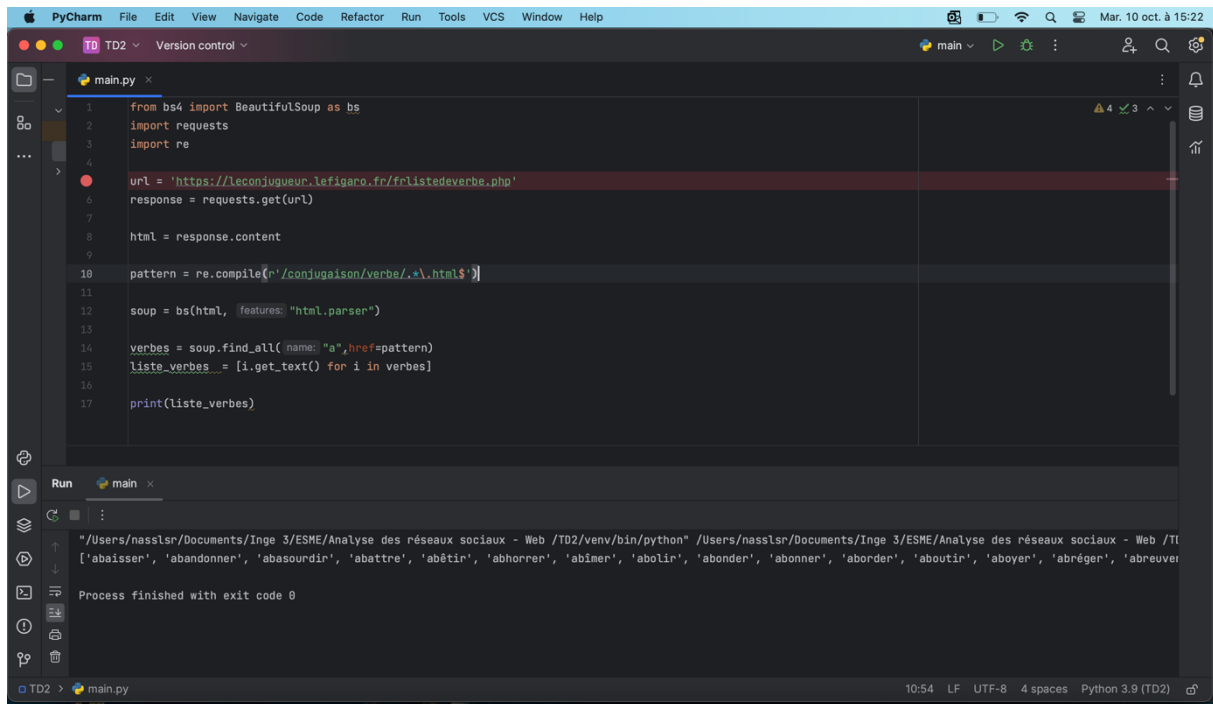


Rendu TD2 LAASRI NASSER

Question 1 :



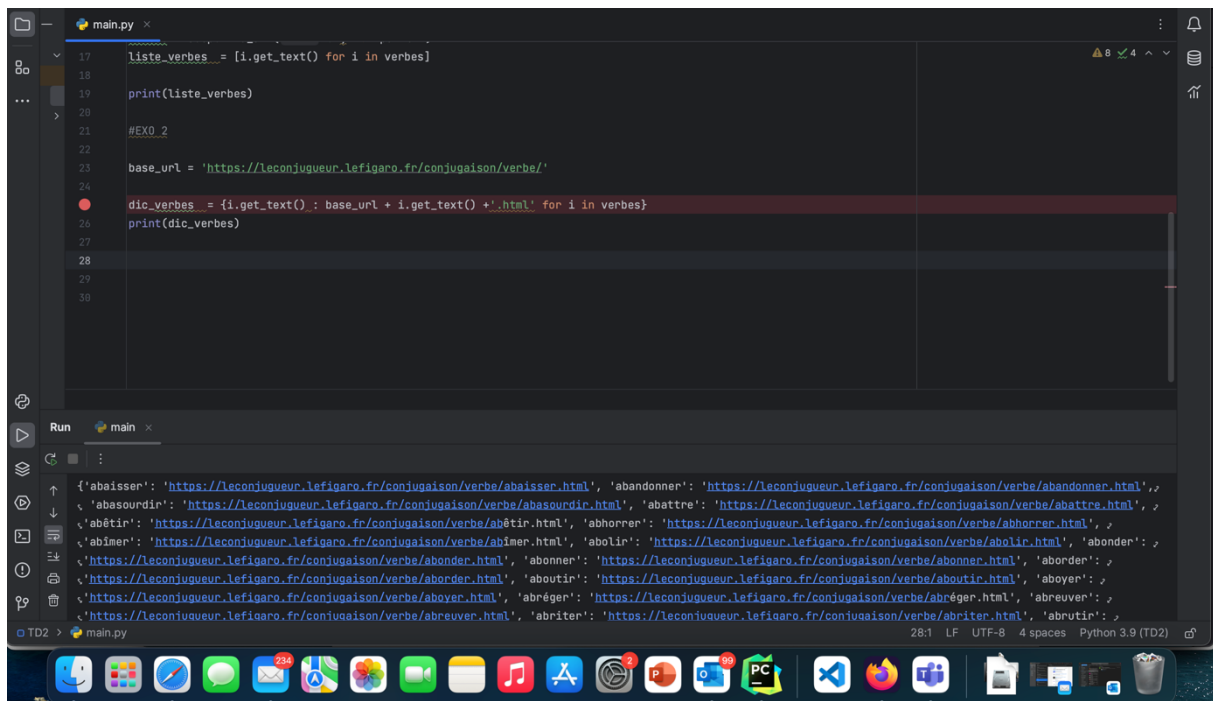
The screenshot shows a PyCharm IDE with a file named `main.py`. The code uses `BeautifulSoup` and `requests` to fetch a webpage and extract a list of verbs. The output of the program is displayed in the Run console.

```
1 from bs4 import BeautifulSoup as bs
2 import requests
3 import re
4
5 url = 'https://leconjugueur.lefigaro.fr/fr/listedeverbe.php'
6 response = requests.get(url)
7
8 html = response.content
9
10 pattern = re.compile(r'"/conjugaison/verbe/.+\.html$')
11
12 soup = bs(html, features="html.parser")
13
14 verbes = soup.find_all(name="a", href=pattern)
15 liste_verbes = [i.get_text() for i in verbes]
16
17 print(liste_verbes)
```

Run console output:

```
"/Users/nasslr/Documents/Inge 3/ESME/Analyse des réseaux sociaux - Web /TD2/venv/bin/python" /Users/nasslr/Documents/Inge 3/ESME/Analyse des réseaux sociaux - Web /TD2/venv/bin/python
['abaissér', 'abandonner', 'abasourdir', 'abattre', 'abêtir', 'abhorrer', 'abîmer', 'abolir', 'abonder', 'abonner', 'aborder', 'aboutir', 'aboyer', 'abrégér', 'abreuver']
Process finished with exit code 0
```

Question 2 :



The screenshot shows a PyCharm IDE with a file named `main.py`. The code builds on the previous one by creating a dictionary of verbs and their full URLs. The output of the program is displayed in the Run console.

```
17 liste_verbes = [i.get_text() for i in verbes]
18
19 print(liste_verbes)
20
21 #EX0_2
22
23 base_url = 'https://leconjugueur.lefigaro.fr/conjugaison/verbe/'
24
25 dic_verbes = {i.get_text(): base_url + i.get_text() + '.html' for i in verbes}
26 print(dic_verbes)
27
28
29
30
```

Run console output:

```
{'abaissér': 'https://leconjugueur.lefigaro.fr/conjugaison/verbe/abaissér.html', 'abandonner': 'https://leconjugueur.lefigaro.fr/conjugaison/verbe/abandonner.html', 'abasourdir': 'https://leconjugueur.lefigaro.fr/conjugaison/verbe/abasourdir.html', 'abattre': 'https://leconjugueur.lefigaro.fr/conjugaison/verbe/abattre.html', 'abêtir': 'https://leconjugueur.lefigaro.fr/conjugaison/verbe/abêtir.html', 'abhorrer': 'https://leconjugueur.lefigaro.fr/conjugaison/verbe/abhorrer.html', 'abîmer': 'https://leconjugueur.lefigaro.fr/conjugaison/verbe/abîmer.html', 'abolir': 'https://leconjugueur.lefigaro.fr/conjugaison/verbe/abolir.html', 'abonder': 'https://leconjugueur.lefigaro.fr/conjugaison/verbe/abonder.html', 'abonner': 'https://leconjugueur.lefigaro.fr/conjugaison/verbe/abonner.html', 'aborder': 'https://leconjugueur.lefigaro.fr/conjugaison/verbe/aborder.html', 'aboutir': 'https://leconjugueur.lefigaro.fr/conjugaison/verbe/aboutir.html', 'aboyer': 'https://leconjugueur.lefigaro.fr/conjugaison/verbe/aboyer.html', 'abrégér': 'https://leconjugueur.lefigaro.fr/conjugaison/verbe/abrégér.html', 'abreuver': 'https://leconjugueur.lefigaro.fr/conjugaison/verbe/abreuver.html', 'abriter': 'https://leconjugueur.lefigaro.fr/conjugaison/verbe/abriter.html', 'abrutir': 'https://leconjugueur.lefigaro.fr/conjugaison/verbe/abrutir.html'}
```

Question 4 :

Différence entre `find` et `find_all` :

- **find** : Il trouve et renvoie le premier élément correspondant au critère de recherche.
- **find_all** : Il trouve et renvoie tous les éléments correspondants à la critère de recherche sous forme de liste.

Question 5 :

Type de sortie de **find_all** :

- Le type de sortie de **find_all** est une liste contenant tous les éléments correspondants trouvés dans le document HTML.

Question 6

Qu'est-ce qu'un parser :

- Un parser est un programme ou une bibliothèque qui analyse et traite des données structurées, telles que des documents HTML ou XML, afin de les rendre accessibles et manipulables. Il permet de extraire des informations spécifiques à partir de ces données, ce qui est utile en scraping web.

Question 7

Différence entre le scraping et le crawling :

- Le scraping (ou web scraping) consiste à extraire spécifiquement des données d'une page web ou de plusieurs pages web, en utilisant un script ou un programme automatisé.
- Le crawling (ou web crawling) est le processus de parcours automatique du web à travers une série de liens pour découvrir et indexer des pages web.

Question 8

Les 2 choses à vérifier si on veut scraper un site :

- Respecter les termes d'utilisation du site : Avant de scraper un site web, il est essentiel de consulter les conditions d'utilisation du site pour vérifier s'il autorise le scraping et s'il existe des restrictions spécifiques.
- Verifier robots.txt