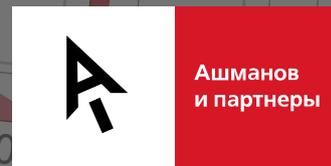


2024



ИНФОРМАЦИОННЫЕ ЗАПРОСЫ:



ФАКТОРЫ РАНЖИРОВАНИЯ
В ЯНДЕКСЕ И GOOGLE

2
часть

МИХАИЛ ВОЛОВИЧ,
ПАВЕЛ ШЕСТАКОВ,
НИКИТА СУШКОВ

ИССЛЕДОВАНИЕ

ashmanov.com



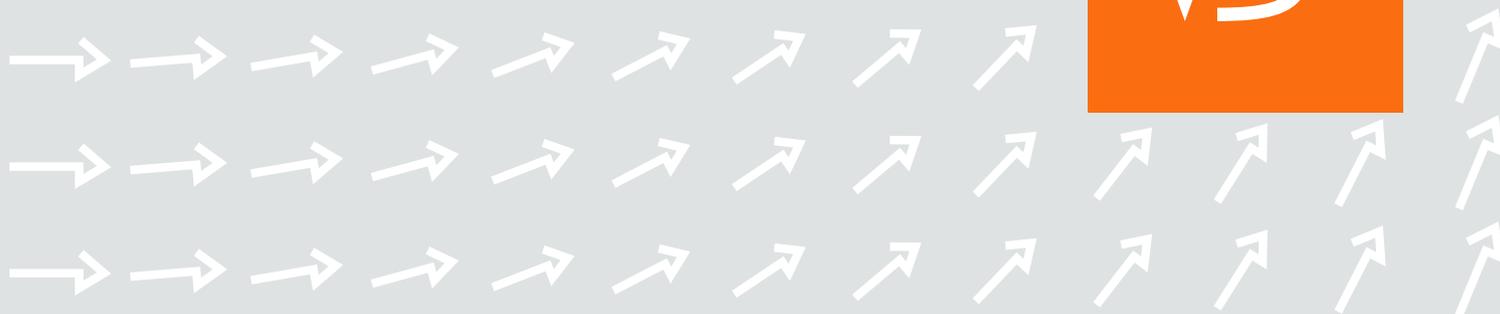
Ашманов
и партнеры

Ашманов и партнеры

Ведущее агентство digital-маркетинга с многолетней историей. Многократный обладатель премии AdIndex. В топ-10 самых цитируемых агентств по версии «Медиалогии».

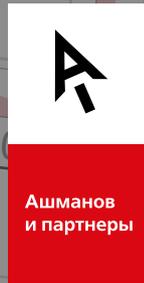
- Оказываем услуги по комплексному продвижению бизнеса в Интернете с 2001 года;
- Применяем продуктовый подход, базирующийся на глубоком изучении пути покупателей и бизнес-задачах клиентов;
- Проводим исследования и разрабатываем технологии и сервисы на базе собственной Лаборатории поисковой аналитики.

Отправьте заявку на [аудит](#) и проверьте соответствие вашего сайта актуальным факторам ранжирования.



Содержание

Ссылочные параметры	04
Ссылки на сайт	06
Ссылки на страницу	09
Ссылочно-текстовые параметры	12
Исходящие ссылки	17
Текстовые параметры	22
Матрица текстовых параметров	24
Логика текстового ранжирования	28
1. Принцип антибуквализма	28
2. Яндекс: контраст с фоном. Google: корреляция с позицией	30
3. Пуд слов россыпью	31
4. Заголовки: всё наоборот	33
5. Какие зоны текста важнее	34
6. SEO-текст и борьба с текстовым спамом	37
А что с информационными запросами	40
Текст страницы	48
Основной текст и «фрагменты»	51
Заголовки <title> и <h1>	54
URL и доменное имя	64
Meta description и keywords	66
Внутренние заголовки	71
Выделения в тексте	77
Ссылки со страницы	80
Списки и атрибуты alt и title	84
Тексты входящих ссылок	92
Малопользней контент	95



Ашманов
и партнеры

Ссылочные параметры



К сожалению, о ссылочных параметрах для информационных запросов мы можем судить только по двум выборкам — *Гаджеты и Наука*.

Для самой большой выборки, *Косметика*, данных по ссылкам мы не собирали (по техническим причинам).

Почти ничего неожиданного со ссылками нам не встретилось. Всё как с коммерческими выборками, только корреляции с позицией в Яндексе посильнее, меньше ссылок на отдельные страницы и еще кое-какие нюансы, не меняющие общей картины.

Ссылки на сайт

Кажется, уже забыты те времена, когда нужно было доказывать, что ссылочные факторы продолжают работать — даже после революции в ранжировании, «замены» ссылок коммерческими, поведенческими и прочими факторами, фильтра «Минусинск» и т. п.

Сайтовые ссылочные параметры очень похожи на ИКС — даже числовые значения корреляции по выборкам примерно такие же. Это неудивительно: корреляция между ними и ИКС весьма высокая. А главное, ссылочные параметры, как и ИКС, связаны с размером сайта и трафиком: чем больше ссылочная масса, тем (обычно) крупнее сайт, а поисковые машины — особенно Яндекс — любят гигантов!

Более или менее всё равно, как оценивать ссылочную массу сайта:

- по общему количеству уникальных ссылок,
- по количеству ссылающихся страниц,
- по количеству ссылающихся доменов

- или же по LinkRank — аналогу гугловского PageRank, учитывающему вес ссылающихся страниц.

Результат получается примерно одинаковым — сила корреляции с позицией (для оценки которой мы используем [ранговый коэффициент Спирмена](#)) или вообще не отличается, или отличается совсем чуть-чуть, на 1-2 сотых. В таблице приведены данные для **количества ссылающихся доменов**, чтобы не пугать читателей сотнями тысяч уникальных ссылок или миллионами (а кое-где и десятками миллионов) ссылающихся страниц — но с тем же успехом можно было бы выбрать и любой другой из перечисленных параметров.

Итак, что мы видим для информационных выборок?

Ссылочные факторы очень сильные в обоих поисковиках, но немного по-разному — так что даже трудно сказать, где они сильнее.

В Яндексе больше контраст внутри топ-30 — между первой тройкой и третьей десяткой. Корреляция с позицией выше, чем для стандартной коммерческой выборки — для сильно монополизированных *Гаджетов* она достигает 0,38, что близко к максимальным заре-

Количество ссылающихся на сайт доменов

Выборка	С корр.	топ-3	4–10	11–20	21-30	топ-30	фон	Ф-30/фон
Яндекс								
Наука	0,29	19943	2167	738	479	821	895	—
Наука ⁻¹	0,16	2785	1505	675	466	627	864	—
Гаджеты	0,38	74871	896	283	150	328	240	
Гаджеты ⁻³	0,25	1744	583	260	145	246	234	—
Станд. комм.	0,22	4547	870	426	329	470	296	
Google								
Наука	0,18	14953	2103	1461	955	1296	581	
Наука ⁻⁴	0,11	3013	1909	1415	940	1128	571	
Гаджеты	0,27	65216	617	427	248	409	182	
Гаджеты ⁻⁵	0,17	1642	489	371	235	330	175	
Станд. комм.	0,23	1400	1266	547	377	520	300	

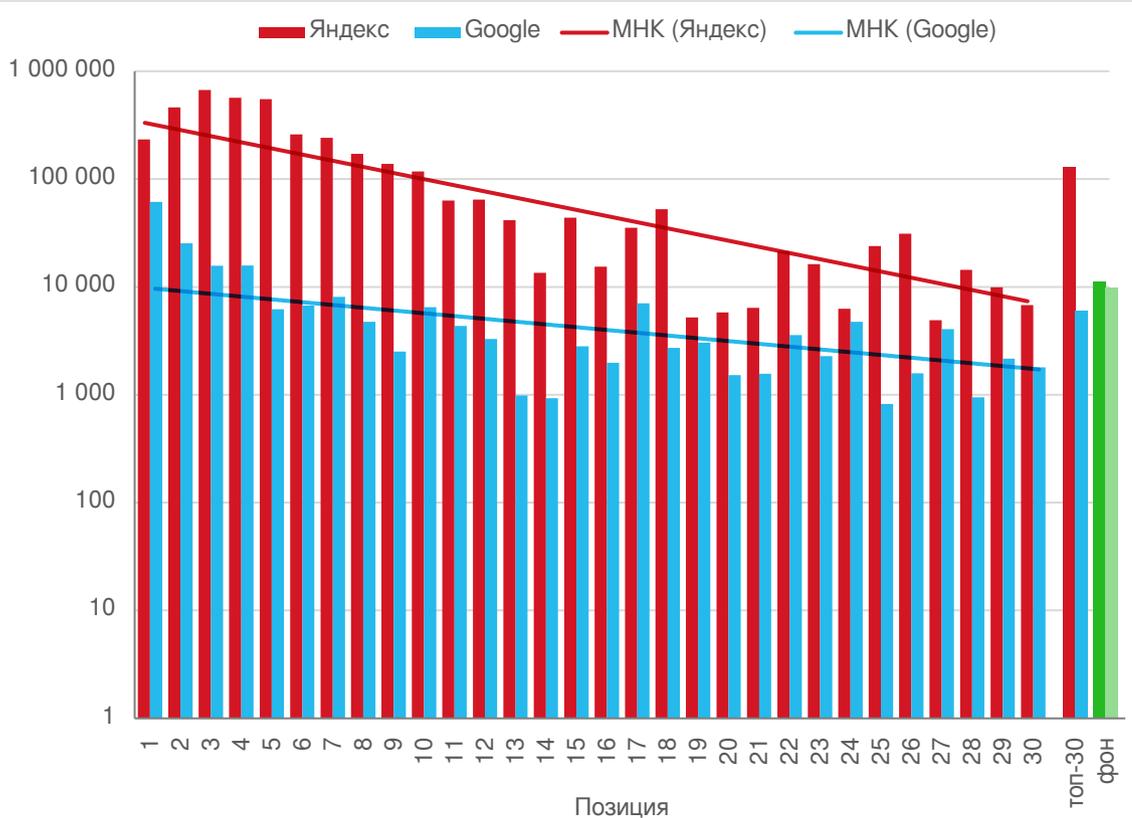
С корр. — коэффициент корреляции с позицией (ранговый критерий Спирмена).

Ф-30/фон — оценка контраста между топ-30 и фоном (точный тест Фишера):

 сильный,  средний,  слабый; прочерк — нет или отрицательный.

Скорректированные выборки: 1: без Википедии, Youtube, Dzen; 3: без Apple, Youtube, Dzen; 4: без Википедии; 5: без Apple, Youtube.

Количество ссылающихся на сайт доменов, Гаджеты



Среднее количество ссылающихся на сайт доменов по позициям в Яндексе и в Google для выборки Гаджеты (логарифмическая шкала). Числа на диаграмме значительно выше чисел в таблице, основанных на медианах по запросам, поскольку на средние значения гораздо сильнее влияют «выбросы» — очень высокие показатели для отдельных сайтов.

Если посмотреть на распределение значений параметра по позициям, мы увидим, что в Google они более или менее равномерно уменьшаются от первой позиции к тридцатой, тогда как в Яндексе максимум достигается только на третьей позиции. Это может быть следствием монополизации выдачи: на одну или несколько самых высоких позиций сайты попадают «за особые заслуги», при этом у них могут быть более низкие значения «общих» факторов, по которым конкурируют остальные. И действительно, для скорректированной выдачи в Яндексе получается диаграмма с максимумом на первой позиции.

гистрированным (для всех параметров в любых выборках) значениям. Даже для скорректированных *Гаджетов* без трех сайтов-лидеров коэффициент корреляции 0,25 — выше, чем для нашей стандартной выборки (0,22). Но зато значения параметра для фона выше, чем для третьей десятки выдачи, и примерно соответствуют второй десятке. То есть получается, что в Яндексе конкуренция по ссылочным параметрам идет внутри топ-30, а фон живет своей отдельной жизнью.

В Google медианные значения ссылочных параметров в топ-30 выше, чем в Яндексе, а уменьшение значений с удалением от первой позиции происходит медленнее — потому и коэф-

фициенты корреляции ниже. Уровень третьей десятки Google примерно соответствует второй десятке выдачи Яндекса — для *Науки* чуть выше, для *Гаджетов* чуть ниже. Зато значения для фона заметно ниже, чем для третьей десятки — но всё равно выше, чем для третьей десятки Яндекса. Получается, что в Google градиент значений ссылочных параметров растянут на более длинный хвост выдачи. Сайты, не попавшие в топ-30, в среднем проигрывают конкуренцию по ссылкам тем сайтам, которые туда попали, — и это, скорее всего, не предварительный отбор, а честная конкуренция на равных.

Ссылки на страницу

Сайтовые параметры почти всегда сильнее страничных — и ссылки не исключение.

Правда, если мы посмотрим на таблицу с **количеством ссылающихся на страницу доменов** (или **уникальных ссылок**, или **ссылающихся страниц**), мы увидим, что уже начиная с 4–10 позиций

в ней сплошные нули, а какие-то отличные от нуля числа — которые тоже все меньше единицы — видны только для первой тройки.

Количество ссылающихся на URL доменов

Выборка	С корр.	топ-3	4–10	11–20	21-30	топ-30	фон	Ф-30/фон
Яндекс								
Наука	0,08	0,01	0	0	0	0	0	
Наука ⁻¹	0,07	0,04	0	0	0	0	0	
Гаджеты	0,11	0,12	0	0	0	0	0	
Гаджеты ⁻³	0,11	0,37	0	0	0	0	0	
Станд. комм.	—	0,07	0	0	0	0	0	
Google								
Наука	0,12	0,04	0	0	0	0	0	
Наука ⁻⁴	0,09	0,03	0	0	0	0	0	—
Гаджеты	0,12	0,13	0	0	0	0	0	
Гаджеты ⁻⁵	0,10	0,23	0	0	0	0	0	
Станд. комм.	0,12	0,48	0,05	0	0	0	0	—

С корр. — коэффициент корреляции с позицией (ранговый критерий Спирмена).

Ф-30/фон — оценка контраста между топ-30 и фоном (точный тест Фишера):

 сильный,  средний,  слабый; прочерк — нет или отрицательный.

Скорректированные выборки: 1: без Википедии, Youtube, Dzen; 3: без Apple, Youtube, Dzen; 4: без Википедии; 5: без Apple, Youtube.

Может показаться, что мы нашли рецепт попадания в топ-3 — как будто бы для этого достаточно просто, чтобы на страницу кто-нибудь ссылался. Но это не так: нули в большинстве клеток таблицы — результат усреднения. На самом деле многие стра-

ницы со ссылками попали и во вторую, и в третью десятку, или вообще не попали в топ-30 — и наоборот, далеко не на все страницы в топ-3 есть ссылки. Наши математические методы работают не с итоговыми нулями, а со всем массивом данных — поэто-

Доля ссылок на главную страницу сайта (чем меньше, тем лучше)

Выборка	С корр.	топ-3	4–10	11–20	21-30	топ-30	фон	Ф-30/фон
Яндекс								
Наука	0,22	0,09	0,15	0,18	0,20	0,17	0,22	
Наука ⁻¹	0,08	0,16	0,17	0,19	0,20	0,19	0,22	
Гаджеты	0,31	0,10	0,19	0,26	0,32	0,25	0,30	
Гаджеты ⁻³	0,17	0,17	0,23	0,27	0,32	0,27	0,30	
Станд. комм.	0,12	0,28	0,37	0,38	0,40	0,37	0,42	
Google								
Наука	0,13	0,14	0,17	0,19	0,21	0,19	0,18	—
Наука ⁻⁴	0,05	0,19	0,18	0,19	0,22	0,20	0,18	—
Гаджеты	0,19	0,15	0,23	0,26	0,30	0,26	0,28	
Гаджеты ⁻⁵	0,08	0,23	0,26	0,27	0,30	0,28	0,28	
Станд. комм.	0,08	0,34	0,34	0,39	0,41	0,38	0,39	—

С корр. – коэффициент корреляции с позицией (ранговый критерий Спирмена).

Ф-30/фон – оценка контраста между топ-30 и фоном (точный тест Фишера):

сильный, средний, слабый; прочерк – нет или отрицательный.

Скорректированные выборки: 1: без Википедии, Youtube, Dzen; 3: без Apple, Youtube, Dzen; 4: без Википедии; 5: без Apple, Youtube.

му, например, можно верить статистике, которая говорит, что в Яндексе нули для топ-30 «больше», чем с виду такие же нули для фона.

Кстати, удивительно, насколько мало от этой таблицы отличается аналогич-

ная таблица для ссылающихся страниц (не говоря уже об уникальных ссылках на URL): в большей части клеток остаются ровно те же значения. Похоже, сквозные ссылки (и даже просто любые ссылки с нескольких страниц одного сайта) редко ведут на те стра-

ницы, которые находятся по запросам наших выборов.

Ссылки на конкретную страницу, хоть и не гарантируют успех, конечно, имеют значение.

Косвенно свидетельствует об этом еще один параметр: **доля ссылок на главную страницу** (см. таблицу на предыдущей странице). Он инвертированный: чем меньше его значение, тем лучше. Это логично: во-первых,

чем больше сайт, тем больше страниц, на которые кто-то может ссылаться; во-вторых, чем больше ссылок на внутренние страницы, тем, значит, больше на сайте контента, достойного того, чтобы на него поставили ссылку.

Для информационных выборок значения параметра ожидаемо ниже (а значит, лучше), чем для коммерческой. В Яндексе параметр сильнее: коэффициенты корреляции выше; контраст с фоном есть для всех выборок, тогда как в Google — только для *Гаджетов*, и то небольшой.

Ссылочно-текстовые параметры

У нас есть целое семейство ссылочно-текстовых параметров.

Они различаются:

- тем, какие ссылки учитываются — на страницу или на сайт (домен) в целом;
- тем, учитываются ли все ссылки, или только предполагаемые SEO-ссылки (с сайтов, на которых возможно платное размещение ссылок), или только «белые» (с прочих сайтов);
- тем, какая степень совпадения с запросом требуется:
 - точное вхождение запроса;
 - вхождение запроса целиком, но не обязательно в той же форме;
 - вхождение всех значимых слов в произвольном порядке;

- вхождение слов запроса «россыпью», включая ссылки, содержащие всего одно значимое слово из запроса;
- то же с учетом синонимов слов запроса;
- то же с учетом слов, выделенных Яндексом (или Google) в снипетах по запросу

Казалось бы, что важнее для ранжирования: ссылки с текстом запроса, ведущие на конкретную найденную страницу — или куда-нибудь, всё равно куда, на сайт? Разумеется, на конкретную страницу важнее, подумаем мы — и будем грубо неправы! Логика поисковых машин противоположная: конечно, сайт важнее. Сайтовые ссылочно-текстовые параметры (каки просто ссылочные) гораздо сильнее страничных.

Если нужна иллюстрация того, что в поисковых системах ранжируются не отдельные страницы, а целые сайты, пожалуй, лучшей, чем ссылочно-текстовые параметры, не найти!

«Белые» ссылки имеют примерно такой же вес, как и все ссылки в целом; «SEO-ссылки» слабее — возможно, потому, что по информационным запросам их просто мало.

Что касается оси степени совпадения с запросом — от точного запроса с сохранением порядка и форм слов до отдельных слов из него (и/или их синонимов или слов, выделенных в снипете), — то тут действует «принцип антибуквализма». Чем дальше от точного запроса, тем больше ссылок — и тем сильнее соответствующие параметры. Важнее не вхождения «ключевиков» в «анкоры», а тематическая релевантность сайта запросу. Для информационных запросов это оказалось так же справедливо, как и для коммерческих, см. об этом ниже, в разделе [Тексты входящих ссылок](#).

Ниже таблицы для одного из сайтовых и одного из страничных параметров. В обоих случаях это параметры, учитывающие любые ссылки (и «белые», и SEO), в которых встречаются слова запроса или их синонимы. Именно эти параметры учитывают больше всего ссылок (хотя в таблице ссылок на страницу всё равно почти везде нули) — и именно они оказываются самыми сильными в своих группах.

Вхождения в ссылки на сайт: слова запроса и синонимы

Выборка	С корр.	топ-3	4–10	11–20	21-30	топ-30	фон	Ф-30/фон
Яндекс								
Наука	0,14	2,43	1,37	0,71	0,46	0,66	0,78	—
Наука ⁻¹	0,14	3,90	1,00	0,68	0,44	0,60	0,78	—
Гаджеты	0,32	189	6,62	2,86	1,21	2,62	3,19	—
Гаджеты ⁻³	0,26	70,6	8,75	2,87	1,18	2,41	3,14	—
Станд. комм.	0,07	14,3	13,8	8,45	7,25	7,65	9,14	—
Google								
Наука	0,15	3,95	1,78	1,13	0,75	1,03	0,47	
Наука ⁻⁴	0,14	6,30	1,82	1,12	0,74	0,98	0,47	
Гаджеты	0,30	708	10,8	5,58	2,59	5,16	1,08	
Гаджеты ⁻⁵	0,22	45,1	11,6	5,81	2,58	4,54	1,04	
Станд. комм.	0,18	24,5	18,1	13,2	9,54	12,6	4,36	

С корр. — коэффициент корреляции с позицией (ранговый критерий Спирмена).

Ф-30/фон — оценка контраста между топ-30 и фоном (точный тест Фишера):

сильный, средний, слабый; прочерк — нет или отрицательный.

Скорректированные выборки: 1: без Википедии, Youtube, Dzen; 3: без Apple, Youtube, Dzen; 4: без Википедии; 5: без Apple, Youtube.

Числа в клетках таблицы показывают, сколько целых запросов с учетом веса слов (более редкие имеют больший вес, чем более частые) можно сложить из вхождений слов запроса в ссылки. Например, число 10 означает не то, что слова запроса вошли в 10 разных ссылок, а то, что их вхождений слов запроса (или их синонимов) в ссылки можно было бы составить 10 целых запросов — т. е. ссылок, скорее всего, значительно больше, чем 10.

Вхождения в ссылки на URL: слова запроса и синонимы

Выборка	С корр.	топ-3	4–10	11–20	21-30	топ-30	фон	Ф-30/фон
Яндекс								
Наука	0,10	0	0	0	0	0	0	
Наука ⁻¹	0,09	0	0	0	0	0	0	
Гаджеты	0,08	0,01	0	0	0	0	0	
Гаджеты ⁻³	0,11	0,11	0	0	0	0	0	
Станд. комм.	0,03	0	0	0	0	0	0	—
Google								
Наука	0,12	0	0	0	0	0	0	
Наука ⁻⁴	0,08	0	0	0	0	0	0	
Гаджеты	0,10	0	0	0	0	0	0	
Гаджеты ⁻⁵	0,10	0,05	0	0	0	0	0	
Станд. комм.	0,09	0,17	0	0	0	0	0	

С корр. — коэффициент корреляции с позицией (ранговый критерий Спирмена).

Ф-30/фон — оценка контраста между топ-30 и фоном (точный тест Фишера):

 сильный,  средний,  слабый; прочерк — нет или отрицательный.

Скорректированные выборки: 1: без Википедии, Youtube, Dzen; 3: без Apple, Youtube, Dzen; 4: без Википедии; 5: без Apple, Youtube.

Для информационных выборок ссылочно-текстовые параметры оказались сильнее, чем для стандартной коммерческой — особенно в Яндексе. Количество ссылок на сайт, имеющих отношение к запросу, в информационных выборках быстрее убывает с уда-

лением от первой позиции. О ссылках на страницу судить трудно ввиду их крайней малочисленности.

Для информационных выборок сработал также параметр, на который мы раньше не обращали внимания,

Доля содержательных ссылок на сайт

Выборка	С корр.	топ-3	4–10	11–20	21–30	топ-30	фон	Ф-30/фон
Яндекс								
Наука	0,11	0,66	0,57	0,55	0,54	0,56	0,61	—
Наука ⁻¹	0,04	0,58	0,57	0,55	0,54	0,55	0,61	—
Гаджеты	0,19	0,55	0,52	0,48	0,42	0,47	0,49	—
Гаджеты ⁻³	0,22	0,58	0,53	0,48	0,42	0,47	0,49	—
Станд. комм.	—	0,54	0,54	0,54	0,53	0,53	0,53	—
Google								
Наука	0,11	0,72	0,63	0,61	0,60	0,62	0,54	
Наука ⁻⁴	—	0,63	0,62	0,60	0,60	0,61	0,54	
Гаджеты	0,13	0,57	0,53	0,52	0,47	0,51	0,43	
Гаджеты ⁻⁵	0,15	0,58	0,54	0,52	0,47	0,51	0,43	
Станд. комм.	—	0,58	0,57	0,54	0,52	0,55	0,51	

С корр. — коэффициент корреляции с позицией (ранговый критерий Спирмена).

Ф-30/фон — оценка контраста между топ-30 и фоном (точный тест Фишера):

 сильный,  средний,  слабый; прочерк — нет или отрицательный.

Скорректированные выборки: 1: без Википедии, Youtube, Dzen; 3: без Apple, Youtube, Dzen; 4: без Википедии; 5: без Apple, Youtube.

т. к. на коммерческой выборке он себя не проявлял: **доля содержательных ссылок на сайт**. Понятно, что чем больше содержательных ссылок (т. е. чем меньше «безанкорных» — ведущих

с URL, с доменного имени, со слов типа *здесь* или *ссылка*), тем выше может быть значение ссылочно-текстовых факторов. В Яндексе есть корреляция с позицией, в Google коэффициенты

корреляции с позицией ниже, но все значения параметра выше и, кроме того, есть контраст с фоном.

Мы [еще вернемся](#) к ссылочно-текстовым параметрам после того, как рассмотрим текстовые параметры.

Исходящие ссылки

И раз уж речь зашла о ссылках, поговорим и о «ссылках наоборот» — исходящих со страниц, попавших в результаты поиска.

И для исходящих внутренних ссылок, и для исходящих внешних — чем их больше, тем лучше, и соответствующие параметры сильнее, чем для коммерческой выборки.

Количество исходящих внутренних ссылок, по-видимому, имеет отношение к размеру сайта и его «внутренней связности». Таких ссылок много — в среднем от 70 до 100 в топ-30 и 120–200 в топ-3; впрочем, это всё равно меньше, чем в коммерческой выборке с ее большими товарными рубриками и каталогом товаров.

Исходящих внешних ссылок на порядок меньше, чем внутренних. У веб-мастеров принято ставить внешние ссылки с опаской — их часто или вообще избегают, или закрывают атрибутом *nofollow*, чтобы не раздавать ссылочный вес. Тем не менее, объективные данные показывают, что исходящие ссылки скорее полезны — причем для информационных запросов в большей степени, чем для коммерческих. Наличие атрибута *nofollow* на пользу от ссылок, похоже, не влияет — впрочем, на информационных сайтах его ставят реже.

Количество исходящих внутренних ссылок

Выборка	С корр.	топ-3	4–10	11–20	21-30	топ-30	фон	Ф-30/фон
Яндекс								
Наука	0,11	192	91,3	81,7	78,5	82,7	73,9	
Наука ⁻¹	0,11	175	99,6	82,8	78,7	82,9	73,6	
Косметика	0,10	120	110	97,9	86,8	96,2	91,3	
Косметика ⁻²	0,15	133	117	100	87,5	98,9	91,7	
Гаджеты	—	76,4	69,3	72,3	69,0	68,0	79,5	—
Гаджеты ⁻³	0,07	99,4	79,5	73,0	69,1	71,1	79,5	—
Станд. комм.	—	278	293	273	266	268	208	
Google								
Наука	0,15	217	86,3	73,0	75,8	80,1	78,4	—
Наука ⁻⁴	0,06	165	81,6	72,5	75,3	74,9	77,6	—
Косметика	0,04	120	102	97,1	94,9	98,5	88,7	
Гаджеты	0,06	127	88,8	76,6	74,4	76,3	63,0	
Гаджеты ⁻⁵	0,12	132	103	82,4	75,3	80,9	63,8	
Станд. комм.	0,08	253	261	244	215	229	258	—

С корр. — коэффициент корреляции с позицией (ранговый критерий Спирмена).

Ф-30/фон — оценка контраста между топ-30 и фоном (точный тест Фишера):

сильный, средний, слабый; прочерк — нет или отрицательный.

Скорректированные выборки: 1: без Википедии, Youtube, Dzen; 2: без Youtube, Dzen; 3: без Apple, Youtube, Dzen; 4: без Википедии; 5: без Apple, Youtube.

Количество исходящих внешних ссылок

Выборка	С корр.	топ-3	4–10	11–20	21-30	топ-30	фон	Ф-30/фон
Яндекс								
Наука	0,14	20,6	7,29	5,99	5,31	6,00	6,40	—
Наука ⁻¹	0,11	14,6	8,31	6,17	5,31	5,99	6,35	—
Косметика	0,09	6,51	4,64	4,21	3,02	3,40	4,59	—
Косметика ⁻²	0,11	7,79	5,36	4,41	3,06	3,67	4,64	—
Гаджеты	0,07	6,71	5,88	5,09	3,83	4,56	5,32	—
Гаджеты ⁻³	0,13	8,69	7,36	5,23	3,85	5,03	5,33	—
Станд. комм.	0,04	6,32	5,77	5,46	5,24	5,43	5,35	—
Google								
Наука	0,17	21,5	8,51	6,59	6,33	7,41	4,80	
Наука ⁻⁴	0,08	12,4	8,01	6,50	6,29	6,72	4,67	
Косметика	0,14	12,9	5,50	4,61	3,93	4,57	2,82	
Гаджеты	0,09	10,1	6,14	5,78	5,17	5,65	3,37	
Гаджеты ⁻⁵	0,12	10,8	7,37	6,38	5,28	6,20	3,48	
Станд. комм.	0,08	8,04	6,19	5,77	5,08	5,66	4,75	

С корр. — коэффициент корреляции с позицией (ранговый критерий Спирмена).

Ф-30/фон — оценка контраста между топ-30 и фоном (точный тест Фишера):

 сильный,  средний,  слабый; прочерк — нет или отрицательный.

Скорректированные выборки: 1: без Википедии, Youtube, Dzen; 2: без Youtube, Dzen; 3: без Apple, Youtube, Dzen; 4: без Википедии; 5: без Apple, Youtube.

Количество исходящих внешних ссылок без nofollow

Выборка	С корр.	топ-3	4–10	11–20	21-30	топ-30	фон	Ф-30/фон
Яндекс								
Наука	0,12	11,4	4,96	4,06	3,59	3,99	4,80	—
Наука ⁻¹	0,09	10,8	5,65	4,20	3,59	3,99	4,76	—
Косметика	0,09	3,56	3,10	2,50	1,56	1,87	2,71	—
Косметика ⁻²	0,10	4,15	3,58	2,63	1,59	1,97	2,72	—
Гаджеты	—	4,02	2,92	2,60	1,94	2,01	2,60	—
Гаджеты ⁻³	0,08	5,20	3,97	2,66	1,95	2,30	2,60	—
Станд. комм.	0,04	4,73	3,24	3,01	3,25	2,98	3,07	—
Google								
Наука	0,15	13,7	6,42	4,89	4,72	5,53	3,05	
Наука ⁻⁴	0,08	10,1	5,95	4,82	4,67	4,91	2,97	
Косметика	0,14	7,76	3,41	2,81	2,24	2,74	1,44	
Гаджеты	0,09	7,07	3,64	2,93	2,84	2,69	1,43	
Гаджеты ⁻⁵	0,11	7,46	4,44	3,27	2,87	3,08	1,44	
Станд. комм.	0,04	5,33	3,87	3,28	3,32	3,39	2,51	

С корр. — коэффициент корреляции с позицией (ранговый критерий Спирмена).

Ф-30/фон — оценка контраста между топ-30 и фоном (точный тест Фишера):

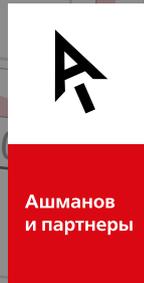
 сильный,  средний,  слабый; прочерк — нет или отрицательный.

Скорректированные выборки: 1: без Википедии, Youtube, Dzen; 2: без Youtube, Dzen; 3: без Apple, Youtube, Dzen; 4: без Википедии; 5: без Apple, Youtube.

Мы, пожалуй, затруднимся объяснить логику поисковых машин, когда они поощряют исходящие ссылки, — хотя в самом факте сомневаться не приходится. Для информационных сайтов (и страниц) могут быть важны ссылки на источники, которые добавляют страницам авторитетности и/или экспертности — но для коммерческих страниц это объяснение вряд ли подходит, а для них исходящие ссылки тоже «работают».

В Google параметры, связанные с исходящими внешними ссылками, сильнее, чем в Яндексе, т. к. здесь к корреляции с позицией — обычно даже более сильной — добавляется контраст с фоном; причем на страницах, не попавших в топ-30, оказывается в среднем меньше ссылок, чем на страницах из третьей десятки выдачи.

См. также ниже в разделе про текстовые параметры подраздел [Ссылки со страницы](#).



Ашманов
и партнеры

Текстовые параметры



Казалось бы, если речь идет об «информационке», что может быть важнее релевантности текста запросу?

Мы ожидали, что для информационных запросов текстовые параметры окажутся более сильными, чем для коммерческих, и будут работать как-то по-другому. Например, что будет более сильная корреляция с позицией, или что те параметры, которые для коммерческих сайтов себя не проявляют, окажутся сильными.

Эти ожидания не вполне оправдались: ранжирование оказалось более похожим на коммерческое, чем можно было предположить.

При этом наши три информационные выборки отличаются друг от друга, пожалуй, сильнее, чем некоторые из них — от стандартной коммерческой.

Поэтому давайте начнем с того, что нам за годы изучения удалось понять о логике текстового ранжирования в основном на коммерческом материале, а потом рассмотрим с этой точки зрения наши информационные выборки.

Матрица текстовых параметров

Мы уже девять лет изучаем текстовое ранжирование для нашей стандартной коммерческой выборки (для которой ежемесячно собираются результаты поиска) и для многих других.

См., в частности, общий [аналитический отчет 2019 года](#), где текстовые факторы разобраны особенно подробно, и тематические отчеты: [e-commerce](#), [финансы](#), [медицина](#), [автомобили](#) и [недвижимость](#).

Текстовые параметры, вместе со ссылочно-текстовыми, составляют почти половину всего нашего «арсенала» — а если считать только автоматические параметры, то больше половины. Они образуют матрицу: *что* считается и *где*. Соответственно, и результаты по ним мы обычно представляем в форме таблицы.

Что:

- точный запрос — с сохранением порядка и форм слов, без пропусков;
- запрос с точностью до форм;

- неточный запрос: все значимые слова запроса недалеко друг от друга в произвольном порядке;
- то же с возможностью замены слов запроса их синонимами;
- слова запроса россыпью: любая часть слов запроса в любом порядке; важно, что результат вычисляется всё равно в пересчете на полные запросы с учетом значимости слов (чем слово более редкое, тем оно значимее);
- то же с учетом синонимов;
- то же с учетом слов, выделяемых в снипетах по запросу Яндексом (или Google); эти слова часто связаны с предполагаемым интендом пользователя: *купить, в Москве, монтаж, своими руками* и т. п., могут включать также слова, родственные словам запроса, синонимы и др.;

- синонимы отдельно;
- выделенные слова отдельно;
- дополнительные технические параметры: входение всех слов запроса (да/нет), расстояние от начала текста или текстовой зоны до входения запроса (точного или неточного), концентрация слов запроса, [BM25](#) и т. п.; они считаются не по всем зонам, и здесь мы о них почти ничего говорить не будем.

Где:

- текст в целом;
- основной текст на странице (для коммерческих страниц — SEO-текст) vs. прочие фрагменты текста, включая различные текстовые блоки, меню и т. п.;
- заголовки: `<title>`; `<h1>`;
- домен и URL;
- внутренние заголовки: `<h2>`–`<h4>`;
- выделения в тексте: `/`; `<i>/`;
- ссылки внутренние (в пределах того же домена) и внешние (на другие сайты);
- списки (``);
- атрибуты тегов: `alt` и `title`;
- метатеги `description` и `keywords`;

- ссылачно-текстовые параметры: входящие ссылки на сайт и страницу, в том числе «серые» (с сайтов, на которых можно поставить ссылку за деньги) и «белые».

Разные выборки отличаются друг от друга тем, насколько плотно оказывается заполнена таблица *Что × Где*, насколько сильны для разных параметров корреляция с позицией и контраст с фоном.

Иногда бывает и отрицательная корреляция (с приближением к первой позиции значение параметра падает), и отрицательный контраст с фоном (значения для фона значимо выше, чем для топ-30). В таблицах такие случаи не отражены, поскольку нет оснований считать, что текстовые параметры могут быть инвертированными. Скорее всего, обратные зависимости возникают случайно (из-за доминирования в выдаче конкретных сайтов) или «наводятся» корреляциями с другими параметрами. В частности, отрицательный контраст с фоном может возникать просто потому, что в другом поисковике данный параметр сильнее.

На следующем развороте сводная таблица параметров для нашей стандартной выборки. Сравните с ней аналогичные таблицы для информационных выборок — скоро мы доберемся и до них.

Текстовые параметры: стандартная коммерческая выборка. Яндекс

Параметры	Выделенные	Слова запроса + выделенные	Синонимы	Слова запроса + синонимы	Слова запроса	Неточный запрос + синонимы	Неточный запрос	Запрос с точностью до форм	Точный запрос	Слова запроса (полных запросов)	
										Топ-3	Топ-30
Текст страницы		0,06	0,04	0,06	0,06	0,06	0,05	0,04	0,05	41,8	33,8
SEO-текст										6,09	6,30
Остальной текст (фрагменты)	0,05	0,07	0,05	0,06	0,06	0,06	0,05	0,05	0,05	35,5	27,8
<title>										0,99	0,97
<h1>	0,03								0,04	0,76	0,72
URL (кроме домена)									0,03	0,64	0,61
Домен										0,05	0,06
<h2>–<h4>		0,04		0,05	0,05	0,05	0,03	0,03	0,04	6,02	3,19
, 					0,03					0,55	0,76
<i>, 										0,08	0,16
Списки: 										4,75	5,75
Атрибут alt			0,03							4,16	5,99
Атрибут title	0,07	0,03	0,06			0,05			0,03	8,37	2,37
Внутренние ссылки		0,03	0,04	0,03	0,03	0,05	0,04	0,04	0,06	17,8	16,1
Внешние ссылки	0,03	0,04	0,04	0,04	0,04	0,03	0,03			0,27	0,21
Description										0,96	1,01
Keywords										0,45	0,77
Ссылки на сайт		0,12		0,07	0,08					10,7	5,35
Ссылки на страницу					0,03					0,00	0,00

– сильный, средний, слабый контраст между топ-30 и фоном (точный тест Фишера).

– сильная, средняя, слабая корреляция с позицией.

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

Текстовые параметры: стандартная коммерческая выборка. Google

Параметры	Выделенные	Слова запроса + выделенные	Синонимы	Слова запроса + синонимы	Слова запроса	Неточный запрос + синонимы	Неточный запрос	Запрос с точностью до форм	Точный запрос	Слова запроса (полных запросов)	
										Топ-3	Топ-30
Текст страницы	0,10	0,11		0,12	0,12	0,07	0,04			35,6	30,2
SEO-текст	0,06	0,07		0,06	0,06	0,04	0,03	0,03		8,06	8,34
Остальной текст (фрагменты)	0,09	0,10	0,04	0,11	0,11	0,06	0,04			28,2	23,1
<title>	0,07	0,04	0,05							1,12	1,05
<h1>	0,04		0,04							0,79	0,79
URL (кроме домена)										0,65	0,65
Домен				0,04	0,05					0,07	0,04
<h2>–<h4>	0,03			0,03	0,04	0,04	0,05	0,05	0,03	3,00	2,44
, 										0,48	0,71
<i>, 										0,20	0,17
Списки: 										6,84	5,99
Атрибут alt										5,73	4,04
Атрибут title										4,75	1,91
Внутренние ссылки	0,04	0,08		0,09	0,08	0,05				16,5	13,1
Внешние ссылки	0,03	0,06		0,05	0,05	0,03	0,04	0,03		0,33	0,10
Description	0,07									0,95	1,02
Keywords										0,52	0,75
Ссылки на сайт		0,22		0,18	0,18	0,08	0,08	0,08	0,07	18,0	8,01
Ссылки на страницу		0,11		0,09	0,10	0,10	0,10	0,10	0,10	0,17	0,00

– сильный, средний, слабый контраст между топ-30 и фоном (точный тест Фишера).

– сильная, средняя, слабая корреляция с позицией.

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

Логика текстового ранжирования

1. Принцип антибуквализма

Самые сильные параметры во всех зонах — обычно те, где учитываются любые вхождения слов запроса, в том числе разрозненные, а также их синонимы или слова, выделенные в снипетах.

Самые слабые — в наших таблицах им обычно соответствуют пустые клетки — это когда учитываются только вхождения запроса целиком, особенно если не допускается даже изменение формы слов (о важных исключениях см. ниже).

Поисковые машины давно не ищут в текстах запрос как он есть: чем дальше от точной формы запроса, тем сильнее параметры.

Архаичная практика насыщения текста «ключевиками» — да еще и, если получится, в исходной форме — потеряла смысл уже много лет назад. Современ-

ные поисковые машины понимают текстовую релевантность очень широко, учитывают не только непосредственно слова запроса, но и многие связанные с запросом слова. Причем не только те, для которых у нас есть специальные параметры (синонимы и выделенные в снипетах), но и другие, для которых параметров нет.

Мы много думали над тем, как сделать текстовые параметры, учитывающие слова, не входящие в запрос, но связанные с ним семантически, контекстно, тематически и т. п., но пока так и не нашли решения. Дело в том, что мы не умеем находить такие слова независимо от поисковых систем. Метод поиска «тематических» слов, полезных для повышения релевантности текста запросу, давно существует — и лучшую, с большим отрывом, его реализацию (из известных нам, конечно) можно найти в нашем же сер-

Пример раскраски текста в результатах LSI-анализа SearchLab

Оценка: 0,04 Уникальность: 100% Уникальность на сайте: 100,0% 9 конкурентов 11.03.2024 Яндекс
1 запрос: текстиль для отелей и гостиниц оптом

ТЕКСТ	РЕКОМЕНДАЦИИ	СТАТИСТИКА	«ЛИШНЕЕ»	КОНКУРЕНТЫ
<p>Title: Купить текстиль для гостиниц и отелей оптом в Москве от производителя Red Star</p> <p>Description: Каталог текстильных товаров для HoReCa от фабрики-производителя в Иваново Red Star ✅ Опт от 30000 рублей ✅ Собственное производство ✅ Пошив на заказ 🚚 Доставка по России</p> <p>H1: Текстиль для отелей и гостиниц оптом</p> <p>Ответственность за результат: Несем финансовые гарантии за качество в рамках договора поставки Помощь в подборе: Не знаете какой текстиль подойдет для вашей деятельности? Предложим наиболее оптимальные варианты Имеем огромный опыт: На рынке мы с 2012 года, за это время накоплен огромный опыт в решении самых нестандартных ситуаций Получить предложение</p>				

Цвет каждого слова соответствует тому, насколько часто оно встречается у конкурентов, попавших в топ-10 по указанным пользователем запросам. Светло-зеленым выделены слова, которые встречаются не менее чем у 50% конкурентов; ярко-красным — которые не встречаются совсем. Жирным шрифтом отмечены слова, входящие в запрос или выделенные в снипетах.

висе LSI-анализа на [SearchLab.ru](https://searchlab.ru). Попробуйте обязательно, если еще не пробовали — вы не только узнаете, каких слов в вашем тексте не хватает, но и увидите, что в нём лишнее. Проблема в том, что сам LSI-анализ опирается на результаты поиска — и поэто-

му делать на его основе параметры неинтересно. Получится замкнутый круг: параметры, учитывающие лексику со страниц, входящих в топ-10, будут сильными автоматически, но высокая корреляция с позицией на самом деле не будет ни о чём гово-

ритель, она такая «по построению». Приходится ограничиваться объективными данными: списками синонимов, которые собраны неза-

висимо от конкретных запросов, и выделением в снипетах, которое делают сами поисковики.

