# Общая часть

Технически сайт http://000000.ru выполнен на хорошем уровне, известные SEO-уязвимости (такие, например, как псевдодубли или возможность дописывания ложных переменных в адресной строке) – закрыты.

Забегая вперед, основной проблемой является попадание сайта в статистическую модель поисковиков «плохой сайт» и SEO-атаки, которым с высокой вероятностью подвергается сайт.

Надо заметить, что в настоящее время Яндекс, Гугл и Бинг, отошли от понятия «формула ранжирования» и не связывают позиции сайта в выдаче с какими-то конкретными факторами. Одним из основных критериев оценки сайта является «статистическая похожесть» ту или иную группу сайтов из огромной базы ранее оцененных сайтов. За десятилетия работы поисковиков статбаза для сравнения и оценки сайтов накопилась более, чем достаточная и она постоянно корректируется.

Одна из проблем заключается в том, что статистически http://000000.ru не соответствует картине «сайта недвижимости», но зато соотвествует картинам «Бесполезного сайта» и «Вредоносного сайта».

Проверка позиций при помощи языка запросом показывает, что сайт находится под очень жесткими санкциями и Яндекса и Гугла.

Простейшие, но показательные примеры.

А) Ни одна из страниц, находящихся в выдаче на верхних позициях не содержит прямых вхождений поискового запроса.

Б) На проиндексированной странице второго уровня http://000000.ru/index.php?do=orders&orders=s\_sale&country=rus&city=moscow имеется фраза с грамматической и пунктуационной ошибками «Продается участок 46 Га, в Москвовской области, с. Ступино, д. Старая Ситня.!», что делает ее абсолютно уникальной и позволяет легко найти в поиске, заключив в кавычки. Однако по запросу в кавычках Яндекс не видит ее на странице, хотя саму страницу видит. А Гугл показывает ее в конце выдачи, хотя в силу ее уникальности, должен показать на 1 позиции.

Более сложные запросы однозначно показывают на жесткую фильтрацию.

Тот факт, что в основном индексе гугла находится менее 1% страниц (494 страницы из 91000) так же косвенно говорит о фильтрации.

К ухудшению положения приводит так же тот факт, что с высокой долей вероятности сайт подвергается перманентным SEO-атакам (в т.ч. и в настоящее время) методом мигающих ссылок.

## Замеченные ошибки

1. В ситуации, когда страница должна отдавать 404 или 410 ошибку на сайте возникает цепочка 30\* редиректов. Так например, при запросе несуществующей страницы http://000000.ru/bla-bla возникает цепочка из трех страниц сразу:
**HTTP/1.1 302 Found
Location:**http://000000.ru/index.php?do=404 **HTTP/1.1 301 Moved Permanently
Location:**http://000000.ru/ **HTTP/1.1 200 OK**Это порождает соответствующую цепочку проблем. Если страницу, дающую код ответа 404 поисковик исключит из индекса, и робот не будет обращаться к ней, даже при условии размещения на нее ссылки, то страницу, отвечающую 302, робот будет обрабатывать каждый раз заново, когда получит ссылку на нее. Обработка цепочки 302-301-200 для одной страницы берет столько же ресурсов поисковика, сколько обработка десятков страниц, отвечающих 200 ОК. Размещая внешние ссылки на такие страницы, отдающие 302 редирект, злоумышленник может серьезно нарушить процесс индексации сайта.
Кроме того, поскольку цепочка ведет на главную страницу, появляется возможность передать на нее ложный вес.
В частности в нашем случае используется именно такая технология. На сайт (и именно на стартовые страницы с 301 редиректом) ссылается большое количество ссылок, содержащие правильные ключевые слова, но в неправильной тематике (конкретный пример: вместо «купить квартиру в Батуми» - «отремонтировать квартиру, бригада маляров из Батуми»), и таких ссылок сотни.
Поскольку данные ссылки, как выяснилось, размещены не владельцем сайта, то очевидно – это атака конкурентов. Это является одной из причин того, что сайт не виден по тематическим запросам, но виден по посторонним.
Поскольку такие ссылки идут через цепочку 301 редиректов, они не отображаются в статистике Яндекса и не были обнаружены.
Список криминальных ссылок прилагается.
2. На сайте свободно, без премодерации можно публиковать любые тексты объявлений, которые тут же появляются в соотвествующих разделах. Последующая модерация и снятие ложных или криминальных объявлений, которая практикуется редакцией сайте - бесполезна, так как профессиональный спамер может заставить робота прочитать страницу в течение нескольких минут после публикации. Чтобы текст спамового или криминального объявления выполнил свою функцию достаточно 10-15 минут.
Несложно заметить, что в индексе и яндекса и гугля есть довльно много криминальных объявлений, которых в реальности на сайте нет.
3. При удалении объявления по оставшемуся адресу отдается 410 код ответа, что совершенно правильно и грамотно. Однако показывается пустая страница – что неправильно. Было бы разумно в такой ситуации показывать главную страницу сайта, сопровождаемую 410 кодом ответа (без каких-либо редиректов). При этом в оглавлениях остаются ссылки на удаленную страницу, что создает дополнительную нагрузку на робота и замедляет индексацию сайта. Так же это дает возможность SEO-атаки на сайт.
Пока что в данное время, как Яндекс так и Гугль обрабатывают 410 ошибку специфически: при наличии на удаленную страницу внешних и внутренних ссылок одновременно, она не удаляется из индекса. Поэтому внутренние ссылки надо автоматически убирать при удалении страницы. Иначе злоумышленник, разместив на такие страницу внешние ссылки, может сильно замедлить или вообще ограничить индексацию сайта.
4. На сайте расположены блоки Google Adsense и Яндекс Директ, показывающие рекламу конкурентов. Наличие их на сайте агентства недвижимости выпадает из статистической модели сайта агентства (очевидно, что честное агентство недвижимости продает не трафик, а квартиры). Конечно само по себе наличие этих блоков не криминал. Однако поскольку Гуглу и Яндексу с определенной вероятностью известны настройки этой контесткной рекламы, отсутствие фильтрации объявлений конкурентов однозначно позволяет им отнести сайт к категории MFA (Made for Adsense), и счесть, что сайт только имитирует агентство недвижимости, и зарабатывает арбиражем трафика, со всеми вытекающими последствиями.
Кроме того, они могут сделать тот же самый вывод исходя их того, что написано дальше.
5. На сайте размещены SEO-статьи, например http://000000.ru/index.php?do=repair\_show&nid=115. Это не такой серьезный криминал, как реклама конкурентов (поисковики допускают, что вебмастера продают ссылки тайком от работодателей и смотрят на это сквозь пальцы), однако в совокупности с другими статистическими данными, он становится серьезным.
6. Статмодель сайта недвижимости подразумевает, причем на 100%, наличие полных контактных данных, включая телефон и адрес. Наличие на странице контактов (а поисковики всегда обрабатывают эту страницу очень детально, по специальным алгоритмам) только e-mail на публичном домене и предложения разместить рекламу на 100% воспринимается как признак ложного сайта недвижимости.
7. С КАЖДОЙ страницы сайта размещено 4-5 ссылок на сайт программы продажи недвижимости ХХХХХХ. Поисковой лицензией Яндекса это прямо запрещено и расценивается как дорвей.
8. Тексты объявлений имеют признаки сгенерированности и статистически недостоверны.
**Пояснение:** В данном случае это означает, что частотный анализ объявлений по сайту не совпадает с частотностью заведомо достоверной выборки, взятой с заведомо трастовых сайтов. Это когда, к примеру, квадратные метры или количество санузлов не соотвествуют заявленным типам квартир в значительной части объявлений. Т.е. если, например, в большинстве «Брежневок» по стране указывается один санузел, а на конкретном сайте 0 или 2, то это признак недостоверности.
Также признаки сгенерированности и недостоверности контента могут отмечаться в том случае, когда база заполняется халтурно, абы как. Например, при заполнении опускается какое-нибудь поле, но на сайте оно выводится пустым или нулевым. И получается квартира на 0 этаже или с 0 санузлов. Данные, в которых присутствует цифра 0, в нашем случае вообще не должны выводится.
9. Данные в оглавлении полностью дублируют данные на карточках объектов (на конечных страницах). Для всех сайтов, а не только для сайтов недвижимости это является признаком MFA и MFS сайтов, так как является самым простым способом увеличения проиндексированных страниц и не несет никакого практического смысла.
Например, в оглавлении:

И при переходе по ссылке «подробнее», ровно то же самое:


# Прочие нестатистические данные

Ниже приводятся данные нестатистического характера, которые могут учитываться поисковиками при наложении санкций. Данные получены от тех же подрядчиков, которые поставляют их Яндексу (по Гуглу таких данных у меня нет). Эти данные МОГУТ учитываться, а могут не учитываться в зависимости от ситуации. С учетом вышеприведенных негативных статистических данных, они так же могли быть также учтены, как данные негативного характера.

