Собрали интересное про Data Science в одном месте

Все потоки Разработка Администрирование Дизайн Менеджмент Маркетинг Научпоп



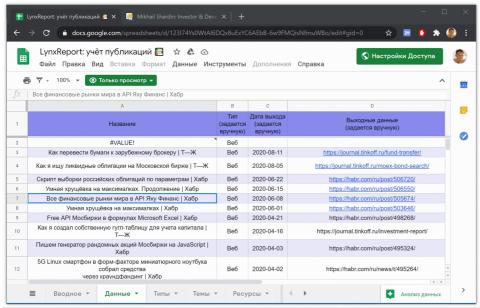


empenoso 17 августа 2020 в 05:15

Как стать экспертом для поисковых систем

JavaScript, Node.JS, Google API, Контент-маркетинг

Не так давно я познакомился с описанием E-A-T алгоритма от Google, который расшифровывается как «Expertise, Authoritativeness, Trustworthiness» (экспертность, авторитетность, достоверность). И мне, как автору, который пишет для разных сайтов стало интересно — насколько я сам соответствую критериям этого алгоритма и могу ли повлиять на текущую ситуацию. Тем более, что некоторые заготовки в виде открытой гугл таблицы для учета и мониторинга собственных публикаций LynxReport уже были.



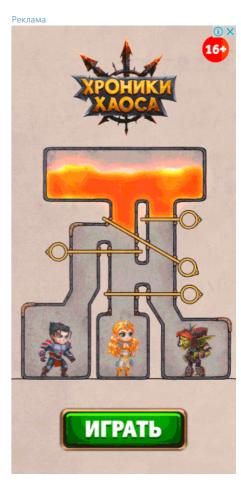
Google Таблицы → Node.js → Google Charts → Сайт-визитка → Ton-3 место в поиске ФИО + специализация

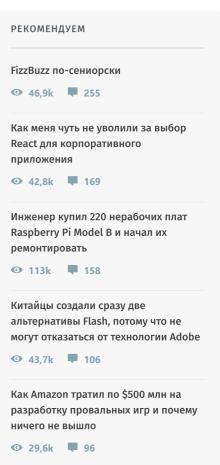
На основании данных таблицы я решил дополнить сайт-визитку, сведениями о публикациях, которые бы генерировались автоматически. Что я хотел получить:

- 1. Актуальную сводку публикаций, расположенную на временной шкале Google Charts.
- 2. Автоматическую генерацию выходных данных и ссылок на статьи из гугл таблицы в html версию визитки.
- 3. PDF версии статей со всех сайтов, из-за опасений закрытия некоторых старых сайтов в будущем.

Как получилось можно посмотреть здесь. Реализовано на платформе Node.js с использованием Bootstrap, Google Charts и Google Таблицы для хранения исходных данных.

Исходные данные о публикациях в Google Spreadsheet





Гугл-таблица LynxReport: учёт публикаций содержит все исходные данные и аналитику по публикациям. Я поддерживаю актуальность сведений на вкладке «Данные», вручную вписывая новые ссылки на статьи, остальное скачивается по большей части автоматически.

	A	В	C	D	Ε	F	G	н		J	K	L
1	Название	Тип (задается вручную)		Выходные данные (задается вручную)	Автор(ы) (задается вручную)	Тема (задается вручную)	Тираж для печатных изданий	Просмотры	Комментарии		Рейтинг (если доступно)	No n/n
3	Как я ищу ликвидные облигации на Московской бирже Т—Ж	Be6	2020-08-05	https://journal.tinkoff.ru/moex-bond-search/	М.В. Шардин	Гугл Таблицы		Loading	Loading	-		2
4	Скрипт выборки российских облигаций по параметрам Хабр	Be6	2020-06-22	https://habr.com/ru/post/506720/	М.В. Шардин	SilverFir		=IF(pages (cotw (cotw No. Account (cotw)		
5	Умная хрущёвка на максималках. Продолжение Хабр	Веб	2020-06-15	https://habr.com/ru/post/506550/	М.В. Шардин	Умный дом		ISNUMBER(I	R(DRVRNOW, 04, *//*[@class"post-stats_views-count']")) UTC UK.CG., *//*[@class"post-stats_views-count']"))			
6	Все финансовые рынки мира в АРІ Яку Финанс Хабр	Be6	2020-06-08	https://habr.com/ru/post/505674/	М.В. Шардин	API		·				
	Умная хрущёвка на максималках Хабр	Be6	2020-06-01	https://habr.com/ru/post/503646/	М.В. Шардин	Умный дом		SUBSTITUTE				
8	Free API Мосбиржи в формулах Microsoft Excel Хабр	Be6	2020-04-21	https://habr.com/ru/post/498268/	М.В. Шардин	API			04, -//*[BC18	iss=-post-1	(ecus (ecus le curus) de l'ecus le curus) de l'ecus de l'ecus de l'ecus de l'ecus de l'ecus (ecus de l'ecus (ecus	
9	Как я создал собственную гутл-таблицу для учета капитала Т—Ж	Be6	2020-04-16	https://journal.tinkoff.ru/investment-report/	М.В. Шардин	SilverFir						
10	Пишем генератор рандомных акций Мосбирки на JavaScript Хабр	Be6	2020-04-03	https://habr.com/ru/post/495324/	М.В. Шардин	JavaScript		<pre>value(sUBSTITUTE(SUBSTITUTE(IMPRORTOM_(ob, "//*[0class='post-stats_v ews-count']") ""."") ""."")>*1000</pre>				
11	5G Linux смартфон в форм-факторе миниатюрного ноутбука собрал средства через краундфандинг Хабр	Be6	2020-04-02	https://habr.com/ru/news/V495264/	М.В. Шардин	Новости						
12	Установка Ubuntu на Microsoft Surface Pro Хабр	Be6	2020-03-30	https://habr.com/ru/post/494210/	М.В. Шардин	Linux		10,800	21	28	+11	11

Часть таблицы LynxReport: учёт публикаций с исходными данными

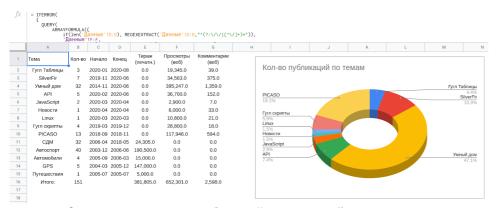
Актуальные данные по просмотрам и комментариям подгружаются через формулы.

Например, чтобы получить количество просмотров со страниц Хабра в ячейке гугл таблиц используется формула:

```
=IF(
ISNUMBER(IMPORTXML(D6, "//*[@class='post-stats_views-count']"))

,
SUBSTITUTE(
IMPORTXML(D6, "//*[@class='post-stats_views-count']")
,",",".")
,
value(SUBSTITUTE(SUBSTITUTE(
IMPORTXML(D6, "//*[@class='post-stats_views-count']")
,"k","")
,",","."))*1000
)
```

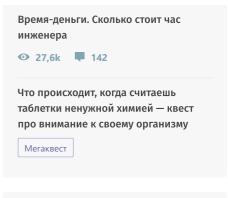
Формулы это не самый быстрый вариант и для того, чтобы получить несколько сотен позиций приходится ждать около получаса. После окончания загрузки можно видеть все цифры как на скриншоте ниже. Они дают ответы какие темы популярны, а какие нет.

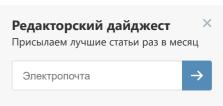


Часть таблицы LynxReport: учёт публикаций с аналитикой

Считывание данных из Таблицы и преобразование в формат Google Charts

Чтобы трансформировать эти сводные данные из гугл таблицы в сайт-визитку мне надо было преобразовать данные в формат временной шкалы Google Charts.







