

my-file

April 27, 2023

```
[1]: import requests
from bs4 import BeautifulSoup
import re

websites = ['https://www.hackerrank.com/', 'https://www.4lotus.co/', 'http://
↳www.hagerstone.com/']

phone_regex = re.compile(r"\b\d{3}[-.]?\d{3}[-.]?\d{4}\b")
email_regex = r"[\w\.-]+@[\w\.-]+\.\w+"

for url in websites:
    print(f"\nScraping data from {url}:")
    reqs = requests.get(url)
    soup = BeautifulSoup(reqs.text, 'html.parser')

    # Title
    print("Title of the website is:")
    for title in soup.find_all('title'):
        print(title.get_text())

    # Phone numbers
    text = reqs.text
    phone_numbers = re.findall(phone_regex, text)
    print("Phone numbers:")
    for number in phone_numbers:
        print(number)

    # Email addresses
    r = requests.get(url)
    emails = re.findall(email_regex, r.text)
    print("Email addresses:")
    for email in emails:
        print(email)
```

Scraping data from <https://www.hackerrank.com/>:

Title of the website is:

Access Denied

Phone numbers:

1682596867

Email addresses:

Scraping data from <https://www.4lotus.co/>:

Title of the website is:

Top Professional Interior Designers & Decorators in Delhi, NCR

Phone numbers:

Email addresses:

info@4lotus.co

info@4lotus.co

Scraping data from <http://www.hagerstone.com/>:

Title of the website is:

Not Acceptable!

Phone numbers:

Email addresses:

[]: