

Building a Simple Web Scraper

TASK-3

Simple Data Manipulation with Pandas

Description:

The intern will create a basic web scraper using Python libraries like BeautifulSoup.

Responsibility:

1. Follow tutorials to understand web scraping principles and techniques.
2. Develop a simple Python script to scrape data from a static web page.
3. Extract relevant information such as text, links, or images from web pages.

```
In [1]: import requests  
from bs4 import BeautifulSoup
```

```
In [2]: # Send a GET request to the webpage  
url = "https://realpython.github.io/fake-jobs"  
response = requests.get(url)  
  
# Parse the HTML content using BeautifulSoup  
soup = BeautifulSoup(response.content, 'html.parser')
```

```
In [3]: # Extract the title of the webpage  
title = soup.find('title').text if soup.find('title') else 'No title found'  
print("Title:", title)
```

Title: Fake Python

```
In [4]: # Extract all links from the webpage  
links = []  
for link in soup.find_all('a'):br/>    links.append(link.get('href'))  
print("Links:")  
for link in links:  
    print(link)
```

```
https://www.realpython.com (https://www.realpython.com)  
https://realpython.github.io/fake-jobs/jobs/python-developer-50.html  
(https://realpython.github.io/fake-jobs/jobs/python-developer-50.html)  
https://www.realpython.com (https://www.realpython.com)  
https://realpython.github.io/fake-jobs/jobs/manufacturing-systems-engineer-51.html  
(https://realpython.github.io/fake-jobs/jobs/manufacturing-systems-engineer-51.html)  
https://www.realpython.com (https://www.realpython.com)  
https://realpython.github.io/fake-jobs/jobs/producer-television-film-video-52.html  
(https://realpython.github.io/fake-jobs/jobs/producer-television-film-video-52.html)  
https://www.realpython.com (https://www.realpython.com)  
https://realpython.github.io/fake-jobs/jobs/scientist-forensic-53.html  
(https://realpython.github.io/fake-jobs/jobs/scientist-forensic-53.html)  
https://www.realpython.com (https://www.realpython.com)  
https://realpython.github.io/fake-jobs/jobs/bonds-trader-54.html (http  
s://realpython.github.io/fake-jobs/jobs/bonds-trader-54.html)  
https://www.realpython.com (https://www.realpython.com)  
https://realpython.github.io/fake-jobs/jobs/editorial-assistant-55.htm  
l (https://realpython.github.io/fake-jobs/jobs/editorial-assistant-55.html)
```

```
In [5]: # Extract all images from the webpage
images = []
for img in soup.find_all('img'):
    images.append(img.get('src'))
print("Images:")
for img in images:
    print(img)
```

Images:


```
In [6]: # Extract text from the page
text = soup.get_text(separator=' ', strip=True)
print("Text:",text)
```

Text: Fake Python Fake Python Fake Jobs for Your Web Scraping Journey Senior Python Developer Payne, Roberts and Davis Stewartbury, AA 2021-04-08 Learn Apply Energy engineer Vasquez-Davidson Christopherville, AA 2021-04-08 Learn Apply Legal executive Jackson, Chambers and Levy Port Ericaburg h, AA 2021-04-08 Learn Apply Fitness centre manager Savage-Bradley East Seanview, AP 2021-04-08 Learn Apply Product manager Ramirez Inc North Jamievew, AP 2021-04-08 Learn Apply Medical technical officer Rogers-Yates Davidville, AP 2021-04-08 Learn Apply Physiological scientist Kramer-Klein South Christopher, AE 2021-04-08 Learn Apply Textile designer Meyers-John son Port Jonathan, AE 2021-04-08 Learn Apply Television floor manager Huges-Williams Osbornetown, AE 2021-04-08 Learn Apply Waste management officer Jones, Williams and Villa Scotttown, AP 2021-04-08 Learn Apply Software Engineer (Python) Garcia PLC Ericberg, AE 2021-04-08 Learn Apply Interpreter Gregory and Sons Ramireztown, AE 2021-04-08 Learn Apply Architect Clark, Garcia and Sosa Figueroaview, AA 2021-04-08 Learn Apply Meteorologist Bush PLC Kelseystad, AA 2021-04-08 Learn Apply Audiological scientist Salazar-Meyers Williamsburgh, AE 2021-04-08 Learn Apply English as a second language teacher Parker, Murphy and Brooks Mitchellburgh, AE 2021-04-08 Learn Apply Surgeon Cruz-Brown West Jessicabury, AA 2021-04-08 Learn Apply Equities trader Macdonald-Ferguson Maloneshire, AE 2021-04-08 Learn Apply Newspaper journalist Williams, Peterson and Rojas Johnston, AA 2021-04-08 Learn Apply Materials engineer Smith and Sons South Davidtown, AP 2021-04-08 Learn Apply Python Programmer (Entry-Level) Moss, Duncan and Allen Port Sara, AE 2021-04-08 Learn Apply Product/process development scientist Gomez-Carroll Marktown, AA 2021-04-08 Learn Apply Scientist, research (maths) Manning, Welch and Herring Laurenland, AE 2021-04-08 Learn Apply Ecologist Lee, Gutierrez and Brown Lauraton, AP 2021-04-08 Learn Apply Materials engineer Davis, Serrano and Cook South Tammyberg, AP 2021-04-08 Learn Apply Historic buildings inspector/conservation officer Smith LLC North Brandonville, AP 2021-04-08 Learn Apply Data scientist Thomas Group Port Robertfurt, AA 2021-04-08 Learn Apply Psychiatrist Silva-King Burnettbury, AE 2021-04-08 Learn Apply Structural engineer Pierce-Long Herberstide, AA 2021-04-08 Learn Apply Immigration officer Walker-Simpson Christopherport, AP 2021-04-08 Learn Apply Python Programmer (Entry-Level) Cooper and Sons West Victor, AE 2021-04-08 Learn Apply Neurosurgeon Donovan, Gonzalez and Figueroa Port Aaron, AP 2021-04-08 Learn Apply Broadcast engineer Morgan, Butler and Bennett Loribury, AA 2021-04-08 Learn Apply Make Snyder-Lee Angelastad, AP 2021-04-08 Learn Apply Nurse, adult Harris PLC Larrytown, AE 2021-04-08 Learn Apply Air broker Washington PLC West Colvin, AP 2021-04-08 Learn Apply Editor, film/video Brown, Price and Campbell West Stephanie, AP 2021-04-08 Learn Apply Production assistant, radio McGee PLC Laurentown, AP 2021-04-08 Learn Apply Engineer, communications Dixon Inc Wrightberg, AP 2021-04-08 Learn Apply Sales executive Thompson, Sheppard and Ward Alberttown, AE 2021-04-08 Learn Apply Software Developer (Python) Adams-Brewer Brockburgh, AE 2021-04-08 Learn Apply Futures trader Schneider-Brady North Jason, AE 2021-04-08 Learn Apply Tour manager Gonzales-Frank Arnoldhaven, AE 2021-04-08 Learn Apply Cytogeneticist Smith-Wong Lake Destiny, AP 2021-04-08 Learn Apply Designer, multimedia Pierce-Herrera South Timothyburgh, AP 2021-04-08 Learn Apply Trade union research officer Aguilar, Rivera and Quinn New Jimmyton, AE 2021-04-08 Learn Apply Chemist, analytical Lowe, Barnes and Thomas New Lucasbury, AP 2021-04-08 Learn Apply Programmer, multimedia Lewis, Gonzalez and Vasquez Port Cory, AE 2021-04-08 Learn Apply Engineer, broadcasting (operations) Taylor PLC Gileston, AA 2021-04-08 Learn Apply Teacher, primary school Oliver, Jones and Ramirez Cindyshire, AA 2021-04-08 Learn Apply Python Developer Rivera and Sons East Michaelfort, AA 2021-04-08 Learn Apply Manufacturing systems engineer Garcia PLC Joybury, AE 2021-04-08 Learn Apply Producer, television/film/video Johnson, Wells and Kramer Emmatown, AE 2021-04-08 Learn Apply Scientist, forensic Gonzalez LLC Colehaven, AP 2021-04-08 Learn Apply Bonds trader Morgan, White and Macdonald Port Coryton, AE 2021-04-08 Learn Apply Editorial assistant Robinson-Fitzpatrick Amyborough, AA 2