Получившаяся временная шкала Google Charts на сайте-визитке

Для того, чтобы корректно отрисовать такой график данные должны быть организованы следующим образом:

```
dataTable.addRows([]

// Начало вставки из сгенерированного файла
["Публикации", "https://journal.tinkoff.ru: 2 шт.", new Date(2020, 3), new Date(2020, 08)],
["Публикации", "https://journal.tinkoff.ru: 2 шт.", new Date(2019, 10), new Date(2020, 06)],
["Публикации", "https://sithub.com: 10 шт.", new Date(2019, 8), new Date(2020, 03)],
["Публикации", "https://medium.com: 1 шт.", new Date(2020, 1), new Date(2020, 02)],
["Публикации", "https://t.me: 4 шт.", new Date(2019, 2), new Date(2020, 01)],
["Публикации", "https://zen.yandex.ru: 1 шт.", new Date(2019, 11), new Date(2019, 12)],
["Публикации", "https://zen.yandex.ru: 13 шт.", new Date(2019, 11), new Date(2019, 11)],
["Публикации", "https://zot.u: 1 шт.", new Date(2019, 10), new Date(2019, 11)],
["Публикации", "Журнал Инновации и инвестиции: 2 шт.", new Date(2018, 3), new Date(2018, 05)],
["Публикации", "Сборник ПНИПУ: 7 шт.", new Date(2012, 3), new Date(2017, 05)],
["Публикации", "https://z-wave.ru: 16 шт.", new Date(2014, 10), new Date(2015, 04)],
["Публикации", "https://z-wave.ru: 16 шт.", new Date(2014, 10), new Date(2015, 04)],
```

Данные для Google Charts на сайте-визитке в html виде

Чтобы выполнять все преобразования автоматически я написал под Node.js скрипт, который доступен на GitHub.

Если вы не знакомы с Node.js, то в своей предыдущей статье я подробно расписал как можно воспользоваться скриптом под разными системами:

- 1. Windows
- 2. macOS
- 3. Linux

Ссылка с инструкциями здесь. Принцип аналогичен.

```
PROBLEMS OUTPUT TERMINAL DEBUGCONSOLE

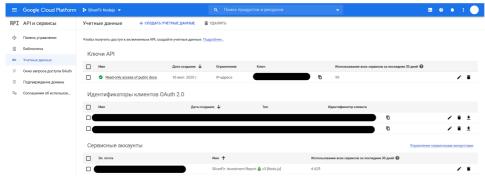
Code

C
```

Работа скрипта по преобразованию в нужный формат данных и генерации pdf версий статей с сайтов (все строки обрабатываются мгновенно — я специально поставил задержку, чтобы записать это видео)

Для того считывать данные из гугл таблицы в автоматическом режиме я пользуюсь авторизацией по ключу.

Получить этот ключ можно в консоли управления проектами гугла:



Учетные данные в Google Cloud Platform

После завершения работы скрипта должны сгенерироваться два текстовых файла с html данными графиков и все pdf копии онлайн статей.

Данные из текстовых файлов я импортирую в html код сайта-визитки.

Генерация pdf копий статей с сайтов

При помощи Puppeteer сохраняю текущий вид статей вместе со всеми комментариями в pdf виде.

Если не ставить задержку, то несколько десятков статей по списку можно сохранить в виде pdf файлов всего за несколько минут.

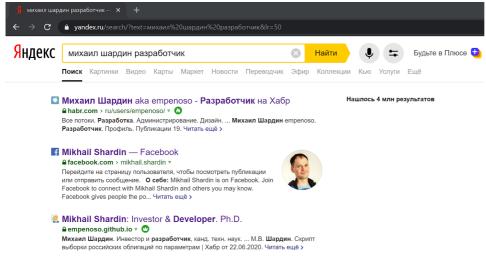
А задержка нужна для того чтобы на некоторых сайтах (например на Т—Ж) успели подгрузиться комментарии.

