

ЯНДЕКС И GOOGLE.

ФАКТОРЫ
РАНЖИРОВАНИЯ

2019

АНАЛИТИЧЕСКИЙ ОТЧЕТ

МИХАИЛ ВОЛОВИЧ,

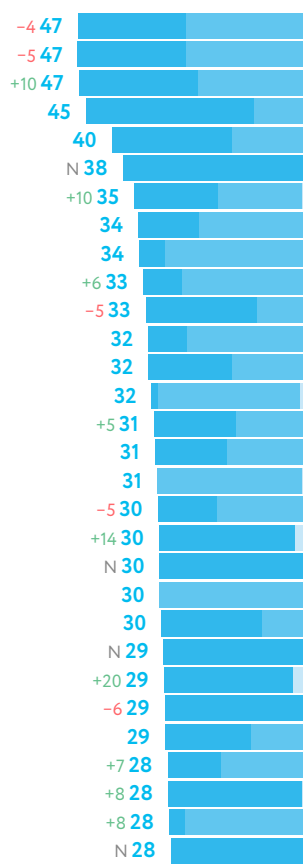
«АШМАНОВ И ПАРТНЕРЫ»



СОДЕРЖАНИЕ

ТОП-75 ПАРАМЕТРОВ ДЛЯ GOOGLE	02	ТЕКСТОВАЯ РЕЛЕВАНТНОСТЬ	31
ТОП-75 ПАРАМЕТРОВ ДЛЯ ЯНДЕКСА	04	Матрица текстовых параметров	32
ВВЕДЕНИЕ	06	Принцип антибуквализма	33
ОТКУДА МЫ ЭТО ЗНАЕМ	08	Где текстовые факторы сильнее?	36
Параметры	09	Слов запроса должно быть много	37
Выборки	11	Длина и концентрация	38
Математическая статистика	13	Витрина как инструмент текстовой оптимизации	38
Корреляции с позицией	13	SEO-тексты и «Баден-Баден»	39
Контраст с фоном	14	«Тургенев», или как не уехать в «Баден-Баден»	39
Статистические метрики, которые мы используем	17	Как SEO-тексты влияют на ранжирование	40
Дополнительные сложности	18	Образцы стиля SEO-текстов	41
Общие и частные факторы	18	Зоны документа	42
Корреляции между параметрами	19	Заголовки текста	42
Инерция ранжирования: преоцененные запросы	19	Метатеги	44
Преодоление инерции: «многорукие бандиты»	20	Домен и URL	44
Рейтинг параметров	21	Концентрация слов запроса по зонам	45
КАК ИСПОЛЬЗОВАТЬ НАШИ РЕЗУЛЬТАТЫ	23	Выделения в тексте, ссылки и списки	46
Не нужно добиваться роста «важных» параметров любой ценой	23	Ссыльно-текстовые факторы	49
Старайтесь понять логику	23	Практические советы	51
Сравнивайте свой сайт с конкурентами	23	РАНЖИРОВАНИЕ САЙТОВ	52
Если параметр для вас неактуален, за его отсутствие вам ничего не будет	24	Размер имеет значение	53
Почему у хорошо ранжируемых сайтов HTML-кода больше	25	Монополизация выдачи	53
Критерии, которые отслеживают поисковики, могут быть формальными	28	ИКС: из чего складывается качество сайта	54
Некоторые параметры могут иметь критическое значение	28	Трафиковые факторы	56
А у нас всё автоматизировано	29	Ссылочные факторы	58
		Ассортимент, размер сайта, присутствие в регионах	59
		Что стоит за сайтом	62
		«Компетентность, авторитетность, надежность»	62
		Контакты и представление компании	63
		Социальные сети	65
		Цены и скидки	66
		Качество сайта	68
		Поведенческие и трафиковые факторы	68
		Требования к функционалу	70
		Описания товаров и услуг	70
		Подбор и сравнение товаров (услуг)	71
		Условия продажи	71
		Технические требования и «продвинутые» сайты	73
		Контент	74
		Практические советы	76

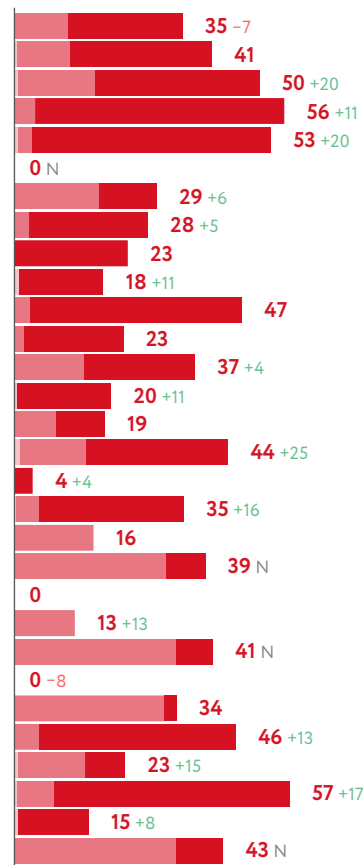
ТОП-75 ПАРАМЕТРОВ ДЛЯ GOOGLE



Google
 корреляция с позицией ■
 связь с попаданием в топ-30 ■
 рост за год ■

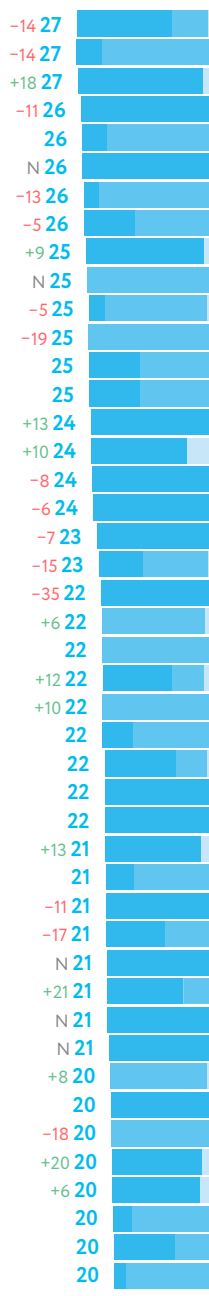
▲ SimilarWeb Country Rank 57-58
 ▲ SimilarWeb Global Rank 57-58
 ▲ Alexa Rank 57, 58
 Суммарный LinkRank ссылок на сайт 59, 54-55
 Ссылающихся на сайт доменов 59, 54-55
 Description: сл. запроса и выдел. Google 44, 33-35
 Телефон 8-800 60-61, 28, 76, 10
 Много адресов 60-61, 55
 Много телефонов 60, 55
 Возраст по упоминаниям на archive.org 08
 Униикальные ссылки на сайт 59, 54-55
 Основная локализация в Санкт-Петербурге
 Несколько фотографий 70
 Более одного адреса на сайте 60
 Постпродажное обслуживание 72, 68
 Видео на сайте 28, 70, 65
 Более одного телефона 60
 Много регионов 60-61, 55
 Description: слова запроса и синонимы 44, 33-35
 Исходящие ссылки: выделенные 46-47, 33-35
 Основной домен
 Рекомендации по выбору на странице 75
 Внутренние ссылки: количество слов 38, 47
 Description: слова запроса 44, 33-35
 Локализация в Москве 14-15, 60, 28
 Ссылающихся на сайт страниц 59, 55
 Рекомендации по выбору 75
 Страниц в индексе Яндекса 60-61, 09, 19, 27, 55
 Возраст домена 08, 53, 76
 Фрагменты: количество слов 32, 27

○ рейтинге 21
 18, 21 — ссылки на страницы отчета.

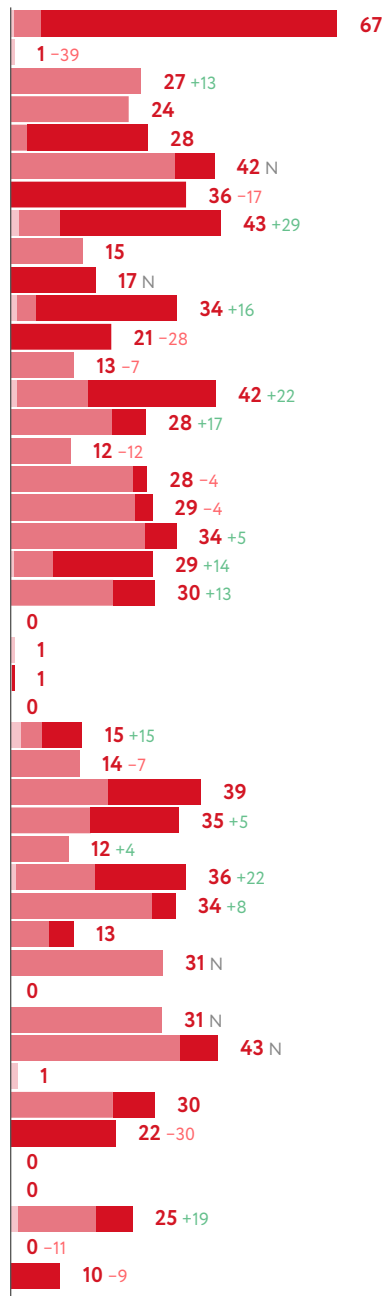


Яндекс
 корреляция с позицией ■
 связь с попаданием в топ-30 ■
 рост за год ■

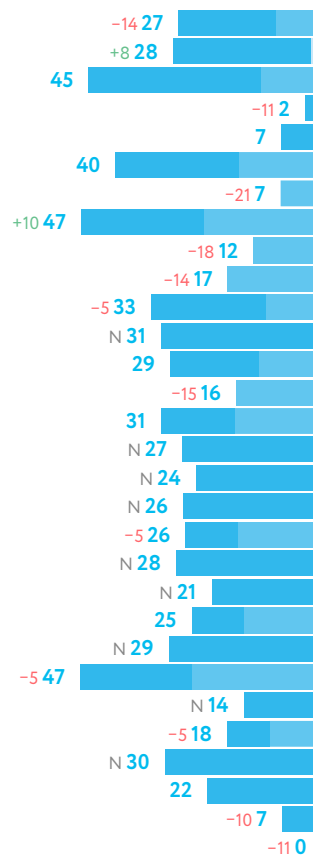
▲ — инвертированный параметр (чем меньше, тем лучше). +8, -28 — изменение за год. N — новый параметр.



- ИКС 54-56, 68, 08
- ТИЦ 54-56
- <title>: выделенные 43, 33-35
- Основная локализация в Москве 60
- Локализация в Санкт-Петербурге 60-61, 55
- Фрагменты: размер текста 32, 27
- Посещения: реклама (display) 70, 55
- Канал на YouTube 65
- Description: неточный запрос + синонимы 44, 33-35
- Исходящие внешние ссылки: количество слов более одного региона 60
- Просмотры страниц за посещение 68-69, 56, 55
- Телефон 8-800 (авто) 10
- Группа в Facebook 65-66
- Внутренние ссылки: выделенные 46-47, 33-35
- Количество знаков препинания в Description 44
- Размер текста страницы 38, 23, 27
- Количество слов в тексте страницы 38, 27
- Текст: слова запроса 33-35
- ▲ SimilarWeb Category Rank 57-58
- Текст: слова запроса и выделенные Google 33-35
- ▲ SEO-текст: плотность стилистических проблем 42
- Информационность страницы
- <title>: неточный запрос + синонимы 43, 33-35
- ▲ SEO-текст: количество стилистических проблем 42
- Сертификаты, лицензии, награды 64, 76
- Длина Description 44
- Размер HTML-кода 25-27, 23, 38, 73
- Найдено документов с сайта в Яндексе 60-61, 27, 55
- Description: слова запроса и выделенные 44, 33-35
- Группы в соцсетях 65-66
- Текст: слова запроса и синонимы 33-35
- Сравнение товаров 71
- Исходящие ссылки: сл. запроса и выдел. Google 33-35
- <title>: слова запроса и синонимы 43, 33-35
- Внутренние ссылки: сл. запроса и выдел. Google 33-35
- Исходящ. ссылки: сл. запроса и выдел. 33-35
- Новости (в интерфейсе) 75
- Текст: выделенные 33-35
- Продолжительность посещения 68-69, 56, 55
- Повторы слов в <title>
- SEO-текст: неточный запрос + синонимы 40, 42, 33-35
- Группы в соцсетях (авто) 65-66
- Длина <title> 43
- Справочные материалы 75



ТОП-75 ПАРАМЕТРОВ ДЛЯ ЯНДЕКСА



Google

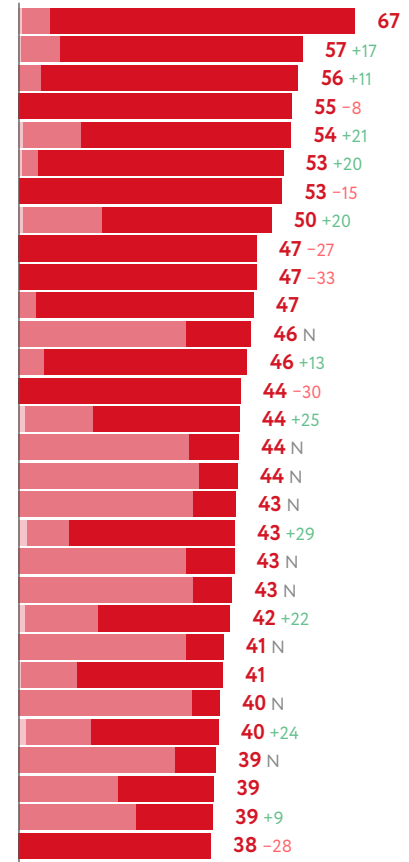
корреляция с позицией ■
связь с попаданием в топ-30 ■
рост за год ■

▲ — инвертированный параметр (чем меньше, тем лучше). +8, -28 — изменение за год. N — новый параметр.

ИКС 54–56, 68, 08
 Страниц в индексе Яндекса 60–61, 09, 19, 27, 55
 Суммарный LinkRank ссылок на сайт 59, 54–55
 Посещения: поиск 58, 54–55
 Огромный ассортимент 54–56, 53
 Ссылающихся на сайт доменов 59, 54–55
 Количество посещений сайта 57–58, 54
 ▲ Alexa Rank 57, 58
 Посещения: прямые переходы 58, 55, 70
 Посещения: почта 58, 55, 70
 Уникальные ссылки на сайт 59, 54–55
 Фрагменты: сл. запроса и выдел. Google 34–35
 Ссылающихся на сайт страниц 59, 55
 Посещения: соцсети 56, 55
 Видео на сайте 28, 70, 65
 Фрагменты: слова запроса и выделенные 34–35
 Фрагменты: слова запроса и синонимы 34–35
 Фрагменты: слова запроса 32, 40, 34–35
 Канал на YouTube 65
 Фрагменты: количество слов 32, 27
 Исходящ. ссылки: сл. запроса и выдел. 33–35
 Группа в Facebook 65–66
 Внутренние ссылки: количество слов 38, 47
 ▲ SimilarWeb Global Rank 57–58
 Атрибуты alt: количество слов 47–48
 Группа в Twitter 65–66
 Исходящие ссылки: выделенные 46–47, 33–35
 Размер HTML-кода 25–27, 23, 38, 73
 Подбор по параметрам 71, 19, 27, 55
 Посещения: ссылки 58, 55

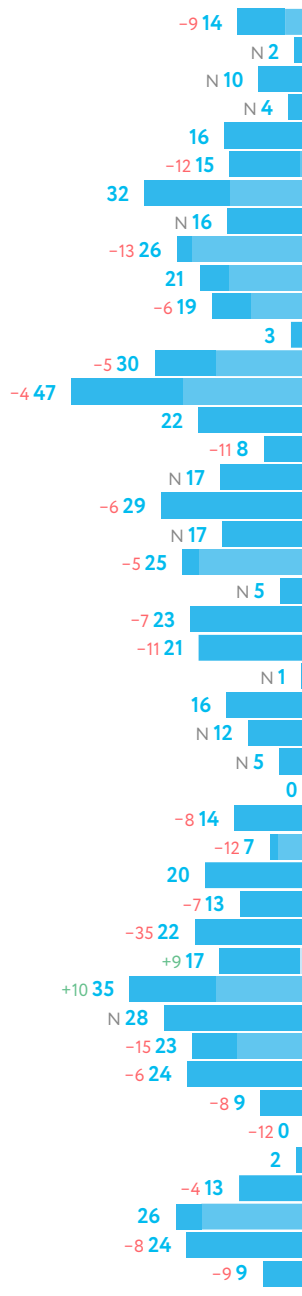
О рейтинге 21

18, 21 — ссылки на страницы отчета.

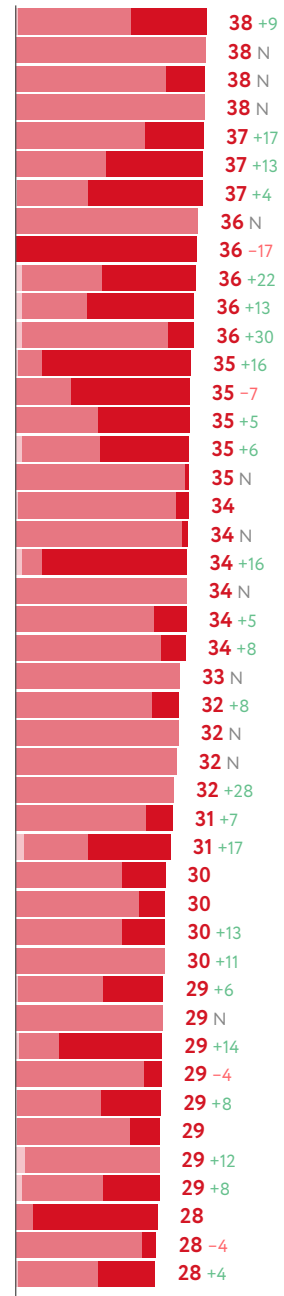


Яндекс

корреляция с позицией ■
связь с попаданием в топ-30 ■
рост за год ■



Оплата картой	24, 71, 70
Цены на странице: медиана	66-67
Фрагменты: неточный запрос + синонимы	34-35
Цены на странице: средняя	66-67
Количество картинок	38, 60, 25, 48
Оплата на странице	71
Несколько фотографий	70
Цены на странице: количество	66-67
Посещения: реклама (display)	70, 55
Группы в соцсетях	65-66
Группа ВКонтакте	65-66
SEO-текст: «тошнота» словосочетаний	42
Много регионов	60-61, 55
SimilarWeb Country Rank	57-58
Найдено документов с сайта в Яндексе	60-61, 27, 55
Доля ссылок на главную страницу сайта	59, 54-56
: общая длина	38, 47
Локализация в Москве	14-15, 60, 28
: количество слов	38, 47
Более одного региона	60
Фрагменты: синонимы	34-35
Текст: слова запроса	33-35
Текст: слова запроса и синонимы	33-35
Цены на странице: наличие	28-29, 66
Исходящие ссылки: слова запроса	46-47, 33-35
Атрибуты title: количество слов	47-48
Цены на странице: максимальная	66-67
SEO-текст: плотность повторов	42
Внутренние ссылки: слова запроса	46-47, 33-35
Группа в «Одноклассниках»	65-66
Текст: выделенные	33-35
Исходящие ссылки: слова запроса и синонимы	33-35
Текст: слова запроса и выделенные Google	33-35
Description: выделенные	44, 33-35
Телефон 8-800	60-61, 28, 76, 10
Фрагменты: выделенные Google	34-35
SimilarWeb Category Rank	57-58
Количество слов в тексте страницы	38, 27
Соцсети («Поделиться»)	65-66, 55
Наличие микроразметки	73
SEO-текст: «академическая тошнота»	42
Отзывы	64, 76, 28, 55
Локализация в Санкт-Петербурге	60-61, 55
Размер текста страницы	38, 23, 27
Facebook («Поделиться»)	65-66



ВВЕДЕНИЕ

За 2019 год Лаборатория поисковой аналитики компании «Ашманов и партнеры» подготовила уже пять аналитических отчетов — это шестой.



Мы подробно рассмотрели особенности ранжирования в отраслях, которым были посвящены прошедшие в этом году конференции eTarget: **электронной коммерции, финансах, медицине, автомобилях и недвижимости**.

Это дало более объемную и разнообразную картину, и теперь пора вернуться к общим закономерностям ранжирования в Яндексе и Google и рассмотреть их с учетом тех тенденций, которые мы видели на отрасле-

вых выборках. На этот раз мы постараемся не только показать, какие параметры статистически связаны с позицией и попаданием в топ, но и приблизиться к ответу на вопрос о том, что действительно важно для ранжирования. Для того, чтобы наши выводы не оставались для вас пустым звуком, а диаграммы — просто цветными картинками, мы решили еще раз вместе с вами разобраться, *откуда мы это знаем* (стр. 08–22), и, главное, *как использовать наши результаты* (стр. 23–30).



Основные группы факторов ранжирования Яндекса и Google мы подробно описывали уже дважды — в аналитических отчетах [Факторы-2017](#) и [Факторы-2018](#). Год на год не приходится — но этот, к счастью, был достаточно спокойным, и за прошедшие месяцы больших изменений в ранжировании мы не увидели. Изменения, конечно, есть (и они видны в нашем рейтинге параметров, стр. 02–05), но они скорее количественные, чем качественные¹, и сказанное в двух предыдущих больших аналитических отчетах в основном не теряет актуальности.

¹ Уже в сентябре, когда работа над этим отчетом кипела, от Google пришли новости о September 2019 Core Update. Мы уже не успеваем учесть изменения, связанные с этим апдейтом, в отчете — но, скорее всего, на коммерческом ранжировании они скажутся не слишком сильно.

Значит, на этот раз мы можем обойтись без уже ставшего привычным последовательного обзора — коммерческие, ссылочные, социальные, трафиковые и поведенческие, технические, текстовые факторы. Вместо этого мы рассмотрим некоторые важные принципы и тенденции ранжирования, в рамках которых нередко поддерживают и дополняют друг друга параметры из разных групп.

Мы подробно остановимся на том, из чего складывается *текстовая релевантность* страницы запросу, (стр. 31–51) а оставшуюся часть отчета посвятим логике многочисленных и разнообразных факторов, отвечающих за *ранжирование сайтов* (стр. 52–76).



ОТКУДА МЫ ЭТО ЗНАЕМ

Мы уже не раз рассказывали о том, как работает Лаборатория поисковой аналитики и на чём основаны наши выводы — но на этот раз хотелось бы остановиться на этом немного подробнее.



Мы постараемся по возможности не очень повторяться — но когда рассказываешь примерно об одном и том же уже восьмой раз, избежать повторений трудно. Поэтому заранее извиняемся перед теми, кто уже изучил один или несколько наших предыдущих отчетов и теперь вынужден читать об этом снова. Надеемся, что их терпение будет вознаграждено, и они найдут здесь новые детали, которые сложатся в более полное представление не только о нашей методике, но и о том, как устроено ранжирование Яндекса и Google.

Кстати, в [Финансах-2019](#) мы собрали вопросы, которые часто задают по нашей методике — и про корреляции, и про карго-культ, и про размер выборки, и про количество параметров, и про исключения из правил. Если после прочтения этого раздела у вас останутся сомнения, загляните туда, в главу *Наша методика в вопросах и ответах*.

Если вопросы всё равно останутся, если вы хотите что-нибудь с нами обсудить или предложить сотрудничество, пишите нам на turgenev@ashmanov.com.

Параметры

Цель наших исследований — увидеть за параметрами сайтов и страниц, которые мы различными правдами и неправдами умеем им приписать, факторы ранжирования, используемые поисковыми машинами при формировании поисковой выдачи.

Прежде всего, нужно было понять, какие характеристики сайтов и страниц могут быть важны для ранжирования, а затем придумать, как их «сосчитать». Сейчас параметров уже больше 600 — разумеется, не все они претендуют на то, чтобы быть факторами ранжирования. Мы собираем² и понемногу реализуем новые идеи, количество параметров медленно, но верно растет.

Большую часть параметров мы сами «снимаем» со страниц, попавших в результаты поиска. Например, мы уверенно определяем размер HTML-кода, долю в нём текста и ссылок, количество вхождений слов запроса в заголовки `<h2>`–`<h4>`, наличие счетчика Google Analytics или протокола `https`. Немного сложнее определить, например, максимальную цену на странице, концентрацию слов запроса и их синонимов в SEO-тексте — для этого нужно научиться находить на странице цены, выделять SEO-текст и приписывать словам запроса синонимы.

Значения некоторых параметров мы получаем от самих поисковиков или сторонних сервисов. Например, значения [ИКС](#) (а раньше ТИЦ), количество страниц в индексе и количество страниц, найденных на сайте, мы узнаем у Яндекса. Количество «шервов» в соцсетях можно узнать ВКонтакте и в «Одноклассниках», а возраст сайта — напрямую на [whois-сервисах](#) или косвенно по первому упоминанию на [Archive.org](#). Данные о ссылках мы получаем от [MegalIndex](#), трафиковые и некоторые поведенческие параметры — от [SimilarWeb](#) и [Alexa.com](#).

И наконец, около ста параметров основаны на экспертных оценках. Их выставляют наши ассессоры, которые для каждой страницы, попавшей в результаты поиска, заполняют специальную анкету. «Ручные» (ассессорские) параметры — в основном коммерческие или социальные. Большая часть из них относятся к сайту в целом. Если какие-либо другие страницы с сайта ранее уже были оценены, ассессор получает уже почти готовую анкету — ему остается заполнить только строки, относящиеся к конкретной странице. Ассессорские оценки есть только для коммерческих сайтов — если ассессор оценил сайт как информационный, то на этом заполнение анкеты заканчивается.

² Если вы хотите предложить новые параметры и у вас есть идея, откуда и как взять для них данные, пишите нам на turgenev@ashmanov.com.

Контакты и пр.				
Адрес	<input checked="" type="radio"/> В обвязке	<input type="radio"/> Есть	<input type="radio"/> Трудно найти	<input type="radio"/> Нет
- Адрес на странице	<input checked="" type="radio"/> Есть	<input type="radio"/> Нет		
- Более одного региона	<input checked="" type="radio"/> Один	<input type="radio"/> Несколько	<input type="radio"/> Много	
- Более одного адреса	<input type="radio"/> Один	<input checked="" type="radio"/> Несколько	<input type="radio"/> Много	
Телефон	<input checked="" type="radio"/> В обвязке	<input type="radio"/> Есть	<input type="radio"/> Трудно найти	<input type="radio"/> Картинкой
- Телефон на странице	<input checked="" type="radio"/> Есть	<input type="radio"/> Нет		<input type="radio"/> Нет
- Более одного телефона	<input type="radio"/> Один	<input checked="" type="radio"/> Несколько	<input type="radio"/> Много	
- Только мобильные	<input type="radio"/> Да	<input checked="" type="radio"/> Нет		
- Телефон 8-800	<input checked="" type="radio"/> Есть	<input type="radio"/> Нет		
Локализация в Москве	<input checked="" type="radio"/> Есть	<input type="radio"/> Вторичная	<input type="radio"/> Нет	<input type="radio"/> Не актуально
Локализация в Санкт-Петербурге	<input type="radio"/> Есть	<input type="radio"/> Вторичная	<input checked="" type="radio"/> Нет	<input type="radio"/> Не актуально
Онлайн-консультант	<input type="radio"/> В обвязке	<input type="radio"/> Есть	<input type="radio"/> Трудно найти	<input checked="" type="radio"/> Нет
Обратный звонок	<input checked="" type="radio"/> В обвязке	<input type="radio"/> Есть	<input type="radio"/> Трудно найти	<input type="radio"/> Нет
- Обратный звонок на странице	<input checked="" type="radio"/> Есть	<input type="radio"/> Нет		

Рис. 1. Фрагмент ассессорской анкеты. Большинство параметров (кроме тех, где прямо указано «на странице») относятся к сайту в целом. Часто предусмотрено несколько вариантов — в таких случаях одна строка анкеты обычно дает материал для двух или больше бинарных параметров. Оценка «не актуально» означает, что сайт (страница) не будет участвовать в подсчетах средних значений параметра.

«Ручные» оценки повышают трудоемкость нашей работы и ограничивают число запросов, которые мы можем оценить за единицу времени, но при этом помогают решить (или, точнее, обойти) множество проблем. Так, они позволяют нам ограничиться выкачкой только тех страниц, которые попали в результаты поиска, и не выкачивать сайты в целом. Но главное, они избавляют нас от необходимости включаться в технологическую гонку с поисковиками и придумывать способы распознавания нетривиальных объектов и ситуаций. Конечно, человек распознаёт наличие информации о гарантии, отзывах или скидках не совсем так, как робот Яндекса, — но скорее всего он это делает не хуже.

Многие ассессорские параметры для отдельных страниц дублируются близкими автоматическими параметрами, однако резуль-

таты по ним не всегда совпадают. Например, ассессор учтет только телефон компании — владельца сайта, а автомат — и телефоны клиентов; поэтому при автоматическом распознавании на сайтах агрегаторов найдется больше телефонов вообще и телефонов 8-800 в частности.

Параметры также можно разделить на *бинарные* (есть или нет у сайта или страницы какое-то свойство) и *числовые*. Значения бинарных параметров в тех случаях, когда речь идет о нескольких результатах поиска (например, о топ-10) и/или о выборке запросов, удобно указывать в процентах: например, 31% означает, что для 31% сайтов, попавших на интересующие нас позиции по данной выборке, значение параметра — *есть*. С числовыми параметрами всё немного сложнее: если разброс значений невелик, мы используем

средние значения; если разброс значительный, мы считаем для каждого из запросов медиану³, а затем усредняем полученные результаты для всех запросов выборки.

Параметры, у которых несколько значений, мы обычно сводим к бинарным — так

³ Для вычисления медианы результаты (в нашем случае, значения параметра для сайтов, попавших в результаты поиска по запросу) упорядочиваются по возрастанию; если их нечетное число, берется средний по порядку элемент в списке; если четное — среднее арифметическое двух средних элементов.

с ними гораздо удобнее работать. Например, в нашей ассессорской анкете предусмотрены следующие варианты для количества брендов в ассортименте: *один, несколько или много*. Мы используем полученные данные из этой строки анкеты в двух бинарных параметрах: *Один бренд* (да, если один; нет, если несколько или много) и *Много брендов* (да, если много; нет, если один или несколько). Есть еще вариант *не актуально*, который просто выключает учет этой группы параметров для сайта.

Выборки

Любое SEO начинается с «семантики» — то есть с выбора запросов. От них зависит очень многое, и априори понятно, что факторы ранжирования могут проявлять себя по-разному в зависимости от того, с каким набором запросов мы работаем.

И действительно, значимость одних и тех же параметров (о том, что мы вкладываем в эти слова, см. ниже) на разных выборках запросов может сильно различаться, что существенно затрудняет «расшифровку» алгоритмов ранжирования. Это может быть связано и с объективными особенностями ранжирования для тех или иных тематик или интенгов (намерений пользователя), и с «ландшафтом» сайтов, конкурирующих за место в топе: параметры, с которыми всё хорошо у сайтов с высокой видимостью, будут казаться важными, даже если поисковая машина на них не обращает внимания. Поэтому важно, во-первых, чтобы

контрольная выборка запросов была достаточно представительной и разнообразной и, во-вторых, чтобы полученные на ней результаты контролировались по другим выборкам.

Сейчас мы работаем с четырьмя типами выборок:

1. «эталонная» политематическая общекоммерческая выборка;
2. эталонные тематические (отраслевые) выборки;
3. запросы, по которым продвигаются сайты наши клиентов;

4. метавыборки, составленные из различных выборок предыдущего типа.

В этом исследовании мы в основном будем анализировать данные нашей эталонной общекоммерческой выборкой из 160 запросов, за которой мы наблюдаем с марта 2015 года.

Осенью 2014 года мы начали с 35 запросов разных тематик, затем каждый месяц их количество увеличивалось и в марте 2015 достигло 160. Мы оценили получившуюся выборку как достаточно представительную и решили на этом остановиться. Во-первых, чем больше запросов, тем больше работы для ассессоров; во-вторых, важно было как можно скорее зафиксировать набор запросов, чтобы в дальнейшем ежемесячные выкачки можно было сравнивать между собой. Сейчас понятно, что у получившейся эталонной выборки есть недостатки: некоторые тематики и типы запросов в ней отсутствуют или представлены недостаточно, а другие, наоборот, можно было бы «проредить»; в нее попало несколько преимущественно информационных запросов, и так далее. Но идеал, по-видимому, всё равно недостижим — а наша выборка по крайней мере дает нам возможность проследить изменения в ранжировании за последние четыре с половиной года.



Везде, где специально не указано иное, данные приводятся по выкачке за сентябрь 2019 для нашей стандартной выборки из 160 запросов, для всех коммерческих сайтов (включая агрегаторы). Изменения — по сравнению с выкачкой за октябрь 2018, которая легла в основу Факторов-2018.

Мы будем сравнивать результаты, полученные на этой выборке, с данными пяти тематических выборок, которые легли в основу наших отраслевых отчетов. Объем этих выборок несколько больше (300–700 запросов), но в целях экономии ресурсов ассессорам была выдана на оценку только часть запросов (от 50 до 70).

Несмотря на то, что ассессорские оценки получены далеко не для всех запросов, «ручные» сайтовые параметры оценены для этих выборок достаточно полно, благодаря тому, что очень многие сайты, попавшие в результаты поиска по неоцененным запросам, найдены и по запросам, для которых оценки есть, или встречались нам раньше. Со страницными параметрами дело обстоит гораздо хуже, т. к. одна и та же страница гораздо реже находится сразу по нескольким запросам, входящим в выборку. Однако они менее важны (коммерческие параметры оценивают прежде всего сайты в целом) и часто дублируются автоматическими параметрами, для которых есть полные данные.

Клиентские выборки и метавыборки в этом отчете мы не используем⁴ — но с их помощью мы не раз убеждались, что выводы, которые мы делаем на основе относительно небольших выборок, в основном подтверждаются и по данным гораздо большего объема.

⁴ В таких выборках ассессирование часто ограничивается одной поисковой машиной, продвижение в которой интересует клиента, поэтому данные получаются неполными. И это далеко не единственный недостаток таких выборок для наших целей.

Математическая статистика

Теперь мы подошли к самому главному. У нас есть выборка, есть значения параметра для всех (ну или почти всех) сайтов/страниц, попавших в топ-30 Яндекса, Google и Mail.ru. Как понять, важен ли параметр для ранжирования в каждом из поисковиков? И что, собственно, значит «важен»?



Сразу оговоримся, что, утверждая важность того или иного параметра, мы не хотим сказать, что доказали факт включения этого параметра в формулу ранжирования.

Мы утверждаем только, что мы видим статистически достоверную связь между значениями параметра и ранжированием. От чего она могла возникнуть и о чём говорит — тут вариантов масса, и мы рассмотрим их подробнее ниже. Прямое включение в формулу ранжирования — только один из них.

Тем не менее, если наши данные показывают, что параметр «важен для ранжирования», то это в большинстве случаев действительно важно, и при поисковой оптимизации сайтов эти выводы полезно учитывать. Хотя и не стоит добиваться улучшения показателей по важным параметрам *любой ценой*.

КОРРЕЛЯЦИИ С ПОЗИЦИЕЙ

Самая простая связь с ранжированием — это когда с приближением к первой пози-

ции значения параметра растут (или падают). Иногда такая зависимость просто очевидна, если посмотреть на диаграмму средних значений параметра по позициям. Однако возможны различные нюансы, особенно с числовыми параметрами, где на общую картину могут сильно влиять «выбросы» значений для отдельных сайтов. Поэтому лучше оценивать зависимость при помощи методов математической статистики.

Мы используем в качестве основного *ранговой критерий Спирмена* (см. врезку на стр. 17) — именно его значение мы имеем в виду, когда говорим о *коэффициенте корреляции* между позицией и значением параметра⁵. Мы считаем корреляцию с позицией сильной, если коэффициент корреляции — 0,10 и больше (или -0,10 и меньше). Ранговой критерий Спирмена используется нами также при оценке корреляций между параметрами.

⁵ С точностью до знака. Если параметр положительно сказывается на ранжировании, коэффициент Спирмена отрицательный: чем *меньше* позиция, тем *больше* значение параметра. Чтобы не запутать читателей, мы называем такую корреляцию «прямой» и убираем минус. Если же при приближении к первой позиции значения параметра падают, мы называем корреляцию «обратной» и минус ставим. Таким образом, наш коэффициент корреляции параметра с позицией — это ранговой критерий Спирмена, умноженный на -1.

При этом мы предпочитаем не полагаться только на одну метрику и контролируем ее при помощи точного теста Фишера (а также U-теста Манна–Уитни). Корреляция засчитывается, только если она «подтверждается» статистически достоверной разницей между значениями параметра для топ-3 или топ-10 и оставшейся частью топ-30.

Чем сильнее корреляция, тем обычно больше перепад значений между топ-3 и третьей десяткой выдачи (исключения могут возникать, например, когда максимум значений приходится не на первые позиции, а на середину первой десятки, как мы это часто видели в отчете по недвижимости). На наших стандартных диаграммах по параметрам мы третью десятку не показываем, но при сильной корреляции разница в средних значениях высока и для топ-3 VS. топ-30.

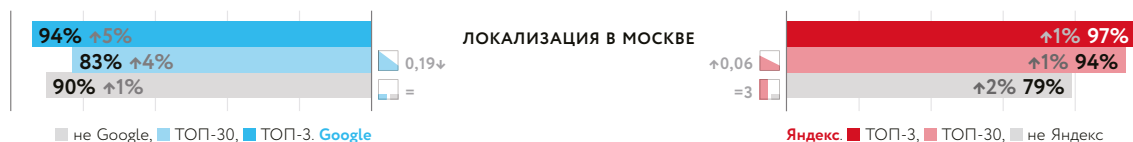
Например, для **локализации в Москве** (см. рис. 2 и стандартную диаграмму ниже) для Яндекса среднее значение для топ-3 — 97%, для топ-30 — 94%, для третьей десятки выдачи (позиции 21–30) — 92%, коэффициент корреляции 0,06. Для Google коэффициент корреляции заметно выше — 0,19; сильнее и разброс значений: топ-3 — 94%, топ-30 — 83%, третья десятка — 74%.

Отметим, кстати, что чем ближе к нулю или к 100%, тем «весомее» разница в значениях параметров. Казалось бы, между 98% и 96%, как между 58% и 56%, — всего два процентных пункта. Но рост с 96 до 98% означает уменьшение доли сайтов, не имеющих параметра, в два раза!

КОНТРАСТ С ФОНОМ

Большая часть запросов в выборках, с которыми мы работаем, — коммерческие и высококонкурентные; по ним нет недостатка в релевантных страницах. Качество поиска у всех трех основных поисковых систем достаточно высокое, чтобы попадание в топ-30 хотя бы одной из них можно было считать гарантией достаточной релевантности запросу и качества. Значит, если страница не попала в топ-30, например, в Google, но попала в Яндекс или в Mail.ru, — то это не потому, что она вообще плохая, а потому, что Google предпочел ей какие-то другие страницы (или сайты), лучше соответствующие каким-то важным для него критериям.

Значит, сравнивая топ-30 поисковика с фоном — со страницами, которых там нет, но которые попали в выдачу по тем



■ 3, ■ 2, ■ 1 — сильная, средняя, слабая связь с попаданием в топ-30, □ — связи с попаданием в топ нет.
 ■■■ — сильная, средняя, слабая корреляция с позицией, 0,08 — коэффициент корреляции с позицией (ранговый критерий Спирмена), □ — связи с позицией нет, ■■■■ — корреляция с позицией с обратным знаком.
 ↑1% — сравнение со стандартной выборкой (% — процентные пункты, = — изменений нет).

На таких диаграммах указаны коэффициенты корреляции, а также сила связи с попаданием в топ-30. Но «сила» параметров обычно видна и без этого. Если значения для топ-3 заметно выше, чем для топ-30, значит, скорее всего, есть сильная корреляция между параметром и позицией. Если значения для топ-30 заметно выше, чем для фона, скорее всего, есть сильная связь с попаданием в топ.

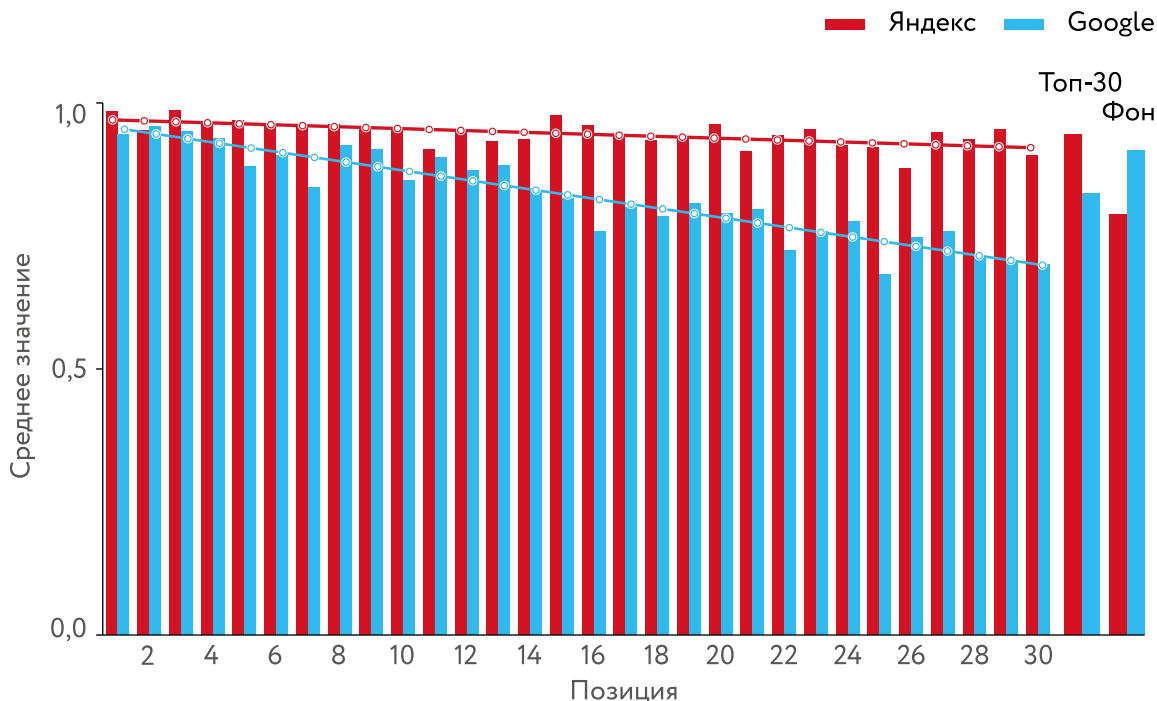


Рис. 2. Параметр **локализация в Москве** (что в нашем случае равносильно локализации в регионе запроса, поскольку выкачка московская). В Google сильная корреляция с позицией: Спирмен $-0,19$, Пирсон $-0,93$, МНК $-0,0085 \times r + 0,95$. В Яндексе корреляция с позицией средней силы, значения всех метрик заметно ниже по модулю: Спирмен $-0,06$, Пирсон $-0,51$, МНК $-0,0017 \times r + 0,97$. Корреляция с позицией в Яндексе (которой, кстати, год назад не было видно) «подтверждается» значениями других метрик: Манн-Уитни $-0,00010$; Фишер для топ-3 VS. 4–30 $-0,00082$, для топ-10 VS. 11–30 $-0,00153$. Для Google у всех этих метрик в первых пяти знаках после запятой только нули. Зато в Яндексе сравнение топ-30 с фоном (т. е. с сайтами, попавшими в выдачу двух других поисковиков, но не данного) дает значение точного теста Фишера, близкое к нулю (нули в 5 знаках после запятой), а для Google — близкое к единице. Это означает, что сайтов с локализацией в Москве в Яндексе статистически достоверно больше, чем в фоне.

же запросам в других поисковых машинах, — мы можем многое узнать о том, какие параметры для него важны. Например, мы можем утверждать, что локализация в Москве важнее для Яндекса, чем для Google, несмотря на то, что корреляция с позицией в Google выше. Потому что в топ-30 Яндекса 94% сайтов локализо-

ваны в Москве, а в топ-30 Google — только 83%. Получается, что в Google доля сайтов, не имеющих московских адресов и телефонов, почти в три раза больше — 17% против 6%.

Наш основной инструмент для оценки разницы между топ-30 и фоном — *точный тест*

Фишера (см. врезку). Чем ближе его значение к нулю, тем меньше вероятность того, что разница в значениях параметров между выдачей поисковой машины и фоном случайна.

Но если связь с попаданием в топ-30 не случайна, то чем она может определяться? Есть несколько основных вариантов, которые, впрочем, могут и дополнять друг друга в различных пропорциях.

1. Предварительный отбор. Современные формулы ранжирования очень сложны — а значит, могут требовать много ресурсов и времени на вычисление. Чтобы уменьшить нагрузку на сервера и ускорить обработку, ранжирование может выполняться в два или больше этапов, причем сначала к большому количеству страниц, поднятых из индекса, применяются упрощенные формулы, оптимизированные под быстрое вычисление. В итоге отбирается относительно немного (например, тысяча⁶) результатов, к которым уже и применяется полная формула ранжирования. Если параметр входит в упрощенную формулу и учитывается на ранних этапах ранжирования, его значения для сайтов, дошедших до последнего этапа, изначально будут в среднем высокими. И даже если на этом последнем этапе параметр уже не учитывается, его влияние на общие результаты ранжирования может быть очень большим.

2. Продолжение градиента. Некоторые из страниц фона на самом деле попадают в выдачу, но ниже 30-й позиции. Если есть корреляция с позицией, т. е. с удалением от вершины значения параметра падают,

⁶ Вероятно, именно предварительным отбором объясняется тот факт, что при декларируемых сотнях тысяч или миллионах результатов поиска по запросу, открывая поисковую выдачу страницу за страницей, можно увидеть лишь порядка тысячи результатов, а затем выдача обрывается. Вероятно, те страницы, которые мы можем увидеть, — это и есть страницы, отобранные на предыдущих этапах ранжирования.

то не удивительно, если сайты (страницы), попавшие на 47-ю или 350-ю позицию, имеют в среднем более низкие значения параметра, чем попавшие на 2-ю или 25-ю.

3. Корреляции с другими параметрами.

Если точный тест Фишера показывает, что распределение не случайно, это означает только, что более высокие значения параметра в топ-30 по сравнению с фоном — не результат случайного выигрыша в лотерею, а следствие каких-то объективных причин. Причины могут быть разными — и они не обязательно связаны с отбором по интересующему нас (или близкому) параметру. Например, в выдаче Google обычно заметно меньше агрегаторов, чем в выдаче Яндекса. Средние значения многих параметров для агрегаторов заметно выше или ниже, чем для других коммерческих сайтов, — и это может сильно влиять на разницу между средними значениями этих параметров для двух поисковиков.

Чем больше контраст между топ-30 и фоном, тем больше вероятность того, что параметр действительно учитывается при ранжировании (или по крайней мере близок к одному из факторов ранжирования). Как всегда, при приближении к нулю или 100% разница весомее. Однако всегда могут найтись альтернативные объяснения — например, связанные с распределением типов сайтов в выдаче.

С другой стороны, отсутствие контраста между топ-30 и фоном или даже более высокие значения в фоне не обязательно говорят о том, что отбора по параметру нет. Если корреляция с позицией зависит только от одной поисковой машины, то фон — от всех трёх. Если кто-то из конкурентов уделяет интересующему нас параметру больше внимания, «наша» поисковая машина

может оказаться в его тени, хотя тоже отдает предпочтение сайтам с высокими значениями параметра. Поэтому отрицательные корреляции мы на диаграммах отмечаем, а отри-

цательную связь с попаданием в топ — нет (за исключением инвертированных параметров вроде рангов, для которых чем меньше, тем лучше).

СТАТИСТИЧЕСКИЕ МЕТРИКИ, КОТОРЫЕ МЫ ИСПОЛЬЗУЕМ*

Ранговый критерий Спирмена

Коэффициент корреляции между рангами величин; показывает, насколько правдоподобно утверждение, что чем выше позиция, тем больше (или меньше) значение параметра. Принимает значения от -1 до 1 ; чем дальше от нуля (в любую сторону), тем проверяемое утверждение правдоподобнее. Не использует никакой информации о точном значении параметра, только порядок, поэтому его особенно удобно использовать для средних из бинарных величин.

Точный тест Фишера

Показывает значимость наличия параметра (для числовых параметров — того, что значение выше медианного) для попадания в определённую группу (топ-10, топ-30 и т. п.), т. е. сравнивает между собой сайты, попавшие и не попавшие в эту группу. Значение от 0 до 1 — вероятность того, что столь же перекошенное разбиение могло получиться случайно. Чем ближе к крайним значениям — нулю или единице, — тем вероятнее, что параметр влияет на разбиение. Обычно различия между группами считаются значимыми, если в точном тесте Фишера для них есть несколько нулей или девяток после запятой (например, $0,00002$ или $0,99998$). Удобен тем, что хорошо работает и при небольшом количестве данных.

Коэффициент корреляции Пирсона

Одна из статистических метрик, наиболее широко используемых в естественных науках, принимает значения от -1 до 1 . Показывает, похожа ли зависимость между позицией в выдаче и некоторым параметром на линейную (чем дальше от нуля, тем больше похожа). Используется в первую очередь для небинарных параметров.

МНК («метод наименьших квадратов», коэффициент классической линейной регрессии)

Нормированный относительно среднего значения градиент прямой наилучшего приближения. Показывает, как (в частности, под каким наклоном) проходила бы прямая, если свести к ней график. Прямые на наших столбчатых диаграммах, например, на рис. 2, — это и есть МНК. Не имеет смысла, если коэффициент Пирсона близок к нулю — то есть если график на прямую совсем не похож. Приблизительно равен относительному изменению значений параметра при сдвиге на соседнюю позицию.

U-тест Манна–Уитни

Статистический тест, который сравнивает между собой позиции сайтов с параметром и сайтов без параметра (или со значением параметра выше/ниже медианного). Оценивает статистическую значимость «превосходства» (в нашем случае, лучшего ранжирования) одной из групп. Значение от 0 до 1 показывает, насколько вероятно, что данные — результат случайного разбиения; чем ближе к нулю, тем увереннее можно говорить, что одна из групп ранжируется выше.

* Написано Сергеем Вальковским для Факторов-2017 (публикуется с небольшими изменениями).

Дополнительные сложности

О некоторых сложностях в интерпретации результатов мы уже писали — но, конечно, всё еще намного сложнее.



ОБЩИЕ И ЧАСТНЫЕ ФАКТОРЫ

Некоторые факторы ранжирования, скорее всего, работают для всех запросов и сайтов — хотя всё равно в разных случаях могут иметь разный вес. Но известно, что поисковые машины уделяют огромное внимание анализу типов запросов, интенгов пользователей и т. п., и для каждого из них формула ранжирования может быть модифицирована — в том числе в нее могут включаться дополнительные частные факторы.



С некоторыми из них нам уже пришлось сталкиваться в отраслевых отчетах. Так, в автомобильном поиске мы видели особое отношение поисковых машин к **сайтам производителей**, которые «подтягивались» в выдачу по многим запросам явно на особых условиях, и к сайтам официальных дилеров.



Неочевидность деления на общие и частные факторы сильно затрудняет анализ: в каких-то случаях связь с ранжированием может остаться незамеченной; в других частный параметр может прикинуться общим и показаться более значимым, чем он этого заслуживает.

Нет сомнений, что единственный путь преодоления этих трудностей — перейти от общей картины к изучению отдельных кейсов, а затем от них вернуться к общей картине. Именно по этому пути мы в этом году и шли.

- На медицинской выборке мы впервые подробно рассмотрели ранжирование по информационным запросам — и увидели там свои закономерности и свои проблемы.
- Поисковая выдача в финансах и, отчасти, в медицине оказалась *химерной*: логику ранжирования в этих отраслях трудно уловить, не разделив выдачу на информационные сайты, агрегаторы и другие коммерческие сайты (банки, клиники и т. п.).
- В недвижимости, напротив, доминирование агрегаторов столь велико, что проще описывать ранжирование без деления на типы сайтов, а агентствам и даже застройщикам можно порекомендовать «прикинуться агрегатором».
- В финансах и в медицине усиливаются факторы ранжирования, связанные с особой ответственностью поисковых машин при ответе на запросы, относящиеся, в терминологии Google, к категории YMYL (*your money or your life, кошелек или жизнь*) — в частности, связанные с «компетентностью, авторитетностью и надежностью» (*E-A-T — Expertise, Authoritativeness, Trustworthiness*⁷).

⁷ См. об этом, например, в записи в блоге Google, посвященной сентябрьскому апдейту 2019 года: webmasters.googleblog.com/2019/08/core-updates.html.

Сейчас пришло время с учетом этого опыта посмотреть на ранжирование коммерческих сайтов в целом.

КОРРЕЛЯЦИИ МЕЖДУ ПАРАМЕТРАМИ

Разбираясь в том, что важно для ранжирования, а что не очень, приходится учитывать два типа корреляций между параметрами (между которыми, конечно, много промежуточных случаев).

Во-первых, более или менее универсальные: в силу объективных, хотя и не всегда очевидных причин наличие одного признака увеличивает (или уменьшает) вероятность второго. Например, наличие связи с Яндекс.Маркетом положительно коррелирует с оплатой и доставкой, подбором по параметрам и сравнением товаров или с количеством страниц в индексе Яндекса. И отрицательно — с наличием списка услуг, размером SEO-текста страницы и количеством иллюстраций. Все эти связи можно если не предсказать, то объяснить: связь с Яндекс.Маркетом характерна для магазинов, и у них, например, обычно больше страниц, чем у сайтов, предлагающих услуги, потому что на каждый товар приходится по странице (а то и по несколько).

Во-вторых, локальные или ситуационные корреляции, действующие в рамках конкретной выборки. Например, в выборке по недвижимости первое место в обоих поисковиках более чем по двум третям запросов занимает один и тот же сайт — Cian.ru. Понятно, что на этой выборке все параметры, по которым у ЦИАН хорошие значения, будут коррелировать не только друг с другом, но и с позицией. Некоторые из этих параметров действительно помогли ЦИАН добиться столь значительных успехов — но зато остальные просто «примазываются». Чем больше сте-

пень монополизации выборки, тем больше на «естественные» корреляции накладываются таких случайных.

К счастью, наша общая выборка в силу своей политематичности должна быть более или менее свободна от случайных корреляций. А общие корреляции мы посчитали (получилась матрица 650×650) и стараемся учитывать. Некоторые примеры такого анализа см. ниже.

ИНЕРЦИЯ РАНЖИРОВАНИЯ: ПРЕОЦЕНЕННЫЕ ЗАПРОСЫ

Большая часть запросов, входящих в нашу общую выборку (да и во многие другие тоже) достаточно высокочастотные, чтобы поисковики не пускали их на самотёк. Формула ранжирования — это хорошо, но есть еще и обратная связь, которую поисковая машина получает от пользователей, анализируя их реакции на выдачу по конкретным запросам. Если те, кто задал запрос, с удовольствием кликают по сайту, долго на нём остаются, открывают другие его страницы, сайт может быть поднят в выдаче. Если наоборот — то наоборот.

Подчеркнём, что это не «общие» поведенческие метрики, а очень локальные (хотя общие из них тоже могут строиться): они влияют на выдачу по конкретному запросу! В результате выдача по таким запросам корректируется и отчасти замораживается: поисковой машине уже не нужно ее просчитывать, а достаточно взять уже готовую из кэша.

Эффект переоценки запросов несколько смазывает картину ранжирования: сайты оказываются расставленными не совсем так, как они «заслуживают» в соответствии со своим набором параметров; к тому же изменения



в формуле ранжирования могут доходить до преоцененных запросов (то есть до тех, с которыми мы имеем дело) с задержкой.

Впрочем, локальные поведенческие факторы, «искажающие» поисковую выдачу, — это просто еще одна группа факторов, которые мы пока не научились оценивать. Весь наш подход направлен на то, чтобы оценивать вклад известных нам параметров, даже если на результат кроме них влияют и многие другие, часть из которых нам вообще неизвестна.

ПРЕОДОЛЕНИЕ ИНЕРЦИИ: «МНОГОРУКИЕ БАНДИТЫ»⁸

Поисковая выдача консервативна — и не только по тем причинам, которые описаны выше.

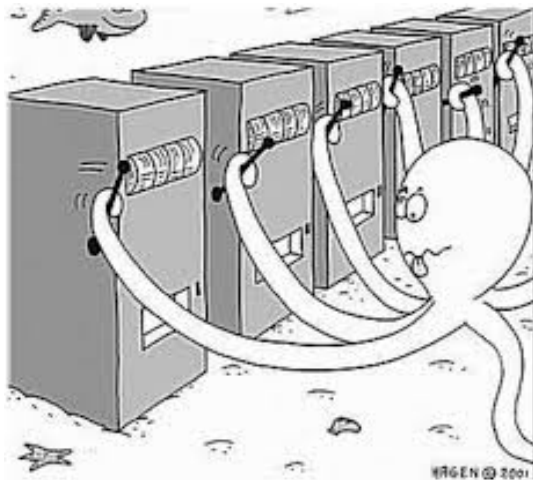


Рис. 3. Название алгоритма «многорукый бандит» возникло из задачи машинного обучения, моделирующей игру с несколькими игровыми автоматами («однорукими бандитами») с различной и при этом заранее не известной вероятностью выигрыша. Первоисточник этого рисунка найти, к сожалению, не удалось.

Сайтам гораздо проще наращивать свой вес (ссылки, аудиторию, трафик, упоминания в соцсетях, ассортимент, объем контента и т. п.), если они уже хорошо ранжируются в поиске. Получается, что «мастодонты» растут быстрее, а новые сайты — медленнее, и им всё труднее преодолевать постоянно растущий порог входа.

Чтобы дать сайтам, не попавшим на высокие позиции «естественным путем», шанс вырваться из этого замкнутого круга (или, точнее, ворваться в него), поисковые машины используют алгоритм «многорукых бандитов»⁸. «Бандиты» «подкидывают» для части пользователей на высокие позиции по запросу или группе запросов сайты, которые обычно ранжируются сильно ниже. Реакция пользователей на модифицированную выдачу сравнивается с их реакцией на обычную — это называется A/B-тестированием. И если эксперимент оказался успешным, подкинутый сайт, за который пользователи проголосовали своими кликами, может закрепиться в топе.

Такое вмешательство в ранжирование — это не учет дополнительных факторов, а локальное искривление формулы. Только для конкретного сайта (и обычно только для некоторого набора запросов) завышается вес поведенческих факторов — на весы кладется гиря, которой у других сайтов нет. Тем не менее, и такие вмешательства в ранжирование для нашей задачи не критичны. Их не может быть слишком много, и влияющие параметры, которые мы исследуем, они не должны сильно исказить.

⁸ См. об этом, например, в блоге Сергея Кокшарова: devaka.ru/articles/multi-armed-bandit.



Кстати, примерно так же, только в другую сторону, работают фильтры вроде «Баден-Бадена» или «Минусинска»: сайт, который в соответствии с обычной формулой ранжирования должен попадать в первую десятку, искусственно отбрасывается на два десятка позиций ниже.

Если бы фильтры применялись массированно, они могли бы серьезно спутать нам карты⁹.

⁹ Не исключено, что на некоторых выкачках, в моменты наиболее активного их применения, фильтры действительно вносят значительные искажения — например, могут повышаться средние значения некоторых параметров в третьей десятке выдачи за счет того, что «хорошие» сайты проваливаются туда из-за фильтра.

Но, к счастью, они созданы не для массового применения — их основная функция не исправительная, а воспитательная. Сама угроза применения «Минусинска» заставляла превентивно снимать закупленные ссылки и даже платить за удаление «вечных» ссылок. Владельцы сайтов, попавших, например, под «Баден-Баден», часто сетуют, что у их конкурентов всё еще хуже, а их не трогают (добавим от себя: пока не трогают). Это потому, что, применяя санкции, Яндекс (или Google) вынужден одновременно заботиться о том, чтобы не просадить качество поиска, — убрать из топа одновременно всех лидеров они себе позволить не могут.

Рейтинг параметров

В этом году он был рассчитан по тем же правилам, что и прошлый раз (см. Факторы-2018), и можно посмотреть, что изменилось за год.

Методика расчета важности параметров, лежащая в основе рейтинга, разработана с участием Сергея Вальковского и реализована им специально для этой серии аналитических отчетов. Рейтинг рассчитан для общей выборки по данным за сентябрь 2019 года для коммерческих сайтов; текущие значения параметров сравниваются со значениями для той же выборки за октябрь 2018. В этом году мы решили внести небольшие изменения в параметры (повышен вес корреляции с позицией, см. ниже), поэтому дан-

ные за 2018 год перед сравнением также были пересчитаны с новыми параметрами. Отмечены изменения значений рейтинга от 4 баллов, а также параметры, для которых изменения не оценивались (нет данных за прошлый год).

Рейтинг не учитывает корреляции между параметрами, и его высокие значения не означают, что параметр входит в формулу ранжирования и непосредственно учитывается Яндексом или Google (как и сильная связь с ранжированием

сама по себе). «Важность» некоторых параметров может определяться особенностями нашей выборки (в частности, распределением по типам сайтов и страниц в Яндексе и Google). Учитывается разница между значениями параметра для топ-30 Яндекса (Google) и фоном, а она зависит не только от той поисковой машины, для которой строится рейтинг, но и от двух других.

Рейтинг для каждого параметра складывается из трех компонент, каждой из которых соответствует своя часть «линейки»:

- корреляция с позицией: вес 55%¹⁰, самый темный цвет;
- связь с попаданием в топ-30: вес 35%, средний оттенок;
- рост за год: вес 10%, самая бледная часть линейки (часто короткая или отсутствует).



Самым трудным было оценить вклад в рейтинг связи с попаданием в топ-30, которая, как мы знаем, может быть даже важнее, чем корреляция с позицией.

В самих отчетах мы оцениваем ее очень грубо: по четырехбалльной шкале, в зависимости от степени близости к нулю значения точного теста Фишера для топ-30 VS. фона.

¹⁰ В прошлом году при расчете рейтинга использовались немного другие веса: по 45% за корреляцию с позицией и за связь с попаданием в топ-30.

Была разработана сложная формула, учитывающая:

- статистическую значимость связи (технически — количество нулей после запятой);
- диапазон изменения параметра: для небольших диапазонов используется понижающий коэффициент, для очень больших — повышающий;
- разницу между значениями для топ-30 и фона в сравнении с размером всего диапазона изменения параметра;
- для бинарных параметров — большую цену этой разницы вблизи границ (0% и 100%).

Вклад корреляции с позицией определялся примерно так же, но статистическая значимость (точный тест Фишера) учитывалась только для контроля, а в качестве основного показателя использовался коэффициент корреляции (ранговый критерий Спирмена).

Наконец, небольшой «довесок» к рейтингу может дать изменение значения параметра за год. Этот бонус считается отдельно для топ-3 и топ-30, но результат представляется одним отрезком линейки.

Значения рейтинга считаются в условных баллах, числа около названий параметров — ссылки на страницы этого отчета.

КАК ИСПОЛЬЗОВАТЬ НАШИ РЕЗУЛЬТАТЫ

Повторим еще раз: мы не утверждаем, что параметры, которые мы называем важными, — точно-точно факторы ранжирования. И что, раз так, с ними делать?

Не нужно добиваться роста «важных» параметров любой ценой

Например, наши данные показывают, что чем больше **размер HTML**, **размер текста** и даже **время загрузки страницы** (только в Яндексе и в разумных пределах — менее секунды без учета скриптов и т. п.), тем выше в среднем ранжируются сайты. Мы уверены, что это нетривиальная, важная и полезная информация, помогающая понять, как устроено ранжирование.



Но горе тому, кто решит оптимизировать свой сайт, специально увеличивая размер HTML или время загрузки!

Как размер HTML может быть связан с ранжированием, см. на стр. 25–27.

Старайтесь понять логику

Может быть, вы заметили, что во всех наших аналитических отчетах и докладах на конференциях мы стараемся не просто зафиксировать корреляции между параметрами и позицией сайта, а увидеть за этим логику поисковых машин. Если так, значит, мы работали не зря.

Иногда разобраться в логике взаимосвязей между параметрами непросто, но без этого оптимизация сайта действительно рискует превратиться в карго-культ (любимое возращение наших критиков).

Сравнивайте свой сайт с конкурентами

Размер HTML вряд ли непосредственно учитывается в ранжировании, но его можно использовать для контроля. Если страницы

вашего сайта значительно меньше, чем страницы конкурентов в топ-3, топ-10 или топ-30, — это повод задуматься, всё ли вы правильно делаете и не нужно ли что-то изменить.

Метод сравнения с конкурентами (а иногда и слепого подражания им) — в SEO, прямо скажем, не новость. Например, именно так работают сервисы, формирующие по списку ключевых слов «задание для копирайтеров». Наши данные помогают выбрать для такого сравнения правильные критерии.



Если параметр для вас неактуален, за его отсутствие вам ничего не будет

Информация о доставке на сайте — несомненно, важный параметр. В Google есть корреляция с позицией, в Яндексе — сильная связь с попаданием в топ-30 (89% против 83% в фоне), а для конкретных видов доставки (**курьер**, **самовывоз**) — и корреляция с позицией. Тем не менее, владельцам

сайтов, предлагающих услуги, о доставке можно не беспокоиться — просто потому, что для них она не актуальна.

Оплата картой — не менее важный параметр. В обеих поисковых машинах на нашей общей выборке видна и сильная корреляция с позицией, и связь с попаданием в топ-30 (в Яндексе тоже сильная). Но вот за **автомобили** расплачиваться картой не принято, да и вообще на автомобильных сайтах вопрос о способах оплаты предпочитают не поднимать. И — кто бы мог подумать! — для автомобилей оплата картой (как и другие параметры, связанные с оплатой) если и коррелирует с позицией, то только отрицательно.

Каким образом поисковые машины поймут, что параметр не актуален, — в конце концов, не так уж и важно (хотя, несомненно, интересно). Можно смело исходить из того, что они справятся: они просто не могут позволить себе в таких вещах сильно ошибаться.

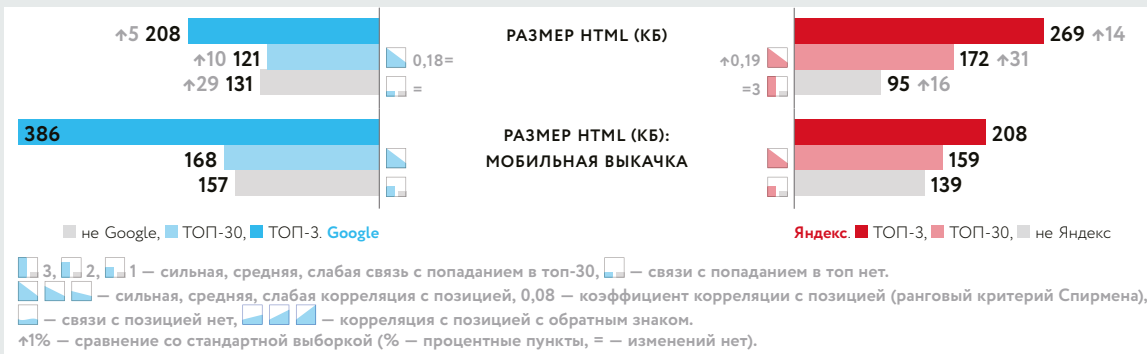


■ 3, ■ 2, ■ 1 — сильная, средняя, слабая связь с попаданием в топ-30, ■ — связи с попаданием в топ нет.
 ■ — сильная, средняя, слабая корреляция с позицией, 0,08 — коэффициент корреляции с позицией (ранговый критерий Спирмена),
 ■ — связи с позицией нет, ■ — корреляция с позицией с обратным знаком.
 ↑1% — сравнение со стандартной выборкой (% — процентные пункты, = — изменений нет).

ПОЧЕМУ У ХОРОШО РАНЖИРУЕМЫХ САЙТОВ HTML-КОДА БОЛЬШЕ

И почему он из года в год растет?

- Интернет давно уже приходит к нам не через модем, и загрузить страницу в несколько сот килобайт — не проблема даже через мобильную сеть. Факторы, ограничивающие размер HTML, во многом потеряли актуальность.
- Размер HTML сильно коррелирует с количеством исходящих внутренних ссылок, картинок, цен и т. п. — т. е. с размером «витрины» (см. в таблице параметры с индексом **В**).
- То, что дело именно в размере витрины, подтверждается корреляциями с параметрами, важными для интернет-магазинов (индекс **М**). Это оплата, оплата картой, доставка (и разные ее варианты), акции и скидки, связь с Яндекс.Маркетом.
- О том, что важен размер витрины, говорят корреляции с ассортиментом, количеством страниц в индексе, трафиком и т. п. — параметрами, связанными с размером сайта (индекс **Р**).
- У магазинов сейчас не текстовые блоки, а именно витрина вносит основной вклад в текстовую релевантность. Здесь не накажут за «ключевики» и повторы, а большой объем не назовут «портянкой» (индекс **Б**, от «Баден-Баден»).
- И действительно, размер HTML коррелирует с количеством вхождений слов запроса в «текстовые фрагменты» (т. е. вне SEO-текста), в ссылки, в атрибуты **alt**; с количеством повторов и «содержательных» слов в тексте страницы (индекс **Т** — параметры, имеющие отношение к текстовому ранжированию).
- Большая витрина — это одновременный и рост количества текста на странице (на каждой карточке товара есть немного текста, и если их много, его объем растет), и уменьшение его доли в HTML-коде. Но размер HTML коррелирует и с другими параметрами, связанными со сложностью кода и технологической продвинутостью сайта (индекс **С**).
- Эта сложность может отвечать, в частности, за улучшение пользовательских характеристик сайта — с поведенческими параметрами (индекс **П**) размер HTML тоже хорошо коррелирует. Правда, эта связь может быть опосредована размером витрины.



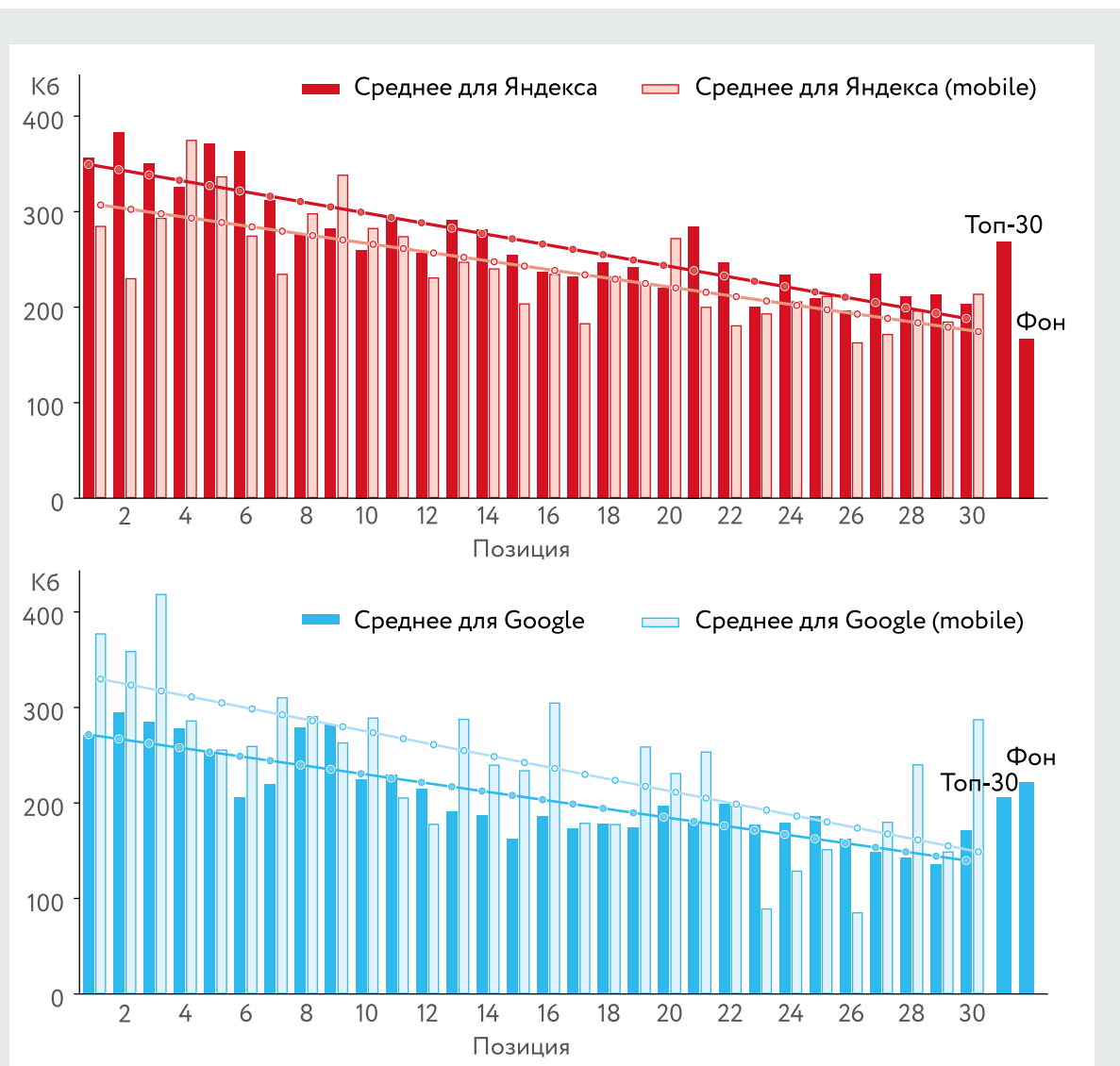


Рис. 4. Средний размер HTML-кода (без учета скриптов, CSS-таблиц и т. п.) для Яндекса и Google, десктопная и мобильная выкачка. Как это часто бывает с числовыми параметрами, средние значения заметно выше медианных, которые показаны на диаграммах (для топ-3, топ-30 и фона) также для обычной и мобильной выкачки. В Яндексе значения для мобильной выкачки в среднем ниже, чем для десктопной, а в Google — часто даже выше, но везде с приближением к первой позиции размер HTML-кода растет.

Параметры	Корреляция*	Группы	Параметры	Корреляция*	Группы
Размер HTML-кода	1,000		Оплата	0,424	М
Количество слов в текстовых фрагментах	0,725	ВТ	Оплата картой	0,423	М
Количество исходящих внутренних ссылок	0,704	В	Вхождения в текст: слова запроса и синонимы	0,412	Т
Количество слов в исходящих внутренних ссылках	0,687	В	Цены на странице: наличие	0,411	М
Доля текста в HTML-коде	-0,683	СТ	Плотность стилистических проблем в тексте	-0,398	ТБ
Количество слов в тексте страницы	0,598	Т	Процент отказов	-0,398	П
Размер текста страницы	0,551	Т	Подбор по параметрам	0,388	М
«Классическая тошнота» текста	0,549	ТБ	Доля содержательного текста	0,385	ТБ
Количество картинок	0,545	В	Количество просмотренных страниц за одно посещение	0,384	П
Вхождения во фрагменты: слова запроса и синонимы	0,534	Т	Ассортимент	0,380	М
Цены на странице: количество	0,527	В	Доставка (авто в тегах)	0,357	М
Плотность повторов	0,524	ТБ	ИКС	0,348	РП
Доля SEO-текста в общем тексте	-0,513	Т	Доля ссылок в тексте	0,345	В
Вхождения в ссылки: слова запроса и синонимы	0,505	ВТ	SimilarWeb Country Rank	-0,338	Р
Найдено документов с сайта в Яндексе	0,500	РВ	Акции, спецпредложения, скидки (авто)	0,335	М
Количество страниц с сайта в индексе Яндекса	0,456	РВ	Продолжительность посещения	0,332	П
Концентрация слов запроса в SEO-тексте	-0,440	Т	Время загрузки страницы	0,316	СП
Количество слов в 	0,431	В	URL с https	0,303	С
Количество слов в атрибутах alt	0,426	В	Доля трафика: display	0,303	М
			Связь с Яндекс.Маркетом	0,298	М

Таблица 1. Бывают параметры, которые мало с чем коррелируют. **Размер HTML-кода**, наоборот, сильно коррелирует со многими параметрами. Чтобы в таблицу попали некоторые важные корреляции, нам пришлось пропустить много параметров, близких к уже встретившимся выше. Корреляции с разными параметрами могут быть связаны с разными аспектами ранжирования, которые в таблице отмечены буквами: **В** – размер витрины; **М** – коммерческие сайты, магазины; **С** – сложность кода, технологическая продвинутость сайта; **Т** – текстовое ранжирование, **Б** – риск «Баден-Бадена», возможные проблемы с контентом; **Р** – размер сайта и трафик; **П** – поведенческие метрики.

* Здесь и далее в подобных таблицах коэффициент корреляции – ранговый критерий Спирмена.

Критерии, которые отслеживают поисковики, могут быть формальными

Представители Яндекса и Google уже много лет говорят красивые слова о том, как важно делать сайты не для поисковых машин, а для людей, и именно это — самый прямой путь к высоким позициям в поиске. Чем дальше, тем больше в этом правды: факторы ранжирования всё более успешно *моделируют* полезность сайтов для человека. Тем не менее, скорее всего, и сейчас значительная часть факторов опирается на простые критерии — примерно такие же, как наши автоматические параметры.

Например, о наличии **отзывов** поисковая машина может судить по заголовкам, содержащим слово *отзывы*, а о наличии **видео** — по присутствию на странице кода видеоплеера. Вполне возможно, в этих конкретных случаях критерии проверки более сложные, — к этому нет препятствий, поскольку индексируемые страницы анализируются офлайн, и поисковики могут позволить себе использовать сложные алгоритмы проверки факторов ранжирования, в том числе и основанные на машинном обучении. Тем не менее, вряд ли тщательно анализируется

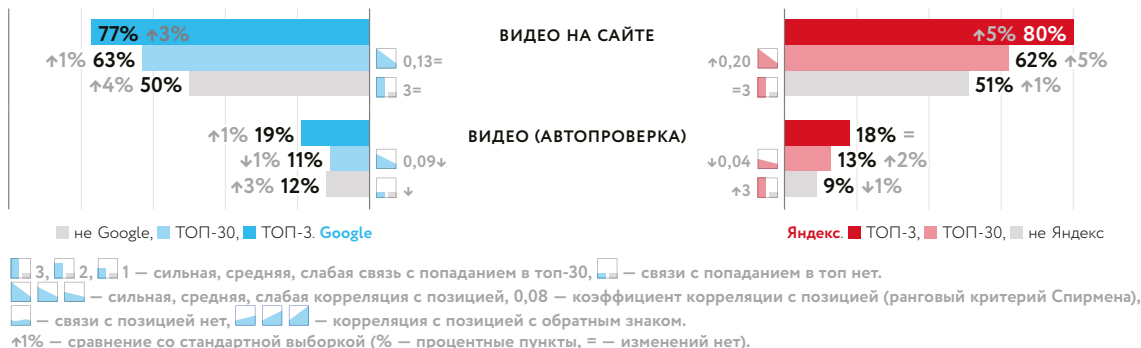
содержание отзывов (хотя теоретически и это возможно) или видеороликов.

Рассчитывать, что вы сумеете легко обмануть поисковик, не стоит, но и совсем не думать о том, как *быстро и дешево* улучшить параметры, по которым вы отстаете, — недальновидно. Мы, разумеется, не призываем вас повесить на сайт фейковый **телефон 8-800** (как это сделали многие года два назад). Но вот найти и вставить в подходящие места релевантные видео, раз уж поисковые машины так хотят их видеть, — точно полезно. Записать свои, вероятно, было бы еще полезнее — но и значительно сложнее, а потому не факт, что эффективнее..

Некоторые параметры могут иметь критическое значение

Если у вас ограниченный **ассортимент**, или нет **локализации в Москве**, или нет **цен**, то по многим запросам у вас просто нет шансов попасть в топ.

Такова ситуация с параметрами, которые учитываются на предварительных этапах ранжирования, (о чём можно догадаться по большому контрасту между значениями параметра в топ-30 и фоне) а также с теми,



у которых доля в топе приближается к ста процентам.

★ Если ваш сайт не в топе (т. е. вообще не ранжируется по запросу — именно тридцатка тут не принципиальна), стоит обратить внимание на параметры, сильно связанные с попаданием в топ, которых у него нет или по которым он заметно отстает от конкурентов.

«Особо важных» параметров особенно много для интернет-магазинов, см. ниже главу *Требования к функционалу*, стр. 70. В этом секторе очень сильна конкуренция, и если ваш сайт не попадет в топ, поисковые машины этого просто не заметят — у них нет недостатка в претендентах на ваше место. **Телефон** и **адрес** в регионе, из которого пришел запрос; как минимум **широкий ассортимент; описания товаров, их фотографии, технические характеристики, цены и кнопка заказа; описание условий оплаты, гарантии и доставки; акции и скидки** — без всего этого интернет-магазин или совсем

не может обойтись, или его шансы крайне низки.

Важно помнить, что в Яндексе факторы, отвечающие за текстовую релевантность, тоже относятся к критически значимым. У страницы, на которой мало слов запроса, есть все шансы остаться за бортом, даже если с другими факторами ранжирования всё замечательно. При этом «мало» — не значит совсем нет. Если у конкурентов **вхождений слов запроса** набирается на сорок запросов, а у вас на пять, вы сильно рискуете.

А у нас всё автоматизировано

Лаборатория поисковой аналитики компании «Ашманов и партнеры» занимается не только исследованиями, и это даже не главная часть нашей работы. Мы разработали практические инструменты для SEO-специалистов — прежде всего нашей компании.

Главный наш продукт предназначен для внутреннего пользования и в компании так и называется без изысков — «Лаборатория».



3, 2, 1 — сильная, средняя, слабая связь с попаданием в топ-30, — связи с попаданием в топ нет.
 3, 2, 1 — сильная, средняя, слабая корреляция с позицией, 0,08 — коэффициент корреляции с позицией (ранговый критерий Спирмена), — связи с позицией нет, — корреляция с позицией с обратным знаком.
 +1% — сравнение со стандартной выборкой (% — процентные пункты, = — изменений нет).

Он позволяет для той или иной выборки запросов собрать информацию, а затем построить нужный отчет для оптимизатора или черновик отчета для клиента (его потом доработает оптимизатор).

Мы сравниваем значения параметров у сайта и его конкурентов из топ-30 и выдаем подробные рекомендации, ранжированные

по важности. Значимость каждого параметра определяется не только его исходной оценкой (основанной на корреляции с позицией и связи с попаданием в топ-30), но и контекстом — тем, как параметр представлен в конкретной выкачке и как его значения у интересующего нас сайта соотносятся со значениями у конкурентов.

Скриншот интерфейса создания отчета. Вверху: название отчета «_conf'19 / Общая выкачка», дата «02.09.2019», город «Москва», тип анализа «Общий анализ», количество запросов «3». Меню: ОТЧЕТЫ, ИНФОРМАЦИЯ, ЗАПРОСЫ. Кнопка «Создание нового отчета».

Название: Отчет за 02.09.2019

Сайт: [выпадающий список]

Тип отчета: Отчет для оптимизаторов (выбрано). Другие варианты: Отчет для клиентов (HTML), Отчет для клиентов (MSWord XML), Задание копирайтерам (CSV), Задание копирайтерам (MSExcel XML).

Поисковая система: Yandex

Конкуренты (с галочками): mircli.ru, 220-volt.ru, mvideo.ru, e-katalog.ru, eco-conditioner.ru, kamindom.ru, kaminch.ru, yseinstrumenti.ru, intelclimate.ru.

Группы факторов (с галочками): Трафиковые метрики, Контакты, Продукт и его представление, Сайт и его наполнение, Связь с социальными сетями, Коммерческие факторы: авто, Сайт для роботов, Текстовые факторы, Качество текста.

Чекбокс: только указанные сайты.

Рис. 5. Карточка создания отчета в интерфейсе Лаборатории поисковой аналитики. Список конкурентов не влияет на рекомендации, но выбранные сайты показываются в таблицах отчета наряду с усредненными данными для топ-10 и топ-30.

ТЕКСТОВАЯ РЕЛЕВАНТНОСТЬ

Мы обычно заканчиваем наш обзор текстовыми факторами — но в этот раз решили начать именно с них. Так логичнее — сразу по многим причинам.

1. Именно с текстового ранжирования начался когда-то поиск, все остальные группы факторов добавились позже.
2. Именно с текстового ранжирования начинается процесс формирования поисковой выдачи. Первая операция — подъем страниц из индекса — прямо связана с текстом запроса (возможно, предварительно «переколдованным»). Если страниц отобрано много, следующим этапом может стать «прореживание» их списка, в котором (по крайней мере, в случае Яндекса) текстовое ранжирование, судя по всему, играет немалую роль.
3. Релевантность страницы запросу — необходимое условие качества поиска. Самая лучшая страница с идеального во всех отношениях сайта окажется бесполезной и только вызовет раздражение, если она недостаточно релевантна. Разумеется, релевантность к текстовым параметрам не сводится, но имеет к ним непосредственное отношение.
4. Текстовые факторы — одна из немногих групп, завязанных преимущественно на отдельные страницы. Ссылочные, социальные, трафиковые, коммерческие факторы, а также многие поведенческие и технические работают главным образом с сайтами, и уже поэтому вторичны.
5. Мы тоже работаем со страницами, поэтому про текстовые факторы знаем больше, чем про другие — нам просто удобнее их изучать.
6. Из наших 650 параметров примерно половина — текстовые.
7. Победив ссылочный спам, Яндекс принялся за текстовый — SEO-тексты, написанные специально для поисковых машин и в лучшем случае не слишком полезные для людей. Фильтру «Баден-Баден» уже два с половиной года, но Яндексу до победы пока далеко, и эта тема по-прежнему остается актуальной. И не только из-за риска попасть под фильтр — «выпадение» SEO-текстов из ранжирования нужно чем-то компенсировать.

Матрица текстовых параметров

Времена, когда релевантность мерилась количеством вхождений «ключа» в текст, давно ушли и вряд ли когда-нибудь вернуться.

Сейчас разработчики поисковых машин любят рассказывать, что релевантность страниц запросу всё равно оценивает искусственный интеллект — а значит, «считать слова» бессмысленно, и чтобы попасть в топ, нужно думать не о роботах, а о пользователях (а уж ИИ как-нибудь там разберется и вас за это наградит). Совет этот не лишен здравого смысла, но практической пользы от него мало. Так что считать слова всё-таки приходится — и это приносит плоды.

Мы не знаем, что и как учитывается при оценке релевантности текста запросу — и, скорее всего, Google и Яндекс делают это по-разному. Но, вооружившись несколькими сотнями параметров, мы можем понять, что важно, а что нет.

Параметров так много, потому что количество вхождений запроса в текст можно посчитать множеством разных способов: с учетом порядка слов или без, разрешая менять формы слов или не разрешая, считая слова вместе или по отдельности, с учетом длины текста (концентрация) или без учета, и так далее. А еще можно учитывать слова, которых в запросе нет — например, синонимы и слова, выделенные в снипетах.

И все эти способы можно приложить не только к тексту страницы в целом, но и к отдельным «зонам», которые принято считать важными — причем часть этих зон даже не входит в «чистый текст», выделяемый из HTML-кода. Это заголовки `<title>` и `<h1>`, заголовки в тексте, ссылки (внутренние и внешние), элементы списков, мета-теги `description` и `keywords`, атрибуты `alt` и `title` и др. А еще есть вхождения слов запроса в домен и в URL. В связи с наступлением «Баден-Бадена» очень актуальным стало деление страницы на «SEO-текст» (самый большой текстовый блок) и «текстовые фрагменты» — остающиеся за его пределами куски текста поменьше, вплоть до отдельных слов.

Получается матрица, в которой по одной оси разные способы оценки, а по другой — зоны страницы. Часть такой матрицы, заполненная для нашей выборки — на следующем развороте. В ней для параметров указаны корреляция с позицией и связь с попаданием в топ-30. Аналогичные таблицы для других выборок можно найти в наших отраслевых отчетах. Все они иногда меньше, иногда больше отличаются друг от друга — но в них много общего.

Принцип антибуквализма

Чем дальше от точного запроса, тем сильнее параметры.

В наших таблицах вхождением **точного запроса** (все слова запроса подряд в исходной форме) соответствует самый правый столбец; чем дальше влево, тем меньше ограничений. **Неточный запрос** — это все значимые слова запроса в произвольном порядке и в произвольной форме недалеко друг от друга (на отрезке не больше чем в два запроса). В **словах запроса** вхождения значимых слов запроса учитываются с весами, соответствующими их частотам¹¹, так, что каждое полное вхождение запроса (точное или неточное) дает единицу, но учитываются и изолированные употребления слов. Колонки **+ синонимы**¹² и **+ выделен-**

ные учитывают и слова, отсутствующие в запросе, но связанные с ним. Наконец, **синонимы** и **выделенные** (а также **выделенные Google**) — это только слова, которых в запросе нет.

В правой части таблицы больше пустых квадратов и даже отрицательных корреляций, а там, где есть положительные корреляции, слева они сильнее. Это подтверждает то, с чего мы начали: современные поисковые машины — не буквалисты, они не считают вхождения «ключевиков». И уж особенно в точной форме — ну кому в XXI веке еще может быть интересна эта точная форма!

Впрочем, есть одно исключение — т. е. кому-то всё-таки интересна. В Google видна сильная связь с попаданием в топ-30 для **запроса в точной форме** в **<title>**: 46% против 38% в фоне.

¹¹ Чем чаще встречается слово, тем меньше его вес. Например, в запросе *запись к педиатру через интернет* слово *запись* учитывается с весом 0,3, *педиатр* — 0,5, *интернет* — 0,2; веса предлогов *к* и *через* равны нулю.

¹² Понятие «синонимов» условное: сюда входят орфографические варианты и транслитерации (*Хендай* — *Хёнда* — *Хьюндай* — *Hyundai*), расшифровки аббревиатур, слова других частей речи, иногда даже антонимы (*купить* — *продажа*).

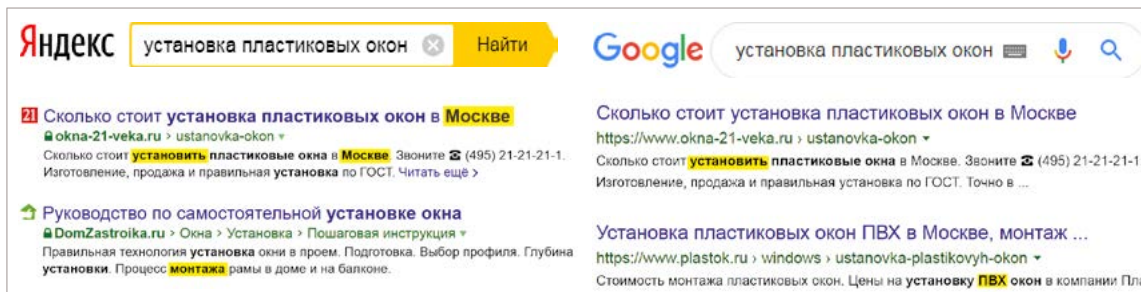


Рис. 6. По запросу *установка пластиковых окон* Яндекс дополнительно выделяет в снипетах глагол *установить* (в запросе — существительное *установка*) и город, из которого задан запрос (в данном случае *Москва*, но это может быть и, например, *Королев*), а также синоним *монтаж*. Google выделяет *установить* и *ПВХ* (контекстный синоним *пластиковых*).

ТАБЛИЦА ТЕКСТОВЫХ ПАРАМЕТРОВ. GOOGLE

	Выделенные	Слова запроса + выделенные	Синонимы	Слова запроса + синонимы	Слова запроса	Неточный запрос + синонимы	Неточный запрос	Запрос с точностью до форм	Точный запрос
Домен				0,06	0,05				
URL (кроме домена)				0,06	0,05				-0,04
<title>	0,20	0,11	0,09	0,10	0,09	0,09	0,04		-0,04
Текст страницы	0,18	0,19	0,06	0,19	0,19	0,10	0,07		
Текст страницы без SEO-текста	0,19	0,18	0,06	0,18	0,18	0,08	0,06		
SEO-текст	0,10	0,10	0,05	0,09	0,09	0,09	0,07	0,06	0,05
<h1>	0,09	0,05	0,04	0,04	0,04	0,05	0,02		-0,05
<h2>-<h4>	0,06	0,06		0,06	0,06	0,07	0,06	0,05	0,05
, 							-0,03	-0,04	
<i>, 					0,02	0,03	0,02		0,03
Внутренние ссылки	0,14	0,15	0,07	0,16	0,15	0,07	0,05		
Списки: 	0,10	0,09	0,05	0,10	0,09				-0,02
Атрибут alt	0,09	0,10	0,06	0,09	0,09	0,05	0,03		
Атрибут title	0,06	0,05	0,06						
Description	0,14	0,12	0,07	0,12	0,11	0,11	0,09	0,04	
Keywords	0,06					0,03	0,02		

– сильная, средняя, слабая связь с попаданием в топ-30.

– сильная, средняя, слабая корреляция с позицией. – корреляция с позицией с обратным знаком.

0,08 – коэффициент корреляции (ранговый критерий Спирмена).

– не видно связи с попаданием в топ-30 и/или корреляции с позицией.

ТАБЛИЦА ТЕКСТОВЫХ ПАРАМЕТРОВ. ЯНДЕКС

	Выделенные	Слова запроса + выделенные	Синонимы	Слова запроса + синонимы	Слова запроса	Неточный запрос + синонимы	Неточный запрос	Запрос с точностью до форм	Точный запрос
Домен									
URL (кроме домена)				0,03					
<title>	0,04					0,02	0,02	0,03	0,03
Текст страницы	0,10	0,11		0,09	0,10	0,06	0,05		0,03
Текст страницы без SEO-текста	0,10	0,12		0,11	0,11	0,07	0,06		0,04
SEO-текст			-0,03	-0,03	-0,03	-0,03	-0,03		
<h1>				0,02		0,03		0,02	
<h2>-<h4>									0,02
, 		-0,08	-0,04	-0,09	-0,09	-0,08	-0,09	-0,07	-0,05
<i>, 		-0,03		-0,05	-0,04	-0,03	-0,03	-0,02	
Внутренние ссылки	0,07	0,10		0,09	0,10	0,07	0,06	0,05	0,05
Списки: 	0,07	0,03		0,02	0,03				
Атрибут alt	0,02	0,06		0,06	0,06	0,06	0,05	0,04	0,04
Атрибут title									
Description	0,03	0,02	-0,03			0,02	0,02	0,03	0,03
Keywords		-0,07	-0,05	-0,07	-0,07	-0,05			-0,02

— сильная, средняя, слабая связь с попаданием в топ-30.

— сильная, средняя, слабая корреляция с позицией. — корреляция с позицией с обратным знаком.

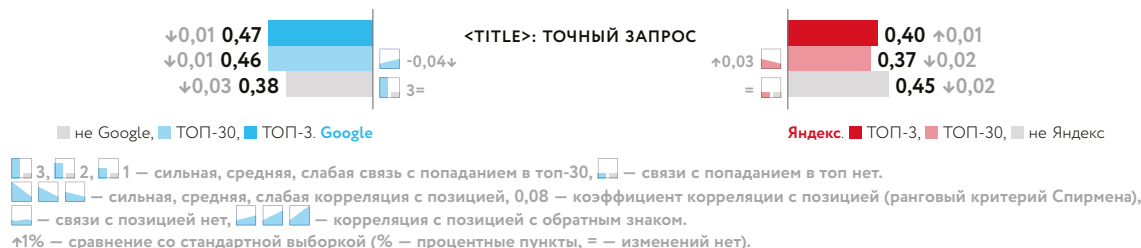
0,08 — коэффициент корреляции (ранговый критерий Спирмена).

— не видно связи с попаданием в топ-30 и/или корреляции с позицией.



В заголовке страницы запросы в точной форме лучше всё-таки включать — персонально для Google.

Связь с попаданием в топ есть и для других параметров с `<title>`, и для **точных запросов** в `<h1>`. Но таблицы для других выборок (и для этой за 2017–2018 годы) наводят на мысль, что это, скорее всего, следствие корреляций между этими параметрами и точным запросом в `<title>`.



Где текстовые факторы сильнее?

Первое, что бросается в глаза при сравнении таблиц для двух поисковиков: в Google преобладают корреляции с позицией, а в Яндексе — связь с попаданием в топ-30.

И это тот случай, когда связь с попаданием в топ важнее — Яндекс явно уделяет вхождению запроса в текст страницы больше внимания, чем Google. Это видно уже по тому, что для большинства параметров медиана для топ-30 Яндекса выше, чем для топ-3 Google!

Впрочем, для некоторых важных параметров корреляции с позицией (правда, более слабые, чем в Google) есть и в Яндексе — и за последний год они усилились. Возможны два объяснения такой картины.

1. Основной отбор по текстовой релеванности проходит на предварительном

этапе; страницы с недостаточным уровнем релевантности (с недостаточным количеством вхождений запроса) имеют меньше шансов попасть в топ — причем не в топ-30, а, условно говоря, в топ-100 или топ-1000. Напомним, что запросы, с которыми мы работаем, достаточно конкурентные, чтобы такой сценарий не казался невозможным: хорошо оптимизированных претендентов на попадание в топ хватает. При окончательном ранжировании уже среди отобранной тысячи результатов текстовые факторы учитываются с меньшим весом — Яндекс может положиться на то, что отобранные *страницы* достаточно релевантны,

и сосредоточиться на том, какой сайт вызывает больше доверия.

- Сильное ранжирование по релевантности затрагивает много сайтов — например, 100 или 1000. Благодаря этому наклон аппроксимирующей прямой¹³ оказывается меньше, и в топ-30 — а не только в топ-3 или топ-10 — попадают преимущественно страницы с большим количеством вхождений (слов) запроса. Сайты из фона в Яндексе ранжируются ниже — а значит, и количество вхождений на них в среднем должно быть ниже. (Эта гипотеза, правда, плохо применима к тем выборкам, где сильная связь с попаданием в топ есть, а корреляции с позицией внутри топ-30 не видно или она отрицательная.)

Чтобы понять, какое из этих объяснений ближе к истине, нужно проанализировать поисковую выдачу на большую глубину, чем наши стандартные 30 результатов. Но, по большому счету, главный вывод от этого не меняется: если на странице, которая продвигается по запросу, слов запроса недостаточно, у вашего сайта мало шансов попасть в топ.

СЛОВ ЗАПРОСА ДОЛЖНО БЫТЬ МНОГО

А достаточно слов запроса — это сколько? Если брать медианные значения (стремиться быть не хуже хотя бы половины сайтов, попавших в соответствующую группу), то мы получаем следующие ориентиры для топ-3 Яндекса: **точных** вхождений запроса — почти 4,5; **с точностью до форм слов** — 8,75; **неточных** (слова запроса недалеко друг от друга) — 12,5; то же **с учетом синонимов** — больше 18.

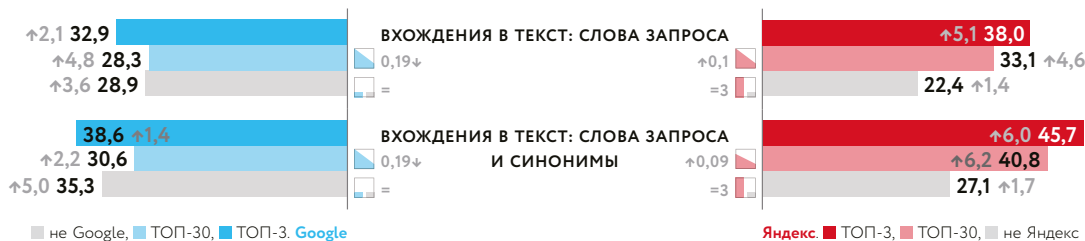


Слов запроса в топ-3 Яндекса набирается уже на 38 запросов, и даже в топ-30 — на 33. А **с учетом синонимов** — на 45 и 40 соответственно.

Разумеется, по каким-то запросам может быть меньше, но по каким-то и больше — это тот случай, когда нужно сравнивать себя с конкурентами для каждого запроса по отдельности. Приведенные выше показатели — средние по нашей политематической выборке из 160 запросов.

В Google числа немного меньше, но тоже впечатляющие. А главное, значения текстовых параметров продолжают расти — за последний год они выросли еще процентов на 10–20.

¹³ Ср., например, наклонные прямые на рис. 4, стр. 26



■ не Google, ■ Топ-30, ■ Топ-3. Google
 ■ Яндекс, ■ Топ-3, ■ Топ-30, ■ не Яндекс

3, 2, 1 — сильная, средняя, слабая связь с попаданием в топ-30, □ — связи с попаданием в топ нет.
 3, 2, 1 — сильная, средняя, слабая корреляция с позицией, 0,08 — коэффициент корреляции с позицией (ранговый критерий Спирмена), □ — связи с позицией нет, 3, 2, 1 — корреляция с позицией с обратным знаком.
 ↑1% — сравнение со стандартной выборкой (% — процентные пункты, = — изменений нет).

ДЛИНА И КОНЦЕНТРАЦИЯ



Чтобы на странице поместился десяток вхождений запроса (пусть даже неточных) и еще запросов на тридцать отдельных слов россыпью, эта страница должна быть очень большой.

И действительно, **размер HTML-кода** (ситуацию с ним мы подробно разобрали на стр. 25–27), **размер текста страницы, количество слов** на ней ведут себя примерно так же, как количество вхождений слов запроса. В Google сильная корреляция с позицией, в Яндексе — сильная связь с попаданием в топ-30 и достаточно сильная (сильнее, чем год назад; слабее, чем в Google) корреляция с позицией. Медианные значения этих параметров для топ-30 Яндекса — лишь немного ниже, чем для топ-3 Google.

Если же посмотреть на **концентрацию слов запроса** в тексте страницы, то оказывается, что их количество на 100 слов текста не растет, а скорее падает, и ни корреляции с позицией, ни связи с попаданием в топ нигде не видно. (В других зонах ситуация может быть иной, см. таблицу на стр. 45).



ВИТРИНА КАК ИНСТРУМЕНТ ТЕКСТОВОЙ ОПТИМИЗАЦИИ

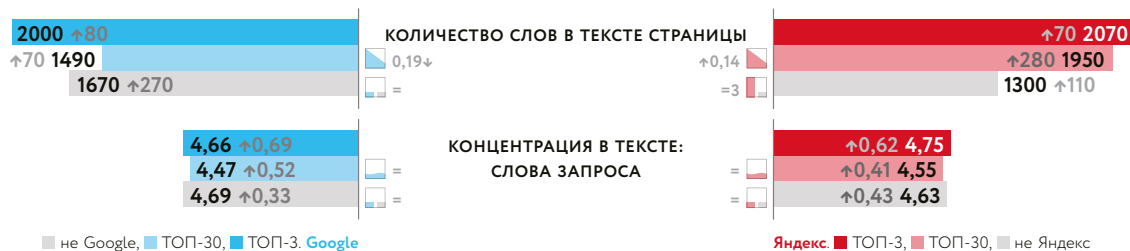
Количество вхождений слов запроса в текст страницы сильно коррелирует с параметрами, которые связаны с размером витрины — например, с количеством **слов в исходящих внутренних ссылках, слов в , картинок, цен на странице**.



Товарная витрина (или ее аналог на сайтах недвижимости, агрегаторах услуг и т. п.) — самое удобное место для того, чтобы наполнить страницу словами запроса.

Здесь создатели сайта меньше ограничены нормами языка, а роботы меньше нацелены на то, чтобы уличить их в излишней оптимизации. Слова запроса и даже запросы целиком здесь выглядят органично. А главное, витрина может содержать и 30, и 50, и 100 карточек, и чтобы еще больше увеличить релевантность страницы, к названиям товаров можно добавить их краткие описания, в которых словам запроса тоже самое место.

Не удивительно, что там, где есть ярко выраженные витрины — в электронной коммерции — количество вхождений запроса ока-



■ 3, ■ 2, ■ 1 — сильная, средняя, слабая связь с попаданием в топ-30, ■ — связи с попаданием в топ нет.
 ■ — сильная, средняя, слабая корреляция с позицией, 0,08 — коэффициент корреляции с позицией (ранговый критерий Спирмена),
 ■ — связи с позицией нет, ■ ■ ■ — корреляция с позицией с обратным знаком.
 ↑1% — сравнение со стандартной выборкой (% — процентные пункты, = — изменений нет).

залось максимальным (так, слов запроса в топ-30 Яндекса хватило бы на 42,5 запроса). В недвижимости значения параметров ниже (28 запросов рассыпью), в медицине и финансах — еще ниже (23,4 и 18,9 соответственно, в основном благодаря агрегаторам),

и наконец, в автомобилях — совсем мало (меньше 8 запросов рассыпью на странице)¹⁴.

¹⁴ Сравнение выборок по этому параметру не вполне корректно — прежде всего потому, что количество вхождений слов запроса в текст сильно зависит от длины запросов и от остроты конкуренции по нему. Отраслевые выборки больше, и в них больше длинных и менее конкурентных запросов.



SEO-тексты и «Баден-Баден»

Мы уже выяснили, что лучшее место для слов запроса — товарная витрина. Худшее место — это, конечно, SEO-текст.

SEO-текст — это блок текста, стоящий над витриной или под ней (реже — и там, и там) и объясняющий — как будто бы посетителям, но на самом деле поисковикам, — про что эта страница и что за товары или услуги на ней предлагаются. Яндекс объявил таким текстам войну, введя фильтр «Баден-Баден». И действительно, есть за что: тексты там часто написаны не совсем для людей — и дело тут не только в ключевиках. Выросло целое поколение копирайтеров, которые готовы писать тексты на любую тему по строгим ТЗ — столько-то тысяч знаков, вхождений ключевиков столько-то, тошнота¹⁵ не более, уникальность не менее... Содержание при этом — какое получится, обычно крайне шаблонное (впрочем, нешаблонная лирика часто еще хуже). Некоторые характерные образцы стиля SEO-текстов — во врезке на следующей странице.

Технически SEO-текст в наших параметрах — просто самый большой блок текста на странице.

Мы выделяем его при помощи довольно простого алгоритма, который, конечно, не всегда срабатывает оптимально. Тем не менее, для интернет-магазинов, агрегаторов и многих других типов коммерческих сайтов это обычно действительно именно SEO-текст. Поскольку мы работаем со статистикой, нам этого достаточно: идеальное выделение SEO-текстов дало бы лучшие результаты, но и имеющиеся достаточно красноречивы.

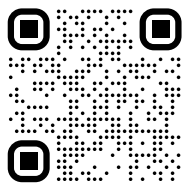
«ТУРГЕНЕВ», ИЛИ КАК НЕ УЕХАТЬ В «БАДЕН-БАДЕН»

Под фильтр «Баден-Баден» попадают преимущественно коммерческие сайты — обычно такие, на которых есть явные SEO-тексты. Аналогичные проблемы на информационных сайтах Яндекс пока текстовым спамом не считает, и под санкции они обычно не попадают. В отличие от более старых фильтров — *переспама* и *переоптимизации*, — «Б-Б» реагирует не только на манипуляции со словами запросов, но и на общее качество и даже размер текста («портянки» и «простыни»).

¹⁵ Для незнакомых с SEO-терминологией поясним, что «классическая» и «академическая тошнота» — параметры, оценивающие количество повторов в тексте.

Яндекс, как обычно, не объясняет подробно, по каким критериям тексты признаются плохими. Возможно, явных критериев и не существует: просто были собраны образцы качественных и некачественных страниц¹⁶, а затем нейронная сеть «сама» научилась отличать одни от других.

А теперь позволим себе чуть-чуть рекламы. **Сервис «Тургенев»** (turgenev.ashmanov.com), запущенный нами два года назад, не только оценивает риск попадания текста под фильтр, но и объясняет, что в нём не так, раскладывая всё по нескольким полочкам: повторы, стилистика, водность, запросы, удобочитаемость. Даже если прямого риска «Б-Б» нет, оценки и подсказки «Тургенева» помогут увидеть недостатки текста и сделать его лучше.



«Тургенев»

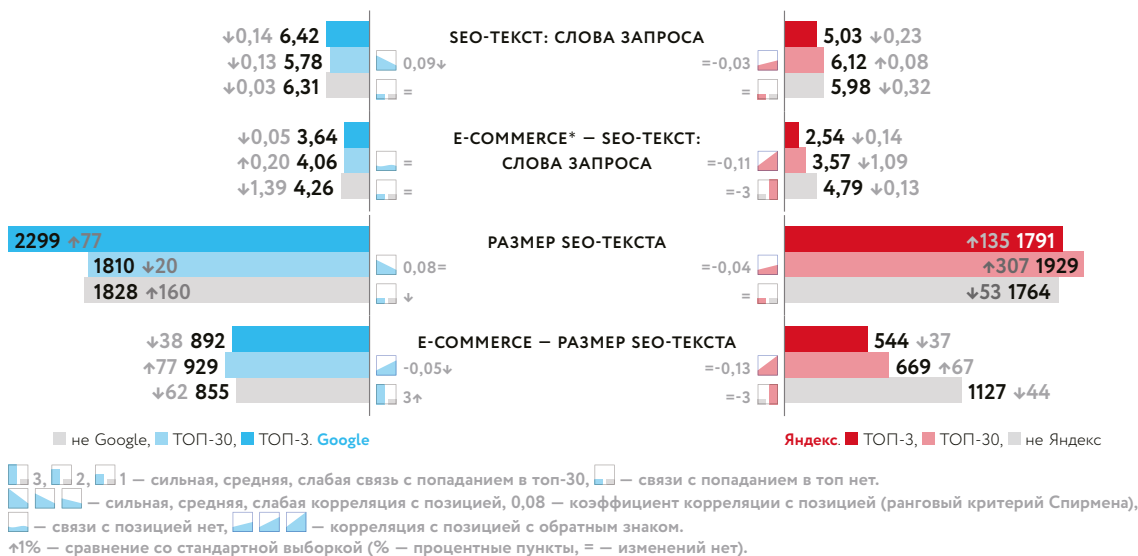
¹⁶ Чтобы ассессоры могли оценить страницы, из которых потом формируются обучающие выборки, какие-то критерии нужны. Но они не обязательно должны быть формальными.

Новая версия «Тургенева» умеет проверять не только тексты — можно загрузить список URL и найти наиболее проблемные страницы сайта. Предусмотрена возможность доступа по API — тексты или страницы можно будет отправлять на проверку прямо из той среды, в которой вы привыкли с ними работать. Правда, новая версия становится платной, хотя некоторые бесплатные возможности в ней остаются.

КАК SEO-ТЕКСТЫ ВЛИЯЮТ НА РАНЖИРОВАНИЕ

Давайте теперь абстрагируемся от риска, который может быть связан с SEO-текстами, и попробуем понять, как они влияют на ранжирование.

Скажем осторожно: наши таблицы не показывают их влияния на ранжирование в Яндексе — а вот для текстовых фрагментов



* E-commerce: электронная коммерция, выкачка за сентябрь 2019; сравнение с выкачкой за март 2019.

ОБРАЗЦЫ СТИЛЯ SEO-ТЕКСТОВ

Хороший головной убор не только спасёт в холодную погоду, но и подчеркнёт стиль и индивидуальность его владельца, поэтому чтобы купить шапку, подходящую именно вам, следует ознакомиться со всеми характеристиками изделия.

