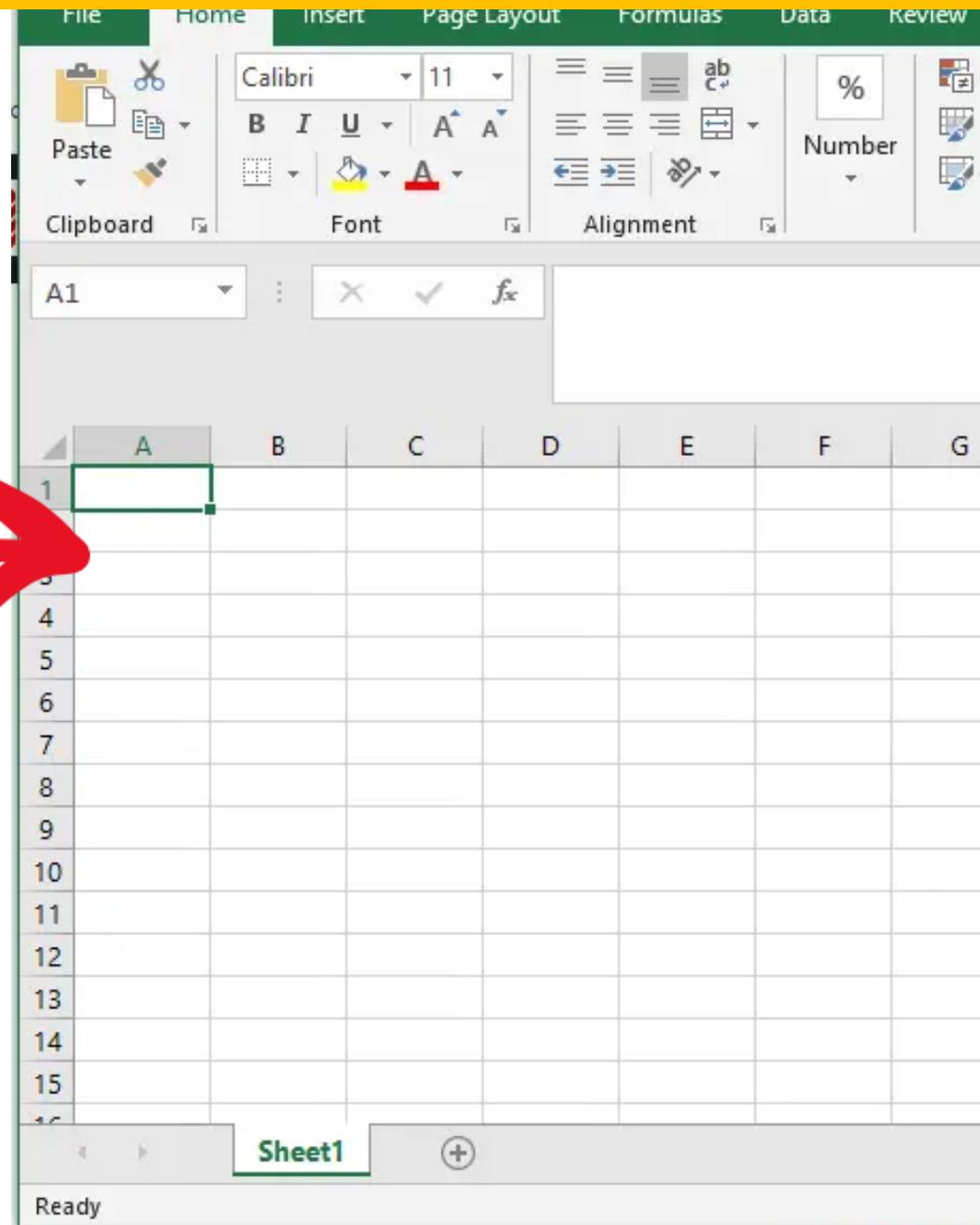
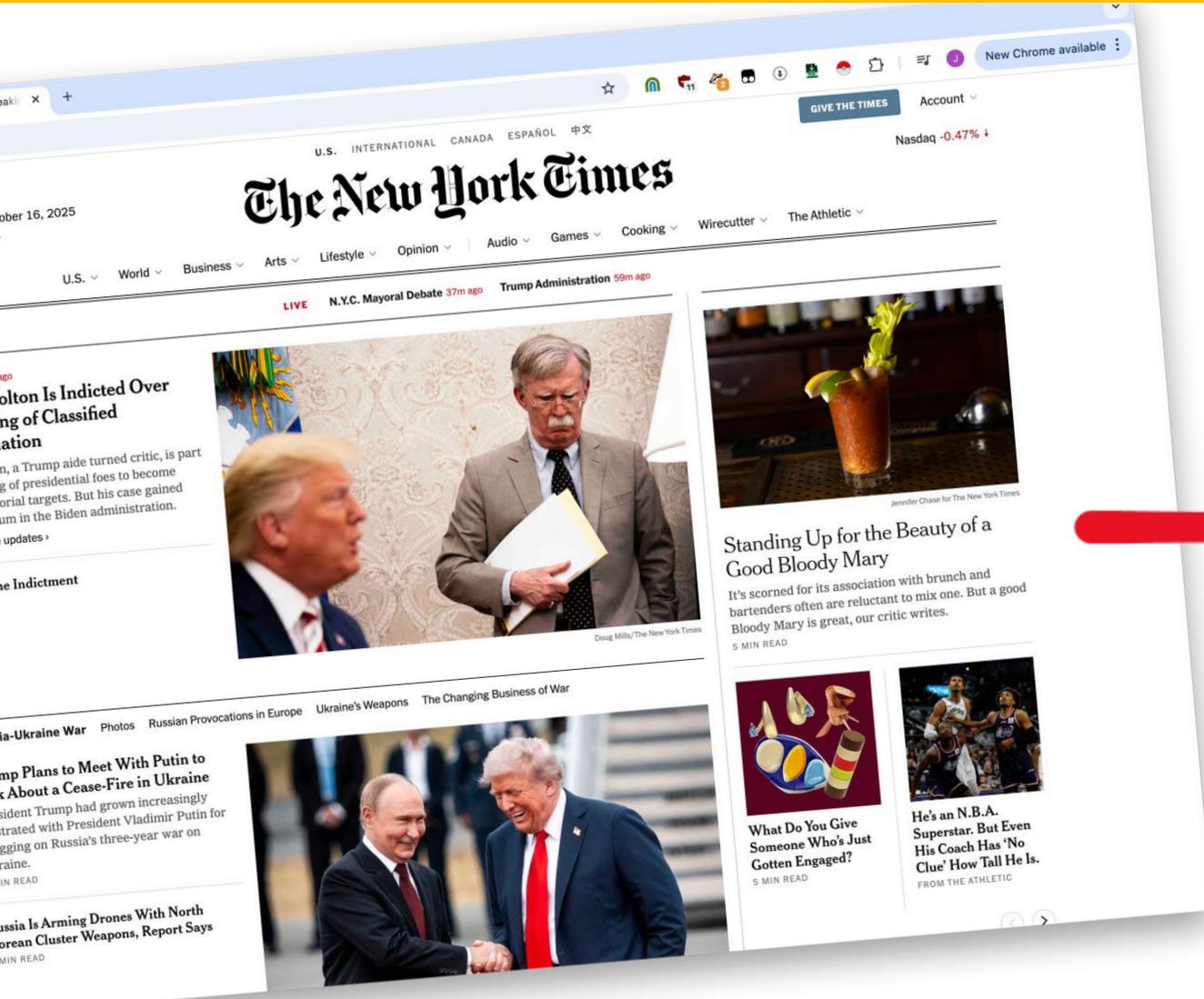


Visit <https://bit.ly/browser-automation> for material

# Scraping!

stealing (unstructured) data  
from the internet for fun

Visit <https://bit.ly/browser-automation> for material



Visit <https://bit.ly/browser-automation> for material

# HTML

the language that all web  
sites are built with

<h1>This is a headline</h1>

<h2>This is a smaller headline</h2>

<h3>And an even smaller headline</h3>

**This is a headline**

**This is a smaller headline**

**And an even smaller headline**

```
<h1>This is a headline</h1>
```

```
<h2>This is a smaller headline</h2>
```

```
<h3>And an even smaller headline</h3>
```

```
<p>This is a paragraph</p>
```

```

```

```
<p>This is a paragraph with
```

```
<a href="google.com">a link</a></p>
```

```
<div>This is ANYTHING</div>
```

```
<div>Anything! Anything in the world</div>
```

# This is a headline

## This is a smaller headline

### And an even smaller headline

This is a paragraph

This is a paragraph with [a link](#)

This is ANYTHING  
Anything! Anything in the world



(imagine there's a photo)

# <h1>An incredible story</h1>

<p>By J. Soma</p>

<p>This is the start of the story.</p>

<p>It is an amazing story!</p>

<p>”It’s incredible,” said the source.</p>

<p>An editor agreed: “it’s true!”</p>

<p>Additional reporting by Mulberry the fat mean cat</p>

<p>Edit: An earlier version said “Jonathan”  
Soma</p>

# An incredible story

By J. Soma

This is the start of the story.

It is an amazing story!

”It’s incredible,” said the source.

An editor agreed: “it’s true!”

Additional reporting by Mulberry the fat mean cat

Edit: An earlier version said “Jonathan” Soma

# <h1>An incredible story</h1>

<p>By J. Soma</p>

<p>This is the start of the story.</p>

<p>It is an amazing story!</p>

<p>”It’s incredible,” said the source.</p>

<p>An editor agreed: “it’s true!”</p>

<p>Additional reporting by Mulberry the fat  
mean cat</p>

<p>Edit: An earlier version said “Jonathan”  
Soma</p>

```
<h1>An incredible story</h1>
```

```
<p id="byline">By J. Soma</p>
```

```
<p>This is the start of the story.</p>
```

```
<p>It is an amazing story!</p>
```

```
<p>"It's incredible," said the source.</p>
```

```
<p>An editor agreed: "it's true!"</p>
```

```
<p class="additional">Additional reporting by  
Mulberry the fat mean cat</p>
```

```
<p class="correction">Edit: An earlier version  
said "Jonathan" Soma</p>
```

# <h1>An incredible story</h1>

<p id="byline">By J. Soma</p>

<p>This is the start of the story.</p>

<p>It is an amazing story!</p>

<p>"It's incredible," said the source.</p>

<p>An editor agreed: "it's true!"</p>

<p class="additional">Additional reporting by  
Mulberry the fat mean cat</p>

<p class="correction">Edit: An earlier version said "Jonathan"  
Soma</p>

	A	B	C	D	E
1	headline	author	text	additional reporting	corrections
2	An incredible story	By J Soma	Blah blah blah blah	Additional reporting by Mulberry the fat mean cat	Edit: An earlier version said "Jonathan" Soma
3					
4					

<p>

<h1>

<h1>An incredible story</h1>

<p id="byline">By J. Soma</p>

<p>This is the start of the story.</p>

<p>It is an amazing story!</p>

<p>"It's incredible," said the source.</p>

<p>An editor agreed: "it's true!"</p>

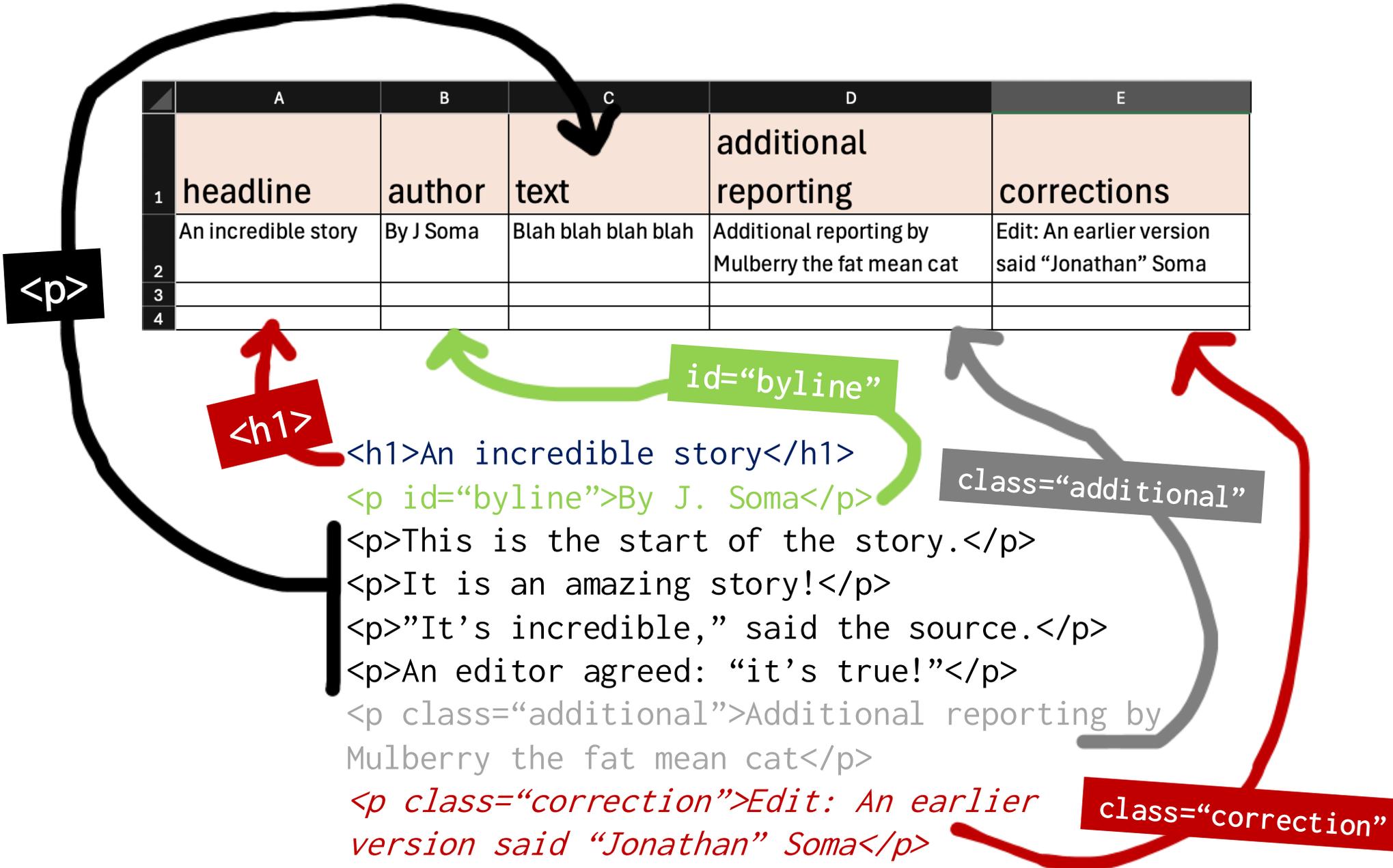
<p class="additional">Additional reporting by Mulberry the fat mean cat</p>

<p class="correction">Edit: An earlier version said "Jonathan" Soma</p>

id="byline"

class="additional"

class="correction"



Let's see it on  
the internet!

BBC Home - Breaking News, x +

bbc.com

Register Sign In

Home News Sport Business Innovation Culture Travel Earth Video Live

**What will be a row in our spreadsheet?**

**LIVE** VP pick Walz to speak as Democrats deploy star guests

The Minnesota governor will run alongside presidential nominee Kamala Harris for November's US election.

Divers find five bodies in wreck of Sicily yacht

**Three things the Democrats have avoided so far at the DNC**

What they have tried to avoid says as much about their weaknesses as what they choose to highlight, writes Anthony Zurcher.

2 hrs ago | US & Canada

- Day three of Democratic Convention features party's rising stars
- Obamas mock Trump over 'black jobs' and crowd sizes

**Gaza nurse says whole family, including quadruplets, killed in air strike**

...wife and six children were ...  
5 hrs ago

**Andrew Tate held overnight after police raid homes in Romania**

The internet personality and his brother have been remanded in custody as part of a probe into new allegations.

44 mins ago | Europe

**Ancient ocean of magma found on Moon south pole**

The findings are from India's historic Chandrayaan-3 mission that landed on the Moon's south pole.

6 hrs ago | Science & Environment

**German Navy blasts out Darth Vader theme on Thames**

BBC Home - Breaking News, x +

bbc.com

Register Sign In

Home News Sport Business Innovation Culture Travel Earth Video Live



**LIVE** VP pick Walz to speak as Democrats deploy star guests

The Minnesota governor will run alongside presidential nominee Kamala Harris for November's US election.



### Three things the Democrats have avoided so far at the DNC

What they have tried to avoid says as much about their weaknesses as what they choose to highlight, writes Anthony Zurcher.

2 hrs ago | US & Canada

- Day three of Democratic Convention features party's rising stars
- Obamas mock Trump over 'black jobs' and crowd sizes

### Gaza nurse says whole family, including quadruplets, killed in air strike

Ashraf El Attar's wife and six children were killed when an air strike destroyed their home in Deir al-Balah.

hrs ago | Middle East

### Andrew Tate held overnight after police raid homes in Romania

The internet personality and his brother have been remanded in custody as part of a probe into new allegations.

44 mins ago | Europe

### Ancient ocean of magma found on Moon south pole

The findings are from India's historic Chandrayaan-3 mission that landed on the Moon's south pole.

6 hrs ago | Science & Environment

### German Navy blasts out Darth Vader theme on Thames



### Divers find five bodies in wreck of family yacht

*What will be a column of data?*

# Andrew Tate held overnight after police raid homes in Romania

The internet personality and his brother have been remanded in custody as part of a probe into new allegations.

44 mins ago | Europe



**📍LIVE VP pick Walz to speak as Democrats deploy star guests**

The Minnesota governor will run alongside presidential nominee Kamala Harris for November's US election.

Browser tabs: Data sets - Open data

Address bar: data.gov.rs/sr/datasets/

# Data sets

2435 sets in total [Export](#) [Download data](#) →

Data search.

