

# Медиана выдачи вместо чек-листа

Что на самом деле отличает топ-20 Google:  
7370 страниц против SEO-чеклиста

Станислав Кириченко

SEO-специалист · sk-seo.ru



160

запросов

7370

страниц

16

ниш

Google · Москва · десктоп · 14.09.2026



**Станислав Кириченко**

ML SEO Tech Lead · Staurus

## КТО ГОВОРИТ

# Двадцать лет в SEO: от своих проектов до enterprise и собственных инструментов

**20 лет**

в SEO, с 2006 года

**200+**

клиентских проектов

**108 млн**

страниц в одном  
аудите

**14 лет**

авторского SEO-  
блога

### ЗАКАЗЧИКИ

Rush Agency: Avito, Hoff, Яндекс.Бизнес, pari.ru. Tea.ru: трафик ×10, прибыль ×20. Интурист: онлайн-заказы +1300%

### СЕЙЧАС

ML SEO Tech Lead: AI/ML-инструменты для SEO, автоматизация, исследования выдачи на живых данных

### ОТКРЫТЫЕ ДАННЫЕ

Препринты на Zenodo, данные и код своих исследований в открытом доступе

### СООБЩЕСТВО

Блог Staurus с 2008 года, супермодератор webmasters.ru. Выступал на Digital Day Belarus и DevCom

# Топ-20 отличает охват темы. Ключи и вес страницы — нет

1

## Охват темы

Из всех текстовых признаков только он пережил все проверки. Ключ в title и H1 порядок мест почти не различает.

2

## Медиана — линейка

Типичная страница выдачи помогает найти провалы. Порогом входа и гарантией она не служит.

3

## Подгоняем своё

Что принадлежит странице (темы, вопросы), подгоняем под медиану. Что принадлежит сайту (вес, скорость), не трогаем.

**Главный посыл:** если со страницей что-то не так, сверьте её с медианой выдачи и подгоните там, где это работает.

# Что мерил и как читать цифры

## ПАСПОРТ ЗАМЕРА

|          |                                                |
|----------|------------------------------------------------|
| Выборка  | 160 запросов, 16 ниш по десять                 |
| Страницы | 7370 адресов, 4432 домена                      |
| Выдача   | Google, Москва, google.ru, десктоп, 14.09.2026 |
| Группы   | топ-20 против позиций 21–50                    |
| Интент   | 1228 документов, слепая оценка                 |
| Считали  | текст, охват темы, вопросы, техничка           |

## ШКАЛА «ИЗ 100 ПАР»

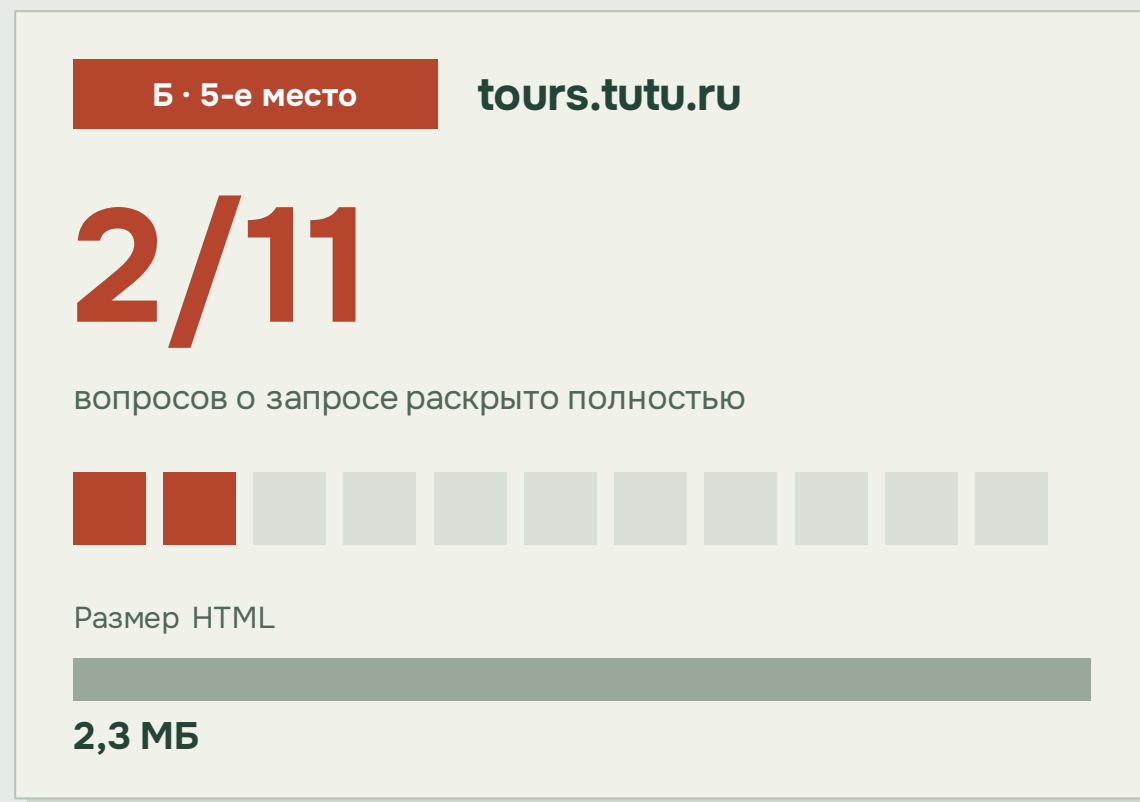
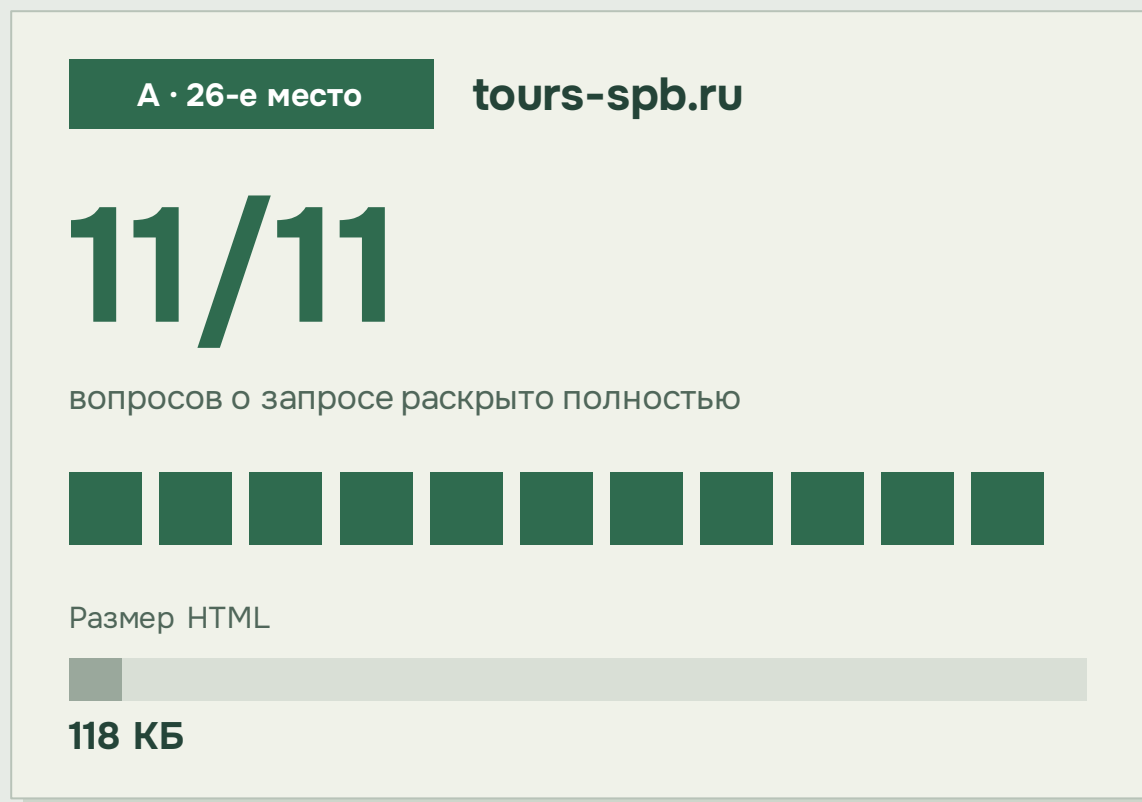
Берём две страницы одной выдачи. У какой признак больше, та стоит выше? Считаем по всем парам.



53,8: по охвату темы верхняя страница угадана в 54 парах из 100, в 46 — нет. Это диагностика, а не прогноз роста.

**Медиана** — типичная страница: у половины страниц значение больше, у половины меньше. Один гигантский сайт её не перекосит.

# На 5-м месте страница Б, на 26-м страница А



«туры в турцию все включено», Google, Москва, 14.09.2026. Одна выдача ничего не доказывает: полнота силу площадки не перебила, но это не значит, что она бесполезна. Проверяем на 160 запросах.

# Ни один из 20 признаков не работает как рычаг: лучший угадывает верхнюю страницу пары в 55 случаях из 100

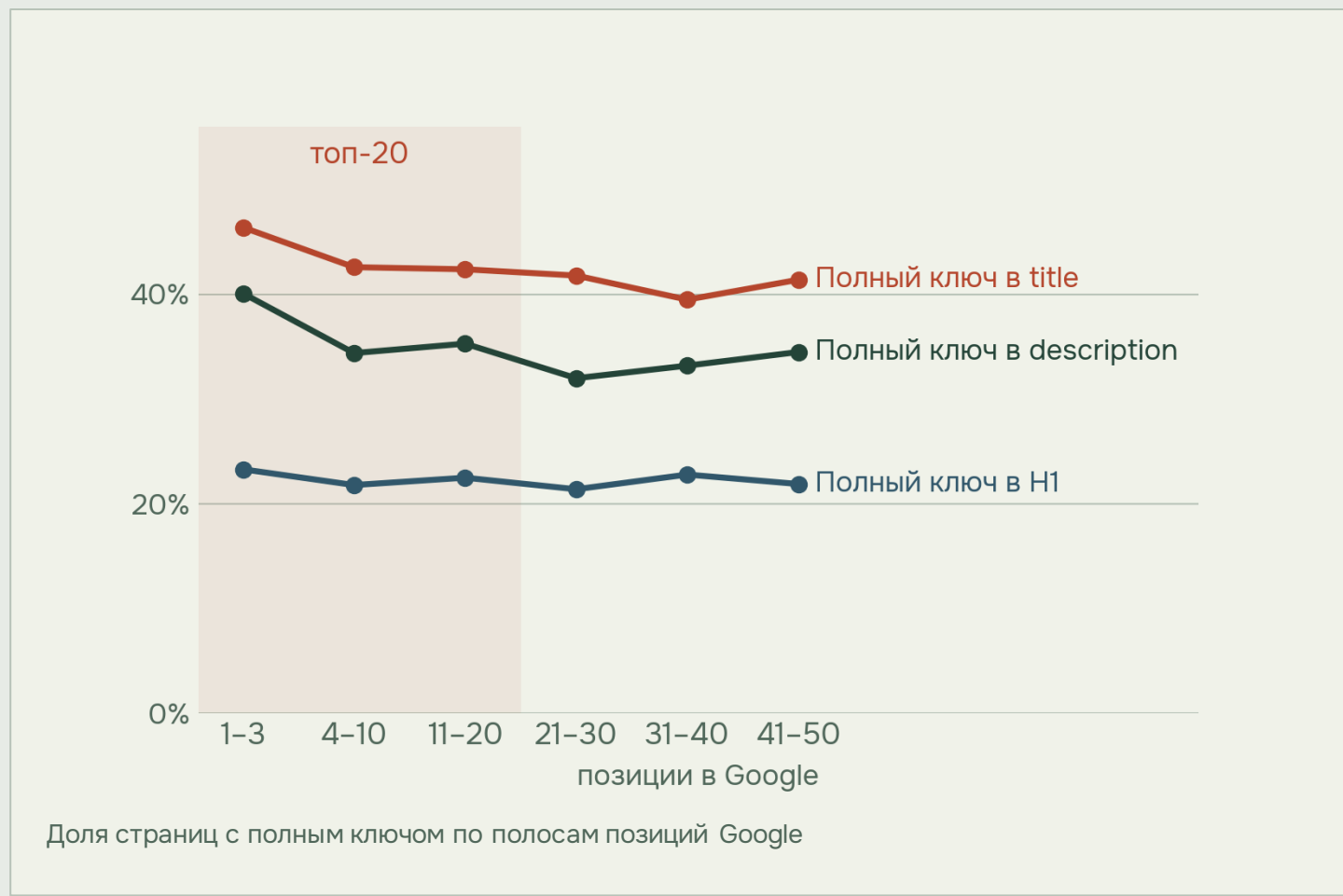


## КАК ЧИТАТЬ

- Техника: слова, DOM, HTML, картинки
- Текст: заголовки, ключи, объём
- Две главные величины: охват темы и полнота интента

**Наверху объёмы.** Из текстовых первым идёт охват темы. **55,4** значит: слов больше у верхней страницы в 55 парах из 100. Ключи лежат на 50, это монетка.

# Ключ в title, H1 и description порядок мест почти не различает



**43,0% и 40,8%**

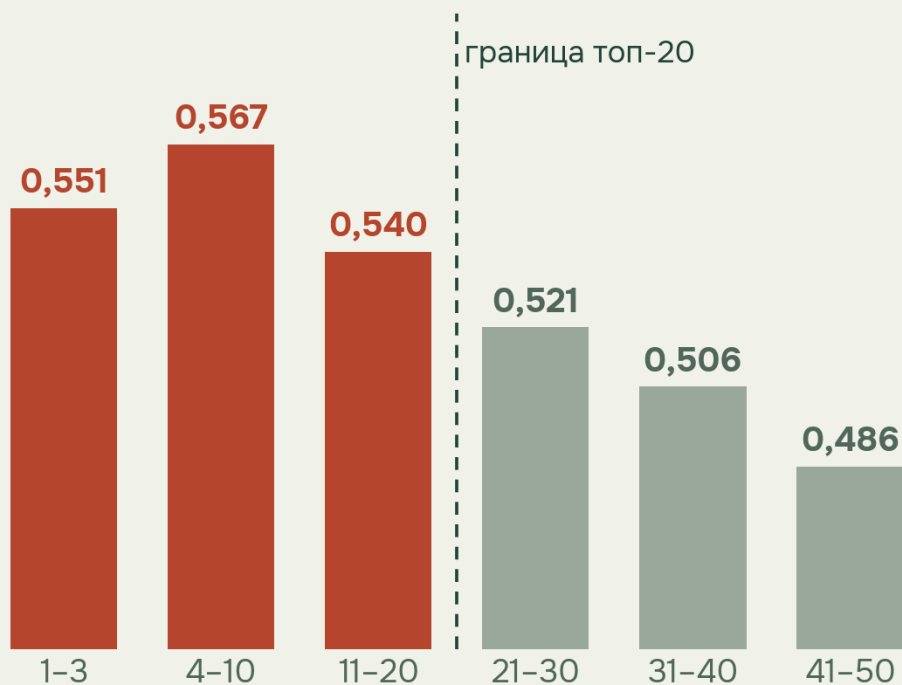
полный ключ в title: у 43 страниц из 100 в топ-20 и у 41 из 100 ниже

**50,0 из 100**

ключ в H1 угадывает верхнюю страницу в 50 парах из 100: монетка. Плотность ключа: 49

**В ТЗ:** плотность ключа убрать. Для страницы, уже вошедшей в топ-50, ещё один ключ в title и H1 рычагом не выглядит. Сам title нужен для релевантности и клика.

# Охват темы — единственный текстовый признак, который пережил все проверки



позиции в Google · медиана доли тематических слов запроса

Внутри топ-20 ровно, дальше каждая страница беднее предыдущей

## 55,3 из 100

в 55 парах из 100 охват шире у верхней страницы, в 45 — у нижней

### ПЫТАЛСЯ СЛОМАТЬ — НЕ СЛОМАЛОСЬ

- ✓ **Просто длинные тексты?**  
при равной длине охват отличается топ
- ✓ **Просто сильные сайты?**  
внутри сайта остаётся половина разницы
- ✓ **Несколько огромных выдач?**  
верх богаче в 111 запросах из 160
- ✓ **Одна ниша?**  
верх богаче в 15 нишах из 16
- ✓ **Агрегаторы и маркетплейсы?**  
без них результат тот же: 55,3

# Охват темы в словах: сколько терминов выдачи должно быть на странице

| Уровень                              | Доля списка | ≈ слов при списке 80 |
|--------------------------------------|-------------|----------------------|
| Красный флаг: нижняя четверть топ-50 | 0,37        | ≈ 30                 |
| Медиана топ-50                       | 0,53        | ≈ 42                 |
| Медиана топ-20                       | 0,56        | ≈ 45                 |
| Верхняя четверть топ-20              | 0,70        | ≈ 56                 |
| Верхняя десятая часть топ-20         | 0,79        | ≈ 63                 |

Слова — это термины из списка выдачи вашего запроса: слово или пара слов. Пересчёт доли в штуки сделан по данным исследования, в статье даны доли.

**СПИСОК СВОЙ У КАЖДОГО ЗАПРОСА**

Медиана списка — около 80 терминов, у средней половины запросов от 67 до 90. Поэтому считайте долю от своего списка. Список из 60 терминов: медиана ≈ 32 слова.

**Топ-20 против 21–50:** медиана 40 найденных терминов против 37,5.

Разница мала: это ориентир для поиска провала, а не норматив.

# Полнота ответа и длина текста слабее, чем ждали

ПОЛНОТА ИНТЕНТА (СЛЕПАЯ ОЦЕНКА, 1228 ДОКУМЕНТОВ)

52,7

в 53 парах из 100 верхняя закрывает больше вопросов: 5 из 10 против 4. Перевес мог дать случай

≈ 0

у страниц одной длины и типа преимущество полноты почти исчезает

56,3

позиции 1–10 против 31–49: верхняя закрывает больше в 56 парах из 100

ОБЪЁМ ТЕКСТА

840 и 700

слов основного текста у типичной страницы топ-20 и у страниц ниже. Подзаголовков H2 шесть против пяти.

**Почему «неясно»:** при одинаковом охвате темы преимущество длины может оказаться случайным. Длинный текст просто закрывает больше тем.

Так же пишет Google: длина сама по себе на ранжирование не влияет.

# Технический вес в основном принадлежит сайту, а не странице

НАСКОЛЬКО HTML ТОП-20 ТЯЖЕЛЕЕ ПОЗИЦИЙ 21–50

**+39,2 КБ** → **+5,8 КБ**

между разными сайтами

внутри одного сайта (505 сайтов)

**606 и 601 мс**

время ответа у топ-20 и у позиций 21–50: верхнюю страницу угадывает в 49,8 парах из 100, то есть монетка

**4 из 5**

страниц топ-20 с претензией краулера. Без единой: 20 из 100 в топ-20 и 23 из 100 ниже

СКОЛЬКО РАЗРЫВА ОСТАЁТСЯ ВНУТРИ САЙТА

Размер HTML 15%

DOM-узлы 12%

Слов на странице 18%

Охват темы (LSI) 52%

**Вывод:** вес и скорость к топу не гоним; недоступность, мягкие 404 и закрытие от индекса чиним всегда. Претензия краулера не равна причине низких позиций.

# Что говорят Google, Backlinko и «Ашманов и партнёры»

## Google Search Central

SEO Starter Guide

**«The length of the content alone doesn't matter for ranking purposes»**

Длина контента сама по себе на ранжирование не влияет

## Backlinko

11,8 млн результатов Google

**«essentially zero correlation between using a keyword in your title tag and higher rankings»**

**«Page HTML size has no relationship with rankings»**

Ключ в title и размер HTML с позицией почти не связаны

## Ашманов и партнёры

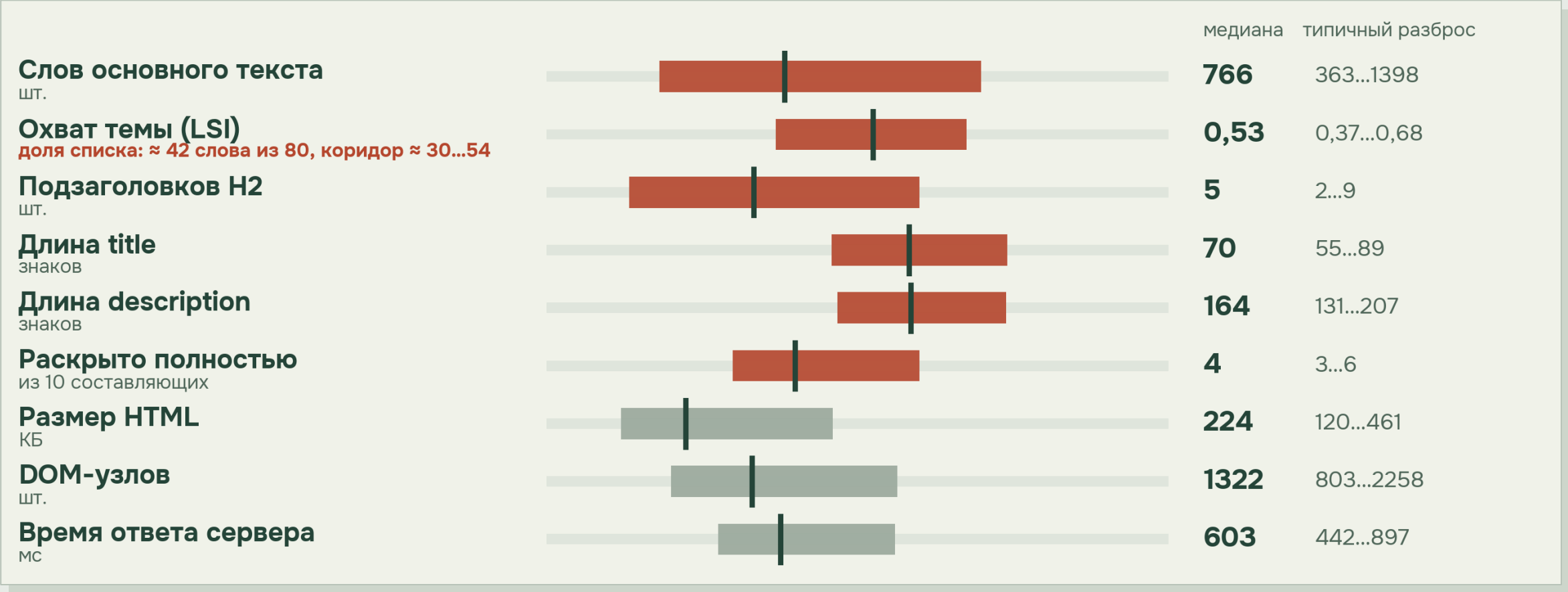
статья от 9 сентября 2024

**«Значения текстовых параметров в Google ниже, но зависимость позиций от них сильнее и наблюдается чаще»**

Зависимость позиций от текстовых параметров в Google сильнее и наблюдается чаще

**Здесь я расхожусь с Backlinko:** у меня страницы топ-20 тяжелее (246 и 207 КБ), но внутри одного сайта разрыв почти исчезает (39,2 → 5,8 КБ).

# Портрет типичной страницы топ-50 Google



Полоса — типичный разброс: внутри неё лежит половина страниц. Чёрная риска — медиана. Позиции 1–50 без деления на верх и низ: текст — 6516 страниц, охват темы — 5964, техничка — 5753, полнота — 1228 документов. **Терракотовым — то, что подгоняем; серым — техничка, которую не гоним.**

# Что подгонять под медиану, а что нет

## Подгонять под медиану выдачи

- Темы, которые раскрывают соседи по выдаче вашего запроса
- Вопросы, на которые не ответил никто из соседей
- Отдельный раздел H2 под каждую тему
- Полнота ответа не ниже типичного соседа

## Обеспечить, не гнать

- Индексируемость и доступность страницы
- Контент отдаётся в HTML
- Страница сделана под этот запрос
- Понятный title под клик

## Не рычаг позиций

- Ключ в title, H1 и description
- Плотность ключа
- Число слов «как у топа»
- Вес HTML и DOM
- Время ответа сервера

В данных это связь, а не эксперимент: поднимет ли правка вашу страницу, покажет только правка ваших страниц.

# Красные флаги и как посчитать медиану своей выдачи

## КРАСНЫЕ ФЛАГИ: НИЖНЯЯ ГРАНИЦА ТИПИЧНОГО РАЗБРОСА

**< 360**

слов основного текста

**< 0,37**

охват темы:  $\approx 30$  слов из списка в 80

**< 2**

подзаголовков H2

**< 3 из 10**

составляющих интента раскрыто полностью

**Это не порог входа.** Ниже каждой границы лежит четверть страниц топ-50.

## КАК ПОСЧИТАТЬ ДЛЯ СВОЕГО ЗАПРОСА

**1**

### Снять выдачу

топ-50 вашего запроса

**2**

### Вырезать основной текст

без меню и подвала

**3**

### Собрать тематические слова

встречаются минимум на 30% страниц и втрое чаще, чем в других нишах

**4**

### Собрать вопросы

Вордстат, подсказки, «Похожие вопросы»; страницы выдачи в список не брать

**5**

### Сравнить страницу с медианой

охват, вопросы, структура H2

**Общая медиана — стартовый ориентир. Мерило — выдача вашего запроса.**

# Четыре ситуации: куда двигаться, сравнив себя с соседями по выдаче

|                | ТЕМА УЖЕ, ЧЕМ У СОСЕДЕЙ                                                                                                                                     | ТЕМА ШИРЕ, ЧЕМ У СОСЕДЕЙ                                                                                                                                                                |
|----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| ВОПРОСЫ ПОЛНЕЕ | <div><div>Добрать темы</div><div>51,6%</div><div>Соберите тематические слова выдачи и допишите разделы под недостающие темы.</div><div>в топ-20</div></div> | <div><div>Текст не узкое место</div><div>56,5%</div><div>Проверьте, та ли это страница сайта под запрос. Затем ссылки, бренд, коммерческие свойства.</div><div>в топ-20</div></div>     |
| ВОПРОСЫ СЛАБЕЕ | <div><div>Дописывать</div><div>46,3%</div><div>Закройте вопросы и темы, которые уже закрыты у соседей по выдаче.</div><div>в топ-20</div></div>             | <div><div>Добрать вопросы</div><div>51,6%</div><div>Тема раскрыта шире соседей, а на вопросы пользователя страница отвечает хуже. Закройте составляющие.</div><div>в топ-20</div></div> |

Доли — сколько документов слепой оценки стоят в топ-20; верх и низ в выборке поровну, в среднем 50,7%. Простыми словами: из 100 страниц, слабых по обоим признакам, в топ-20 стоят около 46, из 100 сильных по обоим — около 57. Обе клетки «по одному признаку» дают 51,6% вместе. Это не прогноз шансов, а разведочный разрез.

# Дешевле всего обойти соседей там, где ни один проверенный документ не ответил

146 из 1533

составляющих интента не раскрыл полностью ни один проверенный документ (9,5%)

258

составляющих раскрыты меньше чем у 15% документов своего запроса

81 из 159

запросов с такой дырой, во всех шестнадцати нишах

ЧТО ОСТАЛОСЬ БЕЗ ОТВЕТА В ТОПЕ

| Запрос                               | Вопрос, на который никто не ответил |
|--------------------------------------|-------------------------------------|
| секция единоборств для детей         | требования к медицинской справке    |
| компьютерная диагностика автомобиля  | бесплатная диагностика              |
| замена масла в двигателе стоимость   | объём масла для замены              |
| лечение кариеса стоимость            | лечение по ОМС                      |
| мультиварка или скороварка что лучше | сравнение с аэрогрилем              |

Прежде чем закрывать дыру, убедитесь, что вопрос относится к вашей странице: «сравнение с аэрогрилем» там может быть лишним.

# remo-nt.ru: 45-е место при странице, которая сильнее соседей по тексту

«СКОЛЬКО СТОИТ РЕМОНТ ОДНОКОМНАТНОЙ КВАРТИРЫ» · СТРАНИЦА УСЛУГИ С ЦЕНАМИ

| Признак                          | Страница             | Коридор топ-20            | Статус           |
|----------------------------------|----------------------|---------------------------|------------------|
| Раскрыто полностью, составляющих | 10 из 11             | 3...6                     | выше края        |
| Охват темы (LSI)                 | 0,88 (79 из 90 слов) | 0,391...0,700 (≈ 31...56) | выше края        |
| Слов основного текста            | 2247                 | 408...1480                | верхняя четверть |
| Размер HTML, КБ                  | 185                  | 141...504                 | в норме          |
| DOM-узлов                        | 1637                 | 899...2419                | в норме          |
| Ответ сервера, мс                | 636                  | 436...899                 | в норме          |
| Полный ключ в title              | нет                  | есть у 43% топа           | не различает     |

Пример подобран вручную уже после данных: это иллюстрация, доказательством он не служит.

## СОСЕДИ В ТОП-20

**6-е место, profi.ru:** 0 из 11 вопросов, охват 0,03.

Агрегатор с узнаваемым брендом стоит выше страницы, которая на вопрос о цене отвечает прямо.

## ЧТО СМОТРЕТЬ ДАЛЬШЕ

- Ссылки на страницу и на сайт
- Бренд и брендовые запросы
- Коммерческие свойства: цены, отзывы, гарантии
- Замер этого не измерял, это гипотезы.**

# Из девяти проверенных пунктов чек-листа во всех проверках устоял один

| Ожидание из чек-листа       | Что в данных                                                   | Итог           |
|-----------------------------|----------------------------------------------------------------|----------------|
| Ключ в title и H1 поднимает | в топ-20 у 43 страниц из 100, ниже у 41; H1: монетка           | не различает   |
| Плотность ключа             | 49 пар из 100                                                  | не различает   |
| Больше слов, как у топа     | 840 и 700 слов; при равном охвате перевес может быть случайным | неясно         |
| Полнее ответ на вопросы     | 5 вопросов из 10 против 4; при равной длине почти ноль         | слабо          |
| Лёгкий HTML                 | топ тяжелее, но внутри сайта разрыв 39,2 → 5,8 КБ              | свойство сайта |
| Быстрый сервер              | 606 и 601 мс                                                   | не различает   |
| Google наказывает ИИ-текст  | выше порога 10,5% и 10,9%                                      | не различает   |
| Меньше воды                 | водность 23,7% и 23,5%                                         | не различает   |
| Шире охват темы             | 55 пар из 100; пережил пять проверок                           | держится       |

# Чего замер не знает и что сделать в понедельник

## ЧЕГО ЗАМЕР НЕ ЗНАЕТ

Ссылки на страницу и на сайт

Бренд и поведение пользователей

Коммерческие свойства, возраст домена

Мобильную выдачу: только десктоп

Один регион и одна дата снятия

**Причинность.** Данные показывают, что идёт вместе. Эксперимента нет.

## В ПОНЕДЕЛЬНИК

**1**

Снимите топ-50 запроса; проверьте индексируемость и что ранжируется нужная страница

**2**

Посчитайте медиану этой выдачи по охвату темы и вопросам

**3**

Ниже по обоим — допишите темы и ответы; по одному — закройте разрыв; выше по обоим — ссылки, бренд, коммерция

**4**

Найдите вопрос, на который не ответил никто из соседей

**5**

Правьте страницы слабее медианы и замеряйте сдвиг: эксперимента в данных нет

МАТЕРИАЛЫ

# Спасибо. Вопросы?

Статья с полными данными

**[sk-seo.ru/blog/что-отличает-top-20-google.html](https://sk-seo.ru/blog/что-отличает-top-20-google.html)**

Препринт, данные по 7370 URL и код

**Zenodo · DOI 10.5281/zenodo.22796116**

SEO-анализ текста

**[sk-seo.ru/tools/seo-analiz-teksta.html](https://sk-seo.ru/tools/seo-analiz-teksta.html)**

Детектор ИИ-текста

**[sk-seo.ru/tools/detektor-ii-teksta.html](https://sk-seo.ru/tools/detektor-ii-teksta.html)**



Статья и все таблицы

Станислав Кириченко

# Приложение

Слайды для вопросов из зала: Google и Яндекс, ИИ-текст и вода, полная таблица медиан, внутри сайта, сверка с чужими данными, методика

# Google и Яндекс: что отделяет верх выдачи

| Признак                                          | Google | Яндекс |
|--------------------------------------------------|--------|--------|
| Полнота интента                                  | 56,3   | 48,3   |
| Покрытие LSI                                     | 57,6   | 55,3   |
| Слов основного текста                            | 56,2   | 51,5   |
| Подзаголовков H2                                 | 56,7   | 53,4   |
| Полный ключ в title                              | 51,9   | 54,4   |
| Полный ключ в description                        | 51,2   | 54,2   |
| Размер HTML                                      | 59,5   | 62,9   |
| DOM-узлов                                        | 58,7   | 59,9   |
| К картинок                                       | 58,6   | 55,2   |
| Время ответа сервера (ниже 50 — наверху быстрее) | 47,1   | 42,4   |

Верх — позиции 1–10, низ — 31–49. Из 100 пар: у скольких верхняя страница выше по признаку. Google — 14.09.2026, Яндекс — 09.09.2026.

## ОБЩАЯ ЧАСТЬ

Охват темы отделяет верх в обоих. По всем парам внутри запроса у каждого поисковика 53,8 из 100.

**Разное:** в Яндексе ещё работает ключ в title и description, в Google — полнота ответа и объём текста.

Топ-10 двух систем по одному запросу делят в среднем 1,86 адреса.

# Детектор ИИ-текста и «вода»: топ от соседей не отличается

**10,5% и 10,9%**  
страниц выше порога 2,0 у топ-20 и у позиций 21–50

**49,7 из 100**  
в паре одного запроса сигнал выше у верхней страницы

**134 и 136**  
сайтов, где верхняя страница «машиннее» и «человечнее» нижней

| Метрика SEO-анализа текста              | Топ-20 | 21–50 |
|-----------------------------------------|--------|-------|
| Водность, %                             | 23,7   | 23,5  |
| Академическая тошнота, %                | 2,4    | 2,5   |
| Средняя длина предложения, слов         | 12,4   | 12,6  |
| Индекс Флеша                            | 4,4    | 3,7   |
| Без «нужно исправить» по шести метрикам | 81,8%  | 83,1% |

Детектор верх выдачи от низа не различает. Вода, тошнота и читабельность позиций тоже не объясняют: при равной длине текста различия сходятся к 48–52 из 100.

# Медианы всех страниц выдачи: Google, позиции 1–50

| Признак                                | Медиана  | Типичный разброс    | Есть на странице                                                                                                                                    | Google |
|----------------------------------------|----------|---------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------|--------|
| Слов основного текста                  | 766      | 363...1398          | Полный ключ в title                                                                                                                                 | 41,8%  |
| Покрытие LSI (доля списка)             | 0,525    | 0,369...0,676       | Полный ключ в H1                                                                                                                                    | 22,2%  |
| Слов из списка выдачи (список ≈ 80)    | ≈ 42     | ≈ 30...54           | Полный ключ в description                                                                                                                           | 34,2%  |
| Подзаголовков H2 / H3                  | 5 / 2    | 2...9 / 0...8       | Блок FAQ                                                                                                                                            | 23,5%  |
| Списков в основном тексте              | 9        | 0...25              | Цены в основном тексте                                                                                                                              | 40,3%  |
| Длина title / description, знаков      | 70 / 164 | 55...89 / 131...207 | Разметка Schema.org                                                                                                                                 | 59,3%  |
| Плотность ключа, %                     | 5,3      | 3,5...7,5           | Индексируемая                                                                                                                                       | 99,4%  |
| Раскрыто полностью, составляющих из 10 | 4        | 3...6               |                                                                                                                                                     |        |
| Размер HTML, КБ                        | 224      | 120...461           | Строки в сравнении с Яндексом: в Яндексе типичный HTML 293 КБ, слов 815, охват 0,556. Полная таблица — в статье, раздел «Типичная страница топ-50». |        |
| DOM-узлов                              | 1322     | 803...2258          |                                                                                                                                                     |        |
| К картинок                             | 32       | 15...70             |                                                                                                                                                     |        |
| Ссылок на странице                     | 158      | 88...288            |                                                                                                                                                     |        |
| Файлов JavaScript / CSS                | 12 / 5   | 6...21 / 2...12     |                                                                                                                                                     |        |
| Время ответа сервера, мс               | 603      | 442...897           |                                                                                                                                                     |        |

# Внутри одного сайта: что остаётся от разницы между сайтами

| Признак              | Между группами | Внутри сайта | Осталось | Верх больше / меньше | p     |
|----------------------|----------------|--------------|----------|----------------------|-------|
| Размер HTML          | +39,2 КБ       | +5,8 КБ      | 15%      | 285 / 219            | 0,004 |
| DOM-узлов            | +237           | +29          | 12%      | 288 / 215            | 0,001 |
| Слов на странице     | +263           | +47,5        | 18%      | 280 / 224            | 0,014 |
| К картинок           | +8             | 0            | —        | 245 / 189            | 0,008 |
| Время ответа сервера | +5 мс          | 0 мс         | —        | 252 / 252            | 1,0   |
| Охват темы (LSI)     | +0,050         | +0,026       | 52%      | 295 / 223            | 0,002 |

Техничка — по 505 сайтам со страницами в обеих группах, охват темы — по 529. Каждый сайт даёт один голос: разница его типичной верхней и нижней страницы.

# Сходится ли это с Google, Backlinko и Ашмановым

| Признак                | Чужие данные                                                                                      | Этот замер                                                                               |
|------------------------|---------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------|
| Полнота раскрытия темы | Backlinko: полный контент обходит неполный. Ашманов: учитываются «другие слова по теме»           | LSI 53,8, держится во всех проверках. Полнота интента 52,3, при равной длине не держится |
| Ключ в title и H1      | Backlinko: связи почти нет. Ашманов: для title в Google корреляция «если и была, то очень слабая» | 50,6 и 50,0 из 100                                                                       |
| Длина текста           | Google: сама по себе не важна. Backlinko: связи нет. Ашманов: «сильный параметр»                  | 53,4; при равном охвате темы может быть случайной                                        |
| Размер HTML            | Backlinko: связи нет                                                                              | 54,5 между сайтами; внутри сайта остаётся 15%                                            |
| Скорость               | Backlinko: связи со скоростью домена нет                                                          | Время ответа сервера 49,4                                                                |

Расходимся дважды: объём текста у Ашманова и размер HTML у Backlinko. Backlinko (11,8 млн результатов Google) сверено с первоисточником; Ашманов – по цитатам из статьи.

# Методика и оговорки

## КАК СЧИТАЛОСЬ

Выдача: XMLRiver, Google, Москва, google.ru, десктоп

Аудит страниц: Site Audit Pro

Основной текст: trafilatura, леммы: pymorphy3

Интент: карты из Вордстата, подсказок и блоков Google, зафиксированы хешем до оценки

Оценщик: языковая модель вслепую; два прогона совпали в 88 случаях из 100 (1877 решений)

