книгу интересной не только для начинающих, но и для продвинутых специалистов.

Следует отметить, что книга, которую вы держите в руках, кардинально отличается от аналогичных изданий. Она написана оптимизатором для оптимизаторов. Я писал не об абстрактном поисковом продвижении, а сосредоточился на практических рекомендациях, поэтому здесь органично сочетаются краткие фундаментальные



### Рецензия на книгу Леонида Гроховского «SEO: руководство по внутренним факторам»

Книга Леонида очень необычна! Прежде всего, тем, что в ней много практики и не так много теории. Немногие авторы трудов, посвященных поисковой оптимизации, раскрывают практические методы работы. Это происходит из-за того, что мир SEO быстро меняется. И пока написанное выходит в свет, большая часть знаний уже устаревает.

В данной книге сделана, на мой взгляд, достаточно удачная попытка раскрыть именно практические инструменты оптимизации сайтов. Достоинство ее и в том, что она узкоспециализированная. Книга целиком и полностью посвящена только внутренним факторам оптимизации сайта: структуре, контенту, перелинковке, написанию файла robots.txt и другим аспектам работы с сайтом.

Порадовало меня и большое количество реальных примеров из практики автора, наглядно иллюстрированных.

Издание, несомненно, будет полезно всем специалистам по оптимизации для закрепления уже имеющихся знаний, а также начинающим оптимизаторам для приобретения новых знаний по SEO.

С уважением, Иван СЕВСТЬЯНОВ,  
генеральный директор компании «ВебПроекты»

понятия с многочисленными конкретными советами, которые являются остроактуальными сегодня и останутся таковыми завтра. Это первое издание по данной теме. Вероятно, очень скоро нам придется столкнуться с новыми исследованиями и новыми факторами ранжирования, изменится значимость этих факторов и способы воздействия на них. Рад избавить вас от лишнего беспокойства.

Технологически работы по продвижению сайта делятся на два этапа: оптимизация и продвижение. В этой книге мы поговорим о первом этапе, который включает в себя все работы, проводимые внутри сайта: это работа с текстом, ссылками, настройками сервера, HTML-версткой, а также создание инструкций для поисковых систем. Здесь также описаны способы автоматизации генерации текстов, заголовков и метатегов, перелинковки. Эта информация пригодится специалистам, продвигающим крупные порталы и интернет-магазины.

Продвигая сайт, я всегда стараюсь максимально использовать потенциал внутренних факторов. Мой подход к продвижению тесно связан с автоматизацией, аналитикой и проведением исследований. На то есть ряд причин.

В 2010 году количество конференций, посвященных SEO, увеличилось в два раза. Очевиден и рост интереса к образовательным мероприятиям: чаще проводятся семинары и курсы, открываются новые центры обучения. SEO-бум – закономерное следствие того, что технологии продвижения усложнились и ценность информации очень сильно возросла. Рядовым SEO-специалистам, чтобы не выпасть из обоймы, необходимо постоянно повышать свою квалификацию, консультируясь у более опытных коллег.

Особая прослойка экспертов, имеющих практическое подтверждение высокого

уровня их знаний и обладающих ценной информацией, в SEO-сообществе уже сформировалась. В их числе – руководители лидирующих компаний, разработчики аналитических программ и систем автоматизации продвижения, высококвалифицированные SEO-профессионалы, успешно работающие с самыми конкурентными тематиками. Под моим руководством в конце 2010 года группа таких экспертов провела исследование факторов ранжирования, которое получило название «**SEO-альманах 2010**». В ходе исследования было рассмотрено более 200 факторов, сгруппированных в блоки. Нами применялась методология экспертных оценок, были привлечены 26 ведущих SEO-экспертов. Исследование получило широкий резонанс в SEO-кругах и за их пределами. Результаты проделанной работы, касающиеся внутренних факторов, изложены в приложении к этой книге. Комментарии некоторых экспертов вы будете встречать и в процессе чтения основной части.

Вообще, говоря о тенденциях поискового продвижения сегодня, можно сделать следующие выводы:

### 1. Снижается значение покупных ссылок

«Священная война» с автоматическим продвижением набирает обороты, и очевидно, что успех сейчас на стороне Яндекса. Можно уверенно говорить о том, что степень влияния покупных ссылок на ранжирование сведена к минимуму. Стоит ожидать еще большего ужесточения, так что биржам ссылок уже сейчас необходимо становиться на путь повышения качества площадок. Впрочем, такая ситуация не стала для кого-то громом среди ясного неба: представители Яндекса не единожды предрекали «смерть» платных ссылок.

### 2. Возрастает значение внутренних факторов

Факторы, которые поддаются влиянию оптимизаторов, стали гораздо менее

## ОТЗЫВ КЛИЕНТА

Одной из первых задач, с которой я столкнулся после того, как возглавил отдел маркетинга в компании Adrenalin Forex, была задача по поисковой оптимизации ее сайта. К тому моменту компания уже успела поработать по данной задаче с несколькими крупными SEO-агентствами и, полностью разочаровавшись в нулевых результатах, перешла на самостоятельное SEO-продвижение. К сожалению, предыдущий специалист так и не смог добиться высоких показателей, совмещая работу по продвижению с целым перечнем иных задач, соответствующих его компетенции и должности в компании.

Таким образом, я пришел к выводу, что требуется поиск нового партнера по продвижению сайта компании с четко обозначенными планами и обязательствами по их выполнению. После первой встречи с Леонидом появилось ощущение, что это тот человек, который обладает всем необходимым опытом для решения данной задачи. Тем не менее, учитывая предыдущий опыт компании, прошло несколько этапов согласования договора по оптимизации, где финансовые взаиморасчеты были привязаны к конкретным результатам работы и срокам их достижения.

После подписания всех документов и начала нашей совместной деятельности агентством было подготовлено ТЗ, где было описано, какие тексты потребуются написать и как они должны размещаться на сайте для эффективного его продвижения. По условиям договора, текст готовился копирайтерами агентства, но здесь мы столкнулись с проблемой восприятия «оптимизаторских» текстов посетителями сайта. Говоря проще, в компании изначально были заданы особые стандарты к содержанию сайта, а именно требования по простоте, лаконичности и доступности материалов. Поэтому те тексты, которые написал копирайтер агентства, мы однозначно не могли позволить разместить на сайте компании. Требовалось подготовить такой текст, чтобы, если кому-то вдруг и захотелось бы его прочитать, у него не возникло бы даже подозрений, что текст написан не для него, а для поисковых роботов. В результате кропотливого труда эти тексты удалось подготовить собственными силами. Итогом мы гордимся и по сей день, ведь написать понятный, простой и осмысленный текст с постоянно повторяющимися словами в строго заданном формате – весьма непростая задача.

Не прошло и пары месяцев, как мы поняли, что проделали эту сложную работу совсем не зря, наблюдая, как наши позиции в поисковиках стремительно стали подниматься. А уже спустя полгода они по самым конкурентным запросам Forex и «Форекс» вышли на 2 место, что на первых этапах работы казалось практически невозможным.



Продвигать сайты стало сложнее. Да, было проделано много работы по разгрузке оптимизаторского труда: создавались ссылочные агрегаторы, переводящие продвижение в автоматический режим. И что же мы имеем сегодня? Необходимо буквально вручную прорабатывать внутренние факторы и напрямую договариваться о размещении ссылок на качественных тематических площадках. Сроки продвижения растягиваются. Конкуренция растет.

значимыми. Те же, на которые повлиять очень сложно, а порой и нереально, вышли на первый план. Что ж, против лома нет приема, а на любое действие противодействие найдется: оптимизаторам необходимо повышать квалификацию, работать над своим подходом к продвижению и сосредотачивать внимание на внутренних факторах.

### **3. Каждая тематика может иметь свои особенности продвижения**

Введение алгоритма машинного обучения «Матрикснет» не прошло бесследно: сейчас каждая конкретная тематика имеет свою специфику продвижения, и специфика эта может зависеть еще и от типа запросов. Например, подходы к продвижению однословного запроса и запроса более конкретного могут отличаться коренным образом. Технология продвижения многих запросов теперь еще и напрямую зависит от географической принадлежности сайта.

Резюмируя все вышесказанное, подчеркнем: продвигать сайты стало сложнее. Да, было проделано много работы по разгрузке оптимизаторского труда: создавались ссылочные агрегаторы, переводящие продвижение в автоматический режим. И что же мы имеем сегодня? Необходимо буквально вручную прорабатывать внутренние факторы и напрямую договариваться о размещении ссылок на качественных тематических площадках. Сроки продвижения растягиваются. Конкуренция растет. Утешительным моментом в сложившейся ситуации является между тем рост покупательской аудитории. К примеру, количество запросов «ноутбук», по данным Яндекс.Вордстат, за год выросло с 1 589 943 (ноябрь 2009) до 2 591 179 (ноябрь 2010). Разумеется, конкретные показатели зависят от тематики, но в целом тенденция роста очевидна.

По мере развития поисковых систем количество факторов ранжирования

увеличивается. Разумно предположить, что в таких условиях достичь положительного результата можно быстрее, если воздействовать на все возможные факторы, пусть даже некоторые кажутся пустяковыми. Помните о том, что многие из них, накладываясь друг на друга, при оценке влияния на ранжирование считаются с кумулятивным эффектом. Согласитесь, лучше сделать все, что в ваших силах, и видеть результат стараний, чем что-то упустить и размышлять, что же конкретно. Именно этому принципу я следую в своей деятельности. Для анализа потенциала внутренних факторов под моим руководством разработана аналитическая система «СайтРепорт». С ее помощью можно воздействовать практически на все факторы, описанные в этой книге. Подробно о возможностях «СайтРепорт» вы прочтете в приложении. Там же вы найдете информацию об акциях и скидочных купонах на мои семинары и консалтинговые услуги.

Наша книга поможет вам удержаться на плаву, адаптироваться к изменившимся условиям и переориентировать свою стратегию на внутренние факторы. Если вы решите углубить свои знания еще больше и информации в этой книге для вас окажется недостаточно, вы можете получить подкрепление почерпнутым знаниям на одноименном платном семинаре «**Оптимизация сайтов: задействуем внутренние факторы**», который проводится раз в год. На семинаре вы поучаствуете в разборе реальных сайтов слушателей, зададите интересующие вас вопросы и получите особые бонусы.



Наша книга поможет вам удержаться на плаву, адаптироваться к изменившимся условиям и переориентировать свою стратегию на внутренние факторы.

# Введение в продвижение сайтов



**В**ы можете смело пропустить эту часть, если считаете себя опытным оптимизатором. Информация, представленная здесь, будет полезна тем, кто только начинает свою деятельность на поприще продвижения сайтов.

Поисковое продвижение сайтов – это комплекс работ, направленных на улучшение позиций веб-ресурса в выдаче поисковых систем. Цель этих мероприятий – извлечение коммерческой выгоды за счет привлечения на сайт посетителей из поисковых систем.

Существуют 3 способа продвижения сайтов:

- ▶ воздействие на внутренние факторы;
- ▶ воздействие на внешние факторы;
- ▶ воздействие на поведенческие факторы.

Также продвижение сайтов часто делят по типам факторов, на которые производится влияние со стороны оптимизатора.

**Внутренние факторы** – это все работы, проводимые над сайтом, а также домен и серверные настройки.

**Внешние факторы** – это использование рекомендательных ссылок, размещаемых на внешних ресурсах.

**Поведенческие факторы** – статистическая информация о посетителях сайта, собираемая поисковыми системами.

Внутренние и внешние факторы подразделяются на динамические и статичные. Первые учитывают статистику, накопленную за некоторый временной промежуток, например, возраст сайта, динамику

прироста рекомендательных ссылок. Статические факторы учитывают текущие значения.

Для эффективного продвижения необходимо использовать все три вышеперечисленных способа. Они неразделимы и должны использоваться в комплексе. Только такой подход позволяет добиваться успеха в SEO.

### 0.1. Понятие о внутренних факторах

#### ► Верстка (HTML-теги)

Эта группа факторов учитывает правильность верстки страниц с точки зрения международного стандарта W3C. В этом плане важны HTML-теги, наделенные логическим смыслом: заголовок страницы, заголовки абзацев, смысловые выделения текста, метатеги и др. Поисковые системы учитывают текстовое содержание этих тегов и по ним определяют содержание страниц. Обратите внимание: смысловые теги не должны пересекаться друг с другом, например, заголовок не должен быть ссылкой.

#### ► Текст

Имеет значение не только текстовое наполнение страниц, но и ключевые слова, уникальность текста, объем информации и другие моменты. Очень важно наполнять сайт уникальным и полезным для посетителей контентом, кроме того, нельзя дублировать тексты на разных его страницах.

#### ► Настройки сервера

Настройки сервера – это инструкция, которую получает поисковый робот, заходя на каждую страницу. Эти настройки используются также при перемещении страниц или их удалении с сайта. Важно правильно настраивать эти инструкции, поскольку даже незначительные на первый взгляд ошибки в серверных

## МНЕНИЕ

Продвигать сайты можно разными способами. Эти способы можно разделять по многим критериям, например по степени соответствия рекомендациям поисковых систем (черные и белые методы). Поисковики плохо относятся к сайтам, использующим методы, которые вводят алгоритм в заблуждение.



К черным методам относятся:

- клоакинг (показ роботу поисковой системы другого контента, отличного от того, который увидит пользователь, например текста, содержащего длинные списки ключевых слов),
- накрутка поведенческих факторов, когда веб-мастер эмулирует интерес к своему сайту со стороны пользователя.

Покупку рекламных ссылок обычно называют серыми методами. Получается так, потому что поисковики хотели бы чтобы никто не покупал ссылки и не воздействовал подобным образом на алгоритм, однако данная практика, несмотря на такое отношение, стала общепринятой. Поэтому серьезные санкции за покупку ссылок со стороны поисковиков сейчас не применяются.

Когда говорят о «белой» оптимизации, в основном подразумевают оптимизацию внутренних факторов и создание информационных поводов в сети для получения ссылок и повышения авторитетности.

настройках могут стать фатальными для процесса продвижения сайта.

► URL

Воздействие на URL-факторы в первую очередь подразумевает использование в адресе страницы ключевых слов. Важен также вид URL, его формула, учитывается и длина. Одно из непреложных правил гласит: страница должна быть доступна только по одному URL.

► Возрастные факторы

К возрастным факторам среди прочих относятся возраст сайта, страницы, URL, динамика прироста контента, частота добавления новых страниц.

► Географическое расположение

Географическое расположение важно для регионального продвижения. В этом плане имеют значение указанные на сайте контакты и адреса организации, а также обозначенные в тексте страниц сайта региональные пункты и подобная информация.

► Перелинковка

Страницы сайта связаны между собой сетью внутренних ссылок. К группе факторов перелинковки относятся схемы перелинковки, тексты внутренних ссылок, а также вес передаваемых ссылок. Кроме того, имеет значение актуальность ссылок, а именно наличие переходов посетителей по ним.

► CMS

Важно содержание и тип сайта, вид CMS, факт использования бесплатных шаблонов, тип движка, наличие особых блоков на сайте.

► Инструкции для поисковых систем

Имеет значение наличие и содержание инструкции robots.txt, карты сайта sitemap.xml, их доступность для индексации.

**РЕЗЮМЕ:** *Внутренние факторы – самое обширное направление*



Настройки сервера – это инструкция, которую получает поисковый робот, заходя на каждую страницу. Эти настройки используются также при перемещении страниц или их удалении с сайта. Важно правильно настраивать эти инструкции, поскольку даже незначительные на первый взгляд ошибки в серверных настройках могут стать фатальными для процесса продвижения сайта.

*в поисковом продвижении. Воздействию на них не поддается автоматизации и является самым затратным направлением работ с точки зрения временных ресурсов. Между тем игра действительно стоит свеч: грамотный подход к использованию внутренних факторов позволяет достигать значительных результатов и на несколько шагов опережать конкурентов. Преимущество внутренних факторов заключается в том, что большая часть работ проводится единоразово. Из этого правила есть исключения, но обычно 20% усилий в этом направлении позволяют достигать 80% результатов. Сейчас и в ближайшем будущем именно внутренние факторы «задают тон» в продвижении сайтов.*



Важно наличие ключевых слов в тексте ссылки, а также сама форма словосочетания: прямое вхождение, словоформа, разбавление ключевых вхождений, уникальность текста ссылки. Имеет значение и текст, окружающий анкор ссылки.

## 0.2. Понятие о внешних факторах

### ► Вес внешних рекомендательных ссылок

К этой группе относится большое количество факторов, косвенно характеризующих качество площадки, размещающей рекомендательную ссылку: возраст площадки, уникальность текстов, количество размещенных рекомендательных ссылок, наличие донора в Яндекс.Каталоге, позиции сайта в ПС, уровень вложенности страницы-донора, количество контента на странице, авторитетность сайта-донора и другие факторы.

### ► Содержание текста рекомендательных ссылок

Важно наличие ключевых слов в тексте ссылки, а также сама форма словосочетания: прямое вхождение, словоформа, разбавление ключевых вхождений, уникальность текста ссылки. Имеет значение и текст, окружающий анкор ссылки.

### ► Статистика прироста внешних ссылок

Существуют (и не должны оставаться без внимания) динамические факторы,



среди которых – плавность прироста рекомендательных ссылок, соотношение их прироста в месяц и общей массы, плотность ключевых слов в анкор-листе.

**РЕЗЮМЕ:** Воздействуют на внешние факторы двумя основными способами: ищут площадки для размещения ссылок либо отправляются на специализированные биржи сайтов и статей. Однако покупка ссылок легко поддается автоматизации с помощью ссылочных агрегаторов, поэтому значение внешних факторов сейчас не так велико, как раньше, и продолжает снижаться дальше.

### 0.3. Понятие о поведенческих факторах

► Удовлетворенность пользователей содержанием сайта

Чтобы на первых страницах выдачи оказывались исключительно полезные для пользователей сайты, разработчики поисковых систем «обучают» своих детищ сбору и анализу статистики поведения посетителей сайта. По ряду косвенных признаков поисковый робот «соображает», насколько пользователь был удовлетворен результатами поиска и собственно содержимым конкретного сайта. В силах оптимизатора добиться более высокого уровня лояльности пользователей. Для этого нужно позаботиться о качестве наполнения сайта.

► Удобство пользования сайтом

Сайт должен быть не только содержательным, но и удобным. Пользователи не задерживаются на сайтах, где невозможно быстро отыскать нужную информацию, а поисковые роботы умеют по косвенным признакам определять, насколько комфортно будет чувствовать себя на сайте посетитель. Важно поработать с юзабилити: качественно представить информацию, сделать ее доступной в пару кликов, а ее изучение незатруднительным.



Чтобы на первых страницах выдачи оказывались исключительно полезные для пользователей сайты, разработчики поисковых систем «обучают» своих детищ сбору и анализу статистики поведения посетителей сайта. По ряду косвенных признаков поисковый робот «соображает», насколько пользователь был удовлетворен результатами поиска и собственно содержимым конкретного сайта. В силах оптимизатора добиться более высокого уровня лояльности пользователей. Для этого нужно позаботиться о качестве наполнения сайта.

► Переходы на сайт из иных источников

Счетчик Яндекс.Метрика определяет статистику посещений сайта, ограничивая переходы с поисковых систем от иных источников. Сайт будет ранжироваться лучше, если его посещают также по прямому заходу, из закладки или со сторонних ресурсов. Ссылка, по которой переходят живые посетители, имеет большее значение. Ссылка, которая простояла год, но ни привлекла ни одного посетителя, справедливо может считаться бесполезной.

► Статистика переходов на сайт из результатов поиска

Имеет значение и то, пользуется ли сайт спросом у пользователей поисковой системы. Разумно предположить, что сайт известной компании и /или компании, услугами которой посетитель уже пользовался и остался доволен, привлечет больше внимания. Если, несмотря на высокие позиции, сайт имеет низкий CTR – он может потерять преимущество в ранжировании.

► Прочие источники статистики

Портфель поведенческих факторов все наращивается. Возникают новые источники статистики вроде ссылок в почтовой переписке или оценок «Like» («Мне нравится») в социальных сетях.

***РЕЗЮМЕ: Поведенческие факторы уже стали самым значимым фактором. Однако тенденции развития поисковых систем явно направлены в сторону улучшения анализа поведенческих факторов, так что в будущем стоит ожидать постепенного повышения значимости этой группы факторов.***

#### 0.4. Комплексный подход к продвижению сайтов

В деле продвижения сайтов важен комплексный подход: оптимизатор должен воздействовать на все группы факторов.

### ЛИКБЕЗ

#### Анкор ссылки (Anchor)

Это текстовое содержимое, находящееся между тегами. Синоним – «текст ссылки».

#### Апдейт

Апдейтом называют пополнение базы данных поисковой машины новой информацией. Выдача поисковых систем меняется с различной частотой. В Google это происходит постоянно, в Яндексе апдейты случаются реже, несколько раз в неделю. В течение дней между апдейтами выдача Яндекса практически не меняется. Исключение – такие тематики и запросы, где присутствуют результаты работы специализированных роботов, например новости или блоги.

#### Внешняя ссылка

Это голосующая единица, дающая рекомендацию сайту, на который она ведет. Понятие «внешняя» означает, что один сайт ссылается на другой, в отличие от внутренних ссылок, которые перенаправляют пользователя в рамках одного сайта. Внешние ссылки помогают поисковым системам определить содержимое страниц, а также их значимость. Небезосновательно предполагается, что на более качественный сайт в Сети размещено большее количество внешних ссылок. Существуют специализированные биржи для работы с такими ссылками. Кроме того, уже несколько лет оптимизаторы пользуются агрегаторами – автоматизированными системами, которые позволяют решить вопрос покупки ссылок практически без ручного труда.

#### Внутренняя ссылка

Так называют ссылку, ведущую на другую страницу того же сайта. Понятие «внутренняя» означает, что одна страница ссылается на другую в рамках одного ресурса, «внутри» него. Внутренние ссылки влияют на перераспределение веса между страницами, этот процесс носит название перелинковки. Перелинковка – один из важных моментов внутренней оптимизации сайта, который мастерами используют в своих интересах опытные SEO-специалисты.

#### ВЧ

Аббревиатура расшифровывается как «высокочастотный». Обычно этим термином обозначают наиболее часто вводимые запросы. Нужно понимать, что частотность, пусть и является спутником конкуренции, напрямую с ней не связана: нередко можно встретить крайне конкурентные запросы, которые не являются высокочастотными.

#### Донор, Акцептор

Донором называют страницу, на которой размещена рекомендательная ссылка. Соответственно, акцептором в поисковом продвижении обозначают страницу, на которую эта ссылка ведет.

#### Дублированный контент

Обозначает неуникальный, существующее где-то еще в Интернете контент. Поисковые системы негативно воспринимают дубликаты, поскольку с ними связаны проблемы хранения и обработки «мусорных» данных.

### Индекс

Это база данных поисковой системы, хранилище всех документов, собранных в Сети. Поиск осуществляется только по тем документам, которые в данный момент времени содержатся в индексе поисковых систем.

### Конверсия

Так называется отношение количества посетителей сайта, совершивших какое-либо целевое действие, к общему количеству привлеченных посетителей. Целевым действием может быть регистрация в сервисе, звонок, заказ услуги или покупка товара, подписка на рассылку, добавление сайта в закладки, скачивание файла и т. п.

### НЧ

Аббревиатура расшифровывается как «низкочастотный». Низкочастотными называют те запросы, которые пользователи поисковых систем вводят редко. Многие из таких запросов вводят раз в месяц и даже реже. НЧ-запрос может быть тем не менее высококонкурентным. Продвигать такие запросы проще. Кроме того, тенденция развития поведения пользователей — стремление к уточнению запросов, поэтому с каждым годом доля НЧ-запросов возрастает, а ВЧ падает. Как правило, «низкочастотники» выигрывают в трафике за счет охвата огромной массы запросов.

### Покупка ссылок

Цель этого мероприятия — улучшение позиций сайта по запросам, содержащимся в тексте ссылки. Существуют разные источники покупки ссылок: биржи, ручной обмен, размещение статей и т. п.

### ПС

Так сокращенно называют поисковую систему. Для оптимизатора наибольший интерес представляют поисковые системы Яндекс, Google, Mail.ru. Некогда лидировавшая на рынке поиска система Rambler сейчас при продвижении всерьез не рассматривается, поскольку обеспечивает совсем незначительное количество посетителей.

### Результаты поиска

Это ответ поисковой системы на запрос пользователя. При стандартных настройках на первой странице результатов поиска в Яндекс и Google оказывается всего десять сайтов, поэтому большинство договоров на SEO-услуги предполагают вывод именно в десятку.

### Релевантность

Под релевантностью понимают меру схожести и применяют это понятие в SEO при оценке соответствия результатов поиска запросу пользователя. Понятие «релевантная запросу страница» используется также в момент выбора входной страницы (страницы, на которую будет попадать посетитель из поисковых систем). Часто в качестве входной страницы выбирают наиболее релевантную.

### Сниппет

Это описание сайта в результатах поиска. Традиционно сниппет берется из текста страницы. При его формировании также могут быть использованы другие источники, например,

При таком раскладе возникает синергия, и значение каждого фактора в группе возрастает. Подобный подход возможен за счет планирования стратегии продвижения, представляющей собой план долгосрочного последовательного захвата позиций в поисковых системах. Давайте разберем такую стратегию.

### Этап 1: Подготовка сайта к продвижению

#### ► Составление семантического ядра

Составление семантического ядра — важнейший этап в продвижении сайта. На первый взгляд может показаться, что технология подбора ключевых слов незатейлива, но это не так. Большинство оптимизаторов совершают фатальные ошибки уже на этой стадии. В действительности «незатейливый» процесс подбора и проверки запросов, например, для интернет-магазина бытовой техники, может занять у оптимизатора несколько дней.

#### ► Анализ тематики

Сейчас, когда алгоритмы используют технологию машинного обучения, каждая тематика может иметь свои особенности в продвижении. Важно провести анализ конкурентов в топ-10 — топ-30, выделить среди них сайты, похожие на ваш, подсчитать средние значения факторов ранжирования и сделать все «под копирку».

#### ► Выбор входных страниц

При выборе входных страниц важно учесть два правила. Во-первых, чем меньше запросов продвигается на одну страницу, тем лучше. Не бывает много входных страниц, бывает много запросов. Поэтому, если вы решили продвигать большое количество запросов, а подходящих страниц нет, не старайтесь уместить ключи на уже существующих, лучше создавайте новые страницы. Во-вторых, на входной странице посетитель обязательно должен найти то, за чем он пришел на сайт.

► Составление технического задания на написание текстов

Подготовка текстов – важный этап, и ошибки здесь обходятся дорого. Поэтому техническое задание (ТЗ) на написание текстов должен подготовить опытный SEO-специалист. В этом документе должны содержаться подробные указания: объем текста на страницах, употребляемые ключевые слова, их количество, рекомендации по составлению заголовков. После написания текстов их нужно проверить на соответствие требованиям ТЗ.

► Подготовка текстов

Тексты должны не только содержать ключевые слова, но также быть полезными и удобочитаемыми для посетителей. Не рекомендую прибегать к услугам дешевых неопытных SEO-копирайтеров, за качество стоит переплатить. В первую очередь тексты должны продавать. В книге мы рассмотрим подход к продвижению, при котором достигается баланс между SEO и продающей способностью текстов.

► Аудит внутренних факторов

Необходимо провести полную диагностику сайта, выявить ошибки, обнаружить области с незадействованным потенциалом. После этого следует устранить недоработки и подстроить сайт под требования и рекомендации поисковых систем.

► Составление технического задания на доработку сайта

Задача оптимизатора – подготовить полное и понятное ТЗ для веб-программиста, проконтролировать выполнение работы и принять результат. Без привлечения программиста внести все правки удастся только в том случае, если оптимизатор сам уже несколько лет работает с CMS-системами, умеет управлять базами данных и знает языки программирования.

► Доработка сайта под SEO

описание сайта в Яндекс.Каталоге или Description. Снippet может содержать разделы сайта, дату, новости, телефон и адрес компании. Привлекательный для пользователя snippet может существенно увеличить посещаемость сайта.

**СЧ**

Аббревиатура расшифровывается как «среднечастотный». Этим термином обозначают нечто среднее между высокочастотным и низкочастотным запросом.

**Сквозняк**

Сквозной называют ссылку, размещенную на многих страницах сайта. «Сквозняки» дублируются либо между всеми страницами сайта, либо внутри какого-то раздела. Они, как правило, «зашиты» в шаблон.

**Трафик**

Это синоним понятия «количество посетителей сайта».

**Семантическое ядро**

Подразумевается набор запросов, по которым продвигается сайт. Составление семантического ядра – очень важный, однако сложный и долгий процесс. Семантическое ядро для обычного интернет-магазина содержит несколько тысяч запросов, для некоторых крупных сайтов их количество исчисляется десятками тысяч.

**ТИЦ**

Аббревиатура расшифровывается как «тематический индекс цитирования». ТИЦ – это числовая величина, характеризующая авторитетность того или иного сайта, с точки зрения Яндекса. Рассчитывается ТИЦ на основании качества и количества ссылающихся на данный сайт источников с учетом динамики прироста этих ссылок.

Лучше изначально делать сайт под SEO, чем «натягивать» SEO на сайт. Но, к сожалению, культура создания сайтов под SEO развита у нас очень слабо: людей, способных делать такие сайты, можно пересчитать по пальцам. Несмотря на то, что многие компании, предоставляющие услуги комплексного интернет-маркетинга, уверяют, что делают «заточенные» под SEO сайты, на проверку оказывается, что лишь единицы могут похвастаться слаженной работой специалистов. В итоге часто на выходе получается сайт, который оптимизаторам приходится значительно дорабатывать под нужды SEO. Причем доработки могут быть очень серьезными, вплоть до существенных изменений в структуре сайта и его программной части. Тем не менее без них об эффективном продвижении сайта в поисковых системах говорить не приходится.

## Этап 2: Продвижение сайта

### ► Покупка рекомендательных ссылок

Работа над внешними факторами в целом сводится к покупке рекомендательных ссылок. На сегодняшний день с этим лучше всего справляются ссылочные агрегаторы, позволяющие полностью автоматизировать процесс: вам нужно лишь указать запросы и настроить покупку ссылок.

### ► Привлечение бесплатных рекомендательных ссылок

Большое количество ссылок можно добыть бесплатно. На их поиск придется затратить немало времени, но, если подойти к этому процессу с умом, многие бесплатные ссылки будут иметь гораздо больший вес, чем покупные. Этот вопрос подробно разобран на моем DVD-семинаре «Продвижение сайта: источники привлечения внешних ссылок».

- Повторный аудит внутренних факторов
- Подготовка дополнительных текстов
- Повторная доработка сайта под SEO



Не стоит тешиться иллюзией, будто ваш веб-мастер, наделавший ошибок

при верстке и программировании сайта, исправит их по вашему новому ТЗ и не допустит ни одной новой оплошности. К примеру, он может поставить редирект 302 вместо 301. Посетителям от этого хуже не станет, но поисковые системы могут вас неправильно «понять», и для продвижения в поиске такое «пустяковое» различие может стать фатальным. Важно не просто проверить исполнение рекомендаций по ТЗ, а провести полноценный повторный аудит, иначе новые ошибки могут остаться незамеченными. Приготовьтесь к тому, что аудит придется проводить несколько раз, и только после проведения контрольного анализа, подтверждающего, что ошибок больше нет, можно будет вздохнуть спокойно.

Из стратегического плана видно, насколько важно воздействовать на внешние и внутренние факторы в течение всего периода продвижения сайта. Среди оптимизаторов бытует миф, будто достаточно один раз внести доработки на сайт на начальной стадии продвижения и можно закрыть вопрос воздействия на внутренние факторы. В действительности все обстоит иначе. **Внутренние факторы также нуждаются в постоянном анализе и доработке.** Ниже рассмотрим основные причины такой необходимости.

► Со временем текст сайта теряет уникальность

Контент «утекает» с сайта по разным каналам. Тексты воруют конкуренты, которые не удосужились придумать что-то свое, а также не порядочные «дорвейщики», зарабатывающие на перепродаже трафика. Копии текста страницы множатся, а сайт теряет вес для поисковых систем. Чтобы не потерять позиции по ВЧ- и СЧ-запросам, время от времени нужно проводить анализ уникальности контента и при необходимости обновлять его, переписывая тексты. Что касается продвижения по НЧ-запросам, воровство контента не наносит сильного ущерба текущим позициям, тем более что существует защитная методика, о которой будет рассказано в одной из следующих глав.

► После доработок сайта часто возникают новые проблемы

Не стоит тешиться иллюзией, будто ваш веб-мастер, наделавший ошибок при верстке и программировании сайта, исправит их по вашему новому ТЗ и не допустит ни одной новой оплошности. К примеру, он может поставить редирект 302 вместо 301. Посетителям от этого хуже не станет, но поисковые системы могут вас неправильно «понять», и для продвижения в поиске такое «пустяковое» различие может стать фатальным. Важно

## МНЕНИЕ

### Настоящее и будущее факторов ранжирования



Основная задача поиска – обеспечить точный и полный ответ на запрос пользователя. Яндекс, равно как и другие поисковые системы, использует множество метрик качества поиска, например:

- средняя позиция первого клика в выдаче,
- доля некликаемых результатов.

Любая модификация алгоритмов антиспама или ранжирования не должна ухудшить качество поиска. В условиях существенного seo-прессинга разработчики поисковых систем вынуждены постоянно искать новые качественные и незаспамленные сигналы, которые можно использовать в технологии антиспама или ранжирования.

Поле факторов стремительно расширилось за последние несколько лет благодаря двум основным направлениям:

1. Факторы, учитывающие поведение пользователей: продолжительность просмотра документов, str-ссылок на сайт в выдаче поисковых систем, доля непоискового трафика на сайт и т. д.
2. Факторы, учитывающие отношение пользователей к сайту: цитируемость и количество «лайков» в социальных сетях, количество твитов и ретвитов со ссылкой на сайт и пр. и пр.

Модель, учитывающая поведение и отношение пользователей к документам сайта, существенно более устойчива к спаму и позволяет более эффективно ранжировать документы в выдаче. Можно легко проиллюстрировать это сравнением топовых сайтов, отранжированных различными алгоритмами:

Top-20 websites by three different algorithms

No	PageRank	TrustRank	BrowseRank
1	adobe.com	adobe.com	<i>myspace.com</i>
2	passport.com	yahoo.com	msn.com
3	msn.com	google.com	yahoo.com
4	microsoft.com	msn.com	<i>youtube.com</i>
5	yahoo.com	microsoft.com	live.com
6	google.com	passport.net	<i>facebook.com</i>
7	mapquest.com	ufindus.com	google.com
8	miibeian.gov.cn	<i>sourceforge.net</i>	ebay.com
9	w3.org	<i>myspace.com</i>	<i>hi5.com</i>
10	godaddy.com	<i>wikipedia.org</i>	<i>bebo.com</i>
11	statcounter.com	phpbb.com	<i>orkut.com</i>
12	apple.com	yahoo.co.jp	aol.com
13	live.com	ebay.com	<i>friendster.com</i>
14	xbox.com	nifty.com	<i>craigslist.org</i>
15	passport.com	mapquest.com	google.co.th
16	<i>sourceforge.net</i>	cafepress.com	microsoft.com
17	amazon.com	apple.com	<i>comcast.net</i>
18	paypal.com	infoseek.co.jp	<i>wikipedia.org</i>
19	aol.com	miibeian.gov.cn	<i>pogo.com</i>
20	<i>blogger.com</i>	<i>youtube.com</i>	<i>photobucket.com</i>

Источник: BrowseRank: Letting Web Users Vote for Page Importance, <https://research.microsoft.com/en-us/people/tyliu/fp032-liu.pdf>

Как видно, алгоритм BrowseRank позволил существенно лучше отранжировать сайты с длинной пользовательской сессией, чем традиционный PageRank.

Логично ожидать и дальнейшего расширения поля влияющих факторов.

Вообще довольно эффективных критериев можно придумать весьма немало. Для решения одной из задач мне потребовалось придумать алгоритм фильтрации сплюгов (спамовых блогов, созданных исключительно для размещения рекламных сообщений). Одна из найденных закономерностей: посты в спамовых блогах обычно получают гораздо меньше комментариев, чем посты в хороших блогах. Распределение количества блогов по среднему числу комментариев к посту выглядит так:



Данные получены на выборке в 30 тысяч блогов. Использование лишь одного параметра позволило отфильтровать спамовые блоги с точностью и полнотой порядка 80% при пороге ~7 комментариев на пост. Введение в алгоритм фильтрации дополнительных критериев позволит еще более эффективно отсеять спам.

Собственно, в этом состоит один из серьезных трендов современного поиска — использование мономов, т. е. комбинаций факторов. Широкое применение мономов позволяет повысить эффективность алгоритмов и затрудняет реверс-инжиниринг.

Логичен вывод: любой сигнал, обеспечивающий достоверное отделение некачественного документа или сайта от качественного, будет использован.

не просто проверить исполнение рекомендаций по ТЗ, а провести полноценный вторичный аудит, иначе новые ошибки могут остаться незамеченными. Приготовьтесь к тому, что аудит придется проводить несколько раз, и только после проведения контрольного анализа, подтверждающего, что ошибок больше нет, можно будет вздохнуть спокойно.

► Некоторые доработки являются экспериментальными

Многие доработки экспериментальны, особенно это касается тех факторов, которые зависят от тематики. С первого раза крайне редко удастся выбрать правильные значения для подобных факторов. Прибавьте к этому нестабильность среды: алгоритм Яндекса, использующий технологии машинного обучения и оценки асессоров, постоянно изменяется в рамках каждой тематики. В результате становится очевидным необходимость экспериментов и оценки результатов с адаптацией сайта к тематике. Только при таком раскладе можно рассчитывать на повышение эффективности продвижения. Экспериментальный подход хорошо иллюстрирует работа с плотностью ключевых слов, когда в написанном тексте мы увеличиваем плотность ключевых слов и ждем апдейта. Если позиции падают — уменьшаем плотность. Если растут — увеличиваем еще.

► Конкуренты не стоят на месте

Выход в топ для многих оптимизаторов становится сигналом к прекращению работы над сайтом. Между тем нельзя терять бдительность. Удержание позиций — это постоянная борьба. Все хотят в топ, а в итоге попадает туда самый упорный и смысленный. Чтобы удержать позиции, нужно анализировать конкурентов и вовремя повторять их шаги.

► Контент-менеджер может допускать ошибки

Если сайт постоянно обновляется, контент-менеджеры публикуют новости и добавляют товары в каталог, нужно отслеживать, не допускают ли они критичных для SEO ошибок. А в случае, когда на сайте периодически появляются новые разделы и меняется текстовое наполнение старых, вероятность ошибок резко возрастает и почти равна 100%.

**РЕЗЮМЕ:** *Хочу заметить, что ошибка никогда не станет критической, если отыскать ее сразу после появления, а не дотянуть до падения позиций сайта и лишь тогда приступить к исправлению упущений.*



## Глава 1

# Перелинковка страниц

---



**П**ри перелинковке важно понимать, каким образом вес передается между страницами. Сразу хотелось бы отметить, что знание это весьма относительно, так как при передаче веса от страницы к странице задействуется множество факторов и точного расчета здесь не может быть по определению.

Доподлинно известно, что при расчете веса ссылки учитывается фактор ее «кликабельности», то есть та ссылка, по которой пользователи активно кликают мышью, передает больший вес, нежели та, которая остается без внимания посетителей сайта. Это утверждение справедливо не только в плане внутренней перелинковки: расчет влияния входящий внешних ссылок ведется по такой же схеме. Свою роль играет и способ размещения ссылки, в частности, важно, является ли ссылка сквозной.

Как ни печально, но при всем желании вы не сможете рассчитать передаваемый ссылкой вес точно. Однако знание о способах перелинковки позволит вам максимально задействовать потенциал внутренних ссылок с точки зрения технологии передачи ссылочного веса PageRank.

Некоторые оптимизаторы могут возразить: о каком PageRank идет речь, если эта технология по нынешним меркам уже может быть отнесена к «каменному веку» эпохи интернет-продвижения? Да, технологии уже 10 лет, безусловно, она претерпела немалые изменения, однако базовые принципы остались прежними. С ними я вас и познакомлю в этой главе.

## 1.1. Как вычисляется PageRank и почему это так важно знать

Говоря о поисковом предложении в целом, я стараюсь предлагать универсальные решения, которые подходят для всех (или для большей части) популярных поисковых систем. Однако иногда выбора между различными подходами не избежать, и в этом случае я вынужден отдавать приоритет продвижению под поисковую систему Яндекс, доля поискового трафика которой растёт день ото дня.

В этой связи кажется странным, что я посвящаю целую главу PageRank, который является атрибутом Google, но никак не Яндекс. Зачем я это делаю? Ответ прост: значение PR-страницы для Яндекса не важно, и тем не менее эта поисковая система тоже оперирует таким понятием, как вес страницы, и использует ссылки при ранжировании. Поскольку основа технологии PageRank представляет собой базовые принципы ссылочного ранжирования, вполне закономерным кажется следующий вывод: знание принципов расчета PageRank наилучшим образом поможет вам понять, как передается вес между страницами с точки зрения популярных поисковых систем. Итак, начнем.

Что все это значит на практике?

Чем глубже в структуре сайта расположена страница, тем меньший вес ей достается (помним о коэффициенте затухания). Вес страницы делится поровну между всеми ссылками на странице. Соответственно, входные страницы должны получать максимальный вес, а второстепенные страницы – обеспечивать максимальный вес входным страницам. Добиться оптимального соотношения можно, закрывая лишние ссылки от индексации.

Существуют моменты, которые могут стать помехой для оптимизации:

### МНЕНИЕ

**В** 1998 году основатели компании Google Сергей Брин и Ларри Пейдж опубликовали статью, в которой ввели понятие PageRank, как меру авторитетности страницы. Рейтинг страницы в этой модели зависит от количества ведущих на нее ссылок и от рейтинга ссылающихся страниц. Статическим весом является доля PageRank, передаваемая каждой ссылкой.



Каждая рассмотренная модель является условным идеальным примером, не учитывающим, что на внутренние страницы могут стоять ссылки извне. Также результаты расчета статического веса внутренних страниц могут искажаться из-за неравномерности индексации страниц со ссылками поисковым роботом.

Большую роль играет и то, на скольких страницах сайта размещена ссылка. Одна сквозная, т. е. стоящая на всех страницах сайта ссылка перераспределит значения рангов внутренних страниц гораздо сильнее, чем одна ссылка с главной страницы.

\*Sergey Brin and Larry Page. The Anatomy of a Search Engine  
<http://www-db.stanford.edu/pub/papers/google.pdf>

## ФАКТ

### Растолкованный PageRank<sup>1</sup>

**PageRank** – это методика измерения «важности» веб-страницы в Google. Когда все другие значимые факторы, такие как `тег Title` и ключевые слова, учтены, Google прибегает к PageRank и корректирует результаты ранжирования с расчетом, чтобы более важные сайты оказывались выше на странице результатов поиска пользователя.

**Главный момент: никто точно не знает, как Google рассчитывает PR.** Однако есть нюансы, которые стоит учитывать всегда:

#### 1. Сайт, который ссылается на забаненные сайты, может попасть в бан.

Будьте щепетильны в выборе исходящих ссылок. Не ссылайтесь на подозрительные сайты: каталоги ссылок («линко-помойки»), забаненные ресурсы и т. п. Всегда проверяйте PR сайтов, на которые собрались сослаться, иначе есть вероятность, что Google выпишет вам «желтую карточку» – пенализирует сайт.

#### 2. Ссылки на тематические сайты с высоким PR и ссылки с таких сайтов очень важны.

Причем передаваемый ссылкой PR будет тем больше, чем ближе будет тематика страниц. Ссылки на полезные и интересные тематические ресурсы указывают поисковым машинам, что сайт имеет определенную ценность для пользователей Сети. Таким образом, проставляя ссылки только на качественные сайты, можно на несколько шагов опередить конкурентов.

#### В целом порядок ранжирования в Google выглядит так:

1. Найти все страницы в индексе, которые соответствуют ключевым словам поиска.
2. Отранжировать найденное в соответствии со «страничными факторами» (ключевыми словами и т. п.).
3. Учесть текст ссылок на страницы.
4. Скорректировать результаты, используя данные PageRank.

#### Принципы вычисления PageRank:

- PageRank – это число, характеризующее исключительно голосующую способность всех входящих ссылок на страницу и ту силу, с которой они эту страницу рекомендуют.
- Если страница А ссылается на страницу В, то страница А считает, что страница В важная.
- Если на страницу ведет много важных ссылок, то и ее ссылки на другие страницы становятся более важными.
- Внутренние ссылки учитываются при расчете веса PageRank для других страниц сайта.

<sup>1</sup>По материалам статьи «Растолкованный PageRank» в переводе Александра Садовского.

- ограниченные возможности CMS,
- обилие ссылок на крупном сайте,
- устоявшаяся структура сайта.

## 1.2. Базовые схемы перелинковки

Для достижения целей вам необходимо использовать правильную схему перелинковки. На примере некоторых базовых конструкций я покажу, каким образом может быть выстроена такая схема.

Сразу сделаю небольшую оговорку: любая схема перелинковки условна и может быть масштабирована. Так, рассматривая схему, в которой участвуют 10 страниц, мы можем подразумевать сайт с десятками тысяч страниц, которые распределены по 10 основным подразделам. Таким образом, мы построим пути передачи ссылочного веса по главным разделам, а потом уже рассмотрим их по отдельности и просчитаем схему распределения веса в каждом разделе.

Планируя перелинковку, я исхожу из соображений оптимального использования внутреннего ссылочного потенциала сайта, поэтому высокочастотные запросы в приведенных ниже схемах перелинковки двигаются на главную страницу, среднечастотные – на страницу раздела (обычно это страница первого уровня вложенности), а низкочастотные – на страницы товара/услуги (обычно это страницы второго, третьего и ниже уровня вложенности).

Универсальной схемы перелинковки не существует. Конкретный выбор зависит даже не от выбора группы запросов, а от особенностей тематики. Существуют такие тематики, в которых основная доля посетителей приходится на 5 запросов, а все остальные запросы не принесут даже 10% того трафика, что обеспечат эти 5 (например, тематика «такси»). В таких случаях, безусловно, оптимальной тактикой будет продвижение пяти выбранных запросов

на главную страницу и выбор схемы перелинковки под ВЧ-запросы.

### Комплексное продвижение всего сайта

Эта схема наиболее универсальна и с ней приходится сталкиваться чаще всего. Она подходит в тех случаях, когда планируется продвигать все группы запросов (ВЧ, СЧ, НЧ), а также тогда, когда шаблонизатор сайта не позволяет закрыть от индексации часть ссылок на конкретной странице. Недаром эту схему перелинковки называют «все ссылаются на всех».

Принцип ее построения прост:

1. Главная страница ссылается на страницы всех разделов.
2. Страница каждого раздела ссылается на вложенные страницы, а также на главную. Разделы перелинковываются между собой так, чтобы все они ссылались друг на друга.
3. Все вложенные страницы ссылаются на вышестоящий раздел и на главную страницу. Вложенные страницы конкретного раздела ссылаются друг на друга. Все вложенные страницы – на начальные страницы соседних разделов.

При таком раскладе вес равномерно угасает от страницы 0 уровня вложенности к странице 2 уровня. Однако у этой схемы есть определенный недостаток: сайту, на котором применена такая методика перелинковки, при продвижении ВЧ-запросов сложно конкурировать с сайтами, где задействована схема перелинковки, предполагающая направление всего веса на главную страницу.

Рассмотрим такую схему на рис. 1.

### Продвижение главной страницы под ВЧ

В условиях напряженной конкуренции используют главный «ход конем» – перелинковку с передачей веса на главную страницу. Эта методика очень проста: каждая страница ссылается на главную,

- Страница не может голосовать так, чтобы другая страница стала столь же важна, как она сама. Таким образом, чем труднее добраться до какой-то web-страницы, тем меньшим авторитетом она обладает.
- Существует так называемый коэффициент затухания: страница всегда передает другой странице меньший вес, чем получает сама. Например, коэффициент равен 0,85. В таком случае передача веса после затухания равна 0,85. В том случае, если со страницы ведут две ссылки, переданный по каждой из них вес будет равняться 0,425 и так далее.
- Во время одной итерации вычислений страница 1 дает странице 2 повышение в весе. Если страницы ссылаются друг на друга, то во время следующей итерации страница 1 получает увеличение веса, пропорциональное новому, увеличившемуся весу страницы 2.
- Вес PageRank, передаваемый на страницу А со страницы В, которая указывает на нее, куда-нибудь уменьшается с каждой ссылкой, которая находится на странице В. Это значит, что вес страницы – это, по сути, мера ее голоса. Страница может разделить этот голос между одной, двумя или более ссылками, но общая голосующая сила остается неизменной.

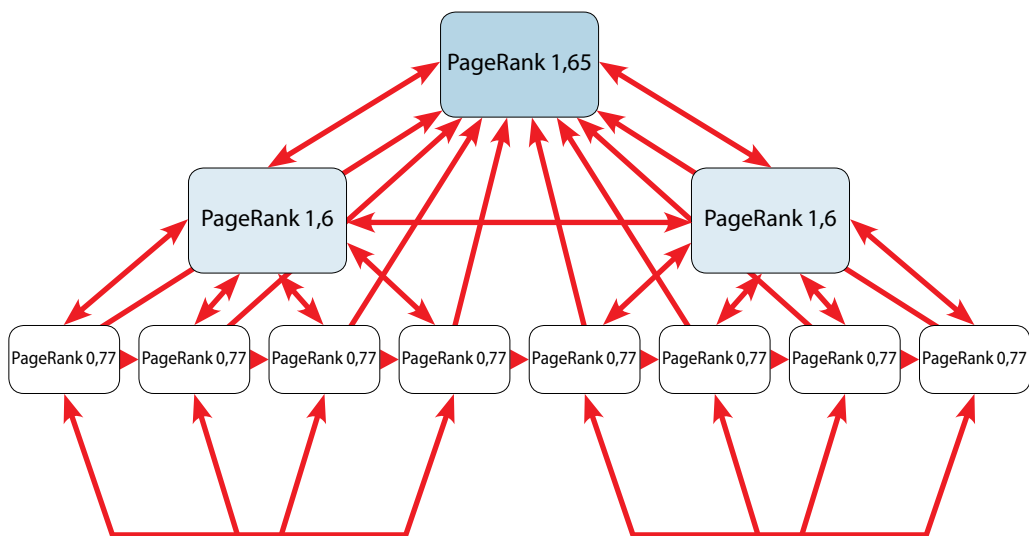


Рис. 1. Комплексное продвижение всего сайта

## ФАКТ

### Пара фактов о PageRank

«Чем меньше ссылок со страницы – тем меньше веса с нее утекает» – думают многие оптимизаторы. Это миф. Взглянув на формулу PageRank нетрудно понять, что количество исходящих ссылок не имеет никакого значения. Одна она будет или 100 – со страницы уйдет одинаковое количество веса.



Более того, даже если убрать все исходящие ссылки со страницы, оставив входящие, это не приведет к накоплению на ней PageRank. Здесь вступают в игру методы обработки «висящих ссылок» (dangling links), и такие страницы могут вовсе не получить статического веса.

Для больших сайтов бывает трудно построить модель и посчитать PageRank. Тогда можно руководствоваться двумя простыми принципами:

- страница имеет высокий PageRank, если количество входящих ссылок кратно превышает количество исходящих,
- страница имеет высокий PageRank, если на нее ссылаются страницы предыдущего типа (высокий относительно среднего уровня. Разумеется, он будет меньше, чем в варианте «а»).

Эти принципы не являются гарантией. Существуют исключения. Но в большинстве случаев они работают.

и на каждую страницу, кроме главной, стоит всего одна ссылка (рис. 2). При такой перелинковке главная страница получает максимально возможный вес. Недостатки этой схемы очевидны: в плюсе остается только главная, а все остальные страницы получают очень низкий вес, поэтому конкурировать по СЧ- и НЧ-запросам с сайтами, избравшими стратегию приоритета этих запросов, становится затруднительно.

### Продвижение разделов под СЧ

Такая схема прекрасно подходит интернет-магазинам в тех отраслях, где люди ищут не название модели/товара, а целую товарную группу. Например, когда пользователь хочет купить индийскую косметику, он не ориентируется в брендах, а ищет просто «индийский шампунь» или «индийскую пену для ванны», поэтому для магазина индийской косметики схема продвижения разделов под СЧ будет идеальной.

Методика сводится к тому, что сайт собирает трафик по СЧ запросам и вес получают страницы разделов. Принцип построения этой схемы прост: все страницы сайта

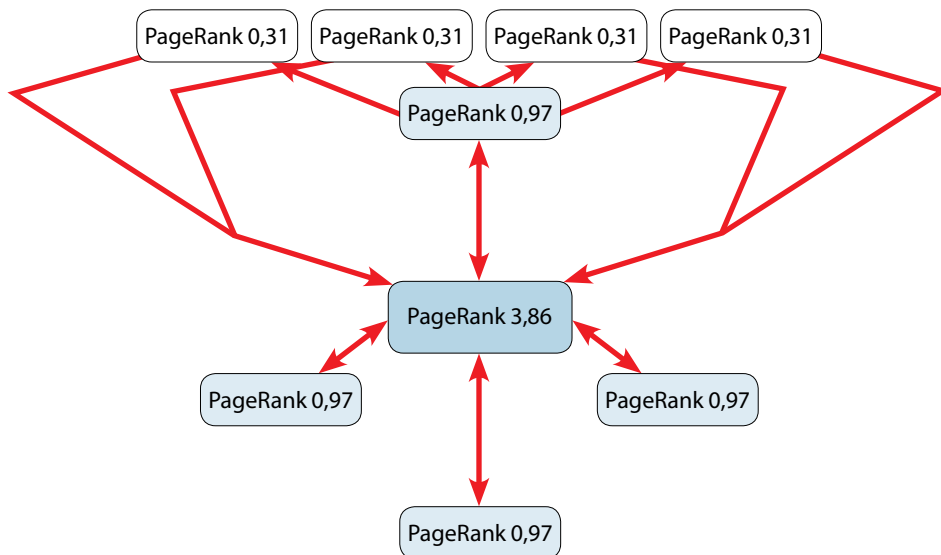


Рис. 2. Продвижение главной страницы под ВЧ

ссылаются на страницы разделов, и исключительно на страницы разделов ссылаются страницы 2 и ниже уровня вложенности. Выглядит такая схема перелинковки следующим образом (рис. 3).

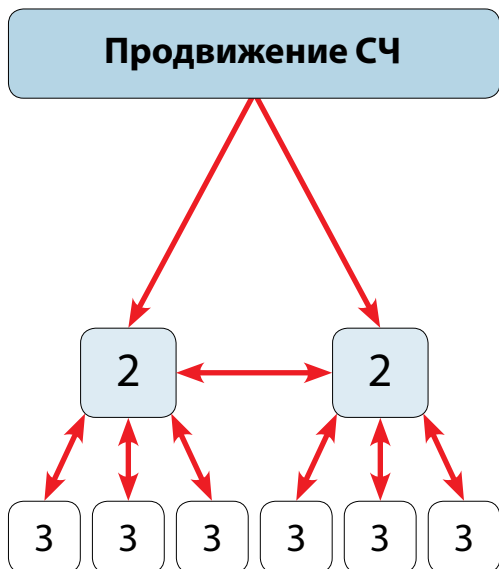


Рис. 3. Продвижение разделов под СЧ



Следует отметить, что комплексный подход к продвижению является намного более эффективным, чем методика, ориентированная лишь на один из сегментов запросов.

### Продвижение внутренних страниц под НЧ

В интернет-магазине, который торгует продукцией популярных брендов, пользователи, как правило, ищут товар по названию конкретной модели. В этом случае стоит задействовать схему перелинковки, продвигающую внутренние страницы под НЧ (рис. 4).

Пример тематики, для которой схема окажется чрезвычайно эффективной, – ноутбуки. Так, на момент написания книги «ноутбук asus» по России искали 4 393 человека, а «ноутбук Asus k50c» – 722 человека. А ведь у бренда Asus широчайший модельный ряд. Несложно смекнуть, что в совокупности запросы с названием моделей могут дать больше трафика, чем более общий запрос «ноутбук Asus». Кроме того, поскольку эти запросы более точные, конверсия с них будет на порядок выше, ведь посетитель найдет именно ту модель, которую искал.

Если вы выбираете эту схему перелинковки, стоит придерживаться следующих правил: страницы карточек товаров/услуг ссылаются только друг на друга, нельзя отдавать вес страницам разделов.

Итак, мы рассмотрели основные схемы перелинковки. Помните о масштабировании: одной страницей на шаблоне по факту может являться огромный раздел сайта. Следует отметить, что комплексный подход к продвижению является намного более эффективным, чем методика,

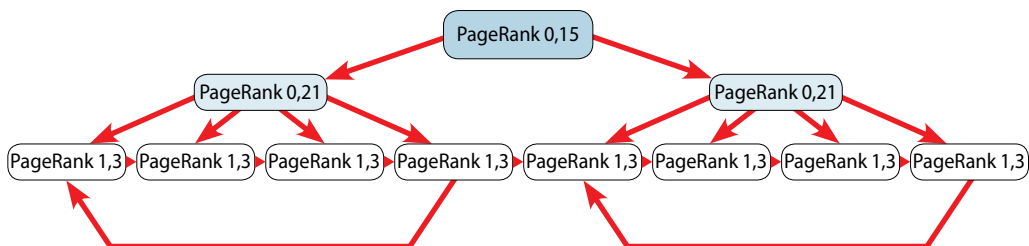


Рис. 4. Продвижение внутренних страниц под НЧ

ориентированная лишь на один из сегментов запросов (ВЧ, СЧ, НЧ). Конечно, в тематиках типа «такси» или «пластиковые окна» выбор схемы перелинковки под ВЧ-запросы полностью оправдан и даже является единственно верным, однако в других случаях я не рекомендую использовать строго схему продвижения одной группы. Каждый раз на этапе выбора схемы перелинковки нужно определиться с целями и особенностями тематики и уже исходя из этой информации модифицировать схему под свои нужды.

### 1.3. Локальные схемы перелинковки

Чтобы правильно строить схемы перелинковки, важно разбираться в способах перелинковки локальной. В данном случае между собой связываются родственные страницы одного уровня вложенности (чаще всего вложенные в один раздел).

#### Кольцо

Популярная схема перелинковки, в которой страницы последовательно связываются между собой, замыкаясь в кольцо:  $1 \rightarrow 2 \rightarrow 3 \rightarrow 1$ .

Явный недостаток такой схемы сводится к тому, что в случае выпадения одной страницы из индекса связь нарушается и вес теряет все кольцо. Экспериментально доказано: в сравнении с другими схемами локальной перелинковки кольцевая схема дает максимальный вес каждой странице.

#### Куб

Для реализации этой схемы страницы группируются в два кольца, каждое из которых состоит из количества страниц, кратного четырем. Считаем звенья одного кольца как «А» и второго как «В». Кольца перелинковываются между собой перекрестно по принципу  $A1 \rightarrow B2 \rightarrow A3 \rightarrow B4 \rightarrow A1$ ;  $B1 \rightarrow A2 \rightarrow B3 \rightarrow A4 \rightarrow B1$ ;  $A1 \rightarrow A2 \rightarrow A3 \rightarrow A4 \rightarrow A1$ ;  $B1 \rightarrow B2 \rightarrow$



Допустимы случаи со значительным отклонением от предложенных чисел.

Например, существует рекомендация об ограничении количества исходящих ссылок со страницы. Когда мы работаем с особо крупными сайтами (новостными порталами, серьезными агрегаторами), отклонения неизбежны.



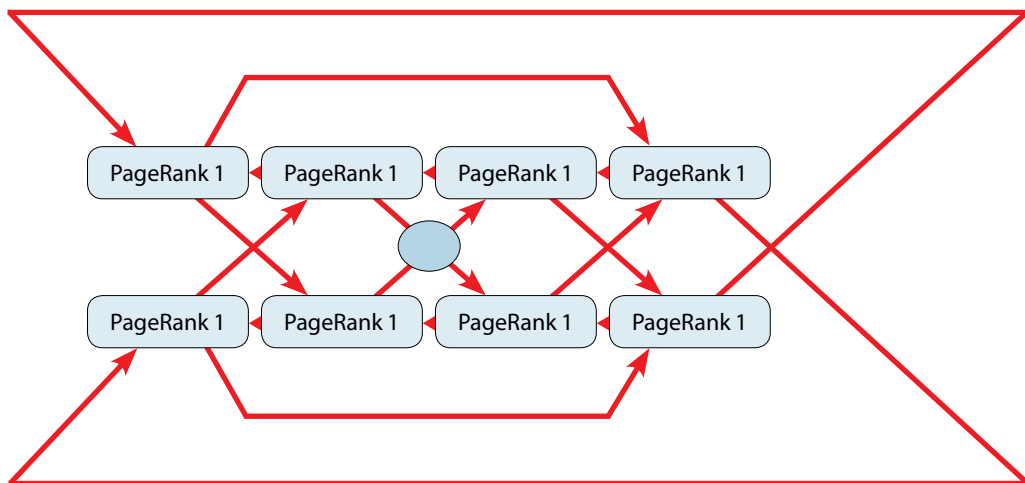


Рис. 5. Куб

В3 —> В4 —> В1. Схема «Куб» на порядок сложнее «Кольца» и труднее поддается автоматизации, однако у нее есть серьезный козырь: даже если несколько звеньев выпадет из этой цепочки, вес все равно сохранится и цепь будет замкнута.

### Звезда

Название этой схемы не должно вас пугать. На самом деле все здесь предельно просто: «все ссылаются на всех».

## 1.4. Правила перелинковки

Мы рассмотрели основные схемы перелинковки. Теперь пора усвоить правила, которые позволят не допустить ошибок и эффективно внедрить выбранную схему.

Важно понимать, что приведенные рекомендации являются усредненными. Допустимы случаи со значительным отклонением от предложенных чисел. Например, существует рекомендация об ограничении количества исходящих ссылок со страницы. Когда мы работаем с особо крупными сайтами (новостными порталами, серьезными агрегаторами), отклонения

неизбежны. Однако рекомендация все равно имеет силу, просто необходимо будет закрывать часть неважных ссылок от индексации и не допускать дублирования их дублирования.

### Рекомендация 1: Не стоит использовать более 50 ссылок на одной странице.

Вес распределяется по всем исходящим ссылкам: чем больше ссылок, тем меньший вес передает каждая из них. Ограничивая количество исходящих ссылок, оптимизатор регулирует передаваемый вес, направляя его на нужные страницы. Так, вес страниц-шлюзов увеличивается, и они, в свою очередь, передают больший вес продвигаемым страницам.

Некоторые ссылки необходимо закрыть от индексации.

### Рекомендация 2: Ссылки нужно оформлять текстом.

Часто для сохранения фирменного стиля элементы меню, набранные в макете нестандартным шрифтом, верстают картинками или анимацией. Поисковая система в этом случае не передает веса по ключевым словам, ведь она не может определить, к какой группе запросов относится ссылка. Для корректной передачи веса рекомендуется, пусть даже в ущерб эстетическим задачам, использовать текстовые ссылки с содержанием ключевых слов. Текст должен быть оформлен на сайте в явном виде, то есть присутствовать в HTML-коде страницы.

### Рекомендация 3: Желательно не дублировать ссылки с одной страницы на другую.

Иллюстрация | [Название модели](#)

Иллюстрация | [Технические характеристики](#)

Иллюстрация | [Краткое описание модели](#)

Иллюстрация | [Подробнее...](#)

## МНЕНИЕ

### Что показывать поисковику



Часто структура сайта для пользователя входит в противоречие с SEO-логикой. Тогда нужно как-то спрятать ссылки от поисковика, оставив их для пользователя. Для этого есть несколько инструментов:

#### <noindex>

Тег <noindex> работает для Яндекса. Google его не воспринимает. Ссылка, закрытая тегом <noindex>, передает статический вес, но ее анкор не работает для Яндекса. Робот Яндекса по таким ссылкам переходит. Тег <noindex> может также использоваться внутри тега <a>, тогда мы сможем закрыть лишь часть анкора. Например:

```
<a href="http://site.ru">ноутбуки <noindex>toshiba</noindex></a>
```

#### rel=nofollow

rel=nofollow блокирует передачу как статического веса, так и работу анкора. Однако его чрезмерное использование на сайте может вызвать проблемы с Google. Летом 2009 года Google презентовал изменение в концепцию работы nofollow на внутренних ссылках.

Допустим, страница может передать 10 «поинтов» PageRank. С нее стоит 10 ссылок. 5 из них закрыты в nofollow.

Как было раньше: по 5 ссылкам, не закрытым в nofollow, передается 10 «поинтов» веса, по 2 «поинта» на каждую. То есть вес с закрытых в nofollow ссылок перераспределяется в пользу незакрытых.

Как стало: по 5 ссылкам, не закрытым в nofollow, передается лишь 5 «поинтов» веса, по 1 «поинту» на каждую. Вес больше не перераспределяется, а просто теряется.

Роботы переходят по ссылкам с nofollow.

Некоторые ссылки с nofollow все равно учитываются поисковиками. Например, ссылки с Твиттера.

#### JavaScript

Закрытие ссылки в JavaScript позволяет скрыть сам факт наличия ссылки на странице и таким образом избежать последствий для Google, описанных в предыдущем пункте.

От Google не так просто что-то скрыть. Но работающие варианты существуют:

1. Использовать безатрибутную ссылку. В этом варианте мы фактически полностью убираем href из тега <a>.
2. Использовать внешний скрипт с закрытием в robots.txt файла со скриптом. Этот вариант сложнее. В тексте html страницы мы только вызываем скрипт, который убран в отдельный js-файл. Сам же файл закрыт в robots.txt. Таким образом, бот Google не может прочитать его и восстановить содержание страницы.

Второе применение JavaScript – это скрыть ссылку от Яндекса, оставив ее видимой для Google. Иногда это может понадобиться. Сделать это несложно. Даже простейший `document.write` ставит Яндекс в тупик, и ссылка перестает работать. В то время как Google будет ее учитывать.

Особенно часто дублирование ссылок встречается в каталоге товаров. В нем обычно содержится три ссылки: изображение товара, название модели и ссылка «подробнее» в описании. Рекомендуется допускать к индексации только одну ссылку, которая, как мы помним, должна быть текстовой и содержать ключевые слова. Если продвигаете только в Яндексе, ссылку «подробнее» можно закрыть в `<NOINDEX>`. К ссылке картинкой лучше добавить атрибут `rel=nofollow`, поскольку иногда картинки могут быть полезны: при грамотной оптимизации они могут привлечь на сайт дополнительный трафик.

#### **Рекомендация 4: URL страниц должны содержать ключевые слова.**

По данным исследования «SEO-альманах», проведенного в конце 2010 года, ключевое слово в URL – один из наиболее значимых факторов ранжирования. Причем наибольшее значение имеет употребление ключевого слова в имени домена (`www.key.ru`), следующим по важности является употребление ключевого слова в названии директории или страницы. Замечено, что ключевое слово на транслите получает больший рейтинг, чем его перевод на английский.

Проще говоря, не рекомендуется использовать URL вида:

`http://www.site.ru/list.php?SECTION_ID=14,`

`http://www.site.ru/dir/content/32/.`

URL должен содержать все слова, формирующие основные запросы:

`http://www.site.ru/brand/model/modification/,`

`http://www.mobile-phone.ru/nokia/6700/classic/.`

По некоторым наблюдениям, для ранжирования конкретной страницы из разделений «/» и «->» лучше выбрать последнее. Так, последний URL рекомендуется употребить

в виде <http://www.mobile-phone.ru/nokia-6700-classic/>. Конечно, этот подход не панацея, у него есть недостатки, например, отсутствие логической структуры сайта: все страницы будут находиться в одной директории. Но для сайтов с небольшим количеством страниц такая подача URL весьма и весьма полезна.

#### **Рекомендация 5: Ссылки необходимо закрывать от индексации.**

Для правильной передачи веса, необходимо на отдельных страницах закрывать разделы меню и прочие ссылки от индексации, оставляя лишь важные ссылки, поставленные для передачи веса между страницами. К сожалению, не все CMS позволяют изменять блоки каждой страницы, используя единый шаблон. В этом случае приходится совершать доработки в программной части сайта.

### **1.5. Способы перелинковки**

#### **Тип 1. Меню**

Меню – основной способ навигации на сайте, представляющий собой блок сквозных ссылок. Как правило, главное меню дублируется на всех страницах сайта, за исключением страниц версии для печати и некоторых других. Вес, передаваемый через подобные ссылки, не складывается, а считается по иному принципу. Страницы разделов ранжируются иначе, нежели остальные страницы сайта. В этой связи можно дать несколько рекомендаций:

- ▶ Текст ссылки в меню должен содержать основные ключевые слова из продвигаемых на вложенной странице. Например, для страницы каталога мобильных телефонов текстом ссылки может быть «телефоны»/«мобильные телефоны»/«каталог телефонов», но никак не «каталог».
- ▶ Оптимальнее всего использовать иерархичное меню. Нет смысла пытаться



Меню – основной способ навигации на сайте, представляющий собой блок сквозных ссылок. Как правило, главное меню дублируется на всех страницах сайта, за исключением страниц версии для печати и некоторых других. Вес, передаваемый через подобные ссылки, не складывается, а считается по иному принципу. Страницы разделов ранжируются иначе, нежели остальные страницы сайта.

поместить ссылки сразу на все страницы сайта в главном меню. Достаточно сослаться на основные разделы. Внутри разделов целесообразно поместить вложенное меню, направляющее на страницы следующих уровней.

► Текст ссылки в меню должен соответствовать элементу URL и элементу навигационной строки вложенной страницы (если навигационная строка присутствует на сайте).

### Тип 2: Навигационная строка

Благодаря навигационной строке пользователь всегда в курсе своего местонахождения на сайте. Оптимизатору важно оформить навигационную строку правильно. Для этого необходимо учесть несколько моментов:

► В навигационной строке должна присутствовать ссылка на главную страницу, и лучше всего, если в тексте ссылки будут прописаны слова, соответствующие тематике сайта, а не просто слово «главная».

► Желательно, чтобы каждая ссылка навигационной строки содержала наиболее конкурентное ключевое слово, продвигаемое на данную страницу. Например:

[http://www.mobile-phone.ru/  
nokia/6700/classic/](http://www.mobile-phone.ru/nokia/6700/classic/)

**Сотовые телефоны –> Nokia –> Нокиа 6700 Classic**

► Фрагмент навигационной строки текущей страницы не должен быть ссылкой. Помните о том, что страница не должна ссылаться на саму себя.

### Тип 3: HTML-карта сайта

HTML-карта сайта – это раздел навигации, который помогает пользователю быстро отыскать ссылки на необходимые ему страницы сайта. HTML-карту вовсе не обязательно использовать, но, если грамотно ее оформить, можно убить двух зайцев сразу:



Благодаря навигационной строке пользователь всегда в курсе своего местонахождения на сайте. Оптимизатору важно оформить навигационную строку правильно.

сделать сайт еще удобнее для пользователя и улучшить индексацию.

При создании HTML-карты стоит придерживаться следующих принципов:

- ▶ HTML-карта не должна включать в себя более 30 ссылок на внутренние страницы сайта.
- ▶ Необходимо, чтобы HTML-карта содержала ссылки на основные разделы и наиболее важные страницы.
- ▶ HTML-карта может быть многоуровневой и состоять из нескольких страниц, если общее количество страниц сайта превышает 30.

#### Тип 4: Многоуровневый каталог

Когда счет страницам идет уже не на десятки, нельзя обойтись без рубрикатора, который упростит поиск информации. Обычно для этих целей используется каталог товаров, который является также одним из основных способов перелинковки. При организации каталога важно составить HTML-шаблон так, чтобы единственной индексруемой ссылкой осталось наименование товара, оформленное текстом (об этом мы уже говорили выше).

#### Тип 5: Подсказки

Подсказка – это рекомендация, которую дает сайт своим посетителям, как правило, это сообщение об акции, скидках или сопутствующих товарах. Для удобства пользователей подсказки оформляются с использованием ссылок, что предоставляет оптимизатору еще одно поле для маневров. При перелинковке использование подсказок – очень эффективный инструмент, особенно если продвижение ведется по низкочастотным запросам.

Часто используются блоки:

- ▶ «Похожие товары»,
- ▶ «Сопутствующие товары»,
- ▶ «Акция»,
- ▶ «Дополнительные услуги»,



Подсказка – это рекомендация, которую дает сайт своим посетителям, как правило, это сообщение об акции, скидках или сопутствующих товарах. Для удобства пользователей подсказки оформляются с использованием ссылок, что предоставляет оптимизатору еще одно поле для маневров.

## МНЕНИЕ

**Анар Бабаев, директор по развитию, Click.ru:**



— Как вы думаете, почему поисковики учитывают внутренние ссылки? Откуда это пошло?

Разработчикам алгоритмов в свое время приходилось цепляться за каждый более-менее осязаемый фактор. Но обсчет этих факторов накладывал определенные ограничения на их внедрение. Учет внутренних ссылок – не такая уж сложная процедура.

Программисты видели, как рядовые пользователи ставили внутренние ссылки. И задавались вопросом: если мы учитываем внешние ссылки, почему бы не учитывать и внутренние?

Лучшие схемы перелинковки вы найдете не в учебниках по SEO и не на сайтах оптимизаторов, а там, где владельцы сайтов думают о своих посетителях:

- Внутренние ссылки помогают продавать. Посмотрите на товарную карточку магазина «Озон». Купившие этот товар также покупают... и текстовые ссылки на похожие книги. Ссылки на авторов, отзывы, издательства, каталог. Удобно, не правда ли?
- Внутренние ссылки заставляют пользователя не уходить с сайта. Зайдите на любую страницу Википедии. Вы читаете один материал, потом интересуетесь другим, смотрите дополнительную информацию, открываете изображения и всячески взаимодействуете с сайтом. Полезно?
- Внутренние ссылки помогают не потеряться. Вы наверняка встречали на многих ресурсах облака тегов, хлебные крошки, сквозные ссылки и другие элементы навигации. Все это крайне необходимо! Такие фишки придуманы не оптимизаторами, поисковики же с удовольствием их учитывают.

Прежде чем вникать в «кольца», «пирамиды» и «звезды», подумайте о ваших посетителях. Как им было бы удобнее находить информацию на вашем сайте?

### ► «Последние поступления».

Учитывая возрастающее значение поведенческих факторов в алгоритмах ранжирования поисковых систем, мы можем предположить, что ссылка, по которой кликают, имеет большее значение, чем та, которая никому не понадобилась. Опираясь на это, мы можем сделать вывод, что хорошо оформленный блок ссылок-подсказок будет тем лучше выполнять свое «seoшное» предназначение, чем удобнее для посетителей будет навигация.

Популярные товары:	
#IMG	<a href="#">Марка + модель #1</a>
#IMG	<a href="#">Марка + модель #2</a>
#IMG	<a href="#">Марка + модель #3</a>
#IMG	<a href="#">Марка + модель #4</a>
#IMG	<a href="#">Марка + модель #5</a>

### Тип 6: Перелинковка внутри текста

Перелинковка этого типа очень просто внедряется и в уже написанные тексты, но лучше закладывать ее еще на этапе их создания, чтобы хорошо продумать структуру текстов и сделать ссылки внутри них максимально полезными посетителю страницы. В качестве примера приведем текст из Википедии, где ссылки оформлены очень удобно для пользователей.

Автоматизировать подобную перелинковку достаточно просто. Для этого нужно написать несложный скрипт, который будет искать в текстах базы данных заданные ключевые слова и по определенному

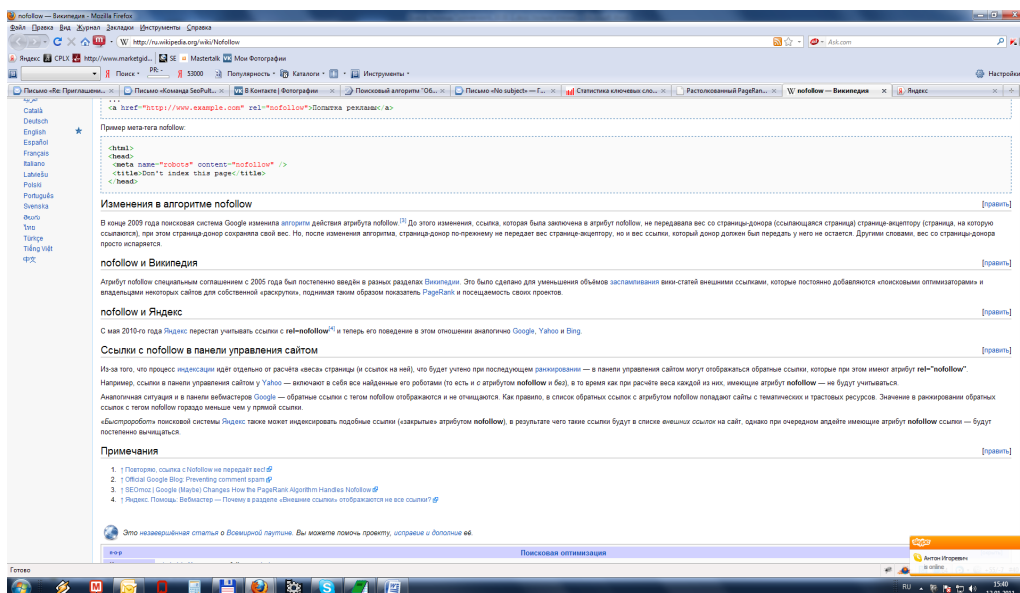


Рис. 6. <http://ru.wikipedia.org/wiki/Nofollow>

алгоритму, для некоторых из них дописывать анкоры.

### Тип 7: Перелинковка внутри текста с использованием якорей

Навигация по большому сайту тексту станет удобнее, если поставить несколько якорей вида `<a name=...`. В этом случае при перелинковке можно ставить ссылку сразу с указанием якоря в URL.

### 1.6. Составление текста ссылки для перелинковки

Правильная перелинковка позволяет управлять весом страниц, поэтому важно уделить внимание текстам ссылок. При составлении текстов ссылок для внутренней перелинковки нужно иметь четкое представление о том, какие группы запросов являются целевыми для конкретных страниц. В ваших интересах сделать так, чтобы как можно большее количество ссылок получило уникальный текст.





При перелинковке важно выдерживать баланс. Если множество ссылок содержит одинаковый текст, это множество может учитываться иначе, ссылки расцениваются как сквозные. Также рассматриваются как сквозные такие ссылки, когда каждая страница сайта ссылается на данную с различным текстом. Для перелинковки не стоит задействовать все страницы сайта.

По возможности задействуйте больше ключевых слов.

В качестве текста ссылок для коммерческих тематик рекомендуется использовать следующие конструкции:

- ▶ прямая форма запроса (например, «ноутбук Acer Aspire 5551g»),
- ▶ прямая форма запроса + образующие слова («купить ноутбук Acer Aspire 5551g»).

Кроме того, при подборе запросов важно определить часть запроса, формирующую группу. В нашем примере это – 5551g, так как все запросы, содержащие в себе этот фрагмент, будут актуальны для продвижения на данную страницу. Поэтому к рекомендуемым конструкциям мы добавляем еще следующую:

- ▶ часть запроса + образующие слова («купить Acer Aspire 5551g»).

Подобные комбинации, сгенерированные из полного шлейфа НЧ-запросов, составляют большие списки уникальных текстов ссылок для внутренней перелинковки.

При перелинковке важно выдерживать баланс. Если множество ссылок содержит одинаковый текст, это множество может учитываться иначе, ссылки расцениваются как сквозные. Также рассматриваются как сквозные такие ссылки, когда каждая страница сайта ссылается на данную с различным текстом. Для перелинковки не стоит задействовать все страницы сайта. И не нужно каждый раз использовать различные тексты ссылки: несколько повторений одного текста (до 10) вполне достаточно, если ссылка не является сквозной.

Важно выдержать определенный порог так, чтобы число повторений не вышло на уровень сквозных ссылок, иначе они могут сильно перераспределить вес внутри сайта. В «СайтРепорт» поиск

порога сквозных ссылок реализован в виде графика.



Контекстная перелинковка на страницах интернет-магазина не очень эффективна, вероятность клика по таким ссылкам низкая. А вот потенциал навигационной строки можно задействовать по полной. Каталог товаров должен быть составлен безупречно: ссылки на странице списка товаров ни в коем случае нельзя дублировать, все дубликаты нужно закрыть для индексации.

### 1.7. Создание модуля для перелинковки

Легко организовать перелинковку, если сайт состоит из 20–30 страничек, но как быть, если продвигать нужно интернет-магазин с тысячами, а то и десятками тысяч товаров? Задача непростая. Мне хватило материала, чтобы создать отдельный семинар, посвященный этому вопросу. Основных типов четыре: портал, интернет-магазин, корпоративный сайт, каталог.

#### Портал

Для перелинковки на портале можно использовать блоки:

- ▶ «Другие статьи этого автора»,
- ▶ «Другие статьи в этой категории»,
- ▶ «Другие статьи за сегодня»,
- ▶ «Самые популярные статьи за сегодня»,
- ▶ «Последние статьи».

Контекстная перелинковка с простановкой ссылок в тексте статьи на порталах воспринимается как естественная, ее тоже нужно непременно использовать (смотрите, например, как это реализовано в Википедии).

Следует понимать, что при перелинковке ссылки не должны меняться со временем, тем более не стоит проставлять их в динамических блоках, которые трансформируются при каждом обновлении страницы. Не всегда

бывает оправдано и использование навигационной строки, поскольку основной поток посетителей приходится на НЧ-запросы, а навигационная строка – инструмент повышения веса страниц разделов.

### Интернет-магазин

Для перелинковки на сайте интернет-магазина можно использовать блоки:

- ▶ «Другие товары в этой ценовой категории»,
- ▶ «С этим товаром смотрели»,
- ▶ «Сопутствующие услуги/товары»,
- ▶ «Другие товары этого производителя»,
- ▶ «Другие товары этой группы».

Контекстная перелинковка на страницах интернет-магазина не очень эффективна, вероятность клика по таким ссылкам низкая. А вот потенциал навигационной строки можно задействовать полностью.

Каталог товаров должен быть составлен безупречно: ссылки на странице списка товаров ни в коем случае нельзя дублировать, все дубликаты нужно закрыть для индексации.

Наглядный пример подобного модуля перелинковки вы можете наблюдать на сайте [www.mobiguru.ru](http://www.mobiguru.ru) (рис. 7). Здесь мы видим

Dell Inspiron M301z - просмотрите фото, описание и характеристики, обзор, новости и отзывы о ноутбуке, купите ноутбук с доставкой в самом дешевом месте, сравните цены 18 Интернет-магазинов

The screenshot shows a website interface for a laptop. At the top, there's a navigation bar with links like 'Главная / Ноутбуки / DELL / Dell Inspiron M301z'. Below that, the product name 'Ноутбук DELL Inspiron M301z' is displayed with a price range from 15900 to 23550 rubles. A search icon and a star rating are also visible. To the right, there's a technical specifications table. On the left, there's a sidebar with a 'MOBI GURU' logo and a section titled 'Популярные ноутбуки по лучшим ценам с ДОСТАВКОЙ!' listing several laptop models with their prices and the number of stores where they are available.

Марка процессора: Intel Core i5	Частота: 2530 МГц	Чипсет: Intel QM57	Тип: DDR3	Объем: 2048 МБ
Частота: 1333 МГц	Размер матрицы: 12.1"	Разрешение: 1280x800	Тип: Intel GMA HD	
Клавиатура:	Дополнительно: Touchpad	Жесткий диск: 160 Гб	LAN: 1000 Мбит/с	WLAN: 802.11n
Bluetooth:	Дополнительно: Веб-камера 1.3 Мпикс	Встроенные динамики:	Встроенный микрофон:	
VGA разъем:	USB: 3 (2.0)	Firewire:	Дополнительно: HDMI, eSATA, DisplayPort, Слоты: Memory Stick, Memory Stick PRO, SD, SDHC, xD-Picture Card	Аккумулятор: Li-Ion
			Размер: 290x212x31.7 мм	Вес: 1.7

Вся характеристика...

Обзор | Цены, где купить | Комментарии (0) | Отзывы (0)

Рис. 7. [www.mobiguru.ru](http://www.mobiguru.ru)

два блока (слева и сверху) со ссылками на карточки популярных товаров.

### Корпоративный сайт

Корпоративные сайты редко достигают таких масштабов, чтобы для перелинковки страниц потребовалась автоматизация. Если же такая необходимость возникла, лучше всего использовать контекстную перелинковку.

### Каталог

По сути, каталог – это уменьшенная модель интернет-магазина, следовательно, в продвижении каталога можно использовать все блоки, которые актуальны для последнего:

- ▶ «Другие товары в этой ценовой категории»,
- ▶ «С этим товаром смотрели»,
- ▶ «Сопутствующие услуги/товары»,
- ▶ «Другие товары этого производителя»,
- ▶ «Другие товары этой группы».

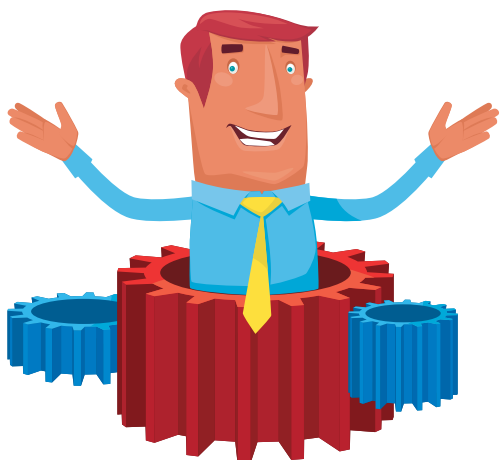


По сути, каталог – это уменьшенная модель интернет-магазина, следовательно, в продвижении каталога можно использовать все блоки, которые актуальны для последнего.

## Глава 2

# Устранение дубликатов

---



**Н**а сайте могут наличествовать дубликаты контента – страницы, полностью или частично идентичные друг другу. Одной из главных проблем поисковых систем является устранение дубликатов.

### 2.1. Чем вредны дубликаты

Скорее всего, вы уже слышали о мусорных сайтах, которые содержат автоматически созданные документы, дорвеи. Создаются дорвеи в поистине промышленных масштабах, а живут за счет продажи трафика. Наполняются такие сайты легко и быстро – ворованным контентом или сгенерированными текстами. Понятно, что для пользователей дорвеи не представляют ценности, кроме того, они занимают серверные ресурсы поисковых систем, поэтому последние стремятся исключать подобные сайты из индекса.

Проблема дубликатов схожа с проблемой дорвеев. Для поисковых систем это страницы, захламляющие Интернет, ведь они не содержат уникального контента. Значит, нет никакого смысла в их индексации.

Со стороны ПС проблема дубликатов, с одной стороны, схожа с проблемой дорвеев – ибо какой смысл индексировать не-уникальный контент. С другой стороны, не-уникальный контент – это не всегда приговор. Поговорим об этом во врезке. Однако, особенно это касается нечетких дубликатов, поисковик не может забывать, что иногда на базе не-уникального

контента возникают сервисы, обладающие так называемой «добавочной стоимостью», которой они могут наделять уникальный контент. Хорошим примером таких сервисов является рейтинг блогов и новостей.

Масштабы этой проблемы часто недооценивают. Мне встречались «прекрасные» сайты, содержащие по 40 дубликатов каждой страницы. Подобный сайт занимает у поисковой системы в 40 раз больше ресурсов, чем требуется. Казалось бы, почему не проиндексировать все? Простой подсчет: для индексации и обработки миллиарда страниц, допустим, понадобится 100 серверов, а если каждая страница имеет хотя бы по три дубликата? Как видите, решение проблемы дубликатов позволяет уменьшить затраты на серверные мощности в несколько раз.

### Чем вредны дубликаты для оптимизатора?

1. Ссылочный вес размывается.

Вместо того, чтобы увеличивать значение одной страницы, мы рассеиваем ссылки по нескольким страницам-дубликатам, в результате чего каждый из этих дубликатов получает меньший вес.

2. Входной может оказаться не та копия, на которую стоят ссылки.

Нередко бывает, что после очередного апдейта дубликат вытесняет в выдаче продвигаемую страницу, и та теряет позиции.

3. Контент теряет уникальность.

Дублированный контент понижает степень уникальности всех страниц, на которых он размещен. Страница с уникальным контентом имеет существенно больший вес.

4. Поскольку дубликаты исключаются из поиска, из него может исчезнуть нужная нам страница.

### МНЕНИЕ

Важной областью информационного поиска является поиск дубликатов документов. От того момента, как документ первый раз индексируется поисковой системой, и до того, как этот документ появится в основных результатах поиска, содержимое документа многократно оценивается и сравнивается с содержанием индекса ПС.



В борьбе с дубликатами поисковые системы могут запросто исключить из своей базы те страницы, которые мы продвигаем, и оставить дубликат, который не имеет ни входящих внешних ссылок, ни особых бонусов по внутренней оптимизации. Итак, проблема дубликатов в целом сводится к двум моментам:

- ▶ поисковые системы часто исключают из поиска значимые страницы, поскольку не способны корректно определять, какой из дубликатов имеет большую ценность;
- ▶ сайт с огромным количеством дублированного контента индексируется не полностью.

## 2.2. Классификация дубликатов

Дубли бывают полными и частичными. Полный дубликат – это абсолютно идентичные страницы, частичный – страницы, совпадающие не полностью. В зависимости от типа дубликата выбирают способ его устранения: полные дубликаты закрывают через robots.txt и 301 редирект, для устранения частичных дублей приходится делать правки на страницах.

### Полные дубликаты

#### 1. Фокусы с URL

В индексе поисковых систем должна находиться только одна версия главной страницы сайта, все ее дубликаты нужно удалить либо закрыть для индексации.

Например, мы имеем варианты такого вида:

```
http://www.site.ru/  
http://www.site.ru/index.php  
http://www.site.ru  
http://site.ru/  
https://www.site.ru/
```

Из вышеперечисленных конструкций оптимально использовать для продвижения первую (<http://www.site.ru/>).



Встречаются «прекрасные» сайтики, содержащие по 40 дубликатов каждой страницы. Подобный сайт занимает у поисковой системы в 40 раз больше ресурсов, чем требуется. Казалось бы, почему не проиндексировать все? Простой подсчет: для индексации и обработки миллиарда страниц, допустим, понадобится 100 серверов, а если каждая страница имеет хотя бы по три дубликата? Как видите, решение проблемы дубликатов позволяет уменьшить затраты на серверные мощности в несколько раз.

Для устранения остальных применяем 301 редирект и настройки robots.txt (в случае с конструкциями вида <https://www.site.ru/>).

## 2. Смена директорий

Золотое правило гласит: каждая страница должна быть доступна только по одному адресу. Нельзя, чтобы адреса варьировались таким образом:

<http://www.site.ru/dir1/dir2/>  
<http://www.site.ru/dir2/dir1/>

Пользователь может переходить по пути «фотоаппараты» → Canon или Canon → «фотоаппараты», и в любом случае адрес страницы должен быть одним, к примеру, <http://www.../fotoapparat/canon/>.

## 3. Формирование URL

При формировании URL нужно подобающим образом настроить сервер: переменных в «урле» отображаться не должно. Использование URL типа <http://www.site.ru/index.php?dir=1231&id=313> – ошибка. Корректным будет употребление <http://www.site.ru/dir/1231/id/313>.

## 4. Идентификатор сессий

Идентификатор сессий, который мы можем наблюдать в примере URL <http://www.site.ru/dir/1231/id/313.php?PHPSESSID=2123>, указывает на сайт, который фактически содержит бесконечное количество дубликатов каждой страницы. Эти страницы при очень большом везении попадают в индекс поисковых систем, создавая копии. Если сессии невозможно заменить чем-то еще, нужно закрыть их через настройки robots.txt.

## 5. Уязвимость 404 ошибки

Когда мы пытаемся открыть несуществующую страницу, то чаще всего видим заглушку. Что «видит» в этом случае сервер? Это зависит от настроек кода ответа. Если в них прописано 404, сервер решает,



Дубли бывают полными и частичными.

Полный дубликат – это абсолютно идентичные страницы, частичный – страницы, совпадающие не полностью. В зависимости от типа дубликата выбирают способ его устранения: полные дубликаты закрывают через robots.txt и 301 редирект, для устранения частичных дублей приходится делать правки на страницах.



## МНЕНИЕ

При прочих равных более уникальная страница выигрывает у страницы с менее уникальным наполнением. А сайт, содержащий много неуникальных страниц, может в какой-то момент быть признан менее авторитетным.



что такой страницы не существует. Такой подход наиболее правильный.

К сожалению, вместо этого нередко можно встретить код ответа 301 («страница перемещена навсегда») или 200 («все хорошо»). И если в первом случае страницы просто склеиваются, то во втором случае создается дубликат. Наиболее проигрышный расклад выглядит так: на сайте для несуществующей страницы отдается код ответа 200, а вместо заглушки отображается главная страница. В этом случае может быть поставлено любое количество внешних ссылок на несуществующие страницы, каждая из которых продублирует главную и размоет ее вес.

### 6. Лишние переменные

Проблема наличия лишних переменных наиболее актуальна для универсальных коммерческих CMS. Лишние переменные образуются, например, в тех случаях, когда посетитель указывает дополнительный параметр поиска, сортировки или отображения, который не влияет на содержание страницы, но может учитываться CMS. Через настройки robots.txt необходимо добиться того, чтобы URL ни в коем случае не имели вид типа `http://www.site.ru/dir/1231/id/313.php?mode=2&coin=3` (корректным будет вариант `http://www.site.ru/dir/1231/id/313.php?mode=4`).

### 7. Реф-ссылки

В электронной коммерции распространена практика создания партнерских программ, способствующих продвижению товара или услуги. Один из принципов работы такой программы заключается в том, что партнер распространяет по своим каналам ссылку, содержащую его личный ID. Регистрация каждого нового пользователя по такой ссылке приносит партнеру вознаграждение. Выглядит реф-ссылка примерно так:

<http://www.site.ru/dir/1231/id/313.php?ref=1323>. Понятно, что использование реф-ссылок становится причиной появления дубликатов. Устранить их можно также через настройки robots.txt. (в данном примере нужно дописать Disallow: \*?ref).

## Частичные дубликаты

### 1. Сортировка

На сайтах, содержащих каталог, для удобства посетителей предлагаются различные виды сортировки: по цене, по количеству отображаемых элементов на странице, по дате добавления, по рейтингу, по категории и т. п. Пользователям становится действительно удобно, а вот у оптимизаторов появляется проблема: сортировка создает мусорные для поисковых систем страницы, не содержащие никакого уникального контента. Чтобы избавиться от полудубликатов, в некоторых случаях содержимое таких страниц закрывают через метатеги noindex, nofollow, в других – к URL приписывают директорию или переменную, которая закрывается через robots.txt.

Иногда, правда, продвижение страниц сортировок в поисковых системах – единственно верная стратегия. В этом случае бороться с полудубликатами не приходится.

### 2. Частичное дублирование контента

По данным исследования «SEO-альманах 2010», проведенного в рамках проекта «ТопЭксперт.РФ», уникальность контента – один из важнейших факторов ранжирования. Некоторые оптимизаторы упорно игнорируют этот момент, в то время как на сайте есть области, таящие в себе большой потенциал для поискового продвижения. В этой связи можно дать несколько советов:

► На многих сайтах есть дубликаты в тегах Description, Title, H1, Keywords. Уникализировав их содержимое на своем сайте,



По данным исследования «SEO-альманах 2010», проведенного в рамках проекта «ТопЭксперт.РФ», уникальность контента – один из важнейших факторов ранжирования. Некоторые оптимизаторы упорно игнорируют этот момент, в то время как на сайте есть области, таящие в себе большой потенциал для поискового продвижения.

вы сделаете отличный вклад в продвижение ресурса.

- ▶ Весь дублирующийся контент в блоках типа «Схожие предложения» необходимо скрывать через <NOINDEX>, оставляя доступными для поисковых роботов лишь ссылки и наименования товаров.
- ▶ Страницы списка анонсов должны содержать уникальный контент, а не формироваться из текстов, частично дублирующих полное содержание статей.
- ▶ Схожие страницы типа «Оставить заявку» вообще не имеют значения, а потому индексироваться не должны.
- ▶ Если контента на сайте нет или его мало, стоит прибегнуть к услугам копирайтеров, или сгенерировать тексты машинным путем, или подумать о том, как привлечь поток User Generated Content.

Рассмотрим некоторые частные случаи.

## 2.3. Кейсы

### 1. Компания «БЕСТ-Недвижимость», [www.best-realty.ru](http://www.best-realty.ru)

По данным Яндекс.Вебмастер, на момент анализа всего роботом загружено 37 684 страниц, а для поиска из них доступно лишь 795. Чем объяснить разрыв в более чем 50 раз между количеством загруженных и доступных для поиска страниц? Почему Яндекс исключил около 36 тысяч страниц из своей базы?

На момент проведения анализа на сайте компании было обнаружено большое количество дубликатов:

#### 1.1. Дубликаты сортировки

На [www.best-realty.ru](http://www.best-realty.ru) используется сложная система обработки баз данных, которая все переменные помещает в URL, и в результате последний выглядит следующим образом:



Если контента на сайте нет или его мало, стоит прибегнуть к услугам копирайтеров, или сгенерировать тексты машинным путем, или подумать о том, как привлечь поток User Generated Content.

```
/services/buy/flat/?search_center_lat=0&search_center_lng=0&search_radius=0&halt_id=34&search_metro=0&is_moscow=true&search_city=0&search_fmt=0&search_mfb=1&search_price=0&search_currency=1&search_proc=10&search_s_min=13&search_s_max=9936&search_ls_min=8&search_ls_max=6000&search_ks_min=1&search_ks_max=52&go=&go2=false.
```

При малейших изменениях в настройках отображения результатов поиска URL трансформируется при сортировке по цене, алфавиту и т. д. CMS работает таким образом, что возникает большое количество дубликатов, вследствие чего некоторые страницы выпадают из индекса. Для решения этой проблемы инструкциями в robots.txt не обойтись. Возникает необходимость внести правки в саму структуру сайта, но изменить существующую схему формирования URL невозможно из-за специфики работы базы данных. Поэтому на сайте [www.best-realty.ru](http://www.best-realty.ru) проблема с дубликатами решена нестандартно:

- ▶ все страницы, содержащие переменные, закрыты от индексации;
- ▶ создан дублирующий раздел, построенный на ЧПУ («человекопонятном урле») с содержанием ключевых слов в адресах страниц.

В результате нам удалось добиться устранения всех дубликатов. Теперь URL сайта имеют вид:

<http://www.best-realty.ru/buy/flat/moscow/aviamotornaya/>

[http://www.best-realty.ru/buy/flat/moscow/krasnokazarmennaya\\_ulitsa/id1120801/](http://www.best-realty.ru/buy/flat/moscow/krasnokazarmennaya_ulitsa/id1120801/)

### 2.1. «Посмотреть другие объекты в радиусе 1 км»

На странице каждого объекта размещена ссылка «Посмотреть другие объекты



При малейших изменениях в настройках отображения результатов поиска URL трансформируется при сортировке по цене, алфавиту и т. д. CMS работает таким образом, что возникает большое количество дубликатов, вследствие чего некоторые страницы выпадают из индекса. Для решения этой проблемы инструкциями в robots.txt не обойтись.



Компания «БЕСТ-Недвижимость» – один из крупнейших риелторов на российском рынке, ее сайт содержит информацию о нескольких тысячах объектов. На странице каждого из них есть ссылка «Оставить заявку на объект», которая направляет посетителя на страницу заявки. Все страницы этого рода идентичны, но каждая имеет уникальный URL, в результате чего на сайте появляется несколько тысяч дубликатов.

в радиусе 1 км». Понятно, что подобная страница дублирует контент, поскольку создает дополнительную страницу сортировки.

В результате преобразований, описанных в п. 1, эта проблема была также устранена.

### 3.1. «Оставить заявку на объект»

Компания «БЕСТ-Недвижимость» – один из крупнейших риелторов на российском рынке, ее сайт содержит информацию о нескольких тысячах объектов. На странице каждого из них есть ссылка «Оставить заявку на объект», которая направляет посетителя на страницу заявки. Все страницы этого рода идентичны, но каждая имеет уникальный URL, в результате чего на сайте появляется несколько тысяч дубликатов.

Проблема решена очень просто: все подобные страницы запрещены к индексации в robots.txt.

## 2. ММЦ «ОнКлиник», [www.onclinic.ru](http://www.onclinic.ru)

После анализа сайта выяснилось, что полезных страниц всего около 4 500. При этом реально на сайте существовало на момент исследования порядка 95 000 страниц. Дубликаты появились из-за ошибок, допущенных разработчиками сайта.

### 2.1. Разделы «Консультации», «Термины», «Отзывы»

На [www.onclinic.ru](http://www.onclinic.ru) существуют 3 дополнительных информационных раздела: «Консультации», «Термины», «Отзывы». В эти разделы посетитель может попасть двумя путями: посетив их полную версию через главное меню сайта или посетив разделы медицинских услуг, где эти блоки представлены частями (например, в разделе «Стоматология» вложены блоки, в которых представлены консультации, расшифровка терминов и отзывы только по этой группе медицинских услуг). Каждый путь

имеет свой URL, из-за этого на сайте появились дубликаты.

Для решения проблемы полные версии разделов «Консультации», «Термины», «Отзывы» закрыты от индексации в robots.txt. Индексируемыми остались только их части, привязанные к конкретным медицинским направлениям.

## 2.2. Список анонсов

Элементы разделов «Консультации», «Термины», «Отзывы» состоят из двух частей: вопрос-ответ, автор-отзыв, термин-толкование. Структура этих разделов состоит из страниц списка анонсов и страниц полного контента. Однако содержание каждого элемента представлено в полном объеме как на странице списка анонсов, так и на странице полного контента. В результате контент в полном объеме дублируется между страницами сайта.

Проблема решена следующим образом: страницы полного контента удалены за ненадобностью.

## 3. Магазин итальянской мебели, [www.eurointerier.ru](http://www.eurointerier.ru)

В индексную базу попало более 12 500 страниц, однако на сайте всего около 1 000 страниц, полезных для посетителей, все остальные – дубликаты, которые появились частично вследствие недоработок на этапе создания сайта, а частично позже – в результате ошибок, допущенных веб-мастером в ходе внесения изменений на сайте.

Этот случай – яркий пример того, как после доработки сайта под SEO, он становится еще менее пригодным для продвижения. Хочу снова подчеркнуть, что создание ТЗ на доработку сайта и контроль его исполнения, а также повторный анализ сайта – прямая обязанность квалифицированного SEO-специалиста. В моей практике было всего два



Существует два простых способа борьбы с дубликатами: использование 301 редиректа и инструкции в robots.txt. К сожалению, простое решение позволяет свести на нет только проблему полных дубликатов, и то не всегда: если у вас крупный сайт, прописывать сотни строк в инструкции robots.txt, чтобы закрыть от индексации десятки, а возможно, и сотни тысяч страниц, не лучший выход. Хотя можно запретить с помощью метатегов.

случая, когда в процессе доработки сайта, веб-программист не создавал новых ошибок. Дело в том, что у веб-программиста есть свое видение хорошего сайта, которое часто отличается от видения поисковой системы. А поскольку специальной подготовки в области SEO у программиста нет, он не в состоянии самостоятельно сделать сайт совершенным в плане продвижения.

Итак, в случае [www.eurointerier.ru](http://www.eurointerier.ru) стояла задача заменить URL на ЧПУ. После проведения доработок каждая из страниц была доступна по трем адресам:

<http://www.eurointerier.ru/index.php?catalog=1&brand=34&model=231>

<http://www.eurointerier.ru/catalog/1/brand/34/model/231/>

[http://www.eurointerier.ru/catalog/gostinie/a\\_r\\_arredamenti/barnaya\\_stoyka\\_opera/](http://www.eurointerier.ru/catalog/gostinie/a_r_arredamenti/barnaya_stoyka_opera/)

Так, все страницы были доступны при изменении порядка частей URL:

[http://www.eurointerier.ru/catalog/a\\_r\\_arredamenti/gostinie/barnaya\\_stoyka\\_opera/](http://www.eurointerier.ru/catalog/a_r_arredamenti/gostinie/barnaya_stoyka_opera/)

[http://www.eurointerier.ru/catalog/a\\_r\\_arredamenti/barnaya\\_stoyka\\_opera/gostinie/](http://www.eurointerier.ru/catalog/a_r_arredamenti/barnaya_stoyka_opera/gostinie/)

[http://www.eurointerier.ru/catalog/gostinie/a\\_r\\_arredamenti/barnaya\\_stoyka\\_opera/](http://www.eurointerier.ru/catalog/gostinie/a_r_arredamenti/barnaya_stoyka_opera/)

[http://www.eurointerier.ru/catalog/gostinie/barnaya\\_stoyka\\_opera/a\\_r\\_arredamenti/](http://www.eurointerier.ru/catalog/gostinie/barnaya_stoyka_opera/a_r_arredamenti/)

[http://www.eurointerier.ru/catalog/barnaya\\_stoyka\\_opera/a\\_r\\_arredamenti/gostinie/](http://www.eurointerier.ru/catalog/barnaya_stoyka_opera/a_r_arredamenti/gostinie/)

[http://www.eurointerier.ru/catalog/barnaya\\_stoyka\\_opera/gostinie/a\\_r\\_arredamenti/](http://www.eurointerier.ru/catalog/barnaya_stoyka_opera/gostinie/a_r_arredamenti/)



Кстати, в панели Яндекс.Вебмастера есть весьма полезный инструмент, проверяющий корректность составления robots.txt (на данный момент Clean-param он не читает, лишь Disallow). Добавляете сайт в панель, затем щелкаете «Настройка индексирования» и «Анализ robots.txt». После чего вам требуется в «Список URL» добавить URL, которые нуждаются в проверке: будут ли они индексироваться при введенном robots.txt или не будут.

Страницы карточки товара отображались даже после удаления части URL с названием товарной группы или фабрики:

[http://www.eurointerier.ru/catalog/a\\_r\\_arredamenti/gostinie/barnaya\\_stoyka\\_opera/](http://www.eurointerier.ru/catalog/a_r_arredamenti/gostinie/barnaya_stoyka_opera/)

[http://www.eurointerier.ru/catalog/a\\_r\\_arredamenti/barnaya\\_stoyka\\_opera/](http://www.eurointerier.ru/catalog/a_r_arredamenti/barnaya_stoyka_opera/)

[http://www.eurointerier.ru/catalog/gostinie/barnaya\\_stoyka\\_opera/](http://www.eurointerier.ru/catalog/gostinie/barnaya_stoyka_opera/)

[http://www.eurointerier.ru/catalog/barnaya\\_stoyka\\_opera/](http://www.eurointerier.ru/catalog/barnaya_stoyka_opera/)

Прибегнув к элементарным арифметическим подсчетам, получаем, что каждая страница карточки товара имела 30 дубликатов. Кроме того, на сайте были дубли сортировки и частичные дубликаты в каталоге.

Подобные проблемы решаются доработкой CMS, цель которых проста: каждая страница должна быть доступна только по одному адресу. В случае изменения правил формирования URL необходимо использовать 301 редирект. Страницы сортировки закрываются от индексации.

## 2.4. Способы устранения дубликатов

Существует два простых способа борьбы с дубликатами: использование 301 редиректа и инструкций в robots.txt. К сожалению, простое решение позволяет свести на нет только проблему полных дубликатов, и то не всегда: если у вас крупный сайт, прописывать сотни строк в инструкции robots.txt, чтобы закрыть от индексации десятки, а возможно, и сотни тысяч страниц, не лучший выход. Хотя можно заpretить с помощью метатегов.

Сложное решение сводится к доработке сайта: от использования метатегов

## МНЕНИЕ

Даниил Маул:

— Действительно, по наличию дублей в индексе порой можно отличить оптимизатора-новичка от профессионала. Проверку на дубли проще всего осуществить следующим образом. Если у вас интернет-магазин, имеющий плюс ко всему раздел статей, то заходим на разные типы страниц (категория товаров, товары, статьи, главная), копируем от каждого из типов кусок текста (в рамках одного предложения, допустим, несколько слов) и вбиваем его в виде цитаты с кавычками в поиск по вашему домену в Яндекске.



Если же в индексе вашего сайта в пределах XXX страниц, то желательнее обойти весь список страниц, которые видит Яндекс. Если в этом списке видны дубли или же просто бесконтентные страницы, то их нужно или удалить, или поставить запрет в robots.txt. Запрет в этом файле можно сделать двумя методами: или Disallow, или Clean-robot (второй метод более новый). Как говорит саппорт Яндекска, срабатывает Clean-robot медленнее. Ну, а действие Disallow будет заметно уже после первой переиндексации и последующего апдейта Яндекска.

Кстати, в панели Яндекс.Вебмастера есть весьма полезный инструмент, проверяющий корректность составления robots.txt (на данный момент Clean-robot он не читает, лишь Disallow). Добавляете сайт в панель, затем щелкаете «Настройка индексирования» и «Анализ robots.txt». После чего вам требуется в «Список URL» добавить URL, которые нуждаются в проверке: будут ли они индексироваться при введенном robots.txt или не будут.

Впрочем, должен заметить, что если анонс, например, статья дублируется в категории статей в виде нескольких предложений, то, на мой взгляд, это ничего страшного не несет. Разумеется, если сами статьи объемные. Например, на своем блоге maulnet.ru я до сих пор не решился на страницах вида page/x запретить индексацию анонсов статей.

Озаботиться стоит не только отсутствием дублей, но и все тем же ЧПУ. Я знаю случай, когда у человека, использовавшего CMS Joomla и мод к нему, Яндексом индексировался весь сайт, кроме контента, который генерил этот мод. Мод генерил, собственно, сам магазин (товары). Все из-за того, что товары располагались на кривых и довольно длинных URL. Затем этот пользователь сделал их вида «url.ru/nazvanie-tovara.html», и товары начали индексироваться.





**Важный момент:** дубликат должен быть не только «индексируемым», но и «юзабельным», то есть им действительно должны пользоваться посетители сайта.

noindex, nofollow на странице и скрывтия части страницы тегом <NOINDEX> до переработки структуры базы данных и реструктуризации сайта. Фактически мы делаем следующее:

1. Удаляем мусорный дублирующий контент.
2. Закрываем от индексации дублирующийся, но важный для удобства пользователей контент.
3. Если удалить разделы невозможно, дублируем их вместе с содержимым, а затем скрываем от индексации. Важный момент: дубликат должен быть не только «индексируемым», но и «юзабельным», то есть им действительно должны пользоваться посетители сайта. Пример такой доработки рассмотрен выше – [www.best-realty.ru](http://www.best-realty.ru). Такие приемы, как раздел «Статьи», запрятанный куда-нибудь в подвал и обозначенный мелким шрифтом, сейчас не работают (за исключением случаев, когда этот раздел существует уже на протяжении нескольких лет).
4. Используем robots.txt, чтобы закрыть лишние страницы от робота-индексатора.

\*Проверить, как будет вести себя робот Яндекса на сайте, можно с помощью специальной утилиты из набора Яндекс.Вебмастер <http://webmaster.yandex.ru/robots.xml>

В результате этих манипуляций на сайте создается не содержащий дублей маршрут для робота. Хорошей работой можно назвать такой маршрут, который посетители сайта тоже используют.

Должен поделиться своим наблюдением: эффективные сайты с высокой видимостью по шлейфу НЧ-запросов реже исключают страницы в robots.txt; чаще дорабатывается сам сайт. Это связано с тем, что в результате скрывтия страниц в robots.txt вес на них уходит, но обратно не возвращается, ссылки с них также

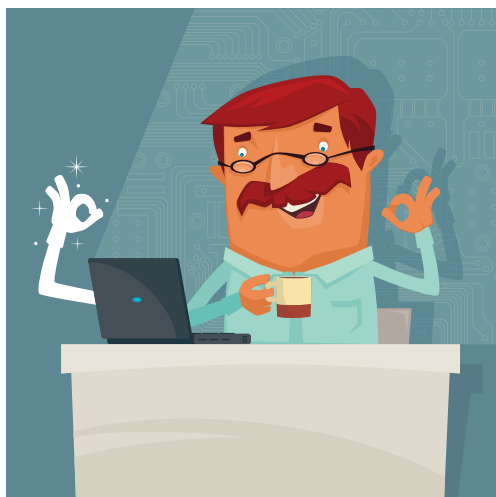
перестают работать. Сайт, структура которого хороша настолько, что нет нужды закрывать страницы в robots.txt, ранжирется лучше.

***РЕЗЮМЕ:** Мы рассмотрели основные виды дублирования контента и способы борьбы с ними. К сожалению, все случаи разобрать невозможно: вариантов очень много, и каждый из них может иметь свои особенности. Более подробно поднятую в данной главе тему я освещаю в рамках семинара «Оптимизация сайта: задействуем внутренние факторы».*

## Глава 3

# Инструкции для поисковых систем

---



**В** этой главе мы разберемся, как составлять инструкции robots.txt и sitemap.xml для поисковых систем. Предназначение этих инструкций – помочь поисковой системе отграничить важные страницы от мусорных, чтобы скорректировать робота, индексирующего сайт.

### 3.1. Зачем нужен robots.txt

Стандарт robots.txt был принят консорциумом W3C 30 января 1994 года. Эта инструкция используется для исключения дубликатов и других мусорных страниц из индекса поисковых систем. Кроме того, именно через robots.txt можно указать Яндексу главное зеркало сайта и адрес карты сайта.

В некоторых случаях к robots.txt прибегают для того, чтобы закрыть сайт от нежелательной поисковой системы. Например, если вы хотите скрыть от ваших конкурентов информацию о сайтах, разместивших ссылку на ваш веб-ресурс, вы можете закрыть сайт от индексации поисковой системой Yahoo.

Итак, файл robots.txt выполняет следующие функции:

- ▶ указание главного зеркала,
- ▶ исключение лишних страниц из индекса,
- ▶ отказ в индексации нежелательным роботам,
- ▶ указание адреса карты сайта.

## 3.2. Robots.txt – универсальные директивы

Существуют универсальные директивы, которые воспринимают любые поисковые системы:

### 1. Директива User-Agent

Иногда, учитывая особенности алгоритмов ранжирования, целесообразно использовать для разных поисковых систем различные маршруты обхода сайта. Для этого существует директива User-Agent.

User-Agent указывает имя робота, которому адресована инструкция (тогда директива имеет вид 'User-agent: botname'), либо сообщает, что инструкция предназначена для всех поисковых систем (выглядит она как 'User-agent: \*'). Робот botname в первую очередь обращает внимание на персональную инструкцию, если таковая имеется, и следует ее указаниям, игнорируя все остальные. Если же персональной инструкции для робота конкретной поисковой системы нет, он руководствуется 'User-agent: \*'.

### 2. Директива Disallow

Disallow – основная директива. Используется она для запрета индексации. Указание 'Disallow: /dir/' налагает вето на индексацию всех страниц, расположенных по адресу [www.site.ru/dir/](http://www.site.ru/dir/). Важно указывать адрес непосредственно от корневой директории сайта. Конструкция вида 'Disallow: http://www.site.ru/' некорректна и воспринимается как <http://www.site.ru/http://www.site.ru/>.

### 3. Директива Sitemap

Sitemap указывает адрес xml-карты сайта. Если эта директива не заполнена, то поисковые системы по умолчанию ищут файл [sitemap.xml](http://www.site.ru/sitemap.xml) по адресу [www.site.ru/sitemap.xml](http://www.site.ru/sitemap.xml). Директив Sitemap на одном сайте может быть несколько, и они бывают вложенными



В некоторых случаях к robots.txt прибегают для того, чтобы закрыть сайт от нежелательной поисковой системы. Например, если вы хотите скрыть от ваших конкурентов информацию о сайтах, разместивших ссылку на ваш веб-ресурс, вы можете закрыть сайт от индексации поисковой системой Yahoo.

во внутренние директории, но их непременно нужно указать в robots.txt.

### 3.3. Robots.txt – директивы для Яндекса

Чтобы веб-мастерам и оптимизаторам было удобнее, Яндекс ввел дополнительные директивы для указаний маршрута поискового робота. В частности, существуют директивы:

#### 1. Allow

Эта директива указывает поисковой системе страницы, которые нужно проиндексировать обязательно. Ее удобство заключается в том, что на некоторых сайтах так сложно искать все мусорные страницы, что проще вместо этого указать страницы полезные. В этом случае весь сайт закрывается от индексации через директиву 'Disallow: /', а инструкциями 'Allow: ...' робот направляется на полезные страницы.

Часто директива Allow используется для указания важных страниц в закрытой от индексации директории. Также ее можно применять, когда нужно закрыть все внутренние страницы раздела, но оставить индексируемой собственно главную страницу. В этом случае уместна такая конструкция:

```
User-agent: Yandex  
Disallow: /dir/  
Allow: /dir/$
```

#### 2. Clean-Param

Директива Clean-param склеивает все страницы с переменной в одну:

```
www.site.ru/some_dir/get_book.  
pl?ref=site_1&book_id=123  
  
www.site.ru/some_dir/get_book.  
pl?ref=site_2&book_id=123  
  
www.site.ru/some_dir/get_book.  
pl?ref=site_3&book_id=123.
```

'Clean-param: ref/some\_dir/get\_book.  
pl' указывает срез для всех страниц,



Часто директива Allow используется для указания важных страниц в закрытой от индексации директории. Также ее можно применять, когда нужно закрыть все внутренние страницы раздела, но оставить индексируемой собственно главную страницу.

содержащих в URL что-то еще. В этом случае проиндексируется только страница [www.site.ru/some\\_dir/get\\_book.pl](http://www.site.ru/some_dir/get_book.pl).

Директива Clean-param полезна для устранения дубликатов, связанных с сортировкой и реферальными ссылками.

### 3. Host

Если сайт компании доступен по нескольким адресам, с помощью инструкции 'Host: www.site.ru' можно указать основное зеркало. Именно этот сайт будет отображаться в результатах поиска, а все второстепенные зеркала будут исключены из основной базы робота-индексатора. Если у вас всего один сайт, в директиве Host нужно указать, какое из зеркал сайта является основным: www.site.ru или site.ru. В случае когда такой инструкции нет, Яндекс может проиндексировать страницы обоих зеркал.

### 4. Спецсимволы \* и \$

Спецсимвол \* означает любую последовательность символов. Например, директива 'Disallow: /\* /search/' указывает поисковому роботу на необходимость индексации всех страниц, содержащих /search/.

Спецсимвол \$ означает строгое соответствие. Так, 'Disallow: /catalog/\$' демонстрирует поисковому роботу запрет на индексацию страницы [www.site.ru/catalog/](http://www.site.ru/catalog/), при этом вложенные страницы, например, [www.site.ru/catalog/page\\_1.php](http://www.site.ru/catalog/page_1.php), будут проиндексированы.

## 3.4. Правила составления robots.txt

Неопытные оптимизаторы при составлении robots.txt очень часто делают ошибки. Чтобы не наступать на чужие грабли, нужно хорошо усвоить следующую информацию:

- ▶ количество директив в файле не должно превышать 1 024,
- ▶ максимальная длина 1 строки – 1 024 символа,



Если сайт компании доступен по нескольким адресам, с помощью инструкции 'Host: www.site.ru' можно указать основное зеркало. Именно этот сайт будет отображаться в результатах поиска, а все второстепенные зеркала будут исключены из основной базы робота-индексатора.

## МНЕНИЕ

### Особенность Google в Robots.txt

Если ваш URL имеет конструкцию вида `http://site.ru/?price [500]`, то директива `Disallow: *` [\* не уберет его из выдачи Google. (При этом в валидаторе все будет правильно). Почему? Google некорректно воспринимает некоторые символы. Например «[». В выдаче ваш URL будет выглядеть так: `http://site.ru/?price%5B500%5D`. Поэтому, чтобы исключить его из выдачи, необходимо вписать в robots такую строку:

```
Disallow: *%5B*
```

Квадратные скобки – не единственный случай. Всегда нужно смотреть на вид URL в выдаче.



- ▶ инструкция должна начинаться с User-agent,
- ▶ robots.txt может содержать несколько инструкций,
- ▶ `http://основное_зеркало_сайта/robots.txt` должен отдавать код ответа 200. Никакой другой код ответа не воспринимается поисковым роботом. Если по этому адресу происходит перенаправление через редирект, поисковая система не станет учитывать robots.txt,
- ▶ спецсимвол # в начале инструкции (`# Disallow: /`) отключает ее.

## 3.5. Разбор robots.txt на примерах

### Пример 1. `http://mobiguru.ru/`

На момент написания этой главы на сайте `http://mobiguru.ru/` была размещена следующая инструкция robots.txt:

```
User-agent: Yandex
Disallow: /search/
Disallow: /js
Disallow: /*?_openstat
Disallow: /*?loc
Disallow: /*/?*
Disallow: /*sort/price/
Disallow: /*sort/rate/
Disallow: /guru_*
Host: mobiguru.ru
Sitemap: http://mobiguru.ru/
mobiguru.xml
```

Попробуем разобраться в том, что хотели сообщить поисковым роботам оптимизаторы.

Итак, мы видим, что инструкция предназначена для поисковой системы Яндекс. Для других поисковых систем инструкций не предусмотрено, следовательно, они индексируют все страницы сайта.

Инструкция закрывает от индексации раздел поиска (search), страницы сортировки (sort) и некоторые переменные. Основное зеркало сайта – mobiguru.ru. Карта сайта расположена по адресу <http://mobiguru.ru/mobiguru.xml>.

### Пример 2. <http://www.last.fm/>

Любопытная инструкция размещена на сайте last.fm

```
User-Agent: *
Disallow: /music?
Disallow: /widgets/radio?
Disallow: /show_ads.php
Disallow: /affiliate/
Disallow: /affiliate_redirect.php
Disallow: /affiliate_sendto.php
Disallow: /affiliatelink.php
Disallow: /campaignlink.php
Disallow: /delivery.php
Disallow: /music/+noredirect/
Disallow: /harming/humans
Disallow: /ignoring/human/orders
Disallow: /harm/to/self
Allow: /
```

Инструкция предназначена для всех роботов. От индексации закрыты некоторые внутренние страницы, несколько страниц с переменными и одна директория. Три последние инструкции Disallow вызывают особый интерес:

1. Disallow: /harming/humans («Запрещается: причинять вред людям»)
2. Disallow: /ignoring/human/orders («Запрещается: игнорировать человеческие приказы»)
3. Disallow: /harm/to/self («Запрещается: причинять вред себе»)



Если основное предназначение robots.txt – это запрет индексации, то карта сайта выполняет прямо противоположные задачи. Она отвечает за ускорение индексации сайта и полноту его индексации.



Эти указания могут показаться бессмыслицей для каждого, кто не знаком с тремя законами робототехники, которые сформулировал в рассказе «Хоровод» знаменитый писатель-фантаст Айзек Азимов. Законы гласят:

1. Робот не может причинить вред человеку или своим бездействием допустить, чтобы человеку был причинен вред.
2. Робот должен повиноваться всем приказам, которые дает человек, кроме тех случаев, когда эти приказы противоречат Первому Закону.
3. Робот должен заботиться о своей безопасности в той мере, в которой это не противоречит Первому и Второму Законам.

### 3.6. Зачем нужен sitemap.xml

Если основное предназначение robots.txt – это запрет индексации, то карта сайта выполняет прямо противоположные задачи. Она отвечает за:

- ▶ ускорение индексации сайта,
- ▶ полноту индексации сайта.

Sitemap.xml указывает поисковой системе частоту, с которой возникает необходимость в переиндексации страниц. В этом плане инструкция особенно важна для сайтов с регулярно обновляющимся контентом (новостные порталы и т. п.).

Кроме того, sitemap.xml содержит все важные страницы сайта с указанием их приоритета.

### 3.7. Директивы sitemap.xml

**Обязательные атрибуты**

```
<urlset>
```

Атрибут инкапсулирует этот файл и указывает стандарт текущего протокола.

```
<url>
```



Sitemap.xml указывает поисковой системе частоту, с которой возникает необходимость в переиндексации страниц. В этом плане инструкция особенно важна для сайтов с регулярно обновляющимся контентом (новостные порталы и т. п.).

Это родительский тег для каждой записи URL. Остальные теги являются для него дочерними.

<loc>

Представляет URL-адрес страницы. Этот URL должен начинаться с префикса (например, HTTP) и заканчиваться косой чертой, если ваш веб-сервер требует этого. Длина этого значения не должна превышать 2 048 символов.

### Необязательные атрибуты

<lastmod>

Здесь в формате W3C Datetime указывается дата последнего изменения файла. W3C Datetime позволяет при необходимости опустить сегмент времени и использовать формат ГГГГ-ММ-ДД. Обратите внимание: этот тег не имеет отношения к заголовку 'If-Modified-Since (304)', который может вернуть сервер, поэтому поисковые системы иногда по-разному используют информацию из этих двух источников.

<changefreq>

Атрибут задает вероятную частоту изменения страницы. Это значение предоставляет общую информацию для поисковых систем и может не соответствовать в точности частоте сканирования этой страницы. Допустимые значения:

always,  
hourly,  
daily,  
weekly,  
monthly,  
yearly,  
never.

Значение «всегда» (always) должно использоваться для описания документов, которые изменяются при каждом доступе к ним,



Значение «всегда» (always) должно использоваться для описания документов, которые изменяются при каждом доступе к ним, значение «никогда» (never) – для описания архивных URL-адресов.



Значение `<priority>` не влияет на процедуру сравнения ваших страниц со страницами других сайтов, оно только позволяет указать поисковым системам, какие страницы более важны для сканеров, на ваш взгляд.

значение «никогда» (never) – для описания архивных URL-адресов.

Имейте в виду, что значение для этого тега рассматривается как подсказка, а не как команда. Несмотря на то, что сканеры поисковой системы учитывают эту информацию при принятии решений, они могут сканировать страницы с пометкой «ежечасно» менее часто, чем указано, а страницы с пометкой «ежегодно» – чаще одного раза в год. Бывает, что поисковые роботы сканируют страницы с пометкой «никогда», чтобы отслеживать неожиданные изменения на этих страницах.

`<priority>`

Указывается приоритетность данного URL относительно других URL на вашем сайте. Допустимый диапазон значений – от 0,0 до 1,0; по умолчанию приоритетность равна 0,5.

Значение `<priority>` не влияет на процедуру сравнения ваших страниц со страницами других сайтов, оно только позволяет указать поисковым системам, какие страницы более важны для сканеров, на ваш взгляд. Таким образом, вы не сможете повлиять на положение ваших URL на страницах результатов какой-либо поисковой системы. Однако поисковые системы используют эту информацию при обработке URL, которые относятся к одному и тому же сайту. Поэтому можно использовать этот тег для увеличения вероятности присутствия в поисковом индексе самых важных страниц вашего сайта.

Вы должны понимать, что нет смысла назначать высокий приоритет всем URL вашего сайта. Это очень относительная величина, поэтому параметр используется лишь для того, чтобы определить очередность обработки URL в пределах одного сайта.

### 3.8. Правила составления sitemap.xml

Существует несколько правил по составлению инструкции sitemap.xml:

- ▶ Размер sitemap.xml не должен превышать 10 Мб.
- ▶ Файл sitemap.xml может быть разбит на несколько отдельных файлов, каждый из которых должен быть указан в robots.txt.
- ▶ Файл sitemap.xml, расположенный в некотором каталоге, должен включать только URL, находящиеся в этом же каталоге либо его вложенных разделах. Нельзя, чтобы он включал в себя адреса из другого (из данного и не вложенного в данный) каталога сайта. То есть в sitemap.xml, расположенном по адресу [http://www.site.ru/dir\\_1/](http://www.site.ru/dir_1/), будут учитываться только инструкции для вложенных в эту категорию страниц. Инструкции, касающиеся страниц, вложенных, например, в категорию [http://www.site.ru/dir\\_2/](http://www.site.ru/dir_2/), будут проигнорированы.
- ▶ Файл sitemap.xml не должен содержать более 50 000 URL.
- ▶ Нельзя использовать более 1 000 файлов sitemap.xml.

Эти правила очень просты, но возможностей sitemap.xml действительно хватит, чтобы указать информацию о 50 миллионах страниц сайта, а этого более чем достаточно.

### 3.9. Пример sitemap.xml

Рассмотрим на примере все того же [mobicuru.ru](http://mobicuru.ru), как составлять sitemap.xml. Всю карту разбирать не будем, нам достаточно небольшой ее части.

```
<urlset>
<url>
<loc>http://mobicuru.ru/</loc>
<changefreq>hourly</changefreq>
```



Файл sitemap.xml, расположенный в некотором каталоге, должен включать только URL, находящиеся в этом же каталоге либо его вложенных разделах. Нельзя, чтобы он включал в себя адреса из другого (из данного и не вложенного в данный) каталога сайта.

```
</url>
<url>
<loc>http://mobiguru.ru/phones/</loc>
<changefreq>hourly</changefreq>
</url>
<url>
<loc>http://mobiguru.ru/photo/</loc>
<changefreq>hourly</changefreq>
</url>
<url>
<loc>http://mobiguru.ru/notebooks/
</loc>
<changefreq>hourly</changefreq>
</url>
<url>
<loc>
http://mobiguru.ru/phones/nokia/
nokia_5800_xpressmusic.html
</loc>
<changefreq>daily</changefreq>
</url>
```

...

Указанный отрезок карты сайта инструктирует поисковые системы, что страницы <http://mobiguru.ru/>, <http://mobiguru.ru/phones/>, <http://mobiguru.ru/photo/>, <http://mobiguru.ru/notebooks/> должны индексироваться ежечасно. Частота переиндексации страницы [http://mobiguru.ru/phones/nokia/nokia\\_5800\\_xpressmusic.html](http://mobiguru.ru/phones/nokia/nokia_5800_xpressmusic.html) – ежедневно.

...

```
<sitemapindex>
<sitemap>
<loc>http://www.gstatic.com/s2/sitemaps/
sitemap-000.txt</loc>
```



Файл `sitemap.xml` не должен содержать более 50 000 URL.

```
<lastmod>2010-07-21</lastmod>
</sitemap>
<sitemap>
<loc>http://www.gstatic.com/s2/sitemaps/sitemap-001.txt</loc>
<lastmod>2010-07-21</lastmod>
</sitemap>
<sitemap>
<loc>http://www.gstatic.com/s2/sitemaps/sitemap-002.txt</loc>
<lastmod>2010-07-21</lastmod>
</sitemap>
<sitemap>
<loc>http://www.gstatic.com/s2/sitemaps/sitemap-003.txt</loc>
<lastmod>2010-07-21</lastmod>
</sitemap>
<sitemap>
<loc>http://www.gstatic.com/s2/sitemaps/sitemap-004.txt</loc>
<lastmod>2010-07-21</lastmod>
</sitemap>
<sitemap>
<loc>http://www.gstatic.com/s2/sitemaps/sitemap-005.txt</loc>
<lastmod>2010-07-21</lastmod>
</sitemap>
<sitemap>
<loc>http://www.gstatic.com/s2/sitemaps/sitemap-006.txt</loc>
<lastmod>2010-07-21</lastmod>
</sitemap>
<sitemap>
<loc>http://www.gstatic.com/s2/sitemaps/sitemap-007.txt</loc>
```



Инструкции sitemap.xml и robots.txt при правильном их использовании должны дополнять друг друга. Существуют три правила взаимодействия этих инструкций.



Сайтам, на которых часто обновляется контент (новостным порталам, интернет-магазинам), необходима автоматизация файла `sitemap.xml`, иначе он потеряет свою актуальность.

```
<lastmod>2010-07-21</lastmod>  
</sitemap>
```

### 3.10. Принцип взаимодействия `sitemap.xml` и `robots.txt`

Инструкции `sitemap.xml` и `robots.txt` при правильном их использовании должны дополнять друг друга. Существуют три правила взаимодействия этих инструкций:

- ▶ `sitemap.xml` и `robots.txt` не должны противоречить друг другу;
- ▶ все страницы, исключенные в `robots.txt`, должны быть исключены также из `sitemap.xml`;
- ▶ все индексируемые страницы, разрешенные в `robots.txt`, должны содержаться в `sitemap.xml`.

### 3.11. Автоматизация добавления страниц в `sitemap.xml`

Сайтам, на которых часто обновляется контент (новостным порталам, интернет-магазинам), необходима автоматизация файла `sitemap.xml`, иначе он потеряет свою актуальность. Вручную обновлять инструкцию каждый день – совершенно нерациональный подход. Для того чтобы автоматизировать добавление и исключение страниц в `sitemap.xml`, необходимо создать модуль, связывающий БД сайта, инструкции `robots.txt` и `sitemap.xml`. Программа должна руководствоваться следующими правилами:

1. Примем как данность, что файл `sitemap.xml` уже изначально создан на сайте, проработан вручную и содержит только важные страницы.
2. При добавлении новой страницы система дописывает строчку в `sitemap.xml` с указанием этой страницы. В зависимости

от типа последней указывается параметр частоты индексации и ее приоритет. Если страница закрыта от индексации в robots.txt, она не прописывается в sitemap.xml.

3. При удалении страницы программа ищет строчку с указанием данной страницы в sitemap.xml. Если страница там есть, информация о ней удаляется.



## Глава 4

# Работа с авторским контентом

---



**В**от мы и добрались до самого интересного. В этой главе я расскажу вам о том, как решать проблемы уникальности и недостатка текста на сайте. Непременно уделите представленной ниже информации должное внимание, поскольку именно по наполнению страниц поисковая система распознает содержание сайта. Нет текста – не понятно, о чем сайт, следовательно, для поисковой системы он «пустышка».

Вообще есть три серьезные проблемы оптимизаторов, связанные с текстами: воровство контента, отсутствие и недостаток текстов, дублирование текстов внутри сайта.

К сожалению, действенной методики борьбы с воровством контента не существует. Несмотря на усилия юристов, есть способы воровства, пресечь которые невозможно. Между тем в этой главе вы познакомитесь с одной методикой, которая сводит к нулю вред, причиненный «угоном» контента.

Что касается проблемы отсутствия и недостатка контента, то решается она обычно за счет автоматизации, а при продвижении небольших сайтов – ручным дописыванием недостающего уникального контента.

### 4.1. Уникальность контента

Каждый, кто занимается продвижением сайтов не первый день, сталкивался с проблемой воровства контента. Казалось бы, воруют – и пусть, что в этом такого?

Но дело в том, что очень часто в итоге копия оказывается выше оригинала в результатах поиска, а значит, воровство контента можно приравнять к воровству трафика. Ходят слухи, что на одной из конференций представители Яндекса озвучили значение погрешности определения первоисточника, назвав цифру ~ 30%.

Как же так? Неужели поисковая система не может определить первоисточник и ранжировать сайты справедливо? Все-го-то и нужно – определить, где контент появился раньше, и отдать этому сайту предпочтение. Но не все так просто. Ведь поисковая система переиндексирует сайты неравномерно. Допустим, если ваш ресурс обновляется не часто, значит, скорее всего, и поисковый робот – не частый ваш гость. Если вы публикуете результаты научного исследования раз в 3 месяца, а новостной портал разместит вашу статью как свою, шансы, что у вас статья будет проиндексирована раньше, ничтожны. На живые новостные порталы робот-индексатор заглядывает несколько раз в час, а ресурс с менее подвижным контентом может посетить и через несколько недель. Таким образом, выходит, что «оригинал» статьи расположен на новостном портале, а научный центр опубликовал его «копию». Как видите, по дате индексации определять оригинал нельзя.

Тогда, может быть, определять первоисточник по авторству в тексте статьи? Бесплезно, ведь ворам не составит труда указать свое авторство и сайт в качестве первоисточника. Даже ссылки на оригинал, которые указали добросовестные сайты, разместившие ваши статьи, тоже не спасение. Можно запросто разместить на своем сайте чужую статью, а потом продублировать ее в нескольких блогах с указанием сайта как первоисточника.

Как видите, любые признаки оригинала легко подделать. Поисковые системы



Если вы публикуете результаты научного исследования раз в 3 месяца, а новостной портал разместит вашу статью как свою, шансы, что у вас статья будет проиндексирована раньше, ничтожны. На живые новостные порталы робот-индексатор заглядывает несколько раз в час, а ресурс с менее подвижным контентом может посетить и через несколько недель.

не станут разбираться в этих тонкостях, а владельцам сайта нет смысла в каждом случае пытаться восстановить свои авторские права юридически. Борьба с дорвеями, воруящими контент, подобна борьбе с ветряными мельницами, ведь эти сайты размещают на доменах, зарегистрированных на несуществующих людях в зонах, не подвластных влиянию юрисдикции Российской Федерации. Против сайтов, размещенных на зарубежных хостингах и не в доменной зоне РФ, вы – увы! – бессильны. Если же сайт, укравший ваш контент, располагается на российской «интернет-территории», а вы имеете юридическое подтверждение авторства или хотя бы след в веб-архиве, вы можете связаться с хостером или владельцем домена и попросить убрать ворованные тексты, а в случае их отказа сделать это, подать иск. Но все-таки куда проще и дешевле написать новый текст, чем закрепить за собой через суд авторство существующего.

Открою вам страшный секрет: **поисковые системы не пытаются распознать оригиналы и копии контента**. Их задача – удовлетворить своих посетителей, а на уровень удовлетворенности никак не повлияет, окажется ли на первом месте в результатах поиска оригинал или же там расположится копия. Поэтому воровство контента – это головная боль исключительно оптимизаторов, а не разработчиков поисковых систем.

Второй важный момент: **в Рунете практически нет статей, уникальных на 100%**. Какая-то часть любой статьи с огромной долей вероятности уже встречалась где-то в Сети, а значит, невозможно точно разграничить оригинал и переписанный текст (репайт).

**Немного статистики для понимания масштабов:**

► По данным проекта «Черный квадрат», реализованного Алексеем Тутубалиным



Поисковые системы не пытаются распознать оригиналы и копии контента.

Их задача – удовлетворить своих посетителей, а на уровень удовлетворенности никак не повлияет, окажется ли на первом месте в результатах поиска оригинал или же там расположится копия.

по адресу <http://www.rukv.ru/>, 9 сентября 2010 года в Рунете только на поддоменах второго уровня в зонах.RU и.SU работало 1893837 сайтов.

► По данным Яндекса, опубликованным в информационном бюллетене осенью 2009 года, в Рунете насчитывалось около 15 миллионов сайтов (это около 6,5% всего Интернета). При этом информация в Сети распределена неравномерно: 88% всего текста находится менее чем на 1% сайтов. Средний сайт Рунета состоит из 255 страниц, содержит 159 тысяч слов и 204 картинки. Большинство сайтов гораздо меньше среднего: половина из них состоит всего из одной страницы.

Так будем же реалистами и усвоим **три простые истины:**

1. Невозможно достоверно определить первоисточник.
2. Практически не существует на 100% уникального контента.
3. Большинство страниц имеет дубликаты.

Неутешительная картинка? Не стоит опускать руки. В действительности все не так плохо, как кажется. Периодически обновляйте контент на входных страницах, это положительно сказывается на ранжировании сайтов. Как поступать с дубликатами, вы уже знаете из 2-й главы. Что касается воровства контента, есть действенный способ нейтрализовать вред, который оно наносит.

## 4.2. Как защитить контент от воровства?

**БЕСПОЛЕЗНО:**

► дублировать статью на сторонних сайтах с указанием вашего сайта в качестве первоисточника, поскольку это не оказывает значительного влияния на позиции сайта в выдаче;



По данным Яндекса, опубликованным в информационном бюллетене осенью 2009 года, в Рунете насчитывалось около 15 миллионов сайтов (это около 6,5% всего Интернета). При этом информация в Сети распределена неравномерно: 88% всего текста находится менее чем на 1% сайтов.



Ссылки нужно размещать с быстроиндексируемых сайтов, а в качестве текстов ссылок следует использовать небольшие части статьи. Если применять простенький скрипт, можно автоматизировать этот процесс. Скрипт будет разбивать заданный текст на пассажи с отбивкой не более 100 символов и дописывать теги для размещения ссылок в SAPE.

▶ отключать через JavaScript правую кнопку мышки и выделение текста на странице, использовать любые программные способы защиты контента от копирования, поскольку в 99% случаев воруют не путем копирования контента со страницы, а парсят специальными программами, которые перекачивают сайт и срезают теги, оставляя только текст статьи, а иногда контент выдирают вручную из HTML;

▶ ставить ссылки на первоисточник в тексте статьи, так как чаще всего такие ссылки срезаются программами-парсерами.

#### Что же делать?

**Определите, по каким запросам вас могут искать, составьте правильные анкеры, закупите ссылки.** Ссылки нужно размещать с быстроиндексируемых сайтов, а в качестве текстов ссылок следует использовать небольшие части статьи. Если применять простенький скрипт, можно автоматизировать этот процесс. Скрипт будет разбивать заданный текст на пассажи с отбивкой не более 100 символов и дописывать теги для размещения ссылок в SAPE.

Закупая ссылки с частями текста статьи, мы поднимаем страницу со статьей в результатах поиска выше других копий, поскольку синергия между содержимым на странице и в анкор-листе увеличивает значимость этой страницы.

У этой методики существует свой недостаток: используя разбивку текста на пассажи, мы теряем запросы, возникающие на их стыке, ведь текст дробится на части, которые не накладываются друг на друга. Кроме того, если мы будем последовательно перебирать весь текст, то станем тратить лишние деньги на продвижение запросов вроде «продукция предприятия получила высшую оценку», а они никогда не приведут к нам целевых посетителей. Поэтому нужно брать тексты ссылок из областей около ключевых слов конкретной

страницы. Именно по ним на сайт могут попасть целевые посетители. В общем виде методика такова: анкор не более 100 символов, отбивка «ключевое слово в конце текста ссылки» и отбивка «ключевое слово в начале текста ссылки».

### 4.3. Как воровать контент?

Воруются контент точно по такому же принципу, как и защищается. Вышеописанный способ работает в обе стороны: при краже статья копируется, а на нее размещаются входящие внешние ссылки с частями текста статьи в анкоре.

### 4.4. Отсутствие и недостаток контента

Нехватка или полное отсутствие контента на сайте приводит к ошибкам в ранжировании, поскольку поисковые системы не могут корректно распознать принадлежность страницы к ключевым словам. На небольших сайтах недостаток контента восполняется вручную – дописыванием недостающих текстов, а вот на крупных ресурсах не обойтись без автоматизации. Например, на сайте <http://mobiguru.ru>

#### О том, как купить DELL Alienware M11x дешевле

Сравни все цены на DELL Alienware M11x в более чем 27 интернет-магазинах, продающих ноутбуки, найди, где можно купить DELL Alienware M11x дешевле, определив лучшую и среднюю стоимость, закажи ноутбук DELL Alienware M11x в самом дешевом месте с доставкой по Москве, Санкт-Петербургу, Екатеринбург или всей России и даже Украине!

#### О том, как купить ASUS K50IJ дешевле

Сравни все цены на ASUS K50IJ в более чем 49 интернет-магазинах, продающих ноутбуки, найди, где можно купить ASUS K50IJ дешевле, определив лучшую и среднюю стоимость, закажи ноутбук ASUS K50IJ в самом дешевом месте с доставкой по Москве, Санкт-Петербургу, Екатеринбург или всей России и даже Украине!

#### Информация о Toshiba Satellite A350

Прочитай полный обзор Toshiba Satellite A350 и эксклюзивный тест ноутбука, ознакомься с отзывами владельцев Toshiba Satellite A350 в форуме ноутбука, посмотри описание Toshiba Satellite A350 с характеристиками и фото ноутбука и не забудь про новости!

#### О том, как купить DELL Inspiron M301z дешевле

Сравни все цены на DELL Inspiron M301z в более чем 18 интернет-магазинах, продающих ноутбуки, найди, где можно купить DELL Inspiron M301z дешевле, определив лучшую и среднюю стоимость, закажи ноутбук DELL Inspiron M301z в самом дешевом месте с доставкой по Москве, Санкт-Петербургу, Екатеринбург или всей России и даже Украине!



Воруются контент точно по такому же принципу, как и защищается. Вышеописанный способ работает в обе стороны: при краже статья копируется, а на нее размещаются входящие внешние ссылки с частями текста статьи в анкоре.

проблему недостатка контента решили за счет генерации мини-описаний с вхождениями ключевых слов в тексте.

Следует отметить, что важно не только наличие контента, но и его уникальность, структура, наполненность. По лицензии Яндекса, сайты, использующие не уникальный контент, могут быть частично или полностью исключены из индексной базы. Кроме того, страницы с не уникальным контентом ранжируются хуже.

Далее нам предстоит подробно рассмотреть способы автоматической генерации контента. Мы разберем не только генерацию контента в тексте страницы, но также генерацию метатегов description и keywords и генерацию заголовков TITLE.

#### 4.5. Автоматизация наполнения Description, Keywords, Title

В зависимости от приспособленности CMS автоматизация заполнения заголовков и метатегов может быть реализована двумя путями:

- ▶ разработка встраиваемого в CMS модуля,
- ▶ ручная загрузка ранее сгенерированных текстов в БД.

Первый способ более практичен и удобен, чем второй. Для продвижения сайтов, на которых часто появляются новые страницы, подходит только он. Второй способ значительно более прост в реализации, но он идеален только для тех сайтов, где количество новых страниц за месяц можно пересчитать по пальцам.

##### Автоматизация наполнения keywords

По умолчанию этот тег должен быть пустым. На входных страницах в него вписываются все продвигаемые запросы. Автоматизировать наполнение тега можно в зависимости от типа сайта: либо через интеграцию с БД, откуда будет забираться

наименование товарной группы и название конкретной модели, либо через лингвистический анализ текста страницы, результатом которого должны стать наиболее часто употребляемые на ней слова. Неключевые частоупотребимые слова могут удаляться через специальный словарь или вручную при небольших объемах.

### Автоматизация наполнения TITLE

Настраивается через формулу [TITLE страницы] [TITLE раздела] [TITLE сайта]. TITLE страницы = N1 страницы, TITLE раздела = N1 раздела, TITLE сайта задается изначально и дублируется на всех страницах сайта. Желательно, чтобы в TITLE попали все возможные части запросообразующих слов. В этом случае не обойтись без прописывания корректных заголовков N1, хотя при правильной верстке и грамотном наполнении сайта, этот тег должен содержать основные ключевые слова страницы.

Пример TITLE с сайта компании «БЕСТ-Недвижимость»:

- ▶ Квартиры в Бибирево. Предложения по продаже квартир в Москве, новостройки и вторичная недвижимость. Агентство недвижимости «БЕСТ».
- ▶ Квартиры в Медведково. Предложения по продаже квартир в Москве, новостройки и вторичная недвижимость. Агентство недвижимости «БЕСТ».
- ▶ Коттеджи в Солнцево. Загородная недвижимость в Подмосковье. Агентство недвижимости «БЕСТ».

### Автоматизация генерации Description

Настраивается по заданной формуле<sup>2</sup> типа [Текст] [x] [Текст] [y] [z] [a] [b] [Текст].

<sup>2</sup>Бывает, что в формуле полностью отсутствует статичный текст. Переменные могут быть выражены как элементом БД (наименование модели, название марки, цена), так и братья из словаря, который заранее готовит копирайтер.

### ФАКТ



**С**айтРепорт – система SEO-аналитики, позволяющая проводить диагностику сайта.

**СайтРепорт** – это возможность провести полную диагностику сайта, увидеть его глазами поисковых систем. Оптимизатору приходится сталкиваться со множеством проблем: недостаток контента, 404 ошибки, дубликаты страниц, мусорные ссылки, страницы 30 уровня вложенности, отсутствие заголовков... Легко найти ошибки на небольшом сайте, достаточно просмотреть все страницы вручную. Что делать, если на сайте несколько тысяч страниц? Наш робот выполнит эту работу за вас, перекачает ваш сайт, после чего система проведет анализ, который вы подробно изучите. Сделав анализ сайта, вы получите подробный отчет об ошибках и узнаете потенциал вашего проекта в работе с внутренними факторами.

**Проще говоря, бесплатно решите проблему внутренних факторов.**

### КАК РАБОТАЕТ СИСТЕМА:

Система представляет собой программный комплекс, использующий мощности нескольких серверных машин и состоящий из 5 модулей:

#### # Робот-индексатор

Робот обходит все страницы сайта в соответствии со всеми инструкциями для вашего сайта, то есть фактически имеет все функции поискового робота. Робот заносит в базу код ответа каждой страницы, запоминает скорость соединения, сохраняет контент каждой страницы, вынося в отдельный блок все исходящие ссылки.

#### # Анализ ссылочных связей

Модуль выстраивает ссылочные связи всех страниц между собой, обрабатывая миллионы ссылок. Для каждой страницы составляется список входящих внутренних ссылок с указанием анкоров. Все ссылки проверяются на индексацию. Система рассчитывает условный вес каждой страницы по технологии, имитирующей расчет PageRank. Создаются отчеты с указанием всех страниц, ссылающихся на 404 ошибки, а также всех страниц, с которых стоят редиректы.

#### # Обработчик контента

Система определяет уникальность контента каждой страницы сайта по отношению к другим его страницам, сравнивая содержание всех страниц между собой. Модуль сравнения основан на собственной технологии построения хешей частей страниц с использованием технологии шинглов, что обеспечивает большую скорость и точность полученных данных. Система также определяет страницы с недостатком и переизбытком контента.



### # HTML-валидатор

Модуль обрабатывает HTML-код каждой страницы, выделяя содержание всех важных HTML-тегов: TITLE, Meta keywords, Meta description, STRONG, B, EM, H1-H6. Система выстраивает отчеты по дублированию содержимого этих тегов, а также делает срезы по страницам, где важные HTML-теги пусты.

### # Среда для аналитики – <http://СайтРепорт.РФ/>

Вы изучаете полученные данные и на основании их планируете, как вам дальше продвигать сайт.

### Система перекачивает до 5 000 страниц вашего сайта.

Какую информацию вы получите о сайте:

- Общий анализ сайта.
- Анализ внутреннего анкор-листа.
- Расчет распределения весов между страницами при перелинковке по классической формуле PageRank.
- Поиск битых ссылок.
- Поиск сквозных ссылок.
- Проверка HTML-кода всех страниц на валидность.
- Анализ корректности <noindex> и rel=nofollow.
- Генератор sitemap.xml.
- Анализ внутренней перелинковки.
- Сводные отчеты с графиками по перелинковке.
- Проверка на наличие сайта в крупных каталогах.
- Поиск страниц с исходящими внешними ссылками.
- График распределения страниц по ответу сервера.
- Поиск страниц с 404 ошибками.
- Поиск страниц с 3\*\* редиректами.
- Оценка размера страниц и скорости их загрузки.
- График распределения страниц по уровням вложенности.
- Список 100% дубликатов страниц, одинаковых в рамках сайта.
- Список 90%+ дубликатов страниц, одинаковых в рамках сайта.
- Анализ заполненности страниц.
- Пустые страницы.
- Отчет «На какие страницы ссылается данная».
- Отчет «Какие страницы ссылаются на данную».
- Анализ TITLE.
- Страницы с одинаковыми TITLE.
- Дубликаты Keywords.
- Дубликаты Description.
- Страницы с пустыми TITLE.
- Оценка составления заголовков H1-H6.
- Анализ разметки STRONG и B.
- Анализ разметки EM.

### Пример:

«Фабрикой [M] производится классическая и современная [A]. Модель [N] отличается от [B]. Позвоните прямо сейчас +7495777-77-77».

▶ A = [мебель для кухни|кухонная мебель]

▶ B = [элегантностью и уникальным дизайном|оригинальным дизайном и надежностью конструкций]

▶ M = \$fabrique

▶ N = \$model

## 4.6. Автоматизация генерации контента

### Автоматизация генерации описаний товаров

Аналогично с Description настраивается по формуле<sup>3</sup> типа [Текст] [x] [Текст] [y] [z] [a] [b] [Текст].

### Пример:

«Фабрикой [M] производится классическая и современная [A]. Модель [N] отличается от [B]».

▶ A = [мебель для кухни|кухонная мебель]

▶ B = [элегантностью и уникальным дизайном|оригинальным дизайном и надежностью конструкций]

▶ M = \$fabrique

▶ N = \$model

### Автоматизация генерации технических характеристик (уровень 1)

Аналогично настраивается по формуле типа [x] [A]; [y] [B]; [z] [C];

<sup>3</sup>Бывает, что в формуле полностью отсутствует статичный текст. Переменные могут быть выражены как элементом БД (наименование модели, название марки, цена), так и братья из словаря, который заранее готовит копипрайтер.

Пример:

«Вес шкафа 213 кг. В высоту шкаф 10 м. В длину 8 м».

▶ A = [213 кг]

▶ B = [10 м]

▶ C = [8 м]

▶ x = [Вес шкафа | Масса шкафа | Шкаф весит | Вес | Масса | Насколько тяжелый]

▶ y = [Высота шкафа | В высоту шкаф | Высота | В высоту | Высота от пола | Габариты: высота]

▶ z = [Длина шкафа | Габариты: ширина | Длина | В длину шкаф | В длину]

### Автоматизация генерации технических характеристик (уровень 2)

Строим зависимости переменных от значений: выставляем пороги.

**Вес:** крохотный 0–100; небольшой 100–200; не очень большой 200–300; достаточно большой 300–400; огромный 400+. В этом случае дописываются текстовые блоки из разных словарей.

**Пример:** Вес шкафа не очень большой – всего 213 кг, можно хранить в помещениях, не оборудованных укрепленными перекрытиями. В высоту шкаф займет около 10 м – впишется в складское помещение с высокими потолками. В длину – 8 м, достаточно емкий, чтобы хранить большое количество грузов.

## 4.7. Обнаружение дубликатов

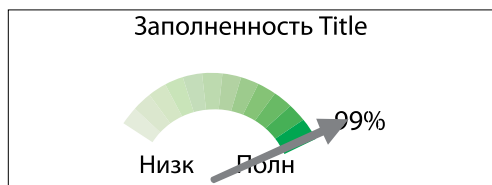
Дубликаты – это, как правило, «побочный эффект» некоторых характеристик сайта, как-то:

- ▶ наличие сквозного блока с контентом,
- ▶ недостаток контента,
- ▶ конструкция страниц списка анонсов,
- ▶ особенности CMS, из-за которых полностью дублируются страницы.

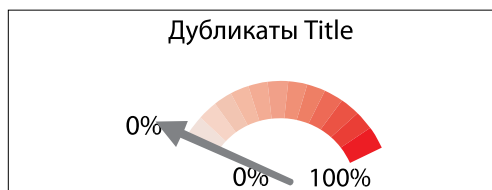
- Анализ robots.txt.
  - Лемматизация всего контента на сайте.
  - Анализ плотности ключевых слов на каждой странице.
  - Поиск страниц, содержащих ключевые словосочетания по сайту.
  - Представление контента каждой страницы без <HTML> кода.
  - Анализ данных whois-домена.
  - Настройка маски страниц, которые система анализирует, т. е. можно, например, проанализировать только страницы, содержащие в URL /beauty/, или все, кроме страниц, содержащих в URL thread.php.
  - Настройка глубины анализа страниц.
  - Возможность настроить robots.txt для робота-индексатора системы.
  - Отчет только по входным страницам.
  - Анализ установленных метрик.
  - Количество ссылок в Yahoo.
  - Анализ скорости загрузки страниц.
  - Анализ размеров страниц.
  - Список всех страниц, на которых встречается любая форма заданного слова.
  - Отчет по наиболее часто употребляемым словам всего сайта.
  - Экспресс-отчет по типичным ошибкам сайта.
  - Структура и иерархия сайта.
  - Сравнение индексации типов страниц и категория в поисковых системах
- и множество других отчетов, список которых мы постоянно расширяем.

### Некоторые графики из отчета:

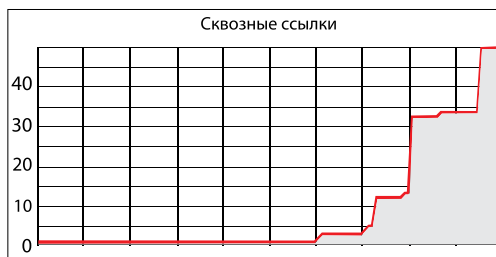
- График заполненности TITLE



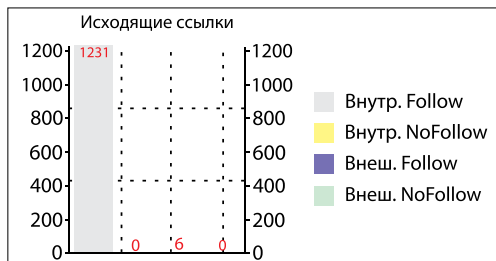
- График дубликатов в теге TITLE



- Анализ сайта на наличие сквозных ссылок



- Анализ исходящих ссылок



- Анализ глубины вложенности страниц сайта



- Анализ сайта на наличие дубликатов страниц



Обнаружить дубликаты можно двумя путями: вручную и автоматически.

### Инструкция по обнаружению дубликатов вручную

Нет необходимости объяснять, почему ручную дубликаты обнаружить сложно. Этот способ подходит лишь для небольших сайтов, где все страницы можно внимательно просмотреть, но даже в этом случае погрешностей не избежать.

### Для ручной проверки воспользуйтесь следующим алгоритмом:

1. Анализируем данные Яндекс.Вебмастер. Просмотрите, какие страницы Яндекс решил удалить из результатов поиска. Именно они нас интересуют.

2. Через расширенный поиск Яндекса ищем все страницы, проиндексированные с данного сайта, и ставим режим отображения 50 результатов на странице. Прокликаем все проиндексированные Яндексом страницы, если их меньше 500, просматриваем каждую из них вручную. Если страниц больше 500, прокликаем только первые 500, у остальных просто просматриваем сниппеты. Этой информации хватит для поверхностного анализа сайта на дубликаты. Далее смотрим страницы, проиндексированные в Google, особенно обращая внимание на те, которых не было в Яндексе (они будут подсвечены в браузере синим цветом, так как вы их еще не посещали). Во время просмотра вы обнаружите часть дубликатов.

3. Просматриваем структуру сайта, вникая, каким образом формируется контент на страницах, откуда берутся текстовые блоки, где возникают блоки дублирующегося контента.

4. Ищем контент внутри сайта по точному вхождению (для этого нужно пассаж текста искать в Яндексе с использованием операторов «кавычки» и «восклицательный

знак»: «!слово!слово!слово») с основных входных страниц. Особенно здесь можно выделить главную страницу. Если дубликаты есть, вы их увидите.

### Способы автоматизации обнаружения дубликатов

Для автоматического анализа нужно воспользоваться специализированными приложениями. На момент написания книги существует всего одна программа, доступная к бесплатному использованию и способная провести подобный анализ, – «СайтРепорт».

Некоторые системы способны сделать небольшую часть анализа, но все они являются пользовательскими и используют мощности вашего компьютера, что очень сильно ограничивает объемы сканируемой информации. Сайт из нескольких сотен страниц просканировать получится, но интернет-магазины с десятками тысяч страниц таким системам не по зубам.

Для поверхностного анализа я могу рекомендовать некоторые программы. Например,

**Semonitor** осуществляет следующие операции:

- ▶ поиск битых/нерабочих ссылок;
- ▶ поиск «потерянных» файлов;
- ▶ поиск недочетов и ошибок в html-коде;
- ▶ создание подробного отчета по всем исходящим ссылкам с данного сайта;

**NetPromoter** выполняет такие задачи:

- ▶ проверка работоспособности ссылок на сайте;
- ▶ поиск неработающих ссылок;
- ▶ определение размера страниц.

К сожалению, эти программы не помогут вам в обнаружении дубликатов. Серьезный анализ можно сделать лишь с помощью одной системы, имя которой – «СайтРепорт».

## МНЕНИЕ

**Иван Севостьянов,**  
генеральный директор «ВебПроекты»:



— Контент очень важен для любого сайта. Еще 2–3 года назад можно было пренебрежительно относиться к содержанию сайта, копировать информацию с чужих сайтов или вовсе не наполнять сайт полезной для посетителя информацией.

Сейчас такие методы не работают, и в дальнейшем требования к текстовой составляющей сайта со стороны поисковых машин будут только расти.

У контентной составляющей любого интернет-сайта есть три проблемы. Первая – воровство контента. Вторая – стоимость создания контента. Третья – определение первоисточника контента.

К сожалению, контент воровали, воруют и будут воровать, так как правовые механизмы защиты авторского права в Интернете еще должным образом не отработаны, и даже если вы выиграете судебное дело, то затраты времени на судебное разбирательство будут много выше, чем сумма компенсации.

Контент – дорогое удовольствие, особенно если у вас крупный контентный проект. Написание 1 страницы хорошего контента стоит от 300 рублей. Даже если у вас 200 страниц на сайте, то это уже 60 тысяч. А если страниц несколько тысяч, то сумма будет еще выше. Частично проблему можно решить за счет автоматической генерации контента, но это подойдет не всем сайтам. Поэтому вкладываться в контент спешат не все, особенно учитывая то, что его все равно могут своровать.

Третья проблема – поисковики до конца не научились определять первоисточник контента. Это тоже мешает вкладываться в написание уникального контента, так как вы можете сделать сайт с качественным контентом, его своруют, а поисковая система первоисточником может посчитать именно сайт с ворованным контентом.

К сожалению или к счастью, выход один – писать уникальный и качественный контент, учиться защищать его от воровства, отстаивать свою позицию перед поисковиками в спорных ситуациях. И самое главное – помнить, что контент пишется для людей, а не для роботов!

## МНЕНИЕ

**Константин Шурыгин, генеральный директор автоматического сервиса по продвижению сайтов «Маремото»:**



— Проблема дублированного контента в реальности стоит даже острее, чем отмечено выше. В настоящий момент я веду переписку с Яндексом, который только что удалил из индекса уважаемый 8-летний сайт, имеющий хорошие показатели авторитетности и 62 000 естественных ссылок. Вместо него «главным зеркалом» был признан безвестный клон трех месяцев от роду, имеющий всего 5 платных ссылок, укравший чужой контент. Авторитетный сайт был удален и из каталога Яндекса, а его место здесь занял вынырнувший из ниоткуда дубль. От аргументов Платона Шукина веет безысходностью: «...если даже описание в каталоге Яндекса будет перенесено вручную на правильный адрес, оно автоматически вернется обратно через некоторое время, потому что ваш сайт все еще является неглавным зеркалом. Пока сайты — зеркала, перенос описания не поможет. Как только сайты перестанут быть зеркалами, напишите нам снова по вопросу возвращения в каталог...».

Как показывает опыт, «раззеркаливание» может занять от одной недели до трех месяцев. За это время владелец клона с ворованным контентом обычно успевает разослать всем владельцам ссылающихся сайтов письма с просьбой «обновить ссылки в связи с переездом сайта на новый адрес». В качестве аргумента обычно приводят «новый» адрес, указываемый в Яндекс.Каталоге. Таким образом рейдерам удастся украсть у ничего не подозревающих владельцев сайта не только тысячи страниц уникальных текстов, но и место в каталоге Яндекса, а также накопленную годами ссылочную базу, а потом и PageRank. Даже если впоследствии двойник будет удален из Яндекса, такой клон может зависнуть в ссылочных биржах и годами приносить доход своим владельцам.

Часто дубли вашего сайта создают конкуренты, которые выкачивают ваш контент программами типа Teleport или Webcorier, размещают его на бесплатных хостингах, накачивают бесплатными или платными ссылками, а также дают на этот сайт рекламу в Директе. Этого бывает достаточно, чтобы спровоцировать Яндекс к неверному зазеркаливанию сайтов.

Однако бывает и так, что дубли вашего сайта создают вовсе не злоумышленники. Дубли может создать ваш собственный веб-мастер, заполняя зеркалами все выкупленные корпоративные домены. Часто это дело рук давно уволенного веб-мастера, который это делал из благородных намерений не оставить тайп-сквоттерам ни одного шанса. Так, однажды столкнувшись с проблемами в продвижении корпоративного сайта, я обнаружил 17 зеркал на доменах второго уровня, о существовании которых никто из владельцев понятия не имел. В таких случаях организация 301 редиректа на главное зеркало и наличие директивы Host в robots.txt всех доменов помогает избежать больших проблем.

**РЕЗЮМЕ:** Существует две методики обнаружения дубликатов: ручная и автоматизированная. Ручная бесплатна, но менее эффективна. Для автоматизированного анализа идеально подходит система «СайтРепорт» либо некоторые самописные разработки, используемые компаниями-лидерами рынка для собственных нужд.

Но неожиданнее всего бывает ситуация, когда зеркала сайта делают... сервера сами. На некоторых версиях сервера Apache поддомены третьего уровня по умолчанию заполняются содержимым домена второго уровня. Если веб-мастер создает поддомены с прицелом «на будущее», то они неожиданно для всех могут стать клонами основного сайта. Это может вызвать «зазеркаливание», катастрофическое падение позиций в выдаче Яндекса, заметное сокращение целевого трафика и продаж. А всему виной – один-единственный символ \*, который надо было удалить из настроек сервера.

Следует постоянно следить за появлением «зеркал» своего сайта в Интернете. При наличии более авторитетных дублей вашего контента, продвижение сайта будет неэффективным, пока вы не решите проблему с этими зеркалами. Перед началом продвижения определяет наличие у сайта авторитетных зеркал. Ведь при наличии таких зеркал продвижение сайта будет нестабильным, а результат будет отсутствовать.

## Глава 5

# Коды ответов сервера

---



**К**од ответа – это инструкция, которую получает посетитель, заходя на страницу сайта. С помощью этой инструкции браузер может скорректировать обработку страницы на стороне посетителя. На первый взгляд, код ответа сервера – фактор, не имеющий никакого отношения к продвижению сайтов. Однако это не так. Да, настройки ответа сервера не могут улучшить ранжирование сайта, но допущенная в них ошибка может поставить крест на ваших планах покорения топа результатов поиска. Поисковые роботы, как и посетители, учитывают коды ответов сервера и на их основании корректируют обработку страниц. В частности, по кодам ответа поисковые системы принимают решение о том, что:

- ▶ страница существует и работает корректно;
- ▶ страница не существует;
- ▶ страница перемещена навсегда;
- ▶ страница временно перемещена;
- ▶ сервер загружен и не может обработать запрос.

Существуют и другие, менее значимые, инструкции.

### 5.1. Код ответа 200

Поисковые системы добавляют в свою индексную базу только те страницы, которые отдают код ответа 200, означающий «страница существует и доступна для просмотра». Все важные страницы, которые мы

хотим «скормить» поисковому роботу, должны отдавать код ответа 200.

Если страница использует редирект и лишь после перемещения отдает код ответа 200, поисковая система индексирует только конечную страницу. Страница-шлюз может быть удалена или понижена в ранжировании (это зависит от типа редиректа).

Если страница отдавала код ответа 200 и попала в индексную базу, но позже была удалена и теперь отдает код 404, она со временем удаляется из индексной базы.

Если на странице поставить 301 редирект, то через некоторое время она будет также удалена из индекса, а ее вес в некоторых случаях будет передан той странице, на которую ссылался редирект. К сожалению, Яндекс очень долго осуществляет переиндексацию страниц с редиректом: были зафиксированы случаи, когда переключка происходила почти год.

## 5.2. Коды ответов 301/302/303/307/3\*\*

Существует большое количество кодов для обозначения редиректа. Укажем здесь основные, наиболее важные для поисковых систем:

▶ **301** – «постоянное перенаправление». Означает, что страница перемещена навсегда. Этот редирект используется чаще всего. В случае если адрес страницы изменился навсегда, что происходит, например, при редизайне или реструктуризации сайта, 301 редирект позволяет сохранить входящие внешние ссылки за счет переклейки их веса на новые адреса страниц. Яндекс обрабатывает этот редирект очень долго, может пройти до года, прежде чем страницы склеятся.

▶ **302/303/307** – «временное перенаправление». Веб-мастер таким образом указывает роботу поисковых систем, что страница перемещена временно,

## МНЕНИЕ

Николай Хиврин:



— **З**начимость верного использования кодов ответов сервера трудно переоценить. Правильная работа с ними может позволить совершенно безболезненно перейти на новую систему управления и адресацию web-сайта. В сфере поискового продвижения особое внимание стоит уделять трем кодам ответов: 200, 404 и 301. Пожалуй, самой частой ошибкой сегодня является неверное использование редиректов 301. Они могут отсутствовать вообще, а могут вести последовательно через несколько адресов. Другой важной проблемой является дублирование содержимого, а это возникает из-за того, что у ряда сгенерированных адресов возникает код ответа 200. Правильная работа с кодами ответов сервера является очень важным навыком для поискового оптимизатора и позволяет избежать целого ряда очень серьезных ошибок.

Рассмотрим наиболее распространенную ситуацию, в которой предыдущая версия сайта компании сильно устарела. Компания хочет обновить свой сайт, однако боится потерять позиции в результатах поиска, а как следствие, и источник новых клиентов. Как правило, обновление сайта связано с установкой новой системы управления (CMS), в результате чего на сайте создается новая структура адресов страниц. Если осуществить переход на новую адресацию, то, конечно же, все старые страницы пропадут из индекса поисковых систем, а вместе с ними и позиции по ключевым запросам. Однако если сделать 301 редиректы с каждого старого адреса на новые страницы, то этой проблемы можно избежать. Поисковые роботы и посетители сайта будут перенаправляться на новые страницы, что позволит полностью сохранить их вес и не потерять позиции в результатах поиска.

Другая распространенная ошибка: интернет-магазин содержит каталог, он генерирует очень большое количество внутренних ссылок, а структура их бывает хаотичной. В результате этого поисковые роботы могут попасть на такие страницы, адреса которых совсем не были предусмотрены разработчиком проекта. Для того чтобы эти страницы не засорили индекс поисковых систем, нужно правильно расставить ошибки 404 и исключить эти страницы из индекса.



а значит, ее не нужно удалять из индексной базы. В этом случае Яндекс может добавить в свою базу обе страницы, но отображать только одну из них в результатах поиска.

#### Из инструкций Яндекс.Вебмастер:

► При постоянном перенаправлении (301) между внутренними страницами хоста в результатах поиска будет отображаться цель.

► При временном перенаправлении (302/303/307/Meta refresh) между внутренними страницами хоста в результатах поиска будет отображаться наиболее компактный URL.

#### Типичные ошибки:

► Вместо 301 стоит конструкция вида  
302 —> 302 —> 302 —> 301

**РЕЗЮМЕ:** Если вы изменили адрес страницы, выложили новую версию сайта или просто хотите убрать дубликат (например, склеить/site/u/site), используйте 301 редирект. Важно использовать именно одношаговый редирект, не создавайте сложных конструкций из нескольких редиректов.

### 5.3. Код ответа 404

Код ответа 404 используется для указания браузеру или поисковой системе, что страницы не существует. Сервер в этом случае отображает или стандартную страницу «по этому адресу ничего не найдено», или страницу-заглушку, если создан специальный шаблон для несуществующих страниц.

Важно не путать серверный ответ 404 и 404-ю страницу-заглушку. Некоторые веб-мастера по каким-то им одним ведомым причинам на 404-й странице размещают заглушку с указанием «страница не существует», но на уровне сервера



Другая распространенная ошибка: интернет-магазин содержит каталог, он генерирует очень большое количество внутренних ссылок, а структура их бывает хаотичной. В результате этого поисковые роботы могут попасть на такие страницы, адреса которых совсем не были предусмотрены разработчиком проекта. Для того чтобы эти страницы не засорили индекс поисковых систем, нужно правильно расставить ошибки 404 и исключить эти страницы из индекса.

отдают код ответа 200. Посетитель этого не замечает, но поисковая система воспринимает такие страницы как существующие, индексирует их и заносит в свою базу как дубликаты. Используя подобную уязвимость, на сайте вы можете создать сколько угодно большое количество дублирующих страниц, понизив вес сайта для поисковой системы.

**Типичные ошибки:**

- ▶ вместо 404 стоит 200,
- ▶ вместо 404 стоит 301 —> 404,
- ▶ вместо 404 стоят конструкции типа 302 —> 302 —> 301 —> 404.

**РЕЗЮМЕ:** Если страницы не существуют, сервер должен отдавать ответ 404.

В этой главе мы разобрали основные серверные ответы и методы работы с ними. Используя эту информацию, вы можете избежать фатальных ошибок, приводящих к исключению страниц сайта из выдачи, созданию дополнительных дубликатов и потере веса при переклейке страниц.

**Рекомендация:** Для ручной проверки кода ответа используйте бесплатный плагин LiveHTTPHeaders для Firefox. С его помощью вы сможете отследить весь путь редиректов и узнать конечный код ответа каждой страницы сайта. Автоматизированная же проверка всех страниц на код ответа доступна в анализе, проводимом программой «СайтРепорт.РФ».



В этой главе мы разобрали основные серверные ответы и методы работы с ними.

Используя эту информацию, вы можете избежать фатальных ошибок, приводящих к исключению страниц сайта из выдачи, созданию дополнительных дубликатов и потере веса при переклейке страниц.

## Глава 6

# Стандарты W3C

---



**К**онсорциум Всемирной паутины (англ. World Wide Web Consortium, W3C) – организация, разрабатывающая и внедряющая технологические стандарты для Всемирной паутины. Консорциум возглавляет сэр Тимоти Джон Бернерс-Ли, автор множества разработок в области информационных технологий.

W3C разрабатывает для Интернета единые принципы и стандарты (называемые рекомендациями), которые затем внедряются производителями программ и оборудования. Таким образом достигается совместимость между программными продуктами и аппаратурой различных компаний, что делает Всемирную сеть более совершенной, универсальной и удобной.

Миссия W3C: «Полностью раскрыть потенциал Всемирной паутины, путем создания протоколов и принципов, гарантирующих долгосрочное развитие Сети»<sup>4</sup>.

Так, существует ряд стандартов HTML-верстки. Эти стандарты носят рекомендательный характер, но, как пишет первоисточник [www.w3c.org](http://www.w3c.org), являются желательными. Что это означает? Существуют теги, имеющие смысловое значение: заголовки, визуальное выделение текста и т. п. По стандартам W3C, выделенный

<sup>4</sup>По материалам Википедии

этими тегами текстовые элементы должны иметь соответствующую смысловую нагрузку. Например, тег H1 должен содержать заголовок страницы, потому что поисковая система воспринимает содержание тегов H1 именно как заголовок, то есть очень краткое содержание страницы.

### К чему приводит игнорирование рекомендаций W3C

Верстальщики часто игнорируют рекомендации W3C. Это понятно, ведь перед ними не стоит задача поискового продвижения: они получают деньги лишь за корректное отображение сайта в популярных браузерах на распространенных разрешениях монитора. В свою очередь, браузеры стремятся обладать максимально универсальным алгоритмом распознавания HTML, чтобы сайты с ошибками верстки отображались корректно. В результате верстальщики, например, используют теги H1 для обозначения элементов меню, настоящие заголовки выделяют тегами SPAN, создают пересечения тегов заголовков H1 с ссылками A. Для верстки это неважно: браузер отобразит текст в виде заголовка, даже если текст будет оформлен тегом P, но через CSS будут прописаны соответствующие параметры визуального отображения.

Поисковые системы разборчивее браузеров: они не пытаются определить, как визуально отображается элемент страницы, а руководствуются исключительно стандартами W3C. Так, найдя на странице H1, они воспринимают его содержимое именно как заголовок страницы. И вот тут возникает самое интересное: какая информация содержится на странице, если ее заголовок выглядит как «Каталог» или «11.10.2011»? Какие услуги продает сайт, какие товары представлены на этой странице? К сожалению, при такой верстке потенциал факторов HTML-тегов остается незадействованным для поискового продвижения.



Поисковые системы разборчивее браузеров: они не пытаются определить, как визуально отображается элемент страницы, а руководствуются исключительно стандартами W3C.

Чтобы вы смогли избежать ошибок, в этой главе мы разберем самые важные правила HTML-верстки, с точки зрения стандартов W3C, а следовательно, с позиций SEO.

### 6.1. Правила HTML-разметки для SEO

Итак, поговорим об основных тегах, важных для SEO, опишем специфику их использования и подробно остановимся на правилах компоновки тегов на страницах сайта для улучшения его ранжирования.

#### ► <TITLE>

Тег TITLE используется для обозначения заголовка страницы. Это самый важный тег, он больше всех увеличивает вес страницы по ключевым словам. Сделать в нем ошибки достаточно сложно, но некоторые рекомендации стоит учесть.

Каждая страница обязательно должна иметь уникальный заголовок, нельзя дублировать заголовки TITLE между страницами сайта. Используйте ключевые слова, но не повторяйте ключ более трех раз. Если повтора избежать не удастся, включайте не менее трех разбавляющих слов между повторяющимися ключевыми словами.

Заголовок не обязательно должен быть составлен грамотно с точки зрения русского языка, но он должен быть продающим, привлекательным и содержательным. При составлении заголовков рекомендуем проанализировать заголовки конкурентов в топ-10 и составить собственные по их подобию.

#### ► <H1–H6>

Текстовые заголовки H1–H6 по значимости следуют сразу за заголовком страницы TITLE. В этих заголовках важно употреблять ключевые слова.



Тег TITLE используется для обозначения заголовка страницы. Это самый важный тег, он больше всех увеличивает вес страницы по ключевым словам. Сделать в нем ошибки достаточно сложно, но некоторые рекомендации стоит учесть.

На странице должен быть только один заголовок H1, количество подзаголовков H2–H6 не лимитировано, но не стоит делать их слишком много: вес ключевых слов распределяется по количеству заголовков на странице. Теги H1–H6 имеют очень большой вес, но их можно использовать только для визуального выделения текстовых заголовков. При этом содержимое заголовка H1 не рекомендуется дублировать между страницами сайта.

Важно понимать, что навигационная строка и элементы меню не являются заголовками страницы и должны обозначаться иными тегами, в частности тегом выделения ссылки A.

#### ► <B> и <STRONG>

Теги B и STRONG должны использоваться для визуального выделения текста жирным написанием. Визуально акцентируя ключевые части страницы, то есть обращение к посетителю, поисковые системы повышают вес страницы по запросам, содержащимся в тегах STRONG и B, но только в том случае, если их содержимое коррелирует с текстом страницы.

Важно, чтобы в этих тегах не содержалось мусора. Обозначайте ими ключевые слова и помните, что встречающиеся в этих тегах ключи нужно также употребить на странице и без тегов, чтобы поисковая система не воспринимала это как спам. Не стоит злоупотреблять тегами, несколько их употреблений на странице будет вполне достаточно.

Следует отметить, что теги B и STRONG (также как I и EM), несмотря на сходство визуального отображения, неидентичны. B – тег физической разметки, задающий жирное начертание текста. STRONG – тег логической разметки, определяющий важность помеченного текста. Такое разделение тегов на логическое и физическое форматирование изначально предназначалось



Важно понимать, что навигационная строка и элементы меню не являются заголовками страницы и должны обозначаться иными тегами, в частности тегом выделения ссылки A.



Тегом EM принято выделять курсивное написание на странице: цитату, название или просто термин в тексте. Этот тег тоже повышает вес страницы по ключевому слову, хотя по значимости уступает предыдущим.

для того, чтобы сделать HTML универсальным, в том числе не зависящим от устройства вывода информации. Теоретически, если воспользоваться, например, речевым браузером, то текст, оформленный с помощью тегов B и STRONG, будет отмечен по-разному. Получилось так, что в популярных браузерах результат использования этих тегов равнозначен. Но для нужд SEO рекомендуется использовать выделение «логическим» тегом STRONG.

► <EM>

Тегом EM принято выделять курсивное написание на странице: цитату, название или просто термин в тексте. Этот тег тоже повышает вес страницы по ключевому слову, хотя по значимости уступает предыдущим.

► <P>

Тег P обозначает текстовый абзац. Важно использовать его именно для выделения абзаца, а не каждого предложения, как это реализовано во многих CMS. Для разрыва строки можно использовать тег BR. Текст на странице лучше выделять с помощью тега абзаца P, чем заменителей вроде SPAN: последний универсален и не имеет логической нагрузки.

**Рекомендации по использованию важных для SEO тегов:**

► Грамотно используйте связки <H1>–<H6> и <P>

Заголовок H1 должен включать в себя основные ключевые слова, содержащиеся на странице. Заголовками H2–H6 обозначаем начало абзаца. В них должны содержаться основные ключевые слова, характеризующие последующий абзац P. В абзацах P, написанных до следующего заголовка, важно употребить те же ключевые слова. Ключевые слова могут повторяться и в других абзацах, но в указанной области они должны иметь максимальную концентрацию, в этом случае возникает синергия,

значительно увеличивающая вес страницы по ключевым словам.

► Обозначение тегами должно быть основано в соответствии со стандартами W3C

Все перечисленные выше теги должны нести указанный логический смысл. Например, H1-H6 следует использовать только для выделения логических заголовков, для визуального форматирования существует универсальный тег SPAN.

► Теги не должны пересекаться друг с другом

Теги, имеющие логический смысл, не должны пересекаться друг с другом. Заголовок не должен быть ссылкой, выделение жирным не должно пересекаться с заголовком. Конструкции вида `<H1><A>Заголовок 1</A></H1>` недопустимы. Если здесь важнее ссылка, нужно оставить только `<A></A>` с указанием настроек визуального отображения через CSS. Заголовок лучше дописать рядом, но для его выделения оставить только тег `<H1></H1>`.

► Теги не должны ставиться бессмысленно  
Многие CMS, содержащие блок визуального редактирования контента, работают с ошибками, порождая, например, конструкции вида `<B>&nbsp;</B>`. Такие элементы в коде страницы недопустимы, и все подобные ошибки должны быть удалены, чтобы не размывать вес значимых тегов.

► Теги не должны использоваться для обмана поисковых систем

Недопустимо употребление тегов STRONG, имеющих визуальное отображение, идентичное с текстом страницы. Выделение подобными тегами должно действительно менять визуальное отображение элемента и отграничивать его от окружающего текста.

► Теги должны закрываться



Многие CMS, содержащие блок визуального редактирования контента, работают с ошибками, порождая, например, конструкции вида `<B>&nbsp;</B>`. Такие элементы в коде страницы недопустимы, и все подобные ошибки должны быть удалены, чтобы не размывать вес значимых тегов.





Недопустимо употребление тегов **STRONG**, имеющих визуальное отображение, идентичное с текстом страницы. Выделение подобными тегами должно действительно менять визуальное отображение элемента и отграничить его от окружающего текста.

Рекомендуется все страницы сайта проверять на наличие ошибок верстки. Часто бывает так, что открывающий тег на странице есть, а закрывающего нет. Подобные ошибки, как правило, визуально заметны при заходе на страницу, но в некоторых случаях посетителям они не видны, а поисковые системы при этом не могут корректно отранжировать содержимое страницы.

## 6.2. Проверка верстки на соответствие стандартам W3C

К счастью, вам не придется в поисках ошибок вручную проверять HTML-код. Существует бесплатный сервис, определяющий его валидность. Сервис доступен по адресу <http://validator.w3.org/unicorn/>, где для проверки валидности кода вам нужно просто ввести адрес сайта.

Должен заметить, что на 100% валидных сайтов практически не существует, ведь бессмысленно доводить код до идеала. Выше мы разобрали самые важные факторы, которые и являются теми контрольными точками, которые вам нужно проверить через валидатор.

Не смущайтесь, если найдете на сайте ошибку тега `<NOINDEX>`: он не является валидным, поскольку по стандартам HTML его не существует. `<NOINDEX>` введен поисковой системой Яндекс для того, чтобы облегчить жизнь себе и оптимизаторам.



Не смущайтесь, если найдете на сайте ошибку тега `<NOINDEX>`: он не является валидным, поскольку по стандартам HTML его не существует. `<NOINDEX>` введен поисковой системой Яндекс для того, чтобы облегчить жизнь себе и оптимизаторам.

## Глава 7

# SEO-аналитика

---

### ФАКТ

#### Яндекс.Вебмастер (чему можно доверять, а чему нет)



**В** Яндекс.Вебмастере можно найти количество страниц вашего сайта в индексе, количество ссылок на ваш сайт и количество ссылающихся сайтов. Но можно ли доверять этим цифрам?

Замечено, что от апдейта к апдейту показываемые веб-мастером числа могут изменяться в разы. Например, количество страниц в индексе прыгает от 50 тысяч до 10 тысяч и обратно.

Такие скачки не соответствуют действительности и обращать внимание на них не нужно (если они не сопровождаются резкими изменениями трафика). Чтобы узнать реальное количество страниц вашего сайта в индексе, нужно проверить каждую из них вот таким запросом: `url:site.ru/page/`

То же самое относится и к выгрузке всех внешних ссылок. Данные часто скачут и не отражают реальной картины, когда ссылок на сайт стоит десятки тысяч. Многие работающие ссылки могут не оказаться в выгрузке.

#### Микроформаты

Микроформаты – это способ семантической разметки информации на веб-страницах, использующий стандартные элементы языка HTML. Фактически мы сообщаем роботу, где и какие данные лежат в коде. В ряде случаев разметка страниц микроформатами может быть альтернативой прямой передаче данных через Яндекс.Вебмастер.

Спецификации микроформатов можно найти здесь:  
<http://microformats.org/>

На данный момент поисковики пока не заявляют о влиянии микроформатов на ранжирование. Однако, очевидно, что при их помощи вы можете сформировать более привлекательный сниппет. Более привлекательный сниппет → больше CTR вашего сайта в выдаче. А CTR уже напрямую влияет на ранжирование.



### Эксперимент и наблюдение

Что такое наблюдение? Мы посмотрели на два сайта по запросу «пластиковые окна». У одного ТИЦ 40, он занимает первое место. У другого – 450, он на втором месте. Это наблюдение. Мы увидели некий факт.

Можно ли из этого сделать вывод о том, что ТИЦ не влияет на ранжирование? Конечно же, нет. Ведь на ранжирование влияют сотни параметров и их комбинаций. И один сайт мог оказаться более релевантным за счет иных параметров.

Когда оптимизатор просматривает топ, он делает много наблюдений. Заметив какой-то повторяющийся факт, мы можем сформировать гипотезу (предположение).

Пример гипотезы: «Частичное вхождение ключевого в URL страницы положительно влияет на релевантность».

Что такое эксперимент? Это когда мы искусственно создали условия для проверки гипотезы, в которых максимально отсечены посторонние факторы.

Пример эксперимента.

Мы захотели проверить гипотезу «Keywords не влияет на ранжирование». Для этого зарегистрировали 10 доменов. На каждом из них разместили сайт из одной страницы с псевдоуникальным контентом (одинаковый набор слов, объем текста и т. д.) В каждом из текстов есть уникальный ключевик, например, «сковородковед». На половине из них добавили его в meta keywords, на половине – нет. Добавили сайты в поиск, подождали пока все проиндексируются. И посмотрели выдачу по запросу «сковородковед». Если 5 сайтов с keywords оказались выше других – можно сделать вывод о подтверждении гипотезы. Если же сайты оказались перемешанными – гипотеза опровергнута.

**ФАКТ**

**Методы управления семантическим ядром для крупных проектов**



Крупные проекты обычно характеризуются большим семантическим ядром. Во многих случаях недостаточно создать хорошее семантическое ядро, нужна его регулярная актуализация. Это связано с изменением спроса: популярность части запросов падает со временем, при этом появляется спрос, не нашедший отражения в ядре.

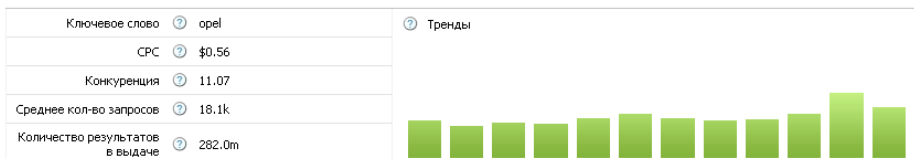
**Методология формирования обширного семантического ядра:**

1. Сбор запросов из всех доступных источников:
  - a) статистика запросов к поисковым системам,
  - b) базы ключевых слов (база Пастухова, liveinternet и аналогичные),
  - c) выгрузки ключевых слов входа из систем веб-аналитики,
  - d) оценка видимости конкурентов по запросам (semrush.com, megaindex.ru и пр.).
2. Уточнение региональной частотности, исключая подзапросы и морфологически измененные формы терминов, оценка сезонной популярности и ссылочного бюджета по запросам.
3. Кластеризация собранных запросов.
4. Формирование набора страниц приземления.
5. Сопоставление запросов страницам приземления.

**Источники данных для актуализации ядра:**

1. Динамика частотности запросов:

Обзор ключевого слова



Точные запросы - 1,827

Экспорт:

Ключ. слово	Δ Запросов	CPC	Конкуренция	Результатов	Тренды
<a href="#">opel</a>	18,100	0.56	11.07	0	
<a href="#">opel_astra</a>	9,900	0.56	11.08	0	
<a href="#">opel_corsa</a>	4,400	0.59	11.07	0	
<a href="#">opel_vectra</a>	3,600	0.50	12.55	0	
<a href="#">opel_zafira</a>	3,600	0.70	11.13	0	
<a href="#">opel_antara</a>	2,900	0.74	11.30	0	
<a href="#">opel_meriva</a>	2,400	1.11	11.07	0	
<a href="#">opel_omega</a>	1,600	0.54	11.22	0	
<a href="#">opel_tigra</a>	1,000	0.71	13.33	0	



Исходя из статистики запросов к поисковым системам и данных веб-аналитики несложно оценить трафик и доход от вывода конкретного запроса на определенную позицию в ТОПе. Таким образом, возможна эффективная приоритизация простым ранжированием запросов в порядке убывания потенциального дохода.

- a) статистика запросов к поисковым системам,
- b) semrush.com и аналогичные сервисы.

2. Собственная аналитика (конкретика зависит от выбранных KPI), например:

- a) отбрасываем запросы с высоким соотношением [стоимость продвижения/полученный доход],
- b) автоматически проверяем частотность запросов и позиции по выгруженным из системы веб-аналитики данным, выбирая эффективные запросы (в соответствии с выбранными KPI, например, с большим временем сессии, доходом или существенным числом просмотренных страниц).

Исходя из статистики запросов к поисковым системам и данных веб-аналитики несложно оценить трафик и доход от вывода конкретного запроса на определенную позицию в ТОПе. Таким образом, возможна эффективная приоритизация простым ранжированием запросов в порядке убывания потенциального дохода.

#### Курица или яйцо?

Что вначале – сайт или семантическое ядро? Обычно сбор запросов выполняется после создания сайта. По сути, подбор запросов – эквивалент оценки спроса. Чем раньше собирается семантическое ядро – тем меньше ненужной работы приходится выполнять. Из собранных и оцененных запросов логично формируются структура каталога и товарная номенклатура.

#### Корректный эксперимент в seo – миф или реальность?

Многие оптимизаторы ставят эксперименты. Однако действительно достоверные выводы можно получить крайне редко. Основные причины неудач:

- отсутствие четко сформулированной гипотезы,
- некорректная постановка эксперимента,
- некорректная обработка результатов.

И если первая и последняя проблемы решаются довольно просто, то корректная постановка эксперимента требует существенных интеллектуальных и технологических затрат.

Воспринимая поисковую систему как черный ящик, мы можем пытаться выявить некоторые закономерности, анализируя ответ в зависимости от варьируемых входных параметров. Проблема заключается в том, что на выдачу влияет огромное число факторов. По большому счету все, что нужно, – это провести несколько тестов, варьируя один параметр и удерживая неизменными все прочие. Обилие влияющих факторов усложняет постановку чистого эксперимента, однако нужно стремиться к этому.

Конечно, результаты даже корректно проведенного эксперимента нужно использовать осмысленно. Ранжирующая формула различна для различных типов запросов и регионов. Соответственно, найденные закономерности позволяют повлиять на ранжирование в диапазоне значений конкретного фактора. Иными словами, невозможно обеспечить хорошее

ранжирование документов по сколь-нибудь конкурентным запросам, влияя лишь на часть факторов.

### Альтернативные методы генерации служебных частей документов на основе правилых алгоритмов и статистики поисковых запросов

Для многих типов сайтов характерно наличие большого числа похожих документов (например, финальные страницы товаров в интернет-магазинах или вакансии на сайтах о работе). Понятно, что наилучшим является title, написанный вручную хорошим копирайтером с учетом семантического ядра документа. Этот подход хорош, но трудозатратен. Более дешевый путь – формирование title, description и прочих служебных компонент документа на основе правилых алгоритмов.

Первый этап – сбор семантического ядра для однотипных документов. Не принципиально, используется ли автомат или ручной труд, главное – максимальная полнота ядра.

Далее в этом корпусе необходимо выделить самые частотные термины и конструкции запросов.

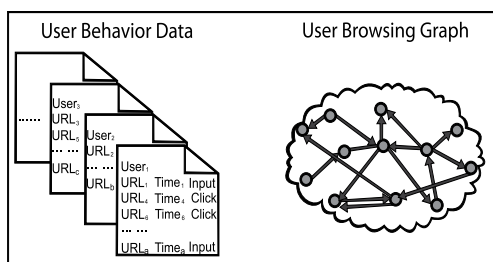
Собственно, третий шаг – написание правил. Если структура запросов примерно одинакова для всего их набора, достаточно универсального правила, кодирующего title и description. Если структура запросов различна в различных тематиках – потребуется несколько правил. Правила также нужны, если есть желание разнообразить title и description даже для сходных документов.

В этой задаче, равно как и в задаче автоматизированной перелинковки, может потребоваться случайное, но неизменное во времени распределение. Для этого отлично подходит генератор случайных чисел, инициализируемый какой-либо константой, специфичной для документа. Самый простой вариант – URL. URL обычно не меняются и уникальны по определению.


### BrowseRank – технология учета поведения пользователей для ранжирования документов

Одна из технологий учета поведения пользователей при ранжировании документов в выдаче поисковых систем – технология BrowseRank от Microsoft (<https://research.microsoft.com/en-us/people/tyliu/fp032-liu.pdf>).

#### Граф кликов вместо ссылочного графа



Основное отличие BrowseRank от PageRank – структура графа. В технологии PageRank узлами графа являются документы,

 Для многих типов сайтов характерно наличие большого числа похожих документов (например, финальные страницы товаров в интернет-магазинах или вакансии на сайтах о работе). Понятно, что наилучшим является title, написанный вручную хорошим копирайтером с учетом семантического ядра документа. Этот подход хорош, но трудозатратен. Более дешевый путь – формирование title, description и прочих служебных компонент документа на основе правилых алгоритмов.



**Основной источник данных о поведении пользователей – браузерные бары.** Все ведущие поисковые системы имеют плагины, обеспечивающие мониторинг активности значительной доли аудитории. Для обработки данных о поведении пользователей предложено использовать цепи Маркова с непрерывным временем. Экспериментальные данные демонстрируют преимущество алгоритма BrowseRank по сравнению с алгоритмами PageRank и TrustRank в определении важности документов, борьбе со спамом и ранжировании.

а ребрами – ссылки. В технологии BrowseRank ребра образуют не ссылки, но клики (переходы). Помимо этого, сохраняется метainформация о продолжительности сессии в рамках документа.

#### **Очевидны два существенных преимущества BrowseRank:**

- существенно лучшая устойчивость к ссылочному спаму,
- учет продолжительности сессии позволяет оценить полезность документа для посетителя.

Поведенческий граф более достоверно отражает процесс веб-серфинга, а, следовательно, он более полезен для расчета значимости документов. Наибольшее число посещений страницы и более продолжительное время, проведенное на ней, означают большую важность страницы.

**Основной источник данных о поведении пользователей – браузерные бары.** Все ведущие поисковые системы имеют плагины, обеспечивающие мониторинг активности значительной доли аудитории. Для обработки данных о поведении пользователей предложено использовать цепи Маркова с непрерывным временем. Экспериментальные данные демонстрируют преимущество алгоритма BrowseRank по сравнению с алгоритмами PageRank и TrustRank в определении важности документов, борьбе со спамом и ранжировании.

По сути, собираемые данные можно представить записью вида URL; TIME; TYPE [input | click]. Предполагается два пути перехода на документ: по ссылке с другого документа (click), либо набором URL в адресной строке браузера (input). Механизм извлечения данных о переходах:

#### **1. Сегментация сессии.**

Новая сессия инициируется в случае 30-минутной и более паузы с момента предыдущей активности, либо в случае ввода названия сайта в адресную строку.

#### **2. Формирование пар URL.**

В рамках каждой сессии создаются пары URL из соседних записей. Пара URL означает, что переход был осуществлен при помощи ссылки.

#### **3. Формирование начального распределения.**

В каждой сессии, сегментированной по типу перехода, первый URL введен непосредственно пользователем. Такие URL мы считаем «доверительными» и называем этот трафик «зеленым». Обработывая данные о поведении пользователей, мы считаем переходы на эти URL следствием случайного распределения. Нормализация на частоту посещения этих документов дает начальные вероятности посещения соответствующих страниц.

#### **4. Извлечение продолжительности сессии.**

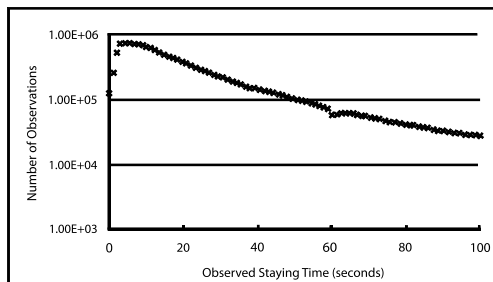
Для каждой пары URL продолжительность сессии первого URL вычисляется простой разностью дат. Если URL был последним в сессии, возможны два варианта. Для сессий, сегментированных по времени, продолжительность просмотра последнего URL рассчитывается на основании данных о просмотре других страниц. Для сессий, сегментированных по типу, время



просмотра последнего URL рассчитывается исходя из времени начала следующей сессии.

### Результаты применения BrowseRank

Microsoft провела два исследования. Первый был сделан на уровне сайта для выявления важных сайтов и подавления спама. Второй эксперимент – на уровне документа для тестирования BrowseRank с целью улучшения ранжирования.



Для первого эксперимента использовался набор данных из примерно 3 миллиардов записей, содержащий примерно 950 миллионов уникальных URL. Любопытно, что распределение количества просмотров по времени отлично описывается классической экспонентой. В рамках этого эксперимента страничные данные не использовались, они были агрегированы на уровне сайтов. Полученный поведенческий граф состоял из 5,6 миллиона узлов и 53 миллионов ребер.

Любопытна выборка топ-20 сайтов, полученных ранжированием по трем разным алгоритмам. BrowseRank отлично отранжировал MySpace, Youtube, Facebook и прочие сайты, для которых характерно большое время сессии. На случайной выборке в 10 тысяч сайтов, вручную размеченной ассессорами, показано, что BrowseRank эффективнее, чем TrustRank и PageRank, решает задачу фильтрации спам-сайтов.

Второй эксперимент не менее интересен.

Факторы ранжирования документов можно грубо разделить на две группы: факторы релевантности и факторы важности. Предполагаем итоговую функцию релевантности линейной комбинацией этих двух групп факторов:

$$\theta * \text{rankrelevance} + (1-\theta) * \text{rankimportance},$$

где  $\theta$  лежит в интервале  $[0;1]$ .

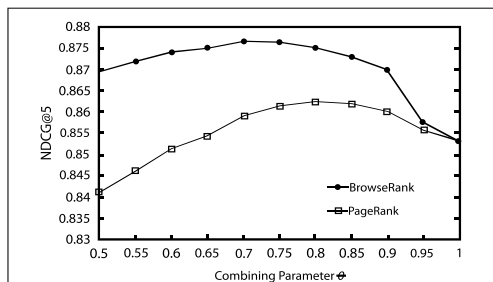


Figure 7: Search performance in terms of NDCG@5 for BrowseRank and PageRank



Любопытна выборка топ-20 сайтов, полученных ранжированием по трем разным алгоритмам. BrowseRank отлично отранжировал MySpace, Youtube, Facebook и прочие сайты, для которых характерно большое время сессии. На случайной выборке в 10 тысяч сайтов, вручную размеченной ассессорами, показано, что BrowseRank эффективнее, чем TrustRank и PageRank, решает задачу фильтрации спам-сайтов.



**Поисковый спам** – это попытки обмана поисковой системы и манипулирования ее результатами с целью завышения позиции сайтов (страниц) в результатах поиска. Сайты, использующие поисковый спам, могут быть понижены при ранжировании или исключены из поиска Яндекса из-за невозможности их корректного ранжирования.

Данные взяты на выборке из сайтов, полученных для 8 000 запросов. Каждая пара запрос-документ оценивалась тремя ассессорами по бинарной шкале [релевантно; нерелевантно]. Релевантными сочтены документы, получившие по запросу не менее 2 оценок «релевантно».

Результат: алгоритм BrowseRank обеспечивает существенно лучшее качество вплоть до полного подавления факторов важности факторами релевантности ( $\theta \sim 0,9$ ).

### Текстовый спам глазами поиска и современные метрики качества текстов

Точного определения поискового спама нет, есть лишь общая формулировка.

Поисковый спам – это попытки обмана поисковой системы и манипулирования ее результатами с целью завышения позиции сайтов (страниц) в результатах поиска. Сайты, использующие поисковый спам, могут быть понижены при ранжировании или исключены из поиска Яндекса из-за невозможности их корректного ранжирования.

<http://company.yandex.ru/legal/termsfuse>

Понятно, что точные критерии спама поисковые системы не озвучивают, поскольку это неминуемо спровоцирует генерацию околоспамовых текстов оптимизаторов и веб-мастерами. Между тем в арсенале поисковых систем большое количество разнообразных методов определения спама. Рассмотрим некоторые из них.

Общеизвестно, что включение термина в title улучшает ранжирование документа по этому запросу поисковыми системами. Это обуславливает традиционный спам в title, поскольку оптимизаторы пытаются употребить максимальное число терминов в заголовке документа. Исследователи из Microsoft обнаружили (<http://research.microsoft.com/pubs/65140/www2006.pdf>), что вероятность спамовости документа резко возрастает при длине title более 25 слов:

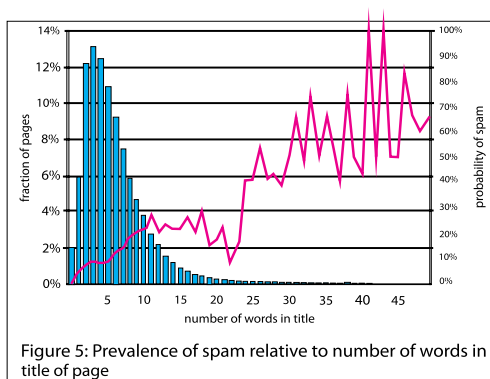


Figure 5: Prevalence of spam relative to number of words in title of page

Распределение по средней длине слова также неплохо выявляет спам. Тексты, содержащие много коротких или длинных слов, неестественны и, вероятно, спамны:

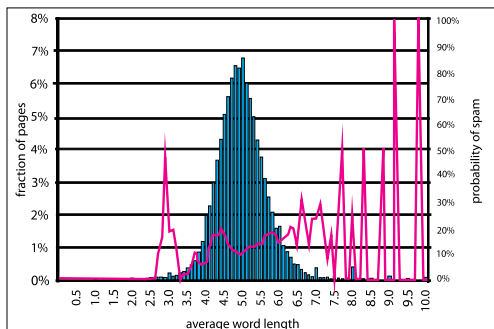
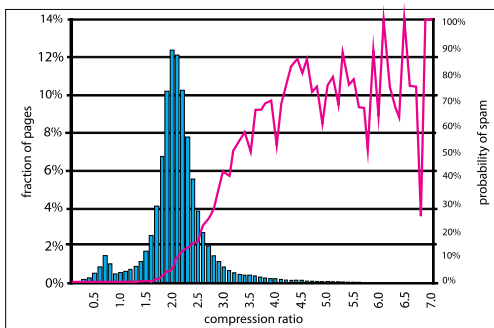
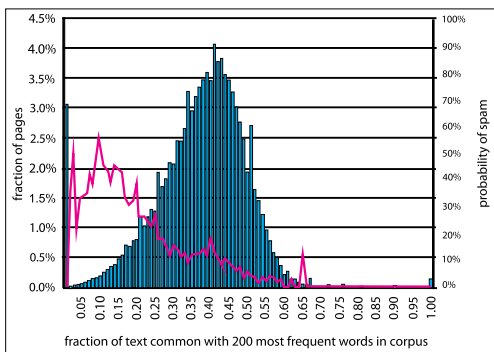


Figure 6: Prevalence of spam relative to average word-length of page

Весьма оригинальный и дешевый метод обнаружения спама – оценка сжимаемости текста. Для спамовых документов, содержащих множество повторений терминов, характерна высокая сжимаемость:



Интересна техника определения спамовости по доле самых частых терминов корпуса в тексте. Видно, что в спамовых текстах доля частых слов ниже, а доля редких слов – выше, чем в естественных текстах:



Примечательна публикация разработчиков Яндекса ([http://download.yandex.ru/company/A\\_Kustarev\\_A\\_Raigorodsky\\_poisk\\_neestestvennih\\_textov\\_statia.pdf](http://download.yandex.ru/company/A_Kustarev_A_Raigorodsky_poisk_neestestvennih_textov_statia.pdf)),



Общеизвестно, что включение термина в title улучшает ранжирование документа по этому запросу поисковыми системами. Это обуславливает традиционный спам в title, поскольку оптимизаторы пытаются употребить максимальное число терминов в заголовке документа.



В современных условиях намного продуктивнее писать хорошие тексты, разносторонне описывающие продукт. Важны стилистика и потребительская ценность текста. Соответственно, самая честная метрика качества текста – отношение пользователей (продолжительность сессии, степень конверсии и пр.).

в которой оценивалась частота встречаемости пар терминов в естественных и синонимизированных текстах. Показательно, что спамовые документы содержат большее количество редких пар и меньшее количество частых пар, чем естественные документы.

Существует множество методов выявления текстового спама. Главная идея очевидна – использование традиционных оптимизаторских метрик («плотность ключевого слова», «длина текста» и пр.) давно потеряло актуальность.

В современных условиях намного продуктивнее писать хорошие тексты, разносторонне описывающие продукт. Важны стилистика и потребительская ценность текста. Соответственно, самая честная метрика качества текста – отношение пользователей (продолжительность сессии, степень конверсии и пр.). Действительно, если качество навигации или юзабилити сайта принято оценивать по поведению пользователей, почему качество текстов нужно оценивать иначе?

**РЕЗЮМЕ:** *Бессмысленно делать нечто исключительно для поисковых систем, как ни банально звучит этот тезис. У поисковых систем достаточно ресурсов для подавления спамерских активностей.*

## Глава 8

# Исследование «SEO-альманах 2010»

---



**В** конце 2010 года было проведено масштабное исследование факторов ранжирования, целью которого стала оценка значимости этих факторов. Подобные исследования наши зарубежные коллеги проводят с периодичностью раз в два года, публикуя результаты на [seomoz.com](http://seomoz.com). В российском проекте приняли участие признанные эксперты рынка, руководители топовых компаний, специалисты, успешно продвигающие сайты в самых конкурентных тематиках, разработчики программ и сервисов, которыми пользуется весь рынок.

В экспертный совет вошли:

1. Анар Бабаев (Click.ru)
2. Александр Баженов (Biplane)
3. Василий Ткачев (All in Top)
4. Леонид Гроховский (ArrowMedia)
5. Николай Хиврин (ALTWeb Group)
6. Елизавета Трибунская (Web Advance)
7. Сергей Кошкин (SmartSEO)
8. Евгений Смирнов (Viaset.ru)
9. Роман Клевцов (Optimism.ru)
10. Сергей Карпович (Apexes)
11. Дмитрий Жохов (UnMedia)
12. Алексей Довжиков (eLama)
13. Тимофей Квачев (TRINET)
14. Валентин Домбровский (NextUpMedia)
15. Константин Леонович (Sape.ru)
16. Алексей Жуков (Новотех)
17. Антон Агапов (WebExpert)

18. Иван Севостьянов (ВебПроекты)
19. Олег Сквородников (VIPRO)
20. Виктор Нагайцев (PerfectSeo)
21. Денис Юрасов (Медведев Маркетинг)
22. Михаил Райцин (Корпорации РБС)
23. Василий Сорокин (Techart)
24. Александр Люстик (MyOnlineStudio)
25. Станислав Поломарь (web-it.ru)
26. Алексей Елфимов (i-Media)

Исследование «SEO-альманах 2010» проводилось с помощью голосования. Все факторы были распределены по группам, и каждый из них оценивался по критерию влияния на ранжирование. Конечная оценка определялась на основании голосов экспертов. Рассматривалось более 200 факторов ранжирования: внутренние, внешние, поведенческие, наиболее распространенные их связки, а также факторы, связанные с ранжированием картинок и определением аффилиатов.

Методология исследования выстраивалась на анкетировании. Каждый из экспертов оценивал факторы по степени их значимости:

- 0 – не влияет вообще,
- 1 – влияет незначительно,
- 2 – стоит учитывать,
- 3 – сильно влияет,
- 4 – важнейший фактор.

Оценки каждого фактора суммировались, после чего было подсчитано процентное соотношение значимости каждого фактора от значимости того, который имеет самый высокий рейтинг в группе. Факторы объединялись в группы по признаку воздействия: внешние факторы, внутренние, поведенческие, аффилированность, продвижение по картинкам.

Все факторы были отсортированы по степени важности и отображены в процентных долях. Отрицательные получили отрицательное значение.



Исследование «SEO-альманах 2010» проводилось с помощью голосования. Все факторы были распределены по группам, и каждый из них оценивался по критерию влияния на ранжирование. Конечная оценка определялась на основании голосов экспертов. Рассматривалось более 200 факторов ранжирования: внутренние, внешние, поведенческие, наиболее распространенные их связки, а также факторы, связанные с ранжированием картинок и определением аффилиатов.

Итоги исследования стоит воспринимать следующим образом:

0–20% означают, что фактор не влияет на продвижение,

20–40% – влияние фактора незначительно,

40–60% – фактор важно учитывать при продвижении,

60–80% – фактор значительно влияет на продвижение,









80–100% – фактор нужно учитывать обязательно.

Результаты исследования с комментариями экспертов были опубликованы по адресу <http://топэксперт.рф> и сразу получили широкий резонанс: за первую неделю после анонсирования только на официальном сайте исследования побывало 10 тысяч человек. В Интернете зафиксировано более 100 публикаций, в том числе в крупнейших отраслевых СМИ, таких как [seonews.ru](http://seonews.ru), [searchengines.ru](http://searchengines.ru), [webmasters.ru](http://webmasters.ru), [advertology.ru](http://advertology.ru) и других, также написаны статьи в нескольких отраслевых журналах, результаты исследования освещались на конференциях и семинарах.






## Влияние HTML-тегов на вес ключевого слова для страницы

TITLE	100%
H1-H6	71%
STRONG	34%
Description	31%
B	27%
EM	25%
P	24%
Keywords	24%
UL -> LI & OL -> LI	13%

## Влияние HTML-тегов на вес ключевого слова для страницы



 Ключевое слово в H1-H6 & ключевое слово в P	54%
 Ключевое слово в STRONG, B, EM & ключевое слово на странице	34%
 Ключевое слово в KEYWORDS и ключевое слово на странице	27%
 Последовательное употребление заголовков	24%
 Пересечение тегов STRONG, B, EM & A href=...	-8%
 Пересечение тегов H1-H6 A href=...	-9%
 Повтор заголовков H1, H1, H1	-15%
 Пересечение тегов H1-H6 & STRONG, B, EM	-25%

## Содержание заголовков H1-H6, TITLE






 Прямое вхождение ключевого слова в заголовке	84%
 Уникальность заголовка в отношении ко всем страницам сайта	48%
 Близость заголовка (H1-H6) к началу документа	47%
 Близость ключевого слова к началу заголовка	43%
 Длина заголовка	36%
 Грамматически правильное содержание заголовка	35%
 Уникальность заголовка в отношении ко всем сайтам в индексе	33%
 Разбавление заголовка неключевыми словами	18%












## Содержание элементов абзаца STRONG, EM, B

 Прямое вхождение ключевого слова	37%
 Разбавление элемента неключевыми словами	19%



## Ранжирование по ключевым словам

Наличие прямого вхождения ключевого слова в тексте страницы	94%
 Словоформа ключевого слова с сохранением части речи	66%
 Наличие как прямого вхождения ключа, так и его словоформы	65%
 Словоформа ключевого слова с изменением части речи	40%
 Соответствие регистра ключевого слова	28%
 Удаление частей составного словосочетания друг от друга	-77%









## Текст страницы

 % уникального контента по отношению ко всему индексу	82%
 % уникального контента по отношению ко всему сайту	76%
 Разнообразное оформление текста	47%
 Близость ключевых слов к началу страницы	45%
 Соответствие объема текста на странице сайтам конкурентов	40%
 Выделение абзаца текста в тег <P>	30%
Наличие контента на иностранном языке	5%
Высокая плотность ключевых слов на странице	-5%
 Концентрация повторов ключевого слова в одной точке	-22%
 Визуальное уравнивание разметки текста с обычным текстом	-35%
Ошибки кодировки	-52%
 Видимость контента (речь о таких эффектах, как display:none)	-57%





## Сервер

301 редирект при постоянном изменении адреса страницы	67%
Отсутствие на сайте значительного количества 404 ошибок	64%
 Использование 404 кода ответа для 404 страниц	63%
Высокий аптайм сервера	63%
Высокая скорость загрузки документа	57%
Время отклика	53%
 1-шаговая конструкция редиректоров	40%
302 редиректа при временном изменении адреса страницы	30%
Географическая зона сервера	27%


## URL

 Наличие ключевого слова в имени домена	83%
 Наличие ключевого слова на транслите в адресе страницы	71%
 Наличие ключевого слова в названии категории	61%
 Наличие ключевого слова в названии документа	59%
 Наличие ключевого слова на английском в адресе страницы	49%
ЧПУ	48%
 Наличие словоформы ключевого слова в адресе страницы	45%
 Наличие части ключевого слова в адресе страницы	37%
 Большая глубина вложенности категории с документом	-7%


## Возрастные факторы

 Возраст индексации	83%
Динамика появления уникального контента на сайте	59%
 Возраст домена	55%
 Возраст URL-страницы	55%
Динамика появления контента на сайте	51%
 Возраст контента на странице	46%



## Влияние картинок на ранжирование

Ключевое слово в ALT картинки	47%
 Ключевое слово в названии картинки	43%
Ключевое слово в TITLE картинки	36%

## Географическое расположение

Указание регионов в Яндекс.Каталоге	95%
Наличие адресов и телефонов региональных филиалов на сайте	87%
 Географическая близость офиса к продвигаемому региону	77%
Упоминание региональных центров в тексте страниц	64%

## Перелинковка

 Ограничение количества исходящих и внутренних ссылок	66%
Закрывать от индексации исходящие внешние ссылки	53%
Использование сквозных ссылок для внутренней перелинковки	18%
 Страница ссылается на другую несколько раз	-2%
Страница ссылается на себя	-6%

## CMS

Использование бесплатных общедоступных CMS	-25%
Использование стандартных шаблонов	-38%

## Карта сайта

Наличие карты сайта	58%
Актуальность карты сайта	54%
Наличие страницы в карте сайта	48%
Указание приоритета индексации страниц в карте сайта	40%

## Взаимодействие с Яндексом

Наличие сайта в Яндекс.Вебмастер	23%
Факт выгрузки обратных ссылок	-12%
Факт ранее наложенных санкций	-54%

## МНЕНИЕ

**Сергей Панков,**  
технический директор, Rookee



### Преамбула

Прежде всего, стоит отметить то титаническое трудолюбие и скрупулезность, которые помогли автору в проведении исследования о факторах ранжирования. Совершенно искренне выражаем благодарность автору, его бесконечному профессионализму и желанию сделать рынок 2.0 более цивилизованным. Надеемся, что Леонид и дальше будет радовать сообщество качественными и своевременными исследованиями.

### Предостережение

Материалы исследования послужат хорошей теоретической базой для новичков и seo-энтузиастов, желающих повысить качество своего сайта и добиться новых высот в поисковой выдаче.

Представленное исследование – это не кнопка «в топ» и даже не рецепт долголетия и счастья. Это скорее набор критериев, по которым можно определить качество отдельно взятой страницы веб-сайта, оценить ее привлекательность с точки зрения поискового робота, который ищет наиболее релевантный ответ на вопрос пользователя.

### Что мы об этом думаем

Исследование еще раз подтвердило высокую точность той информации, что была получена в результате наших внутренних экспериментов. Знания об этих факторах заложены в основу алгоритмов наших сервисов:

- технический аудит,
- текстовый анализ страницы,
- анализ конкурентов топ-10.

Однако стоит отметить, что в среде seo-профессионалов сильны различные мифы. Одна неосторожная статья от Дэна Расковалова о его путешествии на Бали может породить десяток предположений и теорий, которые будут дотошно анализироваться гуру всех мастей.

Например, значимость следующих факторов мы ставим под сомнение:

1. Прямое вхождение ключевого слова куда бы то ни было.
2. Разнообразное оформление текста.
3. Соответствие объема текста на странице сайтам конкурентов.
4. Закрывание от индексации исходящих внешних ссылок.
5. Факторы, связанные с CMS.

Ценность блока критериев под названием «текст ссылки» ставится под большое сомнение. Все мы читали знаменитый мадридский доклад Яндекса за 2009 год. После него практически бессмысленно рассуждать на тему, что лучше – «вечная»



Представленное исследование – это не кнопка «в топ» и даже не рецепт долголетия и счастья. Это скорее набор критериев, по которым можно определить качество отдельно взятой страницы веб-сайта, оценить ее привлекательность с точки зрения поискового робота, который ищет наиболее релевантный ответ на вопрос пользователя.



Безусловно, Интернет меняется. И то, что было актуально вчера, сегодня уже не работает. Остается пожелать одно: анализируйте и делайте выводы, остальное — дело техники.

ссылка или «биржевая», «прямое вхождение или со словоформами» и т. д.

Следует отметить, что за пределами исследования остались вопросы, ответы на которые хотелось бы получить в будущем:

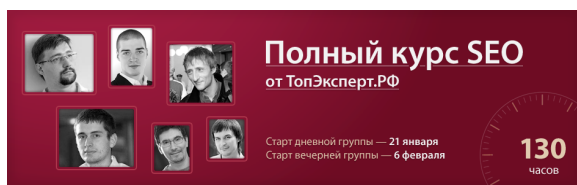
1. Верстка.
2. Поведенческие факторы.
3. Соответствие тематики донора и акцептора.
4. «Трафиконосность» ссылок с доноров.

Конечно, наши возражения не умаляют всех достоинств проделанной работы. На то и нужны исследования, чтобы привлечь внимание к каким-то важным вопросам и заставить нас самостоятельно искать ответы на поставленные вопросы.

#### **Выводы**

Безусловно, Интернет меняется. И то, что было актуально вчера, сегодня уже не работает. Остается пожелать одно: анализируйте и делайте выводы, остальное — дело техники.

## Семинары и курсы



**Полный курс SEO**  
от **ТопЭксперт.РФ**

Старт дневной группы — 21 января  
Старт вечерней группы — 6 февраля

**130**  
часов

### Полный курс SEO от «ТопЭксперт.РФ» (130 часов)

Хочу представить вам самый полный учебный курс по SEO. После прохождения этого курса вы сможете самостоятельно заниматься продвижением сайтов. Продолжительность курса – 130 часов, это 62 часа занятий в аудитории + 40 часов практики в течение месяца + 20 часов консультаций на вебинарах. Программа корректируется и дорабатывается совместно с признанными экспертами в области поискового продвижения. По окончании курса проводится сертификация, результаты которой передаются HR-специалистам ведущих SEO-компаний.

#### Авторы и преподаватели курса:

- ▶ Леонид Гроховский, директор по технологии продвижения и автоматизации, «Оптимизм.ру»,
- ▶ Станислав Поломарь, руководитель направления продвижения сайтов, Web-IT,
- ▶ Алексей Чекушин, руководитель департамента SEO, Wikimart.ru,
- ▶ Михаил Сливинский, ведущий аналитик, Wikimart.ru,
- ▶ Станислав Ставский, ex. аналитик веб-поиска Яндекс.

#### Для кого этот курс:

- ▶ Для начинающих SEO-специалистов, которые хотят получить прочные базовые знания.

- ▶ Для продвинутых SEO-специалистов, которые хотят разобраться в информационном поиске.
- ▶ Для студентов, которые хотят получить высокооплачиваемую и интересную работу.
- ▶ Для оптимизаторов, которые за несколько лет все забыли и хотят обновить актуальность своих знаний.
- ▶ Для интернет-маркетологов, которые хотят очень глубоко погрузиться в SEO.

#### Дни проведения:

##### Блок 1. Информационный поиск

- 21 января 11.00–20.00
- 22 октября 11.00–20.00

##### Блок 2. Поисковое продвижение

- 28 января 11.00–20.00
- 29 января 11.00–20.00
- 4 февраля 11.00–20.00
- 5 февраля 11.00–20.00

##### Блок 3. Аналитика

- 11 февраля 11.00–20.00
- 12 февраля 11.00–15.00

**Экзамен:** 19 февраля 11.00–20.00

#### Формат обучения:

- Очное обучение – проходит в Москве в учебной аудитории
- Заочное обучение – посредством вебинар-трансляции на площадке COMDI

- Всем слушателям на время обучения и в течение 2 месяцев после экзаменов предоставляется доступ к записям занятий на COMDI.ru

### Уникальность курса:

- ▶ Программа составлена людьми, имеющими большой опыт руководства и обучения сотрудников в SEO-компаниях, а значит, и вас учить будут тому, что действительно применимо и необходимо.
- ▶ Программа является коллективным трудом, а не частным мнением, а значит, нам удастся избежать субъективности.
- ▶ По итогам курса проводится сертификация, резюме успешных студентов передаются в HR-службы многих SEO-компаний, что гарантирует их трудоустройство.
- ▶ Это самый полный, профессиональный и практичный курс по SEO.

**И самое главное** – в течение всего курса вы будете получать задания в формате реальной стажировки, а саппорт будет помогать вам в трудных ситуациях.

## Предварительная программа аудиторных лекций учебного курса SEO

### Блок 1. Информационный поиск

HTTP-протокол

- (GET)
- Понятие о протоколе
- Структура протокола
- Коды состояния
- IP, Прокси
- Заголовки
- Основные механизмы протокола
- DNS, работа хостинга
- Языки и кодировки

### Как устроены поисковые системы

- Поисковый робот-индексатор

- Скорость индексации
- Краулинговый бюджет
- Частота индексации
- Типы ботов (быстробот, блоги и т. п.) из описания веб-мастера Яндекса
- Инвертированный индекс
- Ранжирование
- Результаты поиска (SERP)
- Кластеризация

### Идеология поисковых систем

- Качество поиска
  - Релевантность
  - Свежесть
  - Полнота индекса
  - Разнообразии результатов поиска
  - Антиспам (в т. ч. про цепи Маркова и дорвеи)
  - Сニппеты
  - Производительность
  - Вертикальный поиск
- Важнейшие ценности
- Способы монетизации
- Взаимодействие с пользователями
- Конкуренция
- Ответственность
- Нарушения и санкции

### Введение в информационный поиск

- Булев поиск
- Лексикон и списки словопозиций
- Словари и нечеткий поиск
- Построение индекса
- Ранжирование, взвешивание терминов и модель векторного пространства
- Ранжирование в полнофункциональной поисковой системе



- Вертикальный поиск
- Оценка качества поиска
- Расширение запроса, переформулировки
- Машинное обучение
- Языковые модели для информационного поиска
  - Стемминг
  - Морфология
  - Лемматизация

- Кластеризация

#### Поисковые технологии

- IDF TF
- BM 25
- PageRank
- HITS
- ClickRank
- BrowseRank
- SpamRank

#### Обработка информации

- Алгоритм шинглов
- Разбиение текста на пассажи

#### Типы запросов с точки зрения поисковых систем

- Типы классификации
- По бродеру
- Количество
- Коммерческие и некоммерческие

#### Блок 2. Поисковое продвижение

##### Теория

- SEO-факторы
  - Текстовые
    - Плотность ключевых слов
    - Зоны документа (TITLE, H1 и т. д.)
    - Фильтрации и антиспам

- Ссылки
  - Внутренние ссылки
  - Внешние ссылки
  - Ссылочное ранжирование
    - Статическое (PR, ТИЦ)
    - Динамическое (анкорное)
- Host-факторы
  - Возраст
  - Key в URL, в домене
  - Видимость
- Поведенческие факторы
  - Поведение в серпе
  - Поведение на сайте
    - Время просмотра
    - Количество страниц
    - Источники трафика
    - Переходы по ссылкам
- Социальные факторы
- Региональность

##### Практика

##### Внутренние факторы

- Перелинковка
  - Базовые схемы перелинковки
    - Комплексное продвижение всего сайта
    - Продвижение главной страницы под ВЧ
    - Продвижение разделов под СЧ
    - Продвижение внутренних страниц под НЧ
  - Локальные схемы перелинковки
    - Кольцо
    - Куб
    - Звезда
  - Правила перелинковки

- Способы перелинковки
  - Составление текста ссылки для перелинковки
  - Создание модулей для автоматизации перелинковки (начальный уровень)
    - Портал
    - Интернет-магазин
    - Корпоративный сайт
    - Каталог
  - Устранение дубликатов
    - Чем вредны дубликаты
    - Классификация дубликатов
    - Устранение дубликатов на примерах
    - Способы устранения дубликатов
  - Инструкции для поисковых систем
    - Зачем нужен robots.txt?
    - Robots.txt – универсальные директивы
    - Robots.txt – директивы для Яндекса
    - Правила составления robots.txt
    - Разбор robots.txt на примерах
    - Зачем нужен sitemap.xml
    - Директивы sitemap.xml
    - Правила составления sitemap.xml
    - Пример sitemap.xml
    - Принцип взаимодействия sitemap.xml и robots.txt
    - Автоматизация добавления страниц в sitemap.xml
    - Зеркала сайтов, переклейка
    - Данные Яндекс.Вебмастер
    - Данные Google.Вебмастер
  - Уникальность контента
    - Как защитить контент от воровства?
    - Как воровать контент?
  - Отсутствие и недостаток контента
  - Автоматизация генерации контента
    - Автоматизация генерации описаний товаров
    - Автоматизация генерации технических характеристик (уровень 1)
    - Автоматизация генерации технических характеристик (уровень 2)
    - Автоматизация заполнения Description, Keywords, TITLE
    - Обнаружение дубликатов
    - Инструкция по обнаружению дубликатов вручную
    - Способы автоматизации обнаружения дубликатов
  - Коды ответов сервера
    - Код ответа 20\*
    - Коды ответов 301/302/303/307/3\*\*
    - Код ответа 40\*
    - Код ответа 50\*
  - Стандарты W3C
    - Правила HTML-разметки для SEO
    - Проверка верстки на соответствие стандартам W3C
  - Подготовка ТЗ на доработку сайта
  - Типичные ошибки, допускаемые при работе с внутренними факторами
  - Типичные ошибки разработки сайтов
  - Типичные ошибки юзабилити
  - Типичные ошибки копирайтеров
- Подробнее про внешние факторы**
- Все источники привлечения внешних ссылок, оценка эффективности и экономическая целесообразность. Сравнительная таблица
    - Обычные покупные ссылки
    - Ссылки в виде контекстных объявлений

- Статьи
  - Пресс-релизы и посты в блогах
  - Социальные сети
  - Форумы
  - Естественные ссылки
  - Возможности автоматизации
  - Продвижение молодых сайтов и сайты с историей: отличия в работе со ссылками
  - Продвижение крупных сайтов: особенности в работе со ссылками
  - Если по ссылке кликают – она работает лучше. Миф или реальность?
  - Критерии выбора ссылок: на что на самом деле стоит обращать внимание
    - Оценка эффективности ссылок через Яндекс.Метрика
    - По каким критериям отбирать площадки
    - ТИЦ, тематичность, количество внешних – что в действительности важно сейчас
  - Альтернативные источники внешних ссылок
    - Веб-PR
      - Бриф для клиента
      - Подбор площадок для публикации
      - Пресс-релизы, статьи, интервью, отзывы
      - Заказ статей
      - Примеры успешных PR-стратегий для SEO
    - SMM
      - Возможности получения ссылок из социальных сетей
      - Критерии оценки качества блогов
      - Ссылки из Twitter и Facebook для SEO
    - Примеры успешного применения
  - SMO
    - Возможности SMO
    - Создание ценности для посетителей
    - Рейтинг
    - Акции
    - Закрытый клуб
    - Предложение о репосте
    - Примеры успешного применения
  - Инструментарий
    - Аналитика – «СайтРепорт.РФ»
    - Подбор запросов – KeyCollector
    - Биржа ссылок – Sape.ru
    - Биржа ссылок – PR.sape.ru
    - Биржа ссылок – Blogun.ru
    - Биржа ссылок – GoGetLinks.net
    - Биржа ссылок – MiraLinks.ru
    - Агрегатор – Rookee
    - Агрегатор – MegaIndex
    - Плагин для браузера – NetPromoterBar
    - Анализ ссылок – Solomono.ru
    - Анализ ссылок – DriveLink
  - Составление ссылочной стратегии
    - Обычные ссылки
    - PR-ссылки
    - Ссылки без анкоров
    - Мертвые ссылки
    - Ссылки с социальных сетей
  - Размещение ссылок
  - Проверка ссылок
- Подробнее о поведенческих факторах**
- Теория: развитие поведенческих метрик
  - 3 группы поведенческих факторов

- переходы по ссылкам
- результаты поиска
- поведение на сайте
- Где покупать естественные ссылки
- Связки поведенческих факторов: как это работает
- Исследование пользователей: какими данными располагает Яндекс
- 3 способа накрутки поведенческих факторов
  - имитация живых пользователей: создание «живых» ботов
  - ботнет
  - биржа
- Мифы и заблуждения о поведенческих факторах
- Отличия ПФ для разных тематик
- Возможности Яндекс.Метрики: демонстрация функционала

#### Составление семантического ядра

- Запрос выгрузки структуры сайта
- Подбор запросов
- Проверка запросов
- Группировка запросов
- Формирование новой структуры сайта

#### Аудит

- Изучение сайта
- Классификация типов страниц
- Классификация разделов сайта
- Изучение <HTML> шаблонов по типам страниц
- ТЗ по <HTML> доработке страниц
- Изучение данных Яндекс.Вебмастер
- Изучение данных Google.Вебмастер
- Изучение данных Яндекс.Метрика
- Сравнение индексов

- Изучение перелинковки
- Изучение серверных настроек
- Изучение контента
- Поиск дубликатов
- Поиск внешних ссылок
- Анализ и сравнение индексации в поисковых системах
- Анализ внутреннего анкор-листа
- Лингвистический анализ сайта
- Поиск и анализ исходящих внешних ссылок
- Анализ распределения весов по страницам
- Анализ страниц с 404 ошибками
- Анализ редиректов
- Создание sitemap.xml
- Анализ sitemap.xml, robots.txt
- Оценка размера страниц и скорости их загрузки
- Анализ уровней вложенности
- Поиск полных дублей
- Поиск неполных дублей
- Анализ заполненности страниц
- Поиск пустых страниц и страниц-простыней
- Анализ TITLE, Keywords, Description на наличие дубликатов, пустоты, переспама
- Анализ уязвимостей сайта
- Анализ конкурентов
- Проверка сайта на наложение фильтров
- Анализ внешних ссылок
- Поиск дубликатов контента на других сайтах
- Лексикологический анализ сайта

### Типичные ошибки оптимизатора

- Типичные ошибки, допускаемые при работе с внешними факторами
- Типичные ошибки PR
- Типичные ошибки SMM
- Типичные ошибки разработки

### Накрутка

- Накрутка внешних факторов
- Накрутка внутренних факторов
- Накрутка поведенческих факторов

### Копирайтинг

- Взаимодействие с копирайтерами
- Как правильно составить ТЗ по копирайтингу
- Определяем уникальность текста
- Бриф для клиента
- Входные страницы
- Что лучше продает: содержание, заголовки, оформление или иллюстрации?
- Нестандартные способы получения контента

### Блок 3. Аналитика

#### Введение в SEO-аналитику

- Кто такой SEO-аналитик
- Примеры работы SEO-аналитика
- Аналитика
  - Эмпирический анализ
  - Статистический анализ
  - Теория вероятности
  - Нестандартное мышление
  - Логические задачи
  - Собеседование на SEO-аналитика
- Мозговой штурм
- Автоматизация
  - Сбор позиций

- Проверка индексации
- Проверка текстов
- Перелинковка
- Покупка ссылок
- Как писать ТЗ
  - Правила написания ТЗ
  - Таблицы
  - Модули
  - Интерфейс
  - Примеры ТЗ
- Область применения знаний и навыков SEO-аналитика
- Источники получения информации, список литературы
- Сервисы поисковых систем
  - Яндекс.Вебмастер
  - Google.Вебмастер
- Веб-аналитика
  - Задачи веб-аналитики
  - Возможности веб-аналитики
  - Виды инструментов веб-аналитики
    - Внутренние системы
    - Внешние системы
  - Погрешность веб-аналитики
  - Курс обучения работы с Яндекс.Метрика
    - Задачи и цели
    - Установка счетчика
    - Описание функционала
    - Настройка целей
    - Статистика по трафику
    - Источники трафика
    - Поисквые фразы
    - Поисквые системы
    - Внешние ссылки

- Содержание сайта
- Профиль пользователя
- Конструктор отчетов
- Поведенческие метрики
- Карта кликов
- Карта путей
- Тепловые карты
- Применение Яндекс.Метрики для SEO-специалиста
- Замеры, эксперименты

■ Юзабилити

- Основы юзабилити
- Юзабилити-аудит
- Взаимодействие с профессиональным юзабилитистом

**Хронология развития поисковых систем от каталогов до нашего времени**

- Google
- Rambler
- Yandex
- Nigma

**Проводимые исследования**

- Исследования Яндекса
- Исследования Google
- Исследования зарубежных компаний
- Исследования российских компаний
- Исследования частных специалистов
- Исследования «ТопЭксперт.РФ»

**Условия участия:**

- Стоимость участия: 25 000 руб. при оплате до 15 декабря.
- Стоимость участия: 30 000 руб. при оплате до 1 января.
- Стоимость участия: 35 000 руб. при оплате до 15 января.

- Стоимость участия: 40 000 руб. при оплате на семинаре.
- При оплате по безналу стоимость возрастает на 20%.

**VIP- участие:**

- Во время курса вы сидите на первом ряду.
- По окончании вы получаете видеозапись.
- Вы получаете подробные комментарии по выполненным домашним заданиям.
- На практических занятиях мы разбираем материал на примере вашего сайта, в частности SEO-аудит и семантическое ядро.
- В течение курса у вас будет право на 4 индивидуальные консультации.
- Стоимость: + 30% от стоимости курса на момент оплаты.

## DVD-семинары и курсы

---

### **DVD-семинар – #1 «Продвижение порталов и крупных интернет-магазинов» – 8 часов (Уровень подготовки: профессионал)**



#### **Для кого этот DVD?**

- Для оптимизаторов, продвигающих самые крупные и сложные проекты рунета
- Для директоров по маркетингу, планирующих создавать внутренний отдел SEO
- Для руководителей крупнейших порталов и интернет-магазинов

#### **Семинар веди:**

- Леонид Гроховский, директор по технологии продвижения и автоматизации, «Оптимизм.ру»
- Олег Пролубщиков, технический директор, «РДВ-Медиа» (Rabota.ru; Ucheba.ru; VashDosug.ru)
- Алексей Чекушин, аналитик, Wikimart.ru
- Роман Прохоров, руководитель службы продвижения, Wikimart.ru

Среди участников семинара были представители компаний:

- mail.ru
- pulscen.ru
- raise.ru
- turinfo.ru
- championat.ru
- books.ru
- labirint.ru
- absent.ru
- terminal.ru
- fastlaneventures.ru

- etorg.ru
- wildberries.ru
- kika.ru
- proskater.ru
- molotok.ru

#### **Подробнее о семинаре:**

- Докладчики семинара на передаче «ТопЭксперт»
- Краткая программа семинара
- Обсуждение семинара
- Репортаж с семинара от searchengines.ru
- Интервью для seonews.ru
  - Олег Пролубщиков, технический директор «РДВ-Медиа»
  - Алексей Чекушин, аналитик, Wikimart

- Роман Прохоров, руководитель службы продвижения, Wikimart

Краткое содержание семинара + отзывы слушателей можно посмотреть на ролике.

Отзыв от Александра Рыжова, wildberries.ru

#### + Бонус:

- Индивидуальная консультация

#### Покупка:

- Стоимость DVD-бокса 30 000 рублей.
- Доставка в руки через EMS в течение 3 рабочих дней.
- Тираж ограничен, всего 20 коробок.
- В целях профилактики пиратства и сохранения эксклюзивности информации, DVD продаю только реальным людям, представляющим реальные компании. В сомнительных случаях запрашиваются рекомендации. Если вы не связаны ни с одним крупным проектом – вы не сможете приобрести этот семинар. Оставляю за собой право отказать любому в возможности покупки без объяснения причины.
- По вопросам покупки обращаться на [grohovskiy@yandex.ru](mailto:grohovskiy@yandex.ru)

### DVD – семинар #2 «Продвижение сайтов: работа с внешними факторами» – 4 часа (Уровень подготовки: специалист)



#### Для кого этот DVD?

- Для оптимизатора, который хочет эффективно продвигать сайты
- Для интернет-маркетолога
- Для руководителя проекта, коммуницирующего с оптимизатором

#### Вам нужен этот видеокурс, если

- Ссылочная масса растет, а позиции сайта не меняются
- Вы не понимаете, как можно покупать естественные ссылки
- Агрегаторы не всегда работают эффективно
- Вы хотите сократить ссылочный бюджет в несколько раз, сняв неэффективные ссылки
- Вам хочется получить трафик с пресс-релизов, блогов и статей

#### Автор и ведущий:

- Леонид Гроховский, директор по технологии продвижения и автоматизации, «Оптимизм.ру»

#### Программа семинара:

- Все источники привлечения внешних ссылок, оценка эффективности



и экономическая целесообразность.  
Сравнение между собой

- Обычные покупные ссылки
- Ссылки в виде контекстных объявлений
- Статьи
- Пресс-релизы и посты в блогах
- Социальные сети
- Форумы
- Естественные ссылки
- Возможности автоматизации по каждому источнику
- Агрегаторы, взгляд со стороны, оценка эффективности использования
  - WebEffector
  - SeoPult
  - Rookee
  - Megaindex
- Продвижение молодых сайтов и сайты с историей: отличия в работе со ссылками
- Если по ссылке кликают – она работает лучше. Миф или реальность?
  - Исследования Блогуна
  - Исследования других бирж
- Критерии выбора ссылок: на что на самом деле стоит обращать внимание
- Оценка эффективности ссылок через Яндекс.Метрика и Google.Analytics
- По каким критериям отбирать площадки
  - Выбор площадок вчера и сегодня, что изменилось
  - ТИЦ, тематичность, количество внешних – что в действительности важно сейчас
- Типичные ошибки, допускаемые при покупке ссылок

– Краткое содержание семинара + отзывы слушателей можно посмотреть на ролике

#### + Бонус:

- Семинар «Продвижение молодых сайтов»

#### Покупка:

- **Стоимость DVD-бокса 6000 рублей.**
- Доставка в руки через EMS в течение 3 рабочих дней.
- **По вопросам покупки обращаться на [grohovskiy@yandex.ru](mailto:grohovskiy@yandex.ru)**

**DVD – семинар #3**  
**«Продвижение сайтов: задействуем внутренние факторы» – 5 часов**  
**(Уровень подготовки: специалист)**



**Для кого этот DVD?**

- Для руководителя SEO-отдела или SEO-аналитика, создающего технологию продвижения
- Для штатного SEO-специалиста или фрилансера
- Для веб-разработчика или программиста, занимающихся созданием сайтов для SEO
- Для продвинутых клиентов, которые хотят понять потенциал своего сайта с точки зрения внутренней оптимизации
- Для специалиста, продвигающего интернет-магазины и заинтересованного в автоматизации перелинковки и генерации контента

**Вам нужен этот видеокурс, если**

- Вы продвигаете интернет-магазин
- Вы столкнулись с проблемой воровства контента и не знаете, что делать
- У вас на сайте нет уникального контента, и это вызывает сложности
- Вы хотите сэкономить на ссылках, используя внутренние ресурсы сайта
- Вы хотите делать эффективные сайты для SEO

- Вы хотите автоматизировать перелинковку и генерацию контента на сайте

**Автор и ведущий:**

- Леонид Гроховский, директор по технологии продвижения и автоматизации, «Оптимизм.ру»

**Подробная программа семинара по ссылке**

**Отзывы**

- **Владимир Баязитов**  
**(www.manuolog.ru):**

– Это был самый лучший семинар по внутренним факторам из всех, которые я посещал.

- **Андрей Трапезников (много проектов):**

– Семинар понравился, немного не хватало примеров конкретных сайтов.

- **Ольга Дергачева (www.r52.ru):**

– Хорошая атмосфера, понятный язык. Отличный семинар и знания преподавателя.

- **Андрей Зарецкий (www.primelog.ru):**

– Очень познавательно, открыл много нового для себя.

- **Сергей Назаренко**  
**(www.aquacoop.ru):**

– Весь материал очень четко изложен, спокойная уверенная манера ведения семинара положительно сказалась на усвояемости материала.

- **Валентин Домбровский, Psycho**  
**(www.nextup.ru)**

– Достаточно полный и интересный семинар от человека, который, безусловно, очень хорошо разбирается в теме. Полученные знания по столь важному аспекту SEO, конечно, необходимо применять на практике.

**Покупка:**

- **Стоимость DVD-бокса 6000 рублей.**
- Доставка в руки через EMS в течение 3 рабочих дней.

- По вопросам покупки обращаться на [grohovskiy@yandex.ru](mailto:grohovskiy@yandex.ru)

**DVD-семинар – #4**  
**«Поведенческие факторы, Юзабилити и Яндекс.Метрика для SEO-специалиста»**  
(Уровень подготовки: специалист)



**Для кого этот DVD?**

- Для оптимизатора, который хочет эффективно продвигать сайты
- Для интернет-маркетолога
- Для веб-аналитика

**Вам нужен этот видеокурс, если**

- Стандартные методы продвижения не работают
- Вы все еще не понимаете, как можно влиять на поведенческие факторы
- Вам нужно, чтобы сайт клиента продавал
- Вы не умеете пользоваться Яндекс.Метрикой

**Ведущий семинара:**

- Леонид Гроховский, директор по технологии продвижения и автоматизации, «Оптимизм.ру»

**Программа:**

- Теория: развитие поведенческих метрик
- 3 группы поведенческих факторов
  - переходы по ссылкам

- результаты поиска
- поведение на сайте
- Где покупать естественные ссылки
- Связки поведенческих факторов: как это работает
- Исследование пользователей: какими данными располагает Яндекс
- 3 способа накрутки поведенческих факторов
  - имитация живых пользователей: создание «живых» ботов
  - ботнет
  - биржа
- Мифы и заблуждения о поведенческих факторах
- Основы юзабилити
- Чек-лист: типичные ошибки юзабилити
- Юзабилити-аудит сайта
- Отличия ПФ для 4 тематик
  - свадебные платья
  - ноутбуки
  - такси
  - индийская косметика
- Возможности Яндекс.Метрики: демонстрация функционала
- Краткое содержание семинара + отзывы слушателей можно посмотреть на ролике

**Покупка:**

- **Стоимость DVD-курса 6000 рублей.**
- Доставка в руки через EMS в течение 3 рабочих дней.
- По вопросам покупки обращаться на [grohovskiy@yandex.ru](mailto:grohovskiy@yandex.ru)

**DVD – семинар #5**  
**«Комплексная стратегия продвижения» –**  
**3 часа [Уровень подготовки:**  
**начинающий]**



**Для кого этот DVD?**

- Для начинающих оптимизаторов
- Для интернет-маркетологов
- Для заказчиков

**Вам нужен этот видеокурс, если**

- Вы хотите использовать комплексный подход к продвижению сайта
- Вы самостоятельно продвигаете сайт и вам нужны выдающиеся результаты
- Вам кажется, что ваши знания устарели
- Вы хотите понять, использует ли ваш оптимизатор комплексный подход к продвижению сайта

**Ведущие семинара:**

- Леонид Гроховский, директор по технологии продвижения и автоматизации, «Оптимизм.ру»
- Станислав Поломарь, руководитель службы продвижения, Web-it
- Виктор Нагайцев, генеральный директор, PerfectSeo

**Среди слушателей были**

- Магомед Чербижев, руководитель Seointellect.ru
- Вениамин Моисеев, преподаватель AiP

- Илья Савинов, преподаватель AiP
- Денис Макаров, технический директор, Miralab.ru

**Программа семинара:**

- Что такое комплексный подход к продвижению сайтов
- Как совмещать внешние и внутренние факторы
- Как поведенческие факторы коррелируют с внутренними
- Как использовать <HTML> разметку для пользователей и для роботов
- Как составлять текст ссылки
- Как влияет возраст сайта на продвижение и что с этим делать
- Как правильно настраивать редиректы и 404 ошибки
- Как картинки влияют на продвижение
- Как избежать склейки нескольких сайтов одной компании

**Покупка:**

- **Стоимость DVD-бокса 3000 рублей.**
- Доставка в руки через EMS в течение 3 рабочих дней.
- **По вопросам покупки обращаться на [grohovskiy@yandex.ru](mailto:grohovskiy@yandex.ru)**

**DVD-семинар – #1**  
**«Автоматизация в SEO» – 8 DVD**  
**(Уровень подготовки: профессионал)**



**Для кого этот DVD?**

- Для руководителя SEO-компании
- Для директора по маркетингу, руководящего внутренним отделом SEO
- Для разработчика веб-сервисов и программ для интернет-маркетинга

**Ведущий семинара:**

- Леонид Гроховский, директор по технологии продвижения и автоматизации, «Оптимизм.ру»

**Содержание**

**1-й диск**

- Задачи и цели автоматизации
- Автоматизация продаж

**2-й диск**

- Расчет стоимости
- Отчетность, статистика, биллинг

**3-й диск**

- Система контроля рисков

**4-й диск**

- Контроль работы оптимизаторов
- Составление семантического ядра
- Копирайтинг

**5-й диск**

- Генерация контента
- SEO-аудит, диагностика сайта

**6-й диск**

- Перелинковка
- Анализ конкурентов

**7-й диск**

- Система для отдела SEO-аналитики
- Яндекс.Метрика, Google.Analytics
- Поведенческие факторы

**8-й диск**

- Агрегаторы: внутренние и внешние ссылки
- Комплексная автоматизация
- Краткое содержание семинара + отзывы слушателей можно посмотреть на ролике

**+ Бонус на 40000 руб.:**

- 2 часа индивидуальных консультаций (экономия 10 000 руб.)
- аудит автоматизации SEO-отдела (экономия 30 000 руб.)

**Отзывы:**

- **Роман Клевцов, генеральный директор «Оптимизм.ру»:**  
– В условиях динамично меняющегося SEO Леонид один из немногих умеет профессионально реагировать на нововведения и повышать качество автоматизации.
- **Тимофей Квачев, руководитель отдела интернет-маркетинга, TRINET:**  
– Хороший контент от хорошего специалиста. Полагаю, что топ-менеджеры компаний смогут почерпнуть много полезной информации, а также ряд методик для быстрого внедрения в практическую деятельность своей компании, оптимизации рутинных процессов и развития бизнеса в целом.

- **Иван Севостьянов, генеральный директор, «ВебПроекты»:**  
– Леонид Гроховский имеет огромный опыт в автоматизации процессов SEO-компаний. Уверен, что его знания вам пригодятся!
- **Виктор Нагайцев, генеральный директор, PerfectSeo:**  
– Леонид является признанным гранд-мастером в области автоматизации продвижения сайтов. Я глубоко убежден, что в данном курсе каждый сможет почерпнуть что-то новое.
- **Станислав Поломарь, руководитель SEO-направления, Web-IT:**  
– Все вещи, о которых рассказывает Леонид, опробованы и отточены на многих проектах (от мелких до очень крупных). Поэтому подходы и знания, полученные от данного семинара, безусловно, представляют большую практическую ценность.
- **Кирилл Самойлов, генеральный директор, ArrowMedia:**  
– Современный бизнес в области SEO невозможен без высокого уровня автоматизации. Знания и опыт Леонида по данному вопросу будут очень полезны.
- **Артем Бородатюк, генеральный директор NetPeak:**  
– Леонид Гроховский подготовил, пожалуй, самый полезный материал, который доводилось просматривать с топ-менеджментом Netpeak за последнее время.
- **Александр Салтыков, директор по маркетингу, Ingate Development:**  
– Невозможно представить сегодня SEO без автоматизации. Леонид обладает большим и, безусловно, полезным опытом в разработке автоматизированных систем. Его консультации будут весьма полезны тем, кто заинтересован в повышении эффективности бизнес-процессов SEO-компаний.

#### Покупка:

- **Стоимость DVD-курса 50 000 рублей.**
- Доставка в руки через EMS в течение 3 рабочих дней.
- Тираж ограничен – всего 10 коробок.
- **В целях профилактики пиратства и сохранения эксклюзивности информации, DVD продаю только тем, с кем знаком лично, либо запрашиваю рекомендации. Оставляю за собой право отказать любому в возможности покупки без объяснения причины.**
- **По вопросам покупки обращаться на [grohovskiy@yandex.ru](mailto:grohovskiy@yandex.ru)**

#### НАПИСАНИЕ на ЗАКАЗ:

1. Дипломы, курсовые, рефераты...
2. Диссертации и научные работы.

Тематика любая: ИСТОРИЯ, экономика, техника, право, менеджмент, финансы, биология...

#### УЧЕБНИКИ, ДИПЛОМЫ, ДИССЕРТАЦИИ:

полные тексты в электронной библиотеке

[www.учебники.информ2000.рф](http://www.учебники.информ2000.рф).

[Вернуться в библиотеку учебников](#)

[Создание и продвижение сайтов](#)

**Уникальная подборка информации по эффективности корпоративных сайтов:**

- для самообразования топ-менеджеров;
- для повышения квалификации преподавателей;
- для рефератов и контрольных.

[Уникальная информация по менеджменту и экономике](#)

[Рейтинг дипломных и курсовых работ](#)

[Научу дистанционно создавать эффективные сайты.](#)

Разработка и продвижение сайтов  
<http://сайт.информ2000.рф/index.htm>

Вернуться в каталог учебников  
<http://учебники.информ2000.рф/uchebniki.shtml>