Результаты

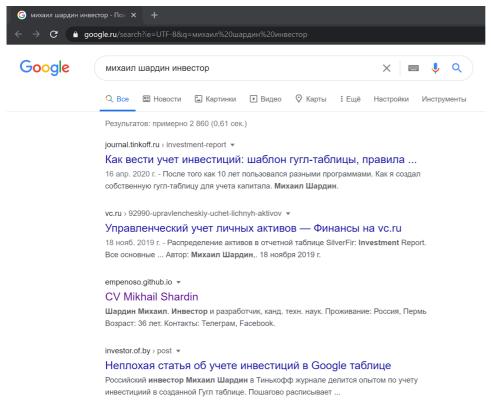
Поскольку написание скрипта затевалось с целью большего соответствия поисковым алгоритмам, то оценить результаты можно воспользовавшись поиском.

Поиск по имени и фамилии + указание специализации в обоих случаях возвращает ссылки на мои статьи и даже сайт-визитку:

В выдаче Яндекса:



В выдаче Гугла:



Пока что не могу решить — стоит ли регистрировать отдельное доменное имя, если визитка empenoso.github.io и так находится на верхних строчках поиска?

Вместо заключения

- 1. Возможно, эта статья заставит кого-то задуматься о том, как он выглядит в интернете.
- 2. Возможно, эта статья поможет кому-то наладить учёт и организацию публикаций.
- 3. Исходный код скрипта расположен на GitHub.

UPD. Доменное имя таки зарегистрировал: https://shardin.name/

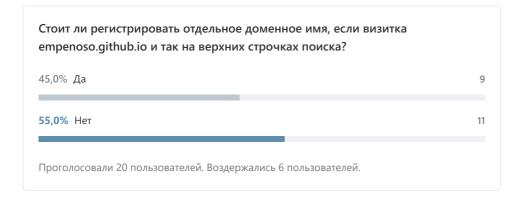
Автор: Михаил Шардин

17 августа 2020 г.

Только зарегистрированные пользователи могут участвовать в опросе. Войдите, пожалуйста.



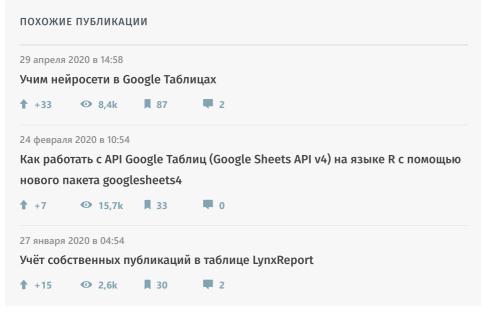
Проголосовал 21 пользователь. Воздержались 5 пользователей.



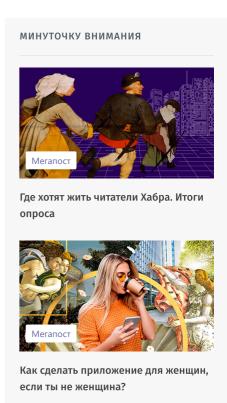
Теги: google sheets, Google Таблицы, статистика, CV, резюме, публикации, работа

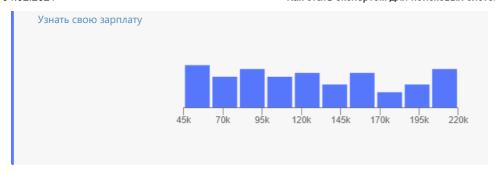
Хабы: JavaScript, Node.JS, Google API, Контент-маркетинг





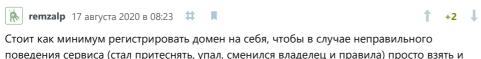












стоит как минимум регистрировать домен на сеоя, чтооы в случае неправильного поведения сервиса (стал притеснять, упал, сменился владелец и правила) просто взять и поднять у другого сервиса. При этом раскрученное доменное имя все еще с вами и оно не изменилось.



