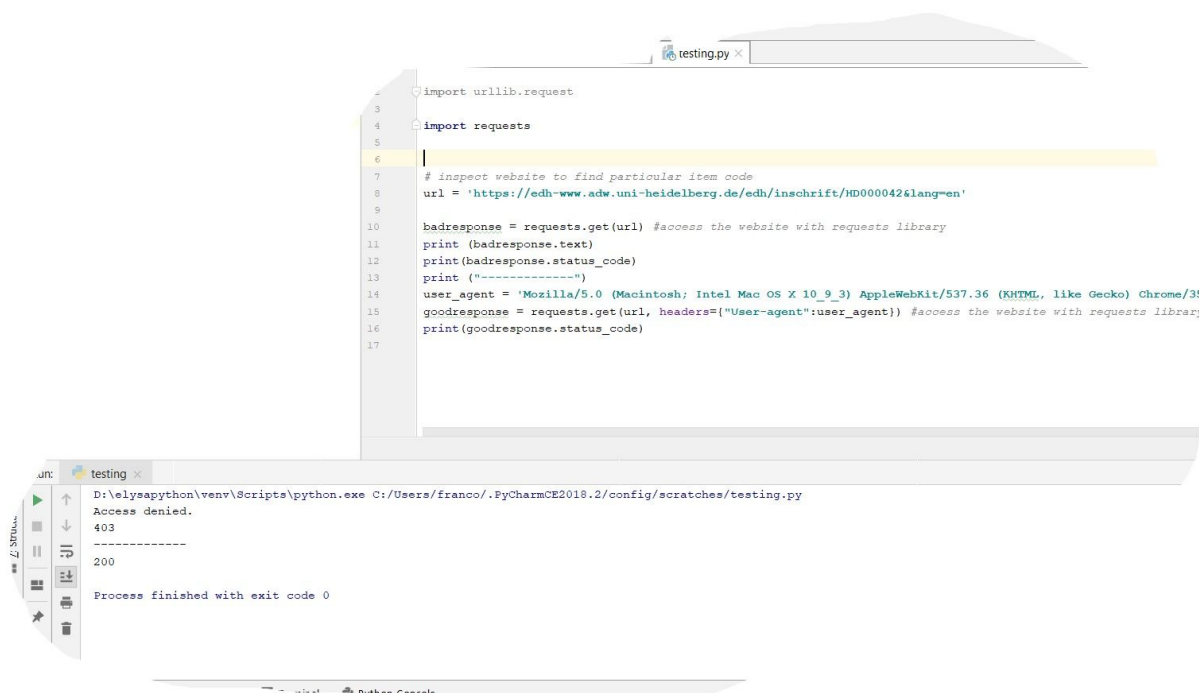


Elisa Brunetta, 866600, 19/11/2019

As I have previously anticipated on Monday's lesson, we have encountered some issues with the assignment. After having listened to your advice on Skype, I kept on working on the code and have reached some conclusions. I have worked on Pycharm to solve this problem. Starting from this code, I have tried to reach a solution.

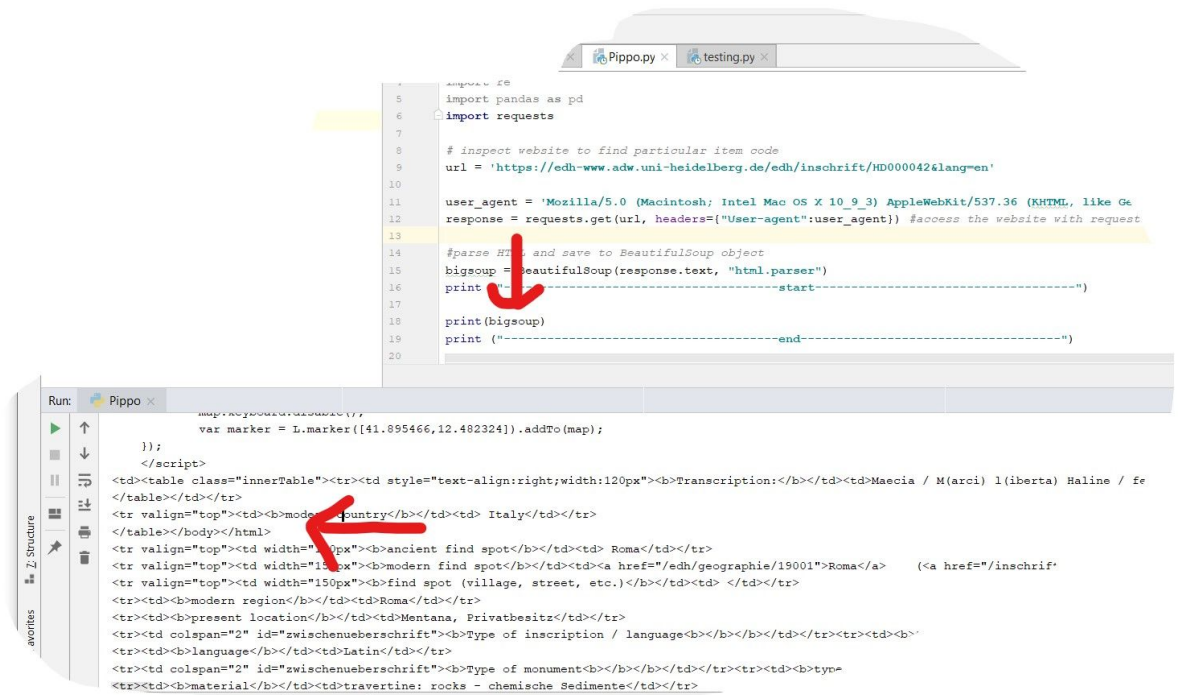
```
import requests
# inspect website to find particular item code
url = 'https://edh-www.adw.uni-heidelberg.de/edh/inschrift/HD000042&lang=en'
response = requests.get(url) #access the website with requests library
response.text
#parse HTML and save to BeautifulSoup object
bigsoup = BeautifulSoup(response.text, "html.parser")
#print(bigsoup)
for td in bigsoup.select("table.innerTable tr td"): #code used in class
    if (td.text == 'language:'):
        #print(td.text)
        tdn = td.fetchNextSiblings()
        print(tdn[0].text)
```

- 1) the first error I got was a message of 403 “Access Denied”. I then figured that the EDH website wasn't consenting the access to the data, and I thus researched how to bypass this problem. Although I seemed to understand this is not the cleanest solution, I gave it a try. I could do this in the following way:



This seemed to work, as I didn't receive an “Access Denied” error again. I thus continued and tried to understand if my code was now working.

- 2) I tried to rerun the code, but it was still not working. I thus decided to analyze the html parser result given by BeautifulSoup, and tried to find the “object” (“language”) I was assigned to. By reading the HTML line by line, I noticed an error: the tags `</body></html>` close the table “too early”, leaving out some information “untagged”.



- 3) I thus thought that the problem was BeautifulSoup. So, I decided to head directly to the EDH detailed page and inspect the HTML code itself, to see if BeautifulSoup was the real problem. I could therefore identify the object “language” inside the class “treffentabelle”. When running the code and looking for that in *for td in bigsoup.select("table.treffentabelle tr td")*, the code could evidently not work because the parsed HTML offered by BeautifulSoup was incorrect.

The screenshot shows a web browser with a table of monument data. The table has the following rows:

|                                   |                        |
|-----------------------------------|------------------------|
| Italic region                     | ↗ Roma                 |
| modern country                    | Italy                  |
| ancient find spot                 | Roma                   |
| modern find spot                  | ↗ Roma ↗ Radial Search |
| find spot (village, street, etc.) |                        |
| modern region                     | Roma                   |
| present location                  | Mentana, Privatbesitz  |
| Type of inscription / language    |                        |
| type of inscription               | epitaph                |
| language                          | Latin                  |
| Type of monument                  |                        |

Below the table, the browser's developer tools are open, showing the HTML source code. Red arrows point to the 'language' field in the table and the corresponding HTML code: `<b>language</b> == $0`.

CONCLUSION: Since apparently BeautifulSoup is not working the way it should be with this particular URL, I have come to the conclusion that I would need a different instrument to parse the HTML, albeit that the EDH website allows access to its data.