

Build Automatizálás Python-ban

4. hét házi feladat – Fehér Péter

Feladat ismertetése

Munkámból adódóan, sokat foglalkozom weboldalakon lévő adatok manipulálásával, statisztikához és egyéb feladatokhoz, ezért csináltam egy Python kódot, ami lekérdezi a Cubix oldalán lévő összes aktív Python képzést, kiírja a nevüket, és a hozzá tartozó linket.

A feladathoz a requests és a BeautifulSoup könyvtárakat használtam.

A requests könyvtár segítségével http kéréseket tudunk küldeni, ezzel ellenőrizve hogy elérhető-e, a cél, a BeautifulSoup pedig lehetővé teszi, hogy weboldalak forráskódjából adatokat emeljünk ki.

```
C:\Users\feherp\Tanulas\Szoftverfejlesztes\.venv\Scripts\python.exe C:\Users\feherp\Tanulas\Szoftverfejlesztes\4.het\Build_hazi\beautifulsoup_demo.py
Cubixon elérhető Python képzések listája:
Képzés címe: Python alapú data analyst képzés adatelemzési és gépi tanulási könyvtárak használatával
Link a képzéshez: https://courses.cubixedu.com/python-alapu-data-analyst-kepzes-pandas-matplotlib-seaborn-scipy-numpy-scikit-learn
Képzés címe: Data Engineer alapképzés
Link a képzéshez: https://courses.cubixedu.com/data-engineer-alapkepzes
Képzés címe: MLOps - Machine Learning modellek üzemeltetése (ML Engineering)
Link a képzéshez: https://courses.cubixedu.com/mlops-machine-learning-modellek-uzemeltetese
Képzés címe: Python programozás alapok [KEZDÉS: 2024.11.18.]
Link a képzéshez: https://courses.cubixedu.com/python-programozas-alapok-online-tanfolyam
Képzés címe: Python fejlesztés középfeladók szinten [INDULÁS: 2024.12.02.]
Link a képzéshez: https://courses.cubixedu.com/python-alapu-szoftverfejlesztes-online-kepzes
Képzés címe: Computer Vision Engineer képzés
Link a képzéshez: https://courses.cubixedu.com/computer-vision-engineer-kepzes
Képzés címe: Bevezetés a Python programozásba
Link a képzéshez: https://courses.cubixedu.com/python-programozasi-nyelv-kezdoknek

Process finished with exit code 0
```

Függőségek

Kézi megvalósítás

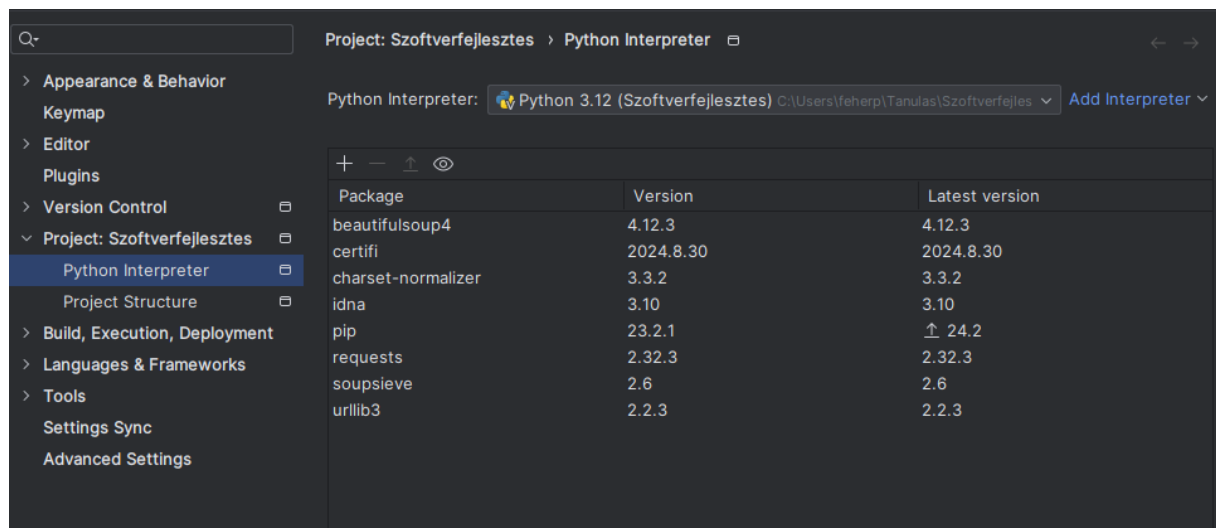
A külső könyvtárak kézi telepítéséhez a terminálba írva

```
pip install requests
```

```
pip install beautifulsoup4
```

parancsokat használtam.

A File->Setting->Project: Projektnév -> Python Interpreter funkciót szoktam többnyire használni, itt látszik egyből a verzió, ha frissítés érhető el, illetve leírás, számomra kényelmesebb.



Automatizált magvalósítás

A külső könyvtárak automatizált kezeléséhez az interneten több megoldást láttam, például a **setup.py**-t:

```
beautifulsoup_demo.py  setup.py x
1  from setuptools import setup, find_packages
2
3  setup(
4      name='Build',
5      version='0.1',
6      author='Fehér Péter',
7      author_email='feher.peter.lazlo@gmail.com',
8      description='Build automatizálás házi feladat',
9      url='https://github.com/FPeterL/Cubix_Software_Dev_Homeworks',
10     packages=find_packages(),
11     classifiers=[
12         'Programming Language :: Python :: 3',
13         'License :: OSI Approved :: MIT License',
14         'Operating System :: OS Independent',
15     ],
16     python_requires='>=3.6', # Python verzió
17     install_requires=open('requirements.txt').read().splitlines(), # Függőségek a requirements.txt-ből
18 )
```

Egy külön python script, illetve egy hozzá tartozó requirements.txt, melyből kiolvassa a telepítendő könyvtárakat. A konfiguráció metaadatokat tartalmaz a projektről, programról, fejlesztőről, illetve magára a pythonra vonatkozóan.

pip install .

paranccsal futtatható, mely kezeli a telepítéseket:

```
Terminal  Local  +
(.venv) PS C:\Users\feherp\Tanulas\Cubix_Szoftverfejlesztes\4_het\Build_hazi> cd..
(.venv) PS C:\Users\feherp\Tanulas\Cubix_Szoftverfejlesztes\4_het> pip install .
Processing c:\users\feherp\Tanulas\Cubix_Szoftverfejlesztes\4_het
  Installing build dependencies ... done
  Getting requirements to build wheel ... done
  Preparing metadata (pyproject.toml) ... done
Collecting beautifulsoup4
  Obtaining dependency information for beautifulsoup4 from https://files.pythonhosted.org/packages/b1/fe/ebc272a95b17eccc5cbf43e1d0b58d8d0b0a3a44cd371c79f4e59167f/beautifulsoup4-4.12.3-py3-none-any.whl.metadata
  Using cached beautifulsoup4-4.12.3-py3-none-any.whl.metadata (3.8 kB)
Collecting requests (from beautifulsoup4)
  Obtaining dependency information for requests from https://files.pythonhosted.org/packages/f9/9b/335f976a76e915ed497fcd0b110f5df0a47b7f757da4a2a0b0d0da4d5/requests-2.32.3-py3-none-any.whl.metadata
  Using cached requests-2.32.3-py3-none-any.whl.metadata (4.6 kB)
Collecting soupsieve>1.2 (from beautifulsoup4->build==0.1)
```

Alternatív megoldásként működhet egy sima szöveges file amibe felsoroljuk a telepítendő könyvtárakat, és a

pip install -r requirements.txt

telepíti őket. Ez a második megoldás, kicsi magán projektek esetén szerintem egyszerűbb, tekintve hogy a setup.py-nál leírások alapján figyelni jobban figyelni kell a mappa struktúrára.

Elérhetőség

https://github.com/FPeterL/Cubix_Software_Dev_Homeworks.git

Forráskód

Setup.py

Leírás

- **name:** A csomag neve, amely alatt a felhasználók telepíthetik.