itsoft 17 августа 2020 в 09:22

Вот я себя также вбил yandex.ru/search/?

text=%D0%98%D0%B3%D0%BE%D1%80%D1%8C%20%D0%A2%D0%B0%D1%80%D0%B0%D1% %D1%81%D0%BF%D0%B5%D1%86%D0%B8%D0%B0%D0%BB%D0%B8%D1%81%D1%82&lr=21

И яндекс про меня всё выдаёт. Гугл выдаёт не только про меня. Но там проблема, что я не один айтишник с таким именем и фамилией.

Не понял я всё же достигнутый результат в условиях того, что есть и другие специалисты с таким же именем. А если их нет, то поисковики вас и так будут выдавать.

Куда эффективнее какой-нибудь никнейм или узкая специализация, тогда и проблем не будет. И даже подсказка в поисковой строке яндекса появляется.

Но и непонятно зачем. Ищут же не автора обычно, а тему, проблему. На автора всем пофиг. Никто не запоминает даже автора книги по айтишной тематике. Кому он нафиг нужен это автор? Он же не Булгаков, чтобы его запоминать. В ИТ нет Булагаковых. Те классики, что есть — выпустили по одной книге. Брукс, Керниган, Ритчи, Кодд, Страуструп, Кнут, Шилд, Котеров. Толку от знания этих фамилий немного, только дань уважения к самим авторам. Новое издание их книги будет на 90% совпадать с предыдущем.

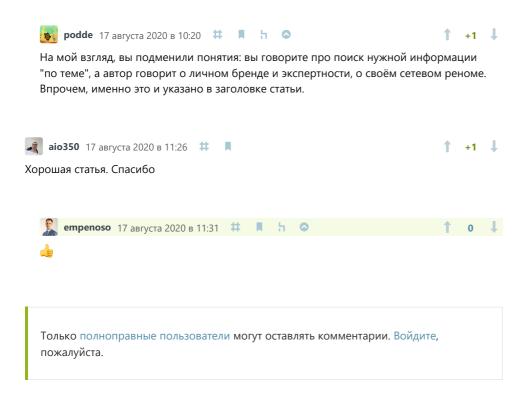


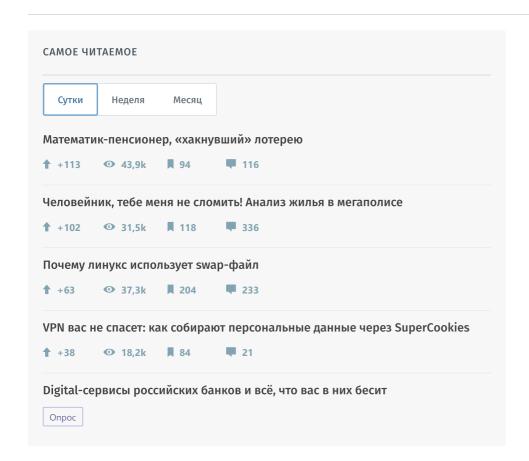
Мне кажется что тут дело в акцентах — когда у человека есть проблема он ищет про неё — тут неважно кто про неё пишет.

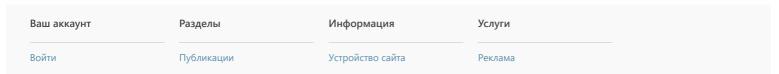
Но если человек хочет узнать кто этот автор, чем он занимается — он ищет по фамилии автора — и вот это уже тот самый случай.



+1







Регистрация Новости Для авторов Тарифы Хабы Для компаний Контент Документы Компании Семинары Пользователи Соглашение Мегапроекты Песочница Конфиденциальность Мерч y f w 🛪 🕨 🕂 © 2006 – 2021 «Habr» **(#)** Настройка языка О сайте Служба поддержки Мобильная версия