22p-9318_Ahmed_BS(SE)-5B

October 14, 2024

[52]:
```
import requests
from bs4 import BeautifulSoup
import os

# University and faculty page URLs
university_name = "University of Management and Technology (UMT)"
faculty_url = "https://www.umt.edu.pk/faculty.aspx"

# Save University Info to a text file
with open('university_info.txt', 'w') as f:
    f.write(f"University Name: {university_name}\n")
    f.write(f"Faculty Webpage: {faculty_url}\n")

# Fetching faculty page
response = requests.get(faculty_url)
soup = BeautifulSoup(response.text, 'html.parser')

# List to store faculty data
faculty_info = []

# Parse faculty information
faculty_rows = soup.find_all('tr')

for row in faculty_rows:
    # Extracting the name, designation, email, etc.
    name = row.find('td').get_text(strip=True) if row.find('td') else 'N/A'
    designation = row.find('td', class_='job-description').get_text(strip=True)␣
 ↪if row.find('td', class_='job-description') else 'N/A'
    email = row.find('a', class_='person-email')['href'].replace('mailto:', '')␣
 ↪if row.find('a', class_='person-email') else 'N/A'
    contact = row.find('td', class_='person-contact').get_text(strip=True) if␣
 ↪row.find('td', class_='person-contact') else 'N/A'

    # Image URL
    img_url = row.find('img')['src'] if row.find('img') else None

    # Append to list
```

```
    faculty_info.append({
        'Name': name,
        'Designation': designation,
        'Email': email,
        'Contact': contact,
        'Image URL': img_url
    })

# Save the extracted data into a CSV file
with open('faculty_data.csv', 'w') as file:
    # Write the headers
    file.write("Name,Designation,Email,Contact\n")

    # Write each faculty's data (excluding the image URL)
    for faculty in faculty_info:
        file.
 ↪write(f"{faculty['Name']},{faculty['Designation']},{faculty['Email']},{faculty['Contact']}\n"

# Creating folder for saving images
if not os.path.exists('faculty_images'):
    os.makedirs('faculty_images')

# Downloading Faculty Images and save them in the 'faculty_images' folder
for idx, faculty in enumerate(faculty_info):
    if faculty['Image URL']:
        img_data = requests.get(faculty['Image URL']).content
        with open(f'faculty_images/faculty_image_{idx + 1}.jpg', 'wb') as␣
 ↪img_file:
            img_file.write(img_data)

print("Scraping completed. Files have been saved.")
```

```
Scraping completed. Files have been saved.
```

[ ]: