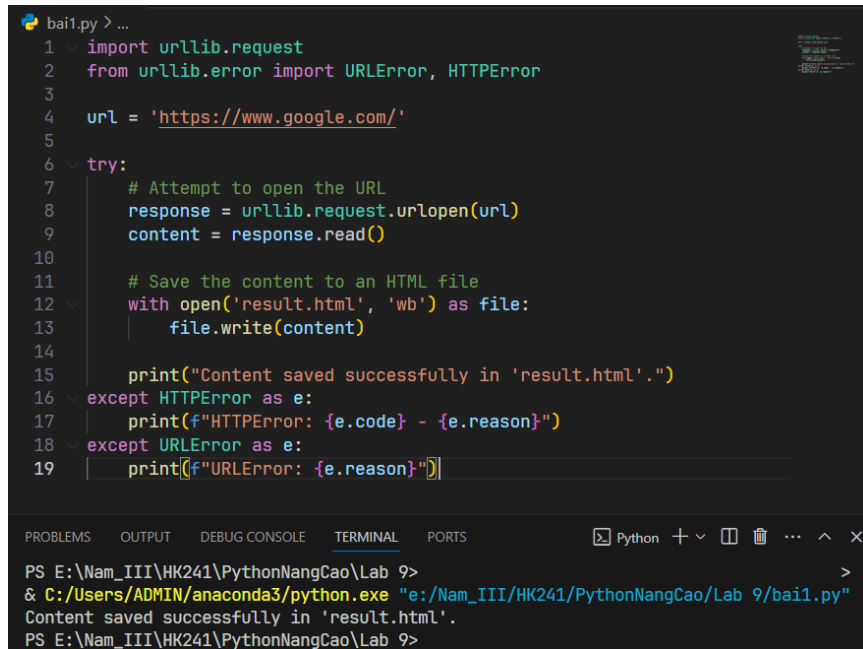


Bài Tập Thực Hành Lab 9

Bài 1: Sinh viên hãy hoàn thành thí dụ 1 bằng cách bổ sung vào đoạn chương trình để lưu kết quả hiện thị trong tập tin có định dạng html, có xử lý ngoại lệ URLError và HTTPError, và hãy nhận xét kết quả nhìn thấy được.



```
bai1.py > ...
1 import urllib.request
2 from urllib.error import URLError, HTTPError
3
4 url = 'https://www.google.com/'
5
6 try:
7     # Attempt to open the URL
8     response = urllib.request.urlopen(url)
9     content = response.read()
10
11     # Save the content to an HTML file
12     with open('result.html', 'wb') as file:
13         file.write(content)
14
15     print("Content saved successfully in 'result.html'.")
16 except HTTPError as e:
17     print(f"HTTPError: {e.code} - {e.reason}")
18 except URLError as e:
19     print(f"URLError: {e.reason}")
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS Python + - [] [] ... ^ x

PS E:\Nam_III\HK241\PythonNangCao\Lab 9>
& C:/Users/ADMIN/anaconda3/python.exe "e:/Nam_III/PythonNangCao/Lab 9/bai1.py"
Content saved successfully in 'result.html'.
PS E:\Nam_III\HK241\PythonNangCao\Lab 9>

Bài 2: Sinh viên hãy cho biết kết quả của câu lệnh sau đây, sinh viên có nhận xét gì về kết quả này

Mở URL: <https://www.google.com/search?q=test>

⇒ **HTTPError 403 (Forbidden):** Google có thể chặn yêu cầu này và trả về mã lỗi 403 vì thiếu User-Agent.

Bài 3: Sinh viên hãy quan sát hai cách truy vấn dữ liệu từ trang <http://python.org> dưới đây, so sánh và nhận xét

Tiêu chí	Cách 1	Cách 2
Cách thực hiện	Sử dụng urlopen() và đọc nội dung trực tiếp từ đối tượng trả về	Sử dụng with để mở URL và tự động đóng kết nối sau khi đọc nội dung
Quản lý tài nguyên	Yêu cầu đóng kết nối thủ công sau khi đọc xong nội dung	Tự động đóng kết nối khi thoát khỏi khối with

Tính an toàn	Có nguy cơ rò rỉ cao nếu quên đóng kết nối	An toàn hơn vì kết nối được tự động đóng
--------------	--	--

⇒ **Cách 2** được khuyến khích hơn vì quản lý tài nguyên tốt hơn, đảm bảo tính an toàn và giảm nguy cơ lỗi khi làm việc với các kết nối mạng.

Bài 4: Đoạn code dưới đây dùng để lưu trữ ảnh (binary_framing2.png) lấy từ địa chỉ https://developer.mozilla.org/en-US/docs/Web/HTTP/Messages/binary_framing2.png, sinh viên hãy phát triển đoạn script này để có thể lưu hết toàn bộ các ảnh tại địa chỉ <https://developer.mozilla.org/en-US/docs/Web/HTTP/Messages/>.

```

1 import urllib.request
2 import re
3 import os
4
5 # URL của trang web cần phân tích
6 url = "https://developer.mozilla.org/en-US/docs/Web/HTTP/Messages/"
7
8 # Tạo thư mục "images" nếu chưa tồn tại
9 if not os.path.exists("images"):
10     os.makedirs("images")
11
12 # Tải nội dung HTML của trang
13 response = urllib.request.urlopen(url)
14 html = response.read().decode("utf-8")
15
16 # Sử dụng regular expression để tìm tất cả các đường dẫn ảnh trong trang
17 image_urls = re.findall(r'src="([^\"]+\.png)"', html)
18
19 # Lặp qua từng URL hình ảnh và tải về
20 for image_url in image_urls:
21     # Nếu đường dẫn không chứa URL đầy đủ, bổ sung phần URL chính
22     if not image_url.startswith("http"):
23         image_url = urllib.parse.urljoin(url, image_url)
24
25     # Lấy tên file từ URL
26     file_name = image_url.split("/")[-1]
27
28     # Đường dẫn lưu file trong thư mục "images"
29     save_path = os.path.join("images", file_name)
30
31     # Tải và lưu ảnh
32     urllib.request.urlretrieve(image_url, save_path)
33     print(f"Tải thành công {file_name} vào thư mục 'images'")
34
35 print("Hoàn thành tải tất cả các ảnh.")

```

```

PS E:\Nam_III\HK241\PythonNangCao\Lab 9> & C:/Users/ADMIN/AppData\Local\Microsoft\Windows\InetRes\
Tải thành công httpmsg2.png vào thư mục 'images'
Tải thành công httpmsgstructure2.png vào thư mục 'images'
Tải thành công http_request_headers3.png vào thư mục 'images'
Tải thành công http_response_headers3.png vào thư mục 'images'
Tải thành công binary_framing2.png vào thư mục 'images'
Hoàn thành tải tất cả các ảnh.

```

Bài 5: Đoạn chương trình dưới đây đọc dữ liệu từ địa chỉ 'http://sixty-north.com/c/t.txt' và kết quả lưu vào trong danh sách lstStory. Sinh viên hãy phát triển đoạn chương trình này để có thể lưu toàn bộ dữ liệu đọc được với định dạng tập tin lưu trữ là csv.

```
1 from urllib.request import urlopen
2 import csv
3
4 # URL của dữ liệu
5 url = 'http://sixty-north.com/c/t.txt'
6
7 # Đọc dữ liệu từ URL và lưu vào danh sách lstStory
8 with urlopen(url) as story:
9     lstStory = story.readlines()
10
11 # Tạo và ghi dữ liệu vào tệp CSV
12 with open('story.csv', mode='w', newline='', encoding='utf-8') as file:
13     writer = csv.writer(file)
14
15     # Ghi từng dòng vào file CSV
16     for line in lstStory:
17         # Chuyển đổi dòng từ bytes sang chuỗi và loại bỏ ký tự xuống dòng
18         writer.writerow([line.decode('utf-8').strip()])
19
20 print("Đã lưu dữ liệu vào tệp story.csv")
```

```
PS E:\Nam_III\HK241\PythonNangCao\Lab 9>
& C:/Users/ADMIN/AppData/Local/Programs/Python/Python38-64/Python.exe story.csv
Đã lưu dữ liệu vào tệp story.csv
```

```
story.csv > data
1 It was the best of times
2 it was the worst of times
3 it was the age of wisdom
4 it was the age of foolishness
5 it was the epoch of belief
6 it was the epoch of incredulity
7 it was the season of Light
8 it was the season of Darkness
9 it was the spring of hope
10 it was the winter of despair
11 we had everything before us
12 we had nothing before us
13 we were all going direct to Heaven
14 we were all going direct the other way
15 in short the period was so far like the present period that some of
16 its noisiest authorities insisted on its being received for good or for
17 evil in the superlative degree of comparison only
18
```

Bài 6: Đoạn mã dưới đây gọi api để lấy kết quả trả về từ iss-pass.json với tham số “lat” là 37.78 và “lon” là -122.41. Kết quả trả về như hình dưới đây. Sinh viên hãy bổ sung đoạn mã để lấy giá trị duration của phần tử đầu tiên.

```
1 import requests
2
3 # Thiết lập tham số lat và lon
4 parameters = {"lat": 37.78, "lon": -122.41}
5
6 # Gọi API và lấy kết quả trả về từ URL mới của NASA
7 response = requests.get("https://api.nasa.gov/planetary/apod", params={"api_key": "DEMO_KEY"})
8
9 # Kiểm tra mã trạng thái phản hồi
10 if response.status_code == 200:
11     try:
12         json_data = response.json()
13         print("Loại dữ liệu:", type(json_data))
14         print("Dữ liệu trả về:", json_data)
15
16         # Kiểm tra nếu dữ liệu có trường duration
17         if "duration" in json_data:
18             first_duration = json_data['duration']
19             print("Duration:", first_duration)
20         else:
21             print("Trường 'duration' không có trong dữ liệu trả về.")
22     except requests.exceptions.JSONDecodeError:
23         print("Lỗi: Không thể phân tích dữ liệu JSON")
24 else:
25     print(f"Lỗi: Mã trạng thái phản hồi là {response.status_code}")
26     print("Nội dung phản hồi:", response.text)
```

