

empenoso
6 янв в 03:23

Как я научился оценивать популярность статей через парсинг показателей сайтов

👉 Простой | 🕒 7 мин | 👁 1.8K


Контент и копирайтинг*, Управление медиа*, Node.JS*, Open source*

Кейс

Я уже давно пишу статьи про различные аспекты IT-технологий, инвестиции, автоматизацию и умные дома на разных площадках: Хабр, Т—Ж, СмартЛаб, Пикабу, VC.ru и других.

За всё время накопилось примерно 250 статей, [которые по итогу свёл в таблицу](#). Но вот задумываться о популярности статей и их реальном эффекте стал относительно недавно.

РЕКЛАМА



Хостинг aéza
Твой проект этого достоин

LynxReport: учёт публикаций [Node.js] ⌵ ⌵ ⌵ ⌵

Файл Правка Вид Вставка Формат Данные Инструменты Расширения Справка

100% р. % 0.00 123 По ум... - 10 + B I A

1	Название	Тип (задается вручную)	Дата выхода (задается вручную)	Выходные данные (задается вручную)	Автор(ы) (задается вручную)	Тема (задается вручную)	Тираж для печатных изданий	Просмотры	Ко
4	Таблицы и скрипты Гугл для бизнеса и анализа	Веб	2024-12-20	https://time.google.com/maps/@37.61737,55.75581,15z	М.В. Шардин	скрипты	-	508	
5	Топ-тренды Смартлаба 2024: что читали больше всего? Смартлаб	Веб	2024-12-19	https://smart-lab.ru/mobile/topic/1096274/	М.В. Шардин	Джава и гугл скрипты	-	4 687	
6	Что читали на Хабре в 2024 году: анализ статей с Node.js, Google Sheets и каплей ChatGPT Хабр	Веб	2024-12-18	https://habr.com/ru/articles/867068/	М.В. Шардин	Джава и гугл скрипты	-	7 109	
7	С бумаги на цифровую карту: генерация файла из таблицы для импорта на карту и геокодирование адресов с помощью Python Хабр	Веб	2024-12-12	https://habr.com/ru/articles/865558/	М.В. Шардин	Картография	-	1 974	
8	С бумаги на цифровую карту: генерация файла из таблицы для импорта на карту и геокодирование адресов с помощью Python Пикабу	Веб	2024-12-12	https://pikabu.ru/story/s_bumagi_na_tsifrovuyu_kartu_generatsiya_fayla_iz_tablitsy_dlya_importa_na_kartu_i_geokodirovanie_adresov_s_pomoshchyu_python_12126449	М.В. Шардин	Картография	-	4 271	
9	Слабуомие и отвага: как найти ликвидные облигации с доходностью до 40% и ежемесячными выплатами Смартлаб	Веб	2024-12-05	https://smart-lab.ru/mobile/topic/1090879/	М.В. Шардин	Инвестиции	-	5 823	
10	Слабуомие и отвага: как найти ликвидные облигации с доходностью до 40% и ежемесячными фиксированными выплатами Хабр	Веб	2024-12-05	https://habr.com/ru/articles/863762/	М.В. Шардин	Инвестиции	-	7 609	
11	Слабуомие и отвага: как найти ликвидные облигации с доходностью до 40% годовых и ежемесячными фиксированными выплатами на Московской бирже Пикабу	Веб	2024-12-05	https://pikabu.ru/story/slaboumie_i_otvaga_kak_nayti_likvidnyie_obligatsii_s_dokhodnostyu_do_40_godovikh_i_ezhemesyachnyimi_fiksirovannymi_vyplatami_na_moskovskoy_birzhe_12093410	М.В. Шардин	Инвестиции	-	3 444	
12	Слабуомие и отвага: как найти ликвидные облигации с доходностью до 40% годовых и ежемесячными	Веб	2024-12-05	https://rationalanswer.club/post/1087/	М.В. Шардин	Инвестиции	-	нд	

+ ≡ Вводное ▾ Данные ▾ Типы ▾ Темы ▾ Популярные ▾ Ресурсы ▾ Нарастающий ▾ Годы ▾ shardin.name ▾ Выборка ▾

Почему я решил собирать статистику публикаций?

Главной целью всех этих публикаций было поделиться своим опытом и при этом попытаться понять насколько вообще это важно и актуально для читателей. Однако обратная связь была не всегда очевидной: где-то комментариев вообще не было, а иногда на Хабре статьи набирали большой рейтинг при малом количестве комментариев (но это редко).

Однако очевидно одно - статистика заставляет посмотреть на материалы со стороны. Например статья про [то, как я при помощи двух скриптов смог автоматически сгенерировать опись документов для 700 страниц на непрофильном Пикабу](#) собрала три месяца назад почти 75 тысяч просмотров и 80 комментариев, а на Хабре эта же тема была не особо популярна.

Или статья про [то, что читали на Хабре в 2024 году: анализ статей с Node.js, Google Sheets и каплей ChatGPT](#) - собрала на Хабре три недели назад рейтинг +68, но «всего» 7 тысяч просмотров.

Зачем я написал Open Source скрипт?



Хостинг aéza

Твой проект этого достоин

```

Файл  Правка  Выделение  Вид  ...  ← →  2020_07_LynxReport учёт публикаций  88
JS update_parsing_sites.js X
JS update_parsing_sites.js > ...
22  async function journalTinkoff(url) {
51  }
52
53  // =====
54  // Парсинг Хабр - Habr
55
56  async function habrCom(url) {
57    try {
58      // 1. Загрузка HTML и извлечение JSON с данными статьи
59      // Загружаем HTML страницы
60      const { data: html } = await axios.get(url);
61
62      // Загружаем HTML в Cheerio для парсинга
63      const $ = cheerio.load(html);
64
65      // Ищем скрипт, содержащий `window.__INITIAL_STATE__`
66      const scriptContent = $('script')
67        .toArray()
68        .map(script => $(script).html())
69        .find(content => content && content.includes('window.__INITIAL_STATE__'));
70
71      if (!scriptContent) {
72        throw new Error('Не удалось найти __INITIAL_STATE__ в HTML');
73      }
74
75      // Извлекаем JSON из содержимого скрипта
76      const jsonMatch = scriptContent.match(/window\.__INITIAL_STATE__\s*=\s*(\{.*\});/s);
77      if (!jsonMatch || jsonMatch.length < 2) {
78        throw new Error('Не удалось извлечь JSON из __INITIAL_STATE__');
79      }
80
81      const initialState = JSON.parse(jsonMatch[1]);
82      const articlesList = initialState.articlesList.articlesList;
83
84
85      // 2. Извлечение нужных данных из JSON
86      // Извлекаем ID публикации из URL

```

Извлечение просмотров, комментариев, закладок и рейтинга из каждой статьи вручную занимало бы много времени, поэтому я решил пойти путём автоматизации. Написал скрипт, который скачивает эти данные по статьям и помещает сразу в одну удобную таблицу, где я вижу, какие темы стоит развивать дальше и на каких ресурсах.

Это [Open Source скрипт](#), размещенный на [Гитхабе](#), который состоит из Google Apps Script и Node.js частей и обе эти части работают с итоговой сводной Google Таблицей.

Если вы автор или агентство и хотите видеть полную картину популярности ваших материалов, выявлять, какие темы интересуют аудиторию, а какие требуют доработки, автоматизированный инструмент сбора данных - это то, что вам нужно. Я уже испытал это на себе и расскажу подробнее об этом в статье.



Хостинг aéza

Твой проект этого достоин

Изначально я использовал Google Apps Script внутри Гугл Таблиц для автоматизации извлечения просмотров, комментариев, закладок, рейтинга. Однако по мере того, как я расширялся на новые платформы, то обнаружил, что некоторые сайты ограничивают доступ к определенным данным, что потребовало от меня перехода на серверный язык JavaScript - Node.js.

Конкретный пример: сайт инвесторов и трейдеров Смартлаб имеет скрытое API для получения просмотров статей.

Если обращение идёт через Google Apps Script внутри Гугл Таблиц, то сервер просто не отдает просмотры, возвращая пустой ответ, потому что заголовки headers и User-Agent не поддерживаются Google Apps Script:

```
function fetchApiResponse() {
  const url = 'https://smart-lab.ru/cgi-bin/gcn.fcgi?list=1083556&func=func8422&_=17319';

  try {
    // Выполняем запрос к API с детализированными заголовками
    const response = UrlFetchApp.fetch(url, {
      method: 'get', // HTTP метод GET
      muteHttpExceptions: true, // Не выбрасывать ошибки для HTTP ответов с кодами, отл
      headers: {
        'Accept': 'text/html,application/xhtml+xml,application/xml;q=0.9,image/avif,image/webp,*/*;q=0.8',
        'Accept-Encoding': 'gzip, deflate, br, zstd',
        'Accept-Language': 'ru-RU,ru;q=0.9,en-US;q=0.8,en;q=0.7',
        'Cache-Control': 'max-age=0',
        'Sec-CH-UA': '"Chromium";v="130", "Google Chrome";v="130", "Not?A_Brand";v="99"',
        'Sec-CH-UA-Mobile': '?0',
        'Sec-CH-UA-Platform': '"Windows"',
        'Sec-Fetch-Dest': 'document',
        'Sec-Fetch-Mode': 'navigate',
        'Sec-Fetch-Site': 'none',
        'Sec-Fetch-User': '?1',
        'Upgrade-Insecure-Requests': '1'
      }
    });

    // Проверяем HTTP код ответа
  }
});
```



Хостинг aéza

Твой проект этого достоин

```

    return;
  }

  // Получаем содержимое в виде обычного текста
  const content = response.getContentText();

  // Логируем необработанный ответ для отладки
  console.log('Необработанный ответ API:', content);

} catch (error) {
  // Обрабатываем ошибки во время выполнения запроса
  console.error('Ошибка при выполнении запроса к API:', error.message);
}
}

```

The screenshot shows a Google Apps Script editor with a function named `fetchApiResponse()`. The function uses `UrlFetchApp.fetch()` to make a GET request to `https://smart-lab.ru/cgi-bin/gcn.fcgi?list=1083556&func=func8422&_=1731928577852`. The headers include `Accept`, `Accept-Encoding`, and `Accept-Language`. Below the code, the 'Журнал выполнения' (Execution Log) shows three entries: '07:42:17 Примечание Выполнение начато', '07:42:18 Информация Необработанный ответ API:', and '07:42:18 Примечание Выполнение завершено'.

```

1 function fetchApiResponse() {
2   const url = 'https://smart-lab.ru/cgi-bin/gcn.fcgi?list=1083556&func=func8422&_=1731928577852';
3
4   try {
5     // Выполняем запрос к API с детализированными заголовками
6     const response = UrlFetchApp.fetch(url, {
7       method: 'get', // HTTP метод GET
8       muteHttpExceptions: true, // Не выбрасывать ошибки для HTTP ответов с кодами, отличными от 200
9       headers: {
10        'Accept': 'text/html,application/xhtml+xml,application/xml;q=0.9,image/avif,image/webp,image/png,*/*;q=0.8,application/signed-exchange;v=b3;q=0.7',
11        'Accept-Encoding': 'gzip, deflate, br, zstd',
12        'Accept-Language': 'ru-RU;q=0.9,en-US;q=0.8,en;q=0.7'
13      }
14    });
15
16    // ... (code from the previous image) ...
17  } catch (error) {
18    // ... (code from the previous image) ...
19  }
20 }

```

Журнал выполнения

07:42:17	Примечание	Выполнение начато
07:42:18	Информация	Необработанный ответ API:
07:42:18	Примечание	Выполнение завершено

Google Apps Script

Если же запрос идёт через Node.js и использование User-Agent, то в ответ возвращается число просмотров даже без использования эмуляции браузера:

```

const axios = require('axios');

async function fetchApiResponse() {
  const url = 'https://smart-lab.ru/cgi-bin/gcn.fcgi?list=1083556&func=func8422&_=173

```



Хостинг aéza

Твой проект этого достоин


```

const options = {
  headers: {
    'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/53
  }
};

// Выполнение запроса к API для получения просмотров
const apiResponse = await axios.get(url, options);
const responseText = apiResponse.data;

// Логируем необработанный ответ для отладки
console.log('Необработанный ответ API:', responseText);

} catch (error) {
  // Обработка ошибок
  console.error(`Ошибка получения количества просмотров: ${error.message}`);
  return null;
}

fetchApiResponse();

```

```

JS update_parsing_sites.js JS test_smarlab.js X
JS test_smarlab.js > fetchApiResponse
1  const axios = require('axios');
2
3  async function fetchApiResponse() {
4    const url = 'https://smart-lab.ru/cgi-bin/gcn.fcgi?list=1083556&func=func8422&_1731928577852';
5
6
7    try {
8      // Установка заголовка User-Agent для имитации браузера Chrome
9      const options = {
10       headers: {
11         'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/91.0.4472.124 Safari/537.3
12       }
13     };
14
15     // Выполнение запроса к API для получения просмотров
16     const apiResponse = await axios.get(url, options);
17     const responseText = apiResponse.data;
18
19     // Логируем необработанный ответ для отладки
20     console.log('Необработанный ответ API:', responseText);
21
22   } catch (error) {

```

ПРОБЛЕМЫ Выходные данные КОНСОЛЬ ОТЛАДКИ ТЕРМИНАЛ ПОРТЫ Code

```

[Running] node "d:\SynologyDrive\docs\2020_07_LynxReport учёт публикаций\test_smarlab.js"
Необработанный ответ API: func8422(1083556,3892);

[Done] exited with code=0 in 0.78 seconds

```

Node.js

**Хостинг aéza**

Твой проект этого достоин

Обратная связь от аудитории может сильно варьироваться в зависимости от платформы, формата подачи материала и времени публикации.

Например, статья, выложенная одновременно на разных сайтах, может получить совершенно разные результаты. Уже упоминаемая выше статья "Как я при помощи двух скриптов смог автоматически сгенерировать опись документов для 700 страниц" собрала:

vc.ru:

225 / 3 / 2 / 2

Просмотры / Комментарии / Закладки / Рейтинг

Пикабу:

74 615 / 76 / 0 / 903

Просмотры / Комментарии / Закладки / Рейтинг

Хабр:

6 620 / 7 / 31 / +12

Просмотры / Комментарии / Закладки / Рейтинг

На Пикабу статья вызвала бурное обсуждение, в то время как на vc.ru аудитория осталась почти равнодушной.

Зачем собирать статистику публикаций?

Анализ метрик (просмотры, комментарии, закладки, рейтинг) помогает понять, что действительно интересует вашу аудиторию и какие темы популярны на различных ресурсах.

- **Просмотры:** показывают, сколько людей заинтересовалось темой.
- **Комментарии:** демонстрируют вовлечённость, обсуждения.
- **Закладки:** отражают ценность материала для долгосрочного использования.
- **Рейтинг:** индикатор общего одобрения.

Анализ этих данных помогают адаптировать стиль, темы и платформы для повышения эффективности публикаций.

**Хостинг aéza**

Твой проект этого достоин

За несколько лет написал [Google Apps Script](#), а затем и [Node.js](#), которые предназначены для автоматического сбора показателей публикаций с различных платформ.

Важно, что скрипты приспособлены к конкретной структуре каждой платформы, используя API (Тинькофф Журнал, Пикабу, Смартлаб) или парсинг HTML-кода (Хабр, vc.ru).

The screenshot shows a GitHub repository page for 'empenoso / LynxReport'. The repository is public and has 1 fork and 3 stars. The main content is a file list:

File Name	Commit Message	Time Ago
Google Apps Script	new	last month
.gitignore	Create .gitignore	5 years ago
HabrAllArticlesToCsv.csv	https://habr.com/ru/articles/8670...	3 weeks ago
HabrAllArticlesToCsv.js	https://habr.com/ru/articles/8670...	3 weeks ago
HabrAllArticlesToCsv_log_202...	https://habr.com/ru/articles/8670...	3 weeks ago
HabrAllArticlesToCsv_start_wi...	https://habr.com/ru/articles/8670...	3 weeks ago
LICENSE	Initial commit	5 years ago
README.md	Update README.md	6 minutes ago
first start.bat	fix	5 years ago

On the right side, there is an 'About' section with the following information:

- LynxReport: учёт публикаций
- Readme
- Apache-2.0 license
- Activity
- 3 stars
- 1 watching
- 1 fork
- Report repository

Below the 'About' section, there are sections for 'Releases' (No releases published) and 'Packages' (No packages published).

Различие Google Apps Script и Node.js версий

Если очень кратко, то версия на Google Apps Script проста в использовании, запускается прямо в браузере, идеальна для быстрой интеграции с таблицами Google, а Node.js версия мощнее, работает на сервере, подходит для более сложных задач и интеграций.

Google Apps Script – это облачное решение, подходящее для задач, не требующих больших вычислительных ресурсов. Он интегрирован в Google Workspace и позволяет легко работать с веб-страницами, API Google, используя встроенные функции. Однако его функционал ограничен, и он не может работать с ресурсоёмкими задачами или сложным парсингом.



Хостинг aéza

Твой проект этого достоин

Сложного парсинга, работы с API, обработки больших объемов данных и реализации

сложных логик. Node.js позволяет использовать различные библиотеки (например, Axios, Cheerio, Puppeteer), расширяя возможности скрипта. Однако Node.js требует установки и настройки на сервере или локально.

Публикация скрипта на GitHub

Публикация скриптов на GitHub в рамках философии Open Source имеет несколько преимуществ. Во-первых, это позволяет другим пользователям использовать и адаптировать скрипт под свои нужды. Во-вторых, открытый исходный код способствует прозрачности и аудиту кода, что помогает обнаруживать и исправлять ошибки. В-третьих, сообщество разработчиков может помочь в улучшении и расширении функциональности скрипта, если захочет :)

Пример анализа собранной статистики для выбора будущих тем

Мне кажется немного некорректно проводить анализ собственных статей здесь и поэтому я оттолкнусь от анализа двух популярных платформ: [Хабра](#) и [Смартлаба](#) по всем их публикациям за 2024 год.



Хостинг aéza

Твой проект этого достоин

Хабр **КАК СТАТЬ АВТОРОМ** Как компании вычисляют накрученный опыт **Войти**

Моя лента Все потоки Разработка Администрирование Дизайн Менеджмент Маркетинг Научпоп

1000+
вакансий с удалёнкой

Хабр Карьера

empenoso 18 дек 2024 в 03:23

Что читали на Хабре в 2024 году: анализ статей с Node.js, Google Sheets и каплей ChatGPT

Простой 13 мин 7.2К

Open source*, Node.JS*, Google API*, Машинное обучение*

Кейс

Сегодня расскажу вам одну занимательную историю о том, как я решил сделать что-то полезное и интересное, а заодно чуть-чуть облегчить жизнь всем нам, кто регулярно зависает на Хабре. Ведь сколько здесь данных? Очень много! Статьи, комментарии, рейтинги... Всё это крутится вокруг нас, но как это можно использовать для анализа и выводов?