2. Яндекс: контраст с фоном. Google: корреляция с позицией

Для Яндекса характерны более высокие значения текстовых параметров, более сильный контраст с фоном, но при этом более слабая (и реже возникающая) корреляция с позицией. Для Google — значения ниже, но корреляция с позицией сильнее и чаще.

Картина примерно такая же, как с регионами в коммерческой выдаче: в Яндексе, по запросам, заданным из Москвы, в топ-30 попадают почти исключительно московские интернет-магазины; в Google в ней есть и магазины из других регионов, но чем ближе к первой позиции, тем меньше вероятность их встретить. Можно предположить, что магазины из других регионов — даже самые лучшие — просто не попадают в верхнюю когорту сайтов, которые претендуют на места в топе выдачи Яндекса. Заметим, что это вполне разумная стратегия: когда в московской или саратовской выдаче Гугла встречаешь магазин с доставкой

по Пензе или Вологде, трудно понять, что он здесь вообще делает.

Аналогичная гипотеза напрашивается и для текстового ранжирования. Судя по всему, в Яндексе текстовые факторы участвуют в предварительном отборе по релевантности. Те сайты, страницы которых набрали достаточное количество «очков», ранжируются уже по другим критериям — например, в соответствии с трафиковыми, коммерческими и т. п. факторами. А все прочие, «неудачники», переходят в следующую когорту в «хвосте» выдачи или вообще не удостоиваются чести попасть в доступную для пользователя часть SERP'a — список найденных сай-

тов при его пролистывании обрывается раньше, чем до них дойдет очередь.

Это не значит, что текстовые факторы в Яндексе слабые — или что они слабее, чем в Google. Как раз наоборот: недобрал текстовых факторов — и всё, ты во второй или третьей лиге, или вообще куришь в сторонке, пока другие делят места на первых страницах выдачи. Куда уж сильнее! Но всё равно: получается, что совсем не текстами и заголовками определяется в конечном счете распределение мест на пьедестале.

Собственно, именно эта особенность текстового ранжирования в Яндексе вызвала у нас особое беспокойство, когда мы взяли за информационные запросы.

Неужели, думали мы, и с ними окажется так же: текстовая релевантность — не главный критерий ранжирования, а только необходимая основа для него?

3. Пуд слов россыпью

А вот чтобы текстовая релевантность оказалась достаточной — это сколько, например, слов запроса должно быть в тексте?

Точно ответить на такие вопросы трудно, но обратите внимание на две правые колонки в наших матрицах параметров. В них — усредненная по запросам выборки оценка* количества слов

запроса в соответствующей зоне документа, в пересчете на целые запросы.

Для разных запросов и, соответственно, выборки получается по-разному, но вот для нашей ком-

* Технически это считается так: для каждого запроса считаем количество слов запроса на каждой из страниц, попавших в топ-30 (или топ-3) результатов поиска; пересчитываем в количество полных запросов с учетом весов слов (обратно пропорциональных их частоте, т. е. вес частотных слов уменьшается, редких увеличивает); берем медиану по запросу; считаем среднее по всем запросам.

мерческой выборки (см. таблицу) в Яндексе для текста в целом выходит больше чем по 40 запросов россыпью в топ-3 и больше чем по 30 — в топ-30! В Google меньше, но не принципиально: более 35 и тоже более 30.

Это прямо очень много!

Такое количество повторений во многом объясняется тем, что для запросов этой выборки стандартный ответ часто — товарная рубрика с большим количеством конкретных товарных позиций (по этой же причине так много вхождений слов запроса во внутренние ссылки). В других выборках — забегая вперед, скажем, что и в наших информационных тоже — результаты могут быть ниже, но даже если выходит не три-четыре десятка запросов россыпью, а полтора-два, это всё равно много.

Понятно, что в основе этой статистики лежат медианы по запросам, т. е. у половины сайтов вхождений всё-таки меньше — но ведь зато у второй половины еще больше. В общем, точные рамки указать трудно, но если не хочешь остаться за бортом, лучше ориентироваться на конкурентов,

попавших в топ, и не слишком от них отставать.

Увеличивать количество вхождений слов запроса можно двумя способами — интенсивным, увеличивая концентрацию нужных слов в тексте, или экстенсивным, увеличивая длину текста (разумеется, можно делать сразу и то, и другое). Наши данные показывают, что поисковики «поощряют» скорее экстенсивный способ: для **концентрации слов запроса в тексте страницы** обычно нет ни значимой корреляции с позицией, ни значимого контраста с фоном, тогда как **размер текста страницы** — обычно сильный параметр. Но это текст страницы в целом — в отдельных зонах может быть и иначе, см. об этом ниже.

4. Заголовки: всё наоборот

В заголовках, кстати, вхождений запроса тоже много — по крайней мере в коммерческой выборке.

Это может отчасти объясняться дисциплинированностью хозяев коммерческих сайтов — уж они-то не забудут напихать в **<title>** всего, что считают полезным, — но ведь и поисковые машины могли бы выбирать произвольные страницы, а выбирают те, где всё в полном соответствии с запросом.

В Яндексе на коммерческой выборке вхождений слов запроса в **<title>** в пересчете на целые запросы 0,97 в топ-30 и 0,99 в топ-3*. Это, конечно, порядочно — но это далеко не предел: у Google в топ-30 в **<title>** в среднем 1,05 запроса россыпью, а в топ-3 — 1,12. А если считать с синонимами, то в Яндексе набирается 1,18–1,2 запроса, а в Google — 1,25–1,34.

То есть по заполнению **<title>** словами запроса Google — вопреки написанному выше о принципе номер два —

заметно обгоняет Яндекс. Но это еще не самое главное.

Если смотреть на точные вхождения запросов (тут их так и хочется назвать «ключевиками»), то оказывается, что в топ-30 Яндекса они есть в 31% **title**'ов (в топ-3 примерно столько же, 32%), а в Google — на треть чаще, в 42% (в топ-3 40%)! Примерно такая же ситуация и с запросом с точностью до форм: 49% в Яндексе (в топ-3 тоже 49%) и 58% в Google (в топ-3 56%).

То есть принцип антибуквализма для заголовков в Google тоже грубо нарушается!

И наконец, никакой корреляции с позицией для точных запросов и близких

* Может создаться впечатление, что почти в каждый **title** входят все слова запроса — но на самом деле это так только в 80% случаев (в топ-3 тоже 80%), а еще почти 20% набираются за счет того, что некоторые слова встречаются в **title** более одного раза. В Google процент «бинго» в **title** выше, но до 100% всё равно не дотягивает: 87% в топ-30 и 86% в топ-3.

к ним параметров нет — только контраст с фоном, ситуация для Google не характерная!

Что касается Яндекса, то в нём почти ни для каких параметров для `<title>` не видно связи с попаданием в топ-30, т. е. значения для фона выше. Но это, скорее всего, не потому, что Яндекс не обращает внимания на входжения слов запроса в заголовки (иначе не было бы результатов, близких к единице), а потому, что сравниваем мы с Google, а в Google еще больше.

Лет пять назад явный контраст с фоном в Google был виден только для одного из параметров, связанных с `<title>`, — для точного запроса, см. [Факторы ранжирова-](#)

[ния — 2018](#). Но уже к концу 2019 года превышение над фоном в Google распространилось и на другие параметры для `<title>` — вплоть до слов запроса рассыпью с учетом синонимов. То есть «экспансия» параметров, связанных с заголовками, в Google шла в терминах наших таблиц справа налево, тогда как обычно, согласно принципу антибукализации, она идет слева направо.

Аналогичная ситуация с тегами `<h1>` — «видимыми» заголовками, которые часто похожи на `<title>`, но короче. Но, возможно, они «светят отраженным светом», т. е. параметры для `<h1>`, которые часто немного слабее аналогичных параметров для `<title>`, «работают» просто из-за высокой корреляции с последними.

5. Какие зоны текста важнее

Тут всё просто: самая главная зона для текстового ранжирования — текст страницы целиком.

Именно для параметров, оценивающих входжения запроса в полный видимый текст страницы*, извлекаемый из HTML-кода, в Google обычно

самая сильная корреляция с позицией, а в Яндексе — самый сильный контраст с фоном (плюс обычно и корреляция с позицией небольшая тоже есть).

* Выделяется весь текст, заключенный в HTML-теги в зоне `<body>`. Обычно пользователи видят именно его — но при этом часть такого текста может быть скрыта CSS-стилями или скриптами, что при выделении текста не учитывается.

Кроме текста в целом, мы смотрим на ряд текстовых зон в HTML-коде. Примерно половина из них не входят в текст страницы: `<title>`, `meta description`, `meta keywords`, атрибуты `title` и `alt`, а также доменное имя и оставшаяся часть URL (с учетом транслитерации). Вторая половина текстовых зон — это части текста страницы; получается, что они учитываются дважды: по отдельности и в его составе. Это заголовки `<h1>` и `<h2>`–`<h4>`, выделения жирностью (``, ``) и курсивом (`<i>`, ``), элементы списков (``), текст внутренних и внешних ссылок (`<a>`).

На все эти зоны в SEO принято обращать внимание еще с тех давних пор, когда позиции в поисковой выдаче определялись в первую очередь релевантностью текста запросу, — но сейчас влияние многих из них на ранжирование, похоже, существенно уменьшилось.

Далее описана ситуация для нашей стандартной выборки — ниже мы уви-

дим, что для информационных выборок некоторые зоны могут быть «сильнее».

Одно из исключений — заголовки `<title>`, для которых в Google виден сильный контраст с фоном, в том числе для запросов в точной форме (см. выше).

Уже роль `<h1>` кажется преувеличенной. Отсутствие этого заголовка, или наличие нескольких `<h1>` на одной странице, или появление других заголовков выше `<h1>` на ранжирование (по крайней мере, для нашей стандартной выборки) особого влияния не оказывает — впрочем, всё это случается весьма редко. Корреляции с позицией у вхождений в `<h1>` не видно ни в Яндексe, ни в Google. Есть контраст с фоном в Google, но он слабее, чем для `<title>`, и вполне может определяться с тем, что `<h1>` очень часто представляет собой сокращенную версию `<title>` (для `<h1>` действуют те же аномалии, что и для `<title>`, см. предыдущий подраздел). Тем не менее, мы пока не находим в себе смелости исключить «гигиену» `<h1>` из рекомендаций по поисковой оптимизации. Забегая вперед, отметим, что информационные выборки скорее подтверждают важность `<h1>`.

Для `<h2>`–`<h4>` (мы рассматриваем их вместе) в обоих поисковиках есть слабая корреляция с позицией — не более того. Выделения курсивом и жирностью «взлетают» редко и сейчас себя ничем не проявляют.

Остальные зоны могут быть более или менее «яркими» в зависимости от двух разных вещей:

- общей «насыщенности» проявления текстовых параметров по выборке в целом: чем сильнее параметры у текста, тем обычно сильнее и у них,
- и от структуры текста и важности в нём зоны, которую мы рассматриваем: например, если среди страниц, попадающих в результаты поиска, преобладают каталоги товаров, то сильными могут ока-

заться элементы списков `` (они часто оформляют товарные позиции), внутренние ссылки (на страницы отдельных товаров), атрибуты `title` и `alt` (у иллюстраций и/или клеток таблиц).

Отдельная история с доменной зоной: «работает» ли она — т. е. есть ли связь между результатами ранжирования и вхождением в домен транслитерированных слов запроса, — сильно зависит от выборки; чаще такой связи нет. Похожая ситуация и с оставшейся частью URL, которую на многих сайтах делают «человекочитаемой». Она обычно похожа на `<title>` и `<h1>` и с ними коррелирует; если они сильны, то часто сильна и она — при условии, что в выдаче мало главных страниц сайтов.

6. SEO-текст и борьба с текстовым спамом

Кроме формальных зон, выделяемых по HTML-коду, веб-страницу можно разделить и на содержательные части: основной связный текст с одной стороны и короткие «текстовые фрагменты» (меню, карточки товаров, перевязочные и другие информационные блоки) — с другой.

Чёткой границы между этими зонами нет: иногда основной текст сам разбит на две или на несколько частей, отделенных друг от друга рекламой или карточками товаров, а то и на небольшие блоки (как это часто бывает на «лендингах»). Как именно поисковые машины выделяют на разных страницах основной текст, разумеется, неизвестно — но это и не имеет большого значения: нюансы в проведении границ на отдельных сайтах не помешают общей картины. Мы используем для его выделения эвристический алгоритм — тот же, который работает в сервисе «Тургенев» при [проверке по URL](#)^{*}.

Это деление важно потому, что именно на основной текст страницы (он же SEO-текст) направлены средства борьбы с текстовым спамом — например, фильтр «Баден-Баден». Не случайно [запись в блоге Яндекса](#) от 23 марта 2017 года, сообщающая о вводе в действие этого алгоритма, начинается с примера SEO-текста^{**}.

Яндекс утверждал, во-первых, что, поскольку SEO-текст играет на коммерческих сайтах далеко не центральную роль, он его при ранжировании и так уже игнорирует — так что вебмастера зря стараются набивать его ключевиками. И во-вторых, что от «некаче-

* Для того, чтобы «погонять» алгоритм выделения SEO-текста, нужна только регистрация в «Тургеневе»: проверка во вкладке «Стилизация» традиционно бесплатная. Вводите URL, смотрите, какой текст с него поступил на проверку. В «Тургеневе» выделение SEO-текста можно корректировать, задавая CSS-дескрипторы; при анализе факторов ранжирования он выделяется полностью автоматически.

** Впрочем, точно так же начиналась и [запись](#) от 13 сентября 2011 г., сообщавшая о первой попытке Яндекса бороться с текстовым спамом.

ственных» SEO-текстов кроме отсутствия пользы может быть и прямой вред: пессимизация как отдельных страниц, так и сайта в целом.

Естественно, мы в Лаборатории поисковой аналитики проверили первое утверждение — и да, действительно оказалось, что текстовые параметры в Яндексе для SEO-текста слабее, чем для «текстовых фрагментов» (в Google, кстати, тоже, хотя и в меньшей степени).

По состоянию [на 2019 год](#) в Яндексе влияние SEO-текста на ранжирование в нашей стандартной коммерческой выборке вообще не прослеживалось, а для e-commerce даже было отрицательным. При этом для «фрагментов» текстовые параметры были сильнее, чем для полного текста. Сейчас на стандартной выборке SEO-текст по корреляции с позицией мало отличается от текста в целом, но вот контраст с фоном в Яндексе для большинства параметров пропадает. В Google корреляция с позицией у текстовых параметров для SEO-текста несколько слабее, чем для текстовых фрагментов и текста в целом.

Что касается второго утверждения, в проверке оно не нуждалось: то здесь, то там сайты теряли позиции в поиске и получали уведомления в Яндекс.Вебмастере; тексты на них действительно были не очень, хотя часто и у их конку-

рентов, не попавших под фильтр, было не лучше.

Но зато мы, изучив ситуацию и проанализировав сайты, «уехавшие в Баден-Баден», сделали сервис [«Тургенев»](#), который помогает оценить риск попадания под фильтр и, если он высок, исправить ситуацию.

Яндекс рассчитывает, насколько та или иная страница заслуживает «Баден-Бадена», как обычно — методом машинного обучения, на основе образцов, признанных текстовым спамом. Соответственно, по большому счету, объяснения, чем плоха та или иная страница, у Платона Щукина нет и быть не может — так алгоритм посчитал, а конкретнее сказать нельзя. И тут приходит на помощь «Тургенев», в котором всё разложено по полочкам. Конечно, это не совсем те критерии, которыми пользуется Яндекс (хотя они и основаны на имеющихся намеках от Яндекса и анализе сотен известных нам страниц и сайтов, попавших под фильтр), — но понять, чем плох текст, очень помогает. Кстати, именно поэтому прогонять через «Тургенев» полезно и информационные страницы, которым «Баден-Баден» в принципе не грозит.

Штрафные баллы за стилистику текста в «Тургеневе» собираются по капле за окрашенные, шаблонные, «канцелярские» и т. п.

Пример анализа текста в «Тургеневе»

The screenshot shows the 'Тургенев' (Turgenev) text analysis tool interface. The main text being analyzed is:
"В современном мире именная вышивка не потеряла своей актуальности. Она стала способом персонализации подарков и предметов интерьера. Вышивка имени на подушке, полотенце или сумке может стать замечательным подарком для близкого человека. Это не только красиво, но и практично – такие вещи легко узнаваемы и не потеряются среди других. Кроме того, именная вышивка может быть использована в качестве семейного артефакта, который передаётся из поколения в поколение. Вышитые имена на семейном стелдарте или скатерти могут стать символом единства и преемственности. Техника именной вышивки может варьироваться от простых крестиков до сложных многоцветных композиций. Современные мастера используют различные стили и методы, чтобы сделать каждую вышитую вещь по-настоящему уникальной. Используя разнообразие нитей – от шёлка до металлизированных – можно достичь удивительных эффектов и сделать вышивку настоящим произведением искусства. Именная вышивка – это не только красиво, но и значимо. Это способ сох..."

The interface includes a sidebar on the left with metrics: ОБЩИЙ РИСК (17), ПОВТОРЫ (5), СТИЛИСТИКА (10), ЗАПРОСЫ, ВОДНОСТЬ (1), and УДОБОЧИТАЕМОСТЬ (1). The top navigation bar contains: ПРОВЕРКА ТЕКСТА, ПРОВЕРКА ПО URL, API, ОБРАТНАЯ СВЯЗЬ, СПРАВКА, and a balance of 579,40 P. The right sidebar shows 'СТИЛИСТИКА' with 10 points and a score of 0.26, and 'ПОДСКАЗКИ' for 'предмет интерьера'.

выражения, каждое из которых само по себе может не составлять проблемы. Если таких капель набирается много, и если к ним добавляются проблемы с повторами, водностью, удобочитаемостью и т. п., то риск попасть под фильтры может быть оценен как высокий или даже критический. Такой текст лучше отредактировать ... или написать заново. Кстати, когда сервис «Тургенев» стал платным, мы решили сделать исключение для «Стилистики» — она и сейчас остается бесплатной. Просто мы очень не любим, когда

так пишут, и хотим, чтобы мир стал чуть-чуть лучше.

За всё время активного сбора информации о сайтах, попавших под «Б-Б», в 2017–2020 гг. мы не видели ни одного достоверного случая применения этого фильтра к информационным сайтам. Официально это прямо вроде бы не говорилось, но на наши вопросы, почему фильтр не применяется

к инфопомойкам, тексты на которых ничем не лучше, представители Яндекса в кулуарах отвечали, что это другое и что с ними они будут бороться отдельно, когда время придет. Вероятно, время, наконец, пришло: в конце 2023 года было объявлено о фильтре «Малопользительный контент».

Для информационных сайтов говорить о SEO-тексте не вполне корректно. Если на коммерческих сайтах текст часто можно просто не читать,

всё понятно и без него (фотографии и характеристики товаров говорят сами за себя), — то на информационных это, как-никак, основное содержание. Но деление на основной текст и текстовые фрагменты всё равно остается актуальным, и один из вопросов, на которые мы хотели получить ответ, когда занялись «информационкой», — какова роль основного текста в текстовом ранжировании.

А что с информационными запросами

В целом картина текстового ранжирования по трем информационным выборкам на удивление хорошо соответствует описанным выше принципам и мало отличается от того, что мы привыкли видеть для коммерческих запросов.

Правда, есть зоны текста — например, списки ``, выделения жирностью и курсивом — где текстовые параметры для информационных запросов заметно сильнее, чем для коммерческих.

У двух выборок из трех матрицы текстовых параметров более насыщенные, чем у стандартной коммерческой: больше заполненных клеток, чаще возникает и в среднем сильнее контраст с фоном (для Яндекса) и корреляции

с позицией (для Google и Яндекса). У третьей выборки матрица, наоборот, значительно менее насыщенная.

Интересно, что две «текстово-сильные» выборки — это *Наука* и *Косметика*, а «слабая» — *Гаджеты*, то есть в этом отношении две коммерческо-информационные выборки оказались резко противопоставлены друг другу. По текстовому ранжированию *Косметика* гораздо ближе к *Науке* и к стандартной коммерческой выборке, чем к *Гаджетам*. Причем в *Гаджетах* текстовые параметры слабее одновременно и в Яндексе, и в Google.

Объяснение слабости текстовых параметров в *Гаджетах* простое: по-видимому, дело в высоком уровне монополизации этой выборки. Когда факторы, связанные с размером сайта, трафиком и т. п., столь сильны, остальным приходится потесниться. При этом, несмотря на общую слабость текстового ранжирования, антибуквализм и другие общие принципы вполне актуальны и для *Гаджетов* — а некоторые из них проявляются здесь даже сильнее (например, контраст с фоном в Яндексе против корреляции с позицией в Google).

Пожалуй, мы ожидали от текстового ранжирования по информационным запросам большего — и даже несколько разочарованы. Казалось бы, что может быть на информационном сайте важнее текста? Но нет, текстовые параметры работают в целом так же, как для коммерческих запросов и сайтов. Высокая текстовая релевантность — условие необходимое, но отнюдь не достаточное.

В частности, в Яндексе, похоже, и для информационных запросов текстовая релевантность учитывается прежде всего на предварительных этапах ранжирования (когда отбираются кандидаты в «широкий топ»), а ее влияние на позицию не велико.

Для информационных запросов, как и для коммерческих, ранжирование — это не конкуренция текстов, а конкуренция сайтов, на которых есть достаточно релевантные тексты.

