



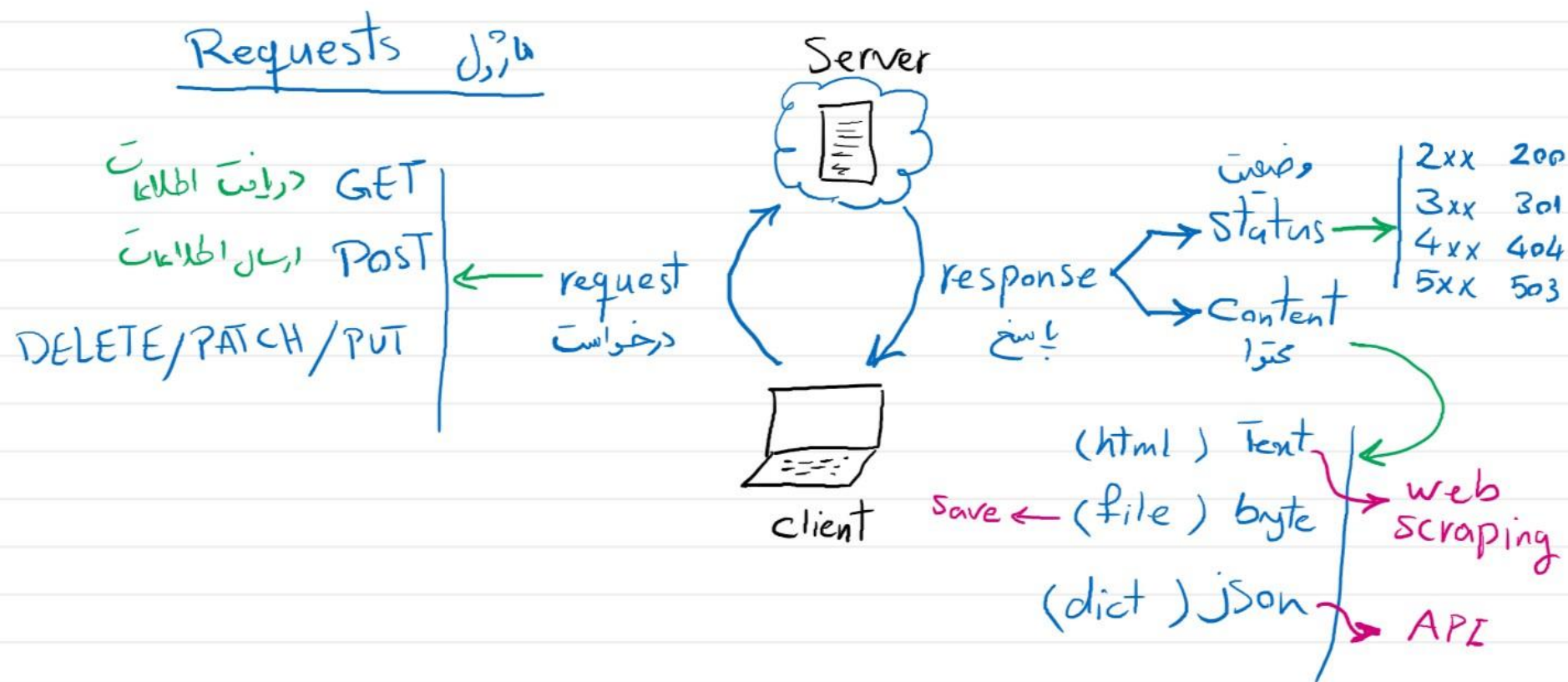
برنامه نویسی با پایتون

سعید محققى

09: Web

ارتباط Client و Server

پروتکل http



- داکيومنت رسمی ماژول Requests: <https://docs.python-requests.org/en/master/index.html>
<https://docs.python-requests.org/en/master/user/quickstart/>
- راهنمای استفاده از Requests: <https://realpython.com/python-requests/>

- سرویس تست HTTP Request & Response: <https://httpbin.org/> و <http://jsonplaceholder.typicode.com/>
- API های رایگان RapidAPI: <https://rapidapi.com/collection/list-of-free-apis>
- لیست API های ایرانی: <https://github.com/Hameds/APIs-made-in-Iran>
- لیست API های عمومی: <https://github.com/public-apis/public-apis>
- مثال:
 - ❖ آب و هوا: <https://www.metaweather.com/api/>
 - ❖ ایجاد تصویر آواتار: <https://avatars.dicebear.com/>

نکات استفاده از Requests

■ برای جلوگیری از مکث طولانی برای سایت‌های فیلتر شده در ایران!

■ `r = requests.get('https://facebook.com', timeout=2)`

```
import requests

try:
    response = requests.get('https://facebook.com', timeout=2)
except:
    print('The request timed out')
```

```
import requests
import time

is_connected = False
for i in range(3):
    try:
        response = requests.get('https://google.com', timeout=2)
        is_connected = True
        break
    except:
        is_connected = False
        time.sleep(3)
if is_connected == True:
    print(f'try {i+1}: Success')
else:
    print(f'try {i+1}: Failed')
```

■ تست اتصال اینترنت



WEB SCRAPING

استخراج اطلاعات از صفحات وب

<!DOCTYPE html>

<HTML>

<HEAD>

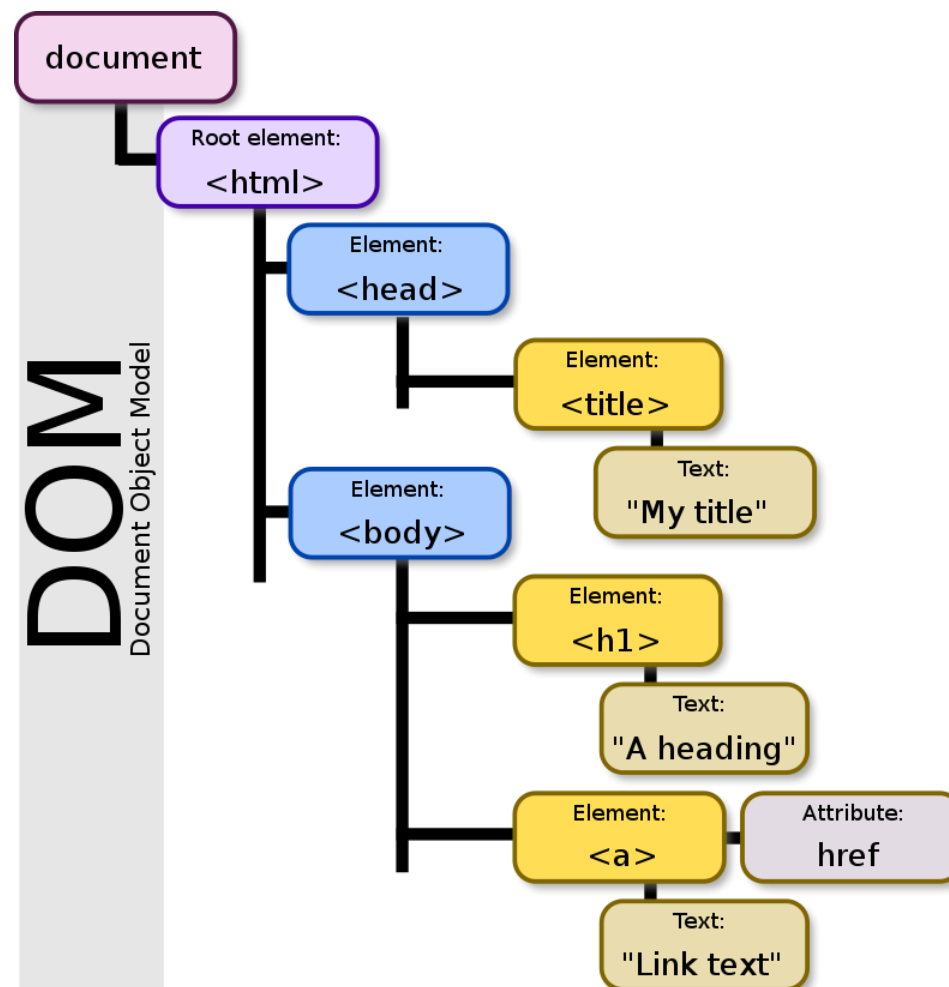
</HEAD>

<BODY>

CONTENT

</BODY>

</HTML>





- ساختار و اسکلت صفحه وب
- چیدمان المان‌ها (مکان، اندازه، جهت و ...)

HTML

- ظاهر و استایل صفحه وب
- نحوه نمایش آیتم‌ها (شکل، رنگ، فونت و ...)

CSS

https://www.w3schools.com/html/html_basic.asp

آشنایی اولیه با HTML

https://www.w3schools.com/html/html_css.asp

آشنایی اولیه با CSS

```
<html>
<body>
  <h1>Welcome to My Web Page</h1>
  <p>This is a paragraph</p>
</body>
</html>
```

Welcome to My Web Page

This is a paragraph

```
<html>
<head>
  <style>
    body {background-color: cyan;}
    h1   {color: magenta;}
    p    {color: red;font-size: 24px}
  </style>
</head>
<body>
  <h1>Welcome to My Web Page</h1>
  <p>This page has a style</p>
</body>
</html>
```

Welcome to My Web Page

This page has a style

تعریف استایل در کلاس‌های مختلف
CSS



اعمال کلاس‌ها به تگ‌های دلخواه
HTML

```
.note {  
  font-size: 120%;  
  color: red;  
}
```



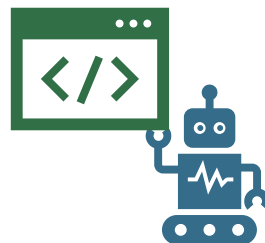
```
<h1 class="note">  
    Important Heading  
</h1>  
<p class="note">  
    Important text  
</p>
```

Important Heading

Important text

تست آنلاین

https://www.w3schools.com/html/html_classes.asp



- استخراج اطلاعات از صفحات وب توسط کد (ربات)
- شناسایی آیتم‌های مورد نظر از طریق تگ‌های HTML و کلاس‌های CSS
- ماژول BeautifulSoup
- <https://beautiful-soup-4.readthedocs.io/en/latest>

■ پروژه 10: ربات بررسی Hacker News (hacker_news_scraper)

What will we learn? (Keywords)

- Requests module
- BeautifulSoup Module
- Tqdm module
- HTML tags
- CSS classes

1. دریافت صفحات مختلف وب سایت [Hacker News](#)

2. جستجوی اطلاعات مورد نظر در صفحات مختلف

3. گزارش نتایج

4. ذخیره نتایج در یک فایل (csv یا xlsx)