Schedule according to... Organizations

---



**CATALOG OF CONTAMINATED LOCATIONS**

The cadastre of contaminated sites represents a set of data on endangered, polluted and degraded lands and it is an integral part of the Land Information System managed by the Environmental Protection Agency. State bodies, i.e. organizations, bodies of autonomous provinces, local self-government units...

resources 0  
use 0  
supporting 0

---



**Soil condition in urban areas**

The data show the degree of soil endangerment from chemical pollution in urban areas in the period 2005-2023. year. Soil testing was carried out at locations in the immediate vicinity of traffic roads, industrial zones, near landfills, water supply sources, educational institutions, recreation...

resources 22  
use 0  
supporting 0

---



**Realization regular program**

A set of data on the performed services of the Animal Hygiene Program.

resources 6  
use 0  
supporting 0

---

<https://data.gov.rs/sr/datasets/stanje-zemljishta-u-urbanim-sredinama/> **Training providers**

resources 1  
use 0

*What will be a row in our spreadsheet?*



What will be a row of data?

## 3,183,312 Titles through 2020

Search Titles

### FILTER SYLLABI

Enter a search keyword or a concept

By Institution ▾

By Field ▾

By Country ▾

By State (US only) ▾

From ▾

To ▾

By School Type (US only) ▾

### FILTER TITLES

RANK	TITLES	APPEARANCES	SCORE
1	<b>Calculus</b> James Stewart	18,299	100
2	<b>The Elements of Style</b> William Strunk	15,020	100
3	<b>Social Research Methods</b> Alan Bryman <i>Oxford University Press</i>	14,704	100
4	<b>Human Anatomy and Physiology Laboratory Manual</b> Elaine Nicpon Marieb	14,542	100
5	<b>English Grammar in Use</b> Raymond Murphy <i>Cambridge University Press</i>	14,101	100
	<b>Research Desian</b>		

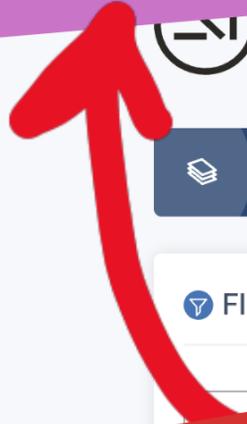
What will be a column of data?



Computers don't see like us  
(usually), we need to see

HTML code

Visit OpenSyllabus



Do this now

The screenshot shows the OpenSyllabus website interface. At the top, there is a navigation bar with 'OS Home', 'Explore', and 'About' links, along with 'SIGN UP / IN' and 'LEARN MORE' buttons. Below the navigation bar, a 'TITLES' button is highlighted. To the left of the main content area is a 'FILTER SYLLABI' sidebar with various dropdown menus for filtering titles. The main content area displays a list of titles with columns for Rank, Title, Appearances, and Score. The top of the list shows '3,183,312 Titles through 2020'.

RANK	TITLES	APPEARANCES	SCORE
1	<b>Calculus</b> James Stewart	18,299	100
2	<b>The Elements of Style</b> William Strunk	15,020	100
3	<b>Social Research Methods</b> Alan Bryman <i>Oxford University Press</i>	14,704	100
4	<b>Human Anatomy and Physiology Laboratory Manual</b> Elaine Nicpon Marieb	14,542	100
5	<b>English Grammar in Use</b> Raymond Murphy <i>Cambridge University Press</i>	14,101	100

RANK

TITLES

APPEARANCES

SCORE

1

**Calculus**

James Stewart

18,299

100

2

**The Elements**

William Strunk

15,020

100

**Qualitative Research**

Bryman

London University

14,704

100

4

**Human Anatomy  
Manual**

Elaine Nicpon M

14,101

100

5

**English Grammar**

Raymond Murph

Cambridge Univ

14,101

100

- Open Link in New Tab
- Open Link in New Window
- Open Link in Incognito Window

- Save Link As...
- Copy Link Address

- Copy
- Copy Link to Highlight
- Search Google for "The Elements of Style"
- Print...
- Translate Selection to English

 View Web Archives

**Inspect**

Right click (or  
command+click)  
and Inspect

**Inspect**

Find the HTML  
for our "item"

Search Titles

3,183,312 Titles through 2020

### Calculus

James Stewart

1

☆ 100 🔍 18,299

div.sc-9d100f21-16.gsKnKH 301x20

William Strunk

2

☆ 100 🔍 15,020

### Social Research Methods

Alan Bryman

Oxford University Press

3

☆ 100 🔍 14,704

Mouse/click  
around

METHOD ONE

Elements

Console

Sources

```
<table class="sc-9d100f21-0 hzyFdQ"> flex
  <thead class="sc-9d100f21-1 kpKKDZ"> ... </thead>
  <tbody class="sc-9d100f21-5 gmZXrp"> flex
    <tr class="sc-9d100f21-6 kbqQUR"> ... </tr> flex
    <tr class="sc-9d100f21-6 duwfJU"> flex
      <td class="sc-9d100f21-14 blmzDX"> ... </td> flex
      <td class="sc-9d100f21-14 BLSKu"> flex
        <ul class="sc-9d100f21-23 jsBOD">
          <li class="sc-9d100f21-24 ldoMe"> ... </li> flex
          <li class="sc-9d100f21-24 cjFXUz"> flex
            <div class="sc-9d100f21-16 gsKnKH"> flex
              <a href="/singleton/authors?id=William%20Strunk"> ... </a> == $0
              <p class="sc-9d100f21-15 OIORn"></p>
            </div>
          </li>
        </ul>
      </td>
    </tr>
  </tbody>
</table>
```

Styles

Computed

Layout

Event Listeners

DOM Breakpoints

Filter

:hov .cls

Find the HTML  
for our "item"

Click  
this

Search +  
click here

METHOD TWO

Search Titles

3,183,312 Titles through 2020

### Calculus

James Stewart

1

☆ 100 🔍 18,299

### The Elements of Style

William Strunk

2

☆ 100 🔍 15,020

### Social Research Methods

Alan Bryman

Oxford University Press

3

☆ 100 🔍 14,704

```
<thead class="sc-9d100f21-1 kpKKDZ">...</thead>
<tbody class="sc-9d100f21-5 gmZXrp"> flex
  <tr class="sc-9d100f21-6 kbqQUR">...</tr> flex
  <tr class="sc-9d100f21-6 duwfJU"> flex == $0
    <td class="sc-9d100f21-14 blmzDX">...</td> flex
    <td class="sc-9d100f21-14 BLSKu"> flex
      <ul class="sc-9d100f21-23 jsBOD">
        <li class="sc-9d100f21-24 ldoMe">...</li> flex
        <li class="sc-9d100f21-24 cjFXUZ"> flex
          <div class="sc-9d100f21-16 gsKnKH"> flex
            <a href="/singleton/authors?id=William%20Strunk">...</a>
            <p class="sc-9d100f21-15 OIOrn"></p>
          </div>
        </li>
      </ul>
    </td>
  </tr>
</tbody>
```

Scraping is in  
connected  
But we need some software  
to do that for us!  
to columns  
in your spreadsheet

What software?



- All
- Images
- Videos
- News
- Books
- Web
- Finance
- Tools

## Scraping libraries

From sources across the web

- |  |   |  |
|--|---|--|
|  Selenium       |  Beautiful Soup |  Scrapy         |
|  Lxml           |  Requests       |  MechanicalSoup |
|  Playwright    |  Puppeteer     |  Urllib3       |
|  ZenRows S.L. |  Cheerio      |  |

[-] **Swingbiter** 70 points 2 years ago

Learn the basic html elements that build up a website.

[-] **coventous** 22 points 2 years ago

I recommend checking out Selenium. It's a tool that lets developers test their websites

[-] **NerdvanaNC** 1 point 2 years ago  
Look into learning and understanding HTML, then BeautifulSoup and Selenium. Web developers test their websites with Selenium, so Google for that. Oh and bonus tip, there's (official?) Docker (or proxy/chrome/etc) which made me very happy (I like Docker!)

[-] **luizv4z** 12 points 2 years ago

From my own research, run away from Selenium. The right direction is CDP (Chrome Developers Protocol). From my own tool, I could scrap Facebook without getting banned.

```
2 all_links = soup.find_all('a')
```

Do python on them until

[-] **riisen** 3 points 2 years ago  
for small projects use built in module req  
for bigger projects go with scrapy, its am  
permalink source embed save save-RES repo

P.S.

permalink source embed save save-RES report reply hide child comments

[-] **ned334** 5 points 2 years ago

Google "Selenium find\_element(By.XPATH, '/XPath/')"

All elements have an XPath that you can copy from chrome by Inspect -> right click on code block -> copy full Xpath.

Scraping solved

permalink source embed save save-RES report reply

Selenium is garbage!

...but BeautifulSoup  
can't scrape all sites.

### Top discounts for you



Continental Grand Prix 5000 Rennrad Faltrreifen // 28-622 (700x28C)   
★★★★☆ 85   
400+ viewed in past month   
\$66.00   
✓prime FREE Delivery Friday, Aug 23



Wink Super X-Large Ultra Thin Lubricated Latex Condoms, Premium Latex for Smooth and Natural...   
★★★★☆ 76   
100+ viewed in past month   
\$14.99 (\$0.30/Count)   
✓prime FREE Delivery Friday, Aug 23



Assortment Rubber Ducks in Bulk, 50-Pack Assorted Mini Duckies Toy for Ducking Cruise Ships, 2" ...   
★★★★☆ 328   
5K+ viewed in past month   
\$24.99   
✓prime Overnight by 8:00 AM



50 Pack Rubber Ducks in Bulk, Jeep Ducks for Ducking, Assorted Rubber Ducks Jeep Ducking, Bab...   
★★★★☆ 569   
4K+ viewed in past month   
\$21.89   
✓prime Overnight by 8:00 AM



ArtCreativity Assorted Rubber Ducks Jeep Ducking (100Pack) - Rubber Ducki...   
★★★★☆ 982   
700+ viewed in past month   
**20% off Limited time deal**   
\$39.98   
List: \$49.99   
✓prime Today by 10:00 PM



Kicko Assorted Rubber Ducks with Mesh Bag - 50 Ducklings, 2 Inch - Jeep Ducks for Kids, Baby Bath...   
★★★★☆ 2,543   
2K+ viewed in past month   
\$28.99   
✓prime FREE Delivery Saturday, Aug 24



Glitter Rubber Ducks in Bulk - (Pack of 50) Assorted 2-inch Duck Toys for Baby Shower Rubber Duckies,...   
★★★★☆ 308   
1K+ viewed in past month   
\$26.79   
✓prime Today by 10:00 PM

### Bulk savings to consider



Chochkees Assorted Rubber Ducks Toy Duckies for Kids and Toddlers, Bath Birthday Baby Showers Classroom,...   
★★★★☆ 1,155



Rubber Ducks Jeep Ducking (50 Pack) - Rubber Duckie...   
★★★★☆ 982   
100+ viewed in past month



Set, Mini Colorful Rubber Duckies Bath Toy for Child,Float & Squeak Tiny...   
★★★★☆ 763



Bulk, Jeep Ducks for Ducking, Assorted Rubber Ducks Jeep Ducking, Bab...   
★★★★☆ 569



Ducks: Fun Unique Military-Inspired Bath Toys for Jeep Ducking or Play - 2 inches   
800+ viewed in past month



Bulk, Jeep Ducks for Ducking, Assorted Rubber Ducks for Jeeps, Bath Toy...   
★★★★☆ 569



Glitter Rubber Duck Toy Assortment Duckies for Kids, Bath Birthday Gifts...   
★★★★☆ 1,094

Some sites need interaction



# TEXAS DEPARTMENT OF LICENSING & REGULATION

## TDLR License Data Search (Active Licenses only)

[Search Help](#) | [Download License files](#) | [Download Other](#) | [Questions/Comments](#)

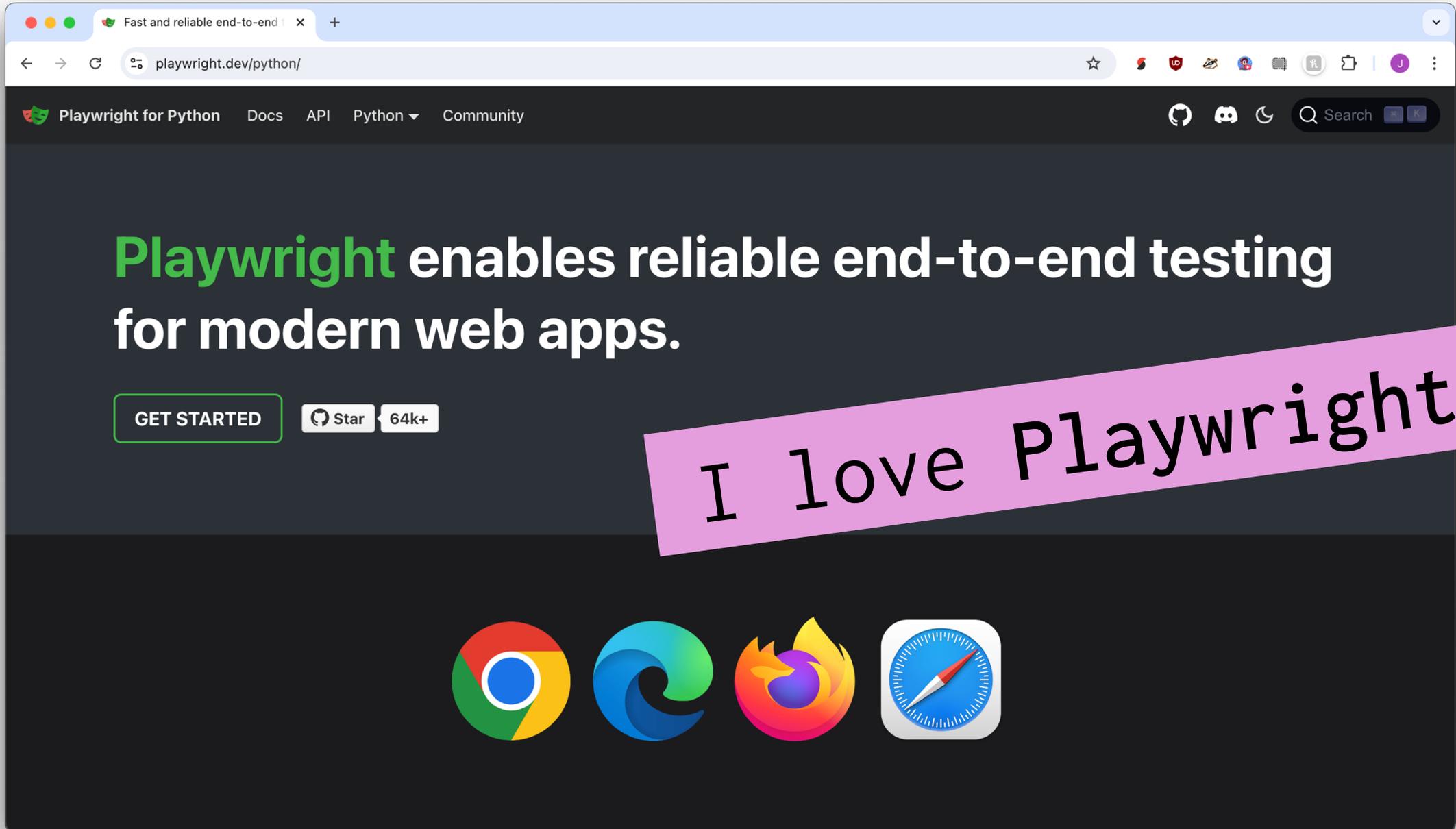
<b>Inquire by License Type</b>	<b>Inquire by License #</b>
Choose One (Optional) <input type="text"/>	<input type="text"/> (Numeric only)
<b>Inquire by Expiration Date</b>	
<input type="text"/> (mmddyyyy)	
<b>Inquire by Name (Last, First) or by Business Name</b>	
<input type="text"/>	
<b>Inquire by Location (City)</b>	
Choose One (Optional) <input type="text"/>	Type the first letter to scroll down.
<b>Inquire by County</b>	
Choose One (Optional) <input type="text"/>	Type the first letter to scroll down.
<b>Inquire by Zip Code</b>	
<input type="text"/>	
<input type="button" value="Search"/> <input type="button" value="Reset"/>	

If license not found, please contact Customer Service at 800-803-9202

Data last updated: 8/21/2024 06:01

[Bookmark This Page](#)

Some sites need interaction



**Playwright** enables reliable end-to-end testing for modern web apps.

GET STARTED

Star 64k+

I love Playwright!



Playwright is perfect!

*But!* It's new, so

ChatGPT isn't perfect

at it. *But!* We try.

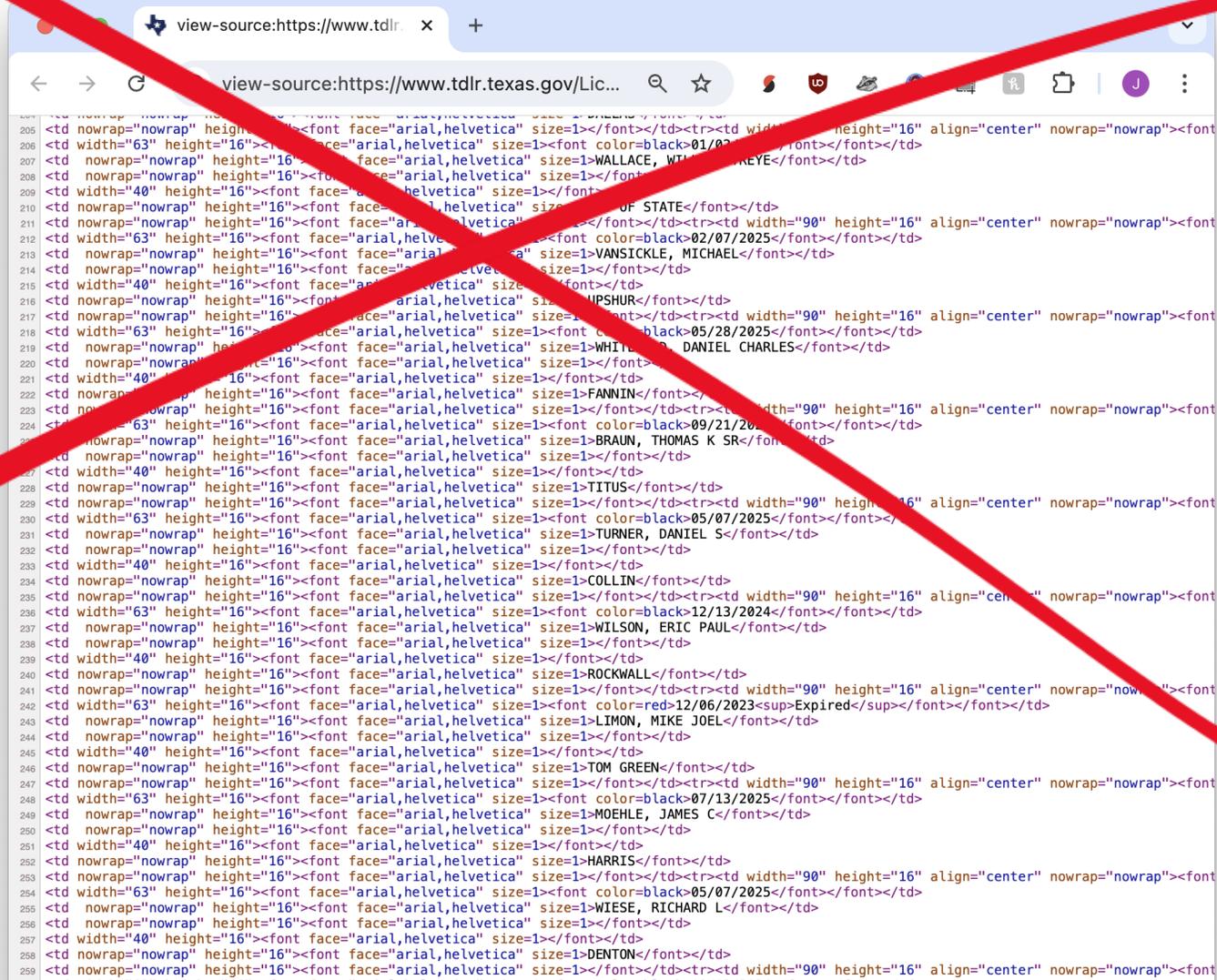
Playwright + ChatGPT +  
pasting samples of HTML

=

**infinite scrapers**

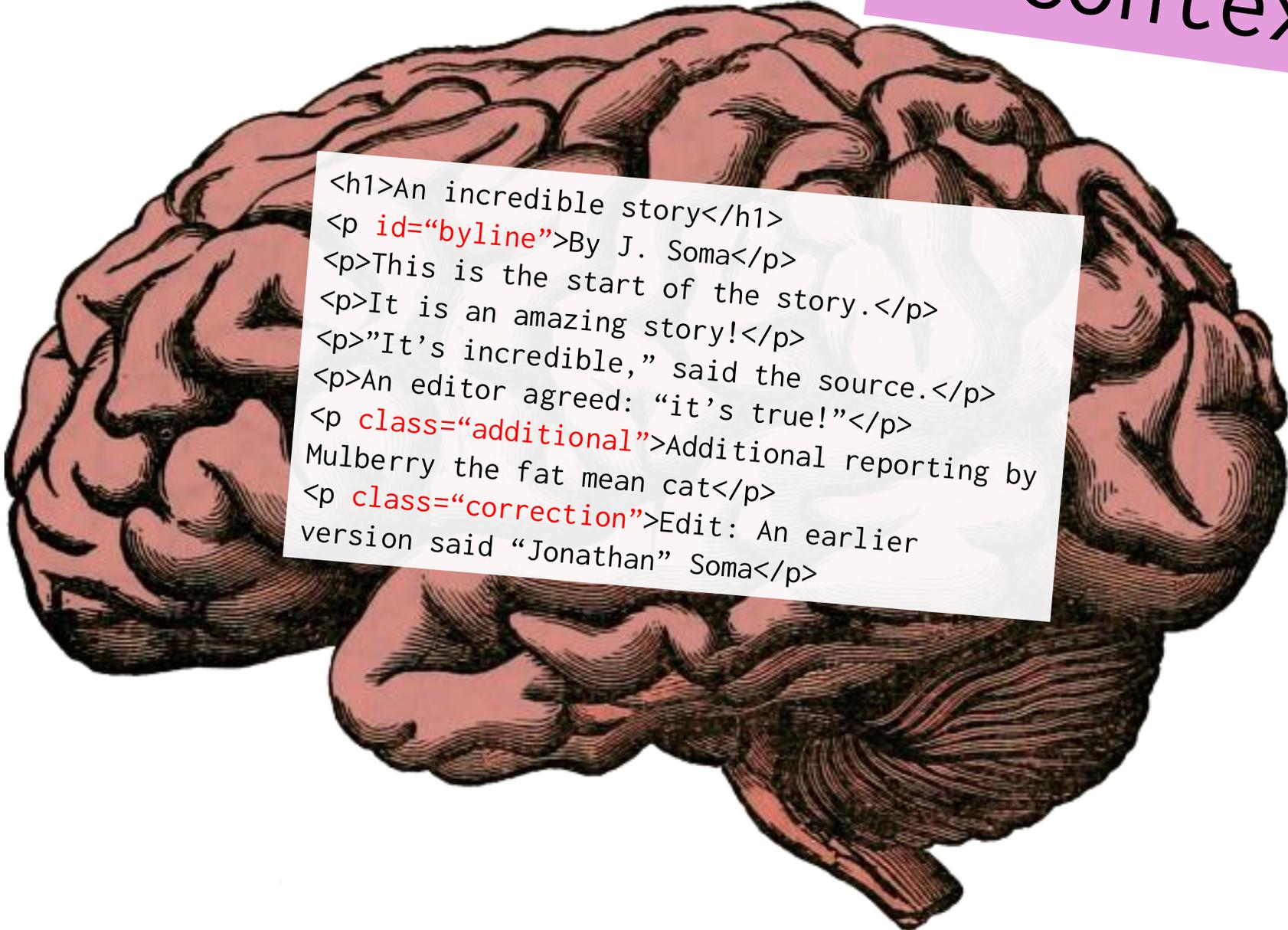


“Write a scraper for these search results:”



```
205 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1</font></td><tr><td width="90" height="16" align="center" nowrap="nowrap"><font
206 <td width="63" height="16"><font face="arial,helvetica" size=1</font color=black>01/03/2025</font></td>
207 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>WALLACE, WILSON KREYE</font></td>
208 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1</font>
209 <td width="40" height="16"><font face="arial,helvetica" size=1</font>
210 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>OF STATE</font></td>
211 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1</font></td><tr><td width="90" height="16" align="center" nowrap="nowrap"><font
212 <td width="63" height="16"><font face="arial,helvetica" size=1</font color=black>02/07/2025</font></td>
213 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>VANSICKLE, MICHAEL</font></td>
214 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1</font></td>
215 <td width="40" height="16"><font face="arial,helvetica" size=1</font></td>
216 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>MPSHUR</font></td>
217 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1</font></td><tr><td width="90" height="16" align="center" nowrap="nowrap"><font
218 <td width="63" height="16"><font face="arial,helvetica" size=1</font color=black>05/28/2025</font></td>
219 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>WHITFIELD, DANIEL CHARLES</font></td>
220 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1</font>
221 <td width="40" height="16"><font face="arial,helvetica" size=1</font></td>
222 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>FANNIN</font></td>
223 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1</font></td><tr><td width="90" height="16" align="center" nowrap="nowrap"><font
224 <td width="63" height="16"><font face="arial,helvetica" size=1</font color=black>09/21/2025</font></td>
225 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>BRAUN, THOMAS K SR</font></td>
226 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1</font></td>
227 <td width="40" height="16"><font face="arial,helvetica" size=1</font></td>
228 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>TITUS</font></td>
229 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1</font></td><tr><td width="90" height="16" align="center" nowrap="nowrap"><font
230 <td width="63" height="16"><font face="arial,helvetica" size=1</font color=black>05/07/2025</font></td>
231 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>TURNER, DANIEL S</font></td>
232 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1</font></td>
233 <td width="40" height="16"><font face="arial,helvetica" size=1</font></td>
234 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>COLLIN</font></td>
235 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1</font></td><tr><td width="90" height="16" align="center" nowrap="nowrap"><font
236 <td width="63" height="16"><font face="arial,helvetica" size=1</font color=black>12/13/2024</font></td>
237 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>WILSON, ERIC PAUL</font></td>
238 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1</font></td>
239 <td width="40" height="16"><font face="arial,helvetica" size=1</font></td>
240 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>ROCKWALL</font></td>
241 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1</font></td><tr><td width="90" height="16" align="center" nowrap="nowrap"><font
242 <td width="63" height="16"><font face="arial,helvetica" size=1</font color=red>12/06/2023<sup>Expired</sup></font></td>
243 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>LIMON, MIKE JOEL</font></td>
244 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1</font></td>
245 <td width="40" height="16"><font face="arial,helvetica" size=1</font></td>
246 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>TOM GREEN</font></td>
247 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1</font></td><tr><td width="90" height="16" align="center" nowrap="nowrap"><font
248 <td width="63" height="16"><font face="arial,helvetica" size=1</font color=black>07/13/2025</font></td>
249 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>MOEHLE, JAMES C</font></td>
250 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1</font></td>
251 <td width="40" height="16"><font face="arial,helvetica" size=1</font></td>
252 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>HARRIS</font></td>
253 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1</font></td><tr><td width="90" height="16" align="center" nowrap="nowrap"><font
254 <td width="63" height="16"><font face="arial,helvetica" size=1</font color=black>05/07/2025</font></td>
255 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>WIESE, RICHARD L</font></td>
256 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1</font></td>
257 <td width="40" height="16"><font face="arial,helvetica" size=1</font></td>
258 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>DENTON</font></td>
259 <td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1</font></td><tr><td width="90" height="16" align="center" nowrap="nowrap"><font
```

# “Context window”



```
<h1>An incredible story</h1>
<p id="byline">By J. Soma</p>
<p>This is the start of the story.</p>
<p>It is an amazing story!</p>
<p>"It's incredible," said the source.</p>
<p>An editor agreed: "it's true!"</p>
<p class="additional">Additional reporting by
Mulberry the fat mean cat</p>
<p class="correction">Edit: An earlier
version said "Jonathan" Soma</p>
```



## What we might need:

- Start URL
- Forms to fill out (*optional*)
- “Rows” of our spreadsheet
- Pagination/”next” pages (*optional*)
- A **magical prompt**

Playwright gets blocked 😞

Camoufox solves it,  
but ChatGPT and AI  
tools are very very  
very very very bad  
at it.

The screenshot shows the Camoufox website's introduction page. The browser's address bar displays 'camoufox.com'. The page features a navigation menu with links for GitHub, Issues, Discussions, PyPI, and Donate. A search bar is located in the top right corner. The main content area is titled 'Introduction' and includes a large banner with the Camoufox logo and the text 'An open source anti-detect browser with robust fingerprint injection. 🦊'. Below the banner is a badge indicating it is the '#3 Repository Of The Day' on GitHub Trending. A warning box states: 'Warning: Camoufox is in active development! Releases are available, but may not be suitable for production use.' The 'Features' section lists several capabilities, such as being invisible to anti-bot systems and having robust fingerprint injection without JS injection. A sidebar on the left contains a filter and a list of navigation items including Introduction, Features List, Stealth Performance, Python Interface, Installation, Usage, GeolP & Proxy Support, Main World Execution, Remote Server, Virtual Display, BrowserForge Integration, Passing Config, Fingerprint Injection, Development, Liability Disclaimer, and Donate. The footer mentions 'Powered by RETYPE'.