



PYTHON DAY - 11

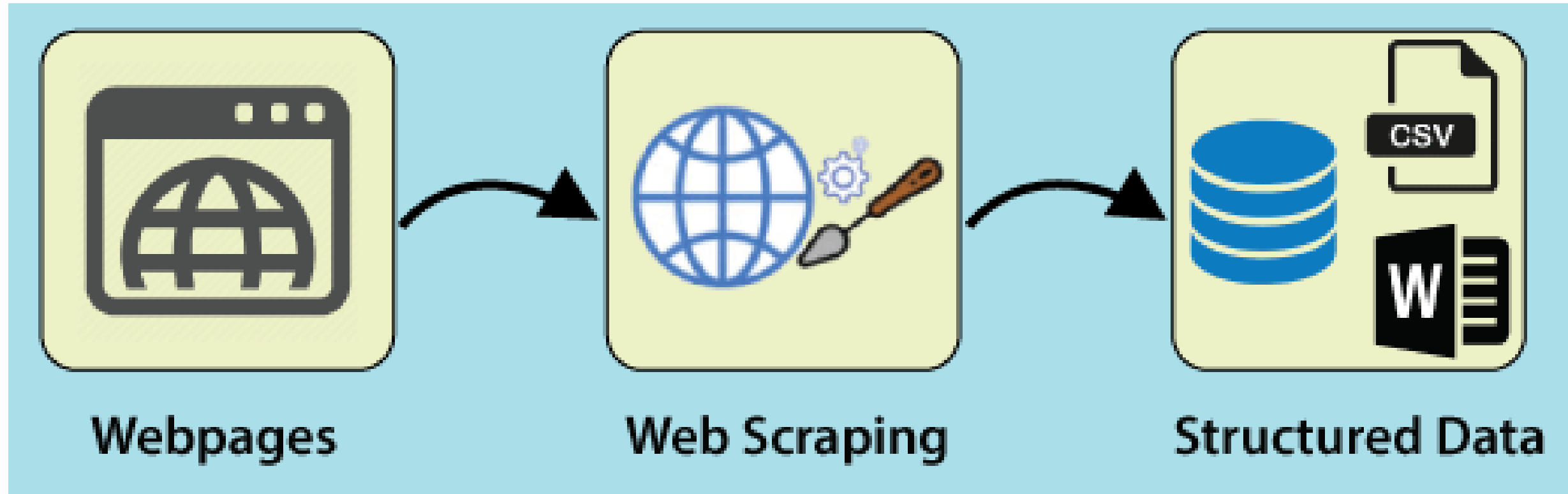




WEB SCRAPING



WEB SCRAPING



WEB SCRAPING



- Web scraping is an automatic method to obtain large amounts of data from websites.
- Most of this data is unstructured data in an HTML format which is then converted into structured data in a spreadsheet or a database so that it can be used in various applications.
- BeautifulSoup is a Python library that is used to pull data of HTML and XML files. It is mainly designed for web scrapping.

INSTALLATION



```
pip install beautifulsoup4
```

IMPORT



```
from bs4 import BeautifulSoup
```

BEAUTIFUL SOUP



```
sample = open("/content/home.html", "r")  
  
content = sample.read()  
  
soup = BeautifulSoup(content)  
  
print(soup)
```