Всё началось с простого вопроса: «Как узнать, какие статьи на Хабре самые популярные?» Казалось бы - просто посмотреть на рейтинг или количество просмотров? Но ведь тут есть куча нюансов: например, у некоторых статей большой рейтинг, но они почти не комментируются. Или

+68 62 22 +22

Откуда вообще я взял исходные данные для анализа всей платформы?

Исходные данные были получены мной путем парсинга двух ресурсов. Использовался автоматизированный сбор данных с использованием Node.js для парсинга HTML-страниц и последующего извлечения информации о каждой статье. Данные о просмотрах, комментариях, рейтинге и закладках позволяли оценить популярность и вовлеченность аудитории.

Фильтр по минимальному количеству просмотров (5 000 на Смартлабе, 30 000 на Хабре) и рейтингу (более 30 на Хабре) позволил отфильтровать статьи, которые не соответствовали критериям популярности.



Хостинг aéza

Твой проект этого достоин

SMART-LAB
Мы делаем деньги на бирже

Михаил Шардин
личный блог

19 декабря 2024, 10:22

Топ-тренды Смартлаба 2024: что читали больше всего?

Трейдеры и инвесторы живут в мире данных. Цены, отчёты, комментарии, прогнозы — всё это нас окружает, но как извлечь из этого пользу? Сегодня расскажу, как решил сделать что-то полезное для сообщества трейдеров и инвесторов Смартлаб.

№	Дата	Время	Автор	Имя пользователя	Тема	Просмотры	Комментарии	Лайки	Рейтинг	
6115	1075596	26 октября 2024, 14:32	Инвестор Сергей	Investor_Sergei	users	О чем говорит на конференции Смартлаба?	36	3	37	6134
6116	1075600	26 октября 2024, 14:57	Kurdit	Kurdit	users	Эффективные управленцы убьют рынок. Лу...	75	2	38	6466
6117	1075604	26 октября 2024, 16:12	Олег Коневи	konev1	users	Что я делаю не так ?	147	5	18	7693
6118	1075607	26 октября 2024, 16:26	INVEST_KITCHEN	VitaliyBykov_f72	users	В чем припарковать свободную часть денег?	9	2	3	8850
6119	1075610	26 октября 2024, 16:38	Kiplinger	Kiplinger	users	Дилемма: продать акции или подождать?	18	2	9	5906
6120	1075624	26 октября 2024, 18:10	Екатерина Идагулулина	Investrivet	users	Почему надо продавать флэттеры. Конфере...	50	9	47	12251
6121	1075645	26 октября 2024, 20:02	Fond&Flow	Mavrodytrade	users	Заседание ЦБ: Почему 21% — это не пре...	31	5	25	6252
6122	1075647	26 октября 2024, 20:35	KAPITAL	KAPITALS	users	Исторический минимум индекса RGIВ	34	1	18	7672
6123	1075648	26 октября 2024, 20:37	tomas_kub	tomtomtom	users	Снова голодаю, неделя -0,43%, когда там ра...	34	2	15	5314
6124	1075649	26 октября 2024, 20:58	Коплю на мечту	leva	users	Корейцы в Нурсе	162	4	27	6188
6125	1075650	26 октября 2024, 21:03	T-Инцидентия	t_invest	company	«Денги не спят»: разбираемся с ключевой...	9	0	6	6212
6126	1075652	26 октября 2024, 21:44	Оля "Наре"...	OlyaPavlyatenko	users	Стоимость нефтегазовой инфраструктуры...	17	0	12	5843
6127	1075671	27 октября 2024, 05:58	Михаил Шардин	empenoso	users	Я был спикером на 358 конференции Смарт...	140	6	116	11320
6128	1075675	27 октября 2024, 07:54	igotsochi	igotsochi	users	Ключевая ставка 21%, доллар не по 100, инд...	8	7	26	7387
6129	1075676	27 октября 2024, 08:04	Евгений Павлик	EvgeniyPavlik	users	Очередные покупки дивидендных акций. Ки...	27	4	24	7823
6130	1075687	27 октября 2024, 09:57	Леонид Чарнов	JBagle	users	Впечатления от диалогов завтрака в СБере "	46	8	73	6012
6131	1075694	27 октября 2024, 10:54	Aleximex	news	users	"Часовые авиалинии" прекращают существов...	8	0	5	8915
6132	1075698	27 октября 2024, 11:41	Oleg Koe	ChicagosBull	users	Страны с самым большим ВВП на душу насе...	153	2	30	11945
6133	1075708	27 октября 2024, 12:37	Krechetov	Krechetov	users	Российский рынок пробил дно. И дальше бу...	0	1	7	7328
6134	1075710	27 октября 2024, 12:48	Rationalist	Rationalist	users	Стоит ли инвестировать в Россию?	138	7	30	7439
6135	1075713	27 октября 2024, 12:58	CyberWish	MikhailVlasenko	users	"Вынимка" из выступления эмитентов - Рос...	11	1	21	5318
6136	1075720	27 октября 2024, 14:09	Инвестор Сергей	Investor_Sergei	users	ФОТО и ТЕЗИСЫ Смартлаба-2024: как прош...	51	3	49	8312

Определение трендов и предпочтений аудитории для разных платформ: Хабр и Смартлаб.

Анализ показал, что у каждой платформы свои тренды.

Хабр более ориентирован на технические темы, такие как настройка программного обеспечения, обходы блокировок, разработка и советы по программированию. Наиболее популярные темы за 2024 год на Хабре - проблемы блокировки YouTube, решение проблем с GoodbyeDPI, уязвимости безопасности.

На Смартлабе, наоборот, доминируют темы, связанные с инвестициями, экономикой и геополитикой. Трендовые темы Смартлаба - санкции, рубль, ключевая ставка, дедолларизация, анализ рынка акций и облигаций.



Хостинг aéza

Твой проект этого достоин

акценты.

Для эффективной публикации на каждой платформе важно учитывать интересы ее аудитории. На Хабре следует фокусироваться на конкретных технических проблемах и их решениях, практических советах и "человеческом" стиле. Статьи с интересными заголовками, посвященные узким, конкретным темам, вероятнее всего, получат большую вовлеченность.

На Смартлабе, ключевым является анализ геополитических и экономических трендов, связанных с финансами. Оптимально создавать качественный аналитический контент, предлагающий рекомендации по инвестированию, с подкреплением фактами и экспертизой.

Выводы

Мой Open Source скрипт для автоматического сбора статистики (просмотры, комментарии, закладки, рейтинг) позволяет анализировать популярность статей на разных платформах, выявлять тренды и понять что разные платформы могут требовать разного контента.

Работа с Open Source инструментом делает этот процесс масштабируемым и полезным для других авторов и агентств.

Автор: [Михаил Шардин](#),

6 января 2025 г.

Теги: парсинг, парсинг сайтов, СмартЛаб, Хабр, Пикабу, Т-Ж, smart-lab

Хабы: Контент и копирайтинг, Управление медиа, Node.JS, Open source

Редакторский дайджест

Присылаем лучшие статьи раз в месяц



Хостинг aéza

Твой проект этого достоин

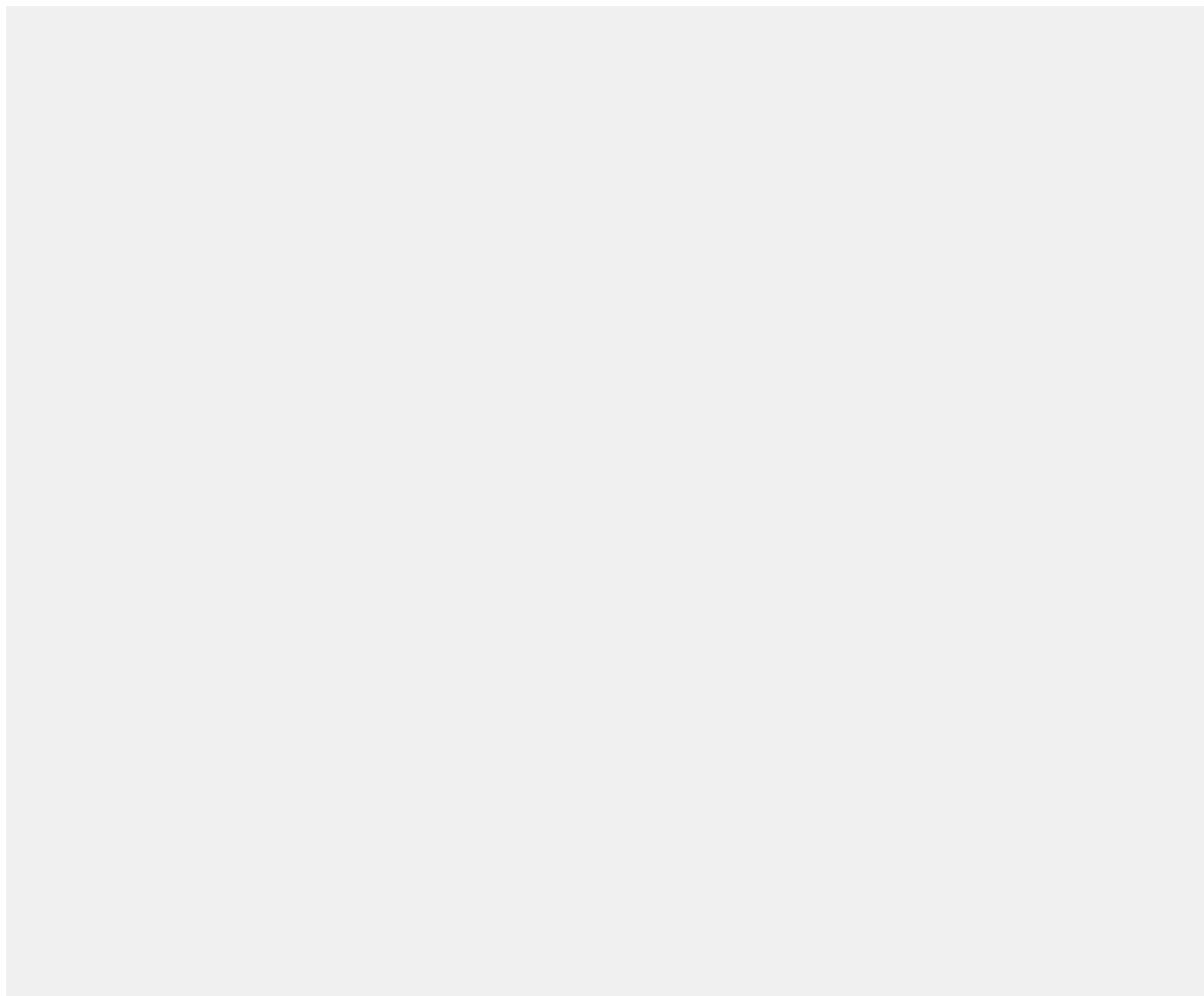
Михаил Шардин @empenoso

Автоматизация / Данные / Финансы / Умные дома

Подписаться



[Сайт](#) [Сайт](#) [Github](#)



 Комментарии 11

Публикации



Хостинг aéza

Твой проект этого достоин

**labyrinth**

20 часов назад

Зачем Яндекс.Браузеру эти данные?

**Простой**

2 мин



16K

Кейс

◆ +146

📖 50

💬 84

**nmgtech**

19 часов назад

Я перешёл на Firefox и обратно возвращаться не намерен

**Простой**

5 мин



34K

Мнение

Перевод

◆ +86

📖 66

💬 285

**BabayMazay**

19 часов назад

Простая откачка и наполнение самодельных газоразрядных ламп

**Средний**

8 мин



1.8K

Тutorial

◆ +46

📖 16

💬 5

**vital_pavlenko**

14 часов назад

Как AI захватывает Хабр, и почему это всех бесит

**Простой**

3 мин



3.5K

Мнение

◆ +43

📖 11

💬 75

**Хостинг éza**

Твой проект этого достоин

Liberux NEXX: Linux-смартфон с 32 ГБ ОЗУ и аппаратными переключателями

🕒 4 мин 👁 11K

📌 +42

🔖 16

💬 44



laureldo

20 часов назад

Полезные паровозики. Часть 1: введение в железнодорожный моделизм

🟢 Простой 🕒 10 мин 👁 2.4K

Обзор

📌 +41

🔖 23

💬 29



CyberPaul

16 часов назад

Персоналки от Erpson — удивительные компьютеры из прошлого

🟢 Простой 🕒 6 мин 👁 1.8K

Ретроспектива

📌 +33

🔖 9

💬 0



wytear

18 часов назад

Маскирование данных от А до Я

🟢 Простой 🕒 14 мин 👁 1.5K

Обзор

📌 +24

🔖 33

💬 12



PatientZero

21 час назад



Хостинг aéza

Твой проект этого достоин

Простой 8 мин 1.1K

+21 23 5

rakovskij_stanislav
13 часов назад

Вредоносные пакеты deepseeek и deepseekai были опубликованы в Python Package Index

Простой 3 мин 1.8K

+20 6 0

Изучаем в опросе: как о здоровье заботятся хабравчане

[Показать еще](#)

ИСТОРИИ

Хабр Карьера • Где работать в IT

Поговорили с компанией СИГМА, которая разрабатывает IT-решения для цифровизации энергетики и ЖКХ

Где работать в IT в 2025: СИГМА

Хабр Карьера

День святого Валентина

HR — это по любви!

Расскажите, за что вы любите своего HR, и купидоны K-Team доставят ваше послание в любую точку мира!

Отправьте эйчару валентинку

Что для вас важно в вашей работе?

Что для вас важно в работе?

Вышла Janus-Pro-7B

Это улучшенная версия мультимодальной нейросети Janus, выпущенной DeepSeek. Janus-Pro-7B служит для анализа и генерации изображений. По ряду бенчмарков она опережает открытую Stable Diffusion 3 Medium и коммерческую DALL-E 3.

Вышла Janus-Pro-7B

Как вырастить айтишника

Подборка статей для тех, кто хочет научить детей программированию и робототехнике

Как вырастить айтишника

Что

Неде

из бл

Где ком

Хостинг aéza
Твой проект этого достоин

от 2 000 до 4 500 \$ · DataLouna · Можно удаленно

Middle fullstack developer (NodeJS)

до 180 000 ₽ · Тетрика · Москва · Можно удаленно

Саппорт в команду

от 500 \$ · MST · Можно удаленно

Менеджер по продажам сервиса Хабр Карьера

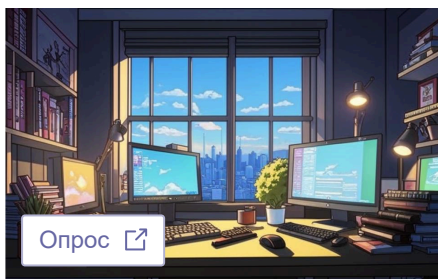
от 60 000 до 120 000 ₽ · Хабр · Москва · Можно удаленно

Инженер по тестированию ПО (ОФИС)

от 80 000 до 110 000 ₽ · SpectrumData · Екатеринбург

[Больше вакансий на Хабр Карьере](#)

МИНУТОЧКУ ВНИМАНИЯ



Печеньки, удаленка, обучение — а что важно в работе для вас?



Послепраздничная история: задачки на внимание и дедукцию



Гиперконвергентная среда: OpenStack или VMware?

РАБОТА

Контент менеджер

16 вакансий

JavaScript разработчик

127 вакансий

Node.js разработчик

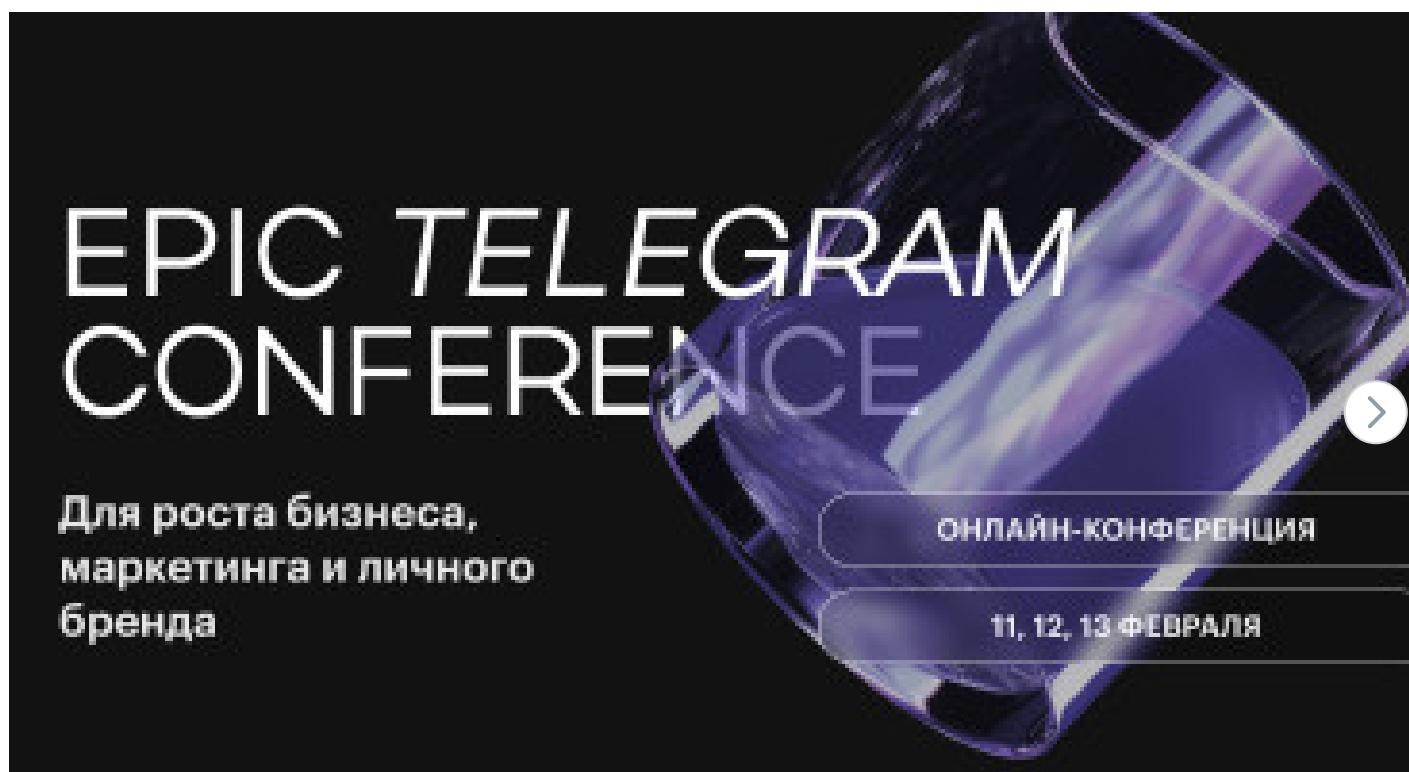
38 вакансий



Хостинг aéza

Твой проект этого достоин

БЛИЖАЙШИЕ СОБЫТИЯ



11 – 13 февраля

Epic Telegram Conference

Онлайн

Маркетинг

Больше событий в календаре

Хабр



Хостинг aéza

Твой проект этого достоин

Техническая поддержка

© 2006–2025, Habr



Хостинг aéza

Твой проект этого достоин