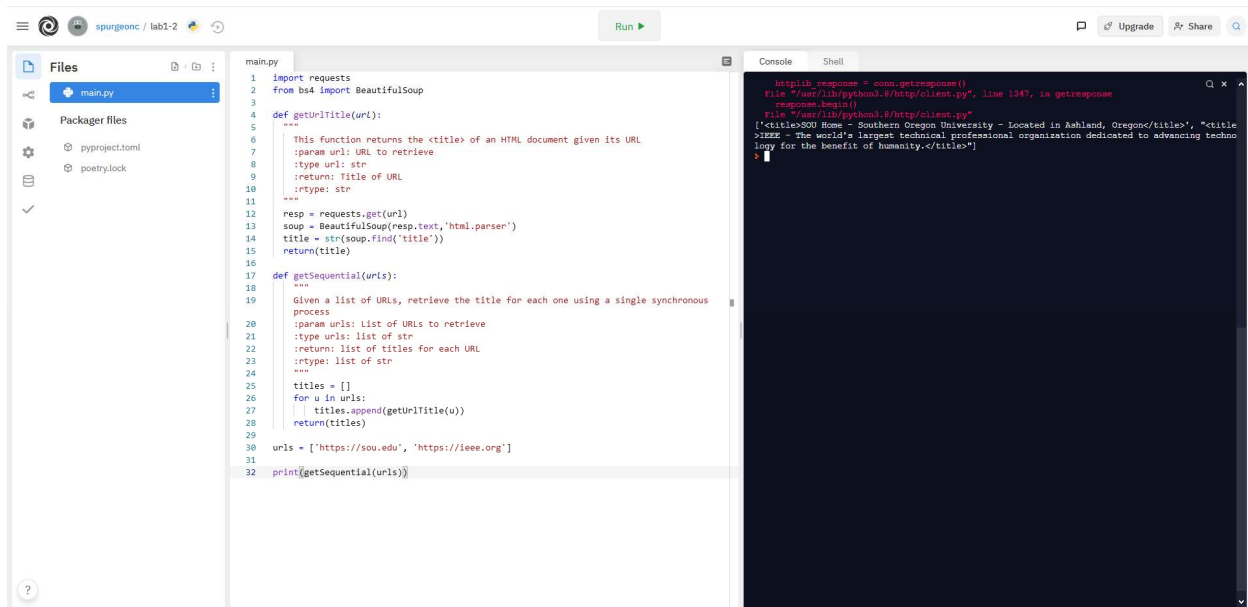


CS455 – Lab 1-2 – Chance Spurgeon-Couraud, 940458367

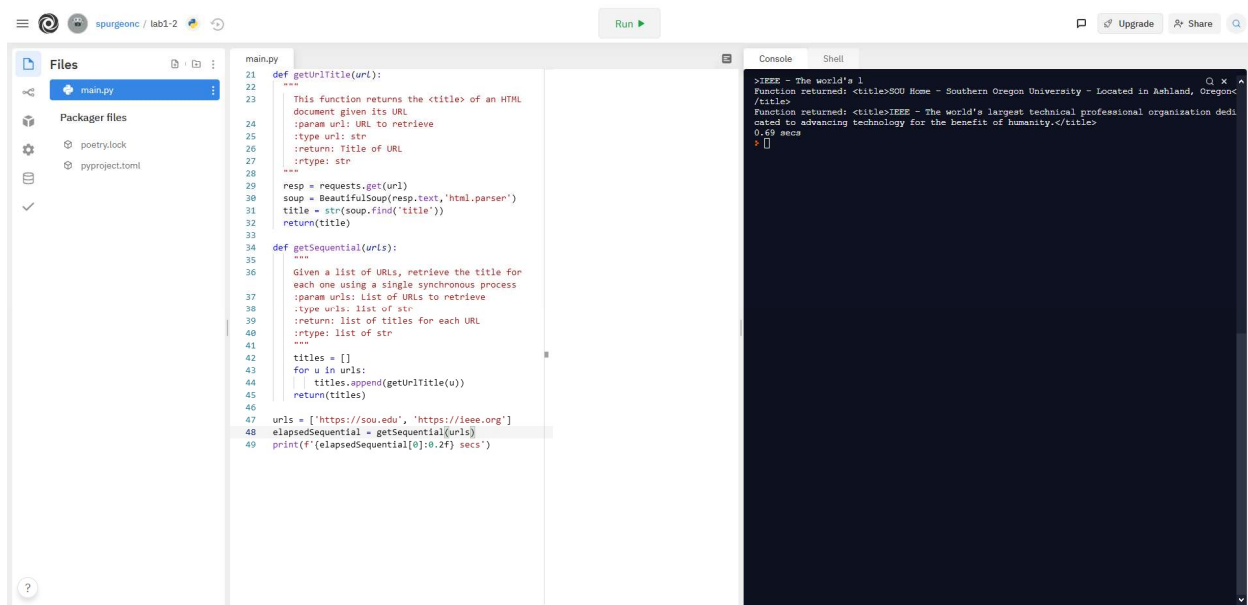


The screenshot shows a JupyterLab environment with a file explorer on the left, a code editor in the center, and a console on the right. The file explorer shows a file named `main.py`. The code editor displays the following Python code:

```
1 import requests
2 from bs4 import BeautifulSoup
3
4 def getUrlTitle(url):
5     """
6     This function returns the <title> of an HTML document given its URL
7     :param url: URL to retrieve
8     :type url: str
9     :return: Title of URL
10    :rtype: str
11    """
12    resp = requests.get(url)
13    soup = BeautifulSoup(resp.text, 'html.parser')
14    title = str(soup.find('title'))
15    return(title)
16
17 def getSequential(urls):
18     """
19     Given a list of URLs, retrieve the title for each one using a single synchronous
20     process
21     :param urls: List of URLs to retrieve
22     :type urls: list of str
23     :return: list of titles for each URL
24     :rtype: list of str
25     """
26    titles = []
27    for u in urls:
28        titles.append(getUrlTitle(u))
29    return(titles)
30
31 urls = ['https://sou.edu', 'https://ieee.org']
32 print(getSequential(urls))
```

The console on the right shows the output of the script:

```
httpLib_response = conn.getResponse()
File "/usr/lib/python3.8/http/client.py", line 1347, in getResponse
response.begin()
File "/usr/lib/python3.8/http/client.py"
[{"title>SOU Home - Southern Oregon University - Located in Ashland, Oregon</title>", "title
>IEEE - The world's largest technical professional organization dedicated to advancing technol
ogy for the benefit of humanity.</title>"}]
```



The screenshot shows a JupyterLab environment with a file explorer on the left, a code editor in the center, and a console on the right. The file explorer shows a file named `main.py`. The code editor displays the following Python code:

```
21 def getUrlTitle(url):
22     """
23     This function returns the <title> of an HTML
24     document given its URL
25     :param url: URL to retrieve
26     :type url: str
27     :return: Title of URL
28     :rtype: str
29     """
30    resp = requests.get(url)
31    soup = BeautifulSoup(resp.text, 'html.parser')
32    title = str(soup.find('title'))
33    return(title)
34
35 def getSequential(urls):
36     """
37     Given a list of URLs, retrieve the title for
38     each one using a single synchronous process
39     :param urls: List of URLs to retrieve
40     :type urls: list of str
41     :return: list of titles for each URL
42     :rtype: list of str
43     """
44    titles = []
45    for u in urls:
46        titles.append(getUrlTitle(u))
47    return(titles)
48
49 urls = ['https://sou.edu', 'https://ieee.org']
50 elapsedSequential = getSequential(urls)
51 print(f'{elapsedSequential[0]:0.2f} secs')
```

The console on the right shows the output of the script:

```
>IEEE - The world's l
Function returned: <title>SOU Home - Southern Oregon University - Located in Ashland, Oregon<
/ title>
Function returned: <title>IEEE - The world's largest technical professional organization dedi
cated to advancing technology for the benefit of humanity.</title>
0.69 secs
[]
```

spurgeonc / lab1-2

Run

Files

main.py

```
36 """
37 Given a list of URLs, retrieve the title for
38 each one using a single synchronous process
39 :param urls: list of URLs to retrieve
40 :type urls: list of str
41 :return: list of titles for each URL
42 :rtype: list of str
43 """
44 titles = []
45 for u in urls:
46     titles.append(getUrlTitle(u))
47 return(titles)
48
49 @time_decorator
50 def getMulti(urls, num_processes):
51     """
52     Given a list of URLs, retrieve the title for
53     each one using a single synchronous process
54     :param urls: list of URLs to retrieve
55     :type urls: list of str
56     :param num_processes: Number of processes to use
57     :type num_processes: int
58     :return: list of str
59     :rtype: list of str
60     """
61     p = multiprocessing.Pool(num_processes)
62     titles = p.map(getUrlTitle, urls)
63     p.close()
64     return(titles)
65
66 urls = ['https://sou.edu', 'https://ieee.org',
67         'https://google.com', 'https://facebook.com',
68         'https://repl.it', 'https://reddit.com',
69         'https://tumblr.com', 'https://homestuck.com',
70         'https://yahoo.com', 'https://minecraft.net',
71         'https://bing.com']
72
73 elapsedSequential = getSequential(urls)
74 concurrencies = [2, 5, 10]
75
76 for c in concurrencies:
77     fetch_time = getMulti(urls,c)
78     print(f'{c} {fetch_time:0.2f}')
```

Console

```
Function returned: ['<title>SOU Home - Southern Oregon University - Located in Ashland, OR</title>', '<title>IEEE - The world's largest technical professional organization dedicated to advancing technology for the benefit of humanity.</title>', '<title>Google</title>', '<title id="pageTitle">Facebook - Log In or Sign Up</title>', '<title>The collaborative browser has meed IDE - Replit</title>', '<title>reddit: the front page of the internet</title>', '<title id="pageTitle">Tumblr</title>', '<title>Homestuck Official | Webcomics by Andrew Hussie</title>', '<title>Yahoo</title>', '<title>Minecraft Official Site | Minecraft</title>', '<title>Bing</title>']
2 3.80
Function returned: ['<title>SOU Home - Southern Oregon University - Located in Ashland, Oregon</title>', '<title>IEEE - The world's largest technical professional organization dedicated to advancing technology for the benefit of humanity.</title>', '<title>Google</title>', '<title id="pageTitle">Facebook - Log In or Sign Up</title>', '<title>The collaborative browser has meed IDE - Replit</title>', '<title>reddit: the front page of the internet</title>', '<title id="pageTitle">Tumblr</title>', '<title>Homestuck Official | Webcomics by Andrew Hussie</title>', '<title>Yahoo</title>', '<title>Minecraft Official Site | Minecraft</title>', '<title>Bing</title>']
5 3.92
Function returned: ['<title>SOU Home - Southern Oregon University - Located in Ashland, Oregon</title>', '<title>IEEE - The world's largest technical professional organization dedicated to advancing technology for the benefit of humanity.</title>', '<title>Google</title>', '<title id="pageTitle">Facebook - Log In or Sign Up</title>', '<title>The collaborative browser has meed IDE - Replit</title>', '<title>reddit: the front page of the internet</title>', '<title id="pageTitle">Tumblr</title>', '<title>Homestuck Official | Webcomics by Andrew Hussie</title>', '<title>Yahoo</title>', '<title>Minecraft Official Site | Minecraft</title>', '<title>Bing</title>']
10 3.75
```

spurgeonc / lab1-2

Run

Files

main.py

```
104 process
105 :param urls: list of URLs to retrieve
106 :type urls: list of str
107 :param num_processes: Number of processes to use
108 :type num_processes: int
109 :return: list of str
110 :rtype: list of str
111 """
112 p = multiprocessing.Pool(num_processes)
113 titles = p.map(getUrlTitle, urls)
114 p.close()
115 return(titles)
116
117 """plt.scatter(concurrencies, elapsed)
118 plt.title("Chance Spurgeon-Coursaud")
119 plt.xlabel("Number of Processes")
120 plt.ylabel("Retrieval Time")
121 plt.show()"""
122
123 resp = requests.get
124 ('https://gist.githubusercontent.com/demersdesigns/4442cd84c1cc65cdda9b19eac1ba52b/raw/cf68189a805661d012113f9a4473a35e478509/craft-popular-urls')
125 urls = resp.text.split("\n")[1:50]
126
127 elapsedSequential = getSequential(urls)
128
129 """fetch_time = getAsync(urls)
130 print(f'Async version: {fetch_time:0.2f}')"""
131
132 for c in concurrencies:
133     fetch_time = getMulti(urls,c)
134     elapsed.append(fetch_time)
135     print(f'{c} {fetch_time:0.2f}')
136
137 """all commented-out code was causing errors I could not understand"""
```

Console

```
Traceback (most recent call last):
  File "main.py", line 124, in <module>
    resp = requests.get
  File "/usr/local/lib/python3.8/dist-packages/requests/api.py", line 75, in get
    return get(url, params=params, **kwargs)
  File "/usr/local/lib/python3.8/dist-packages/requests/api.py", line 59, in get
    response = session.get(url, **kwargs)
  File "/usr/local/lib/python3.8/dist-packages/requests/sessions.py", line 587, in get
    return self.request('GET', url, **kwargs)
  File "/usr/local/lib/python3.8/dist-packages/requests/sessions.py", line 541, in request
    resp = self.send(prep, **send_kwargs)
  File "/usr/local/lib/python3.8/dist-packages/requests/sessions.py", line 457, in send
    r = adapter.send(request, **kwargs)
  File "/usr/local/lib/python3.8/dist-packages/requests/adapters.py", line 490, in send
    resp = conn.urlopen(method, url, **kwargs)
  File "/usr/local/lib/python3.8/dist-packages/urllib3/connectionpool.py", line 702, in urlopen
    httplib_response = self._make_request(
  File "/usr/local/lib/python3.8/dist-packages/urllib3/connectionpool.py", line 485, in _make_request
    self._raise_timeout(err=e, url=url, timeout_value=conn.timeout)
  File "/usr/local/lib/python3.8/dist-packages/urllib3/connectionpool.py", line 455, in _raise_timeout
    raise ReadTimeoutError(self._pool, message, self._pool._request_timeout)
urllib3.exceptions.ReadTimeoutError: HTTPConnectionPool(host='gist.githubusercontent.com', port=443): Read timed out. (read timeout=30)
```