В процессе эксплуатации пластиковые перила из ПВХ не принесут никаких хлопот своим хозяевам, даже наоборот. ... Никто не поспорит, что пластиковые перила, несмотря на всю свою простоту и относительную дешевизну, являются очень прочными и долговечными изделиями, способными служить своему хозяину не один год.

ЭКО может иметь различную стоимость, которая зависит от множества факторов.

Поэтому, лучше всего довериться уже готовым, профессионально разработанным наборам для вышивания на скатертях. ... Украшая повседневный быт, эти изделия невольно напоминают о тесных связях, которые соединяют человека с его родом и предками.

Отзывы покупателей показывают: необязательно покупать посуду в офлайн-магазине. ... Условия для любимых клиентов при продаже посуды недорогого

Кухонные аксессуары признаны одними из полезных подарков, который не будет пылиться в шкафу или будет передарен. Неужели не найдется применение современной сковородке с технологичным покрытием? Или в дальний угол будет закинута термкружка? Или на столе не найдется места сахарнице с оригинальным дизайном?

Если интересна аренда склада в Подмоскowie, то вы можете снимать склад класса А. ... Стоит заметить то, что качество не уступает складам, находящимся в пределах МКАД, которая находится в чуть более 50 км. При этом стоимость у нас значительно ниже, что несомненно скажется на вашей прибыли.

Шампунь без сульфатов и парабенов — это натуральный продукт, состоящий из экстрак-

тов и масел растений. Они имеют витамины и химические соединения, влияющие на здоровье благоприятным образом.

Засыпая по вечерам и просыпаясь по утрам, ребенок чаще всего смотрит на потолок.

В этом разделе сайта есть возможность снять одноэтажный домик или настоящий особняк в два или три этажа, есть коттеджи неподалеку от моря станции Благовещенская.

Тарифных планов у компании множество, каждый из них способен удовлетворить группу населения.

В первую очередь внимание маленьких покупательниц и их родителей привлекают элегантные маечки и трусы для девочек, к которым предъявляются исключительно высокие требования.

Сделать все тайное явным, поможет анализ на скрытые инфекции.

Фасады из кирпича считаются одними из самых прочных, долговечных и эстетичных строительных конструкций, но и они со временем подвергаются разрушениям, покрываясь трещинами, сколами, прогибами.

По этой причине женщина, которая следит за своим здоровьем, должна быть осведомлена, эндометриоз — что это такое, и как проявляется этот недуг.

Только коттедж из дерева является экологически чистым. Именно поэтому данный материал выбирают для загородной недвижимости люди, заботящиеся о собственном здоровье.

Так, зонты для кафе приятно удивят современными цветовыми вариациями, стилистическими решениями и низкой стоимостью.

Дизайн кухни прямоугольной должен получиться гармоничным, многофункциональным. ... Выбранная цветовая палитра не должна раздражать. Можно подобрать оттенки, которые будут пробуждать аппетит.

(вся страница за вычетом SEO-текста), наоборот, текстовые параметры сильнее, чем для страницы в целом. В Google корреляции с позицией есть, но они заметно слабее, чем для текста страницы в целом.

Количество **вхождений слов запроса в SEO-текст**, как и значения других подобных параметров, при приближении к первой позиции в Яндексе падает (хотя и не очень сильно). Но утверждать, что Яндекс как-то специально игнорирует SEO-тексты, мы, пожалуй, не можем. Дело в том, что параметры этой группы очень сильно коррелируют с **размером SEO-текста**, а он тоже падает.



На других выборках для вхождений в SEO-текст в Яндексе могут наблюдаться сильные отрицательные корреляции (**e-commerce**), а могут и положительные (**автомобили**) — и всегда это определяется тем, падает или растет размер SEO-текста при приближении к топ-1.

Для нас особенно важна выборка по электронной коммерции — это как раз те сайты,

по которым сильнее всего ударил «Баден-Баден». Здесь в Яндексе и для размера SEO-текста, и для вхождений в него слов запроса (и других подобных параметров) видна сильная отрицательная корреляция с позицией и сильная отрицательная связь с попаданием в топ-30; в Google картина неопределенная.



Там, где реально применяется «Баден-Баден», SEO-тексты в Яндексе начинают мешать ранжированию — чем их меньше, тем лучше.

При этом выпадающие вхождения слов, важных для ранжирования, необходимо компенсировать за счет других частей страницы — прежде всего, витрины.

О параметрах, из которых складывается риск «Б-Б», — **стилистические проблемы, плотность повторов, «тошнота»** и др. — см. в справочном разделе к «Тургеневу»¹⁷.

¹⁷ См. turgenev.ashmanov.com/?h=vkladki.

Зоны документа

До сих пор мы говорили почти исключительно о вхождениях в текст страницы и в SEO-текст. А что с другими зонами?

ЗАГОЛОВКИ ТЕКСТА

Начнем, конечно, с **заголовка <title>**. Мы уже упоминали, что у Google к нему особое отношение: вхождения точного

запроса в **<title>** сильно связаны с попаданием в топ-30, что в Google редкость и к тому же нарушает *принцип антибуквализма* (стр. 33). Это особое отношение распространяется и на другие параметры

(вхождения **запроса с точностью до словоформ, неточного запроса, отдельных слов запроса**, в том числе **с учетом синонимов**). Везде значения для топ-30 Google не ниже, чем в топ-3 Яндекс — впрочем, поскольку в Яндексе корреляций с ранжированием нет или они слабые, это почти то же самое, что в топ-30.



Можно констатировать, что Google уделяет заголовкам `<title>` больше внимания, чем Яндекс.

За последний год это усилилось (не за счет роста значений параметров в Google, а за счет небольшого падения в Яндексе). В «тени» Google связи с ранжированием в Яндексе по многим параметрам не видно — но это не значит, что ее нет. Вхождений слов запроса в `<title>` в топ-30 Яндекс набирается в среднем почти на 1,1 запроса. На некоторых других выборках в Яндексе есть и связь с попаданием в топ, и корреляция с позицией (для автомобилей даже достаточно сильная).

Хотя для пользователей обычно именно `<h1>` — «настоящий» заголовок страницы, в SEO он — двойник `<title>` и светит его

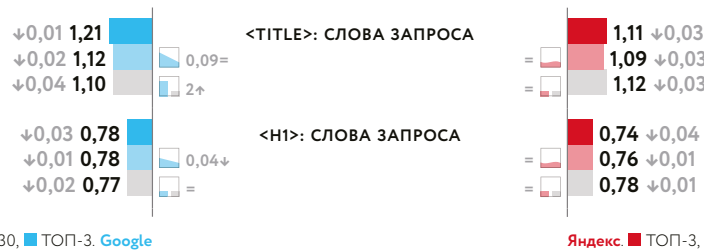
отраженным светом. Аналогичные параметры (например, расстояние до запроса) для `<title>` и `<h1>` обычно сильно коррелируют между собой.

При этом `<title>` ничто не мешает быть **длинным**: увидеть его можно разве что в шапке окна и в результатах поиска, и то часто не целиком. А `<h1>` обычно короче (медианная длина всего около 27 символов) и красуется на самом видном месте. Втиснуть в него все запросы, по которым продвигается страница, затруднительно — в среднем из **отдельных слов** набирается лишь чуть больше трёх четвертей запроса (в Яндексе 76%, в Google 78; с учетом синонимов 88–89%).

Нет сомнений, что заголовки `<h1>` на подавляющем большинстве коммерческих сайтов составлялись с оглядкой на поисковое продвижение. Но насколько сильно они в действительности на него влияют, понять не слишком просто. Судя по всему, *не слишком* сильно. В Google есть связь с попаданием в топ-30 для **точного запроса** — но это может быть следствием аналогичной связи для `<title>`¹⁸. Корреляции с позицией для большинства параметров тоже есть, но они заметно слабее, чем для `<title>` (и в прин-



¹⁸ Чем ближе к точному запросу, тем сильнее корреляции между соответствующими параметрами для `<title>` и `<h1>`.



■ не Google, ■ Топ-30, ■ Топ-3. Google
■ Яндекс, ■ Топ-3, ■ Топ-30, ■ не Яндекс

3, 2, 1 — сильная, средняя, слабая связь с попаданием в топ-30, □ — связи с попаданием в топ нет.
 3, 2, 1 — сильная, средняя, слабая корреляция с позицией, 0,08 — коэффициент корреляции с позицией (ранговый критерий Спирмена),
 □ — связи с позицией нет, □ — корреляция с позицией с обратным знаком.
 ↑1% — сравнение со стандартной выборкой (% — процентные пункты, = — изменений нет).



ципе тоже могут объясняться его влиянием). Ожидаемая связь с попаданием в топ-30 в Яндексе для некоторых параметров есть, но не для всех, и часто слабая — впрочем, на других выборках, особенно для недвижимости и медицины, она сильнее.

МЕТАТЕГИ

Параметры, связанные с **meta description**, включая наличие этого метатега и его длину, в целом ведут себя примерно так же, как многие другие группы текстовых параметров. В Яндексе связь с попаданием в топ-30 (иногда также слабая корреляция с позицией), в Google — корреляция с позицией; всё это с соблюдением принципа антибуквализма. Впрочем, конкретно со **словами запроса** (без синонимов и выделенных) и конкретно для нашей общей выборки в Яндексе что-то пошло не так: связи с попаданием в топ уже второй год не видно.

Интересно, что, хотя **description** обычно намного длиннее, чем **<title>**,¹⁹ разница по количеству вхождений (слов) запроса между ними совсем

¹⁹ Для топ-30 Яндекса и Google на нашей выборке — около 150 и менее 75 символов соответственно. Как всегда в таких случаях для числовых параметров с большим диапазоном значений, мы считаем медиану отдельно для каждого запроса, а затем берем от них среднее.

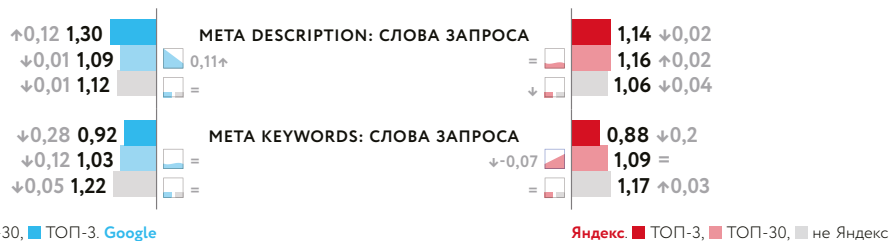
небольшая, см. диаграммы. Возможно, наполнение **description** «ключевиками» сдерживается тем, что этот метатег оптимизируют под показ в результатах поиска в качестве снипета: в этом случае он должен аккуратно дополнять заголовок, а не повторять его.

Для **метатега keywords** сейчас связи с ранжированием обычно не видны (или видны отрицательная) — хотя для отдельных выборок (особенно сильно монополизированных) корреляции с позицией возможны, а в 2017 году они были видны и на общей выборке.

ДОМЕН И URL

Для **вхождений в URL** и **доменное имя**, как ни странно, действуют те же закономерности, что и для других зон: в Яндексе связь с попаданием в топ-30, в Google корреляция с позицией, в обеих поисковых машинах — принцип антибуквализма (для URL).

Даже **количество слов запроса** в URL вполне сравнимо с их количеством, например, в **<h1>** (из которого, впрочем, «человекопонятные URL» часто и строятся). В топ-30 Яндекса их в среднем набирается на 0,6 запроса, а с учетом синонимов — на 0,67, больше двух третей! И это при том, что есть глав-



■ 3, ■ 2, ■ 1 — сильная, средняя, слабая связь с попаданием в топ-30, ■ — связи с попаданием в топ нет.
 ■■■ — сильная, средняя, слабая корреляция с позицией, 0,08 — коэффициент корреляции с позицией (ранговый критерий Спирмена),
 ■■■ — связи с позицией нет, ■■■ — корреляция с позицией с обратным знаком.
 ↑1% — сравнение со стандартной выборкой (% — процентные пункты, = — изменений нет).

ные страницы, сайты с не «говорящими» URL и др.

Количество вхождений слов запроса в доменное имя гораздо скромнее — в топ-30 Яндекса их набирается на 0,07 запроса, а с учетом синонимов — на 0,11. Но если вдуматься, это тоже очень много!

7% от 30 результатов — это уже два сайта с «говорящим» доменом в топе. Но мы считаем не в словах, а в запросах, которые обычно состоят из двух и более слов; значит говорящих доменов еще больше. С учетом синонимов выходит в среднем порядка пяти доменов со словами запроса из тридцати. Но и это еще не всё. Среди главных страниц сайтов, попадающих в результаты поиска, слов запроса в доменных именах в топ-30 Яндекса набирается на 0,22 запроса, а в топ-3 — на 0,25; с учетом синонимов — 0,26 и 0,3, соответственно!



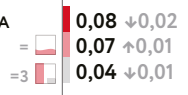
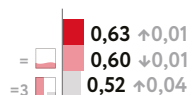
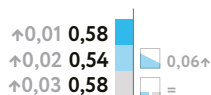
Получается, что доменное имя для многих сайтов — единственный шанс попасть в результаты поиска по конкурентным запросам, на которые они иначе не могли бы претендовать.

Особенно в Яндексе — в Google средние значения ниже (топ-3 Google не всегда дотягивается до топ-30 Яндекса). Примеры

КОНЦЕНТРАЦИЯ СЛОВ ЗАПРОСА ПО ЗОНАМ

Почти во всех рассматриваемых зонах, кроме списков и курсива, концентрация слов запроса выше, чем в тексте, — иногда сильно выше.

Зона	Google		Яндекс	
	Топ-3	Топ-30	Топ-3	Топ-30
Текст	4,27%	4,00%	4,35%	4,04%
SEO-текст	4,95%	4,65%	4,20%	4,44%
<title>	27,3%	28,3%	25,0%	26,0%
<h1>	56,6%	52,9%	55,6%	54,5%
<h2>-<h4>	16,1%	13,7%	15,2%	15,5%
, 	8,11%	7,52%	5,26%	7,06%
<i>, 	4,03%	4,80%	2,41%	4,27%
Внутренние ссылки	5,73%	5,38%	6,81%	6,05%
Списки: 	4,25%	3,65%	3,69%	3,48%
Атрибут alt	11,4%	9,5%	13,4%	10,3%
Атрибут title	7,12%	5,12%	6,01%	5,07%
Description	13,8%	14,1%	12,6%	13,2%
Keywords	40,0%	38,1%	38,3%	39,2%



■ не Google, ■ Топ-30, ■ Топ-3. Google

■ Яндекс, ■ Топ-3, ■ Топ-30, ■ не Яндекс

3, 2, 1 — сильная, средняя, слабая связь с попаданием в топ-30, □ — связи с попаданием в топ нет.
 □ — сильная, средняя, слабая корреляция с позицией, 0,08 — коэффициент корреляции с позицией (ранговый критерий Спирмена),
 □ — связи с позицией нет, □ — корреляция с позицией с обратным знаком.
 ↑1% — сравнение со стандартной выборкой (% — процентные пункты, = — изменений нет).



сайтов, использовавших такой шанс, можно найти в отчете по недвижимости.

ВЫДЕЛЕНИЯ В ТЕКСТЕ, ССЫЛКИ И СПИСКИ

Нам осталось рассмотреть разнообразные зоны внутри текста — внутренние заголовки, выделения полужирным шрифтом и курсивом, пункты списков, ссылки. А также атрибуты `alt` и `title`, которые в состав текста не входят, поскольку прячутся внутри тегов иллюстраций, ссылок и т. п., но ведут себя похоже.



Для многих из этих зон (не для всех) видна сильная связь с ранжированием; как обычно, в Яндексе преобладает связь с попаданием в топ-30, в Google — с позицией; соблюдается принцип антибуквализма. Тем не менее, на вопрос о том, имеют ли эти зоны самостоятельное значение (т. е. имеют ли слова запроса, попавшие в них, больший вес, чем просто при вхождении в текст страницы), ответить не всегда просто.

Количество вхождений слов запроса в ту или иную зону на странице равно произведению общего количества слов в этой зоне, умноженному на концентрацию слов запроса в этой зоне (мы считаем ее в процентах). Если значения параметра, оценивающего количество вхождений слов запроса в интересующую нас зону, при приближении к первой позиции растет, важно понять, за счет чего происходит этот рост. Если за счет роста концентрации, то, по-видимому, поисковая машина отдает предпочтение страницам, в которых слова запроса чаще попадают в эту зону. Если за счет увеличения размера зоны, то, возможно, количество вхождений слов запроса в зону для поисковой машины роли не играет, а рост параметра «наведен» ростом размера зоны.

Правда, по этой логике рост количества вхождений слов запроса в текст страницы можно свести к росту размера этого текста — что представляется сомнительным.

Для многих параметров, связанных с попаданием (слов) запроса во **внутренние заголовки** `<h2>`–`<h4>`, на нашей выборке в Яндексе есть связь с попаданием в топ-30, в Google — корреляции с позицией. Эти корреляции не очень сильные, но возникают они в основном за счет роста концентрации (на общей выборке количество слов в заголовках если и растет с приближением к топу, то слабо; в Яндексе тоже концентрация слов запроса в заголовках выше, чем в фоне).

Правда, для электронной коммерции и недвижимости ни в Google, ни в Яндексе связи с ранжированием не видно — зато для медицины, финансов и автомобилей в обоих поисковиках есть сильная корреляция с позицией. В разных тематиках разная ситуация с объемом зоны: для первых двух количество слов в `<h2>`–`<h4>` с приближением к топ-1 скорее падает, а для остальных — растет. В e-commerce и недвижимости основной вклад в текстовую релевантность дает витрина, а заголовки «ютятся» в SEO-тексте, объем которого ближе к топу падает не только в Яндексе, но и в Google.

Более однозначная ситуация с выделениями в тексте — **bold** (``, ``) и **italic** (`<i>`, ``). Связи с ранжированием для них просто не видно²⁰, а концентрация с приближением к топу падает (особенно в Яндексе).

Вхождения слов запроса и т. п. в **исходящие внутренние ссылки** (и **исходящие ссылки** вообще) на большинстве выборок сильно

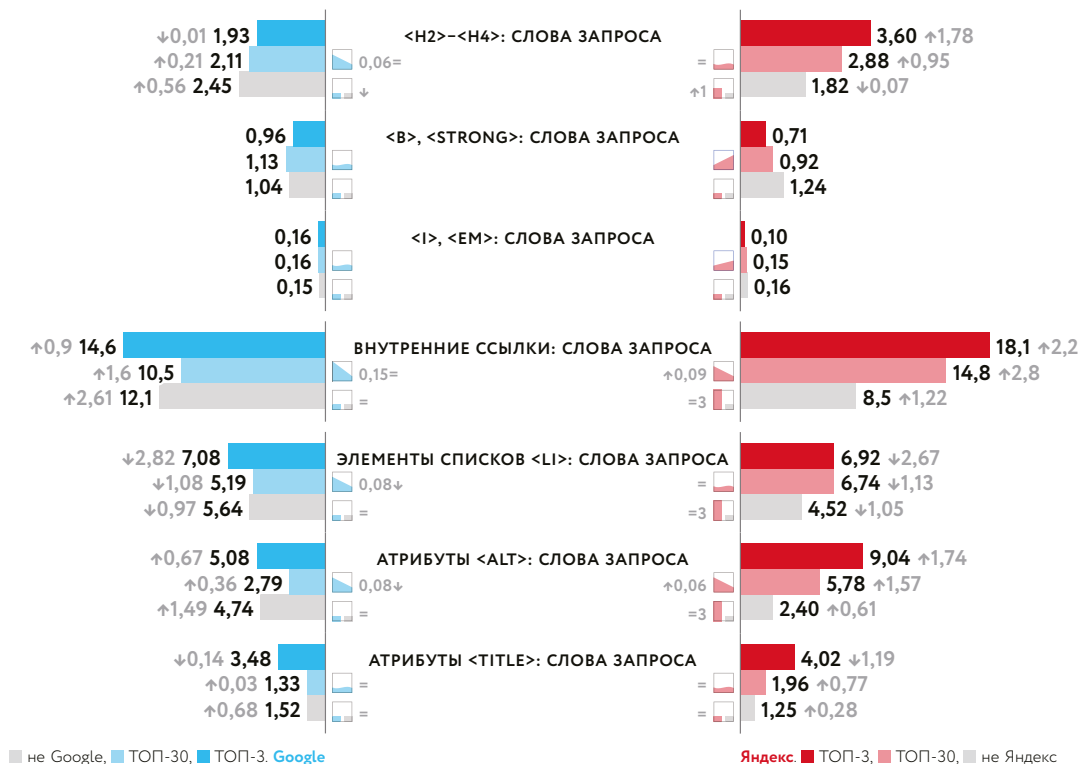
²⁰ Кроме автомобилей и финансов, где корреляции с позицией для bold, по-видимому, связаны с объемом зоны. Мы должны извиниться перед читателями: у нас была ошибка в коде, выделяющем эти зоны, и сведения о них в предыдущих отчетах не верны.

связаны с попаданием в топ-30 в Яндексе и сильно коррелируют с позицией в обеих поисковых машинах (в Google обычно немного сильнее, чем в Яндексе). К тому же в Google при приближении к топ-1 растет концентрация слов запроса во внутренних ссылках. Тем не менее, утверждать, что поисковые машины реагируют именно на слова запроса в ссылках, мы пока не можем. Дело в том, что **количество слов в ссылках** или их **длина** — еще более сильные параметры, чем количество вхождений в них слов запроса: корреляции с ранжированием там

еще сильнее. А они, в свою очередь, связаны с **количеством на странице ссылок**, и через это — с размером витрины и другими важными параметрами (например, количеством в индексе страниц с сайта, отвечающих запросу).



В любом случае, понятно, что «хорошая» с точки зрения поисковых машин страница должна содержать много вхождений запроса (или слов из него) во внутренние ссылки.



■ не Google, ■ Топ-30, ■ Топ-3, ■ Google
 ■ Яндекс, ■ Топ-3, ■ Топ-30, ■ не Яндекс

3, 2, 1 — сильная, средняя, слабая связь с попаданием в топ-30, — связи с попаданием в топ нет.
 — сильная, средняя, слабая корреляция с позицией, 0,08 — коэффициент корреляции с позицией (ранговый критерий Спирмена),
 — связи с позицией нет, — корреляция с позицией с обратным знаком.
 ↑1% — сравнение со стандартной выборкой (% — процентные пункты, = — изменений нет).

Если ваш сайт отстает от конкурентов по этому параметру, нужно разбираться, почему так получилось, и исправлять положение. При этом не так уж и важно, какие параметры непосредственно участвуют в формуле ранжирования, а какие «работают» только за счет корреляций. «Потянув» за любой из них, вы, скорее всего, повлияете и на остальные.

Внешне похожая, но всё-таки немного другая ситуация и с **элементами списков **. Здесь в Яндексе есть сильная связь с попаданием в топ-30, в Google — корреляция с позицией. Но связь с ранжированием не слишком устойчива: для автомобилей не видно связи с попаданием в топ в Яндексе, а для сильно монополизированной недвижимости — и корреляции с позицией в Google. Похоже, высокие значения параметров этой группы в Яндексе и в топ-10 Google определяются почти исключительно объемом зоны — а концентрация слов запроса здесь даже меньше, чем по тексту страницы в целом. Поэтому маловероятно, что поисковые машины придают вхождениям в элементы списка какое-то особое значение — скорее всего, связь с ранжированием здесь наведена другими параметрами, коррелирующими с данным (прежде всего, размером витрины).

Нам осталось рассмотреть две зоны, не учитываемые при выделении текста страницы — текст, скрытый внутри тегов в **атрибутах alt** (обычно в тегах иллюстраций) и **title**

(в тегах ссылок, элементов таблиц, иллюстраций и др.).

Мы, к сожалению, не отслеживали специально, из каких тегов приходят **alt** и **title**. Длина **alt** очень сильно коррелирует с количеством картинок и значительно слабее, но тоже сильно — с количеством ссылок (ссылки часто сопутствуют картинкам — например, на карточках товаров в витрине). Длина **title** коррелирует с количеством ссылок сильнее, чем с количеством картинок, — но всё-таки слабее, чем с ними коррелирует длина **alt**.

У двух атрибутов очень похожие распределения по длине, по-видимому, связанные с распределением тегов, в которых они используются, и вроде бы достаточные для того, чтобы обе группы параметра оказались «сильными» подобно, например, вхождениям в ****. Но **alt** «работает», а **title** — нет.

Поисковые машины собирают тексты в **alt** для поиска по изображениям — похоже, наши данные свидетельствуют о том, что они (в отличие от текстов в **title**) важны и для обычного поиска.

Некоторая логика в том, что **alt** важнее, чем **title**, есть: **alt** *заменяет* текст у изображений, а **title** *дополняет* текст тегов, который есть и так. Не исключено также, что мы просто не замечаем факторов, связанных с атрибутами **title**, из-за того, что не делим их по тегам. Если бы поисковая машина учитывала этот атрибут, например, только для клеток таблиц с числовыми данными, этот эффект, скорее всего, потерялся бы в «шуме» от **title** в других тегах.



Ссылочно-текстовые факторы

Ссылочно-текстовые параметры ведут себя похоже и на ссылочные, и на текстовые.

Как и в случае с «общими» ссылочными параметрами (см. ниже стр. 58), **вхождения (слов) запроса в ссылки на сайт** важны в обеих поисковых машинах, **вхождения в ссылки на URL** — преимущественно в Google.

Как и у текстовых параметров, соблюдается принцип антибуквализма (см. выше стр. 33): чем ближе к точному запросу, тем меньше вероятность, что для параметра есть связь

с попаданием в топ-30, и тем слабее будет корреляция с позицией.

В Google ссылочно-текстовые параметры заметно сильнее, чем в Яндексе, — это относится и к корреляциям с позицией, и к контрасту между значениями для топ-30 и фона.

Это закономерно — и не потому, что Яндекс якобы придает ссылкам меньшее значение.

ТАБЛИЦА ССЫЛОЧНО-ТЕКСТОВЫХ ПАРАМЕТРОВ. **GOOGLE**

	Слова запроса + выделенные	Слова запроса + синонимы	Слова запроса	Неточный запрос + синонимы	Неточный запрос	Запрос с точностью до форм	Точный запрос
Ссылки на сайт	0,24	0,23	0,22	0,15	0,13	0,13	0,13
Ссылки на страницу	0,15	0,15	0,15	0,14	0,13	0,12	0,10

ТАБЛИЦА ССЫЛОЧНО-ТЕКСТОВЫХ ПАРАМЕТРОВ. **ЯНДЕКС**

	Слова запроса + выделенные	Слова запроса + синонимы	Слова запроса	Неточный запрос + синонимы	Неточный запрос	Запрос с точностью до форм	Точный запрос
Ссылки на сайт	0,17	0,14	0,13	0,05	0,03	0,03	0,04
Ссылки на страницу	0,05	0,05	0,05	0,03	0,03	0,03	0,03

— сильная, средняя, слабая связь с попаданием в топ-30.

— сильная, средняя, слабая корреляция с позицией.

0,08 — коэффициент корреляции (ранговый критерий Спирмена).

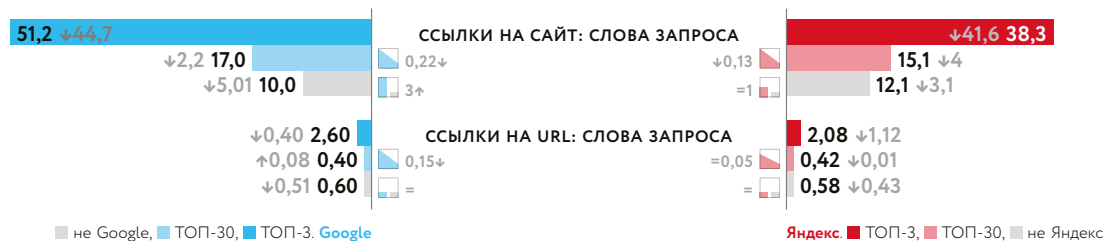
Google использует ссылочные факторы как для оценки сайта, так и для оценки конкретной страницы. Яндекс — преимущественно для оценки сайта: страничные ссылочные параметры для большинства выборок очень слабые. Но ссылочно-текстовые параметры оценивают ссылки с точностью до запроса, а значит, даже в «сайтовых» параметрах содержится привязка к конкретной странице (той, которая находится по данному запросу).