BEAUTIFUL SOUP



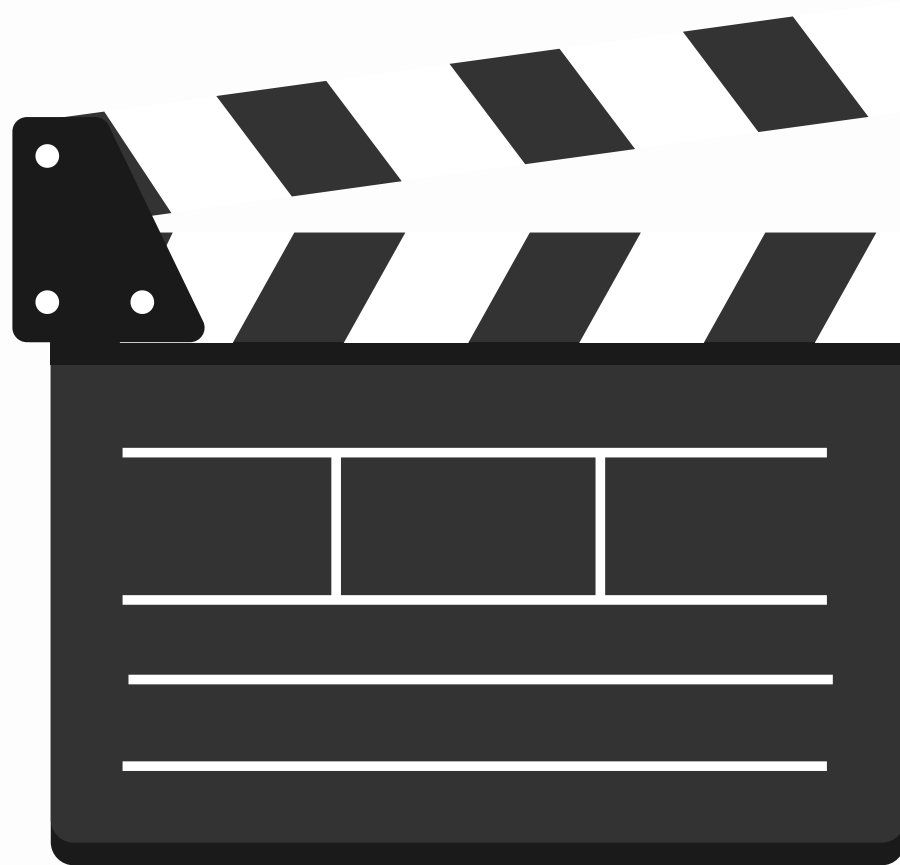
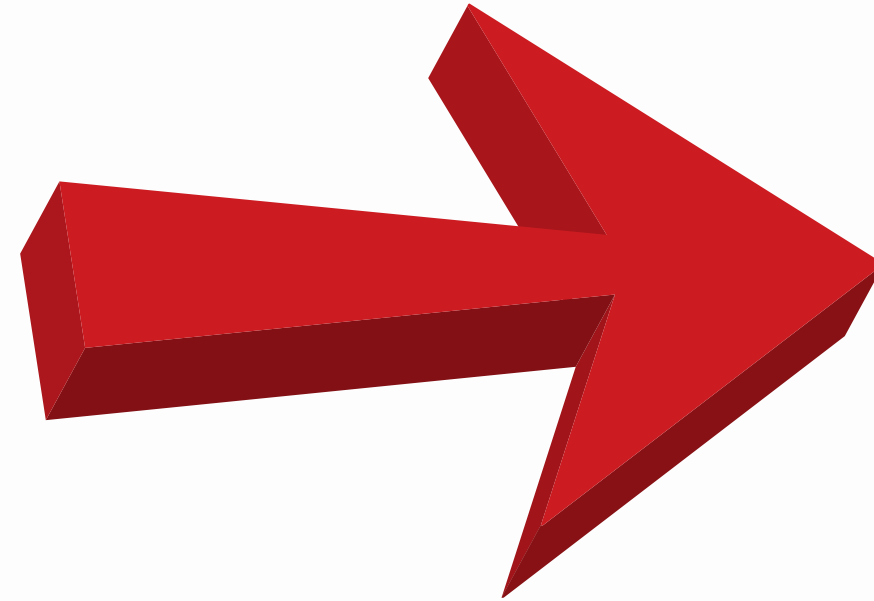
```
soup.find(tag)
```

```
soup.find_all(tag)
```

```
soup.find(tag, class_=" ")
```

```
soup.find_all(tag, class_=" ")
```


SCRAPING IMDB WEBSITE



GET HTML OF A WEBSITE



```
pip install requests
```

```
pip install lxml
```

```
import requests
```

```
import lxml
```

GET HTML CODE USING REQUESTS



```
url = ""
```

```
html = requests.get(url, parser).text
```



BEAUTIFUL SOUP FOR SCRAPPING



```
soup = BeautifulSoup(html)
```

```
print(soup.prettyfy())
```

```
divide = soup.find("div", class_=" ")
```



STORING IN CSV FILE

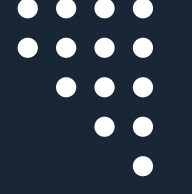


```
import csv

file = open("Movie.csv", "w")

writer = csv.writer(file)

writer.writerow(["Movie name", "year", "rating"])
```



THANKS FOR WATCHING

