

Project objectives:

1. Fetch the earliest comments of 1 blog without logging in
2. Log in to the website
3. Post a comment (logging in is necessary): comment on the current blog

Edit by David Yang Mar 24, 2023

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

Crawl the earliest comments

```
In [2]: # create a comments list for all comments in the earliest page
data_list = []
```

```
In [3]: comments_url = 'https://wordpress-edu-3autumn.localprod.oc.forchange.cn/all-about-the-future_01/#comments/'

# get the request header
headers = {
    'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/70.0.3538.110 Safari/537.36'
}

res = requests.get(comments_url, headers = headers)
bs = BeautifulSoup(res.text, 'html.parser')
```

```
In [4]: # creat a infor dict for one comment
info_dict = {}
```

```
In [5]: # find the positions of all comments
```

```
comment_list = bs.find('ol', class_='comment-list').find_all('li')
```

In [6]:

```
# implement each comment
for comment in comment_list :
    user = comment.find('b').text

    comment_date = comment.find('time')

    comment_content = comment.find('p').text

    info_dict['user'] = user
    info_dict['datetime'] = comment_date
    info_dict['content'] = comment_content

    data_list.append(info_dict)
```

In [7]:

```
# save the comments into csv file
with open('comments.csv', 'w', encoding='utf-8') as f:
    f_csv = csv.DictWriter(f, fieldnames=['user', 'datetime', 'content'])
    f_csv.writeheader()
    f_csv.writerows(data_list)
```

Log in and obtain the cookies

In [8]:

```
login_url = 'https://wordpress-edu-3autumn.localprod.oc.forchange.cn/wp-login.php'
```

In [9]:

```
username = input('Pls input username: ')
password = input('Pls input password: ')
```

Pls input username: yuangangyang608@gmail.com

Pls input password: 123456

In [10]:

```
# get the request body of Login
login_data = {'log': username,
              'pwd': password,
              'wp-submit': '登录',
```

```
'redirect_to': 'https://wordpress-edu-3autumn.localprod.oc.forchange.cn',  
'testcookie': 1}
```

```
In [11]: login_res = requests.post(login_url, headers = headers, data = login_data)  
  
cookies = login_res.cookies
```

Send a comment and confirm

```
In [12]: comment_url = 'https://wordpress-edu-3autumn.localprod.oc.forchange.cn/all-about-the-future_01/'  
comment = input('Pls input comments: ')
```

Pls input comments: It's great!

```
In [13]: # get the request body of comment page  
comment_data = {'comment': comment,  
                'submit': '发表评论',  
                'comment_post_ID': 13,  
                'comment_parent': 0}  
  
comment_res = requests.post(comment_url, headers=headers, data=comment_data, cookies=cookies)
```

```
In [14]: if comment_res.status_code == 200:  
          print('sent comments successfully! ')  
        else:  
  
          print('sent failed! requests status code is {}'.format(comment_res.status_code))
```

sent comments successfully!

```
In [ ]:
```