

Idea for the week

- ① figure out basic function
- ↳ functionality
 - ↳ database (source)
 - ↳ platform → web scraper.

coupon website

web scraping in javascript

↳ Javascript has only used a single thread and performed blocking operations

↳ use Node.js

↳ we don't need multithreading in JS

```
const http = require('http');
const PORT = 3000;

const server = http.createServer((req, res) => {
  res.statusCode = 200;
  res.setHeader('Content-Type', 'text/plain');
  res.end('Hello World');
});

server.listen(port, () => {
  console.log(`Server running at PORT:${port}/`);
});
```

→ not a blocking call
(returns immediately)

↳ need to use a HTTP client
Node.js has an HTTP library by default.

1. Built-in HTTP client

```
const http = require('http');

const req = http.request('http://example.com', res => {
  const data = [];
  res.on('data', _ => data.push(_))
  res.on('end', () => console.log(data.join()))
});

req.end();
```

push wanted data?

2. Fetch API

↳ still considered experimental
↳ also opt for polyfill/wrapper library

```
async function fetch_demo()
{
  const resp = await fetch('https://www.reddit.com/r/programming.json');

  console.log(await resp.json());
}

fetch_demo();
```

json object for response object

↳ accepts additional options argument

3. Axios

↳ similar to fetch
↳ Promised based HTTP client
must install

```
npm install axios
```

```
const axios = require('axios')

axios
  .get('https://www.reddit.com/r/programming.json')
  .then((response) => {
    console.log(response)
  })
  .catch((error) => {
    console.error(error)
  });
```

call this function

```

async function getForum() {
  try {
    const response = await axios.get(
      'https://www.reddit.com/r/programming.json'
    )
    console.log(response)
  } catch (error) {
    console.error(error)
  }
}

```

↳ superAgent
 ↳ has a fairly straightforward API
 like Axios but has more dependencies

```

const superagent = require("superagent")
const forumURL = "https://www.reddit.com/r/programming.json"

// callbacks
superagent
  .get(forumURL)
  .end((error, response) => {
    console.log(response)
  })

// promises
superagent
  .get(forumURL)
  .then((response) => {
    console.log(response)
  })
  .catch((error) => {
    console.error(error)
  })

// promises with async/await
async function getForum() {
  try {
    const response = await superagent.get(forumURL)
    console.log(response)
  } catch (error) {
    console.error(error)
  }
}

```

Benefits: extensibility — list of plugins
 allows for tweaking
 of a request or response
 ex. superagent-throttle plugin will allow you
 to define throttling rules for your request

50 Request

```
const request = require('request')
request('https://www.reddit.com/r/programming.json', function (
  error,
  response,
  body
) {
  console.error('error:', error)
  console.log('body:', body)
})
```

con: not actively maintained anymore

Google Map API ideas

- Add your own map markers
- Convert coordinates to an address
- Add click events to the map
- show place details
- add data layers

Web scraping

- Use axios call the js file "scraperapi.js"

```
1 const axios = require('axios');
2 const url = 'https://www.turmerriy.com/collections/organic-cotton-sheet-sets/prod
3   axios(url)
4     .then(response => {
5       const html = response.data;
6       console.log(html);
7     })
8     .catch(console.error);
```

After running scraper using "node scraperapi.js" in the terminal, it will pull a long unreadable string of HTML

```
<div class="tt-price" itemprop="price" content="99.0">
  <span class="sale-price">$99.00</span> == $0
  <span class="old-price">$119.00</span>
  <p class="price-fs hide" style="display: inline-block;
    ertant; /*font-size: 16px; color: #191919; */> + Free
    Shipping </p>
</div>
```

next step is to identify the elements we want to extract from the HTML

--use browser dev tools

-open the inspector tool of the website where you will see a corresponding line of code where you want get the element class

```
<div class="tt-price" itemprop="price" content="99.0">
  <span class="sale-price">$99.00</span> == $0
  <span class="old-price">$119.00</span>
  <p class="price-fs hide" style="display: inline-block;
    ertant; /*font-size: 16px; color: #191919; */> + Free
    Shipping </p>
</div>
```

Parse the HTML using cheerio

```
1 const axios = require('axios');
2 const cheerio = require('cheerio')
3 const url = 'https://www.turmerriy.com/collections/organic-cotton-sheet-sets/prod
4   axios(url)
5     .then(response => {
6       const html = response.data;
7       const $ = cheerio.load(html)
8       const salePrice = $('span.sale-price').text()
9       console.log(salePrice);
10     })
11     .catch(console.error);
```

After running your program, it will print the content tagged as .sale-price to the console:

```
PS C:\Users\N N\Desktop\firstscraper> node scraperapi.js
$99.00
PS C:\Users\N N\Desktop\firstscraper>
```

google api

① look into data sourcing
and how we are
going to integrate it

② algorithm → javascript
source code → html
↳ set up our algorithm.

↳ integrate the API
from Google maps to
JS.

③ UI parts what we want
to see
↳ website URL

↳ coupon
↳ store into different categories
↳ groceries
↳ coordinate value
↳ string.

dictionary

text
file