021-04-08 Learn Apply Photographer Waters, Wilson and Hoover Reynoldsville
e, AA 2021-04-08 Learn Apply Retail banker Hill LLC Port Billy, AP 2021-0
4-08 Learn Apply Jewellery designer Li-Gregory Adamburgh, AA 2021-04-08 L
earn Apply Ophthalmologist Fisher, Ryan and Coleman Wilsonmouth, AA 2021-
04-08 Learn Apply Back-End Web Developer (Python, Django) Stewart-Alexand
er South Kimberly, AA 2021-04-08 Learn Apply Licensed conveyancer Abbott
and Sons Benjaminland, AP 2021-04-08 Learn Apply Futures trader Bryant, S
antana and Davenport Zacharyport, AA 2021-04-08 Learn Apply Counselling p
sychologist Smith PLC Port Devonville, AE 2021-04-08 Learn Apply Insuranc
e underwriter Patterson-Singh East Thomas, AE 2021-04-08 Learn Apply Engi
neer, automotive Martinez-Berry New Jeffrey, AP 2021-04-08 Learn Apply Pr
oducer, radio May, Taylor and Fisher Davidside, AA 2021-04-08 Learn Apply
Dispensing optician Bailey, Owen and Thompson Jamesville, AA 2021-04-08 L
earn Apply Designer, fashion/clothing Vasquez Ltd New Kelly, AP 2021-04-0
8 Learn Apply Chartered loss adjuster Leblanc LLC Lake Antonio, AA 2021-0
4-08 Learn Apply Back-End Web Developer (Python, Django) Jackson, Ali and
McKee New Elizabethside, AA 2021-04-08 Learn Apply Forest/woodland manag
er Blankenship, Knight and Powell Millsbury, AE 2021-04-08 Learn Apply Cli
nical cytogeneticist Patton, Haynes and Jones Lloydton, AP 2021-04-08 Lea
rn Apply Print production planner Wood Inc Port Jeremy, AA 2021-04-08 Lea
rn Apply Systems developer Collins Group New Elizabethtown, AA 2021-04-08
Learn Apply Graphic designer Flores-Nelson Charlesstad, AE 2021-04-08 Lea
rn Apply Writer Mitchell, Jones and Olson Josephbury, AE 2021-04-08 Learn
Apply Field seismologist Howard Group Seanfurt, AA 2021-04-08 Learn Apply
Chief Strategy Officer Kramer-Edwards Williamburg, AA 2021-04-08 Learn Ap
ply Air cabin crew Berry-Houston South Jorgeside, AP 2021-04-08 Learn App
ly Python Programmer (Entry-Level) Mathews Inc Robertborough, AP 2021-04-
08 Learn Apply Warden/ranger Riley-Johnson South Saratown, AP 2021-04-08
Learn Apply Sports therapist Spencer and Sons Hullview, AA 2021-04-08 Lea
rn Apply Arts development officer Camacho-Sanchez Philipland, AP 2021-04-
08 Learn Apply Printmaker Oliver and Sons North Patty, AE 2021-04-08 Lear
n Apply Health and safety adviser Eaton PLC North Stephen, AE 2021-04-08
Learn Apply Manufacturing systems engineer Stanley-Frederick Stevensland,
AP 2021-04-08 Learn Apply Programmer, applications Bradley LLC Reysesstad,
AE 2021-04-08 Learn Apply Medical physicist Parker, Goodwin and Zavala Be
llberg, AP 2021-04-08 Learn Apply Media planner Kim-Miles North Johnland,
AE 2021-04-08 Learn Apply Software Developer (Python) Moreno-Rodriguez Ma
rtinezburgh, AE 2021-04-08 Learn Apply Surveyor, land/geomatics Brown-Ort
iz Joshuatown, AE 2021-04-08 Learn Apply Legal executive Hartman PLC West
Ericstad, AA 2021-04-08 Learn Apply Librarian, academic Brooks Inc Tucker
town, AE 2021-04-08 Learn Apply Barrister Washington-Castillo Perezton, A
E 2021-04-08 Learn Apply Museum/gallery exhibitions officer Nguyen, Yoder
and Petty Lake Abigail, AE 2021-04-08 Learn Apply Radiographer, diagnosi
c Holder LLC Jacobshire, AP 2021-04-08 Learn Apply Database administrator
Yates-Ferguson Port Susan, AE 2021-04-08 Learn Apply Furniture designer O
rtega-Lawrence North Tiffany, AA 2021-04-08 Learn Apply Ship broker Fuent
es, Walls and Castro Michelleville, AP 2021-04-08 Learn Apply

In []: