

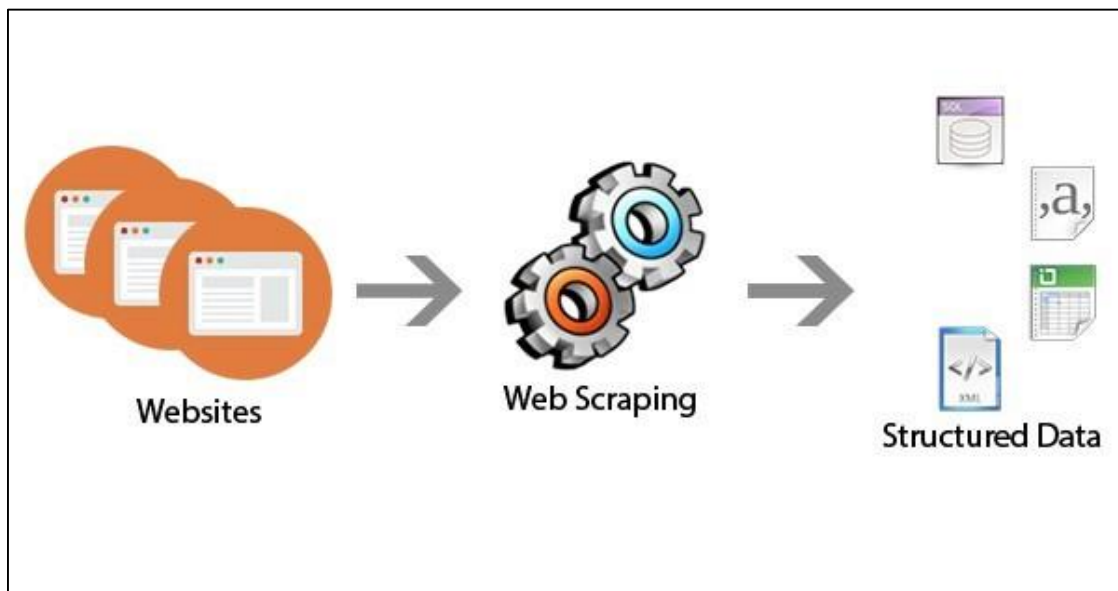
# ASSIGNMENT NO 9

**Title:** Web scraping

**Objective:** Create a review scrapper for any ecommerce website to fetch real time comments, reviews, ratings, comment tags, customer name using Python

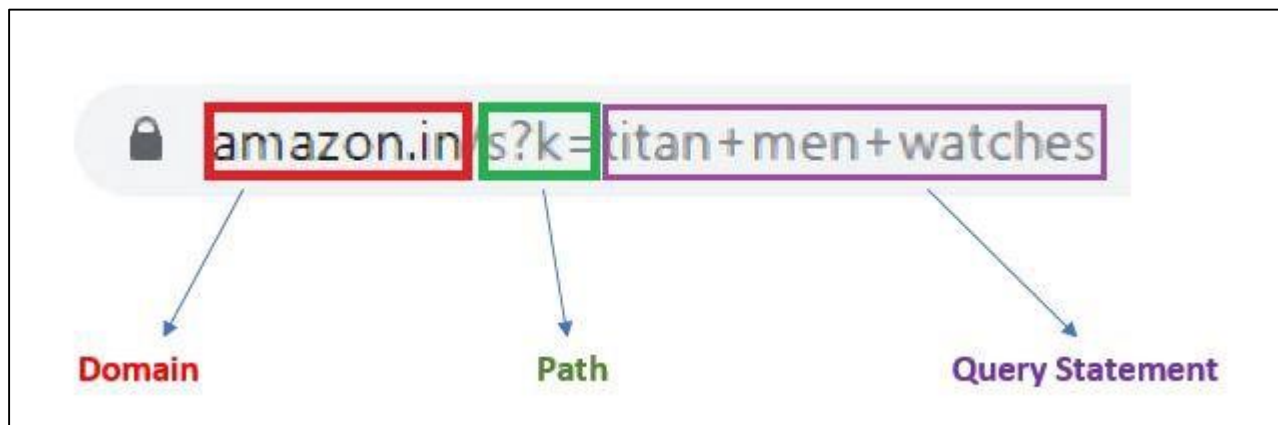
**Theory:**

Web scraping is a technique used to extract large amount of data from websites and store it in your computer . This data can be later used for analysis.



The first step is to check whether the website allows scraping or not. This can be checked by adding **robots.txt** after the website link. <https://www.amazon.in/robots.txt>

A URL can be split into five parts — **protocol, domain, path, query string & fragment**. But we will be focusing mainly on 3 parts — domain, path & query string.



## Steps

1. Get the URL of the page to be scrapped.
2. Inspect the elements of the page and identify the tags required.
3. Access the URL.
4. Get the element from the required tags.

## Code:

```
from selectorlib import Extractor

import requests

import json

from time import sleep

import csv

from dateutil import parser as dateparser

# Create an Extractor by reading from the YAML file e =
Extractor.from_yaml_file('selectors.yml')

def scrape(url):

    headers = {

        'authority': 'www.amazon.com',

        'pragma': 'no-cache',

        'cache-control': 'no-cache',

        'dnt': '1',

        'upgrade-insecure-requests': '1',

        'user-agent': 'Mozilla/5.0 (X11; CrOS x86_64 8172.45.0) AppleWebKit/537.36 (KHTML, li

        'accept': 'text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,image/apn
```

```

'sec-fetch-site': 'none',

'sec-fetch-mode': 'navigate',

'sec-fetch-dest': 'document',

'accept-language': 'en-GB,en-US;q=0.9,en;q=0.8',

}

# Download the page using requests
print("Downloading %s"%url)
r = requests.get(url, headers=headers)

# Simple check to check if page was blocked (Usually 503) if
r.status_code > 500:
    if "To discuss automated access to Amazon data please contact" in r.text: print("Page %s was
        blocked by Amazon. Please try using better proxies\n"%url)
    else:

        print("Page %s must have been blocked by Amazon as the status code was %d"%(url,r.status_code))
# Pass the HTML of the page and create
return e.extract(r.text)

# product_data = []

with open("urls.txt", 'r') as urllist, open('data.csv', 'w') as outfile:

    writer = csv.DictWriter(outfile, fieldnames=["title", "content", "date", "variant", "images",
    writer.writeheader()

    for url in urllist.readlines():

        data = scrape(url)

        if data:

            for r in data['reviews']:

                r["product"] = data["product_title"]

                r['url'] = url

                if 'verified' in r:

```

```
if 'Verified Purchase' in r['verified']:
```

```
r['verified'] = 'Yes'
```

```
else:
```

```
r['verified'] = 'Yes'
```

```
r['rating'] = r['rating'].split(' out of')[0]
```

```
date_posted = r['date'].split('on ')[-1]
```

```
if r['images']:
```

```
r['images'] = "\n".join(r['images'])
```

```
r['date'] = dateparser.parse(date_posted).strftime('%d %b %Y')
```

```
writer.writerow(r)
```

```
# sleep(5)
```

## Output:

```
In [1]: import requests
import bs4
```

```
In [2]: request1=requests.get('https://www.flipkart.com/poco-c31-royal-blue-64-gb/p/itm19effae969b86')
request1
```

```
Out[2]: <Response [200]>
```

```
In [3]: request1.content
```

```
Out[3]: b'<!doctype html><html lang="en"><head><link href="https://rukminim1.flixcart.com" rel="preconnect"><link rel="stylesheet" href="https://static-assets-web.flixcart.com/fk-p-linchpin-web/fk-cp-zion/css/app_modules.chunk.905c37.css"/><link rel="stylesheet" href="https://static-assets-web.flixcart.com/fk-p-linchpin-web/fk-cp-zion/css/app.chunk.4fb0ab.css"/><meta http-equiv="Content-type" content="text/html; charset=utf-8"><meta http-equiv="X-UA-Compatible" content="IE=Edge"><meta property="fb:page_id" content="102988293558"><meta property="fb:admins" content="658873552,624500995,10000233612389"><meta name="robots" content="noodp"><link rel="shortcut icon" href="https://www.promos/new/20150528-140547-favicon-retina.ico"/><link type="application/opensearchdescription+xml" rel="search" href="/osdd.xml?v=2"/><meta property="og:type" content="website"><meta name="og_site_name" property="og:site_name" content="Flipkart.com"/><link rel="apple-touch-icon" sizes="57x57" href="/apple-touch-icon-57x57.png"/><link rel="apple-touch-icon" sizes="72x72" href="/apple-touch-icon-72x72.png"/><link rel="apple-touch-icon" sizes="114x114" href="/apple-touch-icon-114x114.png"/><link rel="apple-touch-icon" sizes="144x144" href="/apple-touch-icon-144x144.png"/><link rel="apple-touch-icon" href="/apple-touch-icon-57x57.png"/><meta name="twitter:card" content="app"><meta name="twitter:site" content="@flipkart"><meta name="twitter:creator" content="@flipkart"><meta name="twitter:title" content="POCO C31 ( 64 GB Storage, 4 GB RAM ) Online at Best Price On Flipkart.com"><meta name="twitter:description" content="Shop for electronics, apparels & more using our Flipkart app Free shipping & COD."><meta property="twitter:image" content="http://rukmini1.flixcart.com/image/300/300/ku4ezrk0/mobile/p/e/4/c31-mzb0a0jin-poco-original-image7bzqxdhgf2n.jpeg"/><meta name="twitter:app:country" content="in"><meta name="al:ios:app_name" content="Flipkart"><meta name="al:ios:app_store_id" content="742044692"><meta name="twitter:app:name:iphone" content="Flipkart"><meta name="twitter:app:id:iphone" content="742044692"><meta name="twitter:app:url:iphone" content="http://dl.flipkart.com/dl/home?"><meta name="twitter:app:name:ipad" content="Flipkart"><meta name="twitter:app:id:ipad" content="742044692"><meta name="twitter:app:url:ipad" content="http://dl.flipkart.com/dl/home?"><meta name="twitter:app:name:googleplay" content="F
```

```
In [6]: soup = bs4.BeautifulSoup(request1.text)
        soup
```

```
Out[6]: <!DOCTYPE html>
<html lang="en"><head><link href="https://rukmini1.flixcart.com" rel="preconnect"><link href="//static-assets-web.flixcart.
com/fk-p-linchpin-web/fk-cp-zion/css/app_modules.chunk.905c37.css" rel="stylesheet"><link href="//static-assets-web.flixcar
t.com/fk-p-linchpin-web/fk-cp-zion/css/app.chunk.4fb0ab.css" rel="stylesheet"><meta content="text/html; charset=utf-8" http-
equiv="Content-type"/><meta content="IE=Edge" http-equiv="X-UA-Compatible"/><meta content="102988293558" property="fb:page_i
d"/><meta content="658873552,624500995,100000233612389" property="fb:admins"/><meta content="noodp" name="robots"/><link href
="https://www.promos/new/20150528-140547-favicon-retina.ico" rel="shortcut icon"/><link href="/osdd.xml?v=2" rel="search" ty
pe="application/opensearchdescription+xml"/><meta content="website" property="og:type"/><meta content="Flipkart.com" name="og
_site_name" property="og:site_name"/><link href="/apple-touch-icon-57x57.png" rel="apple-touch-icon" sizes="57x57"/><link hre
f="/apple-touch-icon-72x72.png" rel="apple-touch-icon" sizes="72x72"/><link href="/apple-touch-icon-114x114.png" rel="apple-t
ouch-icon" sizes="114x114"/><link href="/apple-touch-icon-144x144.png" rel="apple-touch-icon" sizes="144x144"/><link href="/a
pple-touch-icon-57x57.png" rel="apple-touch-icon"/><meta content="app" name="twitter:card"/><meta content="@flipkart" name="t
witter:site"/><meta content="@flipkart" name="twitter:creator"/><meta content="POCO C31 ( 64 GB Storage, 4 GB RAM ) Online at
Best Price On Flipkart.com" name="twitter:title"/><meta content="Shop for electronics, apparels & more using our Flipkart
app Free shipping & COD." name="twitter:description"/><meta content="http://rukmini1.flixcart.com/image/300/300/ku4ezrk0/
mobile/p/e/4/c31-mzb0a0jin-poco-original-imag7bzqxdhgf2n.jpeg" property="twitter:image"/><meta content="in" name="twitter:ap
p:country"/><meta content="Flipkart" name="al:ios:app_name"/><meta content="742044692" name="al:ios:app_store_id"/><meta cont
ent="Flipkart" name="twitter:app:name:iphone"/><meta content="742044692" name="twitter:app:id:iphone"/><meta content="http://
dl.flipkart.com/dl/home?" name="twitter:app:url:iphone"/><meta content="Flipkart" name="twitter:app:name:ipad"/><meta content
```

```
In [8]: #fetching Reviews/Comments
reviews = soup.findAll('div',{'class':'t-ZTKy'});
for review in reviews:
    print(review.get_text()+"\n\n")
```

Very nice poco phone. I am happy this product. color beautiful. sound quality nice. touch screen very good. And all good  
READ MORE

This is best budget phone.... Poco C31, and order this product because I have Poco M2 Reloaded, i was gift this phone to my sis  
ter on her birthday, she loved it Thanks Flipkart for fast delivery Thanks Poco Team  
READ MORE

Very good quality phone but camera quality is pure but okay nd very nice phoneSo beautiful display nd back cover 🧡 So I'm very  
happy 🧡🧡🧡 READ MORE

In this budget it's a good phone , everything is good... Mi is a really nice Android, it's not that slow, cameras are quite goo  
d too... Thanks to flipkart for this on time delivery and I've been using this phone from 2 weeks and it's so good so far.... P  
roduct is value for money. ❤️❤️ READ MORE

Pros : 1. Colour and design.2. Big display at low price.3. Camera as expected at this price range.4. Battery is good. Lasting a  
lmost 1.5 days.Cons : 1. Working slow.2. Ram management is not upto the mark.3. Camera is not good comparison to Redmi and Real  
me. Redmi and realme click far better than this mobile at this price.4. Not for gaming.5. Missing back cover in the package. At  
this price other companies provides back covers.  
READ MORE

The best budget phone. Camera clarity is good. Value for money. Right size phone with light weight. Over all good phone with b  
udget price.  
READ MORE

```
In [10]: #Average Overall rating
ratings = soup.find('div',{'class':'_3LWZlK'}).get_text();
print(ratings)

4.3
```

```
In [12]: #individual ratings
individual_ratings = soup.findAll('div',{'class':'_3LWZlK _1BLPMq'});
for indi_rating in individual_ratings:
    print(indi_rating.get_text()+" \n")
```

5

4

5

5

3

5

5

4

4

4

```
In [32]: #specifications
specifications = soup.find_all('td',{'class':'_1hKmbr'});
for specification in specifications:
    print (specification.get_text())
    answer = specification.findNext('td',{'class':'URwL2w'}).get_text();
    print(answer + "\n\n")

In The Box
Handset, Power Adapter, Micro-USB Cable, SIM Ejector Tool, User Guide and Warranty Card

Model Number
MZB0A0KIN

Model Name
C31

Color
Royal Blue

Browse Type
Smartphones

SIM Type
Dual Sim

Hybrid Sim Slot
No
```

Q Search this file...		
1	title	content
2	Look just like the photo	I love the shoes they are true to size , I wear a 7 1/2 but I ordered a 8 to allow a little extra room and I got just that
3	After 20 returns	Just writing a rare review on these . I love Nike's but my feet don't usually. So I've ordered and returned a lot. Tried again lol and these ARE AMAZING com
4	Half size up :)	I have only run in them a couple of times, but so far they feel great. I normally wear a size 9, and I ordered a 9.5, and I'm glad I did. There is plenty of room
5	Surprisingly comfortable	I have a very hard time finding shoes that are comfortable for me, I have severe fibromyalgia with no help from medications because nothing works for me
6	Cute and Stylish..	Great workout shoes. Very comfortable.
7	Great, Comfortable Shoe	This is my second pair. They are the most comfortable tennis shoes I own, I can be on my feet all day and they don't hurt. I normally wear a 8.5 but in this s
8	Great shoes	This is my second pair. I think these shoes are comfy and aren't hard to keep looking fresh and white. I wear them to work every day and wash them about
9	Loving these shoes	I'm really loving these shoe's. A little pricey. I'm on my feet a lot for work plus I jog. I've worn them three times now and they're very comfortable. I wear a
10	Gorgeous shoes	Gorgeous shoes! I love it. Nice for running
11	Three Stars	Look great but my feet are killing me during my walk and after my walk.
12	4GB RAM NOT 8GB RAM!	This is NOT an 8GB RAM! It is only 4GB RAM! With that being said, it is advertised as an 8GB RAM, packaging was basically triple boxes, packaged very wel
13	Great Laptop, Good Value, well worth it	Amazon delivered my order promptly. This laptop bundle came in a solid double packaging, and it comes with a backpack and a mouse pad. I am so amaz
14	Nice, but faulty!	I got a faulty laptop from day 1! Before I ordered I made sure to ask the seller if it's NEW. Well, I got it and it shuts down every 2 minutes. They offered refu
15	Pretty much everything that i was looking for in a large laptop	the box delivered on time and safe. it was double boxed and packed well. the included laptop bag is good quality and will work well when i need to use it.
16	Excellent	Excellent computer, pretty easy setup, excellent quality backpack. I'd recommend this to anyone looking for a good computer at an affordable price.
17	Nice laptop	Nice looking laptop. Very light. Quality is very good and very happy with this purchase. Got this laptop earlier than expected. Highly recommended seller a
18	To be happy!!	I have nothing to say because since I opened the Laptop, already was defected. I was very disappointed that the next day I have to return it. However, AM
19	Nice laptop bundle	It is an amazing back to school laptop bundle, HP laptop is fast and portable for my school study, and the backpack looks awesome. Overall, it is worth eve
20	Worst purchase of 2019!	Extremely dissatisfied with this laptops lack of speed and how low quality it looks and feels. It truly behaves like a 13 year old slow laptop and I have high s
21	Works very good.	Nothing to dislike about it. Nice laptop for the price.

**Conclusion :** Hence Succesfully performed assignment based on web scrapping using python.