Problem Statement

Build a web crawler in Python which goes to websites of 10 popular newspapers and downloads the editorial posted on that very day. It does this everyday in the morning(lets fix a time like 6 AM) and saves it on your desktop.

Algorithm and code

- Get the link to fetch data from.(Editorials here)

```
import requests
from bs4 import BeautifulSoup

url = "http://www.hindustantimes.com/editorials/"
sourcecode = requests.get(url)
soup = BeautifulSoup(sourcecode,"html.parser")
```

- Now gather links of headlines from there.(Note -Links should have a date and time constraint eg. Data will be fetched at 30/05/2017 06:00 am so datetime constraint :from 29/05/2017 06:01am -30/05/2017 05:59 am)

```
for link in soup.findAll('a'{“class”:something):
          href=link.get('href')
         print(link.get('href'))
```

- Next we should make a loop to go to each of headline links and from the page that opens up it should get the article text.
- It should be done for all those 10 websites.

Websites to get editorials from :

1. **http://www.hindustantimes.com/editorials/**
2. **http://thetelegraph.com/category/editorials**
3. **http://www.thehindu.com/opinion/editorial/**
4. **http://blogs.economictimes.indiatimes.com/et-editorials/**
5. **http://blogs.timesofindia.indiatimes.com/toi-editorials/**
6. **http://indianexpress.com/print/the-editorial-page/**
7. **http://www.tribuneindia.com/list/opinion/editorials/**
8. **https://www.ft.com/opinion**
9. **http://www.newindianexpress.com/Opinions/editorials**
10. **http://www.wsj.com/public/page/editorials.html**