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.

***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 завданнь:

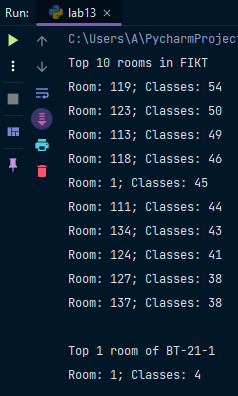
1. Визначте, в яких аудиторіях проводиться більше всього занять на ФІКТі серед усіх груп і усіх курсів.
2. Визначте, які аудиторії обчислювального центру (мають приставку ОЦ перед номером аудиторії) частіше всього залишаються не задіяними.
3. Визначте, в якій аудиторії частіше всього проходять заняття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

***Лістинг програми:***

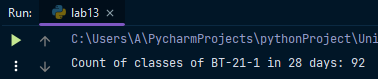
*from* bs4 *import* BeautifulSoup  
*from* requests *import* get  
*from* re *import* findall  
  
*# 1. Визначте, в яких аудиторіях проводиться найбільше занять на ФІКТі серед усіх груп і усіх курсів.  
  
# Parse groups list*request = get('https://rozklad.ztu.edu.ua')  
html = BeautifulSoup(request.content, "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)  
  
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/' + 'ВТ-21-1')  
html = BeautifulSoup(request.content, "lxml")  
vt\_21\_1\_rooms = html.select('.schedule .variative .room')  
print(f'Count of classes of ВТ-21-1 in 28 days: {len(vt\_21\_1\_rooms) \* 2}')  
  
  
*# 3. Визначте, в якій аудиторії частіше всього проходять заняття у вашої групи.  
  
# Parse schedule of ВТ-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 ВТ-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:***

******

***2 завдання:***

******

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