scdtask6

October 4, 2024

[ ]:
```
# pip install requests beautifulsoup4 pandas
# install the Libraries
```

[ ]:
```
import requests
```

[ ]:
```
from bs4 import BeautifulSoup
```

[ ]:
```
response = requests.get("https://uoc.edu.pk/electronicseng.html")
```

[ ]:
```
print(response)
```

```
<Response [200]>
```

[ ]:
```
soup = BeautifulSoup(response.content)
```

[ ]:
```
print (soup)
```

[ ]:
```
#To Download the Images

#url of page
url = 'https://uoc.edu.pk/electronicseng.html'

# Get request to web page
response = requests.get(url)
soup = BeautifulSoup(response.text, 'html.parser')

# my local dir path to save
local_folder = r'C:\Users\Zaibi\Desktop\Task06'
if not os.path.exists(local_folder):
    os.makedirs(local_folder)

# check images folder to save images
image_folder = os.path.join(local_folder, 'faculty_images')
if not os.path.exists(image_folder):
    os.makedirs(image_folder)

# Find all image tags
```

```
images = soup.find_all('img')

# loop all images
for img in images:
    img_src = img.get('src')  # get image URL
    if img_src and 'faculties/electronics' in img_src:
        # Form the full URL for the image
        full_img_url = 'https://uoc.edu.pk/' + img_src

        # get image name from URL
        img_name = os.path.basename(img_src)

        # Save images in faculty_images folder
        img_path = os.path.join(image_folder, img_name)

        # Download and save the image
        img_data = requests.get(full_img_url).content
        with open(img_path, 'wb') as img_file:
            img_file.write(img_data)

        print(f"Downloaded: {img_name}")

print("All images were downloaded")
```

```
Downloaded: mohsin.jpg
Downloaded: asim_quddus.jpg
Downloaded: tariq.jpg
Downloaded: qasim_ali.jpg
Downloaded: Engr. Safia Bibi.jpg
Downloaded: Engr.Athar Baig.jpg
Downloaded: Engr. illahi jan shah.jpeg
Downloaded: Engr.Aneeqa Fakhar.jpg
Downloaded: Engr. Qamar Ud din.jpeg
All images were downloaded
```

[ ]:
```
# Just save the Data of Faculty

# create lists to store data
names = []
designations = []
emails = []

# find all faculty sections
faculty_sections = soup.find_all('div', class_='tg-facultyname')

#for loop
for section in faculty_sections:
```

```
    # Scrape Name
    name_tag = section.find('h4')  # Find the <h4> tag
    if name_tag:
        name_b_tag = name_tag.find('b')  # Find <b> inside <h4>
        name = name_b_tag.text.strip()


    # Scrape Designation
    designation_tag = section.find_next('p')  # Find the <p> tag in the next␣
 ↪<div>
    if designation_tag:
        designation_b_tag = designation_tag.find('b')  # Find <b> inside <p>
        designation = designation_b_tag.text.strip()


    # Scrape Email
    email_tag = section.find_next('span')  # Find the next <span> tag (email)
    email = email_tag.text.strip()

    # Append the data to lists
    names.append(name)
    designations.append(designation)
    emails.append(email)

# Save scraped data into a CSV file
faculty_data = pd.DataFrame({
    'Name': names,
    'Designation': designations,
    'Email': emails
})

# Saving to the local folder
csv_file_path = os.path.join(local_folder, 'faculty_data.csv')
faculty_data.to_csv(csv_file_path, index=False)

print("CSV File Saved in PC")
```

```
CSV File Saved in PC
```

[ ]:
```
# extarct University Name
university_name = soup.find('title').text.split('-')[0].strip()  # get the␣
 ↪title title

# path to save in local pc
university_info_path = r'C:\Users\Zaibi\Desktop\Task06\university_info.txt'

# write in a file txt
```

```
with open(university_info_path, 'w') as f:
    f.write(f"University Name: {university_name}\n")
    f.write(f"URL of the Faculty Webpage: {url}\n")

print("University Info File Saved in PC")
```

```
University Info File Saved in PC
```

[ ]: