

Auto File Downloader using Web Scrapping

Automation script which Extracts URL from specific web site using Web Scrapping and then download that files.

```
1 import os
2 import re
3 import bs4
4 import requests
5 from sys import *
6 from urllib.parse import urlparse
7 from urllib.request import urlopen
8
9 def is_downloadable(url):
10     h = requests.head(url, allow_redirects=True)
11     header = h.headers
12     content_type = header.get('content-type')
13     if 'text' in content_type.lower():
14         return False
15     if 'html' in content_type.lower():
16         return False
17     return True
18
19 def get_filename_from_cd(cd):
20     a = urlparse(cd)
21     return os.path.basename(a.path)
22
23 def MarvellousDownload(url):
24     url = "http:" + url
25     allowed = is_downloadable(url)
26     if allowed:
27         try:
28             res = requests.get(url, allow_redirects=True)
29             res.raise_for_status()
30             filename = get_filename_from_cd(url)
31             fd = open(filename, "wb")
32
33             for buffer in res.iter_content(1024):
34                 fd.write(buffer)
35
36             fd.close()
37             return True
38
39         except Exception as E:
40             return False
41
42 def MarvellousWebScrapper(url):
43     try:
44         connection = urlopen(url)
45
46         raw_html = connection.read()
47
48         connection.close()
49
50         page_soup = bs4.BeautifulSoup(raw_html, "html.parser")
```

```
51
52     container = page_soup.findAll("div",{ "class":"item-container"})
53
54     for elements in container:
55         ret = MarvellousDownload(elements.a.img['data-src'])
56         if ret == False:
57             break
58
59     return ret
60
61 except Exception:
62     return False;
63
64 def main():
65     print("---- Marvellous Infosystems by Piyush Khairnar-----")
66
67     print("Application name : " +argv[0])
68
69     if (len(argv) == 2):
70         if (argv[1] == "-h") or (argv[1] == "-H"):
71             print("This Script is used to download file")
72             exit()
73
74         if (argv[1] == "-u") or (argv[1] == "-U"):
75             print("usage : ApplicationName AbsolutePath_of_Directory Extention")
76             exit()
77
78     url = "https://www.newegg.com/Video-Cards-Video-Devices/Category/ID-38?Tpk=video%20card"
79
80     try:
81         bret = MarvellousWebScrapper(url)
82         if bret == True:
83             print("Files Successfully downloaded")
84         else:
85             print("File Successfully downloaded")
86
87     except Exception:
88         print("Failed to download")
89
90 if __name__ == "__main__":
91     main()
```