



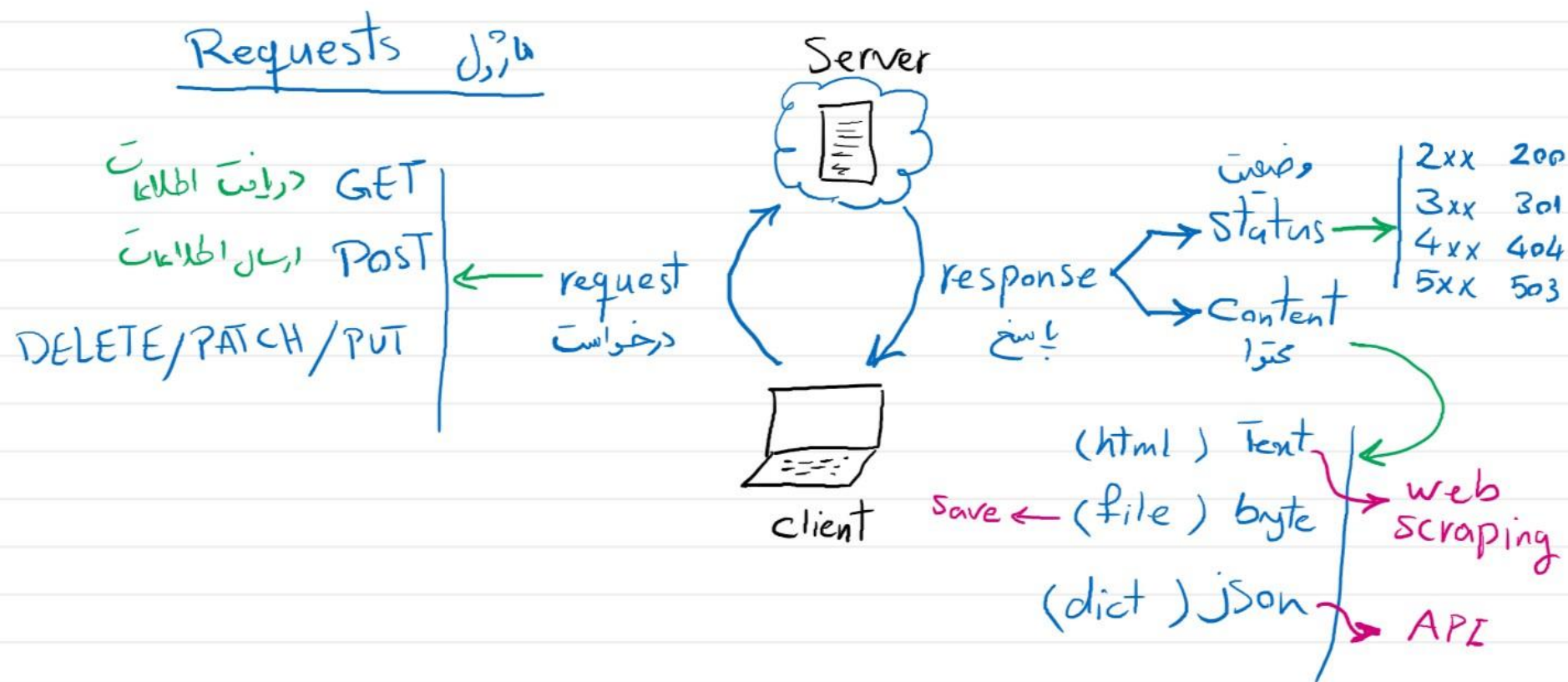
# برنامه نویسی با پایتون

سعید محقق / پاییز 1400

08 : 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/>
- API های رایگان RapidAPI : <https://rapidapi.com/collection/list-of-free-apis>
- لیست API های عمومی : <https://github.com/public-apis/public-apis>
  - ❖ تست کدنویسی : <http://jsonplaceholder.typicode.com/>
  - ❖ 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 Exception e:
    ('The request timed out')
```

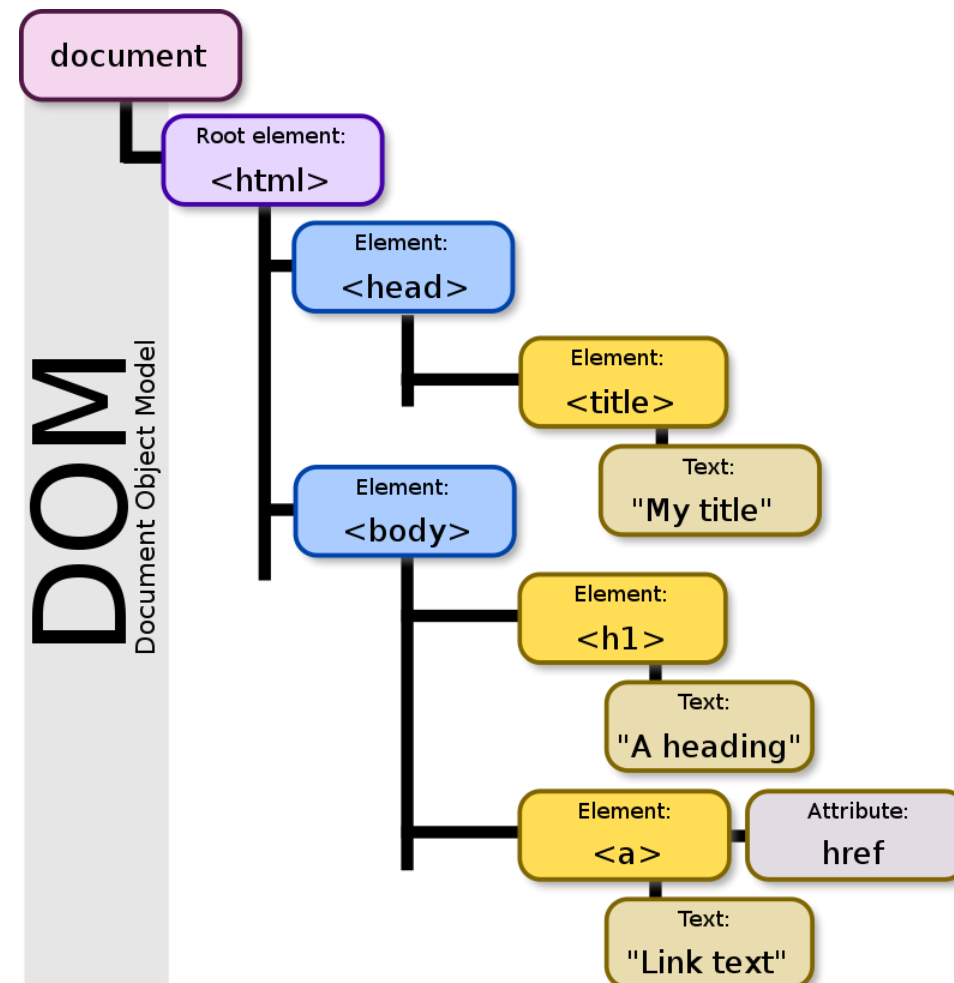
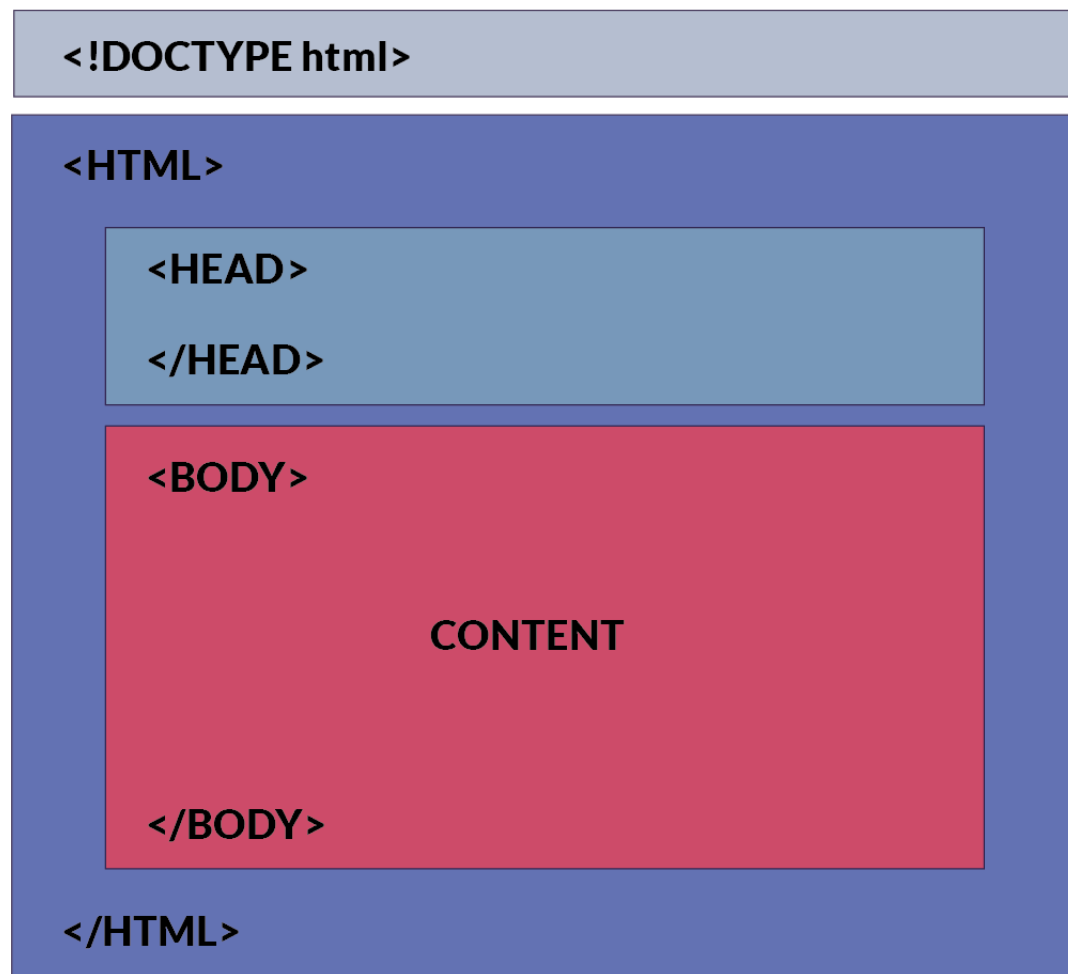
## ■ تست اتصال اینترنت

```
import requests
import time

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

# WEB SCRAPING

استخراج اطلاعات از صفحات وب







- 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

```
<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;  
}
```



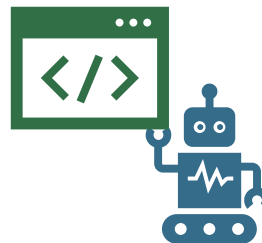
```
<h1 class="note">  
  Important Heading  
</h1>  
<p class="note">  
  Important text  
</p>
```

**Important Heading**

Important text

تست آنلاین

[https://www.w3schools.com/html/html\\_classes.asp](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 12: 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)