

LETTERKENNY INSTITUTE OF TECHNOLOGY**ASSIGNMENT COVER SHEET**

Lecturer's Name: Ruth Lennon

Assessment Title: OOPR For Server Admin

Work to be submitted to: Ruth Lennon

Date for submission of work: 30/11/2021

Place and time for submitting work: Blackboard as per submission link

To be completed by the Student

Student's Name: Luis Gonzalez

Class: OOPR For Server Admin

Subject/Module: OOPR For Server Admin

Word Count (where applicable): N/A

I confirm that the work submitted has been produced solely through my own efforts.

Student's signature: Luis Gonzalez **Date:** 30/11/2021

Notes

Penalties: The total marks available for an assessment is reduced by 15% for work submitted up to one week late. The total marks available are reduced by 30% for work up to two weeks late. Assessment work received more than two weeks late will receive a mark of zero. [Incidents of alleged plagiarism and cheating are dealt with in accordance with the Institute's Assessment Regulations.]

Plagiarism: Presenting the ideas etc. of someone else without proper acknowledgement (see section L1 paragraph 8).

Cheating: The use of unauthorised material in a test, exam etc., unauthorised access to test matter, unauthorised collusion, dishonest behaviour in respect of assessments, and deliberate plagiarism (see section L1 paragraph 8).

Continuous Assessment: For students repeating an examination, marks awarded for continuous assessment, shall normally be carried forward from the original examination to the repeat examination.

Aims/Description

As per Assignment Question:

Code: Using Python on your host (windows) pc scrape the Apache 2 page you just created (or LYIT web page) and parse it minimally for later processing. For example:

- a. What are the headings?
- b. How many times does the word Apache2 appear ?
Think of searching for an event id or user id that is in your system.
- c. Any other item

Use BeautifulSoup or any other parser that you have never used before for this task.

<https://pypi.org/project/beautifulsoup4/>

The purpose of this task is both to install software and to use an api you have not seen before to carry out a minimal task. If apache is properly installed it should be visible from the host by typing in the ip address in your browser. See below.

ip addr show

Use screenshots to demonstrate that this worked. Save them to a file named L0012345_Q2_File_1 where L0012345 is replace by your own L number. Save the script as L0012345_Q2_File_2

Results

1. Library BeautifulSoup4 installed successfully
2. Python Project using BeautifulSoup4 & requests was created
3. Project to scrape websites was tested successfully
4. Current file, script and PyCharm project were uploaded to Student's GitHub Repository into OOPRAssignment_Q2
<https://github.com/L00170299/OOPRForServerAdmin>

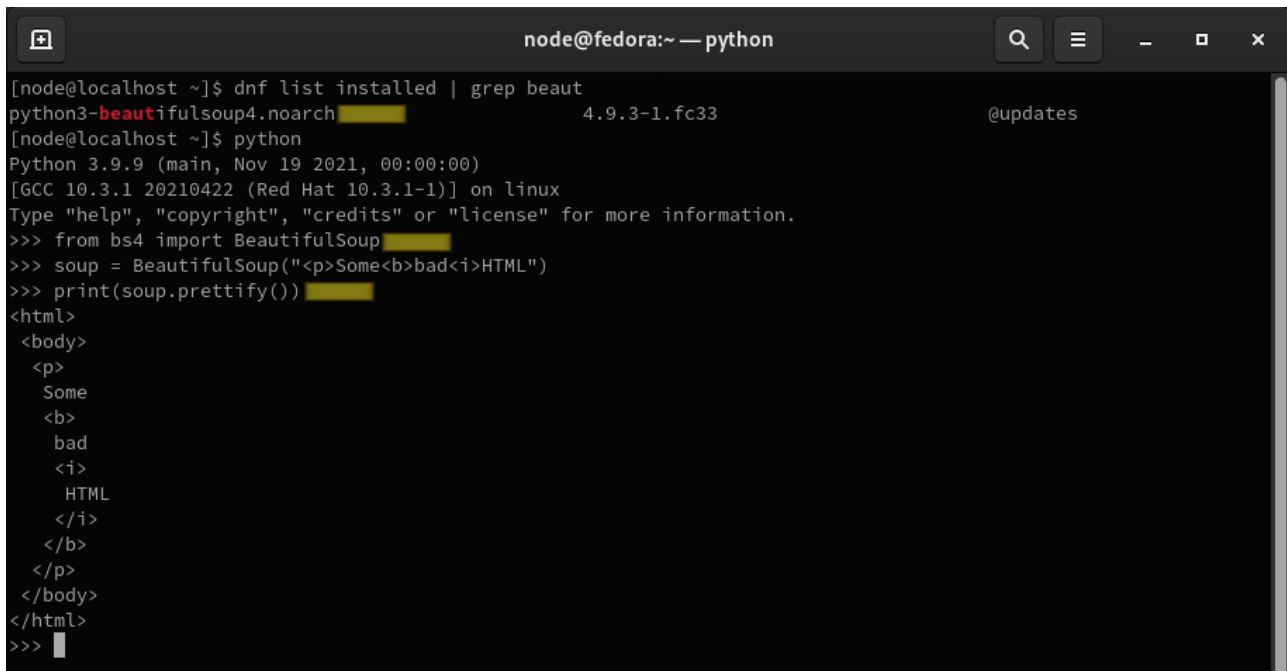
Conclusions

Student found some difficulties importing the new module within PyCharm. Fedora system was reporting that module was installed but project was failing to import it. Student found that system had installed module installed just for one of python versions. Which was updated after project got created.

Student found interesting the use of python to scrape websites, specially because in the past Student needed something similar for a personal project.

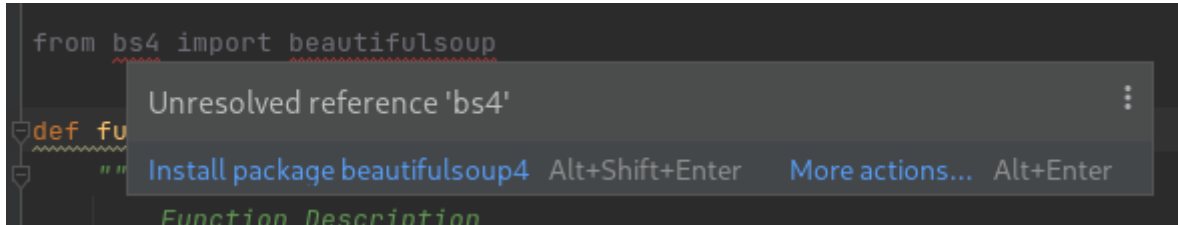
Appendix

Checking if beautifulsoup4 is installed and quick test:



```
node@fedora:~ — python
[node@localhost ~]$ dnf list installed | grep beaut
python3-beautifulsoup4.noarch 4.9.3-1.fc33 @updates
[node@localhost ~]$ python
Python 3.9.9 (main, Nov 19 2021, 00:00:00)
[GCC 10.3.1 20210422 (Red Hat 10.3.1-1)] on linux
Type "help", "copyright", "credits" or "license" for more information.
>>> from bs4 import BeautifulSoup
>>> soup = BeautifulSoup("<p>Some<b>bad<i>HTML")
>>> print(soup.prettify())
<html>
<body>
<p>
  Some
  <b>
    bad
    <i>
      HTML
    </i>
  </b>
</p>
</body>
</html>
>>>
```

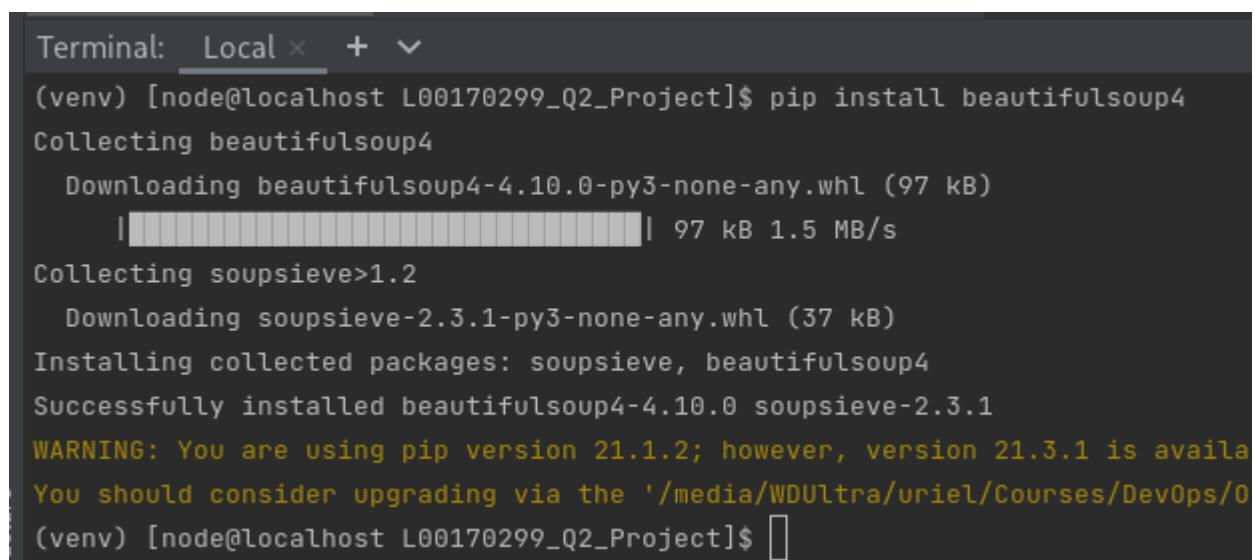
PyCharm wasn't able to load module. So installed within its environment:



```
from bs4 import BeautifulSoup
def fu
    """
    Function Description
```

Unresolved reference 'bs4'

Install package beautifulsoup4 Alt+Shift+Enter More actions... Alt+Enter



```
Terminal: Local x + v
(venv) [node@localhost L00170299_Q2_Project]$ pip install beautifulsoup4
Collecting beautifulsoup4
  Downloading beautifulsoup4-4.10.0-py3-none-any.whl (97 kB)
    | 97 kB 1.5 MB/s
Collecting soupsieve>1.2
  Downloading soupsieve-2.3.1-py3-none-any.whl (37 kB)
Installing collected packages: soupsieve, beautifulsoup4
Successfully installed beautifulsoup4-4.10.0 soupsieve-2.3.1
WARNING: You are using pip version 21.1.2; however, version 21.3.1 is available.
You should consider upgrading via the '/media/WDUltra/uriel/Courses/DevOps/0
(venv) [node@localhost L00170299_Q2_Project]$
```

PyCharm with Project & Code

```

78     pretty_page = get_pretty_response(soup_page)
79     return text in pretty_page
80
81
82 if __name__ == '__main__':
83     """
84
85
86 # List with webpages to scrape
87 url_list = ["http://192.168.0.222/", "http://192.168.0.222:8080/login",
88            "https://lyitbb.blackboard.com/"]
89 text_list = ["Fedora Webserver Test Page", "Welcome to Jenkins!", "contact the helpdesk on 0749186050"]
90
91 # we will scrape each page and store some info from each to use it after
92 for url_path in url_list:
93     # get content of website
94     soup_page = get_soup_response(url_path)
95
96     print(f"Link: {url_path}")
97
98     # from here we can do whatever we want with it

```

Output of code:

```

Run: main
/media/WDUltra/uriel/Courses/DevOps/OOPRForServerAdmin_RuthLennon/GitHub_Projects/OOPRForServerAdmin/OOPRAssignment_Q2/L00170299_Q2-Pro
Link: http://192.168.0.222/
Title: Test Page for the HTTP Server on Fedora
Links: ['https://getfedora.org/', 'https://docs.fedoraproject.org/en-US/quick-docs/getting-started-with-apache-http-server/index.html']
Contains 'Fedora Webserver Test Page': False
Contains 'Welcome to Jenkins!': False
Contains 'contact the helpdesk on 0749186050': False
=====
Link: http://192.168.0.222:8080/login
Title: Sign in [Jenkins]
Links: None
Contains 'Fedora Webserver Test Page': False
Contains 'Welcome to Jenkins!': True
Contains 'contact the helpdesk on 0749186050': False
=====
Link: https://lyitbb.blackboard.com/
Title: Blackboard Learn
Links: ["javascript:loadLocale('en_GB');", "javascript:loadLocale('en_US');", 'https://www.lyit.ie/Staff-Hub/Password-Manager', '/auth-
Contains 'Fedora Webserver Test Page': False
Contains 'Welcome to Jenkins!': False
Contains 'contact the helpdesk on 0749186050': True
=====
Process finished with exit code 0

```