На трех следующих разворотах — сводные матрицы параметров для трех информационных выборок; их можно сравнить друг с другом и с приве-

Текстовые параметры: Наука. Яндекс

Параметры	Выделенные	Слова запроса + выделенные	Синонимы	Слова запроса + синонимы	Слова запроса	Неточный запрос + синонимы	Неточный запрос	Запрос с точностью до форм	Точный запрос	Слова запроса (полных запросов)		
										Топ-3	Топ-30	
Текст страницы					0,03	0,03					20,5	13,5
SEO-текст											11,3	12,1
Текст страницы без SEO-текста	0,04	0,07	0,04	0,07	0,07	0,07	0,06	0,04	0,04		8,76	3,22
<title>		0,06		0,08	0,08	0,08	0,08	0,07	0,06		0,79	0,72
<h1>		0,03		0,03	0,04	0,04	0,04	0,04	0,04		0,61	0,60
URL (кроме домена)											0,41	0,39
Домен											0,01	0,01
<h2>–<h4>											1,00	0,98
, 						0,03	0,03	0,03	0,03		0,98	0,93
<i>, 											0,41	0,36
Списки: 	0,05	0,07	0,06	0,09	0,09	0,08	0,08	0,07	0,06		2,81	0,37
Атрибут alt				0,04	0,04	0,03	0,03	0,03			0,31	0,03
Атрибут title	0,07	0,11	0,06	0,13	0,13	0,11	0,11	0,11	0,10		1,74	0,00
Внутренние ссылки	0,03	0,06	0,05	0,07	0,07	0,06	0,06	0,05	0,04		2,23	0,65
Внешние ссылки	0,08	0,12	0,07	0,13	0,13	0,10	0,10	0,09	0,08		1,06	0,29
Description											0,53	0,61
Keywords		0,04		0,03	0,04	0,04	0,04	0,03	0,03		0,36	0,31
Ссылки на сайт		0,16		0,14	0,14	0,08	0,08	0,07	0,05		1,78	0,45
Ссылки на страницу		0,10		0,10	0,10	0,09	0,08	0,08	0,06		0,00	0,00

– сильный, средний, слабый контраст между топ-30 и фоном (точный тест Фишера).

– сильная, средняя, слабая корреляция с позицией.

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

Текстовые параметры: Наука. Google

Параметры	Выделенные	Слова запроса + выделенные	Синонимы	Слова запроса + синонимы	Слова запроса	Неточный запрос + синонимы	Неточный запрос	Запрос с точностью до форм	Точный запрос	Слова запроса (полных запросов)	
										Топ-3	Топ-30
Текст страницы	0,11	0,15	0,05	0,16	0,16	0,08	0,07	0,05	0,03	19,4	10,3
SEO-текст	0,06	0,09	0,03	0,11	0,11	0,07	0,07	0,06	0,04	11,4	8,90
Текст страницы без SEO-текста	0,10	0,13	0,06	0,14	0,14	0,08	0,07	0,06	0,04	8,84	2,87
<title>		0,04		0,03	0,03	0,03				0,78	0,74
<h1>		0,04		0,03	0,03					0,68	0,60
URL (кроме домена)				0,07	0,06			0,06	0,04	0,48	0,37
Домен										0,01	0,01
<h2>–<h4>	0,06	0,08	0,04	0,08	0,09	0,05	0,05	0,05	0,04	0,91	0,63
, 	0,06	0,09	0,03	0,11	0,11	0,10	0,09	0,10	0,09	1,04	0,70
<i>, 	0,05	0,08	0,03	0,08	0,07	0,05	0,05	0,04		0,43	0,30
Списки: 	0,13	0,17	0,08	0,16	0,16	0,11	0,11	0,10	0,08	2,63	0,24
Атрибут alt	0,03	0,05	0,03	0,03	0,03					0,23	0,01
Атрибут title	0,13	0,09	0,08	0,09	0,09	0,04	0,04	0,04	0,04	1,32	0,00
Внутренние ссылки	0,10	0,13	0,05	0,13	0,14	0,07	0,07	0,06	0,05	2,49	0,55
Внешние ссылки	0,12	0,18	0,09	0,18	0,17	0,15	0,15	0,14	0,13	1,31	0,30
Description										0,44	0,52
Keywords										0,14	0,23
Ссылки на сайт		0,14		0,15	0,15	0,09	0,09	0,08	0,07	2,85	0,71
Ссылки на страницу		0,13		0,12	0,12	0,08	0,08	0,07	0,05	0,00	0,00

– сильный, средний, слабый контраст между топ-30 и фоном (точный тест Фишера).

– сильная, средняя, слабая корреляция с позицией.

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

Текстовые параметры: Косметика. Яндекс

Параметры	Выделенные	Слова запроса + выделенные	Синонимы	Слова запроса + синонимы	Слова запроса	Неточный запрос + синонимы	Неточный запрос	Запрос с точностью до форм	Точный запрос	Слова запроса (полных запросов)	
										Топ-3	Топ-30
Текст страницы		0,03								27,3	21,6
SEO-текст										13,3	12,9
Остальной текст (фрагменты)		0,05	0,03	0,04	0,04	0,03	0,03			13,4	8,87
<title>		0,06		0,07	0,06	0,03	0,03			0,82	0,72
<h1>										0,65	0,62
URL (кроме домена)										0,55	0,55
Домен										0,01	0,01
<h2>–<h4>		0,03		0,03	0,03					3,42	2,73
, 										0,98	0,93
<i>, 										0,22	0,21
Списки: 		0,03		0,03	0,03	0,04	0,04	0,03	0,03	5,81	3,62
Атрибут alt		0,03	0,03	0,03						3,65	2,01
Атрибут title										0,73	0,26
Внутренние ссылки	0,05	0,08	0,04	0,07	0,07	0,04	0,05	0,03		4,88	2,92
Внешние ссылки	0,03	0,07		0,06	0,07					0,07	0,05
Description										0,80	0,79
Keywords		0,05		0,06	0,06	0,05	0,05	0,03	0,03	0,49	0,42

– сильный, средний, слабый контраст между топ-30 и фоном (точный тест Фишера).

– сильная, средняя, слабая корреляция с позицией.

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

Текстовые параметры: Косметика. Google

Параметры	Выделенные	Слова запроса + выделенные	Синонимы	Слова запроса + синонимы	Слова запроса	Неточный запрос + синонимы	Неточный запрос	Запрос с точностью до форм	Точный запрос	Слова запроса (полных запросов)	
										Топ-3	Топ-30
Текст страницы	0,04	0,07	0,04	0,07	0,07	0,04	0,03			23,7	19,4
SEO-текст	0,03	0,06	0,04	0,06	0,06	0,03	0,03			12,7	12,2
Остальной текст (фрагменты)	0,04	0,06	0,03	0,06	0,06	0,04	0,04			11,2	7,74
<title>		0,05		0,04	0,04					0,77	0,73
<h1>		0,04		0,03	0,04					0,66	0,63
URL (кроме домена)				0,06	0,06					0,62	0,56
Домен										0,01	0,01
<h2>-<h4>	0,06	0,10	0,04	0,11	0,10	0,07	0,06	0,03	0,03	3,36	2,44
, 		0,04		0,03	0,03					0,74	0,91
<i>, 										0,24	0,21
Списки: 	0,04	0,08		0,08	0,08	0,07	0,06	0,04	0,03	4,66	3,00
Атрибут alt	0,05	0,07	0,07	0,06	0,06					4,26	1,53
Атрибут title										0,32	0,18
Внутренние ссылки	0,04	0,04		0,05	0,05	0,05	0,05	0,04	0,03	4,12	2,64
Внешние ссылки	0,04	0,10		0,09	0,10					0,09	0,05
Description	0,04	0,06	0,03	0,06	0,05	0,03				0,86	0,81
Keywords		0,05		0,06	0,06					0,57	0,46

– сильный, средний, слабый контраст между топ-30 и фоном (точный тест Фишера).

– сильная, средняя, слабая корреляция с позицией.

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

Текстовые параметры: Гаджеты. Яндекс

Параметры	Выделенные	Слова запроса + выделенные	Синонимы	Слова запроса + синонимы	Слова запроса	Неточный запрос + синонимы	Неточный запрос	Запрос с точностью до форм	Точный запрос	Слова запроса (полных запросов)		
										Топ-3	Топ-30	
Текст страницы											15,5	14,3
SEO-текст											10,1	10,7
Остальной текст (фрагменты)											7,15	5,18
<title>											0,61	0,61
<h1>											0,47	0,56
URL (кроме домена)											0,37	0,48
Домен				0,07	0,07						0,05	0,02
<h2>–<h4>											2,14	2,07
, 											0,94	0,90
<i>, 	0,03	0,04	0,04	0,03							0,44	0,25
Списки: 											3,23	2,40
Атрибут alt								0,03			1,52	0,96
Атрибут title	0,03	0,07	0,03	0,07	0,06	0,05	0,04	0,03	0,03		1,14	0,38
Внутренние ссылки											2,42	1,90
Внешние ссылки		0,03		0,04	0,03						1,27	0,28
Description											0,63	0,59
Keywords	0,07	0,13	0,11	0,14	0,13	0,09	0,08	0,05	0,05		0,71	0,37
Ссылки на сайт		0,29		0,32	0,24	0,09	0,07				93,0	1,39
Ссылки на страницу		0,08		0,09	0,08	0,05	0,05				0,00	0,00

– сильный, средний, слабый контраст между топ-30 и фоном (точный тест Фишера).

– сильная, средняя, слабая корреляция с позицией.

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

Текстовые параметры: Гаджеты. Google

Параметры	Выделенные	Слова запроса + выделенные	Синонимы	Слова запроса + синонимы	Слова запроса	Неточный запрос + синонимы	Неточный запрос	Запрос с точностью до форм	Точный запрос	Слова запроса (полных запросов)	
										Топ-3	Топ-30
Текст страницы				0,03						18,3	13,6
SEO-текст										10,6	9,08
Остальной текст (фрагменты)				0,05	0,03					8,29	5,35
<title>	0,04		0,03							0,61	0,63
<h1>										0,56	0,57
URL (кроме домена)										0,43	0,50
Домен				0,09	0,08					0,06	0,02
<h2>–<h4>		0,04		0,05	0,04					2,59	1,84
, 		0,05	0,03	0,06	0,05	0,04	0,03			1,58	0,92
<i>, 	0,03	0,06	0,03	0,06	0,06	0,03	0,03			0,41	0,23
Списки: 	0,04	0,03		0,04	0,03					4,06	2,34
Атрибут alt	0,04									1,75	0,67
Атрибут title	0,08	0,08	0,09	0,10	0,07	0,05	0,03			1,23	0,36
Внутренние ссылки				0,03						3,06	2,02
Внешние ссылки		0,05		0,05	0,05			0,03	0,03	1,62	0,27
Description										0,47	0,56
Keywords	0,09	0,11	0,09	0,11	0,10	0,08	0,08	0,05	0,05	0,43	0,46
Ссылки на сайт		0,25		0,30	0,23	0,07	0,06	0,03	0,03	275	2,78
Ссылки на страницу		0,10		0,10	0,10	0,06	0,05	0,03	0,03	0,00	0,00

– сильный, средний, слабый контраст между топ-30 и фоном (точный тест Фишера).

– сильная, средняя, слабая корреляция с позицией.

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

денными выше данными для стандартной коммерческой выборки. А дальше мы подробнее остановимся на основных группах текстовых параметров.

Мы долго сомневались, какие таблицы здесь привести — для полных выборок или для скорректированных (где они есть). Они довольно сильно отличаются друг от друга: в Яндексе параметры для скорректированной выборки обычно сильнее, чем для

исходной (за исключением нескольких случаев явной «наведенки»); в Google различия разнонаправленные. Как и следовало ожидать, особенно велики различия для *Гаджетов* с их высокой монополизацией: в таблицах для скорректированной выборки меньше пустых клеток. В результате всё-таки решили показать полные выборки, а скорректированные данные привести ниже в таблицах для отдельных групп параметров.

Текст страницы

Как нам уже известно, текст страницы — главная и самая сильная зона текстового ранжирования. Здесь должен действовать принцип антибуквализма; в Яндексе должен преобладать контраст с фоном, а в Google — корреляция с позицией; должно быть довольно много вхождений запроса.

Всё это мы так или иначе видим на всех информационных выборках.

Для Яндекса есть контраст с фоном, который сильнее всего на самых общих параметрах (там, где считаются любые, даже разрозненные, вхождения слов запроса, а также их синонимов и слов, выделенных в снипетах) — а при приближении к правому краю таблицы (точные вхождения запроса) сходит

на нет. Корреляции с позицией почти нет — она становится заметной только на скорректированных выборках в *Науке* и *Косметике*, но и здесь коэффициент корреляции не превышает 0,06.

В Яндексе текстовые параметры на полных выборках слабее, чем на скорректированных — то есть сайты-лидеры побеждают скорее не благодаря, а вопреки текстовому ранжированию.

Для Google важнее корреляция с позицией, для которой также выполняется принцип антибуквализма, а контраста с фоном почти не видно*. Корреляция сильная в *Науке*, слабее в *Косметике*; в *Гаджетах* на полной выборке она теряется, но на скорректированной вполне заметна (судя по всему, *Apple.com* и *Youtube* попадают на высокие позиции не за выдающуюся текстовую релевантность). Интересно, что в *Науке*, в отличие от *Гаджетов*, текстовые параметры в Google сильнее для полной выборки, чем для скорректированной.

Количество **вхождений слов запроса в текст страницы** в информационных выборках меньше, чем в стандартной коммерческой: в пересчете на полные запросы получается в среднем** 13–21 в Яндексe и 10–19 в Google против 30+ в стандартной выборке.

Но это всё равно довольно много — не один-два запроса и даже не пять-шесть.

Меньше всего в *Науке* (вероятно, из-за минимальной поисковой оптимизации), больше всего в *Косметике*.

То, что на информационных (то есть текстовых) страницах слов запроса оказывается в среднем меньше, чем на коммерческих, где тексты обычно короче, не должно удивлять: в коммерческой выборке большая часть вхождений находится не в тексте как таковом, а в названиях и описаниях товаров, в меню и т. п.

В этом разделе вместо таблиц с подробной информацией по одному параметру мы будем приводить матрицы, во многом аналогичные полным матрицам текстовых параметров, которые вы видели выше. По вертикали в них тоже разные способы подсчета, а вот по горизонтали вместо разных зон страницы — разные выборки. Разумеется, это относится только к собственно текстовым (а также ссылочно-текстовым) параметрам, которых на каждую зону по несколько. Другие параметры — такие, например, как наличие или длина тех или иных заголовков, — мы будем иллюстрировать обычными таблицами.

* Там, где контраст с фоном всё-таки есть, это обусловлено скорее низкими значениями в Яндексe, чем высокими в Google.

** Технически эти числа — результат усреднения медиан для запросов по всем запросам выборки.

Текстовые параметры: текст страницы

Яндекс

Параметры	Выделенные	Слова запроса + выделенные	Синонимы	Слова запроса + синонимы	Слова запроса	Неточный запрос + синонимы	Неточный запрос	Запрос с точностью до форм	Точный запрос	Слова запроса (полных запросов)		
										Топ-3	Топ-30	
Наука											20,5	13,5
Наука ⁻¹		 0,04		 0,06	 0,06	 0,04	 0,04	 0,03			17,9	13,7
Косметика		 0,03									27,3	21,6
Косметика ⁻²		 0,04	 0,06	 0,04	 0,06	 0,06	 0,03	 0,03			29,7	22,2
Гаджеты											15,5	14,3
Гаджеты ⁻³											21,5	15,0
Станд. комм.		 0,06	 0,04	 0,06	 0,06	 0,06	 0,05	 0,04	 0,05		41,8	33,8

Google

Параметры	Выделенные	Слова запроса + выделенные	Синонимы	Слова запроса + синонимы	Слова запроса	Неточный запрос + синонимы	Неточный запрос	Запрос с точностью до форм	Точный запрос	Слова запроса (полных запросов)		
										Топ-3	Топ-30	
Наука	 0,11	 0,15	 0,05	 0,16	 0,16	 0,08	 0,07	 0,05	 0,03		19,4	10,3
Наука ⁻⁴	 0,07	 0,10	 0,03	 0,11	 0,11	 0,05	 0,05	 0,03			15,0	9,95
Косметика	 0,04	 0,07	 0,04	 0,07	 0,07	 0,04	 0,03				23,7	19,4
Гаджеты				 0,03							18,3	13,6
Гаджеты ⁻⁵	 0,06	 0,06	 0,03	 0,09	 0,06	 0,03					19,5	14,3
Станд. комм.	 0,10	 0,11		 0,12	 0,12	 0,07	 0,04				35,6	30,2

– сильный, средний, слабый контраст между топ-30 и фоном (точный тест Фишера).

– сильная, средняя, слабая корреляция с позицией.

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

Скорректированные выборки: 1: без Википедии, Youtube, Dzen; 2: без Youtube, Dzen; 3: без Apple, Youtube, Dzen; 4: без Википедии; 5: без Apple, Youtube.

Основной текст и «фрагменты»

Для информационных страниц некорректно говорить о SEO-тексте — да и Яндекс не применяет к ним фильтр «Баден-Баден». Означает ли это, что основной текст на информационных страницах важнее для ранжирования, чем окружающие его «текстовые фрагменты»?

В Google определено *не* означает.

Как и для стандартной коммерческой выборки, для информационных выборок в Google обычно сильнее всего параметры, относящиеся к тексту страницы в целом; за ними (с небольшим отставанием) следуют параметры для текстовых фрагментов — а вот параметры для основного текста (как и для SEO-текста в стандартной выборке) отстают и от тех, и от других довольно сильно. И это несмотря на то, что в информационных выборках доля текстовых фрагментов на странице значительно ниже, чем в коммерческих, а основного текста — значительно выше.

В Яндексе ситуация еще интереснее. Там однозначно сказать, какая часть текста сильнее, вообще не получается: картина распадается на две составляющие.

«Профильный» для Яндекса контраст с фоном для основного текста сильнее — и по сравнению с фрагментами, и по сравнению с текстом в целом.

А вот корреляции с позицией для основного текста нет вообще нигде — зато для фрагментов она сильнее, чем для текста в целом.

Получается, что гипотезу о том, что в Яндексе текстовая релевантность учитывается преимущественно на предварительных этапах ранжирования, для информационных запросов можно уточнить: это верно в первую очередь для основного текста страницы, а вот роль остального текста

Текстовые параметры: Основной текст

Яндекс

Параметры	Выделенные	Слова запроса + выделенные	Синонимы	Слова запроса + синонимы	Слова запроса	Неточный запрос + синонимы	Неточный запрос	Запрос с точностью до форм	Точный запрос	Слова запроса (полных запросов)	
										Топ-3	Топ-30
Наука										11,3	12,1
Наука ⁻¹										11,7	12,4
Косметика										13,3	12,9
Косметика ⁻²										14,3	13,3
Гаджеты										10,1	10,7
Гаджеты ⁻³										12,8	11,2
Станд. комм.										6,09	6,3

Google

Параметры	Выделенные	Слова запроса + выделенные	Синонимы	Слова запроса + синонимы	Слова запроса	Неточный запрос + синонимы	Неточный запрос	Запрос с точностью до форм	Точный запрос	Слова запроса (полных запросов)	
										Топ-3	Топ-30
Наука	0,06	0,09	0,03	0,11	0,11	0,07	0,07	0,06	0,04	11,4	8,9
Наука ⁻⁴	0,05	0,07		0,08	0,08	0,06	0,06	0,04	0,03	10,3	8,73
Косметика	0,03	0,06	0,04	0,06	0,06	0,03	0,03	0,03	0,03	12,7	12,2
Гаджеты								0,03		10,6	9,08
Гаджеты ⁻⁵	0,04	0,03		0,04	0,04			0,03	0,03	11,2	9,52
Станд. комм.	0,06	0,07	0,03	0,06	0,06	0,04	0,03	0,03	0,03	8,06	8,34

– сильный, средний, слабый контраст между топ-30 и фоном (точный тест Фишера).

– сильная, средняя, слабая корреляция с позицией.

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

Скорректированные выборки: 1: без Википедии, Youtube, Dzen; 2: без Youtube, Dzen; 3: без Apple, Youtube, Dzen; 4: без Википедии; 5: без Apple, Youtube.

Текстовые параметры: Текстовые фрагменты

Яндекс

Параметры	Выделенные	Слова запроса + выделенные	Синонимы	Слова запроса + синонимы	Слова запроса	Неточный запрос + синонимы	Неточный запрос	Запрос с точностью до форм	Точный запрос	Слова запроса (полных запросов)	
										Топ-3	Топ-30
Наука	0,04	0,07	0,04	0,07	0,07	0,07	0,06	0,04	0,04	8,76	3,22
Наука ⁻¹	0,07	0,07	0,03	0,07	0,08	0,07	0,06	0,04	0,03	6,13	3,25
Косметика	0,05	0,03	0,04	0,04	0,03	0,03	0,03	0,03	0,03	13,4	8,87
Косметика ⁻²	0,04	0,09	0,04	0,09	0,09	0,05	0,05	0,03	0,03	15,1	9,27
Гаджеты	0,03	0,03	0,05	0,03	0,03	0,03	0,03	0,03	0,03	7,15	5,18
Гаджеты ⁻³	0,03	0,03	0,05	0,03	0,03	0,03	0,03	0,03	0,03	10,3	5,50
Станд. комм.	0,05	0,07	0,05	0,06	0,06	0,06	0,05	0,05	0,05	35,5	27,8

Google

Параметры	Выделенные	Слова запроса + выделенные	Синонимы	Слова запроса + синонимы	Слова запроса	Неточный запрос + синонимы	Неточный запрос	Запрос с точностью до форм	Точный запрос	Слова запроса (полных запросов)	
										Топ-3	Топ-30
Наука	0,10	0,13	0,06	0,14	0,14	0,08	0,07	0,06	0,04	8,84	2,87
Наука ⁻⁴	0,04	0,06	0,04	0,07	0,07	0,03	0,03	0,03	0,03	5,14	2,65
Косметика	0,04	0,06	0,03	0,06	0,06	0,04	0,04	0,04	0,04	11,2	7,74
Гаджеты	0,05	0,03	0,05	0,11	0,07	0,04	0,03	0,03	0,03	8,29	5,35
Гаджеты ⁻⁵	0,07	0,07	0,05	0,11	0,07	0,04	0,03	0,03	0,03	8,77	5,71
Станд. комм.	0,09	0,10	0,04	0,11	0,11	0,06	0,04	0,04	0,04	28,2	23,1

— сильный, средний, слабый контраст между топ-30 и фоном (точный тест Фишера).

— сильная, средняя, слабая корреляция с позицией.

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

Скорректированные выборки: 1: без Википедии, Youtube, Dzen; 2: без Youtube, Dzen; 3: без Apple, Youtube, Dzen; 4: без Википедии; 5: без Apple, Youtube.

в окончательном ранжировании может быть более высокой.

Заметим, что текстовые фрагменты, окружающие основной текст, часто имеют большее отношение не к самой найденной странице, а к другим страницам сайта и косвен-

но — к сайту в целом: ссылки в меню, перекрестные блоки, анонсы подчиненных страниц (подразделов) и т. п.; см. также ниже главу об [исходящих внутренних ссылках](#). Это хорошо согласуется с общим принципом: ранжируются не страницы, а сайты.

Заголовки <title> и <h1>

С заголовками, как и следовало ожидать, всё наоборот: в Google контраст с фоном, причем (вопреки принципу антибукализма) начиная от самых точных вхождений запроса, а в Яндексе корреляция с позицией.

В Google (здесь, конечно, логичнее начать с него) во всех трех информационных выборках контраст с фоном есть не только для **точных вхождений запроса в <title>**, но и для других параметров, вплоть до отдельных слов запроса **с учетом синонимов или слов, выделенных в снипетах**. Но это не специфика «информационки» — так же сейчас и в коммерческих выборках.

Раньше, оценивая контраст с фоном, мы сравнивали топ-30 Google с не попавшими в него результатами не только от Яндекса, но и от Mail.ru, и тогда контраст с фоном не уходил так далеко влево по столбцам матрицы. Увы, поиск Mail.ru уже давно закрыт, остался только Яндекс. Возможно, отчасти дело в том, что «обогнать» его Гуглу проще,

чем «сборную» из отвергнутых им результатов двух поисковых машин. Впрочем, мы посмотрели динамику нескольких текстовых параметров — и почти нигде не заметили резких изменений в момент «выключения» Mail.ru.

Корреляция с позицией в Google если и есть, то слабая (почти всегда до 0,05) и против принципа антибукализма не бунтующая.

В Яндексе для двух выборок из трех есть корреляция с позицией — причем более сильная, чем в Google. В *Науке* она не только доходит до правого края матрицы, но и остается там более или менее сильной: 0,06 для **точных вхождений запроса**. В *Косметике* корреля-

ция с позицией дополнена контрастом с фоном сразу в нескольких колонках в левой части таблицы — относительно редкий случай, когда он есть одновременно и в Яндексе, и в Google.

По ситуации с текстовыми параметрами для **<title>** *Гаджеты* очень похожи на стандартную выборку: в Google почти везде только контраст с фоном; в Яндексе — ничего*. А вот в *Науке* и *Косметике* строки таблицы и для Яндекса, и для Google заполнены гораздо сильнее, хотя и немного по-разному.

Вообще, если в одной из поисковых машин какие-то параметры для выборки А сильнее, чем для выборки В, то и в другой картина, скорее всего, окажется похожей — хотя «сила» параметров может проявляться по-разному.

В чём все информационные выборки отстают от стандартной коммерческой — так это по количеству вхождений слов запроса в **<title>**.

Если там **отдельные слова запроса в <title>** в пересчете на целые запросы в топ-30 Яндекса дают почти единицу (0,97), а в Google — даже больше (1,05), то в информационных выборках набирается только от 0,6 до 0,75 целого запроса. Меньше всего в *Гаджетах*, примерно одинаково в *Косметике* и *Науке* — возможно, именно с этим и связана большая сила параметров для **<title>** в этих выборках.

Впрочем, если смотреть на количество **вхождений точных запросов в <title>**, картина оказывается несколько другой. Для *Науки* результат вполне сравним со стандартной коммерческой выборкой: 0,28 против 0,31 в Яндексе и 0,34 против 0,42 в Google. А вот в *Косметике* и *Гаджетах* результаты гораздо ниже — от 0,04 до 0,07. Объяснение весьма прозаическое и с различиями между информационными и коммерческими запросами прямо не связанное: просто в *Науке*, как и в стандартной выборке, запросы в среднем сильно короче, чем в *Косметике* и *Гаджетах*. Кстати, этим же можно объяснить и то, что корреляция с позицией в *Науке* заходит так далеко вправо по таблице.

Заголовки <h1> — как обычно, «тень» **<title>**. Они часто совпадают, но иногда в **<h1>** попадает укороченная версия, без повторов и тавтологий, которые для **<title>** характерны и прости-

* Как всегда, это не означает, что вхождения в **<title>** вообще не влияют на отбор релевантных результатов поиска — просто в Google таких вхождений немного больше.

Текстовые параметры: <title>

Яндекс

Параметры	Выделенные	Слова запроса + выделенные	Синонимы	Слова запроса + синонимы	Слова запроса	Неточный запрос + синонимы	Неточный запрос	Запрос с точностью до форм	Точный запрос	Слова запроса (полных запросов)	
										Топ-3	Топ-30
Наука		 0,06		 0,08	 0,08	 0,08	 0,08	 0,07	 0,06	0,79	0,72
Наука ⁻¹		 0,06		 0,07	 0,07	 0,07	 0,06	 0,05	 0,04	0,81	0,72
Косметика		 0,06		 0,07	 0,06	 0,03	 0,03			0,82	0,72
Косметика ⁻²		 0,05		 0,06	 0,06	 0,03	 0,03			0,83	0,72
Гаджеты										0,61	0,61
Гаджеты ⁻⁵										0,60	0,61
Станд. комм.										0,99	0,97

Google

Параметры	Выделенные	Слова запроса + выделенные	Синонимы	Слова запроса + синонимы	Слова запроса	Неточный запрос + синонимы	Неточный запрос	Запрос с точностью до форм	Точный запрос	Слова запроса (полных запросов)	
										Топ-3	Топ-30
Наука		 0,04		 0,03	 0,03	 0,03				0,78	0,74
Наука ⁻⁴		 0,07		 0,05	 0,05	 0,04	 0,04			0,84	0,75
Косметика		 0,05		 0,04	 0,04					0,77	0,73
Гаджеты	 0,04		 0,03							0,61	0,63
Гаджеты ⁻⁵	 0,03									0,62	0,63
Станд. комм.	 0,07	 0,04	 0,05							1,12	1,05

– сильный, средний, слабый контраст между топ-30 и фоном (точный тест Фишера).

– сильная, средняя, слабая корреляция с позицией.

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

Скорректированные выборки: 1: без Википедии, Youtube, Dzen; 2: без Youtube, Dzen; 3: без Apple, Youtube, Dzen; 4: без Википедии; 5: без Apple, Youtube.

Текстовые параметры: <h1>

Яндекс

Параметры	Выделенные	Слова запроса + выделенные	Синонимы	Слова запроса + синонимы	Слова запроса	Неточный запрос + синонимы	Неточный запрос	Запрос с точностью до форм	Точный запрос	Слова запроса (полных запросов)	
										Топ-3	Топ-30
Наука		0,03		0,03	0,04	0,04	0,04	0,04	0,04	0,61	0,60
Наука ⁻¹		0,07		0,08	0,08	0,07	0,07	0,05	0,04	0,68	0,62
Косметика										0,65	0,62
Косметика ⁻²		0,04		0,04	0,04					0,69	0,64
Гаджеты										0,47	0,56
Гаджеты ⁻³										0,59	0,59
Станд. комм.	0,03								0,04	0,76	0,72

Google

Параметры	Выделенные	Слова запроса + выделенные	Синонимы	Слова запроса + синонимы	Слова запроса	Неточный запрос + синонимы	Неточный запрос	Запрос с точностью до форм	Точный запрос	Слова запроса (полных запросов)	
										Топ-3	Топ-30
Наука		0,04		0,03	0,03					0,68	0,60
Наука ⁻⁴		0,04		0,03	0,03					0,69	0,60
Косметика		0,04		0,03	0,04					0,66	0,63
Гаджеты										0,56	0,57
Гаджеты ⁻⁵	0,04									0,59	0,60
Станд. комм.	0,04		0,04							0,79	0,79

— сильный, средний, слабый контраст между топ-30 и фоном (точный тест Фишера).

— сильная, средняя, слабая корреляция с позицией.

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

Скорректированные выборки: 1: без Википедии, Youtube, Dzen; 2: без Youtube, Dzen; 3: без Apple, Youtube, Dzen; 4: без Википедии; 5: без Apple, Youtube.

тельны: всё-таки **<h1>** всегда на виду, а **<title>** на глаза попадается редко и оптимизировать его можно до неприличия. Поэтому длина **<h1>** и количество вхождений запросов в этот заголовок во всех выборках заметно меньше.

Если не считать последние две колонки, где указано количество вхождений, таблица для **<h1>** очень мало отличается от **<title>**. Большая часть параметров слабее (в частности, у полной *Косметики* в Яндексе пропадает корреляция с позицией), скорректированная *Наука* в Яндексе чуть сильнее, но в целом очень похоже.

Можем ли мы из того, что в Яндексе параметры, связанные с **<title>** и **<h1>**, никак себя не проявляют (*Гаджеты*, стандартная коммерческая выборка) или проявляют слабо (*Наука*, особенно с **<h1>**) сделать вывод, что эти зоны документа не важны для ранжирования в Яндексе, по крайней мере для этих выборок? Нет, для такого вывода нет оснований.

Во-первых, многое указывает на то, что в Яндексе основная функция текстовых параметров — предварительный отбор достаточно релевантных кандидатов. Если это действительно

так, то отбор в широкий топ, о котором мы судим по контрасту топ-30 с фоном, важнее, чем корреляция с позицией внутри него. Но влияние параметра на попадание в топ может быть и в тех случаях, когда мы не видим такого контраста. Всё хорошо, если параметр учитывается таким образом только одной из двух поисковых машин — но про заголовки мы как раз знаем, что Google придерживается той же стратегии. В *Науке* и *Косметике* для нескольких текстовых параметров, связанных с **<title>** и **<h1>**, мы видим контраст с фоном (пусть слабый) сразу в обоих поисковиках. В *Гаджетах* Google почти везде обгоняет Яндекс — но совсем чуть-чуть: например, для вхождений разрозненных слов запроса в топ-30 полной выборки значение параметра в Яндексе составляет 0,56 против 0,57 в Google; в топ-30 скорректированной выборки — 0,59 против 0,60 в Google.

Во-вторых, если мы посмотрим на дополнительные («косвенные») параметры, оценивающие **<title>** и **<h1>**, то они скорее подтверждают важность заголовков для ранжирования Яндекса. Таких параметров у нас более десяти; они связаны в основном с наличием и длиной заголовков (которую можно считать по-раз-

ному — в символах, в словах, по количеству частей, разделенных знаками препинания). Чтобы не перегружать это исследование, которое и без того уже разрослось до немислимых размеров, мы приведем таблицы только для немногих из них, а еще про некоторые просто расскажем. Подробности можно пропустить — скажем только, что большинство этих параметров в Яндексе по крайней мере не слабее, чем в Google.

Наличие <title>: в Яндексе во всех информационных выборках средний или слабый контраст с фоном. Но трудно было бы ожидать большего — учитывая, что при округлении до второго знака после запятой для топ-30 и всех его поддиапазонов получается ровно единица (100%), а для фона или тоже единица, или 0,99. Для сравнения, в Google слабая корреляция с позицией в *Науке* (там для первой десятки единица, для второй и третьей — 0,99) и слабый контраст с фоном в стандартной коммерческой выборке.

Наличие <h1>: в Яндексе сильный контраст с фоном в *Науке* (в скорректированной выборке есть также корреляция с позицией 0,06) и в скорректированной *Косметике* (также корреляция с позицией 0,07), средний контраст

с фоном в скорректированных *Гаджетах*. В полных *Косметике* и *Гаджетах*, а также в стандартной коммерческой выборке связи с ранжированием не видно. В Google слабая корреляция с позицией в полной *Науке*, в скорректированных *Гаджетах* и в стандартной выборке.

Наличие двух или более <h1> (инвертированный параметр: чем такого меньше, тем лучше): в Яндексе только слабый контраст с фоном и слабая обратная корреляция с позицией в *Косметике* (полной и скорректированной); в Google только слабый контраст с фоном в *Косметике* же. Но таких случаев просто очень мало: значения во всех диапазонах и в фоне в обоих поисковиках колеблются в узких рамках от 0,00 до 0,01, и только в топ-3 Яндекса для скорректированной *Науки* «взлетают» до 0,02 — так что большего ждать и не приходится.

<H1> после <h2>-<h4> (инвертированный параметр): в Яндексе есть контраст с фоном во всех выборках. В *Науке* он слабый, в других информационных выборках, как и в стандартной коммерческой, — сильный. В Google почти ничего (только корреляция с позицией 0,03 в полной *Науке*). Но такие нарушения формата так-

Длина <title> в символах

Выборка	С корр.	топ-3	4–10	11–20	21–30	топ-30	фон	Ф-30/фон
Яндекс								
Наука	—	51,1	60,0	58,8	56,8	56,8	57,6	—
Наука ⁻¹	0,04	61,9	60,8	59,0	56,9	57,9	57,8	
Косметика	0,05	70,1	67,9	65,7	65,5	65,9	61,4	
Косметика ⁻²	0,05	71,0	67,8	65,7	65,5	65,9	61,4	
Гаджеты	0,03	62,9	62,2	60,6	60,0	60,6	58,7	
Гаджеты ⁻³	—	61,4	61,8	60,6	60,0	60,5	58,6	
Станд. комм.	0,04	83,6	78,5	74,2	76,2	75,5	72,7	
Google								
Наука	—	44,8	55,5	59,2	59,8	56,0	57,1	—
Наука ⁻⁴	—	57,8	57,3	59,6	60,0	58,3	57,4	—
Косметика	—	65,0	64,5	62,9	63,0	63,2	66,2	—
Гаджеты	0,03	62,4	61,3	61,0	59,1	60,3	60,3	—
Гаджеты ⁻⁵	—	62,3	60,3	60,4	59,1	60,0	60,2	—
Станд. комм.	0,04	81,8	74,6	75,4	74,3	75,4	75,0	—

С корр. — коэффициент корреляции с позицией (ранговый критерий Спирмена).

Ф-30/фон — оценка контраста между топ-30 и фоном (точный тест Фишера):

сильный, средний, слабый; прочерк — нет или отрицательный.

Скорректированные выборки: 1: без Википедии, Youtube, Dzen; 2: без Youtube, Dzen; 3: без Apple, Youtube, Dzen; 4: без Википедии; 5: без Apple, Youtube.

Длина <h1> в символах

Выборка	С корр.	топ-3	4–10	11–20	21-30	топ-30	фон	Ф-30/фон
Яндекс								
Наука	—	26,7	36,6	38,2	35,5	34,9	32,8	
Наука ⁻¹	0,04	39,2	40,0	39,1	35,7	37,7	33,1	
Косметика	—	47,3	48,5	48,7	50,5	49,1	44,2	
Косметика ⁻²	—	49,7	50,1	49,3	50,7	49,8	44,3	
Гаджеты	—	38,1	46,5	50,3	48,5	48,5	44,4	
Гаджеты ⁻³	—	49,6	50,3	50,6	48,5	49,9	44,4	
Станд. комм.	0,05	32,0	28,2	27,5	27,5	27,3	28,0	—
Google								
Наука	—	25,7	32,0	33,0	34,4	31,5	36,1	—
Наука ⁻⁴	—	33,3	33,4	33,2	34,4	33,1	36,4	—
Косметика	0,03	45,4	47,1	44,5	44,2	45,0	51,3	—
Гаджеты	—	44,3	45,3	46,2	45,4	45,8	48,3	—
Гаджеты ⁻⁵	—	46,9	48,6	47,8	45,8	47,2	48,7	—
Станд. комм.	—	30,2	28,4	28,8	28,3	27,8	27,7	

С корр. — коэффициент корреляции с позицией (ранговый критерий Спирмена).

Ф-30/фон — оценка контраста между топ-30 и фоном (точный тест Фишера):

 сильный,  средний,  слабый; прочерк — нет или отрицательный.

Скорректированные выборки: 1: без Википедии, Youtube, Dzen; 2: без Youtube, Dzen; 3: без Apple, Youtube, Dzen; 4: без Википедии; 5: без Apple, Youtube.

же очень редки — по разным выборкам и диапазонам от 0,01 до 0,03, в единичных случаях 0,04–0,05.

Длина <title> в символах (см. таблицу): в Яндексе во всех выборках, кроме полной *Науки*, есть контраст с фоном, от слабого до сильного (см. таблицу); в большинстве выборок он дополняется корреляцией с позицией. Аналогичная картина и для **количества слов в <title>**, но здесь слабая корреляция с позицией есть и в полной *Науке*. **Количество знаков препинания в <title>**: в Яндексе сильная корреляция с позицией во всех информационных выборках. Для всех этих параметров в Google только эпизодически встречается корреляция с позицией.

Длина <h1> в символах (см. таблицу) и **количество слов в <h1>**: всё примерно так же, как с <title>. Интересно, что <h1> в информационных выборках обычно длиннее, чем в стандартной коммерческой, при том, что <title>, наоборот, заметно короче. А вот количество вхождений слов запроса в <h1> по сравнению с <title> в информационных выборках и в коммерческой падает примерно одинаково.

Таким образом, судя по «косвенным» параметрам, получается, что в Яндексе значение заголовков <title> и <h1> скорее выше, чем в Google, и при этом в информационных выборках выше, чем в стандартной коммерческой.

Еще один важный дополнительный параметр — **позиция первого слова из запроса в <title>**. Здесь, в отличие от большой группы параметров, которую мы сейчас обсуждали, в Яндексе ничего существенного, а вот в Google есть и контраст с фоном, и (в *Науке* и *Гаджетах*) корреляция с позицией. Естественно, обратная, поскольку и сам параметр инвертированный: чем ближе «ключевые слова» к началу <title>, тем лучше.

Похоже, идеал для Google — <title>, полностью совпадающий с запросом. Отсюда и нарушение принципа антибуквализма (учет точных вхождений запроса), и предпочтение заголовков, где запрос встречается ближе к началу.

Позиция первого встретившегося слова из запроса в <title>
(чем меньше, тем лучше)

Выборка	С корр.	топ-3	4-10	11-20	21-30	топ-30	фон	Ф-30/фон
Яндекс								
Наука	0,05	2,39	2,86	2,96	2,90	2,86	2,82	—
Наука ⁻¹	—	2,75	2,85	2,95	2,90	2,90	2,82	—
Косметика	—	2,57	2,86	2,95	2,87	2,87	2,80	—
Косметика ⁻²	—	2,57	2,82	2,95	2,87	2,86	2,81	—
Гаджеты	—	3,16	3,28	3,29	3,22	3,25	3,25	—
Гаджеты ⁻³	—	3,22	3,25	3,29	3,22	3,25	3,15	—
Станд. комм.	0,03	1,67	1,78	1,74	1,79	1,76	1,76	
Google								
Наука	0,11	1,99	2,55	2,87	2,93	2,70	2,95	
Наука ⁻⁴	0,07	2,25	2,57	2,87	2,93	2,76	2,96	
Косметика	0,03	2,33	2,70	2,77	2,74	2,70	3,00	
Гаджеты	—	2,86	3,26	3,17	3,23	3,19	3,28	
Гаджеты ⁻⁵	—	2,81	3,31	3,16	3,23	3,20	3,29	
Станд. комм.	—	1,68	1,79	1,72	1,73	1,73	1,78	

С корр. – коэффициент корреляции с позицией (ранговый критерий Спирмена).

Ф-30/фон – оценка контраста между топ-30 и фоном (точный тест Фишера):

 сильный,  средний,  слабый; прочерк – нет или отрицательный.

Скорректированные выборки: 1: без Википедии, Youtube, Dzen; 2: без Youtube, Dzen; 3: без Apple, Youtube, Dzen; 4: без Википедии; 5: без Apple, Youtube.

URL и доменное имя

Так же, как `<h1>` — тень `<title>`, тенью `<h1>` можно считать URL, если он содержит кириллицу или транслит.

«Человекопонятные» URL'ы часто формируются автоматически как раз из `<h1>` и с точностью до транслитерации совпадают с ним или, реже, с его началом. Тем не менее, таблица для **вхождений слов запроса в URL** только отдалённо похожа на таблицу для `<h1>`*.

В Яндексе сохраняется противопоставление *Науки* и *Косметики*, где параметр «работает», *Гаджетах* и стандартной коммерческой выборке, где его влияния на ранжирование не видно. Но при этом в *Науке* и *Косметике* почти полностью пропадает корреляция с позицией. В Google корреляция с позицией в тех же *Науке* и *Косметике* сильнее, чем для `<h1>`, но зато контраст с фоном слабее.

Интересно, что количество вхождений слов запроса по сравнению с `<h1>` в разных выборках падает сильно по-разному — хотя падает, конечно, везде, поскольку «говорящие» URL используются далеко не на всех сай-

тах. Сильнее всего падение в *Науке* (как в полной, так и в скорректированной выборке) и в полных *Гаджетах* (в скорректированной выборке падение значительно меньше).

Что касается самих доменов, то это сайтовый, а не страничный параметр, и с заголовками он никак не связан. Заметное (до 0,06 в пересчете на целые запросы) количество **вхождений слов запроса в доменное имя** есть только в *Гаджетах*. Здесь этот параметр оказывается довольно сильным: контраст с фоном и корреляция с позицией до 0,07–0,08 есть и в Яндексе, и в Google. На скорректированной выборке — без сайта *support.apple.com* — сила параметра ожидаемо падает.

В *Науке* и *Гаджетах* вхождений очень мало, 1–2% (в пересчете на запросы) — соответственно, и параметр не «работает».

* Для URL собирается меньше параметров, поэтому сравнить эти таблицы можно только по части столбцов.

Текстовые параметры: URL (кроме домена)

Яндекс

Параметры	Выделенные	Слова запроса + выделенные	Синонимы	Слова запроса + синонимы	Слова запроса	Неточный запрос + синонимы	Неточный запрос	Запрос с точностью до форм	Точный запрос	Слова запроса (полных запросов)	
										Топ-3	Топ-30
Наука										0,41	0,39
Наука ⁻¹										0,44	0,40
Косметика										0,55	0,55
Косметика ⁻²				0,03	0,03					0,64	0,58
Гаджеты										0,37	0,48
Гаджеты ⁻³										0,53	0,52
Станд. комм.								0,03		0,64	0,61

Google

Параметры	Выделенные	Слова запроса + выделенные	Синонимы	Слова запроса + синонимы	Слова запроса	Неточный запрос + синонимы	Неточный запрос	Запрос с точностью до форм	Точный запрос	Слова запроса (полных запросов)	
										Топ-3	Топ-30
Наука				0,07	0,06					0,48	0,37
Наука ⁻⁴				0,04	0,03					0,41	0,36
Косметика				0,06	0,06					0,62	0,56
Гаджеты										0,43	0,50
Гаджеты ⁻⁵										0,55	0,52
Станд. комм.										0,65	0,65

— сильный, средний, слабый контраст между топ-30 и фоном (точный тест Фишера).

— сильная, средняя, слабая корреляция с позицией.

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

Скорректированные выборки: 1: без Википедии, Youtube, Dzen; 2: без Youtube, Dzen; 3: без Apple, Youtube, Dzen; 4: без Википедии; 5: без Apple, Youtube.

Meta description и keywords

С заголовком `<title>` связана еще одна зона — **meta description**, которая часто служит его расширенной версией, вполне «законно» предназначенной уже только для поисковых машин.

Но **вхождений слов запроса в description** обычно меньше, чем в `<title>`; исключение — *Косметика*.

В Яндексе для **description** преобладает контраст с фоном, и только для скорректированной *Науки* видна корреляция с позицией. Кстати, такая же ситуация и с вспомогательными параметрами — **наличие description** и **длина description** (всё равно, **в знаках** или **в словах**): во всех информационных выборках для них характерен сильный контраст с фоном, и только в скорректированной *Науке* он дополняется корреляцией с позицией (0,07 для наличия и 0,08 для длины).

То есть Яндекс, без сомнения, «читает» description, но, судя по картине срабатывания параметров, скорее не как `<title>`, а как продолжение текста страницы.

В Google картина менее определенная. Здесь влияния наличия или длины **description** на ранжирование не видно, а текстовые параметры выглядят убедительно только в *Косметике*, где сильный контраст с фоном от правого края таблицы (как с `<title>`) доходит почти до левого, а корреляция с позицией может достигать 0,06.

Ну и, раз мы обсудили **meta description**, нельзя не сказать несколько слов и о **meta keywords**.

«Ключевые слова» — первая зона в HTML-коде, которую еще на заре Интернета, на рубеже веков или даже раньше, тогдашние веб-мастера научились «спамить», а тогдашние поисковые машины — игнорировать. Немного трудно поверить, что сейчас Яндекс или Google ее учитывают — хотя, с другой стороны, и манипулировать ими через **keywords** сейчас мало кто пытается, так что почему бы и нет.

Текстовые параметры: meta description

Яндекс

Параметры	Выделенные	Слова запроса + выделенные	Синонимы	Слова запроса + синонимы	Слова запроса	Неточный запрос + синонимы	Неточный запрос	Запрос с точностью до форм	Точный запрос	Слова запроса (полных запросов)		
										Топ-3	Топ-30	
Наука											0,53	0,61
Наука ⁻¹		 0,05		 0,06	 0,06	 0,06	 0,05	 0,04	 0,03		0,75	0,64
Косметика	 0,03										0,80	0,79
Косметика ⁻²	 0,04	 0,03				 0,03					0,83	0,80
Гаджеты											0,63	0,59
Гаджеты ⁻³											0,61	0,58
Станд. комм.											0,96	1,01

Google

Параметры	Выделенные	Слова запроса + выделенные	Синонимы	Слова запроса + синонимы	Слова запроса	Неточный запрос + синонимы	Неточный запрос	Запрос с точностью до форм	Точный запрос	Слова запроса (полных запросов)		
										Топ-3	Топ-30	
Наука											0,44	0,52
Наука ⁻⁴		 0,04	 0,04	 0,06	 0,05						0,65	0,55
Косметика	 0,04	 0,06	 0,03	 0,06	 0,05	 0,03					0,86	0,81
Гаджеты											0,47	0,56
Гаджеты ⁻⁵											0,46	0,55
Станд. комм.	 0,07										0,95	1,02

— сильный, средний, слабый контраст между топ-30 и фоном (точный тест Фишера).

— сильная, средняя, слабая корреляция с позицией.

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

Скорректированные выборки: 1: без Википедии, Youtube, Dzen; 2: без Youtube, Dzen; 3: без Apple, Youtube, Dzen; 4: без Википедии; 5: без Apple, Youtube.

Длина meta description в знаках

Выборка	С корр.	топ-3	4–10	11–20	21-30	топ-30	фон	Ф-30/фон
Яндекс								
Наука	—	109	128	119	110	120	78,2	
Наука ⁻¹	0,08	155	128	119	110	123	79	
Косметика	—	152	154	151	152	153	140	
Косметика ⁻²	—	152	154	151	152	152	140	
Гаджеты	—	115	130	132	135	137	122	
Гаджеты ⁻³	—	107	127	131	135	135	122	
Станд. комм.	—	157	176	168	167	167	156	
Google								
Наука	—	64,7	95,4	92	92,3	88,8	121	—
Наука ⁻⁴	0,04	129	102	93,3	93,2	99,5	122	—
Косметика	—	151	143	146	145	148	151	—
Гаджеты	—	96,3	124	127	127	128	139	—
Гаджеты ⁻⁵	—	93,5	115	122	126	123	138	—
Станд. комм.	—	155	161	163	154	159	169	—

С корр. — коэффициент корреляции с позицией (ранговый критерий Спирмена).

Ф-30/фон — оценка контраста между топ-30 и фоном (точный тест Фишера):

сильный, средний, слабый; прочерк — нет или отрицательный.

Скорректированные выборки: 1: без Википедии, Youtube, Dzen; 2: без Youtube, Dzen; 3: без Apple, Youtube, Dzen; 4: без Википедии; 5: без Apple, Youtube.

Таблицу текстовых параметров, связанных с **meta keywords**, по-видимому, следует рассматривать всё-таки как иллюстрацию того, что относиться к иконкам и числам в наших таблицах надо с большой осторожностью: у явно не работающих параметров может быть и заметная корреляция с позицией, и контраст с фоном.

Самые «сильные» строки — очевидная «наведенка». Судя по всему, и для Яндекса, и для Google высокая корреляция с позицией для **вхождений слов запроса, слов запроса и синонимов**, просто **синонимов** и т. п. в **keywords** создается одним конкретным сайтом — Youtube. Все следы «важности» **keywords** теряются и для скорректированной выборки *Косметика* в Яндексе — по той же причине, из-за удаления Youtube.

В целом, количество вхождений в **keywords** очень низкое, почти везде заметно меньше, чем в **description**. У многих сайтов этой зоны в HTML-коде вообще нет — и в такой ситуации случайное распределение этого параметра у сайтов, попавших в топ, может сильно влиять на расклад. Так, для *Науки* (в том числе скорректированной) контраст с фоном есть в Яндексе, а для *Косметики* и скорректированных *Гад-*

жетов — в Google. При таком небольшом среднем числе вхождений для этого, по-видимому, достаточно случайных перекосов. При отсутствии значимых свидетельств участия **keywords** в ранжировании для коммерческих выборок, разнонаправленные «сигналы» для информационных выборок вряд ли стоит принимать во внимание.

Мы прекрасно понимаем, что аналогичные аргументы можно применить и к другим параметрам — в том числе тем, где мы сами считаем «срабатывание» параметров не случайным. Именно поэтому мы стараемся быть осторожными в своих выводах и не торопимся называть наши параметры — даже сильные — факторами ранжирования. С удовольствием обсудим со всеми, кому это интересно, наши и ваши гипотезы; особенно будем признательны за аргументированную критику: в спорах рождается истина! Связаться с нами проще всего по электронной почте: info@searchlab.ru.

Текстовые параметры: meta keywords

Яндекс

Параметры	Выделенные	Слова запроса + выделенные	Синонимы	Слова запроса + синонимы	Слова запроса	Неточный запрос + синонимы	Неточный запрос	Запрос с точностью до форм	Точный запрос	Слова запроса (полных запросов)	
										Топ-3	Топ-30
Наука		0,04		0,03	0,04	0,04	0,04	0,03	0,03	0,36	0,31
Наука ⁻¹		0,03			0,03	0,03	0,03			0,30	0,28
Косметика		0,05		0,06	0,06	0,05	0,05	0,03	0,03	0,49	0,42
Косметика ⁻²										0,23	0,32
Гаджеты	0,07	0,13	0,11	0,14	0,13	0,09	0,08	0,05	0,05	0,71	0,37
Гаджеты ⁻³		0,04	0,04	0,05	0,04	0,03	0,03			0,21	0,23
Станд. комм.										0,45	0,77

Google

Параметры	Выделенные	Слова запроса + выделенные	Синонимы	Слова запроса + синонимы	Слова запроса	Неточный запрос + синонимы	Неточный запрос	Запрос с точностью до форм	Точный запрос	Слова запроса (полных запросов)	
										Топ-3	Топ-30
Наука										0,14	0,23
Наука ⁻⁴										0,21	0,24
Косметика		0,05		0,06	0,06					0,57	0,46
Гаджеты	0,09	0,11	0,09	0,11	0,10	0,08	0,08	0,05	0,05	0,43	0,46
Гаджеты ⁻⁵	0,06	0,06	0,04	0,06	0,06	0,05	0,05			0,28	0,27
Станд. комм.										0,52	0,75

– сильный, средний, слабый контраст между топ-30 и фоном (точный тест Фишера).

– сильная, средняя, слабая корреляция с позицией.

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

Скорректированные выборки: 1: без Википедии, Youtube, Dzen; 2: без Youtube, Dzen; 3: без Apple, Youtube, Dzen; 4: без Википедии; 5: без Apple, Youtube.

Внутренние заголовки

Если заголовок <h1> в норме один на странице, то заголовков <h2>–<h4>, если они вообще есть, обычно несколько: они предназначены для членения текста на содержательные зоны.

Внутренние заголовки — это способ обозначить структуру текста, маркировать (то есть выделить) наиболее важные слова, суммирующие смысл частей основного текста или отдельных блоков текста на странице. В SEO с давних, еще доссылочных* времен принято обращать на них внимание и следить, чтобы «ключевые запросы» были в них в достаточной степени представлены.

Таблица для <h2>–<h4> (мы не рассматриваем заголовки этих уровней по отдельности и не учитываем редко встречающиеся заголовки более низких уровней, <h5> и <h6>) подтверждает, что и современные поиско-

вые машины, судя по всему, учитывают вхождения во внутренние заголовки.

Впрочем, вопрос об их роли в ранжировании нужно поставить более строго. Поскольку заголовки <h2>–<h4> — часть текста страницы, очевидно, что они вносят свой вклад в срабатывание факторов, основанных на учете этого текста в целом. Но есть ли у них какая-то специальная роль? Связаны ли с ними дополнительные факторы ранжирования — пусть даже не очень сильные? Или (что почти то же самое) учитываются ли они с большим весом, чем другие слова на странице? По крайней мере для Яндекса на эти вопросы, кажется, можно ответить положительно.

* Возможно, не все уже помнят, что ссылочное ранжирование (алгоритм PageRank) был тем самым секретом, который помог «взлететь» и победить большинство конкурентов поиску Google, созданному в 1996–1998 годах. В те легендарные времена ссылки еще никто не спамил, и основанное на них ранжирование давало существенный прирост качества. У поисковых машин — предшественников Google — ранжирование опиралось в основном на тексты, которыми к тому времени уже повсюду манипулировали веб-мастера.

Таблица для внутренних заголовков, если смотреть на информационные выборки, вообще радует глаз: она близка к образцовой в том смысле, что на ней удобно демонстрировать описанные выше принципы текстового ранжирования.

В Яндексе явно преобладает контраст с фоном, а в Google — корреляция с позицией. Принцип антибуквализма налицо — правда, в Яндексе сильный контраст с фоном доходит до правого края таблицы не только в *Науке*, где много коротких запросов, но и в *Косметике*, где количество вхождений слов запроса во внутренние заголовки выше, чем в других выборках (2,5–3 в пересчете на целые запросы в топ 30, против примерно 1 в *Науке* и 2 с небольшим в *Гаджетах*). В Google присутствует слабый контраст с фоном в правых столбцах.

Наконец, таблица подтверждает еще один принцип текстового ранжирования, на который мы обратили внимание уже в рамках этого исследования (см. выше в разделе [Заголовки <title> и <h1>](#)): если на данной выборке какая-то зона текста сильнее про-

являет себя в Яндексе, то то же самое, скорее всего, верно и для Google, и наоборот — хотя в Яндексе и в Google сила параметров может проявляться совершенно по-разному. В *Косметике* параметры этой группы сильнее, в *Гаджетах* слабее — и это в равной степени относится и к Яндексу, и к Google. Похоже, эти параллельные усиления или ослабления параметров в двух совершенно разных поисковиках как-то определяются объективными свойствами выборок (или, точнее, сайтов, которые релевантны их запросам). Чем больше в выборке средний (или медианный) объем той или иной текстовой зоны, тем выше вероятность, что параметры для нее будут сильными.

В Яндексе таблица для **<h2>–<h4>** не слишком похожа на аналогичную таблицу для текста страницы в целом — что представляется косвенным подтверждением самостоятельной роли заголовков в ранжировании. Контраста с фоном заметно больше, а корреляции с позицией меньше. В Google разница между таблицами менее выраженная, но она тоже есть: в *Науке* корреляция с позицией для внутренних заголовков слабее, чем для текста страницы, а в *Косметике* и *Гаджетах* она, наоборот, сильнее. Для текста страницы корреляции с позицией на полной выборке *Гаджетов* вообще не видно, а здесь она появляется сразу для трех параметров (хоть и слабая).

Текстовые параметры: <h2>-<h4>

Яндекс

Параметры	Выделенные	Слова запроса + выделенные	Синонимы	Слова запроса + синонимы	Слова запроса	Неточный запрос + синонимы	Неточный запрос	Запрос с точностью до форм	Точный запрос	Слова запроса (полных запросов)	
										Топ-3	Топ-30
Наука										1,00	0,98
Наука ⁻¹										0,97	1,00
Косметика		 0,03		 0,03	 0,03					3,42	2,73
Косметика ⁻²	 0,04	 0,06		 0,06	 0,06	 0,04	 0,03	 0,03		3,68	2,81
Гаджеты										2,14	2,07
Гаджеты ⁻³										2,72	2,18
Станд. комм.		 0,04		 0,05	 0,05	 0,05	 0,03	 0,03	 0,04	6,02	3,19

Google

Параметры	Выделенные	Слова запроса + выделенные	Синонимы	Слова запроса + синонимы	Слова запроса	Неточный запрос + синонимы	Неточный запрос	Запрос с точностью до форм	Точный запрос	Слова запроса (полных запросов)	
										Топ-3	Топ-30
Наука	 0,06	 0,08	 0,04	 0,08	 0,09	 0,05	 0,05	 0,05	 0,04	0,91	0,63
Наука ⁻⁴	 0,04	 0,05	 0,03	 0,05	 0,05	 0,03	 0,03			0,69	0,60
Косметика	 0,06	 0,10	 0,04	 0,11	 0,10	 0,07	 0,06	 0,03	 0,03	3,36	2,44
Гаджеты		 0,04		 0,05	 0,04					2,59	1,84
Гаджеты ⁻⁵		 0,08	 0,03	 0,09	 0,08	 0,03				2,73	1,93
Станд. комм.	 0,03			 0,03	 0,04	 0,04	 0,05	 0,05	 0,03	3,00	2,44

— сильный, средний, слабый контраст между топ-30 и фоном (точный тест Фишера).

— сильная, средняя, слабая корреляция с позицией.

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

Скорректированные выборки: 1: без Википедии, Youtube, Dzen; 2: без Youtube, Dzen; 3: без Apple, Youtube, Dzen; 4: без Википедии; 5: без Apple, Youtube.

Косвенные параметры тоже подтверждают самостоятельное значение `<h2>`–`<h4>`, особенно для Яндекса.

Для **наличия заголовков `<h2>`–`<h4>`** в Яндексе есть контраст с фоном для всех информационных выборок, кроме полных *Гаджетов*, причем в *Науке* и *Косметике* сильный. Например, в *Науке* на страницах, попавших в топ-30 Яндекса, внутренние заголовки есть в 79% случаев (в скорректированной выборке в 81%), а в фоне — только в 73%. В Google есть корреляция с позицией от 0,03 до 0,08 в трех из пяти строк для информационных выборок: полная *Наука*, *Косметика* и скорректированные *Гаджеты*.

Для **длины `<h2>`–`<h4>` в символах** (см. таблицу) или **в словах** ситуация похожая. Обратите внимание, что и в Яндексе, и в Google почти во всех выборках есть контраст между топ-3 и остальной частью топ-30 (хотя полноценная корреляция с позицией есть далеко не везде). Контраст с фоном в Яндексе на большинстве выборок очень сильный: в некоторых даже топ-3 Google недотягивает до топ-30 Яндекса.

Таким образом, в топ-30 Яндекса по сравнению с фоном больше доля страниц, на которых есть внутренние заголовки; при этом

там, где такие заголовки есть, больше их средняя длина.

**Но это еще не всё:
концентрация слов запроса
в этих заголовках тоже
значимо выше, чем в фоне —
то есть вхождения запроса
в Яндексе наступают по всем
фронтам.**

В принципе, это логично: раз для вхождений слов запроса в `<h2>`–`<h4>` есть контраст с фоном, можно ожидать, что и концентрация (доля в процентах) слов запроса в топ-30 окажется выше, чем в фоне. Но дело в том, что количество вхождений может расти не только за счет повышения концентрации, но и за счет увеличения размера зоны при одинаковой концентрации. Именно так это происходит с текстом страницы: рост числа вхождений слов запроса здесь «экстенсивный», а не «интенсивный» — он достигается за счет большего объема текста, а не за счет большей плотности вхождений в текст того же объема. То же самое и с `<title>` и с `<h1>` — даже в тех выборках, где в этих зонах для слов запроса в Яндексе есть контраст с фоном, для концентрации слов запроса мы контраста с фоном не наблюдаем. А вот в `<h2>`–`<h4>` «растут» одновременно и суммарная длина заголовков, и концентрация слов запроса в них.

В Google, где для вхождений слов запроса в `<h2>`–`<h4>` характерна

Общая длина <h2>-<h4> в символах

Выборка	С корр.	топ-3	4–10	11–20	21-30	топ-30	фон	Ф-30/фон
Яндекс								
Наука	—	249	121	135	136	126	68,2	
Наука ⁻¹	—	169	136	138	136	127	67,6	
Косметика	—	417	353	340	352	338	225	
Косметика ⁻²	—	460	380	351	356	351	226	
Гаджеты	0,03	315	254	289	296	266	231	
Гаджеты ⁻³	—	441	303	292	296	283	231	
Станд. комм.	0,05	339	160	151	145	150	139	—
Google								
Наука	0,09	223	94,4	75,2	77,8	80,1	132	—
Наука ⁻⁴	—	100	85,5	73,7	76,8	69,1	130	—
Косметика	0,05	326	304	260	238	259	356	—
Гаджеты	—	397	224	238	240	233	285	—
Гаджеты ⁻⁵	0,06	417	265	257	245	249	290	—
Станд. комм.	—	189	156	139	139	133	158	—

С корр. — коэффициент корреляции с позицией (ранговый критерий Спирмена).

Ф-30/фон — оценка контраста между топ-30 и фоном (точный тест Фишера):

сильный, средний, слабый; прочерк — нет или отрицательный.

Скорректированные выборки: 1: без Википедии, Youtube, Dzen; 2: без Youtube, Dzen; 3: без Apple, Youtube, Dzen; 4: без Википедии; 5: без Apple, Youtube.

Концентрация слов запроса в <h2>-<h4> (%)

Выборка	С корр.	топ-3	4–10	11–20	21-30	топ-30	фон	Ф-30/фон
Яндекс								
Наука	—	6,14	8,45	8,17	7,90	7,92	6,35	
Наука ⁻¹	—	7,88	8,58	8,21	7,93	8,15	6,39	
Косметика	0,07	16,1	13,5	12,5	12,4	13,0	12,2	
Косметика ⁻²	0,06	15,9	13,5	12,5	12,4	13,0	12,2	
Гаджеты	—	9,71	9,23	8,64	8,60	8,81	7,81	
Гаджеты ⁻³	—	9,76	9,26	8,64	8,60	8,82	7,81	
Станд. комм.	—	15,2	15,2	14,5	14,0	14,6	13,4	
Google								
Наука	0,06	6,58	7,82	6,64	5,95	6,68	8,19	—
Наука ⁻⁴	0,06	9,10	8,18	6,69	5,98	7,00	8,26	—
Косметика	0,10	17,8	13,5	12,9	12,3	13,3	12,0	
Гаджеты	0,06	9,10	9,51	8,15	8,09	8,50	8,40	—
Гаджеты ⁻⁵	0,06	9,10	9,51	8,15	8,09	8,50	8,40	—
Станд. комм.	0,04	14,7	13,9	14,4	13,3	13,9	14,1	—

С корр. — коэффициент корреляции с позицией (ранговый критерий Спирмена).

Ф-30/фон — оценка контраста между топ-30 и фоном (точный тест Фишера):

сильный, средний, слабый; прочерк — нет или отрицательный.

Скорректированные выборки: 1: без Википедии, Youtube, Dzen; 2: без Youtube, Dzen; 3: без Apple, Youtube, Dzen; 4: без Википедии; 5: без Apple, Youtube.

корреляция с позицией, корреляция с позицией есть и для концентрации слова запроса, причем местами даже более сильная.

