

# FinTech @ IU Python Session #6 – Web Libraries

## *Prepared by*

---

**Cameron Nelson**

***Fund Equities Lead – FinTech @ IU***

<https://www.linkedin.com/in/cameron-j-nelson/>

**Gabriel Shores**

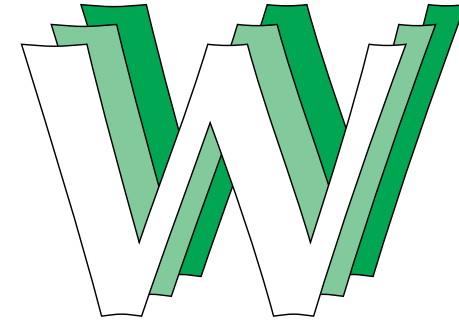
***Director of Technology – FinTech @ IU***

<https://www.linkedin.com/in/gabriel-shores-379b81291/>

# Why?

- Data could not be more important
- The internet has a significant amount of data, some that isn't neatly formatted for you (or the nearly formatted version is behind a paywall)
- The web is also really important, and knowing how to interact with it in Python will enable you to write more interesting programs

*Let's Share What We Know*



**World Wide Web**



# OpenWeather

# HTML & JSON

- Web requests are going to be answered in one of two formats (most likely): HTML or JSON
- HTML is a markup language (hence the ML) that is used to display the vast majority of web pages
  - Not neatly formatted. Significantly harder to parse since it is inconsistently defined & implemented
- JSON is basically a way to describe a dictionary or array through text
  - Easy to parse. For this reason, APIs will usually be in JSON

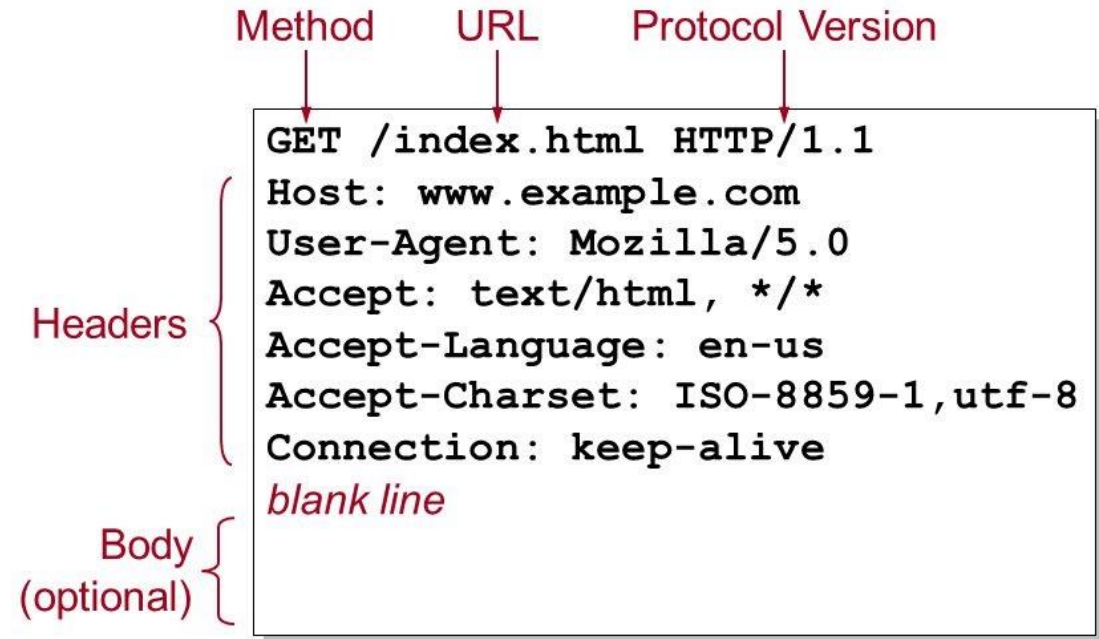
```
1 <!DOCTYPE html>
2 <html>
3   <head><title> All About Caves</title>
4   </head>
5   <body>
6     <h1>Welcome to my website about caves!</h1>
7     <!-- Navigation bar goes below -->
8     <a href="BlueSpring%20Caverns.html">BlueSpring Caverns</a>
9     <a href="Mammoth%20Cave.html">Mammoth Cave</a>
10    <a href="Luray%20Caverns.html">Luray Caverns</a>
11    <h3>How many facts do you know about caves? Did you know...</h3>
12    <h4> <ul>
13      <li>A cave (or cavern) is a naturally occurring area or space under the
14        surface of the Earth.</li>
15      <li>Stalagmites are created as the mineral rich water droplets fall to the
16        cave floor leaving mineral deposits that grow to become pillar-like formations.
17      </li>
18      <li>Stalactites are created by the continuous dripping of mineral rich water,
19        which little by little leaves mineral deposits on the ceiling as the minerals
20        harden. </li>
21      <li> The deepest known cave is Voronya Cave in Georgia, at 2,197 m (7,208 ft).
22    </li>
23    </ul></h4>
24    <h3> Use the navigation bar at the top to look at three of my favorite
25    caves.</h3>
26    <!-- Link back to main page -->
27    <a href="index.html"> Return to Home Page</a>
28  </body>
29 </html>
```

```
{
  "bands": [
    {
      "name": "Metallica",
      "description": "Metallica is an American heavy metal band. The band was formed in 1981 in Los Angeles by vocalist/guitarist James Hetfield and drummer Lars Ulrich, and has been based in San Francisco for most of its career.",
      "image": "metallica.jpg",
      "year_established": 1981,
      "albums": [
        {
          "name": "Kill 'Em All",
          "year": 1983
        },
        {
          "name": "Ride the Lightning",
          "year": 1984
        }
      ]
    },
    {
      "name": "U2",
      "description": "U2 are an Irish rock band from Dublin, formed in 1976. The group consists of Bono (lead vocals and rhythm guitar), the Edge (lead guitar, keyboards, and backing vocals), Adam Clayton (bass guitar), and Larry Mullen Jr. (drums and percussion).",
      "image": "u2.jpg",
      "year_established": 1976,
      "albums": [
        {
          "name": "Boy",
          "year": 1980
        },
        {
          "name": "October",
          "year": 1981
        }
      ]
    }
  ]
}
```

# request & urllib

- request is a simple library for making HTTP(S) requests
- Can be as simple as `requests.get(url)`
- Can be used to read either json or html
- urllib3 is a lower level library for more options or more specific requests.
- You can see by the code we'll look at that it is a little more complex than requests. But it gives you more flexibility over the requests you make!
- If you search for urllib, the first page is docs to a previous version. These are the correct docs:  
<https://urllib3.readthedocs.io/en/1.26.x/>

## HTTP Request



# Beautiful soup

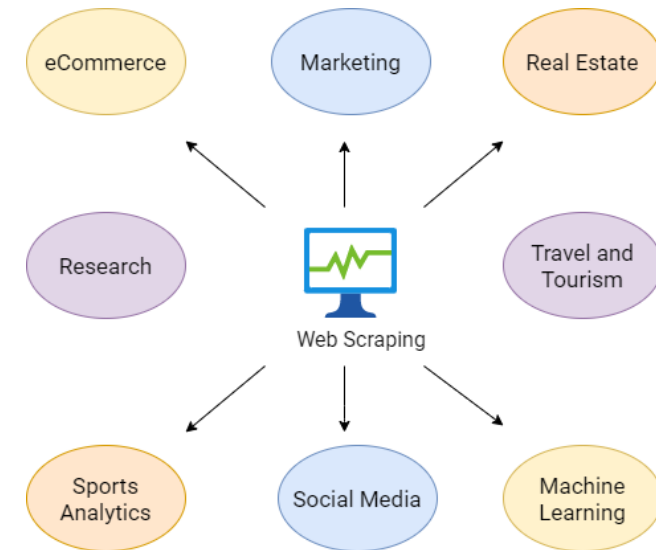
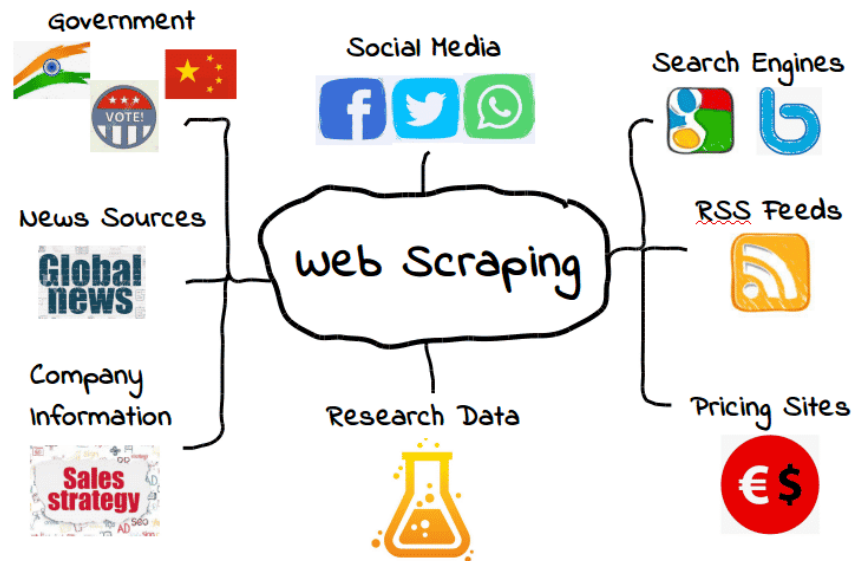
- BeautifulSoup, also called bs4, is a library to help parse html.
- You can either feed bs4 with an open file or a string containing html code
- Can be combined with request/urllib3 to parse pages on the internet
- Very complex library that is immensely useful for web scraping

```
from bs4 import BeautifulSoup
import requests

url = "https://example.com"
response = requests.get(url)
soup = BeautifulSoup(response.text, 'html.parser')
links = soup.select('.link')
for link in links:
    print(link.get('href'))
```

# Applications

- HTML requests are the basis of the modern web
- request and urllib are two useful libraries to get data from the Internet
- BS4 can help you parse poorly-formatted HTML data



# The End!

- Thank you!