

Author - Shashank Chhoker

Topic : Wikipedia Scraper

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

```
In [ ]: title=str(input("Enter the topic :")).replace(' ','+')
link='https://www.google.com/search?q='+ title + 'wikipedia'

res=requests.get(link)
soup=BeautifulSoup(res.text,'html.parser')

for sp in soup.find_all('div'):
    try:

        link=(sp.find('a').get('href'))
        if ("en.wikipedia.org" in link):

            break
    except:
        pass
link = link[7:].split('&')[0]
```

```
In [ ]: Title=title.replace('+',' ')
print(Title)
```

```
In [ ]: soup
```

```
In [ ]: print(soup.text.replace('\n\n',''))
```

```
In [ ]: res=requests.get(link)
soup=BeautifulSoup(res.text,'html.parser')

corpus= ''

for p in soup.find_all('p'):
    corpus +=p.text
    corpus +='\n'
corpus=corpus.strip()
print(corpus)
```

```
In [ ]: for i in range (3,467):
        corpus=corpus.replace('['+str(i)+']','')
```

```
In [ ]: print(corpus)
```

```
In [ ]: fd=open(Title + '.txt', 'w', encoding="utf-8")  
        fd.write(corpus)  
        fd.close()
```

```
In [ ]:
```