Отметим, что заголовки `<h2>`–`<h4>` с концентрацией слов запроса поряд-

ка 10% занимают промежуточное положение между текстом страницы (концентрация около 3%) и основными заголовками (порядка 20% в `<title>` и 25–30% в `<h1>`).

Выделения в тексте

Выделения жирностью и курсивом — менее значимые, чем внутренние заголовки, и обычно для них влияния на ранжирование мы не видим. Однако для информационных выборок намеки на такое влияние есть.

Как обычно, в Яндексе преобладает контраст с фоном, а в Google — корреляция с позицией. Правда, параметры «срабатывают» не на всех выборках. В Яндексе для жирности не видно связи с ранжированием в *Гаджетах*, а для курсива — в *Косметике*. В Google параметры в *Гаджетах* сильнее, чем в *Науке*, а в *Косметике* они самые слабые.

Впрочем, **вхождений слов запроса в *bold* и в *italic*** вообще немного: в тексте, выделенном жирным шрифтом (теги `` и ``) их набирается примерно на один полный запрос, а в выделенном курсивом (теги `` и `<i>`) — еще в два-четыре раза меньше. Но и вообще выделенного тега-

ми текста обоих видов тоже не так уж и много — меньше, например, чем текста, выделенного тегами внутренних заголовков. По концентрации слов запроса жирный и курсивный текст занимают промежуточное положение между текстом страницы в целом и внутренними заголовками: первый ближе к заголовкам, второй — к тексту страницы в целом.

В Яндексе большее количество вхождений слов запроса в выделенный текст достигается за счет большего объема выделенного текста; в Google — также и за счет большей концентрации (особенно для жирного шрифта).

Текстовые параметры: жирность (и)

Яндекс

Параметры	Выделенные	Слова запроса + выделенные	Синонимы	Слова запроса + синонимы	Слова запроса	Неточный запрос + синонимы	Неточный запрос	Запрос с точностью до форм	Точный запрос	Слова запроса (полных запросов)	
										Топ-3	Топ-30
Наука						 0,03	 0,03	 0,03	 0,03	0,98	0,93
Наука ⁻¹										0,84	0,94
Косметика										0,98	0,93
Косметика ⁻²	 0,03	 0,03								1,04	0,95
Гаджеты										0,94	0,90
Гаджеты ⁻³						 0,03	 0,03			1,19	0,94
Станд. комм.					 0,03					0,55	0,76

Google

Параметры	Выделенные	Слова запроса + выделенные	Синонимы	Слова запроса + синонимы	Слова запроса	Неточный запрос + синонимы	Неточный запрос	Запрос с точностью до форм	Точный запрос	Слова запроса (полных запросов)	
										Топ-3	Топ-30
Наука	 0,06	 0,09	 0,03	 0,11	 0,11	 0,10	 0,09	 0,10	 0,09	1,04	0,70
Наука ⁻⁴				 0,03	 0,03	 0,04	 0,03	 0,04	 0,03	0,78	0,66
Косметика		 0,04		 0,03	 0,03					0,74	0,91
Гаджеты		 0,05	 0,03	 0,06	 0,05	 0,04	 0,03			1,58	0,92
Гаджеты ⁻⁵	 0,03	 0,07	 0,05	 0,08	 0,07	 0,05	 0,04			1,67	0,97
Станд. комм.										0,48	0,71

– сильный, средний, слабый контраст между топ-30 и фоном (точный тест Фишера).

– сильная, средняя, слабая корреляция с позицией.

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

Скорректированные выборки: 1: без Википедии, Youtube, Dzen; 2: без Youtube, Dzen; 3: без Apple, Youtube, Dzen; 4: без Википедии; 5: без Apple, Youtube.

Текстовые параметры: курсив (и <i>)

Яндекс

Параметры	Выделенные	Слова запроса + выделенные	Синонимы	Слова запроса + синонимы	Слова запроса	Неточный запрос + синонимы	Неточный запрос	Запрос с точностью до форм	Точный запрос	Слова запроса (полных запросов)	
										Топ-3	Топ-30
Наука										0,41	0,36
Наука ⁻¹										0,36	0,36
Косметика										0,22	0,21
Косметика ⁻²										0,24	0,22
Гаджеты	 0,03	 0,04	 0,04	 0,03						0,44	0,25
Гаджеты ⁻³	 0,04	 0,06	 0,05	 0,06	 0,04					0,56	0,26
Станд. комм.										0,08	0,16

Google

Параметры	Выделенные	Слова запроса + выделенные	Синонимы	Слова запроса + синонимы	Слова запроса	Неточный запрос + синонимы	Неточный запрос	Запрос с точностью до форм	Точный запрос	Слова запроса (полных запросов)	
										Топ-3	Топ-30
Наука	 0,05	 0,08	 0,03	 0,08	 0,07	 0,05	 0,05	 0,04		0,43	0,30
Наука ⁻⁴		 0,05		 0,05	 0,04	 0,03				0,32	0,28
Косметика										0,24	0,21
Гаджеты	 0,03	 0,06	 0,03	 0,06	 0,06	 0,03	 0,03			0,41	0,23
Гаджеты ⁻⁵	 0,04	 0,07	 0,04	 0,07	 0,07	 0,04	 0,03			0,44	0,24
Станд. комм.										0,20	0,17

— сильный, средний, слабый контраст между топ-30 и фоном (точный тест Фишера).

— сильная, средняя, слабая корреляция с позицией.

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

Скорректированные выборки: 1: без Википедии, Youtube, Dzen; 2: без Youtube, Dzen; 3: без Apple, Youtube, Dzen; 4: без Википедии; 5: без Apple, Youtube.

Ссылки со страницы

Внутренние ссылки со страницы — одна из самых активных зон текстового ранжирования.

Вероятно, это связано с тем, что для ранжирования важен «релевантный размер» сайта — количество страниц, имеющих отношение к запросу. Мы говорили об этом в первой части исследования, когда обсуждали параметры, связанные с размером сайта.

Для информационных сайтов здесь всё примерно так же, как для коммерческих, если не считать *Гаджетов*, где из-за высокой монополизации текстовые параметры вообще слабые.

Количество ссылок меньше, чем в стандартной коммерческой выборке, но всё равно довольно большое: порядка 70–100 ссылок (в стандартной выборке в три раза больше), их общая длина исчисляется сотнями слов и тысячами символов. При этом и **количество исходящих внутренних ссылок** (см. таблицу выше в подразделе [Исходящие ссылки](#) раздела про ссылочные параметры), и их **общий объем в символах** и **в словах** — сильные параметры (кроме *Гаджетов* в Яндексе).

Концентрация слов запроса во внутренних ссылках — тоже сильный параметр, то есть выигрыш в ранжировании дают и увеличение общей длины внутренних ссылок, и увеличение доли в них слов запроса.

В общем, релевантные запросу внутренние ссылки на продвигаемой странице, несомненно, полезны.

При этом концентрация слов запроса во внутренних ссылках в информационных выборках заметно меньше, чем в стандартной коммерческой (для *Науки* — более чем в три раза), и сравнима с концентрацией слов запроса в тексте страницы. Дело в том, что на страницах коммерческих сайтов, попадающих в результаты поиска в стандартной выборке, часто есть большой блок, наполненный «релевантными» внутренними ссылками — ссылки с карточек товарного каталога на страницы самих товаров. А на информационных сайтах такого блока нет, большая часть внутренних ссылок относится к раз-

Текстовые параметры: исходящие внутренние ссылки

Яндекс

Параметры	Выделенные	Слова запроса + выделенные	Синонимы	Слова запроса + синонимы	Слова запроса	Неточный запрос + синонимы	Неточный запрос	Запрос с точностью до форм	Точный запрос	Слова запроса (полных запросов)	
										Топ-3	Топ-30
Наука	0,03	0,06	0,05	0,07	0,07	0,06	0,06	0,05	0,04	2,23	0,65
Наука ⁻¹		0,04	0,04	0,06	0,06	0,04	0,04	0,03	0,03	1,55	0,66
Косметика	0,05	0,08	0,04	0,07	0,07	0,04	0,05	0,03		4,88	2,92
Косметика ⁻²	0,07	0,11	0,06	0,11	0,11	0,06	0,06	0,03	0,03	5,46	3,08
Гаджеты										2,42	1,90
Гаджеты ⁻³				0,04						3,40	2,04
Станд. комм.		0,03	0,04	0,03	0,03	0,05	0,04	0,04	0,06	17,8	16,1

Google

Параметры	Выделенные	Слова запроса + выделенные	Синонимы	Слова запроса + синонимы	Слова запроса	Неточный запрос + синонимы	Неточный запрос	Запрос с точностью до форм	Точный запрос	Слова запроса (полных запросов)	
										Топ-3	Топ-30
Наука	0,10	0,13	0,05	0,13	0,14	0,07	0,07	0,06	0,05	2,49	0,55
Наука ⁻⁴	0,04	0,06	0,03	0,06	0,06					1,39	0,47
Косметика	0,04	0,04		0,05	0,05	0,05	0,05	0,04	0,03	4,12	2,64
Гаджеты				0,03						3,06	2,02
Гаджеты ⁻⁵	0,03	0,05	0,04	0,08	0,04					3,24	2,20
Станд. комм.	0,04	0,08		0,09	0,08	0,05				16,5	13,1

— сильный, средний, слабый контраст между топ-30 и фоном (точный тест Фишера).

— сильная, средняя, слабая корреляция с позицией.

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

Скорректированные выборки: 1: без Википедии, Youtube, Dzen; 2: без Youtube, Dzen; 3: без Apple, Youtube, Dzen; 4: без Википедии; 5: без Apple, Youtube.

личным меню сайта, и лишь немногие ведут на тематически связанные страницы — они-то и помогают ранжированию.

Исходящих внешних ссылок обычно гораздо меньше: на многих страницах их нет совсем. Если верить описаниям алгоритма PageRank, внешние ссылки (по крайней мере, не защищенные NoFollow) раздают ссылочный вес другим сайтам, уменьшая тем самым вес своего сайта, — поэтому SEOшники, при всей их любви к входящим ссылкам, к исходящим ссылкам на своих сайтах относятся с подозрением и стараются их не ставить. Тем не менее, достаточно посмотреть на Википедию, чтобы убедиться в том, что исходящие ссылки — это далеко не всегда плохо.

Собственно, посмотреть на Википедию можно в *Науке*: именно она обеспечивает здесь (для полной выборки) столь высокую силу параметров, связанных с внешними ссылками, — не удивительно, что для скорректированной выборки эти параметры резко слабеют.

Интересно, что почти для всех информационных выборок контраст с фоном, пусть даже не всегда сильный, есть одновременно и в Яндексе, и в Google — а это обычно свидетельствует о важ-

ности параметра для обоих поисковиков.

Вполне вероятно, что внешние ссылки поисковые машины могут интерпретировать как ссылки на источники (часто это они и есть на самом деле, особенно в *Науке*), что может способствовать утверждению экспертного статуса, который для информационных страниц далеко не лишний. (Впрочем, — как мы уже упоминали выше в разделе [Исходящие ссылки](#) — «работают» и исходящие ссылки с коммерческих сайтов, для которых экспертность не так важна.)

В общем, если вы хотите поставить ссылку на внешний источник информации, не отказывайте себе в этом удовольствии, и вам зачтется при ранжировании.

Текстовые параметры: исходящие внешние ссылки

Яндекс

Параметры	Выделенные	Слова запроса + выделенные	Синонимы	Слова запроса + синонимы	Слова запроса	Неточный запрос + синонимы	Неточный запрос	Запрос с точностью до форм	Точный запрос	Слова запроса (полных запросов)	
										Топ-3	Топ-30
Наука	0,08	0,12	0,07	0,13	0,13	0,10	0,10	0,09	0,08	1,06	0,29
Наука ⁻¹	0,03	0,03	0,03	0,04	0,04					0,22	0,18
Косметика	0,03	0,07	0,03	0,06	0,07					0,07	0,05
Косметика ⁻²	0,03	0,07	0,03	0,06	0,07					0,06	0,05
Гаджеты		0,03		0,04	0,03					1,27	0,28
Гаджеты ⁻³		0,06	0,03	0,06	0,05					1,61	0,30
Станд. комм.	0,03	0,04	0,04	0,04	0,04	0,03	0,03			0,27	0,21

Google

Параметры	Выделенные	Слова запроса + выделенные	Синонимы	Слова запроса + синонимы	Слова запроса	Неточный запрос + синонимы	Неточный запрос	Запрос с точностью до форм	Точный запрос	Слова запроса (полных запросов)	
										Топ-3	Топ-30
Наука	0,12	0,18	0,09	0,18	0,17	0,15	0,15	0,14	0,13	1,31	0,30
Наука ⁻⁴	0,04	0,06		0,06	0,05	0,04	0,04	0,04	0,04	0,25	0,15
Косметика	0,04	0,10		0,09	0,10		0,04			0,09	0,05
Гаджеты		0,05	0,03	0,05	0,05			0,03	0,03	1,62	0,27
Гаджеты ⁻⁵	0,03	0,06	0,03	0,06	0,07			0,03	0,03	1,71	0,28
Станд. комм.	0,03	0,06		0,05	0,05	0,03	0,04	0,03		0,33	0,10

— сильный, средний, слабый контраст между топ-30 и фоном (точный тест Фишера).

— сильная, средняя, слабая корреляция с позицией.

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

Скорректированные выборки: 1: без Википедии, Youtube, Dzen; 2: без Youtube, Dzen; 3: без Apple, Youtube, Dzen; 4: без Википедии; 5: без Apple, Youtube.

Списки `` и атрибуты `alt` и `title`

На коммерческих сайтах эти элементы HTML-кода характерны прежде всего для карточек товаров в товарных рубриках — с чем мы всегда и связывали их важность для ранжирования. А что с информационкой, где никаких карточек товаров нет?

С элементами списков, вводимыми тегом ``, ситуация отчасти парадоксальная. В информационных выборках с ними всё хорошо, а вот как раз в стандартной коммерческой — уже давно не очень.

В 2015–2017 годах текстовые параметры, связанные с ``, были сильными и в Яндексе, и в Google, а с 2018 года в стали постепенно ослабевать. Корреляция с позицией для числа **вхождений слов запроса в ``** в Яндексе сошла на нет к концу 2018 года, в Google — примерно к концу 2022-го (но к началу 2024 года снова немного «отросла»). Контраст с фоном в Яндексе держался до середины 2021 года, потом стал

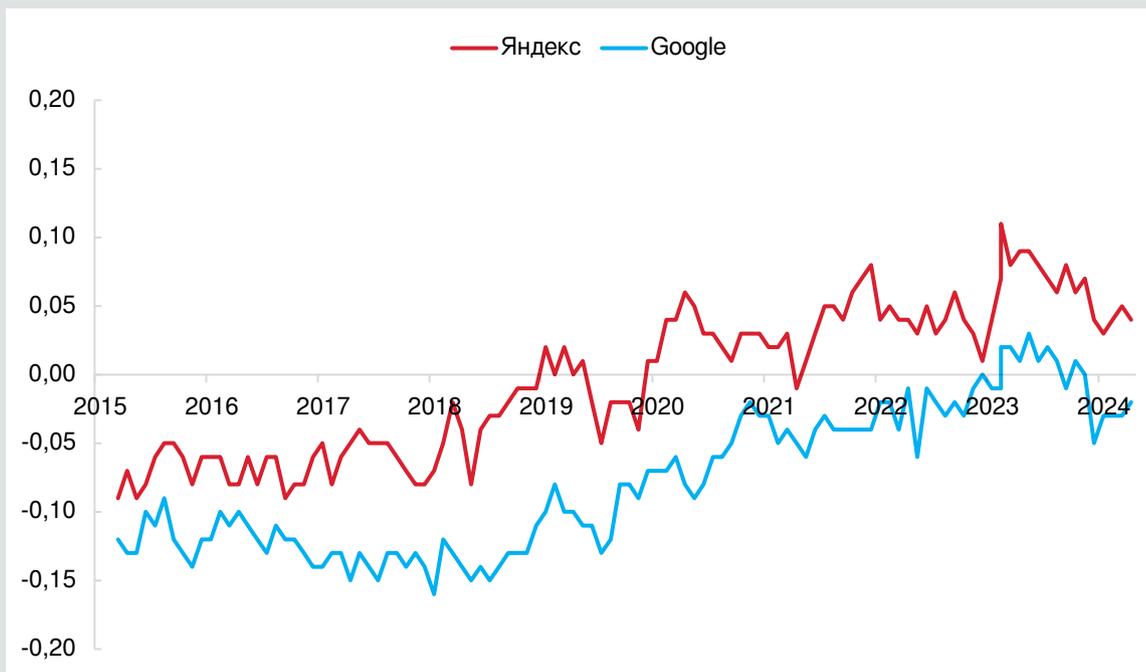
неустойчивым, а в конце 2022 года пропал окончательно.

Тем не менее, на информационных выборках параметры для `` не выглядят слабыми. Высокими показателями в Google полная выборка *Наука* обязана исключительно Википедии: на скорректированной выборке (т. е. без учета как раз Википедии) результаты резко падают. В *Гаджетах* параметры для `` (как и многие другие текстовые параметры) слабее, чем в двух других выборках, — особенно в Яндексе, так что в Google для скорректированных *Гаджетов* даже появляется контраст с фоном.

Косвенные данные также подтверждают неслучайность срабатывания текстовых параметров для элементов списков. В Яндексе **общая длина ``** и **количество слов в ``** для *Нау-*

Вхождения слов запроса в , стандартная коммерческая выборка

А. Корреляция с позицией (ранговый критерий Спирмена)

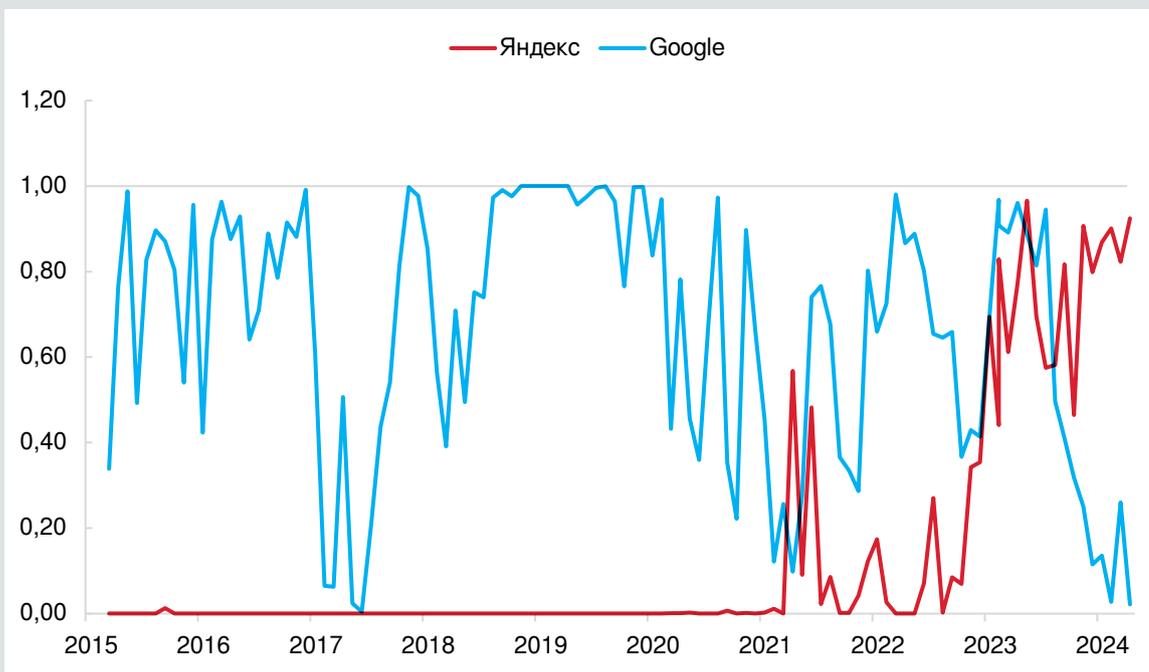


Динамика изменения коэффициента корреляции с позицией (ранговый критерий Спирмена) для количества вхождений слова запроса в элементы списков (тег) в Яндексе и Google для стандартной коммерческой выборки за девять лет наблюдений, 2015–2024. Отрицательные значения соответствуют «правильной» корреляции — то есть росту значений параметра к первой позиции. Положительные значения означают «обратную» корреляцию: чем выше результат, тем в среднем меньше значение параметра.

