



**Code
Academy**



1 LYGIS

15 paskaita. Informacijos traukimas iš interneto (WEB SCRAPING)



Šiandien išmoksite

01

Susipažinsite su WEB SCRAPING'u

02

Ištraukti informaciją iš norimo internetinio puslapio



```
from bs4 import BeautifulSoup
import requests

source = requests.get('https://www.delfi.lt/').text
soup = BeautifulSoup(source, 'html.parser')
blokas = soup.find('div', class_='headline')
print(blokas.prettify())
```

Kaip nuskaityti norimą svetainės bloką (pirmą)



```
from bs4 import BeautifulSoup
import requests

source = requests.get('https://www.delfi.lt/').text
soup = BeautifulSoup(source, 'html.parser')
blokas = soup.find('div', class_ = 'headline')

kategorija = blokas.find('div', class_ = 'headline-category').text.strip()

tekstas = blokas.find('a', class_ = 'CBarticleTitle').text.strip()
print(kategorija)
print(tekstas)
```

Kaip išrinkti norimą informaciją iš bloko



```
from bs4 import BeautifulSoup
import requests

source = requests.get('https://www.delfi.lt/').text
soup = BeautifulSoup(source, 'html.parser')
blokai = soup.find_all('div', class_ = 'headline')

for blokas in blokai:
    try:
        kategorija = blokas.find('div', class_ = 'headline-category').text.strip()
        tekstas = blokas.find('a', class_ = 'CBarticleTitle').text.strip()
        print(kategorija)
        print(tekstas)
    except:
        pass
```

Kaip gauti visų blokų informaciją



```
from bs4 import BeautifulSoup
import requests
import csv

source = requests.get('https://www.delfi.lt/').text
soup = BeautifulSoup(source, 'html.parser')
blokai = soup.find_all('div', class_ = 'headline')

with open("delfi naujienos.csv", "w", encoding="UTF-8", newline='') as failas:

    csv_writer = csv.writer(failas)
    csv_writer.writerow(['Kategorija', 'Tekstas'])
    for blokas in blokai:
        try:
            kategorija = blokas.find('div', class_ = 'headline-category').text.strip()
            tekstas = blokas.find('a', class_ = 'CBarticleTitle').text.strip()
            csv_writer.writerow([kategorija, tekstas])
        except:
            pass
```

Kaip įrašyti gautą informaciją į csv failą



```
import csv

from bs4 import BeautifulSoup
import requests

source = requests.get('https://shop.telialt/telefonai/?filter=brand:samsung').text
soup = BeautifulSoup(source, 'html.parser')

blokai = soup.find_all('div', class_='card card__product card--anim js-product-compare-product')

with open("Telia Samsung telefonai.csv", "w", encoding="UTF-8", newline='') as failas:
    csv_writer = csv.writer(failas)
    csv_writer.writerow(['Modelis', 'Mėnesio kaina', 'Kaina'])

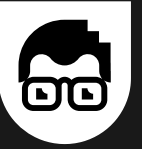
    for blokas in blokai:
        try:
            pavadinimas = blokas.find('a', class_='card__title js-product-name-truncate').text.strip()
            men_kaina = blokas.find('div', class_='pull-left').span.text.strip()
            kaina = blokas.find('span',
                                class_='price--marker').next_element.next_element.next_element.next_element.span.text.strip()
            csv_writer.writerow([pavadinimas, men_kaina, kaina])
        except:
            pass
```

Pavyzdys, kaip gauti norimo gamintojo parduodamų telefonų kainas iš svetainės



Užduotis nr. 1

Parašyti programą, kuri atspausdintu visas puslapio www.15min.lt "Redakcija rekomenduoja" skilties antraštes



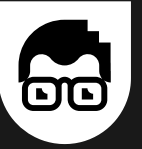
Užduotis nr. 2

Parašyti programą, kuri suskaičiuotų kiek šios dienos www.delfi.lt puslapyje yra antraštinių straipsnių



Užduotis nr. 3

Parašyti programą kuri iš puslapio <https://www.telia.lt/prekes/mobilieji-telefonai/samsung> rastų pigiausią ir brangiausią telefono modelį



Namų darbas

Užbaigti klasėje nepadarytas užduotis



Išspręsti paskaitos uždaviniai (įkelti pirmadienį)

<https://github.com/aurimas13/Python-Beginner-Course/tree/main/Programs>

SQLAlchemy

SQLAlchemy aprašymas

<https://docs.sqlalchemy.org/en/13/core/engines.html>

**Naudinga
informacija**