



Web Scraping by using Regular Expressions:

The process of collecting information from web pages is called web scraping. In web scraping to match our required patterns like mail ids, mobile numbers we can use regular expressions.

Eg:

```
1) import re,urllib
2) import urllib.request
3) sites="google rediff".split()
4) print(sites)
5) for s in sites:
6)     print("Searching...",s)
7)     u=urllib.request.urlopen("http://"+s+".com")
8)     text=u.read()
9)     title=re.findall("<title>.*</title>",str(text),re.I)
10)    print(title[0])
```

Eg: Program to get all phone numbers of redbus.in by using web scraping and regular expressions

```
1) import re,urllib
2) import urllib.request
3) u=urllib.request.urlopen("https://www.redbus.in/info/contactus")
4) text=u.read()
5) numbers=re.findall("[0-9-]{7}[0-9-]+",str(text),re.I)
6) for n in numbers:
7)     print(n)
```

Q. Write a Python Program to check whether the given mail id is valid gmail id or not?

```
1) import re
2) s=input("Enter Mail id:")
3) m=re.fullmatch("\w[a-zA-Z0-9_]*@gmail[.].com",s)
4) if m!=None:
5)     print("Valid Mail Id");
6) else:
7)     print("Invalid Mail id")
```

Output:

```
D:\python_classes>py test.py
Enter Mail id:durgatoc@gmail.com
Valid Mail Id
```