

Visit <https://bit.ly/dh2025-playwright> for material

## Scraping dynamic websites with Playwright!

stealing (??) data from  
the internet

# 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

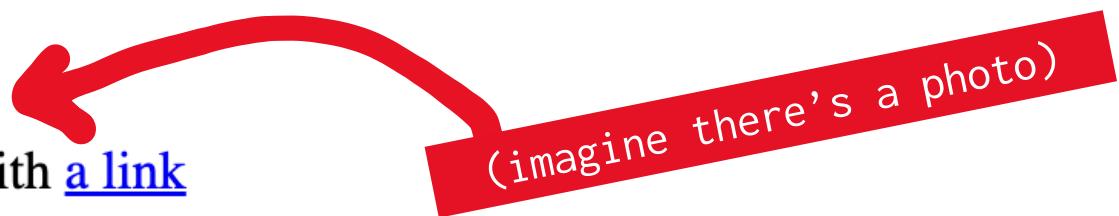
### And an even smaller headline

This is a paragraph

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

This is ANYTHING

Anything! Anything in the world



```
<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>
```

# 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*

	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 >

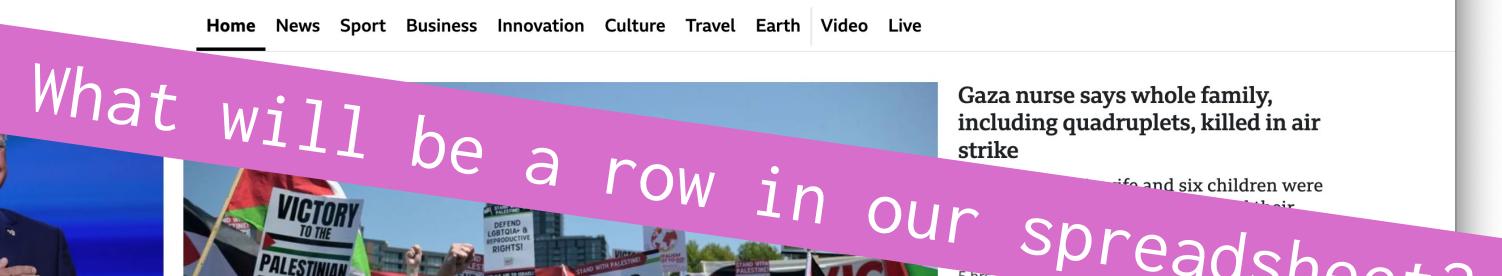
< p class="correction" > Edit: An earlier  
version said "Jonathan" Soma </ p >

id="byline"

class="additional"

class="correction"

Let's see it on  
the internet!



BBC

Home News Sport Business Innovation Culture Travel Earth Video Live

Register Sign In

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.

Divers find five bodies in wreck of Sicily yacht

• 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

Andrew Tate held overnight after police raid homes in Romania

Ancient ocean of magma found on Moon south pole

► German Navy blasts out Darth Vader theme on Thames

BBC Home - Breaking News | +

bbc.com

☰ 🔍 BBC 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



Divers find five bodies in wreck of sunken sail yacht



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.

7 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.

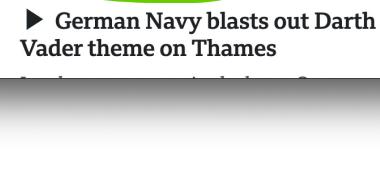
44 mins ago | Europe



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

*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.

The screenshot shows the Open Syllabus Analytics website interface. At the top, there's a navigation bar with links for "OS Home", "Explore", and "About". Below the navigation is a large pink banner with the text "What will be a row in our spreadsheet?". On the left, there's a sidebar titled "FILTER SYLLABI" with dropdown menus for "Enter a search keyword or a concept", "By Institution", "By Field", "By Country", "By State (US only)", "From" and "To" date fields, and "By School Type (US only)". The main content area is titled "3,183,312 Titles through 2020" and features a search bar labeled "Search Titles". A table lists the top 5 most frequent 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

Open Syllabus Analytics

analytics.opensyllabus.org/record/works

SVG Crowbar SVG Crowbar 2 Top Shot SVG Crowbar popup with tags Coding Viz gamedev bread lede elrn make All Bookmarks

# OPEN SYLLABUS ANALYTICS LITE

TITLES

SIGN UP / IN LEARN MORE

## 3,183,312 Titles through 2020

Search 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

Research Design

Enter a search keyword or a concept

By Institution

By Field

By Country

By State (US only)

From To

By School Type (US only)

What will be a column of data?

rank	2	title	The Elements of Style	appearances	15,020	score	100
		author	William Strunk				

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

HTML code

# Visit OpenSyllabus



Do this now

The screenshot shows the OpenSyllabus Analytics Lite homepage. At the top, there's a navigation bar with links for "OS Home", "Explore", "About", "SIGN UP / IN", and "LEARN MORE". Below the navigation is a large pink banner with the text "Visit OpenSyllabus". On the left, there's a sidebar with a "TITLES" dropdown menu and a "FILTER SYLLABI" section containing dropdowns for "By Dept", "By Field", "By Country", "By State (US only)", "From", "To", and "By School Type (US only)". To the right, a large box displays the text "3,183,312 Titles through 2020" above a search bar labeled "Search Titles". Below this is a table listing the top 5 most appearing 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

RANK	TITLES	APPEARANCES	SCOR
1	<b>Calculus</b> James Stewart	18,299	100
2	<b>The Elements</b> William Strunk	15,020	100
3	Social Research Bryman Oxford University	14,704	100
4	<b>Human Anatomy Manual</b> Elaine Nicpon M	14,101	100
5	<b>English Grammar</b> Raymond Murphy Cambridge Univ		

**Right click (or command+click) and Inspect**

The Elements

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**

Open Syllabus Analytics

# Find the HTML for our “row”

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

Human Anatomy and Physiology Laboratory Manual  
Elaine Nicpon Marieb

Mouse/click around

Elements Console Sources > 3

```
<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 OI0Rn"></p>
            </div>
          </li>
        </ul>
      </td>
    </tr>
  </tbody>
</table>
```

Styles Computed Layout Event Listeners DOM Breakpoints >

Filter

```
element.style {
}
.gsKnKH > a {
  width: fit-content;
}
a {
```

METHOD ONE

67ddd51f19edbb1.css:1

Open Syllabus Analytics

# Find the HTML for our “row”

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

Human Anatomy and Physiology Laboratory Manual  
Elaine Nicpon Marieb

Elements    Console    Sources    Click this

Click this

Search + click here

METHOD TWO

```
<thead></thead>
<tbody>
  <tr>
    <td>...</td>
    <td>...</td>
    <ul>
      <li>...</li>
      <li>...</li>
    </ul>
  </tr>
</tbody>
```

```
.duwfJU {
  width: 100%;
  display: flex;
  -webkit-box-pack: start;
  justify-content: flex-start;
}
```

Scraping is just  
connecting to a website  
and pulling data  
into columns in  
your spreadsheet

But we need some software  
to do that for us!

# What software?

The screenshot shows a search results page for "Scraping libraries". The top navigation bar includes a search bar with a microphone icon, a camera icon, and a magnifying glass icon, along with a user profile icon. Below the search bar are filters for "All", "Images", "Videos", "News", "Books", "Web", and "Finance". A "Tools" section is also present. The main content area displays nine scraping libraries, each with its logo, name, and a dropdown arrow:

- 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

with

Do

[–] NerdvanaNC 1 point 2 years ago

Look into learning and understanding HTML, then BeautifulSoup and Selenium.

sites

developers test their

Google for that. Oh and bonus tip, there's (official?)

(Docker!) which made me very happy (I like Docker!)

html\_data = response.text

[–] luisv4z 12 points 2 years ago

From my own research, run away from Selenium. The right direction is CDP (Chrome Developers

tool, I could scrap Facebook without getting banned

2 all\_links = soup.find\_all(name

Do python on them until

BeautifulSoup

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. selenium, unless

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

Page 1 of 3



Continental Grand Prix  
5000 Rennrad Faltreifen //  
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 Overnight by 8:00  
AM



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  
2-inch Duck Toys for Baby  
Shower Rubber Duckies,...  
 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

Page 1 of 7



Chockees Asso  
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



ire24-scraping/06-Maryland

TDLR Active License Search

tdlr.texas.gov/LicenseSearch/

**TEXAS DEPARTMENT OF LICENSING & REGULATION**

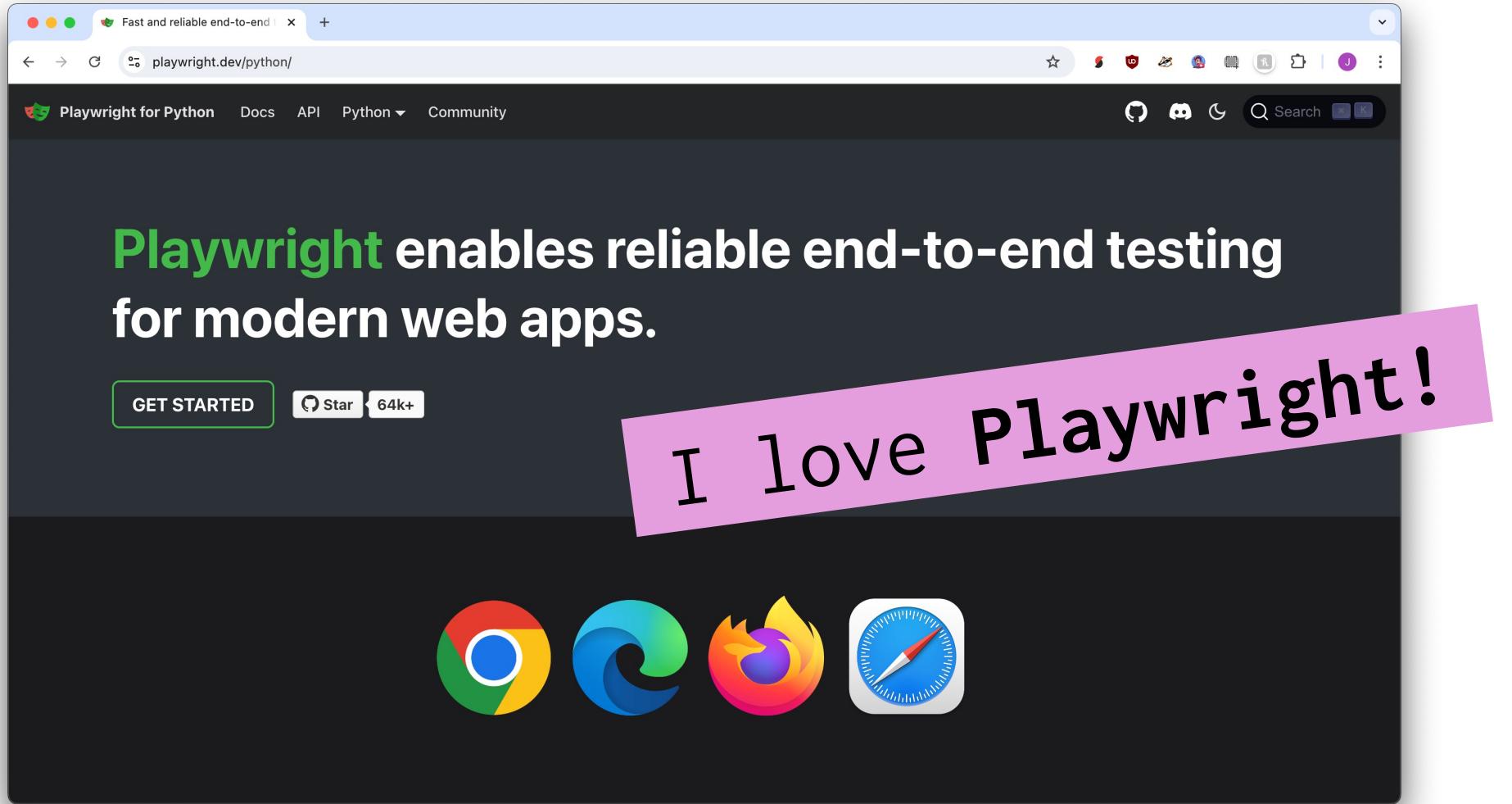
**TDLR License Data Search** (Active Licenses only)

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

<b>Inquire by License Type</b> <input type="text"/> Choose One (Optional)	<b>Inquire by License #</b> <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> <input type="text"/> Choose One (Optional) <input type="button"/> Type the first letter to scroll down.	
<b>Inquire by County</b> <input type="text"/> Choose One (Optional) <input type="button"/> Type the first letter to scroll down.	
<b>Inquire by Zip Code</b> <input type="text"/>	

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 is perfect!

*But!* It's new-ish, so  
ChatGPT isn't very good  
at it. *But!* We try.

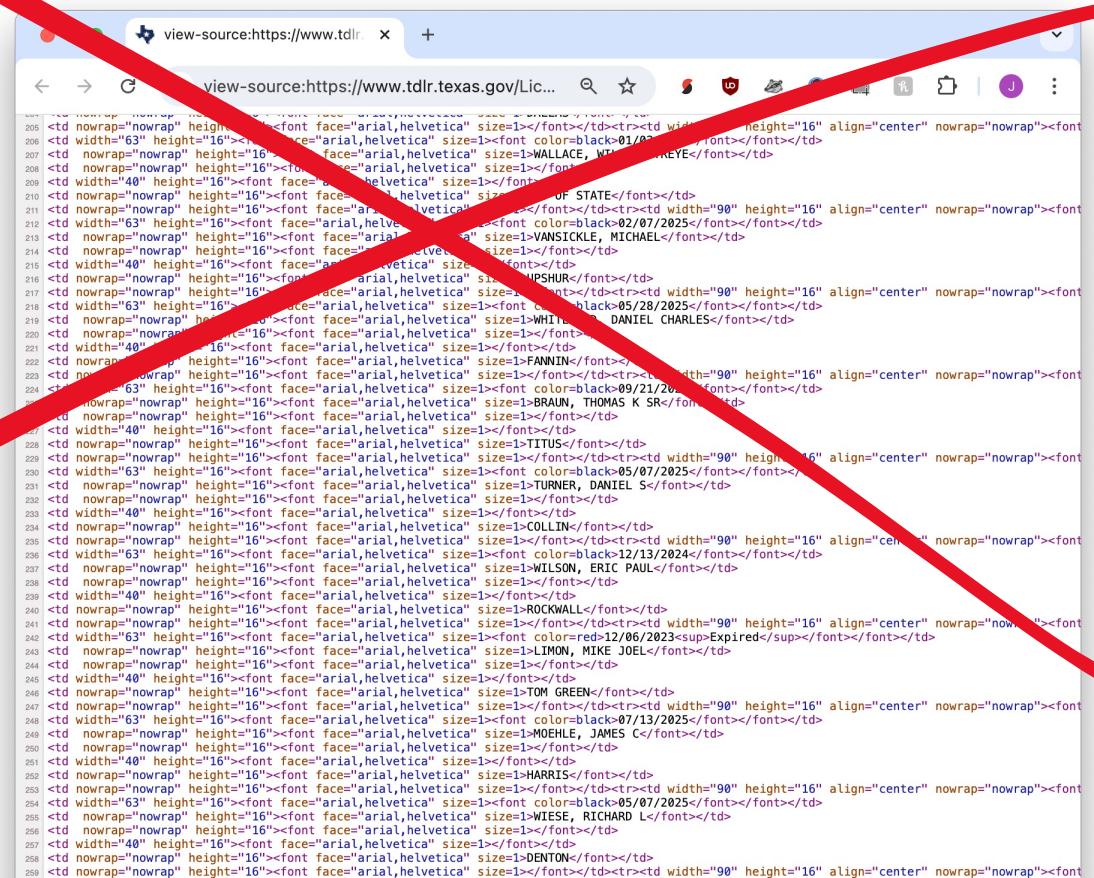
Playwright + ChatGPT +  
pasting samples of HTML

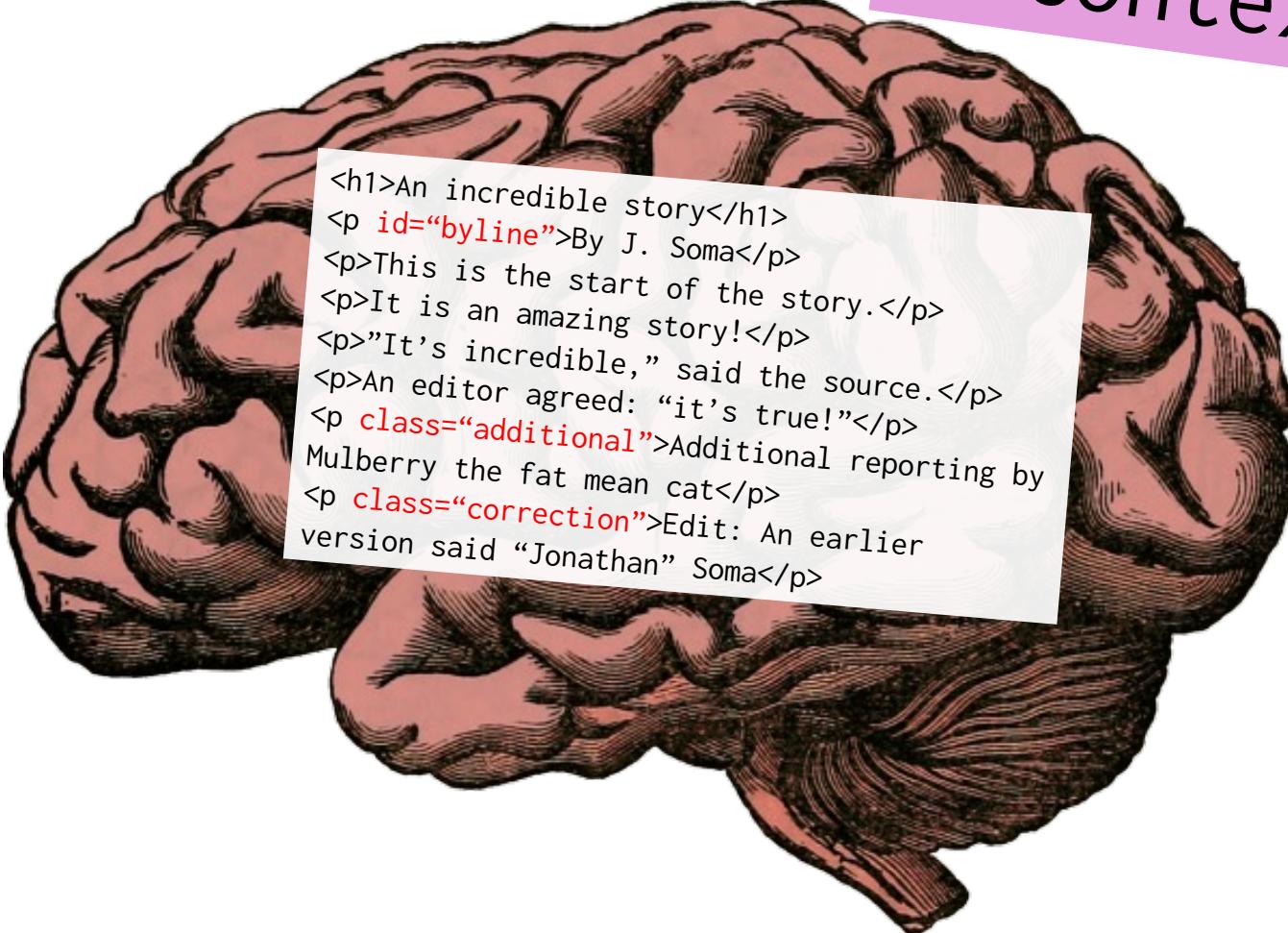
=

infinite scrapers



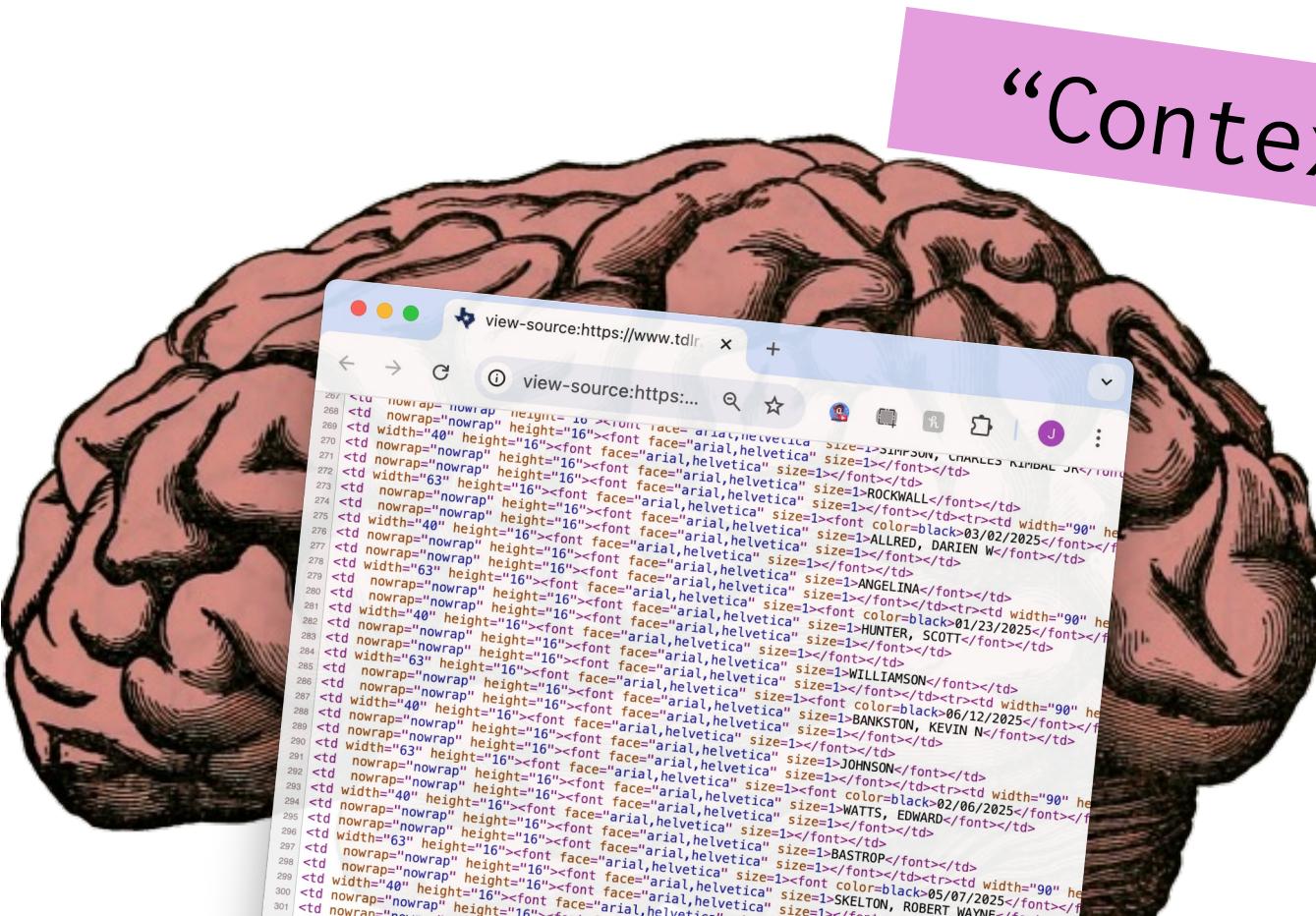
~~“Write a scraper for these search results:”~~





# “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>
```



A screenshot of a web browser window titled "view-source:https://www.td...". The page displays a large amount of HTML code, specifically a table structure with many rows and columns. The code is rendered in a monospaced font, showing various tags like <td>, <font>, and <tr>. The content of the table appears to be a list of names and dates, likely from a database or log file. The browser interface includes standard controls (back, forward, search, etc.) and a status bar at the bottom.

