

Отчёт

По лр-8

**Дисциплина «Парадигмы и конструкции языков
программирования»**

Студент: Якубович Анна

Группа: ИБМ3-23Б

Преподаватель: Гапанюк Ю.Е.

Задание:

1. Осуществите разбор данных сайта с использованием библиотеки BeautifulSoup.
2. Сформируйте датасет табличного типа с использованием Python.
3. Сохраните датасет в формате CSV.

Код программы:

```
import requests

from bs4 import BeautifulSoup

import pandas as pd

url = 'http://books.toscrape.com/'

response = requests.get(url)

if response.status_code == 200:

    print("Страница успешно загружена!")

else:

    print(f"Ошибка! Код: {response.status_code}")

    exit()

soup = BeautifulSoup(response.text, 'lxml')

books = soup.find_all('article', class_='product_pod')

books_data = []

for book in books:

    title = book.h3.a['title']

    price = book.find('p', class_='price_color').text

    rating_class = book.find('p', class_='star-rating')['class']

    rating_map = {'One': 1, 'Two': 2, 'Three': 3, 'Four': 4, 'Five': 5}

    rating = rating_map.get(rating_class[1], 'N/A')

    link = url + book.h3.a['href']

    books_data.append({

        'Title': title,

        'Price': price,

        'Rating': rating,

        'Link': link

    })
```

```
df = pd.DataFrame(books_data)

print(df.head())

df.to_csv(r'C:\Users\Anna\OneDrive\Desktop\Учёба Бауманка\2 курс\Пял\лаба 4\books_dataset.csv', index=False,
encoding='utf-8')

print("Датасет успешно сохранен в файл 'books_dataset.csv'")
```

Пример результата работы программы:

```
Страница успешно загружена!
```

	Title	Price	Rating	Link
0	A Light in the Attic	Â£51.77	3	http://books.toscrape.com/catalogue/a-light-in...
1	Tipping the Velvet	Â£53.74	1	http://books.toscrape.com/catalogue/tipping-th...
2	Soumission	Â£50.10	1	http://books.toscrape.com/catalogue/soumission..
3	Sharp Objects	Â£47.82	4	http://books.toscrape.com/catalogue/sharp-obje...
4	Sapiens: A Brief History of Humankind	Â£54.23	5	http://books.toscrape.com/catalogue/sapiens-a-...

Датасет успешно сохранен в файл 'books_dataset.csv'