

# scrape

April 4, 2022

```
[ ]: #Mitre Crawler
      #CVE Details Crawler
      import requests
      from bs4 import BeautifulSoup
      from tqdm import tqdm
      import pandas as pd
      url = 'https://www.cvedetails.com/vulnerability-list/'

      #initial Request
      first_page = requests.get(url).text
      #Parse initial Page
      soup = BeautifulSoup(first_page, 'html.parser')
```

```
[ ]: #How many vulnerabilities are on cvedetails.com?
      vulnerabilities = soup.select('div[class="paging"]')[1].select('b')[0].getText()
      print('There are',vulnerabilities,'known vulnerabilities on cvedetails.com')
```

There are 173313 known vulnerabilities on cvedetails.com

```
[ ]: #How many pages are on cvedetails.com?
      pages = len(soup.select('div[class="paging"]')[1].select('a'))
      print('There are',pages,'pages on cvedetails.com')
      #Compile Link List
      link_list = []
      base_url = '/'.join(url.split('/')[0:3])
      for url in soup.select('div[class="paging"]')[1].select('a'):
          link_list.append(base_url+url['href'])
```

There are 3467 pages on cvedetails.com

```
[ ]: #print CVE Infos
      def obtain_cve_info(url):
          #Perform HTTP Request
          html = requests.get(url).text
          #Parse html
          soup = BeautifulSoup(html, 'html.parser')
          #Locate Table
          table = soup.select('table[class="searchresults sortable"]')
          #What headers exist?
```

```

headers = []
for header in table[0].findAll('th')[1:]:
    headertext = header.get_text().strip()
    headers.append(headertext)
headers.append('Description')
#Get Data Elements
i = 0
cve_elements = []
for cve in table[0].select('tr[class="srrowsns"]'):
    cve_element = {}
    subid = 0
    for subelement in cve.select('td')[1:]:
        cve_element[headers[subid]] = subelement.get_text().strip()
        subid += 1
    cve_element[headers[subid]] = table[0].
→select('td[class="cvesummarylong"]')[i].get_text().strip()
    i += 1
    cve_elements.append(cve_element)
return (cve_elements)

```

```

[ ]: cve_list = []
#Remove [:5] for full output
for link in tqdm(link_list[:5]):
    cve_list.extend(obtain_cve_info(link))

```

100%| | 5/5 [00:14<00:00, 2.95s/it]

```

[ ]: df = pd.DataFrame(cve_list)
#save as csv
df.to_csv('cves.csv')
#print output
df

```

```

[ ]:

```

	CVE ID	CWE ID	# of Exploits	Vulnerability Type(s)	Publish Date	\
0	CVE-2022-28391			Exec Code	2022-04-03	
1	CVE-2022-28390				2022-04-03	
2	CVE-2022-28389				2022-04-03	
3	CVE-2022-28388				2022-04-03	
4	CVE-2022-28381			Exec Code Overflow	2022-04-03	
..	...	...	...	...	...	
245	CVE-2022-26290	78			2022-03-24	
246	CVE-2022-26289	78			2022-03-24	
247	CVE-2022-26285	89		Sql	2022-03-21	
248	CVE-2022-26284	89		Sql	2022-03-21	
249	CVE-2022-26283	89		Sql	2022-03-21	

  

	Update Date	Score	Gained Access	Level	Access Complexity	Authentication	\
0	2022-04-03	0.0		None	???	???	???

1	2022-04-03	0.0		None	???	???	???
2	2022-04-03	0.0		None	???	???	???
3	2022-04-03	0.0		None	???	???	???
4	2022-04-03	0.0		None	???	???	???
..	...	...	...	...	...	...	...
245	2022-03-29	10.0		None	Remote	Low	Not required
246	2022-03-29	10.0		None	Remote	Low	Not required
247	2022-03-29	7.5		None	Remote	Low	Not required
248	2022-03-29	7.5		None	Remote	Low	Not required
249	2022-03-29	7.5		None	Remote	Low	Not required

	Conf.	Integ.	Avail.	\
0	???	???	???	
1	???	???	???	
2	???	???	???	
3	???	???	???	
4	???	???	???	
..	...	...	...	
245	Complete	Complete	Complete	
246	Complete	Complete	Complete	
247	Partial	Partial	Partial	
248	Partial	Partial	Partial	
249	Partial	Partial	Partial	

	Description
0	BusyBox through 1.35.0 allows remote attackers...
1	ems_usb_start_xmit in drivers/net/can/usb/ems_...
2	mcba_usb_start_xmit in drivers/net/can/usb/mcb...
3	usb_8dev_start_xmit in drivers/net/can/usb/usb...
4	Mediaserver.exe in ALLMediaServer 1.6 has a st...
..	...
245	Tenda M3 1.10 V1.0.0.12(4856) was discovered t...
246	Tenda M3 1.10 V1.0.0.12(4856) was discovered t...
247	Simple Subscription Website v1.0 was discovere...
248	Simple Client Management System v1.0 was disco...
249	Simple Subscription Website v1.0 was discovere...

[250 rows x 15 columns]