

Глава 2

Ранжирование в деталях

Ранжирование — это процесс выстраивания веб-страниц в порядке убывания от наиболее соответствующих запросу (релевантных) к наименее соответствующим. Для установления степени релевантности сайта в лабораториях поисковых систем разрабатывают алгоритмы, которые учитывают, что должно влиять на позицию при определенных запросах и условиях. Для предоставления актуальных и полезных посетителю сведений механизмы ранжирования постоянно совершенствуются: добавляются новые факторы, усложняются механизмы обработки информации. В данной главе речь пойдет как о релевантности и классификации запросов, схемах ранжирования, так и о других моментах, оказывающих прямое влияние на позиции сайта в поисковых системах. Предлагаем рассмотреть эту сложную и интересную тему на примере поисковой системы Яндекса. Во-первых, ее алгоритмы подвергаются существенным изменениям чаще, чем алгоритмы других поисковых систем. Во-вторых, [Яндекс](http://yandex.ru/)¹ — первая в рунете система, работающая с коммерческими, или продающими, запросами.

¹ <http://yandex.ru/>

Факторы ранжирования

Факторы ранжирования — это характеристики алгоритма поисковой системы, по которым оценивается страница сайта и определяется ее позиция в выдаче. Яндекс использует в своих алгоритмах несколько тысяч факторов и их сочетаний. Поисковая система делит их на три большие группы:

- 1) **статические факторы**, которые связаны с самой страницей (например, количество ссылок на данный документ в интернете);
- 2) **динамические факторы**, связанные одновременно с запросом и страницей (например, присутствие в документе слов запроса, их количество и расположение);
- 3) **запросные факторы**, учитывающие признаки поискового запроса (например, геозависимость).

Если поисковая система классифицирует факторы по принадлежности к запросу или странице, то SEO-специалисты — по их влиянию на поисковую выдачу, а именно:

- факторы, оценивающие техническую составляющую сайта, которая в первую очередь влияет на возможность и качество индексации;
- факторы, оценивающие текстовую составляющую интернет-ресурсов, которая показывает релевантность содержимого запросу;
- факторы, оценивающие ссылочную составляющую как страницы, так и всего сайта (внешнюю и внутреннюю, анкорную и трастовую);
- факторы, оценивающие поведенческую составляющую (нравится ли сайт посетителям, удобен ли он, отвечает ли их требованиям);
- факторы дополнительного назначения (соответствие сайта региону, наличие аффилиатов и т. д.).

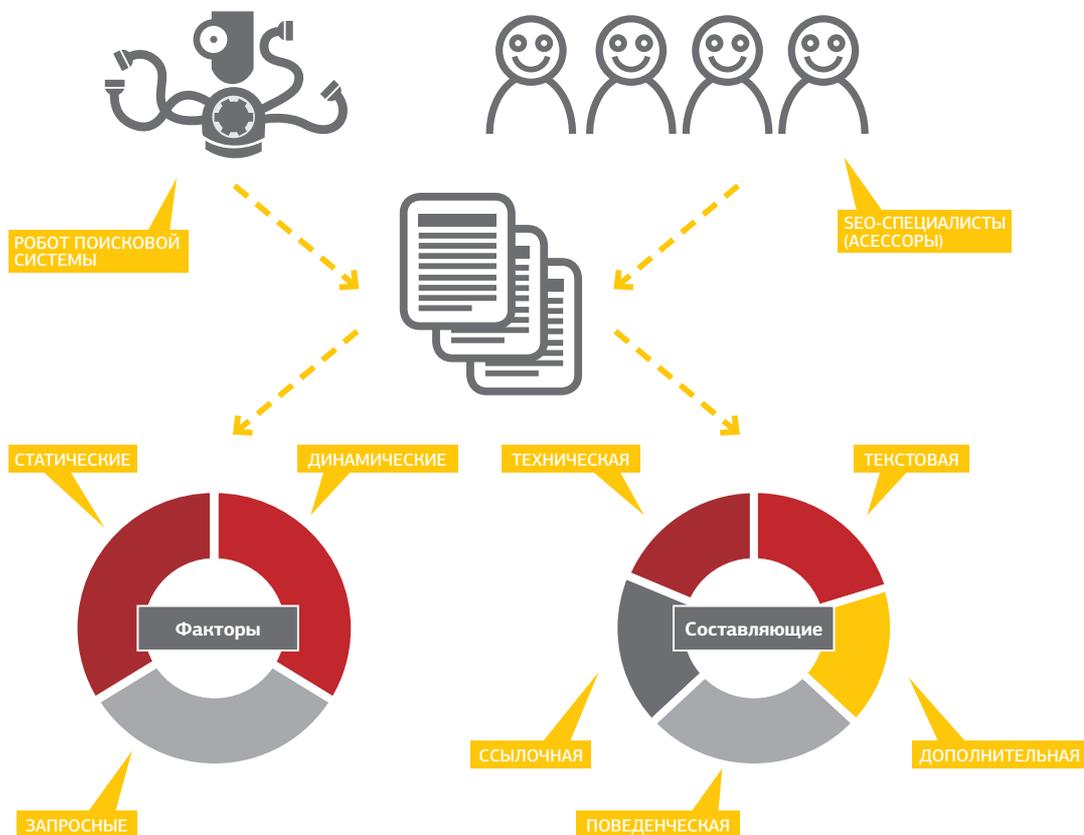


Рис. 2.1. Факторы и составляющие ранжирования

Технические характеристики сайта

Технические характеристики сайта — это визитная карточка ресурса для поисковой системы. Важнейшими техническими характеристиками для продвижения сайта являются его доступность, корректность настроек, скорость ответа сервера, корректность кодов ответа сервера, настройка зеркал и др. Технические проблемы влекут за собой трудности с индексацией, что, в свою очередь, сильно влияет на продвижение: неиндексированные страницы не принимают участия в ранжировании. По этой причине диагностика работоспособности сайта и технических неполадок является важным моментом в поисковой оптимизации ресурса.

Большинство проблем, возникающих при работе сайта, можно разделить на пять категорий:

1. Сайт недоступен. Сайт должен работать 24 часа в сутки 7 дней в неделю и быть постоянно доступным как для пользователя, так и для поисковой системы. Если на сайт в момент даже кратковременного сбоя попытается зайти поисковый робот и не сможет этого сделать, ресурс не будет проиндексирован. Это может вызвать падение позиций сайта по ключевым словам или исключение всех его страниц из индекса.

Если же во время сбоя на сайт зайдет не робот, а пользователь, это может быть еще более критично. Крупным компаниям даже 1 минута неработоспособности ресурса может нанести ущерб в миллионы рублей.

Если сайт регулярно «падает» и возникают серьезные технические неполадки в его работе, стоит задуматься о смене хостинга или аренде выделенного сервера.

2. Медленный или некорректно настроенный сайт. Скорость ответа сервера, на котором находится сайт, не менее важна, чем стабильность его работы. В последние несколько лет поисковые системы стали уделять этому показателю гораздо больше внимания, чем прежде. Яндекс индексирует сайты, не отвечающие его представлениям о скорости, в десятки раз медленнее, чем аналогичные сайты с хорошими показателями. Google за «медлительность» не поместит ваш сайт на высокие позиции в выдаче.

Снижение скорости работы может быть вызвано перегрузкой сервера, который не рассчитан на приведенное количество трафика, неоптимизированным кодом страницы, неработающими скриптами, подгружающимися изображениями или flash-объектами большого размера. Кроме низкой скорости на индексацию и ранжирование могут влиять некорректно настроенные правила противостояния сетевым атакам. В таком случае запросы роботов поисковых систем воспринимаются как нападение — сервер блокирует их активность, и индексация не завершается.

Среди причин повышенного внимания к скорости загрузки можно указать следующие:

1. Медленно работающий ресурс автоматически подпадает под определение «не для всех», что означает неспособность обслуживать большое число пользователей.
2. Любые изменения или новые страницы на таких сайтах могут индексироваться месяцами вместо обычных дней или недель.

Выдержка из рекомендаций Яндекса: «Для индексирования важен размер документа — документы больше 10 Мбайт не индексируются».

Особенно часто страницы 404 появляются на сайтах интернет-магазинов. Когда товар снимают с продажи, страницу с его описанием удаляют. В этом случае поисковой системе необходимо передать информацию о том, что такой страницы на сайте больше нет и ее следует исключить из индекса.

Как проверить код ответа сервера в Яндекс.Вебмастере:

1. Авторизоваться в сервисе Яндекс.Вебмастер (рис. 2.2).
2. Перейти на вкладку «Мои сайты». Если сайтов несколько, выбрать нужный.
3. Перейти на вкладку «Проверить ответ сервера».
4. Указать URL сайта в специальной строке.
5. Нажать кнопку «Проверить».

Если новые страницы продвигаемого сайта долго не появляются в выдаче или часто исчезают из нее, а большие нагрузки приводят к появлению ошибок, рекомендуется провести полную диагностику кода сайта и подгружаемых элементов. При необходимости следует сменить хостинг, оптимизировать код, картинки, скрипты и т. д.