- **version:** A csomag verziószáma, amely általában a szemantikus verziózás szabályai szerint van megadva.
- **author:** A csomag szerzőjének neve.
- **author_email:** A szerző e-mail címe.
- **description:** A csomag rövid leírása, amely összefoglalja, hogy mi a csomag célja.
- **url:** A csomag weboldala vagy GitHub repository linkje.
- **packages:** A csomagolt modulok listája. A setuptools find_packages() függvény az automatikusan megkeresi ezeket.
- **classifiers:** A csomaghoz tartozó kategóriák listája, amely segít másoknak megtalálni a csomagot. A Python Package Index (PyPI) számára releváns információkat tartalmaz.
- **python_requires:** A Python verzióra vonatkozó követelmények. Meghatározza, hogy a csomag mely Python verzióval használható.
- **install_requires:** A csomag függőségei, amelyek telepítésre kerülnek a pip által. Ezek a csomagok szükségesek a csomag működéséhez.

Kód

```
from setuptools import setup, find_packages
```

```
setup(
    name='Build',
    version='0.1',
    author='Fehér Péter',
    author_email='feher.peter.lazlo@gmail.com',
    description='Build automatizálás házi feladat',
    url='https://github.com/FPeterL/Cubix_Software_Dev_Homeworks',
    packages=find_packages(),
    classifiers=[
        'Programming Language :: Python :: 3',
        'License :: OSI Approved :: MIT License',
        'Operating System :: OS Independent',
    ],
    python_requires='>=3.6', # Python verzió
    install_requires=open('requirements.txt').read().splitlines(), # Függőségek a requirements.txt-ből
)
```

beautifulsoup.py

Kód

```
#pip install -r requirements.txt --alternatív mód a könyvtárak telepítésére
```

```
import requests
```

```
from bs4 import BeautifulSoup
```

```
#Cubix python képzések szűrés linkje
```

```
url =
```

```
"https://courses.cubixedu.com/kepzesek?_gl=1%2a1ussfu4%2a_ga%2aMTAyMjg2NjYxNS4xNzE5MzA5ODM4%2a_ga_KG4N6G61R4%2aMTcyODQ1NDA0MS4yMS4wLjE3Mjg0NTQwNTkuNDIuMC45NDYyNDc2NTc.%2a_gcl_au%2aOTgzMjU5OTY4LjE3Mjg0OTM2MzA.&tags=python"
```

```

# Az oldal alap URL-je a relatív hivatkozásokhoz
base_url = "https://courses.cubixedu.com"

# Weboldal lekérése
response = requests.get(url)

# Oldal válaszána ellenőrzése
if response.status_code == 200:
    print("Cubixon elérhető Python képzések listája:")
    # HTML feldolgozása BeautifulSouptal
    soup = BeautifulSoup(response.text, 'html.parser')

    # Képzések keresése: 'course-box-v2' osztályú elemek, amelyek <a> tageket tartalmaznak
    courses = soup.find_all('div', class_='course-box-v2')

    # Képzések URL-jeinek és a belső címek lekérése
    for course in courses:
        link = course.find('a', href=True) # Keresünk <a> tagot href attribútummal
        if link:
            # Ellenőrizzük, hogy a link relatív-e, és ha igen, csak akkor fűzzük hozzá a base_url-t
            if link['href'].startswith('http'):
                course_url = link['href'] # Teljes URL, nincs szükség a base_url-re
            else:
                course_url = base_url + link['href'] # Relatív URL, fűzzük hozzá a base_url-t
            # Az egyes képzések oldalának lekérése
            course_response = requests.get(course_url)
            if course_response.status_code == 200:
                # Képzések oldala feldolgozása
                course_soup = BeautifulSoup(course_response.text, 'html.parser')
                # Az 'h1' elem keresése (a képzés címe)
                h1_tag = course_soup.find('h1')
                if h1_tag:
                    print(f"Képzés címe: {h1_tag.get_text().strip()}")
                    print(f"Link a képzéshez: {course_url}")
                else:
                    print("Nem található 'h1' elem az oldalon.")
            else:
                print(f"Nem sikerült elérni a képzés oldalát. HTTP hiba: {course_response.status_code}")
        else:
            print(f"Nem sikerült elérni az oldalt. HTTP hiba: {response.status_code}")

```