Количество вхождений слов запроса в ссылки на сайт сильнее коррелирует с количеством ссылок на страницу, чем с количеством ссылок на сайт.

С другой стороны, ссылочно-текстовые факторы позволяют оценить релевантность запросу не только одной «лучшей» страницы, которая попадет в результаты поиска, но и сайта в целом. Потому параметры этой группы всё-таки важны для Яндекса.

Похожую роль, кстати, могут играть и исходящие внутренние ссылки на странице, которую Яндекс (или Google) находит по запросу: чем больше среди них релевантных, тем больше, следовательно, релевантных страниц на сайте, и тем более релевантен запросу сайт в целом. Не случайно между параметрами этих двух групп — вхождения в ссылки на сайт и вхождения в исходящие внутренние ссылки на странице — есть достаточно сильные корреляции.



■ не Google, ■ ТОП-30, ■ ТОП-3, ■ Google
 ■ Яндекс, ■ ТОП-3, ■ ТОП-30, ■ не Яндекс

■ 3, ■ 2, ■ 1 — сильная, средняя, слабая связь с попаданием в топ-30, ■ — связи с попаданием в топ нет.
 ■ — сильная, средняя, слабая корреляция с позицией, 0,08 — коэффициент корреляции с позицией (ранговый критерий Спирмена),
 ■ — связи с позицией нет, ■ ■ ■ — корреляция с позицией с обратным знаком.
 ▲1% — сравнение со стандартной выборкой (% — процентные пункты, = — изменений нет).

Практические советы

1. Если у конкурентов много слов запроса (обычно так и бывает!), а у вас мало, это может быть критически важно.
2. Научитесь смотреть на страницы своего сайта глазами ассессора. Сейчас это стало просто: инструкция для ассессоров Google опубликована и даже переведена на русский. Потренируйтесь на страницах из топ-3 – они должны быть релевантны запросу (иначе они бы сюда не попали).
3. Если с вашей экспертной точки зрения страница релевантна и слов запроса на ней не меньше, чем у конкурентов, — о «ключевиках» можно не беспокоиться: поисковые машины разберутся и без них.
4. Бойтесь SEO-текстов. Избегайте лишней «лирики», «воды», канцелярита и штампов. Проверяйте страницы сайта «Тургеневым». Если без SEO-текста можно обойтись, уберите его — или хотя бы сократите.
5. Убирая лишнее, контролируйте значения параметров, чтобы не просадить релевантность.
6. Не нужно специально вставлять слова запроса в заголовки, ссылки, элементы списков и т. п. — но если по этим группам параметров вы отстаёте от конкурентов, разберитесь, почему.
7. Если на странице есть витрина, именно она должна давать основной вклад в текстовую релевантность. Увеличьте количество позиций на странице. Указывайте не только модель, но и тип товара. Расширьте названия товаров (услуг) и добавьте их краткие характеристики. Пропишите картинкам атрибуты `alt`.
8. Не пренебрегайте «человекопонятными URL».
9. Важны не только слова запроса. Синонимы, «спектральные» слова (*купить, недорого*), географическая привязка (*в Москве*) могут быть даже важнее — особенно если поисковик выделяет их в снипетах.
10. Релевантная страница не должна быть одинокой. Она не «взлетит», если не будет поддержана множеством подчиненных и/или соседних страниц.

РАНЖИРОВАНИЕ САЙТОВ

Ранжируются не страницы, а сайты. Страница, которую вы видите в результатах поиска, — только верхушка айсберга. Если есть из чего выбирать, поисковые машины предпочитают выбрать не лучшую верхушку, а лучший айсберг.

Релевантность страницы запросу — необходимое, но отнюдь не достаточное условие. Большая часть факторов ранжирования занимается не страницей, а сайтом в целом. Их назначение — оценить массу, объем, чистоту, популярность, удобство и т. п. «айсбергов» и выбрать из них самый достойный. Это делается разными методами — но, казалось бы, далекие друг от друга параметры часто бьют в одну цель и связаны между собой сильными корреляциями.

В этом году мы решили рассмотреть параметры из разных групп именно под этим углом зрения. Последовательного обзора не будет — он дан в двух предыдущих наших больших отчетах, и повторять его пока нет смысла, благо изменилось не так уж и много. Вместо этого мы, как и с текстовыми фак-

торами, сделаем упор на логике ранжирования и рассмотрим основные тенденции и направления отбора сайтов.



Краткое содержание предыдущих серий

Факторы-2017:

подробно о коммерческих; социальных, ссылочные

Факторы-2018:

трафиковые и поведенческие, технические, коммерческие, социальные, ссылочные.

Размер имеет значение

По-видимому, самый простой и при этом распространенный принцип ранжирования у поисковых машин — отдавать предпочтение большим сайтам перед средними, а средним — перед маленькими.

Измерять сайты можно разными способами — прежде всего по посещаемости и количеству страниц. А также по ассортименту (если речь идет о магазинах), по количеству ведущих на них ссылок и популярности в соцсетях, и даже по возрасту. И в рамках каждого из этих подходов можно считать разные вещи и по-разному. Но как ни считай, результаты будут если не одинаковыми, то похожими: все пути сходятся у одних и тех же сайтов.

Чем больше у сайта страниц, тем больше на него (при прочих равных) ссылок, тем больше его находят поисковые машины, тем выше посещаемость. Чем старше сайт, тем больше у него было времени на то, чтобы накопить ссылочную массу и собрать аудиторию — не говоря уже о количестве страниц. Чем больше ассортимент, тем больше страниц, и тем больше посетителей из соцсетей и по ссылкам. Чем выше посещаемость, тем больше ресурсов и стимулов для того, чтобы расти. Разумеется, все эти цепочки могут прерваться в любом месте — но если они работают, сайт становится большим во всех отношениях.

МОНОПОЛИЗАЦИЯ ВЫДАЧИ

Выйти в лидеры сложно, но если сайту это удалось, дальше положительная обрат-

ная связь действует с удвоенной силой, и основным ее мотором становятся поисковые машины.

Мы видели в недвижимости, что из 300 запросов нашей выборки в 246 случаях в Яндексе и в 252 — в Google в топ попадает один и тот же сайт — Cian.ru (включая поддомены). Причем более чем в 200 случаях в каждом из поисковиков — на первую позицию.



Спору нет, ЦИАН — отличная база данных объектов недвижимости и полезный сайт, но трудно представить себе, что ни по одному из этих 200 запросов не нашлось страницы лучше, чем на ЦИАНе. При этом есть запросы, где у Яндекса Cian.ru на первом месте, а у Google — вообще не попал в топ-30, или наоборот. Скорее всего, ранжирование здесь работает так: если странице с сайта-тяжеловеса удалось пройти предварительный отбор и попасть в пул сайтов, претендующих на высокие позиции (в условные топ-30, или топ-100, или топ-1000), сайтовые факторы ранжирования поднимают его наверх и «прибивают гвоздями» к топу. Скорее всего, для конкретной группы запросов (тематика + интент пользователя) сила этих факторов при переранжировании²¹ увеличивается. Нечто подобное иногда происходит, например, с Википедией по информационным

²¹ См. tools.pixelplus.ru/news/pereocenska-serp-patent-google.

запросам. Но для нас сейчас важнее положительная обратная связь, которая при этом возникает.

Сайт, который по многим запросам ранжируется столь высоко, получает очень много трафика. Это может влиять на его позиции и непосредственно²², и опосредованно — через несколько каналов: ссылки, упоминания в соцсетях, отзывы; рост дохода и, соответственно, возможностей для дальнейшего наращивания ассортимента, контента, рекламного и поискового продвижения.

В результате небольшое количество сильных игроков начинает доминировать в поисковой выдаче, оставляя для всех остальных нишевые запросы, до которых у «крупняка» не дошли руки, или где у сайтов поменьше есть специальные конкурентные преимущества (например, «говорящий» домен). И эта ситуация имеет тенденцию воспроизводиться и усиливаться; мы наблюдаем ее во многих отраслях.



Иногда, например для информационных запросов по медицине в Яндексе, сайтовые факторы специально ослабляются, чтобы более релевантные страницы не были вытеснены страницами с крупных сайтов (как это происходит по аналогичным запросам у Google).²³

ИКС: ИЗ ЧЕГО СКЛАДЫВАЕТСЯ КАЧЕСТВО САЙТА

В Яндексе в топе нашего рейтинга (см. стр. 04) много параметров, по которым

²² Хочется верить, что переходы с результатов поиска самой поисковой машины на позиции в ней прямо влиять не должны — это было бы совсем уж карикатурой. Но переходы из других поисковиков учитываться вполне могут.

²³ Впрочем, хотя результаты Яндекса по таким запросам получаются совсем не похожими на результаты Google, и те и другие не слишком удачны; см. наш аналитический отчет по факторам ранжирования в медицине.

крупные сайты отличаются от небольших: **количество страниц в индексе**, **огромный ассортимент**, количество ссылок на сайт (**ссылающиеся домены**, **общей LinkRank**), **количество посещений сайта** (общее и **из поиска**).

Возглавляет рейтинг **ИКС** — «индекс качества сайта», новая «пузомерка» от Яндекса, заменившая **ТИЦ**. Год назад ИКС уже был самым сильным параметром, сейчас его вес остался приблизительно на том же уровне (контраст с фоном немного упал, но корреляция с позицией выросла), а вес «архивного»²⁴ ТИЦ упал практически до нуля²⁵).

Давайте посмотрим, из каких компонентов, по мнению Яндекса, складывается качество сайта. Непосредственно мы этого, конечно, видеть не можем, но многое становится понятным из рейтинга параметров, с которыми коррелирует ИКС, см. врезку.

Самые сильные корреляции — с трафиковыми параметрами: **рейтинги SimilarWeb** и **Alexa**, **количество посещений** (общее и из разных источников). Дальше идут сайтовые ссылочные факторы (количество **уникальных ссылок на сайт** и др.), затем **количество страниц в индексе** и **огромный ассортимент**. Всё это параметры, оценивающие размер сайта и поднимающие наверх «крупняк».

И только после двадцатой позиции начинаются параметры, имеющие отношение к тому, что естественно было бы назвать *качеством*

²⁴ Яндекс больше не рассчитывает ТИЦ для сайтов и не публикует его, поэтому мы используем последние значения ТИЦ, которые были ранее получены. Правда, мы успели получить некоторое количество нулей для новых сайтов, которые должны «смазывать» картину.

²⁵ Корреляция с позицией еще видна, хотя и не очень сильная: 0,06 (год назад было 0,10, два года назад — 0,17), но точный тест Фишера уже не подтверждает «неслучайность» распределения значений. Значения для топ-30 отстали от фона (причем это отставание как раз статистически значимо).

Название новое	ИКС	Изм. поз.	Название новое	ИКС	Изм. поз.
ИКС	1,000		Оплата картой	0,399	-4
SimilarWeb Global Rank	-0,944		Доля трафика: реклама (display)	0,397	+13
SimilarWeb Country Rank	-0,941		Канал на YouTube	0,394	+4
Alexa Rank	-0,938		Ассортимент	0,391	
Количество посещений: ссылки	0,907		Контакты (в интерфейсе)	-0,385	+5
Количество посещений сайта	0,889		Телефон на странице	-0,378	+5
Количество посещений: прямые переходы	0,885		Группа в «Одноклассниках»	0,370	
Количество посещений: поиск	0,846	+4	Информационность сайта	0,362	-3
Количество посещений: почта	0,842		Самовывоз	0,354	+3
Количество посещений: соцсети	0,841		Размер HTML-кода	0,348	-9
SimilarWeb Category Rank	-0,795		Видео	0,334	+3
Уникальные ссылки на сайт	0,729		Несколько фотографий	0,322	+14
Страниц с сайта в индексе Яндекса	0,702		Много адресов	0,318	-4
Ссылающиеся на сайт домены	0,697		Facebook («Поделиться»)	0,318	+20
Ссылающиеся на сайт страницы	0,684		Реклама	0,316	+22
Общий LinkRank ссылок на сайт	0,664		URL с https	0,316	-8
Количество посещений: display	0,564		Соцсети («Поделиться»)	0,313	+19
Огромный ассортимент	0,552	+4	Возраст домена	0,311	
Доля ссылок на главную страницу	-0,483		Много регионов	0,308	-6
Продолжительность посещения	0,477		ВКонтакте («Поделиться»)	0,305	+8
Доля трафика: соцсети	0,475	+4	Локализация в СПб	0,302	+9
Количество SEO-доноров сайта	0,467		Отзывы	0,301	-6
Процент отказов	-0,453		Специалисты	0,300	+6
Группа в Facebook	0,439	+7	Исходящие внешние ссылки	0,286	-29
Группа ВКонтакте	0,436	+5	Подбор по параметрам	0,278	-7
Найдено документов с сайта в Яндексе	0,432	-8	Оплата	0,277	-10
Регистрация (в интерфейсе)	0,431		Доля текста в HTML-коде	-0,275	-14
Просмотры страниц за посещение	0,422	-5	Много телефонов	0,261	-15
Группы в соцсетях	0,420	+5	Доля SEO-текста в тексте	-0,236	-14
Доля трафика: почта	0,404		...		
			ТИЦ	0,229	-23

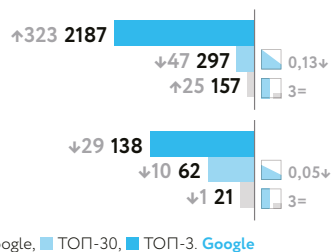
Таблица 2. Корреляции с ИКС. Показаны корреляции параметров с «Индексом качества сайта». Некоторые похожие параметры пропущены. Показаны изменения позиций в рейтинге (от 3 вверх или вниз) по сравнению с октябрем 2018 (см. [Факторы-2018](#)).

сайта без характерного для представителей поисковых машин лицемерия — **доля ссылок на главную страницу** сайта²⁶, поведенческие метрики (**продолжительность посещения, процент отказов, просмотры страниц за посещение**). Сюда же относятся доли переходов из «хороших» источников: **из соцсетей, из почты, из рекламы**. Далее следуют параметры, связанные с соцсетями, и параметры, описывающие функциональность сайта: **регистрация, оплата картой, доставка (самовывоз)** и др.

Приоритеты Яндекса при оценке «качества» сайтов очевидны: хороший сайт — это большой сайт, с огромным трафиком, огромным количеством страниц и огромным ассортиментом.

Всё остальное, конечно, тоже важно, но только при прочих равных — на порядок величины ИКС другие параметры повлиять не могут. На первом месте с большим отрывом — размер аудитории.

²⁶ Чем больше доля ссылок на внутренние страницы, тем лучше: это означает, что на сайте много полезных страниц. Впрочем, это параметр говорит и о размере сайта, см. стр. 59.



■ не Google, ■ ТОП-30, ■ ТОП-3, ■ Google
 ■ 3, ■ 2, ■ 1 — сильная, средняя, слабая связь с попаданием в топ-30, ■ — связь с попаданием в топ нет.
 ■ — сильная, средняя, слабая корреляция с позицией, 0,08 — коэффициент корреляции с позицией (ранговый критерий Спирмена),
 ■ — связи с позицией нет, ■ — корреляция с позицией с обратным знаком.
 ↑1% — сравнение со стандартной выборкой (% — процентные пункты, = — изменений нет).

ТРАФИКОВЫЕ ФАКТОРЫ

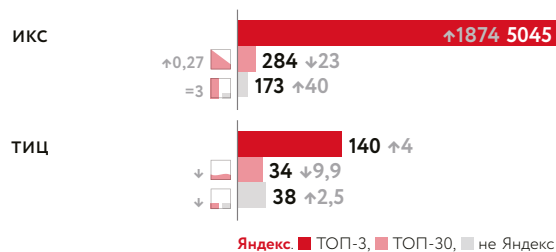


Поисковые машины работают для пользователей, а пользователи не могут ошибаться: они голосуют своими кликами, это и есть демократия!²⁷

Положиться на выбор миллионов и предпочитать сайты с максимальной посещаемостью — самый простой способ избежать грубых ошибок вроде показа спамерских, мошеннических или просто ненадежных сайтов.

С информационными запросами это часто не срабатывает: мы видели на медицинской выборке, как у Google экспертами по лечению всех болезней стали «Комсомольская правда» и «Аргументы и факты». Но с коммерческими сайтами такой подход может давать неплохие результаты, помогая выбрать из множества похожих примерно одинаково релевантных результатов наиболее надежный. Лучшую цену так не получить, но зато вероятность серьезных разочарований резко падает: товар, скорее всего, по крайней мере будет в продаже по той цене, которая указана.

²⁷ «Все ресурсы являются равноправными с точки зрения алгоритма расчета индекса качества», — гордо пишет Яндекс на странице *Что такое индекс качества сайта*.



■ Яндекс, ■ ТОП-3, ■ ТОП-30, ■ не Яндекс

К тому же посещаемость слабо поддается воздействию: накрутки легко ловятся; покупать дорого; чудеса случаются, но редко. В отличие, например, от ссылок, которые можно получить раз и навсегда, за посещаемость нужно бороться постоянно.

По этой же причине рекомендация увеличивать аудиторию для владельцев сайтов не слишком полезна: и так понятно, что надо, а как этого добиться? Каким бы ни было поисковое продвижение сложным и дорогим, оно — едва ли не самый эффективный способ привлечения аудитории: завоевав позиции в поиске один раз, ты надолго обеспечишь себе приток посетителей. И вот здесь возникает замкнутый круг: трафик, полученный от поисковых машин, помогает улучшить позиции в поиске и получить еще больше трафика. Чем больше ранжирование опирается на трафиковые факторы, тем больше опасность монополизации выдачи, о которой мы говорили выше.

У поисковых машин почти неограниченные возможности для отслеживания трафиковых метрик — свои браузеры, счетчики, бары, десктопные и мобильные приложения. Они могут по-разному учитывать переходы из разных источников: прямой ввод адреса,

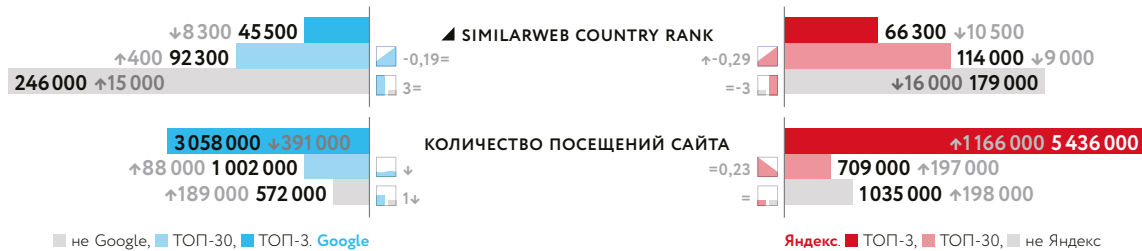
переходы из соцсетей, из почты, по ссылкам, с баннеров и т. п. Могут опираться на размер аудитории и различные его срезы (например, сколько посетителей заходит на сайт каждую неделю в течение месяца).

Наши источники данных о посещаемости — гораздо более скромные и приблизительные: только то, что предоставляют сервисы SimilarWeb и Alexa. Но этого достаточно, чтобы констатировать: трафиковые параметры очень сильны в обеих поисковых машинах — несмотря на то, что, по словам представителей Google, «посещаемость сайта не является фактором ранжирования»²⁸.



Напрямую посещаемость сайта, скорее всего, действительно не учитывается (и в Яндексе, надеемся, тоже) — но факторы ранжирования, связанные с трафиком, сильно коррелируют между собой, и так или иначе сайты с более высокой посещаемостью получают преимущество при ранжировании.

²⁸ "Traffic to a website isn't a ranking factor."
См. searchengines.ru/traffic-google.html.



■ 3, ■ 2, ■ 1 — сильная, средняя, слабая связь с попаданием в топ-30, ■ — связи с попаданием в топ нет.
■ — сильная, средняя, слабая корреляция с позицией, 0,08 — коэффициент корреляции с позицией (ранговый критерий Спирмена),
■ — связи с позицией нет, ■ ■ ■ — корреляция с позицией с обратным знаком. ▲ — инвертированный параметр (чем меньше, тем лучше).
 ↑1% — сравнение со стандартной выборкой (% — процентные пункты, = — изменений нет).

Ранги по посещаемости **Alexa Rank**, **SimilarWeb Global**, **Country** и **Category Rank**, конечно, не могут быть факторами ранжирования, но они хорошо показывают важность трафиковых факторов: для обоих поисковиков видна и сильная корреляция с позицией (естественно, обратная: чем ниже ранг, тем лучше), и сильная связь с попаданием в топ-30. Корреляции сильнее в Яндексе, контраст между значениями для топ-30 и фона — в Google.

Наши параметры, учитывающие общее **количество посещений** и посещения из различных источников (**поиск**, **прямые переходы**, **ссылки** и др.), к сожалению, основаны на неполных данных: SimilarWeb не выдает количество посещений для сайтов, у которых оно не превышает 60-70 тысяч в месяц. Получается, что мы можем сравнивать данные о посещаемости только для сайтов, у которых она выше некоторого порога — но уж что есть, то есть (правда, в результате число посещений выходит сильно завышенным). Здесь для Яндекса видна сильная корреляция с позицией, для Google — сильная связь с попаданием в топ-30.

ССЫЛОЧНЫЕ ФАКТОРЫ

Ссылки долго были главной мерой важности сайтов — и до сих пор сохраняют эту функцию, хотя и разделили ответственность с другими группами факторов.

Ссылочные факторы в своё время (на рубеже веков) произвели в поиске революцию, на волне которой поднялся Google. Алгоритм PageRank дал возможность учитывать не просто текстовую релевантность страниц, но и их «вес»; аналогичным образом можно было считать вес и для сайтов. В Яндексе приблизительно тогда же поя-

вился тИЦ, который на протяжении двух десятков лет оставался главной регалией для сайтов на одной шестой части суши. Помните «денежки» Яндекса в подвале многих сайтов?²⁹

Поначалу ссылочные факторы работали очень хорошо, но ими быстро начали манипулировать; появился рынок «продажных ссылок» и биржевая торговля на нём, которая в России стала инновационным бизнесом и достигла рекордных объемов. Возможности поискового продвижения определялись бюджетом на ссылки.³⁰ Google, не сталкивавшийся в других странах с такими объемами ссылочного спама, был довольно пассивен — их «Пингвин» нашей «Сапе»³¹ был как слону дробина. Но Яндекс объявил ссылкам войну (предварительно научившись заменять их другими факторами) и вышел из нее победителем. Фильтр «Минусинск», применявшийся очень выборочно, но угрожавший буквально всем коммерческим сайтам, заставил массово снимать ранее купленные ссылки и отказываться от приобретения новых. Рынок ссылочного продвижения не исчез, но съезжился и вошел в цивилизованные рамки.



Даже на пике борьбы со ссылками они продолжали учитываться, в том числе и для коммерческих сайтов — см. об этом в **Факторах-2017** и **Факторах-2018**.

²⁹ См. devaka.ru/articles/yandex-cy-buttons.

³⁰ Ситуация, когда выше ранжируется тот, кто может вложить больше денег в продвижение, сама по себе не так уж и плоха — именно поэтому поисковые машины продолжали учитывать платные ссылки, хотя распознать их как таковые было не сложно.

³¹ Google Penguin — алгоритм (фильтр), наказывающий сайты за некачественную ссылочную массу. Sape.ru — биржа ссылок.

Сейчас ссылочные и ссылочно-текстовые параметры — уже не главная, но важная группа факторов ранжирования. Сайтовые ссылочные параметры «работают» в обеих поисковых машинах, страничные — только в Google (впрочем, слабая корреляция с позицией есть и в Яндексе).

Корреляция между разными ссылочными параметрами — количеством **уникальных ссылок, ссылающихся доменов и страниц, суммарным LinkRank** — настолько велика, что более или менее всё равно, что измерять. И в Google, и в Яндексе сильны и связь с попаданием в топ-30, и корреляция с позицией — но, как и в случае с трафиковыми факторами, в Google выше разброс значений между топ-30 и фоном, а в Яндексе — внутри топ-30 (между топ-3 и третьей десяткой выдачи).

Сила сайтовых параметров в Яндексе за последний год выросла; их значения выросли везде — судя по всему, Рунет окончательно оправился после «Минусинска» и принялся наращивать ссылочную массу. Впрочем, значения страничных ссылочных

параметров, как и всех ссылочно-текстовых, скорее падает, а **доля безанкорных ссылок** растет. О ссылочно-текстовых параметрах см. стр. 49.

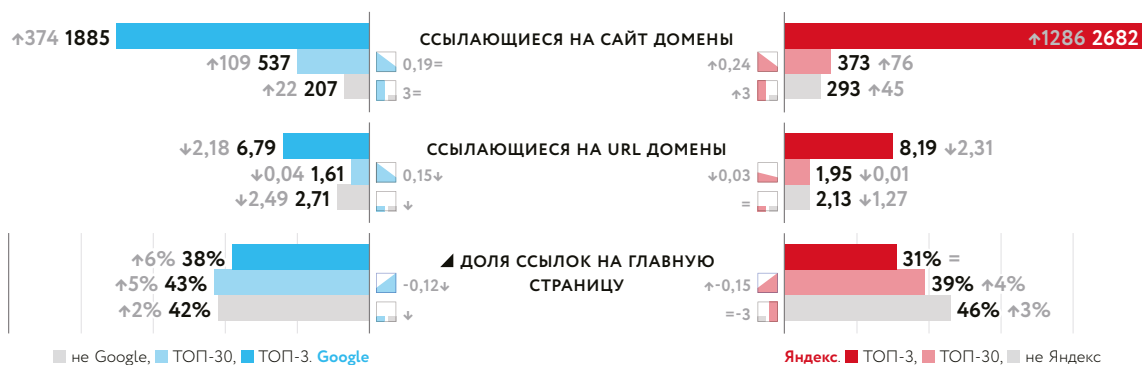
Косвенный, но весьма сильный ссылочный параметр — **доля ссылок на главную страницу** сайта. Чем она ниже, тем больше вес внутренних страниц. Низкая доля ссылок на главную сильно коррелирует с посещаемостью, большим ассортиментом, размером сайта, общим количеством ссылок и т. п.

АССОРТИМЕНТ, РАЗМЕР САЙТА, ПРИСУТСТВИЕ В РЕГИОНАХ

Для интернет-магазинов (и не только) размер бизнеса важнее, чем размер сайта. И когда уходящим в тень ссылочным факторам подставляли плечо факторы новые, коммерческие, поисковые машины стали пытаться его оценивать — правда, в основном по косвенным признакам.



Один из таких критериев — ассортимент. Поисковики могут оценить его и по объему



↑3, ↑2, ↑1 — сильная, средняя, слабая связь с попаданием в топ-30, □ — связи с попаданием в топ нет.
 □, □, □ — сильная, средняя, слабая корреляция с позицией, 0,08 — коэффициент корреляции с позицией (ранговый критерий Спирмена),
 □ — связи с позицией нет, □, □, □ — корреляция с позицией с обратным знаком. ▲ — инвертированный параметр (чем меньше, тем лучше).
 ↑1% — сравнение со стандартной выборкой (% — процентные пункты, = — изменений нет).

витрин (если отдельной страницы для каждого предложения нет), и, конечно, по количеству и разнообразию страниц с описаниями товарных или иных позиций. У нас ассортимент сайта оценивают ассесоры, и из нескольких вариантов в ассессорской анкете формируются два бинарных параметра (поисковые машины, скорее всего, оценивают его количественно).

Огромный ассортимент имеет прямое отношение к монополизации выдачи, это очень сильный параметр, и его значения из года в год продолжают расти. Даже на общей выборке в топ-3 Яндекса таких сайтов уже почти 40%, а в топ-30 — почти 20%, а в электронной коммерции — больше половины и больше трети соответственно; в сильно монополизированной недвижимости — 86% и 54%. Сайты с **широким ассортиментом** получают в топ-30 Яндекса более двух третей мест, в электронной коммерции — 92%, в недвижимости — 96%! Для сайтов со средним и тем более ограниченным ассортиментом остается совсем мало места.

Можно было бы предположить, что причинно-следственная связь обратная: чем больше ассортимент, тем по большему количеству запросов на сайте есть релевантные страницы, и тем больше его видимость в поиске. Но корреляции с позицией показывают, что сайты с большим ассортиментом в среднем ранжируются выше. К тому же наша основная выборка — политематическая, и запросов по каждой из тематик в ней мало, так что мультипликатор по количеству релевантных запросов, который для отраслевых выборок, несомненно, играет свою роль, здесь практически не работает.

В Google значения этих параметров ниже: он с подозрением относится к агрегаторам и оставляет место в результатах поиска для **монобрендовых** сайтов: в топ-30 Яндекса их всего 10%, в Google — 17%. Доля сайтов,

в ассортименте которых много брендов, в Яндексе составляет 79%, в Google — 69%. Соответственно, на сайты с несколькими брендами и там, и там остается меньше 15% выдачи.

У коммерческих сайтов, особенно интернет-магазинов, об ассортименте можно судить и просто по размеру сайта: количество **страниц с сайта в индексе Яндекса** — сильный параметр в обоих поисковиках.



Впечатляет разница между медианными значениями для топ-3 и топ-30 Яндекса и Google — в сто раз, на два порядка! И в топ-3 большой рост за год, который, вероятно, еще продолжится.

Количество **найденных страниц с сайта** для коммерческих сайтов отражает ассортимент в узкой области, связанной с запросом пользователя³². Кроме того, с размером витрины и через него с ассортиментом связаны количества **цен на странице, картинок, исходящих внутренних ссылок** (см. об этом также в разделе про текстовые факторы).

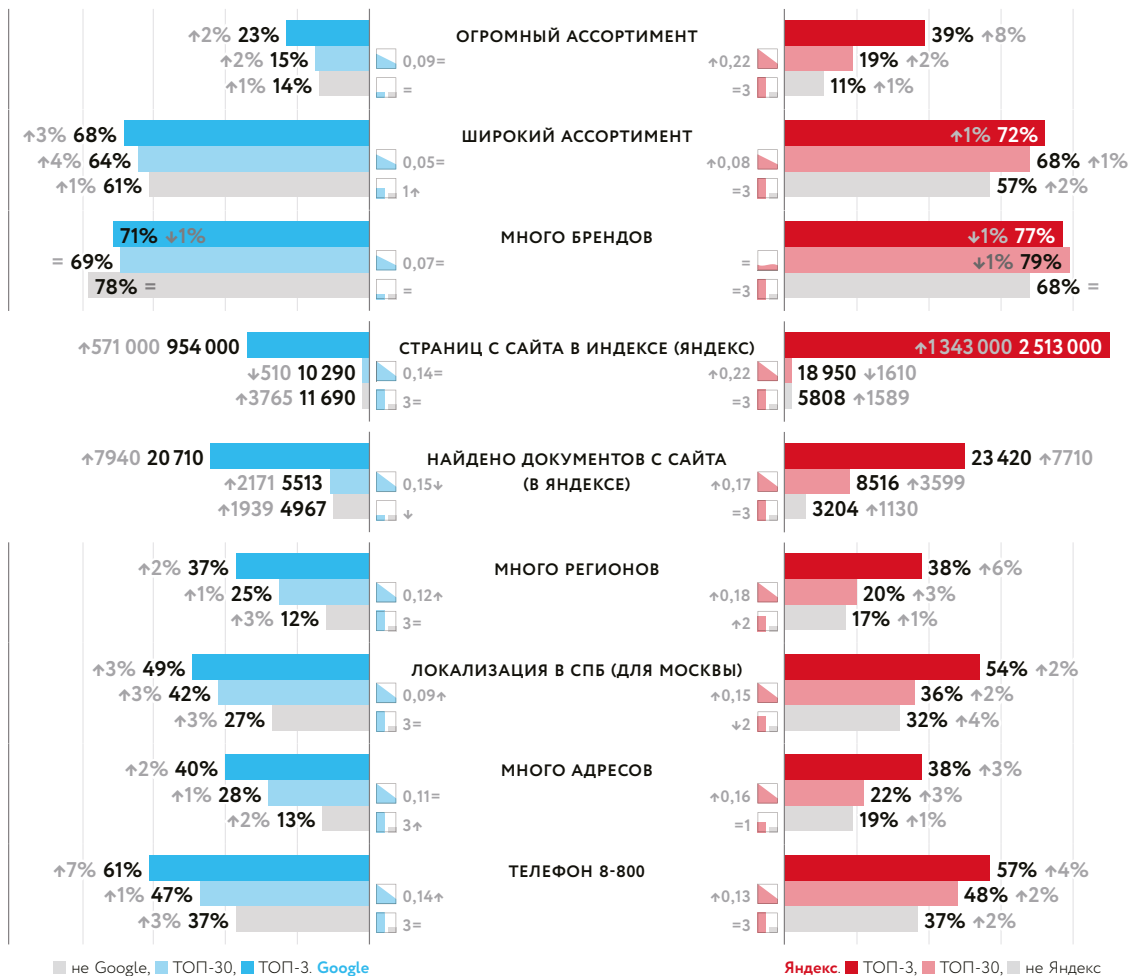
Еще один признак крупного, надежного бизнеса для поисковых машин — представленность во **многих** или хотя бы **нескольких регионах**. Для Москвы показателем последнего может быть **локализация в Санкт-Петербурге** (наиболее частый дополнительный регион для бизнесов, активных в столице). При этом **основная локализация в Москве** в Яндексе есть у 88% коммерческих сайтов, а просто **локализация в Москве** — у 94%, это фактически необходимое условие для попадания в топ.

³² Правда, сюда попадают и информационные страницы, но они тоже важны как косвенный признак наличия «экспертизы» в данной области.



Похожие параметры, косвенно связанные с размером бизнеса и его распространённостью в регионах — **много адресов** и **много телефонов**. Примерно эта же идея стоит и за знаменитым **номером 8-800** (это едва ли не самый хрестоматийный при-

мер коммерческого параметра). Его основной смысл — облегчить обратную связь для жителей других городов, для которых звонки в Москву могут быть дороги. Предпочтения поисковых машин дают на бизнес, и доля сайтов с таким номером в топ-30 обоих



■ 3, ■ 2, ■ 1 — сильная, средняя, слабая связь с попаданием в топ-30, □ — связи с попаданием в топ нет.
 ■■■ — сильная, средняя, слабая корреляция с позицией, 0,08 — коэффициент корреляции с позицией (ранговый критерий Спирмена),
 □ — связи с позицией нет, ■■■ — корреляция с позицией с обратным знаком.
 ↑1% — сравнение со стандартной выборкой (% — процентные пункты, = — изменений нет).



поисковиков уже вплотную приблизилась к половине; в электронной коммерции — почти 60%, в финансах — 65–74%. В топ-3 Яндекса на общей выборке 57% сайтов с бесплатными для пользователя

общероссийскими телефонными номерами, в топ-3 Google — 61%, причем рост за год здесь составил 4–7 процентных пунктов (это тоже признак растущей монополизации выдачи).

Что стоит за сайтом

Склепать сайт и начать продвигать его может кто угодно, в наше время это дело не хитрое. Важно, есть ли у сайта что-то за душой.

В те времена, когда мы занимались детекцией поискового спама, массово изготавливались сайты, выглядящие как корпоративные, но представляющие собой пустую оболочку, — зато идеально оптимизированные. Они успешно продвигались в топ, получали из воздуха поиска трафик и продавали его тем, кто готов был за него платить. Сейчас подобные технологии работают с информационными сайтами, см. раздел *Ранжирование по информационным запросам* в нашем медицинском отчете.



Коммерческий сайт ценен для пользователей, если за ним стоит реальный бизнес — работающий с клиентами, дающий гарантии и обеспечивающий постпродажное обслуживание, отвечающий за актуальность цен и других данных на сайте и доставляющий покупки вовремя и «с надлежащим качеством». Информационный — если его делает серьезная редакция, обладающая необходимым уровнем того, что сейчас называют «экспертизой», и порождающая оригинальный, качественный контент. Даже агрегаторы и доски объявлений

бывают очень разными — одни надежные и действительно полезные, другие сделаны на коленке (причем часто оптимизаторской).

В принципе, чем крупнее бизнес, тем он надежнее, поэтому многое из сказанного выше о размере сайтов, трафике, ассортименте и т. п. имеет отношение и к этому направлению отбора. Но есть и параметры, с размером напрямую не связанные.

«КОМПЕТЕНТНОСТЬ, АВТОРИТЕТНОСТЬ, НАДЕЖНОСТЬ»

Google в руководстве для ассессоров³³ ставит во главу угла при оценке любой страницы *компетентность, авторитетность, надежность*³⁴ сайта и/или создателей контента (далее для краткости будем говорить

³³ См. его русский перевод от Ant-Team Александра Чепукайтиса: ant-team.ru/public/GoogleAnt-Team.ru.pdf. Цитаты ниже взяты из этого перевода.

³⁴ Англ. E-A-T — Expertise, Authoritativeness, Trustworthiness.

об авторитетности). Особое значение эти критерии имеют для сайтов, попадающих в категорию «кошелек или жизнь» — англ. *YMYL, your money or your life*. Определение этой категории достаточно широкое, оно включает, например, покупку товаров и «новости о важных темах», но в некоторых отраслях (медицина, финансы и др.) факторы, связанные с авторитетностью сайта, получают действительно первостепенное значение. Именно их усиление лежало в основе сразу нескольких важных апдейтов Google за последние два года, включая недавнее сентябрьское обновление основного алгоритма. Подробнее об этом см. во Введении к нашему [медицинскому отчету](#).

Среди общих критериев, на которых может основываться оценка авторитетности коммерческого сайта, называются:

- положительная репутация компании или сайта;
- опыт работы в соответствующей сфере;
- исчерпывающая информация о компании;
- подробная информация по обслуживанию клиентов;
- отзывы пользователей;
- сайт производителя или компании, официально признанной производителем (например, официальные дилеры для [автомобилей](#));
- регулярное редактирование и обновление информации.

Для некоторых отраслей и типов контента применяются особые критерии. Например, на медицинских сайтах весь контент «может быть написан или составлен только людьми или организациями с соответствующей медицинской компетенцией или аккредитацией».

Всё это — критерии оценки для ассессоров Google. Но они не имели бы смысла, если бы не были реализованы в факторах ранжирования. Несмотря на некоторые нюансы в приоритетах (опять вынуждены сослаться на наш [медицинский отчет](#)), в Яндексе многим из этих критериев также соответствуют факторы ранжирования, что подтверждают и наши данные.



КОНТАКТЫ И ПРЕДСТАВЛЕНИЕ КОМПАНИИ

В качестве эпиграфа — цитата из примера, приведенного в руководстве для ассессоров Google:

Это страница «о нас» на сайте ресторана. Здесь представлена информация о том, когда ресторан открылся и что ждет в нём посетителей. На других страницах веб-сайта находятся сведения о ресторане, включая адрес, меню, прочую контактную информацию и т. д. Этот веб-сайт является очень авторитетным, поскольку сам рассказывает о себе всё.

На сайте — но не обязательно на каждой странице — должны быть хотя бы один **телефон** и офлайн-**адрес** — в обеих поисковых машинах доля сайтов, где они есть, не ниже 94%. Для агрегаторов и производителей могут быть некоторые послабления, но продавцам (поставщикам услуг) без этих контактных данных путь в топ-30 фактически закрыт: телефон есть на 99-100% сайтов, адрес — на 97-98%. Примерно в 80% случаев в меню сайта есть раздел **«Контакты»**, но заметной связи с ранжированием для него не видно. **E-mail** присутствует только примерно на половине сайтов и на ранжирование, судя по всему, не влияет, а вот возможность заказать **обратный звонок** важна



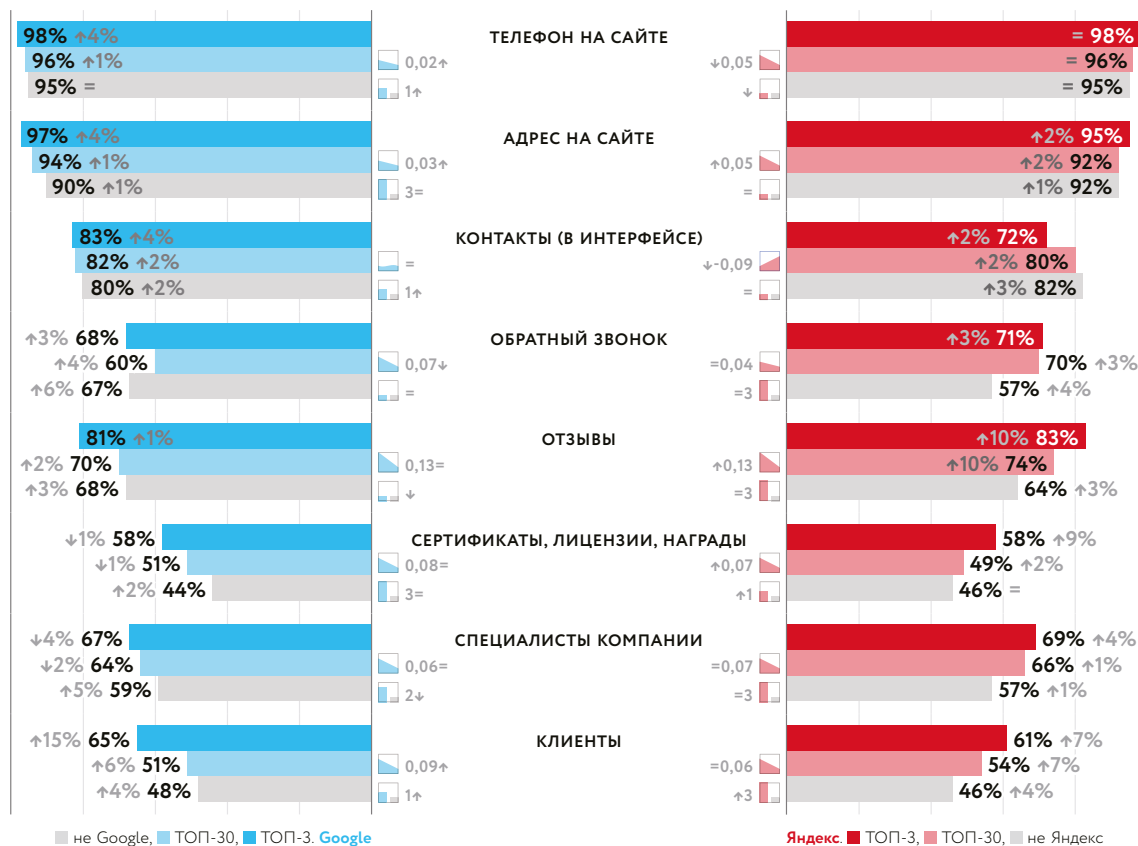
и в Яндексе (сильная связь с попаданием в топ-30), и в Google (корреляция с ранжированием).

Очень важно, чтобы на сайте были **отзывы**: в обоих поисковиках сильная корреляция с позицией; в Яндексе на нашей выборке доля сайтов с отзывами — 74% (видна сильная связь с попаданием в топ-30); в Google — лишь немногим меньше, 70%. Для клиник



в топ-30 Яндекса сайтов с отзывами почти 90%, в топ-3 — 95%.

Сертификаты, лицензии, награды демонстрируют около половины сайтов; в Google этот параметр немного сильнее, но его рост в этом году здесь вроде бы прекратился, а Яндекс понемногу «подтягивается». Подтверждения авторитетности предсказуемо более важны для сайтов из областей «кошелек или жизнь»: банки и другие финан-



■ 3, ■ 2, ■ 1 — сильная, средняя, слабая связь с попаданием в топ-30, ■ — связи с попаданием в топ нет.
 ■ — сильная, средняя, слабая корреляция с позицией, 0,08 — коэффициент корреляции с позицией (ранговый критерий Спирмена),
 ■ — связи с позицией нет, ■ — корреляция с позицией с обратным знаком.
 ↑1% — сравнение со стандартной выборкой (% — процентные пункты, = — изменений нет).


совые сайты выкладывают их в 66–75% случаев, клиники — в 88–90%.

Примерно в трети случаев сайты представляют **специалистов** компании, но и значения параметра, и их важность сильно зависят от отрасли. В сфере услуг сведения о специалистах есть чаще (недвижимость — 77–84%, клиники — 95–96%), на сайтах автосалонов — заметно реже (26–27%, и связи с ранжированием не видно). Как ни странно, для интернет-магазинов этот параметр оказывается достаточно важным: в обоих поисковиках 44–46% в топ-30 (при 34–38% в фоне) и 52–54% в топ-3.

Сведения о **клиентах** компании на общей выборке встречаются примерно в половине случаев — в основном за счет сайтов услуг (например, маркетинг, строительство) и сложных продаж (станки, генераторы). На наших отраслевых выборках (кроме недвижимости) их заметно меньше, но везде, кроме автомобилей, где раскрывать клиентов, по-видимому, не принято, сохраняется связь с ранжированием.

Практически для всех упомянутых выше параметров важно именно наличие информации на сайте — на всех страницах поисковики ее увидеть не ожидают.

Впрочем, об офлайновой составляющей бизнеса поисковые машины сейчас могут судить не только по сайтам.

 Большую роль может играть, например, присутствие на картах и в других сервисах Яндекса и Google.

Например, Яндекс предлагает пользователям оставлять отзывы о компаниях и, с недавнего времени, о сайтах, — и, по-видимому, учиты-

вает их при ранжировании. «Пользователи могут оставлять отзывы непосредственно на сервисах Яндекса или на сайтах-партнерах.»³⁵

СОЦИАЛЬНЫЕ СЕТИ

Группы в соцсетях сейчас есть у 71% сайтов в топ-30 Яндекса и Google по нашей общей выборке (год назад было 66–68%); в обеих поисковых машинах, особенно в Яндексе, это важный параметр с тенденцией к росту. Для отраслевых выборок значения еще выше — до 88–90% в топ-30 у клиник; сильнее всего связь с ранжированием для недвижимости, слабее всего — для автосалонов.

Из конкретных социальных сетей сейчас по нашим данным важнее всего **Facebook** и **Youtube** (но канал в Youtube коррелирует с наличием на сайте **видео**, которое важно и само по себе).

Ссылки «Поделиться» тоже важны, хотя на тех страницах, которые обычно попадают в результаты поиска по коммерческим запросам, они стоят редко: примерно 20% на найденных страницах против примерно 60% на сайте.

Как всегда, наши параметры — только первое приближение к данным и алгоритмам, которые могут использовать поисковые машины, грубая проверка гипотез о том, что может быть важно. Скорее всего, «социальные» сигналы используются более активно — могут быть важны и размер групп в соцсетях (социальная аудитория), и ссылки из соцсетей, и тональность комментариев. Данных об активности пользователей соцсетей у нас немного: только **«шеры» ВКонтакте** и **в**



³⁵ См. yandex.ru/support/webmaster/service/reviews.html.

«Одноклассниках»³⁶. Это страничные параметры — сайтовые были бы, конечно, полезнее. Корреляция с позицией видна для них и в Яндексе (слабая), и в Google (немного более сильная).

ЦЕНЫ И СКИДКИ

Попадали ли вы в ситуацию, когда хотели что-то купить дешево, нашли в Сети много вроде бы подходящих предложений, но только потратили зря время и так ничего и не купили? Дозвониться нельзя, вообще контактных данных не найти, «пива нет и неизвестно», цена изменилась, доставка через месяц и в два раза дороже, модель не та, состояние не то, еще что-то не так?



³⁶ Есть еще «Мой мир@Mail.ru», немногочисленные «шеры» в котором важны в поиске Mail.ru, но, похоже, совсем не интересуют Яндекс и Google. Данные о «шерах» в Facebook были доступны в прошлом году, по ним картина примерно такая же, как по «ВКонтакте».

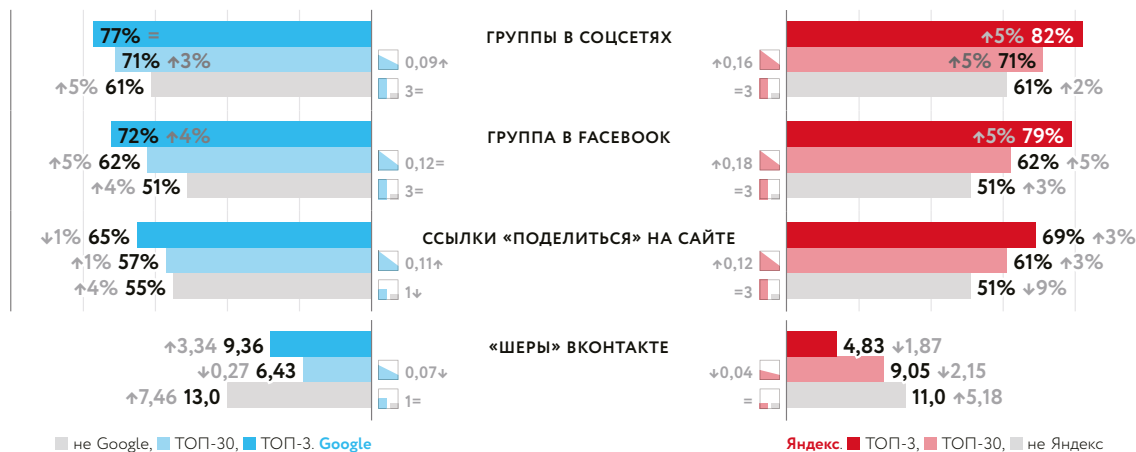
И хорошо еще, если вы сами выискивали дешевые варианты. А если они на первой странице поисковой выдачи — можно ли такие результаты поиска признать качественными?



Вот поэтому поисковики и не гонятся за дешевой, надежностью для них важнее.

Наличие **цен на сайте** крайне важно, на **найденной странице** — тоже важно (см. диаграммы выше на стр. 29). Важно и **количество цен** — правда, это связано скорее с размером витрины: на сайтах электронной коммерции цен раза в два больше, чем на общей выборке.

Однако нельзя сказать, что при приближении к топ-1 цены падают. Какую цену ни взять — **среднюю, медианную, максимальную** или



■ 3, ■ 2, ■ 1 — сильная, средняя, слабая связь с попаданием в топ-30, ■ — связи с попаданием в топ нет.
 ■ — сильная, средняя, слабая корреляция с позицией, 0,08 — коэффициент корреляции с позицией (ранговый критерий Спирмена),
 ■ — связи с позицией нет, ■ — корреляция с позицией с обратным знаком.
 ↑1% — сравнение со стандартной выборкой (% — процентные пункты, = — изменений нет).

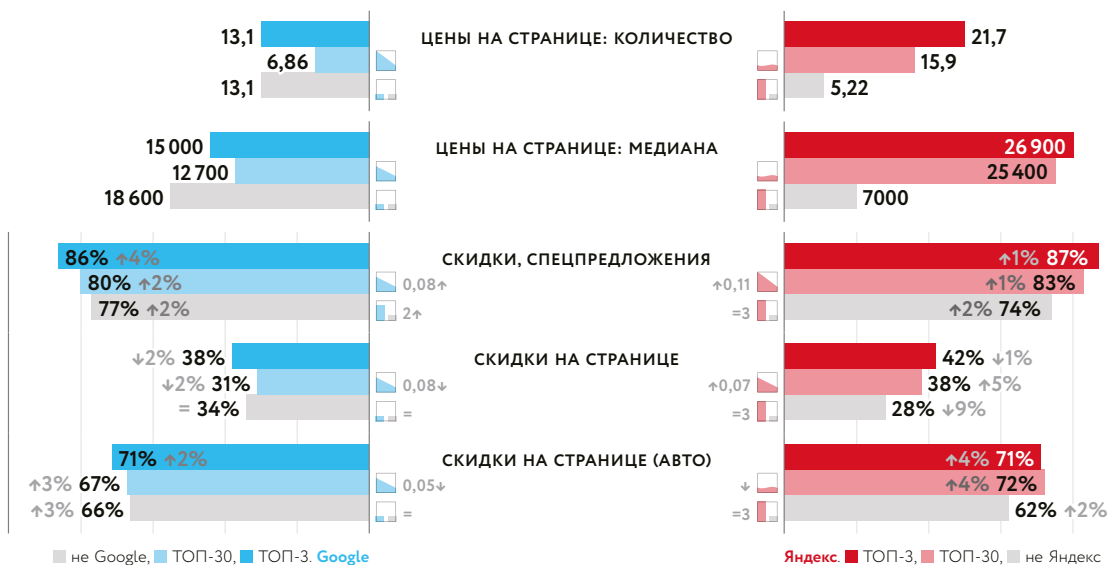
минимальную³⁷ — она обычно, наоборот, растет, особенно в Google. Правда, пик цен часто достигается не на самой первой позиции, а на третьей или пятой. Для Яндекса рост цен внутри топ-30 менее характерен, но зато в топ-30 цены обычно намного выше, чем в фоне.³⁸

Зато очень важно, чтобы на сайте были **акции, спецпредложения или скидки** — если нельзя добиться действительно низких цен, то нужно по крайней мере продемонстрировать заботу о пользователе. Это один из тех параметров, которые были сильными с момента начала наблюдений, но при этом их значения продолжают расти. Сайтов со скидками в топ-30 Яндекса и Google на нашей общей выборке 80–83%, в электронной коммерции — 96–97%, а среди автосалонов — 99%! На найденной **странице** акции и скидки как таковые встречаются относительно редко, но **упоминаются** гораздо чаще.



³⁷ Последний параметр самый ненадежный, т. к. часто минимальная цена на странице относится не к основному товару, а аксессуару или услуге (например, это может быть стоимость доставки).

³⁸ Более высокий уровень цен связан и с тем, что в топ попадают сайты с более качественным, а потому и дорогим, ассортиментом.



■ не Google, ■ TOP-30, ■ TOP-3. Google
 ■ Яндекс, ■ TOP-3, ■ TOP-30, ■ не Яндекс

■ 3, ■ 2, ■ 1 — сильная, средняя, слабая связь с попаданием в топ-30, ■ — связи с попаданием в топ нет.
 ■ — сильная, средняя, слабая корреляция с позицией, 0,08 — коэффициент корреляции с позицией (ранговый критерий Спирмена),
 ■ — связи с позицией нет, ■ — корреляция с позицией с обратным знаком.
 ↑1% — сравнение со стандартной выборкой (% — процентные пункты, = — изменений нет).

Качество сайта

Крупные сайты, запрос за запросом попадающие в топ, — это не просто крупные бизнесы, это и сайты «правильные». У них есть чему поучиться.

Впрочем, провести границу между сайтом и бизнесом тут трудно. Как, например, понять, что важнее — чтобы гарантия была на сайте или чтобы она была в жизни (то есть реально работала)? Для пользователей, возможно, второе. Для поисковиков — скорее всего, первое.

Читая этот раздел, важно помнить, что для вашего сайта актуально только то, что не противоречит его тематике, типу (жанру) и запросу, и подтверждается на сайтах конкурентов, занимающих высокие позиции.

ПОВЕДЕНЧЕСКИЕ И ТРАФИКОВЫЕ ФАКТОРЫ

В том, что поисковые машины используют поведенческие и трафиковые метрики, нет никаких сомнений. Всякий, у кого стоит Яндекс.Метрика, может убедиться в этом, открыв главную страницу счетчика. Доли трафика из различных источников; время, проведенное на сайте; глубина просмотров; процент отказов...

Очевидно, что поисковые машины считают данные о поведении пользователей важными, собирают их и используют в ранжировании. Делать это можно двумя способами — напрямую (используя поведенческие метрики как факторы ранжирова-

ния) или опосредованно, используя эти данные для обучения системы наряду с ассессорскими оценками и добиваясь от других факторов ранжирования такого оптимального сочетания, при котором наверху окажутся сайты, которые нравятся пользователям. Скорее всего, применяются оба подхода; в какой пропорции — для нас не так уж и важно. Главное, что поведенческие метрики *вливают* на ранжирование, а значит, их улучшение — важное направление работы по поисковому продвижению сайта (не говоря уже о том, что оно может сделать сайт более эффективным и в других отношениях).

В рейтинге параметров, коррелирующих с ИКС (см. стр. 55), поведенческие факторы идут почти сразу за трафиковыми — скорее всего, при его формировании они учитываются. Это первые параметры в этом рейтинге, имеющие непосредственное отношение к качеству сайта. **Продолжительность посещения, процент отказов** (посещений, за которые была просмотрена одна страница), количество **просмотренных страниц** сильно коррелируют с позицией в Яндексе. В Google корреляции с позицией нет, но значения для топ-30 значительно лучше, чем для фона. В Яндексе связи с попаданием в топ не видно (из-за того, что значения в Google больше), но она, скорее всего, есть: за последний год улучшились значения для всех групп позиций, и едва ли не сильней

всего для топ-30, где они уже почти такие же, как в Google.³⁹

Обратите внимание, что Google считает «отказом» визит, за который была просмотрена одна страница, независимо от его продолжительности, а Яндекс — визит, который продолжался не более 15 секунд и не сопровождался никакими действиями, кроме закрытия страницы; в Яндекс.Метрике процент отказов может быть в несколько раз меньше, чем в Google Analytics. SimilarWeb, у которого мы получаем данные для поведенческих параметров, считает, по-видимому, как Google — то есть если пользователь прочитал страницу от начала до конца и закрыл ее, это всё равно отказ. Значения этого параметра (а заодно и двух других) сильно зависят от задачи, стоящей перед пользователем, — а значит, от запроса, типов сайта и страницы и тематики. Для интернет-магазинов хорошая доля отказов — около

40%, для автосалонов — около половины, для медицинских информационных сайтов — около 60% в Google и около 70% в Яндексе (сайты, которые они выдают, очень разные).

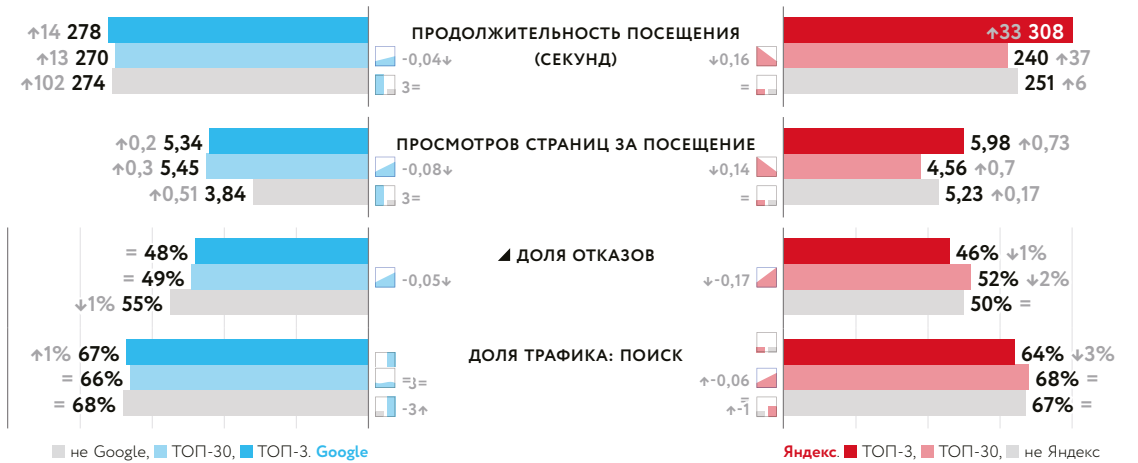


Оценивая поведенческие метрики своего сайта, важно сравнивать его с похожими сайтами, а не с сайтами других типов и тематик — результаты могут быть несопоставимы.

Для поисковых машин могут быть важны источники трафика, приходящего на сайт. Например, высокая доля переходов из социальных сетей может говорить о наличии у сайта собственной аудитории и о готовности пользователей его рекомендовать. К сожалению, хотя наши данные показывают здесь корреляцию с позицией для Яндекса и связь с попаданием в топ-30 для Google, значения параметра для разных групп позиций здесь из-за округления практически не отличаются друг от друга. Везде



³⁹ Формально коэффициент корреляции с позицией в Яндексе упал, но это из-за того, что значения для второй и третьей десятки «подтянулись» к значениям для топ-3 и топ-10; в такой ситуации можно говорить скорее об усилении параметров.



↑3, ↑2, ↑1 — сильная, средняя, слабая связь с попаданием в топ-30, □ — связи с попаданием в топ нет.
 □, □, □ — сильная, средняя, слабая корреляция с позицией, 0,08 — коэффициент корреляции с позицией (ранговый критерий Спирмена), □ — связи с позицией нет, □, □, □ — корреляция с позицией с обратным знаком.
 ↑1% — сравнение со стандартной выборкой (% — процентные пункты, = — изменений нет).



по два, максимум три процента, и только в топ-3 предельно монополизированной выборки по недвижимости как недостижимый идеал сияют немислимые 5%. Похожая ситуация с другими «хорошими» источниками трафика — **почта, прямые переходы, реклама (display)**. Доля **переходов по ссылкам** с других сайтов, по-видимому, не считается положительной характеристикой: в Google для нее связи с ранжированием не видно, а в Яндексе видна только связь с попаданием в топ (при разнице в значениях между топ-30 и фоном по всем выборкам в 1–2%). Отрицательный герой в этой истории — **переходы из поиска** (см. диаграмму): чем выше их доля, тем меньше у сайта собственного трафика.

Параметры, описывающие количество визитов из разных источников, сильно коррелируют между собой (и с общим количеством посещений). Поэтому, хотя все они занимают высокие места в рейтингах параметров для Яндекса и Google, судить о том, какой именно трафик предпочитают поисковые системы, по ним трудно.

ТРЕБОВАНИЯ К ФУНКЦИОНАЛУ

Поисковые машины предъявляют к коммерческим сайтам длинный список требований⁴⁰, который, впрочем, может существенно корректироваться в зависимости от запроса, типа сайта и т. п. Некоторые из этих требований принципиальные — например, наличие **«кнопки заказа»** или **цен** (см. диаграмму на стр. 29) для интернет-магазинов. Если их нет, сайт, скорее всего, скорее всего, просто не будет отнесен к нужному типу и не попадет в выдачу. Но большинство — мягкие: если «галочка» в соответствующей



⁴⁰ К информационным, по-видимому, тоже, но другой, и мы пока не пытались его «прочитать».

графе не стоит, занять хорошую позицию можно и без нее, но шансы понижаются.

При этом «работают» только те коммерческие факторы, которые актуальны в вашей области и для вашего типа сайтов: клиникам не нужна **доставка**, а на сайтах автосалонов не принято обсуждать способы **оплаты** и тем более возможность **оплаты картой**. Поэтому наборы важных параметров будут очень разными для разных отраслей — а иногда и для разных типов сайтов внутри одной отрасли.



Коммерческие факторы вообще не слишком универсальны — но это не значит, что ими можно пренебрегать.

Вот краткий обзор требований к сайту, для которых на нашей общей выборке видна сильная связь с ранжированием.

Описания товаров и услуг

- Собственно текстовые **описания** — кажется тривиальным, но они есть почти в 100% случаев, и попытка обойтись без них может оказаться фатальной.
- **Фотографии товаров** — очень близко к 100% (см. диаграмму на стр. 29).
- **Несколько фото на товар**: важно наличие функционала, а не присутствие на всех страницах, включая найденную. На общей выборке более 70%, для автосалонов — около 100%.
- **Видео** (см. диаграмму на стр. 28) — более 60% сайтов в топ-30 и 75% в топ-3; для автосалонов — 85–90% и 98–99%. Не обязательно должно быть на страни-

цах товаров или услуг, но часто встречается именно там.

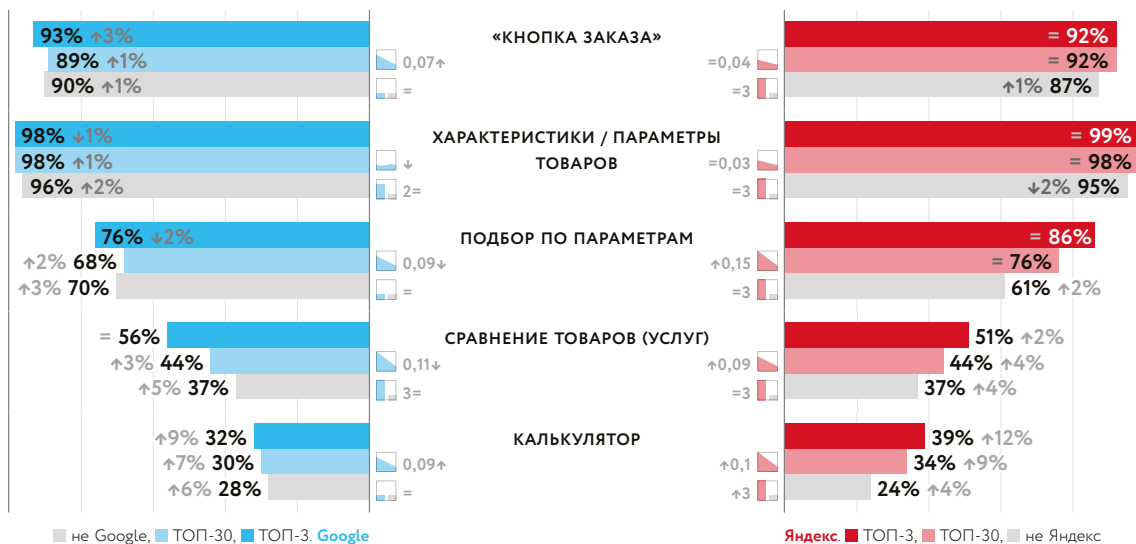
Подбор и сравнение товаров (услуг)

- **Характеристики (параметры)** товаров или услуг: средние значения обычно близки к 100%, но может быть не актуально (например, в медицине).
- **Подбор по параметрам и сравнение товаров** или услуг. Сильные и активно растущие параметры в Яндексе. В Google тоже есть корреляция с позицией, а для сравнения товаров — и сильная связь с попаданием в топ-30.
- **Калькулятор** (на сайте): аналог подбора по параметрам, актуальный там, где товар или услуга формируется для клиента

индивидуально: автомобили, недвижимость, финансы. Но и на общей выборке связь с ранжированием заметна, особенно в Яндексе.

Условия продажи

- Описание способов **оплаты** товара или услуги: порядка 60% в топ-30 на общей выборке, более 90% для интернет-магазинов (и 97–98% в топ-3).
- Возможность безналичной оплаты: банковская карта, **Яндекс.Деньги**, **Webmoney** очень приветствуется. **Оплата картой** (диаграмма на стр. 24) на общей выборке есть примерно в половине случаев; для интернет-магазинов — более чем в 80%. Но для некоторых отраслей, наоборот, совершенно не актуально.



■ 3, ■ 2, ■ 1 — сильная, средняя, слабая связь с попаданием в топ-30, □ — связи с попаданием в топ нет.
 ■■■ — сильная, средняя, слабая корреляция с позицией, 0,08 — коэффициент корреляции с позицией (ранговый критерий Спирмена), □ — связи с позицией нет, ■■■ — корреляция с позицией с обратным знаком.
 ↑1% — сравнение со стандартной выборкой (% — процентные пункты, = — изменений нет).

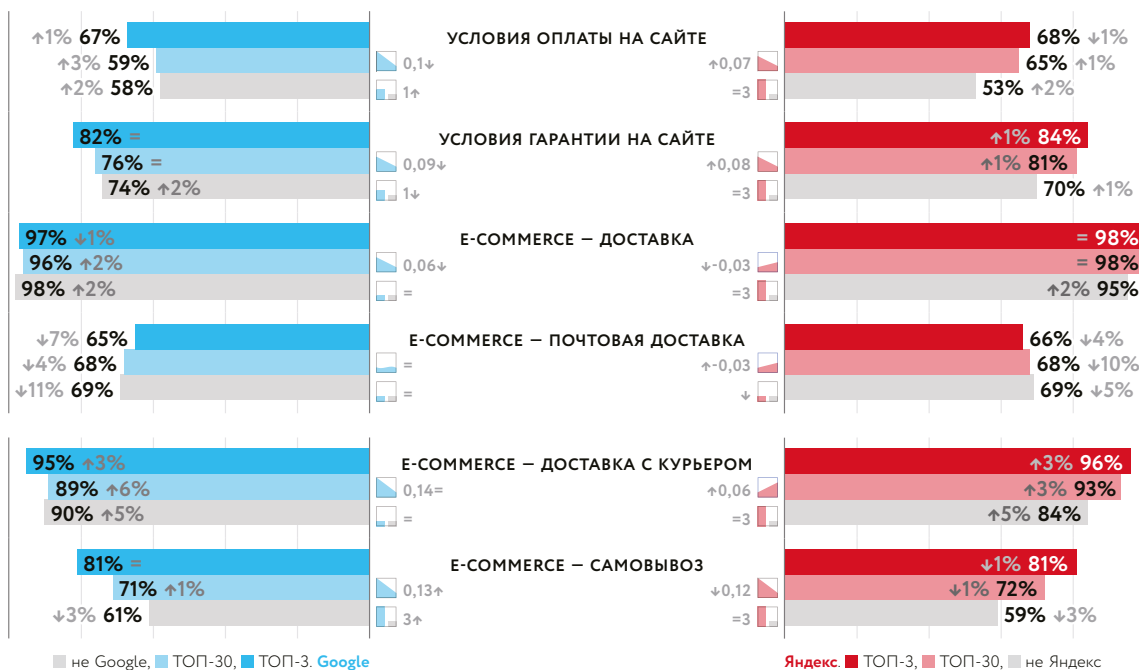


- Описание условий **гарантии** (а также **постпродажного обслуживания**): около 80% в топ-30 Яндекс и Google на нашей общей выборке; заметно больше для интернет-магазинов и автосалонов, меньше для других отраслей, особенно для недвижимости, где преобладают агрегаторы⁴¹. Сведения о гарантии на многих сайтах сводятся к дежурным фразам о том, что гарантийное обслуживание осуществляется в сервис-центрах производителей, — тем не менее, наличие хотя бы такой информации для интернет-магазинов необходимо.

⁴¹ Параметры этой группы вообще малозначимы и часто не актуальны для агрегаторов.

- Информация о **доставке** актуальна прежде всего для интернет-магазинов: 98% в топ-30 Яндекса, 96% в Google (см. тж. диаграмму для общей выборки на стр. 24). Здесь, по-видимому, уже произошло насыщение раствора — но, как и с оплатой, поисковые машины, похоже, готовы вдаваться в детали. Если для «дальней» доставки (**почта, транспортные компании**) связи с ранжированием не видно, то для «локальной» (**курьер** и особенно **самовывоз**) эта связь очень сильная⁴².

⁴² Пользователи в России (в отличие, например, от США, где почтовая доставка очень популярна) предпочитают локальные магазины и локальную доставку, и поисковикам приходится под них подстраиваться.



■ 3, ■ 2, ■ 1 — сильная, средняя, слабая связь с попаданием в топ-30, ■ — связи с попаданием в топ нет.

■ — сильная, средняя, слабая корреляция с позицией, 0,08 — коэффициент корреляции с позицией (ранговый критерий Спирмена), ■ — связи с позицией нет, ■ ■ ■ — корреляция с позицией с обратным знаком.

↑1% — сравнение со стандартной выборкой (% — процентные пункты, = — изменений нет).

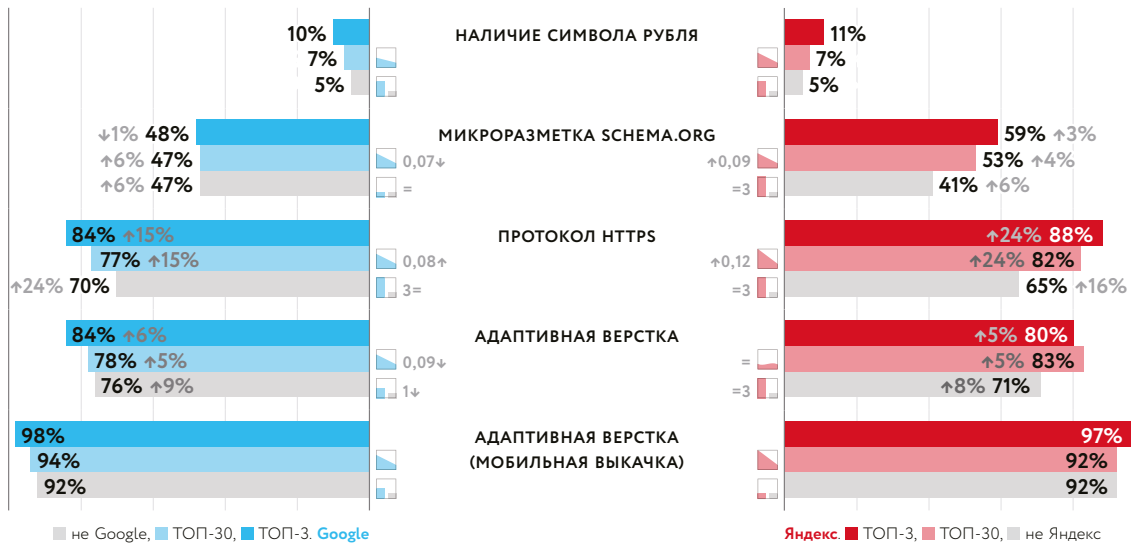
ТЕХНИЧЕСКИЕ ТРЕБОВАНИЯ И «ПРОДВИНУТЫЕ» САЙТЫ

В одном шаблоне рекомендаций для клиентов мы встретили предостережение от использования **символа рубля** ₺: если поисковые машины не распознают этот знак как маркер цены, это может повредить ранжированию. Мы решили проверить, как наличие символа рубля на странице связано с ранжированием, и увидели, что связь есть, и она вопреки опасениям вполне положительная (хотя и не очень сильная). Из других параметров наличие символа рубля сильнее всего коррелирует с поведенческими факторами, с количеством цен на странице и с огромным ассортиментом. То есть использование символа рубля характерно скорее для «продвинутых» сайтов и явно ничему не вредит — да и странно было

бы, если бы за шесть лет после утверждения этого символа Центробанком Яндекс и Google не научились бы его понимать.

Вряд ли поисковые машины специально поощряют сайты за символ рубля, но этот пример показывает, что не надо бояться следовать «новомодным веяниям», если они дают выигрыш, например, в удобстве для пользователей. Среди таких современных тенденций: наличие **микроразметки** (**Schema.org** для поисковых машин и **Open Graph** для соцсетей), большой **размер HTML-кода** (см. подробнее стр. 25–27), **низкая доля в тексте в HTML**.

Защищенный **протокол https** и **адаптивная верстка** уже перестали быть признаками прогрессивного сайта и становятся стандар-



■ 3, ■ 2, ■ 1 — сильная, средняя, слабая связь с попаданием в топ-30, ■ — связи с попаданием в топ нет.
 ■ — сильная, средняя, слабая корреляция с позицией, 0,08 — коэффициент корреляции с позицией (ранговый критерий Спирмена),
 ■ — связи с позицией нет, ■ — корреляция с позицией с обратным знаком.
 ↑1% — сравнение со стандартной выборкой (% — процентные пункты, = — изменений нет).

том для всех — сейчас они есть уже примерно на 80% сайтов.



В Яндексе доля сайтов, следующих этому новому стандарту, сейчас заметно выше, чем в Google, который в последние годы активно его пропагандировал.



В результатах по отраслевым выборкам запросов сайтов с https еще больше. Например, в e-commerce, где конкуренция очень высокая, https в топ-30 Яндекса и Google есть уже в 90–92% случаев, а в топ-3 — в 95%.



При выкачке «с мобильного устройства» (с указанием малого размера экрана) в выдаче обоих поисковиков резко возрастает количество сайтов с адаптивной версткой. При этом в Яндексе резко меняется характер связи между параметром и ранжированием: пропадает связь с попаданием в топ-30,

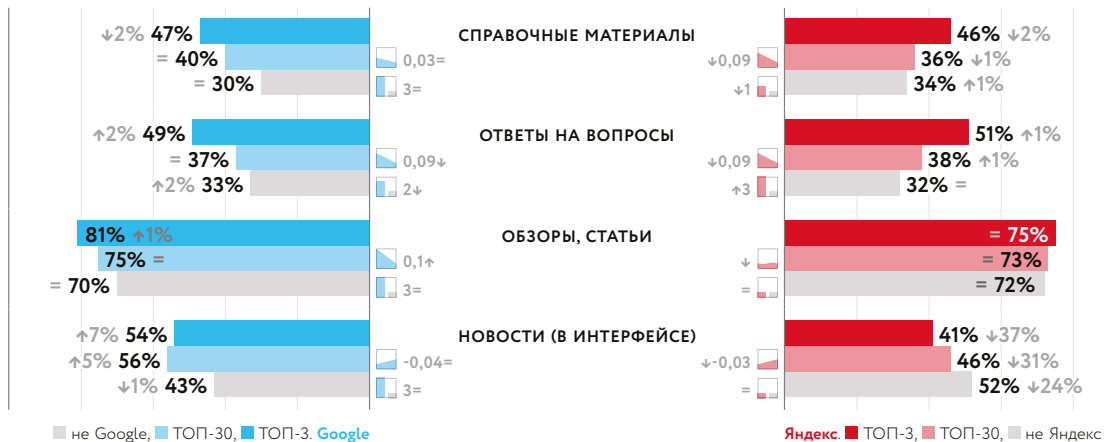
появляется сильная корреляция с ранжированием. А в Google всё остается так же, как и при обычной (десктопной) выкачке.

КОНТЕНТ

Поисковые машины в целом неохотно ранжируют по информационным запросам страницы с коммерческих сайтов⁴³. Тем не менее, вкладываться в него имеет смысл.

Дело в том, что высококачественный релевантный контент может серьезно повысить оценку авторитетности и компетентности сайта поисковыми машинами, которые собирают информацию с сайта и анализируют ее. Еще важнее может быть косвенное положительное влияние контента: увеличивается количество страниц в индексе; среди них много релевантных нужным запросам; если

⁴³ Впрочем, некоторый прогресс в этой области есть: например, в Google в последнее время стали лучше находиться информационные страницы с сайтов [клиник](#).



3, 2, 1 — сильная, средняя, слабая связь с попаданием в топ-30, — связи с попаданием в топ нет.

 — сильная, средняя, слабая корреляция с позицией, 0,08 — коэффициент корреляции с позицией (ранговый критерий Спирмена),
 — связи с позицией нет, — корреляция с позицией с обратным знаком.
 ↑1% — сравнение со стандартной выборкой (% — процентные пункты, = — изменений нет).

контент полезный, растет ссылочная масса, накапливаются упоминания в соцсетях, начинает расти аудитория.

Справочные материалы и **ответы на вопросы** присутствуют на 36–40% сайтов, попавших в топ-30 Яндекса или Google по запросам общей выборки, и примерно на половине сайтов, попавших в топ-3. Есть корреляция с позицией и (обычно) связь с попаданием в топ. Похожая ситуация и с **возможностью задать вопрос**, которая встречается примерно в два раза реже, и с **рекомендациями по выбору** (около 60% в топ-30, 70% в топ-3; в Google важны и **рекомендации по выбору на странице**). По отраслевым выборкам значения могут быть выше (медицина) или ниже (электронная коммерция), но параметры обычно даже сильнее. Исключение составляют **автосалоны**, где контент встре-

чается редко, а связь с ранжированием скорее отрицательная.

Обзоры, статьи присутствуют на сайтах значительно чаще, но в Яндексе связи с ранжированием для них на общей выборке не видно; впрочем, на большинстве отраслевых выборок связь есть⁴⁴. В Google и корреляция с позицией, и связь с попаданием в топ сильные.

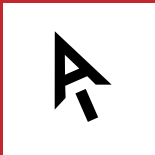
Наконец, **новости** (детектируются автоматически в меню и т. п.) в Яндексе ничего хорошего для ранжирования не обещают. В Google есть сильная связь с попаданием в топ-30 — но еще нужно разобраться, откуда эта связь берется: это Google так любит новости или Яндекс так их избегает.

⁴⁴ Сильнее всего корреляция с позицией на автомобильной и финансовой выборках, где для справочных материалов связь с ранжированием слабее всего.



Практические советы

1. Даже если вы не можете влиять на трафик своего сайта и его возраст (впрочем, тут время работает на вас), наращивать ссылочную массу и ассортимент — вполне в вашей власти.
2. Подумайте — может быть, с источниками трафика всё не так уж и безнадежно? Например, если сделать какой-нибудь полезный сервис, устроить конкурс, раскрутить инстаграм...
3. Размер витрины — косвенный показатель ассортимента.
4. Количество страниц на сайте вообще и релевантных запросу в особенности может быть очень важно. Если для группы моделей (или вариантов услуги) можно сделать одну общую страницу, а можно несколько отдельных, — второй вариант лучше.
5. Во многих отраслях (например, в e-commerce) *большой* — значит работающий по всей стране. Возможно, у вас есть хотя бы партнеры или точки самовывоза в других городах? Их адреса (телефоны) могут оказаться очень кстати. Телефон 8-800 и даже предложение клиентам заказать обратный звонок — тоже.
6. Адрес(а) и телефон(ы) важны и сами по себе — как доказательство вашего существования в офлайне. По возможности подкрепляйте их присутствием на картах и в других сервисах Яндексa и Google.
7. Для чувствительных отраслей вроде медицины и финансов нужны еще доказательства вашей компетентности — сертификаты, свидетельства о регистрации, информация о специалистах, качественный тематический контент из бонуса превращаются в необходимость. Изучите руководство для ассессоров Google, чтобы понять, что важно для вашей отрасли.
8. Если вы надеетесь на чудо, социальные сети — одно из тех мест, где чудеса случаются. Но даже если нет, пренебрегать ими уже нельзя. Как и отзывами.
9. Низкие цены — не главное. Главное — скидки.
10. Составьте список важных параметров и сравните по ним свой сайт и конкурентов. Или приходите к нам за аудитом.



Ашманов
и партнеры

WWW.ASHMANOV.COM

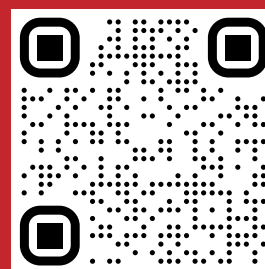
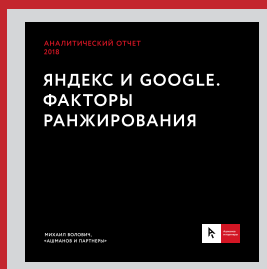
8 495 126-02-61

ПРОДВИЖЕНИЕ БИЗНЕСА В ИНТЕРНЕТЕ

Продвижение сайтов — задача, с которой мы помогли сотням компаний. Нам есть чем гордиться: 18 лет назад мы сформировали рынок поискового маркетинга и всё это время удерживаем лидерство, подтверждая его рейтингами и премиями, кейсами и достижениями. Станьте лидерами вместе с нами!

ЯНДЕКСЕ И GOOGLE. ФАКТОРЫ РАНЖИРОВАНИЯ

Уже три года с вами



КАК ИСПОЛЬЗОВАТЬ НАШИ РЕЗУЛЬТАТЫ

Если параметр для вас неактуален, за его отсутствие вам ничего не будет. стр. 23

ПОДРОБНО О ТЕКСТОВЫХ ФАКТОРАХ

Времена, когда релевантность мерилась количеством вхождений «ключа», давно ушли и вряд ли когда-нибудь вернуться. стр. 31

ИКС: ИЗ ЧЕГО СКЛАДЫВАЕТСЯ КАЧЕСТВО САЙТА

И только после двадцатой позиции начинаются параметры, имеющие отношение к тому, что естественно было бы назвать качеством сайта... стр. 54