Шесть гипотез записаны до данных; разведка помечена в тексте

## ОГРАНИЧЕНИЯ

Один регион, одна дата, десктоп

Корреляция без причинности

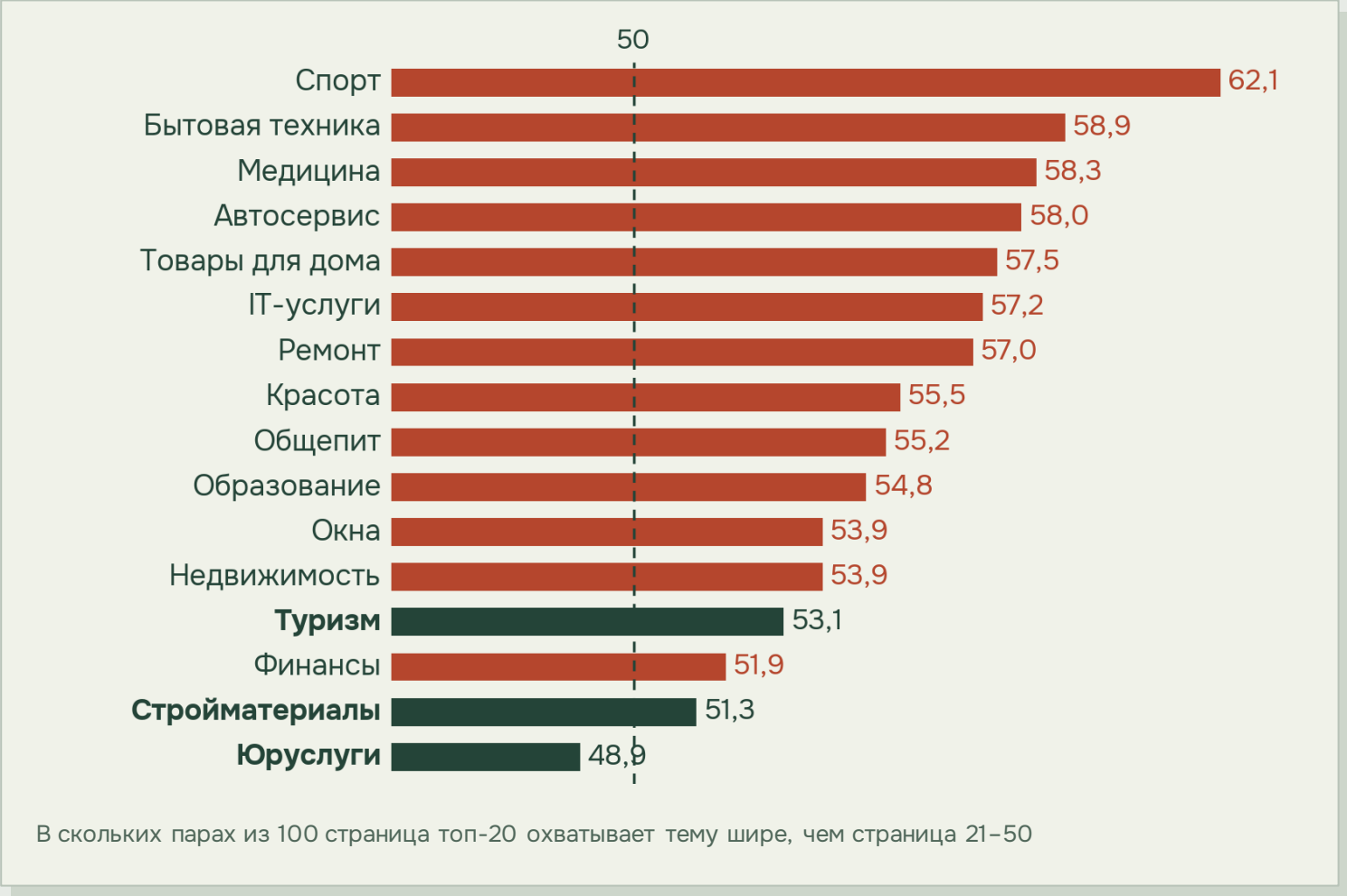
Оценщик — модель, человеческой разметкой не калибровался

Текст длиннее 2500 слов оценщик видел обрезанным

На запрос проверено восемь документов Google, не вся выдача

Снимок Яндекса сделан на пять дней раньше

# Где охват темы сильнее разделяет группы: спорт против юруслуг



**ОХВАТ ТЕМЫ РАЗДЕЛЯЕТ СИЛЬНЕЕ**

Спорт, бытовая техника, медицина, автосервис: наверху статьи и страницы услуг. Недостающие темы закрывать есть смысл.

**ТЕКСТОМ НЕ ОБОГНАТЬ**

**Стройматериалы и туризм:** каталоги и агрегаторы занимают 65% и 49% проверенного топа (в медицине 5%). Считайте ассортимент, цену, фильтры, силу сайта.

Юруслуги: охват верх не отличает совсем.

# Разведка: триплеты «субъект — связь — объект» топ-20 от позиций 21–50 не отличаются

| Показатель                                         | Топ-20 | 21–50 | Из 100 пар |
|----------------------------------------------------|--------|-------|------------|
| Триплетов на 1000 слов                             | 18,7   | 19,1  | 49,6       |
| Триплетов с прямым дополнением на 1000 слов        | 7,0    | 7,2   | 49,5       |
| Разных сущностей (субъекты и объекты) на 1000 слов | 25,9   | 26,3  | 49,5       |
| Доля триплетов «мы ...» (сайт о себе)              | 3%     | 4%    | 49,7       |
| Разных связей (глаголов) на страницу               | 13     | 11    | 52,6       |

Медианы по страницам. «Из 100 пар» — та же мерка, что в статье: 50 — монетка. Интервалы у первых пяти строк включают 50. Строка про глаголы выше у топа, но это длина текста, а не густота фактов.

Метод — быстрая эвристика, а не языковая модель: строго верны 56% триплетов, у прямых дополнений 71% (90% с допуском). Гипотеза придумана после данных.

**114 677**  
триплетов на 5964  
страницах, 3773 сайта

**Из 51 частой связи**  
(предлагать, зависеть, иметь, работать...) ни одна не отличает топ после поправки на множественные проверки.