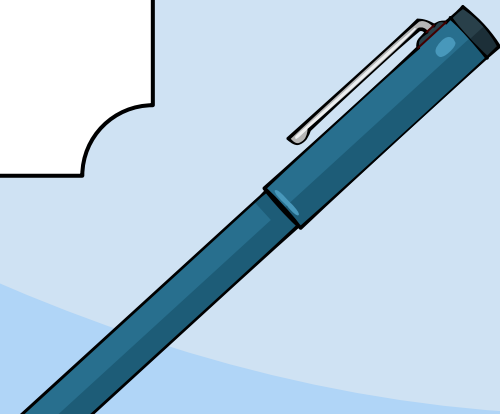


Web Scraping

(معرفی Web Scraping Tools)

سارا خرّم



مباحث جلسه گذشته (جلسه دوم)

• معرفی HTML

• HTML Tags

• HTML Elements

• HTML Attributes

سرفصل مطالب



Web Scrapping Tools معرفی •

Scrapy •

BeautifulSoup •

Selenium •

Static vs. Dynamic Webpages •



معرفی Web Scraping Tools Scrapy

Open source framework •

معایب

• نامناسب برای افراد مبتدی

مزایا

- Performance بالا
- مصرف کمتر حافظه و CPU
- Good Community

معرفی Web Scrapping Tools BeautifulSoup

- استخراج داده از فایل‌های HTML و XML
- وابسته به سایر ابزارها : کتابخانه‌های requests و urllib

• انواع parser:

• html.parser

• lxml

• lxml-xml

• html5lib



BeautifulSoup

معرفی Web Scrapping Tools

معایب

- سرعت پایین
- مناسب برای پروژه‌هایی با Scale کوچک

مزایا

- یادگیری و استفاده آسان
- User-Friendly Documentation
- Good Community

BeautifulSoup

Web Scrapping Tools معرفی

```
1 import requests
2 from bs4 import BeautifulSoup
3
4 page = requests.get('http://books.toscrape.com/')
5 soup = BeautifulSoup(page.text, 'html.parser')
6
7 links = soup.find_all('a')
8 links_list = [link.get('href') for link in links]
```

```
1 import urllib.request
2 from bs4 import BeautifulSoup
3
4 page = urllib.request.urlopen("http://books.toscrape.com/")
5 soup = BeautifulSoup(page, 'html.parser')
6
7 links = soup.find_all('a')
8 links_list = [link.get('href') for link in links]
```



معرفی Web Scrapping Tools

Browser automation tool for testing •

معایب

- سرعت پایین
- مناسب برای پروژه‌هایی با Scale کوچک

مزایا

- Load کردن JavaScript
- یادگیری و استفاده آسان

معرفی Web Scrapping Tools: جمع‌بندی



Scrapy

Vs

Selenium



Vs

BeautifulSoup

Static vs. Dynamic Webpages

Static Website



• **Static :**

• ارائه محتوای یکسان برای کاربران مختلف

• مثال: ویکی‌پدیا

• مثال: <https://www.cars.com/>

Static vs. Dynamic Webpages

• **: Dynamic**

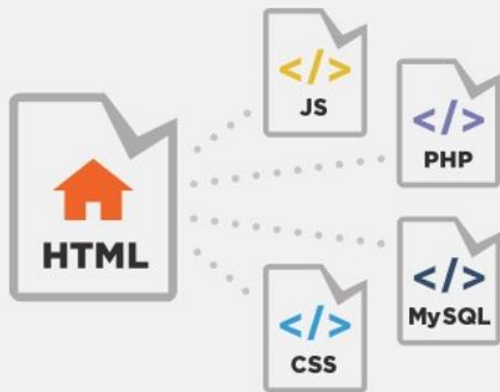
• مختصر کابر

• از JavaScript برای مدیریت داده‌ها در بخش Front End استفاده میشود.

• مثال: Google Search_Images Tab

• مثال: <https://www.pinterest.com/>

Dynamic Website



Static and Dynamic Webpages Scraping Tools



Static Webpages

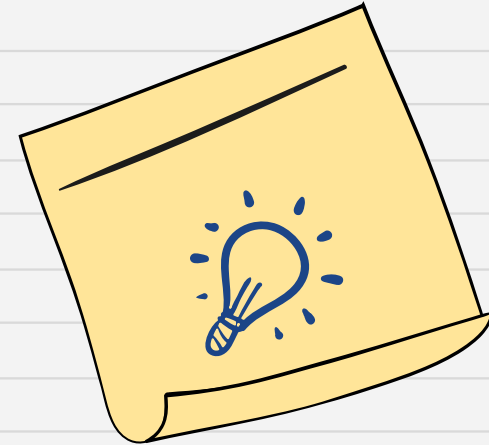
- **Scrapy Framework**
- **Requests + BeautifulSoup**
- **Selenium**

Dynamic Webpages

- **Selenium**
- **Splash**

مصادر

- <https://medium.com/analytics-vidhya/scrapy-vs-selenium-vs-beautiful-soup-for-web-scraping-24008b6c87b8>
- <https://www.bestproxyreviews.com/scrapy-vs-selenium-vs-beautifulsoup-for-web-scraping/>
- <https://towardsdatascience.com/choose-the-best-python-web-scraping-library-for-your-application-91a68bc81c4f>
- <https://www.youtube.com/watch?v=zucvHSQsKHA>
- <https://aminboutarfi.com/what-is-the-difference-between-scraping-dynamic-websites-and-static-websites/>



Thanks!

آموزش های
رایگان بیشتر

www.data-hub.ir

www.t.me/data_hub_ir

www.linkedin.com/company/data-hub-ir