

Practical No – 7

Title – Build the web crawler to pull product information and links from an e-commerce website. (Python)

Program :

```
import requests
```

```
from bs4 import BeautifulSoup
```

```
def fetch_urls(url):
```

```
    try:
```

```
        response = requests.get(url)
```

```
        response.raise_for_status() # Raise an error for bad responses
```

```
        webpage = response.text
```

```
        soup = BeautifulSoup(webpage, 'html.parser')
```

```
        links = []
```

```
        # Find all anchor tags and extract their href attributes
```

```
        for a in soup.find_all('a', href=True):
```

```
            href = a['href']
```

```
            if href.startswith('http://') or href.startswith('https://'):
```

```
                links.append(href)
```

```
        return links
```

```
    except Exception as e:
```

```
        print(f"Exception caught: {e}")
```

```
    return []
```

```
if __name__ == "__main__":
```

```
    url = "https://www.amazon.in/?tag=googmantxtmob170-21&ascsubtag=_k_EAlaIQobChMIofD-g7TsiAMVZ6RmAh0fPSrNEAAYASAAEgJdJvD_BwE_k_"
```

```
    urls = fetch_urls(url)
```

for u in urls:

print(u)

print(f"Total URLs Fetched are {len(urls)}")

OUTPUT :

```
PS C:\Users\admin\Desktop\ISR> python ISR7.py
https://www.amazon.in/ap/signin?openid.pape.max_auth_age=0&openid.return_to=https%3A%2F%2Fwww.amazon.in%2F%3F_encoding%3DUTF8%26ref_%3Dnav_ya_signin&openid.identity=http%3A%2F%2Fspecs.openid.net%2Fauth%2F2.0%2Fidentifier_select&openid.assoc_handle=inflex&openid.mode=checkid_setup&openid.claimed_id=http%3A%2F%2Fspecs.openid.net%2Fauth%2F2.0%2Fidentifier_select&openid.ns=http%3A%2F%2Fspecs.openid.net%2Fauth%2F2.0
https://www.amazon.in/s?_encoding=UTF8&i=apparels&bbn=100443351031&rh=n%3A1068120031%2Cp_36%3A59000_%2Cp_85%3A10440599031%2Cp_n_pct-off-with-tax%3A2665401031&s=apparels&dc=&hidden-keywords=-woman-women-boy-girl-kid&qid=1726813166&rnid=2665390031&ref=sr_nr_p_n_pct-off-with-tax_4&ds=v1%3A8gmH8KJGza4M92F21Z6kPpVA7OmIfWdkJ4zQkTiub%&pd_rd_w=51Sup&content-id=amzn1.sym.66e21b5a-c704-4638-9291-e1a98f6a43db&pf_rd_p=66e21b5a-c704-4638-9291-e1a98f6a43db&pf_rd_r=1064XRMD2581F8287E0F&pd_rd_wg=MudXZ&pd_rd_r=0d960dc5-c7dc-4f18-bc58-212c32f3c385&ref=pd_hp_d_btf_unk
https://www.amazon.in/s?_encoding=UTF8&i=fashion&bbn=100766813031&rh=n%3A100766813031%2Cp_36%3A75000_%2Cp_85%3A10440599031%2Cp_n_pct-off-with-tax%3A2665401031&s=exact-aware-popularity-rank&hidden-keywords=-Dhoti-Lungi-8075R81P75-B09C2351M4&pd_rd_w=51Sup&content-id=amzn1.sym.66e21b5a-c704-4638-9291-e1a98f6a43db&pf_rd_p=66e21b5a-c704-4638-9291-e1a98f6a43db&pf_rd_r=1064XRMD2581F8287E0F&pd_rd_wg=MudXZ&pd_rd_r=0d960dc5-c7dc-4f18-bc58-212c32f3c385&ref=pd_hp_d_btf_unk
https://www.amazon.in/s?_encoding=UTF8&i=apparels&bbn=100431596031&rh=n%3A1968024031%2Cp_85%3A10440599031%2Cp_n_pct-off-with-tax%3A2665401031&s=apparels&dc=&ds=v1%3Aya1Wlu0%2FTZe70IseE9HwPF5BaOn5DTNju853QFxnHQ&qid=1722497966&rnid=2665390031&ref=sr_nr_p_n_pct-off-with-tax_4&pd_rd_w=51Sup&content-id=amzn1.sym.66e21b5a-c704-4638-9291-e1a98f6a43db&pf_rd_p=66e21b5a-c704-4638-9291-e1a98f6a43db&pf_rd_r=1064XRMD2581F8287E0F&pd_rd_wg=MudXZ&pd_rd_r=0d960dc5-c7dc-4f18-bc58-212c32f3c385&ref=pd_hp_d_btf_unk
https://www.amazon.in/s?_encoding=UTF8&i=apparels&bbn=100361673031&rh=n%3A100361673031%2Cp_85%3A10440599031%2Cp_n_pct-off-with-tax%3A2665402031&s=apparels&dc=&qid=1722924233&rnid=4595883031&ref=sr_nr_p_36_0_0&pd_rd_w=51Sup&content-id=amzn1.sym.66e21b5a-c704-4638-9291-e1a98f6a43db&pf_rd_p=66e21b5a-c704-4638-9291-e1a98f6a43db&pf_rd_r=1064XRMD2581F8287E0F&pd_rd_wg=MudXZ&pd_rd_r=0d960dc5-c7dc-4f18-bc58-212c32f3c385&ref=pd_hp_d_btf_unk
https://www.aboutamazon.in/?utm_source=gateway&utm_medium=footer
https://amazon.jobs
https://press.aboutamazon.in/?utm_source=gateway&utm_medium=footer
https://www.amazon.science
https://www.amazon.in/gp/redirect.html/ref=footer_fb?location=http://www.facebook.com/AmazonIN&token=2075D5EAC788214089728E2183FD391706D41E9486
https://www.amazon.in/gp/redirect.html/ref=footer_twitter?location=http://twitter.com/AmazonIN&token=A309DFBFCB1E37A808FF531934855DC817F130B686
https://www.amazon.in/gp/redirect.html?location=http://www.instagram.com/amazondotin&token=264882C912E90095CB1D9B61F12E125D5DF9BFC7&source=standards
https://accelerator.amazon.in/?ref=_map_1_b2b_GW_FT
https://brandservices.amazon.in/?ref=AOINABRLGNRFOOT&ld=AOINABRLGNRFOOT
https://sell.amazon.in/grow-your-business/amazon-global-selling.html?ld=AZIN_Footer_V1&ref=AZIN_Footer_V1
https://supply.amazon.com/?ref=footer_sta&lang=en-IN
https://affiliate-program.amazon.in/?utm_campaign=assocshowcase&utm_medium=footer&utm_source=GW&ref=footer_assoc
https://services.amazon.in/services/fulfillment-by-amazon/benefits.html/ref=az_footer_fba?ld=AWRGINFBAfooter
https://advertising.amazon.in/?ref=Amz.in
https://www.amazonpay.in/merchant
https://www.amazon.in/your-product-safety-alerts?ref=footer_bsx_ypsa
https://www.abebooks.com/
https://aws.amazon.com/what-is-cloud-computing/?sc_channel=EL&sc_campaign=IN_amazonfooter
https://www.audible.in/
https://www.imdb.com/
https://www.shopbop.com/
Total URLs Fetched are 26
PS C:\Users\admin\Desktop\ISR> []
```