3. Коды ответа страниц сайта. Страницы сайта должны не только быстро открываться, но и корректно отвечать поисковым системам. Код ответа рабочей страницы должен быть 200, а код несуществующей страницы — 404. Статус-коды делятся на 3 группы: перенаправление, ошибка сервера и ошибка клиентского программного обеспечения. Данная информация является служебной и не показывается посетителю непосредственно на странице сайта. Чтобы узнать, как сайт отвечает Яндексу при запросе той или иной страницы, можно воспользоваться специальным сервисом в панели Яндекс.Вебмастера.

Для этого требуются регистрация и подтвержденные права на сайт в сервисе [Яндекс.Вебмастер](https://webmaster.yandex.ru/)¹. Однако для диагностики сайта можно использовать и сторонние сервисы, например [Rooletka.ru](http://www.rooletka.ru)².

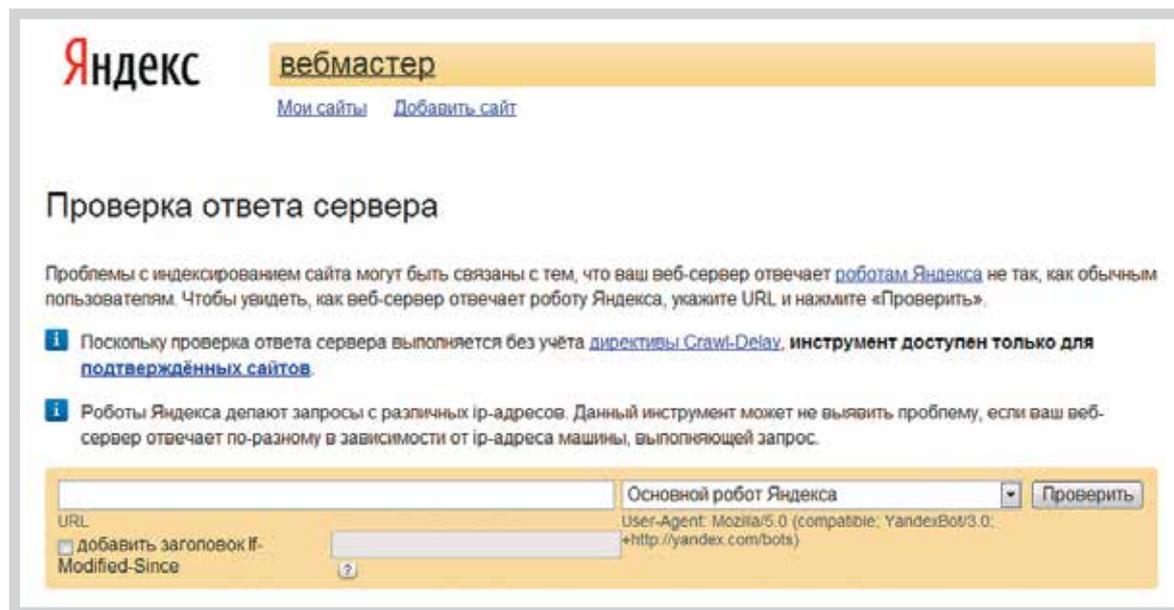
Некорректный статус страницы может привести к появлению дублей, неверному распределению ссылочного веса, неправильной склейке зеркал. Следует иметь это в виду при изменениях на сайте и своевременно настраивать корректные статусы страниц. Страницы могут становиться «несуществующими» по различным причинам:

- из-за опечаток в адресах сайтов при наборе;
- при переходах из закладок браузеров на страницы, которых больше нет;
- при переходах с внешних сайтов по некорректным или устаревшим ссылкам;
- при полной смене структуры сайта.

¹ <http://webmaster.yandex.ru/server-response.xml>

² <http://www.rooletka.ru>

Статус-коды будут подробно рассмотрены в главе 5.



Сразу настроив правила на сайте и предусмотрев перечисленные ситуации, можно оптимизировать затраты на исправление технических неполадок. Периодическая проверка сайта на битые ссылки (ссылки на несуществующие страницы) позволит оперативно их находить и эффективно работать с весом сайта.

4. Настройка зеркал. Каждый ресурс изначально доступен по двум разным адресам: с `www` и без `www`. Поисковая система перед индексацией воспринимает эти зеркала как два разных сайта, и для достижения высоких результатов продвижения необходимо провести так называемую склейку их адресов, или настройку [зеркал](#)¹. Склейка — это процесс соединения двух копий сайта, после которого устанавливается переадресация пользователя с одного адреса на другой (например, с ресурса, адрес которого включает `www`, пользователи будут перена-

Рис. 2.2. Внешний вид панели Яндекс.Вебмастера

¹ <http://www.rookee.ru/post/2011/12/23/zhizn-sajta-posle-audita.aspx>

правлены на ресурс, адрес которого не содержит www, или наоборот).

Указать поисковой системе, какое зеркало считать главным, следует в файле robots.txt при помощи директивы Host. Далее надо настроить переадресацию, используя 301 редирект. Данный код сообщает о перемещении страницы и указывает ее новый адрес. Использование 301 редиректа является надежным способом, но в случае отсутствия доступа к настройкам сервера альтернативой ему является настройка главного зеркала в панели вебмастера Яндекса и Google.

5. Другие технические характеристики. В качестве дополнительных рекомендаций технической оптимизации можно указать настройку robots.txt и sitemap.xml, рекомендации по работе с flash и frame, обоснование формирования ЧПУ («человекопонятных урлов»), устранение идентификаторов сессий и динамических URL и др. Выполнение всех рассмотренных рекомендаций требуется в первую очередь для того, чтобы обеспечить полноценное индексирование ресурса.

Текстовая составляющая

В механизме ранжирования одной из самых важных составляющих является оценка релевантности содержания страницы запросу, который ввел пользователь. Для составления индексных баз и определения релевантности поисковые системы используют контент, текст запроса и некоторые элементы HTML-разметки. Поэтому особое внимание при продвижении ресурса следует уделить текстовой составляющей веб-страницы.

Текст должен быть интересен и полезен посетителю, при этом количество вхождений ключевых слов и их расстановка должны соответствовать требованиям поисковой системы. Оптимизация касается всего содержимого страницы: основного текста, заголовков текста и страницы, меню, подписей к картинкам, метатегов и т. д. (рис. 2.3).



Изменения, связанные с контентом, вступают в силу быстрее других факторов, влияющих на ранжирование: иногда сразу же после переиндексации, иногда через 2–3 недели — из-за пересчета и снятия фильтров. Поскольку содержимое документа и HTML-разметка влияют на релевантность, а результат от изменений контента становится заметен быстро, существует соблазн добавить больше ключевых слов, выделить значимыми тегами продвигаемые слова или написать внушительный объем

Рис. 2.3. Пример оптимизации содержимого страницы

текста. Но поисковая система считает такую «псевдооптимизацию» недопустимым инструментом развития сайта и применяет санкции, то есть фильтры (подробнее об этом читайте далее), которые вызывают снижение позиций. Фильтры снимаются, когда переоптимизированные тексты будут исправлены и произойдет пересчет соответствующих параметров сайта.

Ссылочная составляющая

Сайт-акцептор (принимающая сторона) — это ресурс, на который ведет ссылка. Страницу с размещенной на ней ссылкой, ведущей на сайт-акцептор, называют **страницей-донором**.

Наличие большого количества ссылок, ведущих на определенный сайт с других ресурсов, увеличивает шансы попадания на него посетителя. Поисковые системы учитывают переходы пользователей для выделения одного документа среди множества, то есть используют принцип цитируемости.

Индекс цитирования (ИЦ) — показатель, который говорит о значимости данной страницы, вычисляется путем учета ссылающихся на нее ресурсов. Данная характеристика лишь учитывает количество ссылок на ресурс, но не отражает их структуру в каждой тематике. Ссылки с малой и большой значимостью могут иметь одинаковый индекс цитируемости, что неправильно. Поэтому был введен термин «фактор популярности» (англ. Popularity Factor), он же «взвешенный индекс цитирования» или «вес ссылки». В разных поисковых системах этот фактор называется по-разному: PageRank — в Google, ВИЦ — в Яндексе, ИЦ — в Апорте. Также существует **тематический индекс цитирования**¹ (ТИЦ), учитывающий тематику сайтов, ссылающихся на ваш ресурс.

Алгоритмы, применяемые современными поисковыми системами для оценки ссылочной составляющей, претерпели большие изменения, но качество и количество ссылок напрямую влияет на выделение страницы из множества копий. Чем авторитетнее

¹ <http://www.rookee.ru/post/2010/11/12/Поднимаем-ТИЦ-актуальные-методы.aspx>

будет страница-донор, схожая с сайтом по тематике, тем больший вес она ему передаст.

Существует множество способов наращивания ссылочной массы: размещение статей, новостей, пресс-релизов, написание постов в блогах, социальных сетях и работа с биржами ссылок. При подборе площадок-доноров необходимы опыт оценки, аналитические способности при сборе параметров и тщательный отбор необходимых сайтов. Эта функция представлена в сервисах автоматизированного продвижения сайтов: система [Rookee](http://www.rookee.ru/)¹ не только помогает подбирать площадки, но и автоматизирует работу с ними, экономя бюджет и время.

Большое значение имеют не только внешние ссылки, но и внутренняя [перелинковка](#)². Если правильно распределить вес всего сайта по целевым (продвигаемым) страницам, то можно существенно сократить затраты на внешние ссылки.

Гезависимость и геонезависимость запроса

Яндекс стремится персонализировать выдачу. Одним из шагов на пути к этому является учет региональности в результатах поиска. Учет региональности состоит из определения следующих параметров:

- гезависимости запросов;
- региона сайта;
- региона пользователя.

Гезависимость запроса

Гезависимость запроса показывает, важен ли пользователю при поиске региональный таргетинг. Пример геонезависимого запроса — «рецепт приготовления пиццы». Вводя гезависи-

Если сайт относится к конкретным регионам, проследите за тем, чтобы на нем было размещено как можно больше информации, интересной пользователям, живущим именно на этой территории. Чем больше будет таких данных, тем выше будут позиции ресурса при поиске по этим регионам. Также обратите внимание: если на страницах сайта указан только адрес, например в Московской области, то ресурс может быть не присвоен региону «Москва», необходимому для продвижения.

¹ <http://www.rookee.ru/>

² <http://prozhector.ru/publications/vypusk-3-ot-03-aprelya/master-klass-vnutrennyaya-perelinkovka-sayta/>

мый запрос «заказ пиццы», пользователь, скорее всего, желает увидеть в результатах выдачи сайты того региона, в котором он находится. Коммерческие запросы, как правило, являются геоинформационными.

У поисковой системы Яндекс существует специальный [сервис проверки геоинформационности запроса](#)¹ (рис. 2.4).

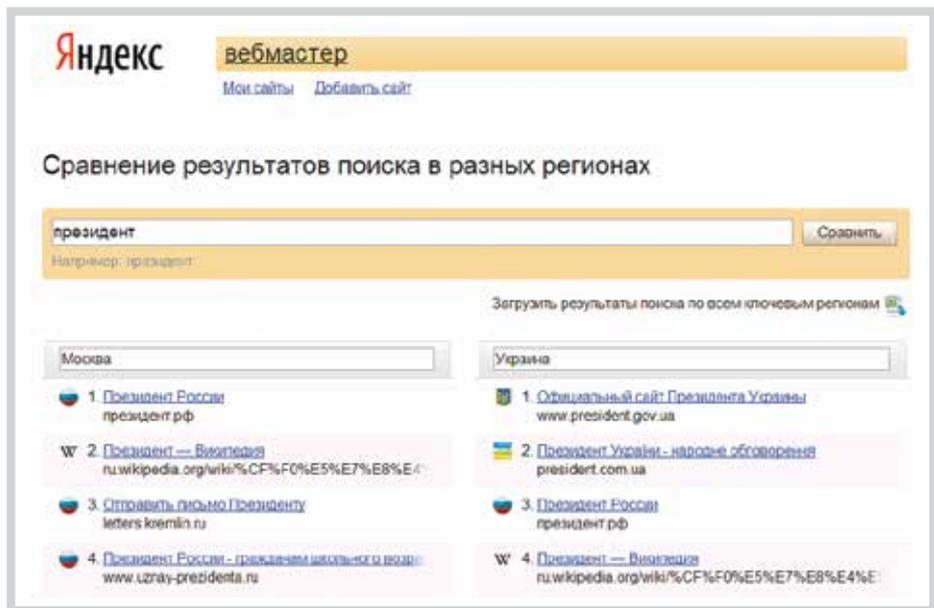


Рис. 2.4. Сравнение результатов поиска в разных регионах

Проверить геоинформационность запроса можно и без данного сервиса. Для этого необходимо:

- проверить, есть ли внизу страницы с результатами выдачи фраза: «При поиске отдано предпочтение сайтам из ...». Если данная фраза присутствует, то запрос геоинформационный, если нет — геоинформационнонезависимый;
- проверить, указан ли рядом с сайтами город. Если указан, то запрос является геоинформационным, если не указан — геоинформационнонезависимый. (К данному способу проверки не относятся запросы с указанием города, например «пластиковые окна Москва», так как они уже содержат регион и не являются геоинформационными.)

¹ http://webmaster.yandex.ru/compare_regions.xml