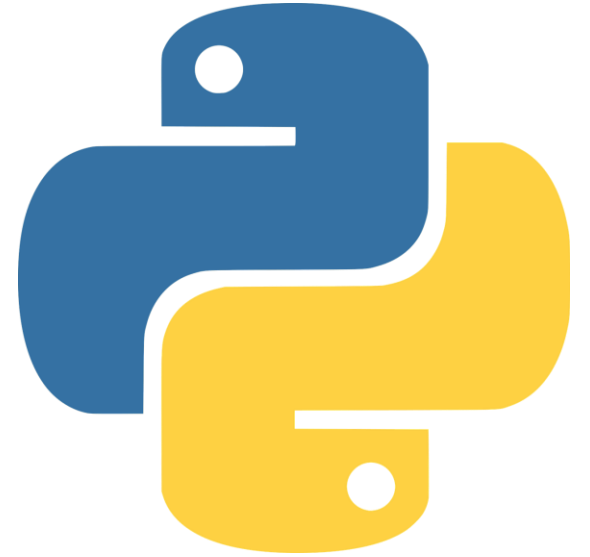


Web Scraping with Python Using BeautifulSoup

I'll be showing you how we can find the cheapest, in-stock, products on Newegg.

You'll be able to modify the code I show in this video to apply it to other websites and combine them together when searching for a product.



BeautifulSoup

Getting Website HTML

Continuing the “family tree” analogy, every tag and every string has a parent: the tag that contains it.

web_scraping4.py

```
from bs4 import BeautifulSoup
import requests
import re

search_term = input("What product do you want to search for? ") # ex: 3090

url = f"https://www.newegg.ca/p/pl?d={search_term}&N=4131"
page = requests.get(url).text
doc = BeautifulSoup(page, "html.parser")
```

Querying Multiple Pages

Finding count of the pages with specific product

web_scraping4.py

```
from bs4 import BeautifulSoup
import requests
import re

search_term = input("What product do you want to search for? ") # ex: 3080

url = f"https://www.newegg.ca/p/pl?d={search_term}&N=4131"
page = requests.get(url).text
doc = BeautifulSoup(page, "html.parser")

page_text = doc.find(class_="list-tool-pagination-text").strong
pages = int(str(page_text).split("/")[-2].split(">")[-1][: -1])
print(pages)
```

Finding Products

Printing list of products

web_scraping4.py

```
from bs4 import BeautifulSoup
import requests
import re

search_term = input("What product do you want to search for? ")

url = f"https://www.newegg.ca/p/pl?d={search_term}&N=4131"
page = requests.get(url).text
doc = BeautifulSoup(page, "html.parser")

page_text = doc.find(class_="list-tool-pagination-text").strong
pages = int(str(page_text).split("/")[-2].split(">")[-1][: -1])
print(pages)

for page in range(1, pages + 1):
    url = f"https://www.newegg.ca/p/pl?d={search_term}&N=4131&page={page}"
    page = requests.get(url).text
    doc = BeautifulSoup(page, "html.parser")

    div = doc.find(class_="item-cells-wrap border-cells items-grid-view four-cells expulsion-one-cell")
    items = div.find_all(text=re.compile(search_term))

    for item in items:
        print(item)
```

Finding Products

web_scraping4.py

Printing list of the products:

Name, Price and link

```
from bs4 import BeautifulSoup
import requests
import re

search_term = input("What product do you want to search for? ")

url = f"https://www.newegg.ca/p/pl?d={search_term}&N=4131"
page = requests.get(url).text
doc = BeautifulSoup(page, "html.parser")

page_text = doc.find(class_="list-tool-pagination-text").strong
pages = int(str(page_text).split("/")[-2].split(">")[-1][:1])
print(pages)

items_found = {}

for page in range(1, pages + 1):
    url = f"https://www.newegg.ca/p/pl?d={search_term}&N=4131&page={page}"
    page = requests.get(url).text
    doc = BeautifulSoup(page, "html.parser")

    div = doc.find(class_="item-cells-wrap border-cells items-grid-view four-cells expulsion-one-cell")
    items = div.find_all(text=re.compile(search_term))

    for item in items:
        parent = item.parent
        if parent.name != "a":
            continue

        link = parent['href']
        next_parent = item.find_parent(class_="item-container")
        try:
            price = next_parent.find(class_="price-current").find("strong").string
            items_found[item] = {"price": int(price.replace(",", "")), "link": link}
        except:
            pass

print(items_found)
```

