

Institute of Computer Technology
B. Tech Computer Science and Engineering
Sub: Data Mining and Warehousing (2CSE60E27)

PRACTICAL 1: WEB SCRAPPING USING BEAUTIFULSOUP

Suppose you wanted to buy a product online (example. a mobile phone or Television). So, for that you will check the e-commerce website from where you wanted to buy. As an alert customer, you will be checking all the items available for different brands and specifications, varying in cost. All those details you need to go through one by one for all the products to select the best one. It will take loss of your time and efforts. So now, to ease your work, extract all the information from website and store that details in the tabular format or in an excel file. (Example: Storing all the details of phone like brand name, specification and price in a table) So that it will be easy to analyze.

Note: You can choose any product of your choice and e-commerce website as well.

<http://localhost:8888/notebooks/Desktop/dishwa/dmw/Practical%201/dishwa%20dmw%201.ipynb>

Code:

```
import bs4
from bs4 import BeautifulSoup
from urllib.request import urlopen

scrap_page="https://webscraper.io/test-sites/e-commerce/allinone/computers/laptops"
fetch_url=urlopen(scrap_page)
page_html=fetch_url.read()
soup=BeautifulSoup(page_html)
print(soup.prettify)

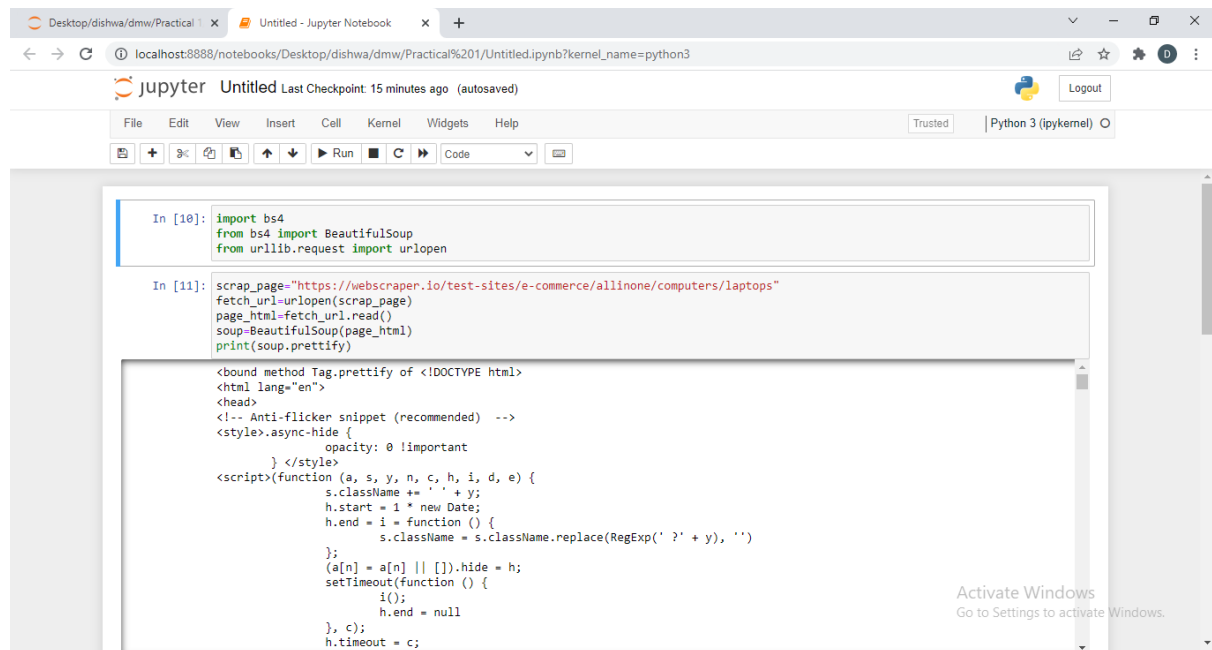
titles=[]
for x in soup.find_all('a',{'class':'title'}):
    titles.append(x.text.strip())
prices=[]
for x in soup.find_all('h4',{'class':'pull-right price'}):
    prices.append(x.text.strip())
description=[]
for x in soup.find_all('p',{'class':'description'}):
    description.append(x.text.strip())
reviews=[]
for x in soup.find_all('div',{'class':'ratings'}):
    reviews.append(x.text.strip())

import pandas as pd
```

```
scrap_data=pd.DataFrame({"Name of the
Product":titles,"Price":prices,"description":description,"reviews":reviews})
file_name="Scrap_data.xlsx"
scrap_data.to_excel(file_name)
print("File created Successfully")
```

```
df = pd.read_excel ('Scrap_data.xlsx')
print (df)
```

Output:

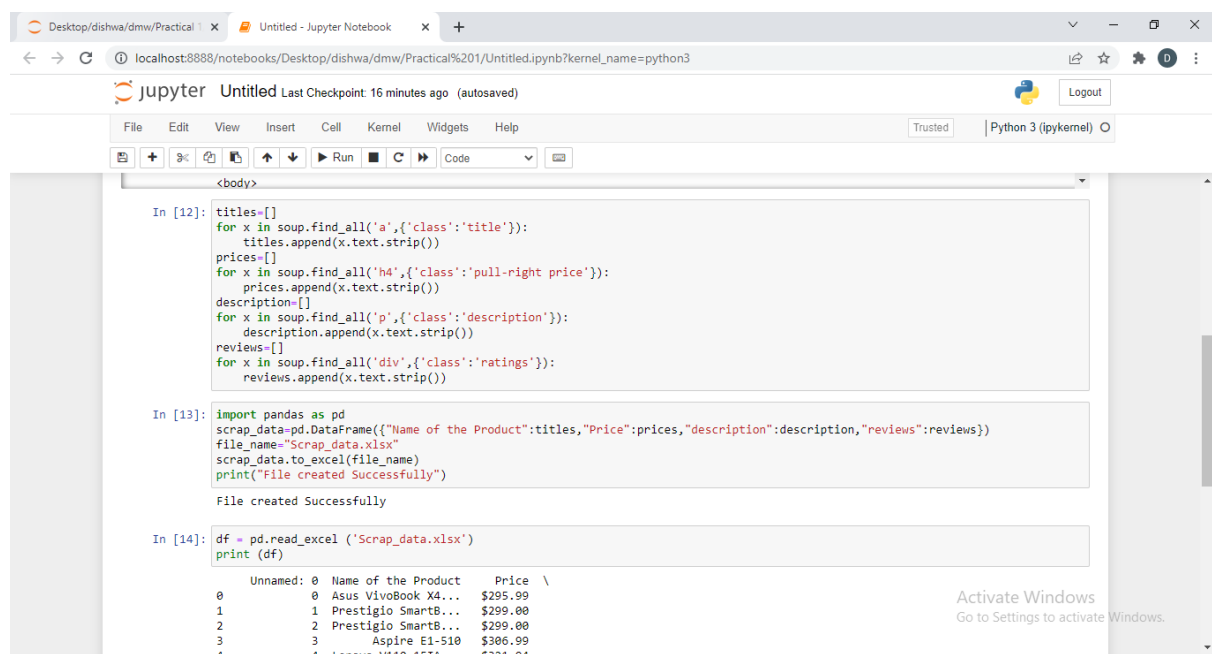


```
In [10]: import bs4
from bs4 import BeautifulSoup
from urllib.request import urlopen

In [11]: scrap_page="https://webscraper.io/test-sites/e-commerce/allinone/computers/laptops"
fetch_url=urlopen(scrap_page)
page_html=fetch_url.read()
soup=BeautifulSoup(page_html)
print(soup.prettify())

<bound method Tag.prettify of <!DOCTYPE html>
<html lang="en">
<head>
<!-- Anti-flicker snippet (recommended) -->
<style>.async-hide {
  opacity: 0 !important
}</style>
<script>(function (a, s, y, n, c, h, i, d, e) {
  s.className += ' ' + y;
  h.start = 1 * new Date;
  h.end = i = function () {
    s.className = s.className.replace(RegExp(' ?' + y), '');
  };
  (a[n] = a[n] || []).hide = h;
  setTimeout(function () {
    i();
    h.end = null
  }, c);
  h.timeout = c;

```



```
In [12]: titles=[]
for x in soup.find_all('a',{'class':'title'}):
    titles.append(x.text.strip())
prices=[]
for x in soup.find_all('h4',{'class':'pull-right price'}):
    prices.append(x.text.strip())
description=[]
for x in soup.find_all('p',{'class':'description'}):
    description.append(x.text.strip())
reviews=[]
for x in soup.find_all('div',{'class':'ratings'}):
    reviews.append(x.text.strip())

In [13]: import pandas as pd
scrap_data=pd.DataFrame({"Name of the Product":titles,"Price":prices,"description":description,"reviews":reviews})
file_name="Scrap_data.xlsx"
scrap_data.to_excel(file_name)
print("File created Successfully")

File created Successfully

In [14]: df = pd.read_excel ('Scrap_data.xlsx')
print (df)

Unnamed: 0  Name of the Product  Price \
0          0  Asus VivoBook X4...  $295.99
1          1  Prestigio SmartB...  $299.00
2          2  Prestigio SmartB...  $299.00
3          3  Aspire E1-510      $386.99
4          4  Asus VivoBook X4...  $331.99
```

Desktop/dishwa/dmw/Practical x Untitled - Jupyter Notebook x +

localhost:8888/notebooks/Desktop/dishwa/dmw/Practical%201/Untitled.ipynb?kernel_name=python3

jupyter Untitled Last Checkpoint: 16 minutes ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

```
In [14]: df = pd.read_excel('Scrap_data.xlsx')
print(df)
```

Unnamed: 0	Name of the Product	Price	
0	Asus VivoBook X4...	\$295.99	
1	Prestigio SmartB...	\$299.00	
2	Prestigio SmartB...	\$299.00	
3	Aspire E1-510	\$306.99	
4	Lenovo V110-15IA...	\$321.94	
...	
112	Lenovo Legion Y7...	\$1399.00	
113	Asus ROG Strix G...	\$1399.00	
114	Asus ROG Strix G...	\$1769.00	
115	Asus ROG Strix G...	\$1769.00	
116	Asus ROG Strix S...	\$1799.00	

	description	reviews
0	Asus VivoBook X441NA-GA190 Chocolate Black, 14...	14 reviews
1	Prestigio SmartBook 133S Dark Grey, 13.3" FHD ...	8 reviews
2	Prestigio SmartBook 133S Gold, 13.3" FHD IPS, ...	12 reviews
3	15.6", Pentium N3520 2.16GHz, 4GB, 500GB, Linux	2 reviews
4	Lenovo V110-15IAP, 15.6" HD, Celeron N3350 1.1...	5 reviews
...
112	Lenovo Legion Y720, 15.6" FHD IPS, Core i7-770...	8 reviews
113	Asus ROG Strix GL702VM-GC146T, 17.3" FHD, Core...	10 reviews
114	Asus ROG Strix GL702ZC-GC154T, 17.3" FHD, Ryz...	7 reviews
115	Asus ROG Strix GL702ZC-GC209T, 17.3" FHD IPS, ...	8 reviews
116	Asus ROG Strix SCAR Edition GL503VM-ED115T, 15...	8 reviews

[117 rows x 5 columns]

Activate Windows
Go to Settings to activate Windows.

Scrap_data - Excel (Product Activation Failed)

File Home Insert Page Layout Formulas Data Review View Tell me what you want to do...

Clipboard Font Alignment Number Styles Cells Editing

	A	B	C	D	E	F
		Name of the Product	Price	description	reviews	
2	0	Asus VivoBook X4...	\$295.99	Asus VivoBook X441NA-GA190 Chocolate Black, 14", Celeron N3450, 4GB, 128GB SSD, Endless OS, ENG kbd	14 reviews	
3	1	Prestigio SmartB...	\$299.00	Prestigio SmartBook 133S Dark Grey, 13.3" FHD IPS, Celeron N3350 1.1GHz, 4GB, 32GB, Windows 10 Pro + Office 365 1 gadam	8 reviews	
4	2	Prestigio SmartB...	\$299.00	Prestigio SmartBook 133S Gold, 13.3" FHD IPS, Celeron N3350 1.1GHz, 4GB, 32GB, Windows 10 Pro + Office 365 1 gadam	12 reviews	
5	3	Aspire E1-510	\$306.99	15.6", Pentium N3520 2.16GHz, 4GB, 500GB, Linux	2 reviews	
6	4	Lenovo V110-15IA...	\$321.94	Lenovo V110-15IAP, 15.6" HD, Celeron N3350 1.1GHz, 4GB, 128GB SSD, Windows 10 Home	5 reviews	
7	5	Lenovo V110-15IA...	\$356.49	Asus VivoBook 15 X540NA-GQ008T Chocolate Black, 15.6" HD, Pentium N4200, 4GB, 500GB, Windows 10 Home, En kbd	6 reviews	
8	6	Hewlett Packard...	\$364.46	Hewlett Packard 250 G6 Dark Ash Silver, 15.6" HD, Celeron N3060 1.6GHz, 4GB, 128GB SSD, DOS	12 reviews	
9	7	Acer Aspire 3 A3...	\$372.70	Acer Aspire 3 A315-31 Black, 15.6" HD, Celeron N3350 1.1GHz, 4GB, 128GB SSD, Windows 10 Home	2 reviews	
10	8	Acer Aspire A315...	\$379.94	Acer Aspire A315-31-C33J Black 15.6", HD, Celeron N3350, 4GB DDR3L, 128GB, Windows 10 Home, ENG	0 reviews	
11	9	Acer Aspire ES1...	\$379.95	Acer Aspire ES1-572 Black, 15.6" HD, Core i3-6006U, 4GB, 128GB SSD, Linux	9 reviews	
12	10	Acer Aspire 3 A3...	\$391.48	Acer Aspire 3 A315-31 Black, 15.6" HD, Pentium N4200 1.1GHz, 4GB, 128GB SSD, Windows 10 Home	10 reviews	
13	11	Acer Aspire 3 A3...	\$393.88	Acer Aspire 3 A315-21, 15.6", AMD A4-9120. 4GB. 128GB SSD, Linux	9 reviews	
14	12	Asus VivoBook Ma...	\$399.00	Asus VivoBook Max X541NA-GQ041 Black Chocolate, 15.6" HD, Pentium N4200 1.1GHz, 4GB, 500GB, Windows 10 Home	4 reviews	
15	13	Asus VivoBook E5...	\$399.99	Asus VivoBook E502NA-GO022T Dark Blue, 15.6" HD, Pentium N4200 1.1GHz, 4GB, 128GB SSD, Windows 10 Home, En/Ru kbd	3 reviews	
16	14	Lenovo ThinkPad...	\$404.23	Lenovo ThinkPad E31-80, 13.3" HD, Celeron 3855U 1.6GHz, 4GB, 128GB SSD, Windows 10 Home	12 reviews	
17	15	Acer Aspire 3 A3...	\$408.98	Acer Aspire 3 A315-31 Black, 15.6" HD, Pentium N4200 1.1GHz, 4GB, 128GB SSD, Windows 10 Home	10 reviews	
18	16	Lenovo V110-15IS...	\$409.63	Lenovo V110-15ISK, 15.6" HD, Core i3-6006U, 8GB, 128GB SSD, Windows 10 Home	9 reviews	
19	17	Acer Aspire ES1...	\$410.46	Acer Aspire ES1-732 Black, 17.3" HD+, Celeron, N3350, 4GB, 1TB, Windows 10 Home	14 reviews	
20	18	Asus VivoBook 15...	\$410.66	Asus VivoBook 15 X540NA-GQ026T Chocolate Black, 15.6" HD, Pentium N4200, 4GB, 128GB SSD, Windows 10 Home, En/Ru kbd	4 reviews	
21	19	Packard 255 G2	\$416.99	15.6", AMD E2-3800 1.3GHz, 4GB, 500GB, Windows 8.1	2 reviews	
22	20	Asus EeeBook R41...	\$433.30	Asus EeeBook R416NA-FA014T, 14" FHD, Pentium N4200, 4GB, 128GB eMMC, Windows 10 Home, Eng kbd	1 reviews	
23	21	Acer Aspire 3 A3...	\$436.29	Acer Aspire 3 A315-51, 15.6" HD, Core i3-6006U, 4GB, 1TB, Windows 10 Home	1 reviews	

Sheet1

Ready

Activate Windows
Go to Settings to activate Windows.