1. По данным Whois более 75% доменов принадлежащих администратору 000000.ru регистрируются на год, потом отпускаются. 25% является порогом, говорящим о недобросовестных целях регистрации.
2. Домен XXXXXXXX.ru, принадлежащий владельцу 000000.ru находится в блок-листе спамхауса, как сайт распространяющий вредоносное ПО. Попадание в базу спамхауса происходит в случае если есть основания намеренного распространения вредного ПО.
3. Владельцу домена 000000.ru принадлежит большое число доменов (зарегистрированных на тот же емэйл), запаркованных на SEDO – около полутора тысяч.

## Возможная SEO-атака

При анализе динамики ссылочной массы за последний месяц зарегистрировано очень сильное несоответствие, даваемое разными способами ее оценки. Общее количество «прочитанных» роботами ссылок на сайт http://000000.ru почти на порядок превышает количество реально размещенных ссылок. Это можно видеть даже в таком грубом и инерционном инструменте, как панель вебмастера Яндекса, задав диапазон показа обратных ссылок за две недели.

Это является признаком SEO-атаки мигающими ссылками. Суть ее заключается в том, что размещенная ссылка снимается до того, как она попадет в индекс, сразу после считывания роботом, либо по фиксации факта НПС, либо просто по среднему времени известному для конкретного сайта-донора.

При этом ссылки ведут на страницы, посещение которых поисковым роботом нежелательно, в вашем случае это страницы, отдающие 302 и 410 код ответа. См. выше.

В данном случае – поскольку удалось отловить лишь небольшое количество исчезнувших ссылок, факт атаки является только достаточно достоверным предположением. Если у вашего вебмастера есть доступ к логам посещаемости сайта, то наличие в нем обращений поисковых роботов к страницам, дающим код ответа 302 или большого количества обращений ИЗВНЕ на удаленные страницы, дающие код ответа 410, говорит о ведущейся атаке.

В реальной ситуации обращений от роботов на страницы, дающие 302 вообще быть не должно, ведь эти страницы искусственные, и роботу о них неоткуда узнать, кроме как от криминальной ссылки. На удаленные 410 страницы в статистике должны быть только внутренние переходы с ЕДИНИЧНЫМИ внешними переходами. Переходов с внешних страниц должно быть не более процента от внутренних. Если количество внешних переходов на 410 страницы сопоставимо с внутренними или тем более – превышает их – это однозначно говорит о SEO-атаке.

## Всякие мелочи

То, что собрано в данном разделе не могло повлиять на ситуацию, но в случае работ по ее исправлению, все что ниже – должно быть учтено.

1. Не стоит помещать ссылки в тег ноуиндекс, без использования параметра rel nofollow. В настоящее время этот тег никак не влияет на закрытие ссылок и их индексацию (кроме предотвращения показа в выдаче текста ссылки). В определенных статистических условиях ссылки в nouindex без nofollow могут быть восприняты, как криминал.
2. Для продвижения данного сайта активно используется прогон по доскам объявлений и прочим ресурсам, где можно легко разместить ссылку на свой сайт. Само по себе это не страшно и даже помогает во многих случаях, но в сочетании с вашей статмоделью, – этого делать не стоит.

## Рекомендации по исправлению

В данном случае малоперспективно что-либо делать самим сайтом, не отвязав его от других доменов данного администратора и от самого администратора. Поэтому все исправления крайне желательно делать на новом домене (не обязательно свежезаригстрированном, но на «чистом»), и если важен ТИЦ и пр. характеристики старого домена, с последующей подклейкой старого домена к новому. Если ТИЦ и пр. можно пренебречь, просто забываем про старый домен.

Разумеется, рекомендация по размещению на новом домене – опциональна. Все нижеперечисленные мероприятия по устранению недостатков можно проводить и с текущим сайтом, однако шансы на успех будут существенно ниже.

Если предполагается, что сайт должен зарабатывать именно на продаже объектов недвижимости, то от любого побочного заработка на арбитраже трафика и продаже ссылок следует отказаться.

Если предполагается, что сайт должен зарабатывать именно на арбитраже трафика и ссылках – это не лучшая идея, псевдосайты агентств недвижимости очень легко статистически выявляются и их создание требует специфической и высокой квалификации исполнителя.

**Все, что описано ниже в п 16-25, надо проделать независимо от того, на каком домене будет размещен новый сайт.**

1. В нашем случае, однотипные и связанные проекты одного владельца следует размещать не на разных доменах, а на разных поддоменах одного сайта.
2. Ссылка на программу Риэлтор Про может быть только в нерекламном виде, и только одном экземпляре и только в подвале сайта, например, в виде копирайта или ремарки «Построено на базе Риэлтор Про». Ссылка должна быть index, follow. На период реконструкции сайта ее следует вообще убрать.
3. Для исключения SEO-атак отсутствующие страницы должны выдавать 404 код ответа. Если страница была удалена и выдает 410 код ответа – на нее не должно оставаться внутренних ссылок на сайте.
4. Публикация любых объявлений на сайт и текстов должна быть либо премодерируемой, либо фильтроваться хорошим роботом.
5. Никаких контекстных объявлений на сайте показываться не должно.
6. На сайте должны быть полные контактные данные, включая адрес офиса и обязательно телефон городского кода, который должен соотвествовать указанному региону сайта. Желательна схема проезда, размещенная через сервис карт Яндекса или Гугля (что именно – не важно). E-mail должен быть на домене самого сайта, а не на публичном домене. Желательно несколько e-mail разных специалистов агентства. Наличие не засвеченного в плохих делах телефона в коде 800 – несомненный плюс.
7. Сайт следует добавить в карты Яндекса и Гугля. Обязательно.
8. Если сайт должен приносить доход именно за счет продажи недвижимости, его обязательно надо добавить в каталог Яндекса на платной основе (после устранения всех указанных недостатков).
9. Почистить базу, чтобы в ней оставались только реальные валидные объекты и живые агенства, контактные данные которых были бы валидны и не дублировались бы в поиске в качестве контактных данных других лиц.
Если вдруг на сайте действительно использовался сгенерированный контент, убрать его или по крайности сменить поставщика такого контента.
Если использовались базы данных внешних поставщиков, убрать из показа все нулевые параметры (типа 0 комнат). Если параметр действительно имеет нулевое значение не отображать его или отображать текстом «Санузел отсуствует», вместо «Санузлов 0».
10. Убрать дублирование описание объектов недвижимости в оглавлениях и на карточках объектов. Либо более сжатое описание оглавления, либо более детальная карточка.

**Внимание, пункты 25-27, касаются только стратегии восстановления с новым доменом. Если делаем на старом – их можно опустить.**

1. В случае если делаем новый домен, он не должен содержать никакой привязки по данным whois к другим доменам владельца. Крайне желательна регистрация на юрлицо, не обязательно российское, сойдет и офшор. Так же желательно передать старый сайт на данные нового администратора.
Для информации – при подаче документов на регистрацию от имени офшора российские регистраторы не проверяют валидность документов. Т.е. офшор может быть липовым с нарисованной печатью. Но и потом, случись чего, в суде такой офшор ничего не сможет оспорить, так что лучше – настоящий офшор.
Если регистрация на юрлицо неприемлема, регистрируем на Private Person c новыми, незасвеченными e-mail и телефоном.
2. Если делаем новый домен, с целью подклеить потом старый, то начинаем заполнять его только новыми объектами, с учетом всех нижеприведенных рекомендаций, не перенося пока данные со старого сайта. Новых объектов должно быть не менее 10% от объема старого сайта.
Одновременно чистим старый сайт, удаляя с него все лишние объекты, так, чтобы осталось только необходимое, ОБЯЗАТЕЛЬНО учитывая нижеприведенные рекомендации. После того, как новый сайт проиндексируется и появится в индексе, не ранее, чем через 3 месяца синхронизируем его по содержанию со старым сайтом, предварительно указав в robots.txt директиву
**HOST: новый домен**.
После полной синхронизации сайтов и переклейки Яндексом основного домена на новый сайт поставить со старого домена 301 редирект на новый.
Ссылку на программу Риеэлтор Про, на новом домене не размещать вплоть до полной переклейки.
3. Сайт должен создаваться на новом домене с учетом вышеперечисленных рекомендаций.