AMAZON WEB SCRAPING

- Web scraping (also called web data extraction or data scraping) is an automated process that extracts data from a website and
- exports it in a structured format

In [5]:
```
#importing the necessary libraries
```

In [10]:
```
from bs4 import BeautifulSoup
import requests
import pandas as pd
import numpy as np
```

In [11]:
```
#functions to extract product title,product price,product rating,number of user reviews
```

In [12]:
```
# Function to extract Product Title
def get_title(soup):

    try:
        # Outer Tag Object
        title = soup.find("span", attrs={"id":'productTitle'})

        # Inner NavigatableString Object
        title_value = title.text

        # Title as a string value
        title_string = title_value.strip()

    except AttributeError:
        title_string = ""

    return title_string

# Function to extract Product Price
def get_price(soup):

    try:
        price = soup.find("span", attrs={'id':'priceblock_ourprice'}).string.strip()

    except AttributeError:

        try:
            # If there is some deal price
            price = soup.find("span", attrs={'id':'priceblock_dealprice'}).string.strip(

        except:
            price = ""

    return price

# Function to extract Product Rating
def get_rating(soup):

    try:
        rating = soup.find("i", attrs={'class':'a-icon a-icon-star a-star-4-5'}).string.

    except AttributeError:
        try:
            rating = soup.find("span", attrs={'class':'a-icon-alt'}).string.strip()
        except:
            rating = ""
```

```
    return rating

# Function to extract Number of User Reviews
def get_review_count(soup):
    try:
        review_count = soup.find("span", attrs={'id':'acrCustomerReviewText'}).string.st

    except AttributeError:
        review_count = ""

    return review_count

# Function to extract Availability Status
def get_availability(soup):
    try:
        available = soup.find("div", attrs={'id':'availability'})
        available = available.find("span").string.strip()

    except AttributeError:
        available = "Not Available"

    return available
```

In [13]:
```
#add the user agent,the url of web page we need to scrape on and connect a request to th
```

In [14]:
```
if __name__ == '__main__':

    # add your user agent
    HEADERS = ({'User-Agent':'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.

    # The webpage URL
    URL = "https://www.amazon.com/s?k=playstation+4&ref=nb_sb_noss_2"

    # HTTP Request
    webpage = requests.get(URL, headers=HEADERS)

    # Soup Object containing all data
    soup = BeautifulSoup(webpage.content, "html.parser")

    # Fetch links as List of Tag Objects
    links = soup.find_all("a", attrs={'class':'a-link-normal s-no-outline'})

    # Store the links
    links_list = []

    # Loop for extracting links from Tag Objects
    for link in links:
            links_list.append(link.get('href'))

    d = {"title":[], "price":[], "rating":[], "reviews":[],"availability":[]}

    # Loop for extracting product details from each link
    for link in links_list:
        new_webpage = requests.get("https://www.amazon.com" + link, headers=HEADERS)

        new_soup = BeautifulSoup(new_webpage.content, "html.parser")

        # Function calls to display all necessary product information
        d['title'].append(get_title(new_soup))
        d['price'].append(get_price(new_soup))
        d['rating'].append(get_rating(new_soup))
        d['reviews'].append(get_review_count(new_soup))
        d['availability'].append(get_availability(new_soup))
```

```
amazon_df = pd.DataFrame.from_dict(d)
amazon_df['title'].replace('', np.nan, inplace=True)
amazon_df = amazon_df.dropna(subset=['title'])
amazon_df.to_csv("amazon_data.csv", header=True, index=False)
```

In [15]:
```
#print all the details of the product
```

In [16]:
```
amazon_df
```

Out[16]:

	title	price	rating	reviews	availability
0	Sony Playstation PS4 1TB Black Console		4.7 out of 5 stars	970 ratings	Not Available
1	PlayStation 4 Slim 1TB Console		4.7 out of 5 stars	15,349 ratings	
2	PlayStation 4 500GB Console [Old Model][Discon...		4.6 out of 5 stars	13,633 ratings	Only 8 left in stock - order soon
3	PlayStation 4 Slim 1TB Limited Edition Console...		4.7 out of 5 stars	806 ratings	Only 1 left in stock - order soon
4	DualShock 4 Wireless Controller for PlayStatio...		4.7 out of 5 stars	144,503 ratings	Not Available
5	PlayStation PS5 Console – God of War Ragnarök ...		4.8 out of 5 stars	4,948 ratings	In Stock
6	2 Pack Wireless Controller for Playstation 4, ...		4.3 out of 5 stars	475 ratings	Not Available
7	TXTHcpo Wireless Controller for PS4 Remote, P4...		Previous page of related Sponsored Products		Not Available
8	TIANHOO Wireless Controller Compatible with PS...		4.3 out of 5 stars	1,152 ratings	Not Available
9	DualShock 4 Wireless Controller for PlayStatio...		4.7 out of 5 stars	144,503 ratings	Not Available
10	DualShock 4 Wireless Controller for PlayStatio...		4.7 out of 5 stars	144,503 ratings	Not Available
11	TotalMount for PlayStation 4 Pro (Mounts PS4 P...		4.5 out of 5 stars	2,733 ratings	Not Available
12	Marvel's Spider-Man: Miles Morales (PS4)		4.8 out of 5 stars	5,366 ratings	Not Available
13	Hogwarts Legacy: Standard Edition - Xbox Serie...		4.1 out of 5 stars	20 ratings	Available now
14	DOYO 1080° Gaming Racing Wheel with Pedals and...		3.9 out of 5 stars	521 ratings	Not Available
15	$100 PlayStation Store Gift Card [Digital Code]		4.7 out of 5 stars	253,563 ratings	Available now
16	TOPAD Wireless Game Controller Compatible for ...		4.0 out of 5 stars	22 ratings	Not Available
17	Sony Playstation 4 Dual Shock 4 Controller		4.6 out of 5 stars	2,060 ratings	Not Available
18	PS4 Fan Cooling Fan with Controller Charger Co...		4.3 out of 5 stars	1,257 ratings	Not Available

19	MageGee Mini 60% Gaming Keyboard, RGB Backlit ...	4.3 out of 5 stars	1,084 ratings	Not Available
20	Sony Dualshock 4 Wireless Controller for PlayS...	4.7 out of 5 stars	3,093 ratings	Not Available
21	Winshall Wireless Controller Compatible with P...	Previous page of related Sponsored Products		Not Available
22	Fantech Sonata MH90 All-Platform Gaming Headse...	4.6 out of 5 stars	21 ratings	Not Available
23	TXTHcpo Wireless Controller for PS4 Remote, P4...	Previous page of related Sponsored Products		Not Available