Dear Candidate,

We want you to write an article scraper for https://www.cermati.com/artikel

No need to scrape

only scrape article below

Catat! Begini Cara Mudah Investasi Reksa Dana Agar Cuan Maksimal

Investasi itu penting untuk mewujudkan berbagai tujuan keuangan. Investasi reksa dana bisa menjadi pilihan yang pas karena mudah, modal kecil dengan hasil cuan maksimal. Selengkapnya

sponsored Investasi • 15 Juli 2020

Kurangi 3 Kebiasaan Boros Ini Agar Bisa Menabung dan Investasi

Kendali keuangan ada pada Anda. Niatkan untuk memiliki keuangan yang sehat, bebas hutang, banyak menabung dan investasi. Simak tiga kebiasaan boros finansial dan solusinya. Selengkapnya

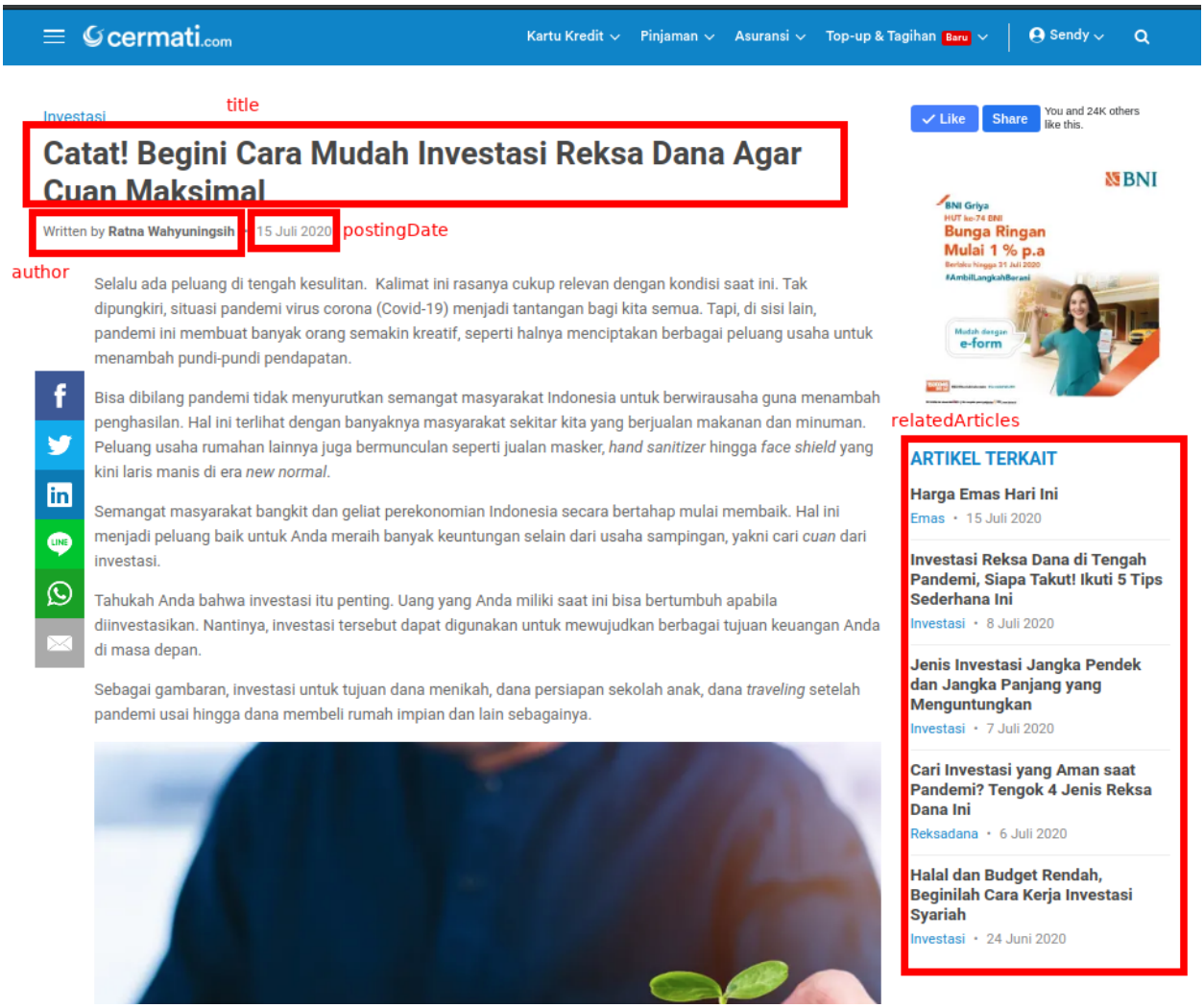
Gaya Hidup • 29 Juli 2020

Makin Mudah, Begini Cara Mencicil Elektronik Tanpa Kartu Kredit

As you can see in the red box, there are many articles on the website, the articles that you see on the screenshot might be different. You don't need to scrape articles in the blue box, only in the red box. Please scrape a list of articles on the cermati website, the scraper **code must be written in NodeJS**. Do not hard code the url of each article. But it's ok to hard code the main URL => this URL https://www.cermati.com/artikel as the starting point, **you only need to scrape the articles <u>on the first page</u>.**

You also need to scrape the details of each article:

- url
- title
- author
- postingDate
- relatedArticles



Put all the info in a single json file.
The JSON file has a format like this:

```
{
  “articles”: [
    {
      “url”: “...”,
      “title”: “...”,
      “author”: “...”,
      “postingDate”: “...”,
      “relatedArticles”: [
        {
          “url”: “...”,
          “title”: “...”
        },
        ….
      ]
    },
    { nextArticleObject... },
    { nextArticleObject... },
    { nextArticleObject... }
  ]
}
```

At Cermati, we value individuals who like challenges, are fast learners, and have an awesome work ethic.

I will share with you several npm packages that might be useful: (you don't have to use this though)

1. HTTP Client https://www.npmjs.com/package/superagent
2. https://www.npmjs.com/package/cheerio (to parse the html and pull data from it)
3. https://www.npmjs.com/package/bluebird (Promise library to handle the asynchronous nature of javascript)

Promise has a parallel function. So try to use the "parallel" way of scraping to make scraping faster.

How to submit the solution (PLEASE READ THIS IS IMPORTANT)

Please submit your solution before the due date that is mentioned in the email.

Please zip your entire solution directory and attach it into this email. Don't include the result json into this solution.

- Your code should be named 'solution.js' and we will run it like this: `node solution.js`
- make sure you also include your package.json and node_modules directory in the zip file.
- When running your 'solution.js', we expect the scraped_data will be put into a file called 'solution.json'.
- Please make sure the solution runs on node version 8. This is what we are going to use to determine the correctness of your solution.

Thanks, let us know if you have questions. And if you want to know more about our engineering team, please visit engineering.cermati.com