```
<tr>
<td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>ROBERTSON, CHARLES RIMBAL JR</font></td>
<td width="40" height="16"><font face="arial,helvetica" size=1></font></td>
<td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1>ROCKWALL</font></td>
<td width="63" height="16"><font face="arial,helvetica" size=1></font></td><tr><td width="90" he
<td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1></font></td>
<td width="40" height="16"><font face="arial,helvetica" size=1>ALLRED, DARIEN W</font></td>
<td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1></font></td>
<td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1></font></td>
<td width="63" height="16"><font face="arial,helvetica" size=1>ANGELINA</font></td>
<td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1></font></td>
<td width="63" height="16"><font face="arial,helvetica" size=1><font color=black>03/02/2025</font></font></td>
<td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1><font color=black>01/23/2025</font></font></td>
<td width="40" height="16"><font face="arial,helvetica" size=1>HUNTER, SCOTT</font></td>
<td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1></font></td>
<td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1></font></td>
<td width="63" height="16"><font face="arial,helvetica" size=1>WILLIAMSON</font></td>
<td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1></font></td>
<td width="63" height="16"><font face="arial,helvetica" size=1><font color=black>06/12/2025</font></font></td>
<td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1><font color=black>02/06/2025</font></font></td>
<td width="40" height="16"><font face="arial,helvetica" size=1></font></td>
<td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1></font></td>
<td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1></font></td>
<td width="63" height="16"><font face="arial,helvetica" size=1>JOHNSON</font></td>
<td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1></font></td>
<td width="63" height="16"><font face="arial,helvetica" size=1></font></td>
<td width="40" height="16"><font face="arial,helvetica" size=1><font color=black>02/06/2025</font></font></td>
<td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1></font></td>
<td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1></font></td>
<td width="63" height="16"><font face="arial,helvetica" size=1>BASTROP</font></td>
<td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1></font></td>
<td width="40" height="16"><font face="arial,helvetica" size=1><font color=black>05/07/2025</font></font></td>
<td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1></font></td>
<td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1><font color=black>05/21/2025</font></font></td>
<td width="63" height="16"><font face="arial,helvetica" size=1></font></td>
<td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1></font></td>
<td width="40" height="16"><font face="arial,helvetica" size=1><font color=black>05/21/2025</font></font></td>
<td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1></font></td>
<td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1></font></td>
<td width="63" height="16"><font face="arial,helvetica" size=1>BEVAR</font></td>
<td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1></font></td>
<td width="40" height="16"><font face="arial,helvetica" size=1><font color=black>05/21/2025</font></font></td>
<td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1></font></td>
<td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1><font color=black>05/21/2025</font></font></td>
<td width="63" height="16"><font face="arial,helvetica" size=1>HARRIS</font></td>
<td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1></font></td>
<td width="40" height="16"><font face="arial,helvetica" size=1><font color=black>05/07/2025</font></font></td>
<td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1></font></td>
<td width="63" height="16"><font face="arial,helvetica" size=1>ROBERTSON, MICHAEL JAY</font></td>
<td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1></font></td>
<td width="40" height="16"><font face="arial,helvetica" size=1></font></td>
<td nowrap="nowrap" height="16"><font face="arial,helvetica" size=1></font></td>
<td width="63" height="16"><font face="arial,helvetica" size=1>OUT OF COUNCIL</font></td>
```

# “Context window”

Visit <https://bit.ly/dh2025-playwright> for material

## What we might need:

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

Visit <https://bit.ly/dh2025-playwright> for material

The screenshot shows the JupyterLab interface. On the left, there's a sidebar with 'Start' and 'Recent sessions' sections. The 'Start' section includes options like 'New notebook...', 'New session...', 'Open...', and 'Connect...'. The 'Recent sessions' section lists several notebooks. In the center, a code cell in an untitled notebook is displayed, demonstrating how to use the playwright library to scrape data from a URL. A large pink banner at the bottom asks 'Open Jupyter? JupyterLab?'.

```
[1]: from playwright.async_api import async_playwright

# Start Playwright
playwright = await async_playwright().start()
browser = await playwright.chromium.launch(headless=False)
page = await browser.new_page()

# Navigate to the URL
await page.goto("https://www.tdlr.texas.gov/LicenseSearch/")

[2]: import pandas as pd

# Wait for the results to be visible after form submission
await page.wait_for_selector("table")

# Scrape table rows
rows = await page.query_selector_all("table tr")

data = []
for row in rows:
    cells = await row.query_selector_all("td")
    if len(cells) > 0:
        license_name = await cells[0].inner_text()
        expiration_date = await cells[1].inner_text()
        license_holder = await cells[2].inner_text()

    data.append({
        "license_name": license_name,
        "expiration_date": expiration_date,
        "license_holder": license_holder
    })

df = pd.DataFrame(data)
df.head()
```

Open Jupyter? JupyterLab?

amazon.com  
search results

Belgium CBE  
business search

items on  
Mercari

Sales at Lidl

# Now try it.

Anywhere!!!!

Wikipedia's  
Fictional Big  
Cats

List of amphibians  
from Germany's Red  
List Center

School Board Minutes  
from Grand Island Public  
Schools in Nebraska