

Міністерство освіти і науки України
Державний університет “Житомирська політехніка”

Кафедра інженерії програмного забезпечення
Група: ВТ-21-1[1]

Програмування мовою Python
Лабораторна робота № 13
«Парсинг сайтів»

Виконав:

Бабушко А. С.

Прийняв:

Морозов Д. С.

					«Житомирська політехніка».22.121.01.000–Лр13		
Змн.	Арк.	№ докум.	Підпис	Дата	Звіт з лабораторної роботи		
Розроб.		Бабушко А.С.					
Перевір.		Морозов Д.С.					
Керівник							
Н. контр.							
Затверд.					ФІКТ Гр. ВТ-21-1[1]		
					Літ.	Арк.	Аркушів
						1	4

Мета роботи: ознайомитися з роботою з бібліотек для парсингу сайтів.

Хід роботи:

Використовуючі будь-які бібліотеки для парсингу сайтів створіть програму для пошуку інформації на сайті розкладу занять денної форми Державного університету «Житомирська політехніка» <https://rozklad.ztu.edu.ua/> для вирішення наступних завдань:

1. Визначте, в яких аудиторіях проводиться більше всього занять на ФІКТі серед усіх груп і усіх курсів.
2. Визначте, які аудиторії обчислювального центру (мають приставку ОЦ перед номером аудиторії) частіше всього залишаються не задіяними.
3. Визначте, в якій аудиторії частіше всього проходять заняття у вашої групи

Лістинг програми:

```
from bs4 import BeautifulSoup
from requests import get
from re import findall

# 1. Визначте, в яких аудиторіях проводиться найбільше занять на ФІКТі серед усіх груп і
усіх курсів.

# Parse groups list
request = get('https://rozklad.ztu.edu.ua')
html = BeautifulSoup(request.content, "lxml")

groups_list: list[str] = []
groups = html.select('body > div:nth-child(15) .collection .collection-item')
for group in groups:
    groups_list.append(group.text)

# Parse rooms
rooms_dict: dict = {}
for group in groups_list:
    request = get('https://rozklad.ztu.edu.ua/schedule/group/' + str(group))
    html = BeautifulSoup(request.content, "lxml")
    rooms = html.select('.schedule .variative .room')

    for room in rooms:
        room_number = findall(r'\d+', room.text).pop()

        if room_number in rooms_dict:
            rooms_dict[room_number] = rooms_dict[room_number] + 1
        else:
            rooms_dict[room_number] = 1

# Find top 10 max rooms
top_10_rooms: dict = {}
rooms_keys: list = list(rooms_dict.keys())
rooms_values: list = list(rooms_dict.values())
for i in range(0, 10):
    max_index = rooms_values.index(max(rooms_values))

    top_10_rooms[rooms_keys[max_index]] = rooms_values[max_index]

    rooms_keys.pop(max_index)
    rooms_values.pop(max_index)
```

		Бабушко А.С.			«Житомирська політехніка».22.121.01.000 – Лр13	Арк.
		Морозов Д.С.				2
Змн.	Арк.	№ докум.	Підпис	Дата		

```

print('Top 10 rooms in FIKT')
for key, value in top_10_rooms.items():
    print(f'Room: {key}; Classes: {value}')

# 2. Розрахувати кількість занять за місяць у своїй групі
request = get('https://rozklad.ztu.edu.ua/schedule/group/' + 'BT-21-1')
html = BeautifulSoup(request.content, "lxml")
vt_21_1_rooms = html.select('.schedule .variative .room')
print(f'Count of classes of BT-21-1 in 28 days: {len(vt_21_1_rooms) * 2}')

# 3. Визначте, в якій аудиторії частіше всього проходять заняття у вашій групі.

# Parse schedule of BT-21-1
vt_21_1_rooms_dict: dict = {}

for room in vt_21_1_rooms:
    room_number = findall(r'\d+', room.text).pop()

    if room_number in vt_21_1_rooms_dict:
        vt_21_1_rooms_dict[room_number] = vt_21_1_rooms_dict[room_number] + 1
    else:
        vt_21_1_rooms_dict[room_number] = 1

# Find top 1 classroom
vt_21_1_rooms_values: list = list(vt_21_1_rooms_dict.values())
vt_21_1_max_index = vt_21_1_rooms_values.index(max(vt_21_1_rooms_values))
print('\nTop 1 room in BT-21-1')
print(f'Room: {list(vt_21_1_rooms_dict.keys())[vt_21_1_max_index]}; Classes: {vt_21_1_rooms_values[vt_21_1_max_index]}')

```

Результат програми:

1, 3 завдання:

```

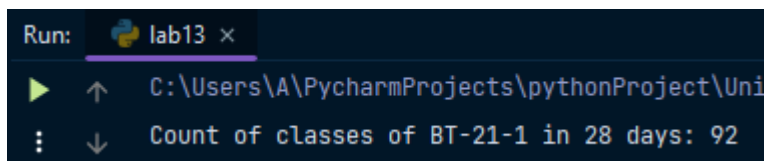
Run: lab13 x
C:\Users\A\PycharmProject
Top 10 rooms in FIKT
Room: 119; Classes: 54
Room: 123; Classes: 50
Room: 113; Classes: 49
Room: 118; Classes: 46
Room: 1; Classes: 45
Room: 111; Classes: 44
Room: 134; Classes: 43
Room: 124; Classes: 41
Room: 127; Classes: 38
Room: 137; Classes: 38

Top 1 room of BT-21-1
Room: 1; Classes: 4

```

		Бабушко А.С.			«Житомирська політехніка».22.121.01.000 – Лр13	Арк.
		Морозов Д.С.				3
Змн.	Арк.	№ докум.	Підпис	Дата		

2 завдання:



```
Run: lab13 x
C:\Users\A\PycharmProjects\pythonProject\Uni
Count of classes of BT-21-1 in 28 days: 92
```

Висновок: під час виконання лабораторної роботи було отримано навички парсингу сайтів за допомогою бібліотеки BeautifulSoup.

		Бабушко А.С.			«Житомирська політехніка».22.121.01.000 – Лр13	Арк.
		Морозов Д.С.				
Змн.	Арк.	№ докум.	Підпис	Дата		4