

Web Scraping Using Python - Amazon Product Information:

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

```
In [2]: File = open("out.csv", "a")
HEADERS = ({'User-Agent': 'Mozilla/5.0 (X11; Linux x86_64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/44.0.2403.157 Safari/537.36',
            'Accept-Language': 'en-US, en;q=0.5'})
```

```
In [3]: URL = "https://www.amazon.com/dp/B0BYV51VBN/ref=sspa_dk_detail_3?pd_rd_i=B0BYV51VBN&pd_rd_w=V7rwe&content-id=amzn1.sym.386c274b-4a11-4c11-8030-000000000000"
webpage = requests.get(URL, headers=HEADERS)
```

```
In [4]: # Making the HTTP Request
webpage = requests.get(URL, headers=HEADERS)
```

```
In [5]: # Creating the Soup Object containing all data
soup = BeautifulSoup(webpage.content, "lxml")
```

```
In [8]: try:
        title = soup.find("span", attrs={"id": 'productTitle'})
        title_value = title.string
        title_string = title_value.strip().replace(',', ' ')

        except AttributeError:
            title_string = "NA"

        print("Product Title = ", title_string)
```

Product Title = NA

```
In [ ]: title = soup.find('span', attrs={'id': 'productTitle'}).get_text().strip()
```

```
In [9]: # saving the title in the file
File.write(f"{title_string},")
```

```
Out[9]: 3
```

```
In [10]: # retrieving price
try:
    price = soup.find("span", attrs={'id': 'priceblock_ourprice'}).string.strip().replace(',','')
    # we are omitting unnecessary spaces
    # and commas form our string
except AttributeError:
    price = "NA"
print("Products price = ", price)
```

Products price = NA

```
In [11]: # saving
File.write(f"{price},")
```

Out[11]: 3

```
In [12]: # retrieving product rating
try:
    rating = soup.find("i", attrs={'class': 'a-icon a-icon-star a-star-4-5'}).string.strip().replace(',','')

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

Overall rating = NA

```
In [13]: File.write(f"{rating},")
```

Out[13]: 3

```
In [14]: try:
    review_count = soup.find("span", attrs={'id': 'acrCustomerReviewText'}).string.strip().replace(',','')

except AttributeError:
    review_count = "NA"

print("Total reviews = ", review_count)
```

Total reviews = NA

```
In [15]: File.write(f"{review_count},")
```

```
Out[15]: 3
```

```
In [16]: # print availability status
try:
    available = soup.find("div", attrs={'id': 'availability'})
    available = available.find("span").string.strip().replace(',',' ')

except AttributeError:
    available = "NA"
print("Availability = ", available)

Availability = NA
```

```
In [17]: # saving the availability and closing the line
File.write(f"{available},\n")
```

```
Out[17]: 4
```

```
In [18]: # closing the file
File.close()
```

```
In [ ]: if __name__ == '__main__':
# opening our url file to access URLs
    file = open("url.txt", "r")

# iterating over the urls
for links in file.readlines():
    main(links)
```