```
Loại dữ liệu: <class 'dict'>
Dữ liệu trả về: {'copyright': '\nFényes Lóránd\n', 'date': '2024-11-04', 'explanation': 'The Great Nebula in Orion, an immense, nearby starbirth region, is probably the most famous of all astronomical nebulas. Here, glowing gas surrounds hot young stars at the edge of an immense interstellar molecular cloud only 1500 light-years away. In the featured deep image in assigned colors highlighted by emission in oxygen and hydrogen, wisps and sheets of dust and gas are particularly evident. The Great Nebula in Orion can be found with the unaided eye near the easily identifiable belt of three stars in the popular constellation Orion. In addition to housing a bright open cluster of stars known as the Trapezium, the Orion Nebula contains many stellar nurseries. These nurseries contain much hydrogen gas, hot young stars, protoplanets, and stellar jets spewing material at high speeds. Also known as M42, the Orion Nebula spans about 40 light years and is located in the same spiral arm of our Galaxy as the Sun.', 'hdurl': 'https://apod.nasa.gov/apod/image/2411/Orion_Lorand_1992.jpg', 'media_type': 'image', 'service_version': 'v1', 'title': 'M42: The Great Nebula in Orion', 'url': 'https://apod.nasa.gov/apod/image/2411/Orion_Lorand_1992.jpg'}
Trường 'duration' không có trong dữ liệu trả về.
```

Bài 7: Sinh viên hãy xem hướng dẫn tạo API KEY ở mục lưu ý tại trang 13, phát triển thí dụ 9 trong bài (trang 12) để truy vấn tập các videos với videoid = [7cmvABXyUC0, 9eH-7x7swEM, JndzGxbwvG0, 10P5_E6J_g0] và lưu kết quả thành tập tin csv gồm những thông tin như bảng dưới đây:

```
1 import requests
2 import pandas as pd
3
4 # API Key của bạn
5 API_KEY = 'AIzaSyA0XSwWPf-mm1qWt1M-ZQ6LoMwRhXTm28'
6
7 # Các video ID cần truy vấn
8 video_ids = ['7cmvABXyUC0', '9eH-7x7swEM', 'Jndz6xbwv60', 'l0P5_E6J_g0']
9
10 # URL để lấy thông tin video
11 url = "https://www.googleapis.com/youtube/v3/videos"
12
13 # Danh sách để lưu kết quả
14 items = []
15
16 # Lặp qua từng video ID và lấy thông tin
17 for video_id in video_ids:
18     params = {
19         'part': 'statistics',
20         'id': video_id,
21         'key': API_KEY
22     }
23     response = requests.get(url, params=params)
24
25     # Kiểm tra nếu yêu cầu thành công
26     if response.status_code == 200:
27         data = response.json()
28         # Kiểm tra nếu dữ liệu tồn tại
29         if 'items' in data and len(data['items']) > 0:
30             stats = data['items'][0]['statistics']
31             # Lấy thông tin và thêm vào danh sách items
32             items.append({
33                 'videoid': video_id,
34                 'viewCount': stats.get('viewCount', 0),
35                 'likeCount': stats.get('likeCount', 0),
36                 'dislikeCount': stats.get('dislikeCount', 0),
37                 'favoriteCount': stats.get('favoriteCount', 0),
38                 'commentCount': stats.get('commentCount', 0)
39             })
40         else:
41             print(f"Lỗi: Không thể lấy thông tin cho video ID {video_id}")
42
43 # Chuyển đổi danh sách items thành DataFrame và lưu thành file CSV
44 fields = ['videoid', 'viewCount', 'likeCount', 'dislikeCount', 'favoriteCount', 'commentCount']
45 summaryResults = pd.DataFrame(items, columns=fields)
46 summaryResults.to_csv('video_statistics.csv', index=False)
47 print("Dữ liệu đã được lưu vào file video_statistics.csv")
```

```
PS E:\Nam_III\HK241\PythonNangCao\Lab 9>
& C:/Users/ADMIN/anaconda3/python.exe "e:/Nam_III/HK241/PythonNangCao/Lab 9/bai7.py"
Dữ liệu đã được lưu vào file video_statistics.csv
PS E:\Nam_III\HK241\PythonNangCao\Lab 9>
```

```
video_statistics.csv > data
1 videoid,viewCount,likeCount,dislikeCount,favoriteCount,commentCount
2 7cmvABXyUC0,2983,58,0,0,2
3 9eH-7x7swEM,281,5,0,0,0
4 Jndz6xbwv60,612,13,0,0,0
5 l0P5_E6J_g0,173892,2844,0,0,19
```