ки и *Косметики* — сильные параметры: есть контраст с фоном, а в *Науке* также и корреляция с позицией. В Google для всех информационных выборок есть корреляция с позицией, хотя в боль-

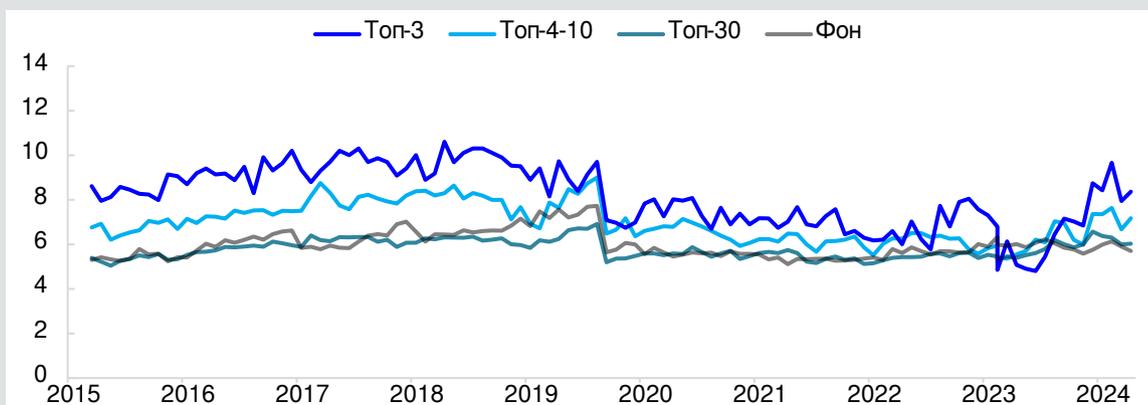
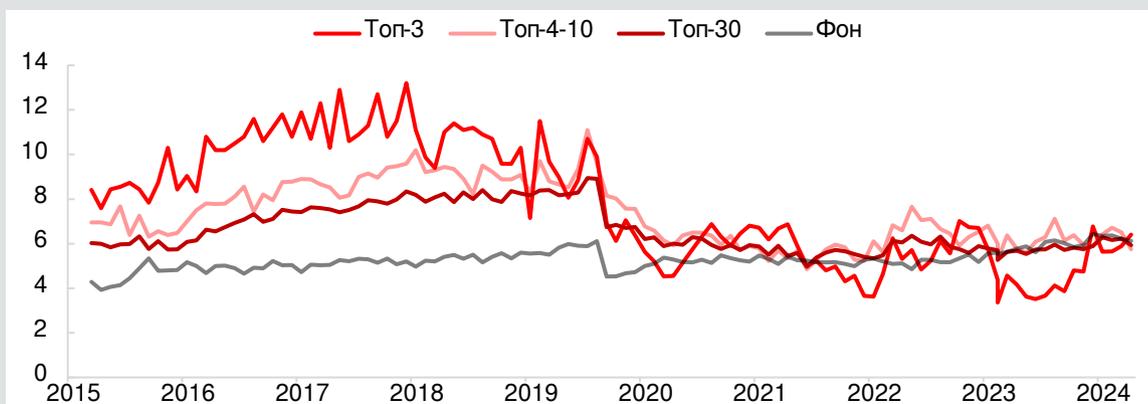
шинстве случаев и слабая (0,03). **Концентрация слов запроса в ** небольшая (на уровне текста страницы в целом), но в Яндексе в *Науке* и *Косметике* сильный контраст с фоном и кор-

Б. Топ-30 VS. фон (точный тест Фишера)



Вхождения слов запроса в `` для стандартной коммерческой выборки: динамика изменения точного теста Фишера для топ-30 Яндекса и Google в сравнении с фоном — страницами, не попавшими в топ-30, но «найденными» вторым поисковиком. Это показатель, по которому мы судим о наличии контраста с фоном. Значимый контраст с фоном есть только там, где точки на графике очень близки к нулю — как это было в Яндексе до середины 2021 года. Если график вплотную приближается к единице, показатели для топ-30 значительно ниже, чем для фона. Любые промежуточные значения указывают на отсутствие статистически значимого контраста.

В. Средние значения по диапазонам



Количество вхождений слов запроса в по диапазонам (топ-3, топ-4-10, топ-30, фон) в Яндексe (вверху) и в Google (внизу).

Синхронное резкое падение на графиках в сентябре 2019 года – явный артефакт. Мы не сразу сообразили, откуда он взялся, а когда сообразили, решили всё-таки показать здесь эти графики как есть, без пересчета. Не читайте дальше, если хотите сами догадаться, с чем могло быть связано резкое падение показателей при изменении способа расчета (вы можете, мы в вас верим!).

(Оно связано с переходом при подсчете этого параметра от среднего по запросу к медиане по запросу. При подсчете по среднему результаты с максимальными значениями параметра оказывают на итоговое значение большее влияние, чем при подсчете по медиане, где «выбросы» срезаются. Именно поэтому для параметров с большим разбросом значений мы вместо среднего используем медиану.)

Текстовые параметры: элементы списков ()

Яндекс

Параметры	Выделенные	Слова запроса + выделенные	Синонимы	Слова запроса + синонимы	Слова запроса	Неточный запрос + синонимы	Неточный запрос	Запрос с точностью до форм	Точный запрос	Слова запроса (полных запросов)	
										Топ-3	Топ-30
Наука	0,05	0,07	0,06	0,09	0,09	0,08	0,08	0,07	0,06	2,81	0,37
Наука ⁻¹		0,05	0,04	0,06	0,06	0,05	0,05	0,04		1,62	0,37
Косметика		0,03		0,03	0,03	0,04	0,04	0,03	0,03	5,81	3,62
Косметика ⁻²	0,04	0,06		0,06	0,07	0,05	0,05	0,04	0,03	6,46	3,82
Гаджеты										3,23	2,40
Гаджеты ⁻³				0,03						4,36	2,68
Станд. комм.										4,75	5,75

Google

Параметры	Выделенные	Слова запроса + выделенные	Синонимы	Слова запроса + синонимы	Слова запроса	Неточный запрос + синонимы	Неточный запрос	Запрос с точностью до форм	Точный запрос	Слова запроса (полных запросов)	
										Топ-3	Топ-30
Наука	0,13	0,17	0,08	0,16	0,16	0,11	0,11	0,10	0,08	2,63	0,24
Наука ⁻⁴	0,06	0,08	0,04	0,07	0,07	0,04	0,04	0,03		1,33	0,18
Косметика	0,04	0,08		0,08	0,08	0,07	0,06	0,04	0,03	4,66	3,00
Гаджеты	0,04	0,03		0,04	0,03					4,06	2,34
Гаджеты ⁻⁵	0,06	0,07	0,05	0,08	0,07	0,03				4,30	2,60
Станд. комм.										6,84	5,99

– сильный, средний, слабый контраст между топ-30 и фоном (точный тест Фишера).

– сильная, средняя, слабая корреляция с позицией.

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

Скорректированные выборки: 1: без Википедии, Youtube, Dzen; 2: без Youtube, Dzen; 3: без Apple, Youtube, Dzen; 4: без Википедии; 5: без Apple, Youtube.

реляция с позицией, а в Google — везде корреляция с позицией.

Атрибуты **alt** и **title** характерны для иллюстраций (чаще **alt**), клеток таблиц (чаще **title**) и некоторых других тегов.

Эти атрибуты «прячутся» внутри тегов и в текст страницы не входят — и потому дают дополнительную возможность подтвердить или повысить текстовую релевантность.

Заключенный в них текст становится видимым для пользователя при наведении мыши на элемент, кодируемый тегом, — например, на картинку. Такого текста набирается обычно не так уж и мало (особенно в хорошо оптимизированных *Косметике* и *Гаджетах*). В **alt** — обычно больше, чем в заголовках **<h2>**–**<h4>**, в **title** — примерно на том же уровне. **Объем текста в alt** и **в title** (в символах или в словах) для

большинства выборок в обоих поисковиках — довольно сильные параметры, так что их «заполнением» лучше не пренебрегать. **Концентрация слов запроса в alt** и **в title** заметно выше, чем в тексте страницы (хотя и ниже, чем во внутренних заголовках); в Яндексе на большинстве выборок для нее есть контраст с фоном, обычно сильный, и/или корреляция с позицией (в Google нет).

В Яндексе текстовые параметры для атрибутов **alt** и **title** достаточно сильные, хотя и по-разному на разных выборках. В Google параметры для **title** значительно слабее: корреляция с позицией есть фактически только для *Гаджетов* — если не считать полной *Науки*, которая «выезжает» только на Википедии (поскольку на скорректированной выборке никакой связи вхождений в **title** с ранжированием не видно). Заметим, что для стандартной коммерческой выборки связи с ранжированием в Google нет не только для **title**, но и для **alt**.

Текстовые параметры: атрибуты alt

Яндекс

Параметры	Выделенные	Слова запроса + выделенные	Синонимы	Слова запроса + синонимы	Слова запроса	Неточный запрос + синонимы	Неточный запрос	Запрос с точностью до форм	Точный запрос	Слова запроса (полных запросов)	
										Топ-3	Топ-30
Наука										0,31	0,03
Наука ⁻¹										0,70	0,05
Косметика										3,65	2,01
Косметика ⁻²										4,19	2,19
Гаджеты										1,52	0,96
Гаджеты ⁻³										2,13	1,13
Станд. комм.										4,16	5,99

Google

Параметры	Выделенные	Слова запроса + выделенные	Синонимы	Слова запроса + синонимы	Слова запроса	Неточный запрос + синонимы	Неточный запрос	Запрос с точностью до форм	Точный запрос	Слова запроса (полных запросов)	
										Топ-3	Топ-30
Наука										0,23	0,01
Наука ⁻⁴										0,43	0,02
Косметика										4,26	1,53
Гаджеты										1,75	0,67
Гаджеты ⁻⁵										1,87	0,79
Станд. комм.										5,73	4,04

– сильный, средний, слабый контраст между топ-30 и фоном (точный тест Фишера).

– сильная, средняя, слабая корреляция с позицией.

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

Скорректированные выборки: 1: без Википедии, Youtube, Dzen; 2: без Youtube, Dzen; 3: без Apple, Youtube, Dzen; 4: без Википедии; 5: без Apple, Youtube.

Текстовые параметры: атрибуты title

Яндекс

Параметры	Выделенные	Слова запроса + выделенные	Синонимы	Слова запроса + синонимы	Слова запроса	Неточный запрос + синонимы	Неточный запрос	Запрос с точностью до форм	Точный запрос	Слова запроса (полных запросов)	
										Топ-3	Топ-30
Наука	0,07	0,11	0,06	0,13	0,13	0,11	0,11	0,11	0,10	1,74	0,00
Наука ⁻¹										0,28	0,00
Косметика										0,73	0,26
Косметика ⁻²										0,79	0,21
Гаджеты	0,03	0,07	0,03	0,07	0,06	0,05	0,04	0,03	0,03	1,14	0,38
Гаджеты ⁻³		0,04		0,04	0,03					1,16	0,33
Станд. комм.	0,07	0,03	0,06			0,05			0,03	8,37	2,37

Google

Параметры	Выделенные	Слова запроса + выделенные	Синонимы	Слова запроса + синонимы	Слова запроса	Неточный запрос + синонимы	Неточный запрос	Запрос с точностью до форм	Точный запрос	Слова запроса (полных запросов)	
										Топ-3	Топ-30
Наука	0,13	0,09	0,08	0,09	0,09	0,04	0,04	0,04	0,04	1,32	0,00
Наука ⁻⁴										0,15	0,00
Косметика										0,32	0,18
Гаджеты	0,08	0,08	0,09	0,10	0,07	0,05	0,03			1,23	0,36
Гаджеты ⁻⁵	0,06	0,06	0,07	0,07	0,05	0,03				1,25	0,31
Станд. комм.										4,75	1,91

— сильный, средний, слабый контраст между топ-30 и фоном (точный тест Фишера).

— сильная, средняя, слабая корреляция с позицией.

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

Скорректированные выборки: 1: без Википедии, Youtube, Dzen; 2: без Youtube, Dzen; 3: без Apple, Youtube, Dzen; 4: без Википедии; 5: без Apple, Youtube.

Тексты входящих ссылок

Выше мы уже посмотрели на ссылочно-текстовые параметры как на ссылочные — и убедились, что ссылки на сайт гораздо важнее ссылок на страницу. Теперь мы можем рассмотреть их и с точки зрения принципов текстового ранжирования.

Матрицы ссылочно-текстовых параметров (напомним, что по техническим причинам они есть только для двух из трех информационных выборок — *Науки* и *Гаджетов*) во многом похожи на аналогичные таблицы для текстовых параметров — но есть и отличия.

Во-первых, корреляции с позицией здесь есть и в Яндексе, и в Google — причем даже для ссылок на страницу по меркам текстовых параметров они весьма сильные, а уж о корреляциях с коэффициентом 0,25, 0,30 или 0,32, которые встречаются для ссылок на сайт, текстовые параметры и мечтать не могут.

Во-вторых, контраст с фоном здесь больше характерен для Google, чем для Яндекса. В этом отношении ссылочно-текстовые параметры похожи на просто ссылочные (связанные с общим количеством ссылок безотно-

сительно к их текстам), см. выше раздел [Ссылки на сайт](#).

Для ссылочно-текстовых параметров, как и для просто текстовых, действует принцип антибуквализма — хотя и здесь есть нюансы.

Самые сильные параметры, как обычно — самые далекие от точного запроса: это слова запроса «россыпью» и они же с учетом синонимов. Но и на точном конце спектра здесь часто сохраняется корреляция с позицией или контраст с фоном, хотя и более слабые. Как и с текстовыми параметрами, в Google это случается чаще, чем в Яндексе, а из информационных выборок это более характерно для *Науки*, поскольку в ней выше доля коротких (одно- и двухсловных) запросов.

Ссылочно-текстовые параметры: ссылки на сайт

Яндекс

Параметры	Выделенные	Слова запроса + выделенные	Синонимы	Слова запроса + синонимы	Слова запроса	Неточный запрос + синонимы	Неточный запрос	Запрос с точностью до форм	Точный запрос	Слова запроса (полных запросов)	
										Топ-3	Топ-30
Наука		0,16		0,14	0,14	0,08	0,08	0,07	0,05	1,78	0,45
Наука ⁻¹		0,15		0,14	0,15	0,09	0,09	0,08	0,06	3,90	0,41
Гаджеты		0,29		0,32	0,24	0,09	0,07			93,0	1,39
Гаджеты ⁻³		0,24		0,26	0,22	0,09	0,09			34,6	1,28
Станд. комм.		0,12		0,07	0,08					10,7	5,35

Google

Параметры	Выделенные	Слова запроса + выделенные	Синонимы	Слова запроса + синонимы	Слова запроса	Неточный запрос + синонимы	Неточный запрос	Запрос с точностью до форм	Точный запрос	Слова запроса (полных запросов)	
										Топ-3	Топ-30
Наука		0,14		0,15	0,15	0,09	0,09	0,08	0,07	2,85	0,71
Наука ⁻⁴		0,13		0,14	0,15	0,10	0,09	0,09	0,07	4,51	0,67
Гаджеты		0,25		0,30	0,23	0,07	0,06	0,03	0,03	275	2,78
Гаджеты ⁻⁵		0,20		0,22	0,19	0,06	0,06	0,03	0,03	27,0	2,40
Станд. комм.		0,22		0,18	0,18	0,08	0,08	0,08	0,07	18,0	8,01

— сильный, средний, слабый контраст между топ-30 и фоном (точный тест Фишера).

— сильная, средняя, слабая корреляция с позицией.

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

Скорректированные выборки: 1: без Википедии, Youtube, Dzen; 3: без Apple, Youtube, Dzen; 4: без Википедии;

5: без Apple, Youtube.

Ссылочно-текстовые параметры: ссылки на страницу

Яндекс

Параметры	Выделенные	Слова запроса + выделенные	Синонимы	Слова запроса + синонимы	Слова запроса	Неточный запрос + синонимы	Неточный запрос	Запрос с точностью до форм	Точный запрос	Слова запроса (полных запросов)	
										Топ-3	Топ-30
Наука		0,10		0,10	0,10	0,09	0,08	0,08	0,06	0	0
Наука ⁻¹		0,07		0,08	0,08	0,07	0,07	0,06	0,05	0	0
Гаджеты		0,08		0,09	0,08	0,05	0,05	0,05	0,05	0	0
Гаджеты ⁻³		0,09		0,11	0,09	0,06	0,06	0,06	0,06	0,06	0
Станд. комм.					0,03					0	0

Google

Параметры	Выделенные	Слова запроса + выделенные	Синонимы	Слова запроса + синонимы	Слова запроса	Неточный запрос + синонимы	Неточный запрос	Запрос с точностью до форм	Точный запрос	Слова запроса (полных запросов)	
										Топ-3	Топ-30
Наука		0,13		0,12	0,12	0,08	0,08	0,07	0,05	0	0
Наука ⁻⁴		0,10		0,08	0,08	0,06	0,05	0,05	0,04	0	0
Гаджеты		0,10		0,10	0,10	0,06	0,05	0,03	0,03	0	0
Гаджеты ⁻⁵		0,10		0,10	0,10	0,06	0,06	0,04	0,04	0,05	0
Станд. комм.		0,11		0,09	0,10	0,10	0,10	0,10	0,10	0,17	0

– сильный, средний, слабый контраст между топ-30 и фоном (точный тест Фишера).

– сильная, средняя, слабая корреляция с позицией.

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

Скорректированные выборки: 1: без Википедии, Youtube, Dzen; 3: без Apple, Youtube, Dzen; 4: без Википедии; 5: без Apple, Youtube.

Малопольный контент

Увы, данные по информационным выборкам были собраны раньше, чем Яндекс ввел в действие обновленный фильтр «Малопольный контент», так что по нашим результатам судить о нём нельзя.

Но о фильтрах вообще трудно судить по общей картине ранжирования, особенно вскоре после начала их применения. Дело в том, что фильтры в принципе предусматривают точечные и отчасти непредсказуемые санкции для тех, кто слишком прямолинейно использует приемы, которые, как показывает практика, помогают продвигать сайты. Если бы поисковая машина могла исправить нежелательные эффекты корректировкой основной формулы, без дополнительных механизмов (фильтра), — это, конечно, было бы сделано.

То есть санкции по фильтру всегда работают в направлении, противоположном основной формуле ранжирования, и смазывают картину.

Оптимизированные тексты на сайте — это хорошо и помогает попасть в топ и/или занять более высокую позицию, но только до тех пор, пока поисковая машина не решает, что ею пытаются манипулировать, и не отбрасывает сайт далеко назад.

Пессимизация отдельных сайтов за то, что обычно помогает попадать в топ и/или улучшать позиции, уменьшает корреляцию соответствующих параметров с позицией. Риск такой пессимизации, висящий над всеми, приводит к снижению значений параметров — особенно у успешных сайтов, уже занявших высокие позиции, которым есть что терять и которые надеются сохранить позиции за счет других факторов ранжирования или просто по инерции (которая часто достаточно сильна). Это тоже ослабляет те факторы, со злоупотреблением которыми борется фильтр.

Мы пока не планируем подробно изучать «Малополезный контент» и создавать сервис для оценки риска применения этого фильтра. В отличие от прогремевшего в свое время «Баден-Бадена», ввод «Малополезного контента», кажется, прошел более тихо, и поток жалоб с мест оказался недостаточным для того, чтобы можно было подробно изучить особенности «правоприменения» в этом случае — для этого нам просто не хватает материала.

Тем не менее, если ваш сайт попал под «Малополезный контент», напишите нам — нам будет интересно изучить ваш кейс, и мы с удовольствием обсудим с вами свои соображения и выводы.

Главная рекомендация, которую мы можем дать, — пользоваться «Тургеневым». Всё, что мы имели сказать

о плохих текстах, мы уже сказали — и, хотя «Баден-Баден» к информационным сайтам не применяется, проверять их тексты очень даже имеет смысл — хотя бы для того, чтобы почистить от канцелярита и тавтологий. Если «Тургенев» показывает для информационного текста высокие штрафные баллы, это означает, что риск «малополезного контента» для него далеко не нулевой.

Впрочем, качество текста — важная, но не единственная составляющая такого риска. Фильтр применяется в первую очередь к сайтам, на которых много рекламы, контент которых оценивается как неоригинальный или недостаточно авторитетный; поэтому следует обращать внимание также на всё, что связано с E—A—T-факторами.

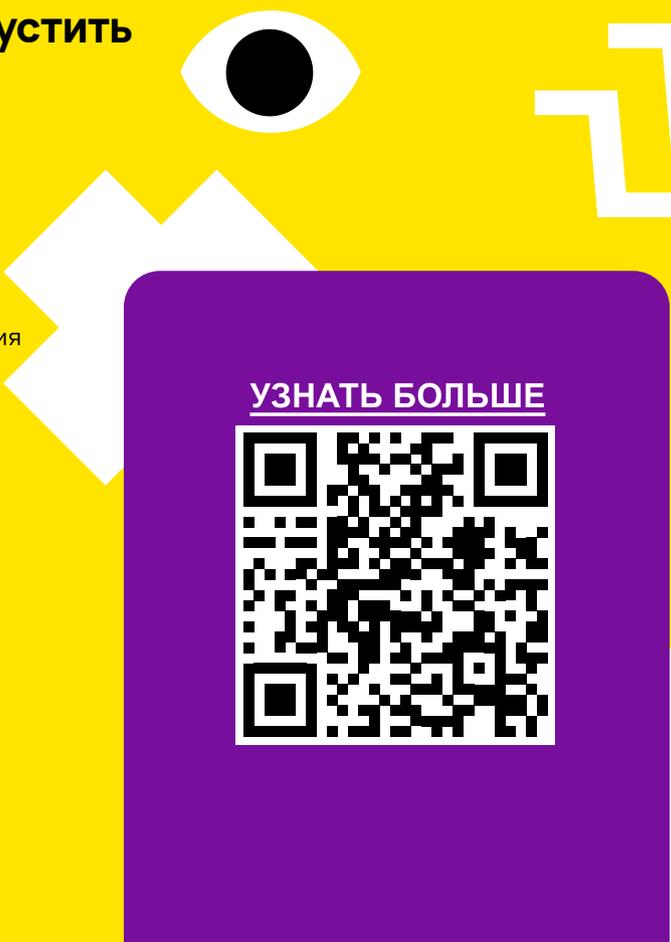


OPTI
MIZA
TION

OPTIMIZATION

КРУПНЕЙШАЯ ПРАКТИЧЕСКАЯ КОНФЕРЕНЦИЯ
ПО ПОИСКОВОМУ МАРКЕТИНГУ

Почему конференцию нельзя пропустить



1.

Знания и польза

Глубокий экспертный контент, который расширит ваши профессиональные знания и поможет найти точки роста для вашего бизнеса

2.

Знакомства и общение

Во время конференции предусмотрены мероприятия и механики, которые помогут участникам завести полезные знакомства! Общайтесь и расширяйте свою сеть профессиональных контактов.

3.

Вечеринка

Легендарная вечеринка Optimization – это возможность пообщаться с коллегами по цеху в неформальной обстановке. А также посмотреть премию Optimization, которой наградят лучшие по видимости сайты в разных тематиках.

[УЗНАТЬ БОЛЬШЕ](#)





Ашманов
и партнеры

В третьей части

Коммерческие сайты в информационной выдаче

- Какие страницы находятся
- Антикоммерческие факторы: Яндекс
- Сайтовые или страницные?
- Чего не должно быть на информационных страницах
- Телефон 8-800
- Антикоммерческие факторы в Google?

Про Науку

- «Научные авторитеты»
- Коммерческие сайты в Науке