Sorting Products By Price

```
from bs4 import BeautifulSoup
import requests
import re

search_term = input("What product do you want to search for? ")

url = f"https://www.newegg.ca/p/pl?d={search_term}&N=4131"
page = requests.get(url).text
doc = BeautifulSoup(page, "html.parser")
page_text = doc.find(class_="list-tool-pagination-text").strong
pages = int(str(page_text).split("/")[2].split(">")[-1][: -1])

items_found = {}

for page in range(1, pages + 1):
    url = f"https://www.newegg.ca/p/pl?d={search_term}&N=4131&page={page}"
    page = requests.get(url).text
    doc = BeautifulSoup(page, "html.parser")

    div = doc.find(class_="item-cells-wrap border-cells items-grid-view four-cells expulsion-one-cell")
    items = div.find_all(text=re.compile(search_term))

    for item in items:
        parent = item.parent
        if parent.name != "a":
            continue

        link = parent['href']
        next_parent = item.find_parent(class_="item-container")
        try:
            price = next_parent.find(class_="price-current").find("strong").string
            items_found[item] = {"price": int(price.replace(",", "")), "link": link}
        except:
            pass

sorted_items = sorted(items_found.items(), key=lambda x: x[1]['price'])

for item in sorted_items:
    print(item[0])
    print(f"${item[1]['price']}")
    print(item[1]['link'])
    print("-----")
```

web_scraping4.py

Sorting Products By Price

```
Run: web_scraming4 x
C:\Users\MarizzaMil\PycharmProjects\BeautifulSoup4\venv\Scripts\python.exe C:/Users/MarizzaMil/PycharmProjects/BeautifulSoup4/web_scraming4.py
What product do you want to search for? 3090
ASUS TUF Gaming GeForce RTX 3090 24GB GDDR6X PCI Express 4.0 SLI Support Video Card TUF-RTX3090-024G-GAMING
$3999
https://www.newegg.ca/asus-geforce-rtx-3090-tuf-rtx3090-o24g-gaming/p/N82E16814126454?Description=3090&cm\_re=3090\_-\_14-126-454\_-\_Product
-----
GIGABYTE GeForce RTX 3090 GAMING OC 24G Video Card, GV-N3090GAMING OC-24GD
$3999
https://www.newegg.ca/gigabyte-geforce-rtx-3090-qv-n3090gaming-oc-24gd/p/N82E16814932327?Description=3090&cm\_re=3090\_-\_14-932-327\_-\_Product
-----
EVGA GeForce RTX 3090 FTW3 ULTRA GAMING Video Card 24G-P5-3987-KR 24GB GDDR6X iCX3 Technology ARGB LED Metal Backplate and Intel Core i9-11900K
$4128
https://www.newegg.ca/Product/ComboDealDetails?ItemList=Combo.4422668
-----
GIGABYTE GeForce RTX 3090 EAGLE 24GB Video Card, GV-N3090EAGLE-24GD,NOT LHR
$4174
https://www.newegg.ca/p/1FT-000A-002E0?Description=3090&cm\_re=3090\_-\_9SIAPBWF056643\_-\_Product
-----
GIGABYTE GeForce RTX 3090 EAGLE OC 24GB Video Card, GV-N3090EAGLE OC-24GD
$4179
https://www.newegg.ca/p/1FT-000A-002E3?Description=3090&cm\_re=3090\_-\_9SIAPBWF056657\_-\_Product
-----
ASUS TUF Gaming GeForce RTX 3090 DirectX 12 TUF-RTX3090-024G-GAMING 24GB GDDR6X PCI Express 4.0 x16
$4199
https://www.newegg.ca/p/1DW-0027-001C9?Description=3090&cm\_re=3090\_-\_9SIA9U1F3X3771\_-\_Product
-----
```