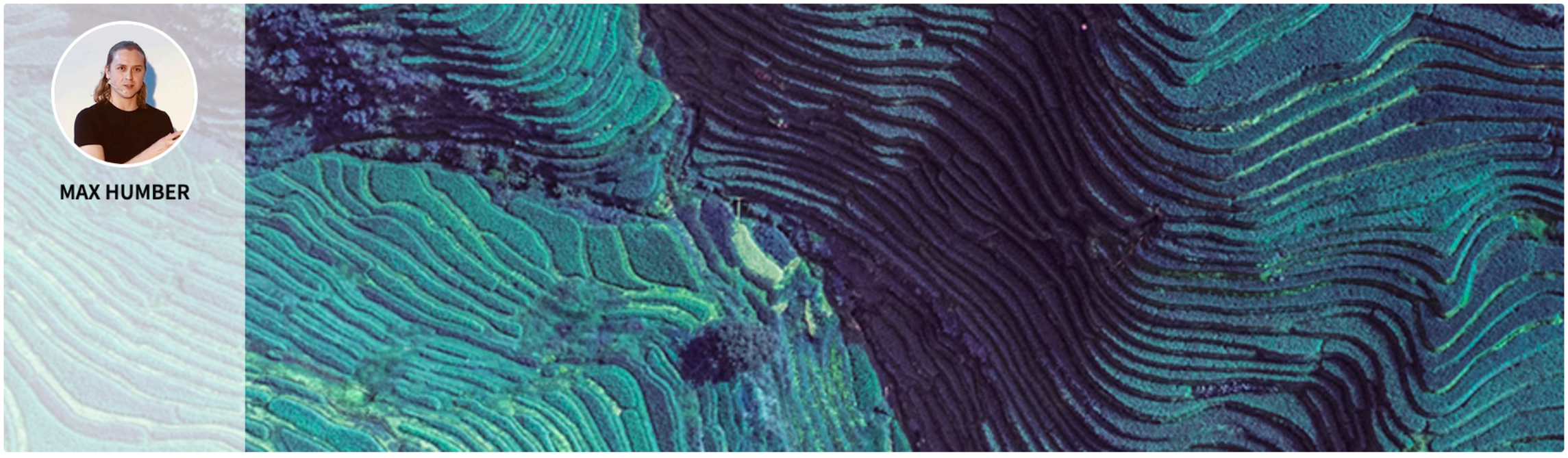
LIVE ONLINE TRAINING

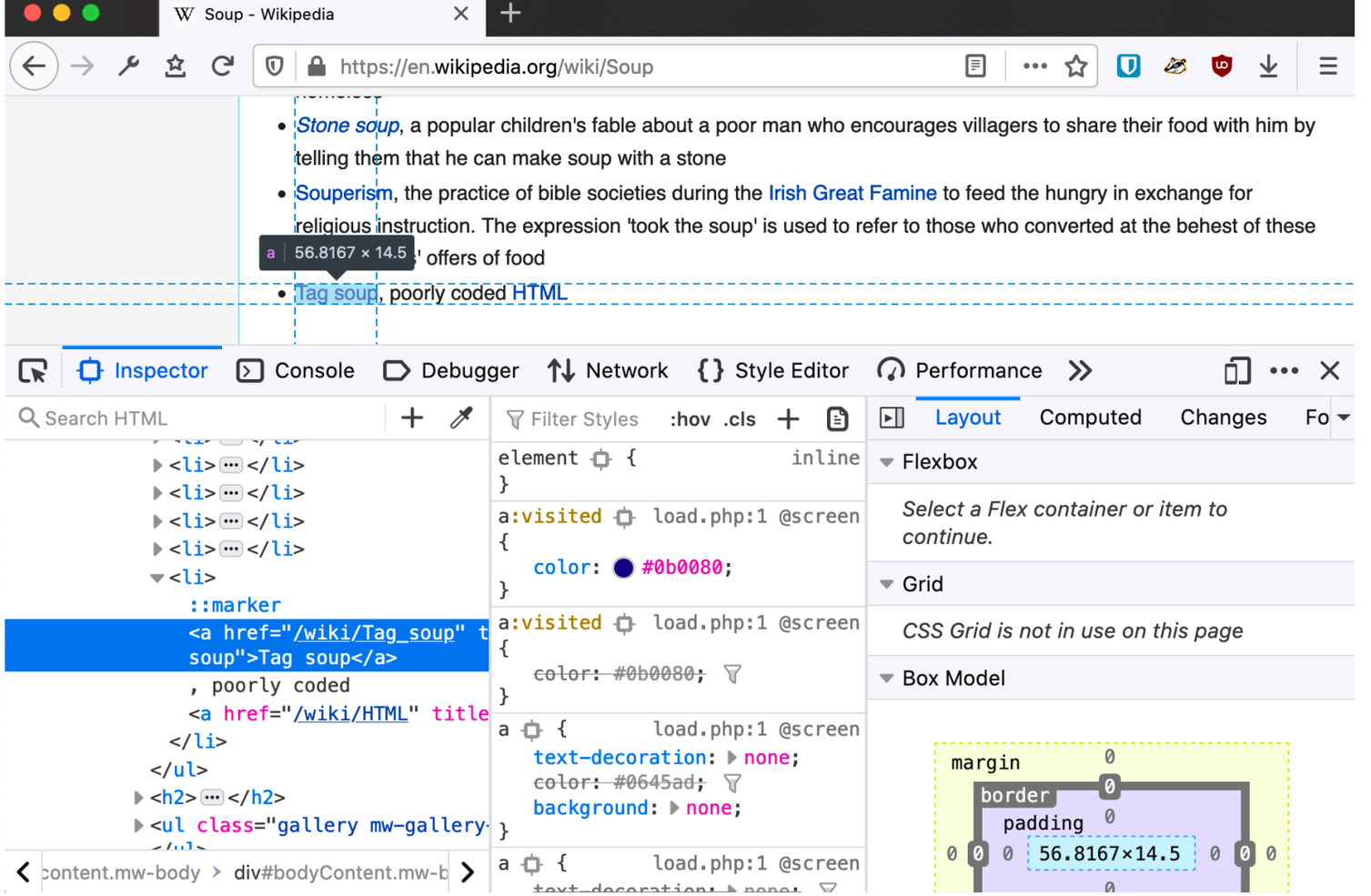
Advanced Web Scraping



https://resources.oreilly.com/binderhub/advanced-web-scraping

November 10, 2020

12:00pm – 1:00pm EST


Soup - Wikipedia
https://en.wikipedia.org/wiki/Soup
Stone soup, a popular children's fable about a poor man who encourages villagers to share their food with him by telling them that he can make soup with a stone
Souperism, the practice of bible societies during the Irish Great Famine to feed the hungry in exchange for religious instruction. The expression 'took the soup' is used to refer to those who converted at the behest of these
a 56.8167 × 14.5
' offers of food
Tag soup, poorly coded HTML
Inspector
Console
Debugger
Network
Style Editor
Performance
Search HTML
Filter Styles
:hov .cls
Layout
Computed
Changes
Flexbox
Select a Flex container or item to continue.
Grid
CSS Grid is not in use on this page
Box Model
margin
border
padding
56.8167×14.5
element { inline
a:visited load.php:1 @screen
color: #0b0080;
a:visited load.php:1 @screen
color: #0b0080;
a { load.php:1 @screen
text-decoration: none;
color: #0645ad;
background: none;
a { load.php:1 @screen

::marker
<a href="/wiki/Tag_soup" t
soup">Tag soup
, poorly coded
<a href="/wiki/HTML" title

<h2>
<ul class="gallery mw-gallery
content.mw-body
div#bodyContent.mw-b

Soup - Wikipedia

https://en.wikipedia.org/wiki/Soup

- *Stone soup*, a popular children's fable about a poor man who encourages villagers to share their food with him by telling them that he can make soup with a stone
- Souperism, the practice of bible societies during the Irish Great Famine to feed the hungry in exchange for religious instruction. The expression 'took the soup' is used to refer to those who converted at the behest of these ' offers of food
- Tag soup, poorly coded HTML

a | 56.8167 × 14.5

Inspector | Console | Debugger | Network | Style Editor | Performance

Search HTML

```
<li> … </li>
<li> … </li>
<li> … </li>
<li> … </li>
<li>
  ::marker
  <a href="/wiki/Tag_soup" t
  soup">Tag soup</a>
  , poorly coded
  <a href="/wiki/HTML" title
</li>
</ul>
<h2> … </h2>
<ul class="gallery mw-gallery-
```

content.mw-body › div#bodyContent.mw-b

Filter Styles :hov .cls

```
element { inline
}
a:visited load.php:1 @screen
{
  color: #0b0080;
}
a:visited load.php:1 @screen
{
  color: #0b0080;
}
a { load.php:1 @screen
  text-decoration: none;
  color: #0645ad;
  background: none;
}
a { load.php:1 @screen
```

Layout | Computed | Changes | Fo

Flexbox

Select a Flex container or item to continue.

Grid

CSS Grid is not in use on this page

Box Model

margin 0
border 0
padding 0
0 0 0 56.8167×14.5 0 0 0
0

Soup - Wikipedia

https://en.wikipedia.org/wiki/Soup

- *Stone soup*, a popular children's fable about a poor man who encourages villagers to share their food with him by telling them that he can make soup with a stone
- Souperism, the practice of bible societies during the Irish Great Famine to feed the hungry in exchange for religious instruction. The expression 'took the soup' is used to refer to those who converted at the behest of these ' offers of food
- Tag soup, poorly coded HTML

a | 56.8167 × 14.5

Inspector | Console | Debugger | Network | Style Editor | Performance

Search HTML

```
<li> … </li>
<li> … </li>
<li> … </li>
<li> … </li>
<li>
  ::marker
  <a href="/wiki/Tag_soup" t
  soup">Tag soup</a>
  , poorly coded
  <a href="/wiki/HTML" title
</li>
</ul>
<h2> … </h2>
<ul class="gallery mw-gallery-
```

content.mw-body › div#bodyContent.mw-b

Filter Styles :hov .cls

```
element { inline
}
a:visited load.php:1 @screen
{
  color: #0b0080;
}
a:visited load.php:1 @screen
{
  color: #0b0080;
}
a { load.php:1 @screen
  text-decoration: none;
  color: #0645ad;
  background: none;
}
a { load.php:1 @screen
```

Layout | Computed | Changes | Fo

Flexbox

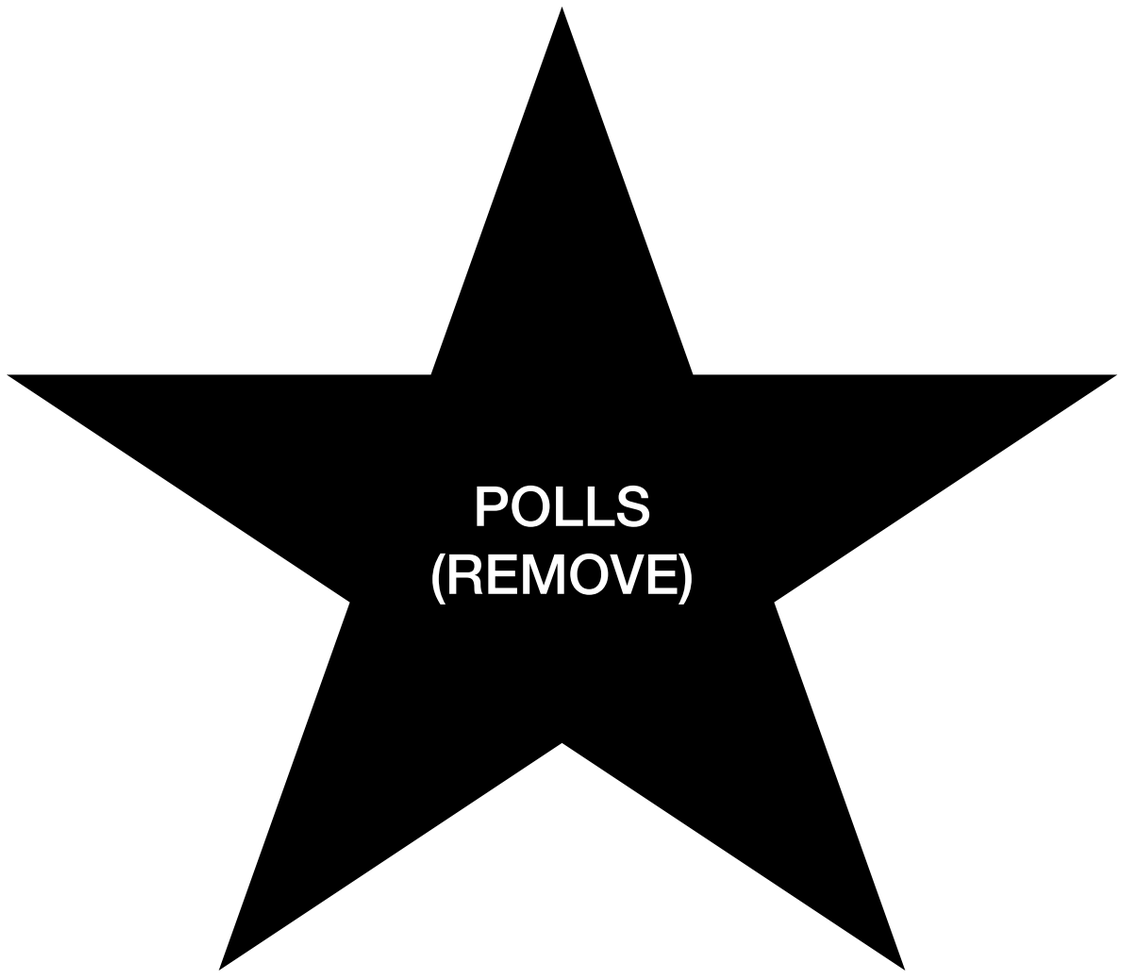
Select a Flex container or item to continue.

Grid

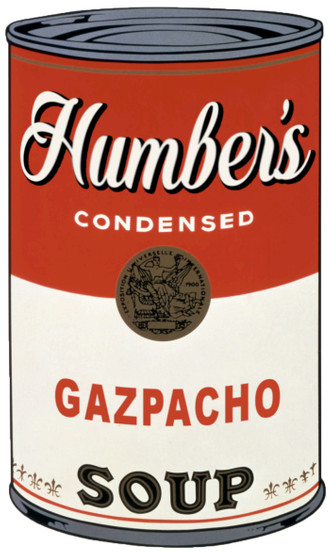
CSS Grid is not in use on this page

Box Model

margin 0 border 0 padding 0 0 0 0 56.8167×14.5 0 0 0 0



- Poll: How many websites have you scraped before? {0, 1, 10, 100+}
- Poll: Is your interest in web scraping professional or personal? {professional, personal}


Humber's
CONDENSED
GAZPACHO
SOUP

33

5

57

BeautifulSoup is so 2000-and-late: Web Scraping in 2020

#python #webscraping #gazpacho #hacktoberfest

https://dev.to/maxhumber/beautifulsoup-is-so-2000-and-late-web-scraping-in-2020-2528

1. No dependencies
2. **Batteries included**
3. **1 find method**
4. Production ready
5. PEP 561 compliant
6. Automatic formatting
7. **Speed (50% faster)**
8. **Partial matching**
9. Debt-free (3 first)
10. Open (and Friendly)

> 500/week
> 450 now

Max Humber
Human
1mo •

🍾 gazpacho JUST TURNED 1 🍾

Some highlights from this year:

⬇️ 20,000 installs
⭐ 300 stars
💻 Used by teams at Apple, Microsoft, and Facebook

To mark the occasion I've just bumped the library out of ZeroVer hell and released version 1.0!

Improvements include:

❓ type: hints = everywhere
🖼️ The ability to parse malformed void tags
🍲 A new way to initialize a parser with: Soup.get("url")

Details: **https://gazpacho.xyz**/

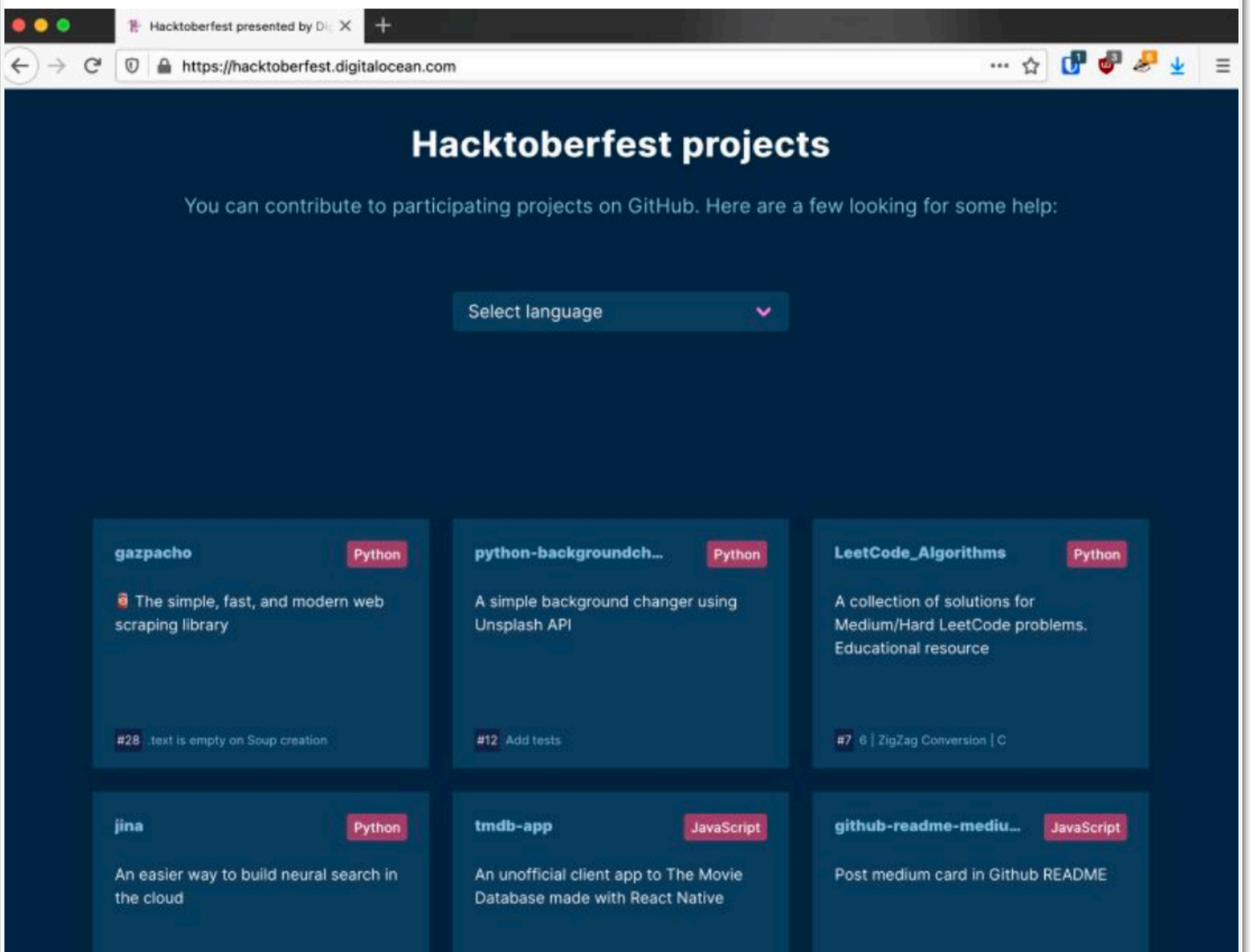
And if you'd like to help build out the roadmap for gazpacho please fill out this quick (seven questions, all optional) survey:

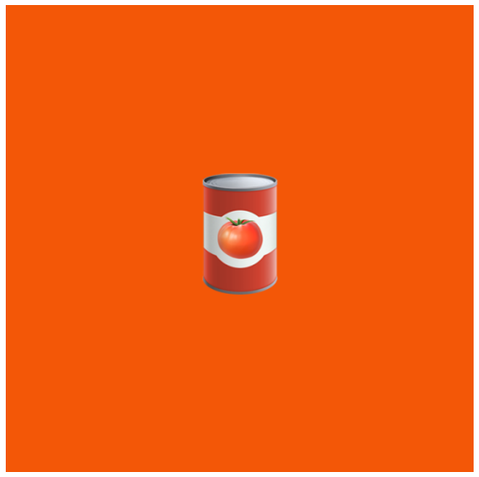
https://lnkd.in/gKWzRqM

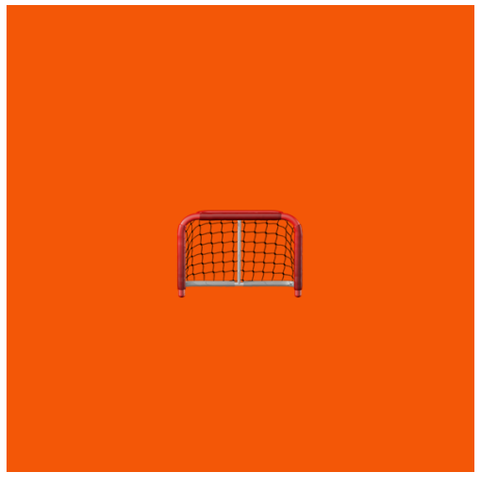
Max Humber
Human
2w •

gazpacho is #1 on **#Hacktoberfest** right now

After a full year of development it finally feels like the *most beautiful soup* 😉 has some real momentum behind it







404

https://www.scrape.world

Welcome to Scrape World

This website is meant to be scraped.

The HTML on this website is garbage...

This is on purpose *wink*

Each page is a challenge.

To start a challenge click on the accordion:

Show Challenge

Try to complete each challenge *before* peeking at the solution.

Happy scraping!

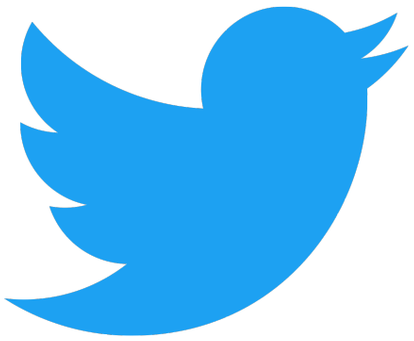

LETS GO
4TH 2:32
24

Q&A

That's all Folks!

https://github.com/maxhumber/gif

https://github.com/maxhumber/marc



twitter.com/maxhumber

www.linkedin.com/in/maxhumber