Hindi Language_NLP.

- Prepared by: Sagun Shakya (https://github.com/sagsshakya)
- MSc. Data Science
- GITAM Institute of Science, Visakhapatnam.
- Email: sags.shakya@gmail.com (mailto:sags.shakya@gmail.com)

- Here, we create a year - wise data out of day - wise data and convert into a pic

Importing necessary libraries.

In [1]:
```
import pandas as pd
import numpy as np
from bs4 import BeautifulSoup as BS
import requests
import os
```

In [2]:
```
wd = r'C:\Users\acer\Desktop\PythonProgramming\Nepali_Hindi Language Classificati
os.chdir(wd)
```

Fetching the 'a' elements from the website (https://www.dainiktribuneonline.com).

In [34]:
```
def fetchData(url):
    page = requests.get(url)
    soup = BS(page.text, 'html.parser')
    div_text = soup.find_all('h3', class_ = 'width_100')
    return [div_text[ii].a.text for ii in range(len(div_text))]
```

Function to generating the day - wise - URLs using a given

In [36]:
```
from datetime import date,timedelta

def getURL(epoch_year, getDates = False):

    gap = int(str(date(int(epoch_year),12,31) - date(int(epoch_year),1,1))[:3])

    date_list = [str(date(int(epoch_year),1,1) + timedelta(ii)) for ii in range(g
    date_list = [dates.replace('-','/') for dates in date_list]

    url_list = [r'https://www.dainiktribuneonline.com/' + dates for dates in date

    if getDates:
        return date_list, url_list
    else:
        return url_list
```

In [37]:
```
date_list_2017, url_list_2017 = getURL(2017, True)
```

In [38]:
```
date_list_2018, url_list_2018 = getURL(2018, True)
```

In [39]:
```
date_list_2019, url_list_2019 = getURL(2019, True)
```

Generating a dictionary.

In [40]:
```
# Generating a dictionary for 2017.
date_wise_text_2017 = dict()
for ii in range(0,len(date_list_2017),3):
    date_wise_text_2017[date_list_2017[ii]] = fetchData(url_list_2017[ii])
```

In [45]:
```
# Generating a dictionary for 2018.
date_wise_text_2018 = dict()
for ii in range(0,len(date_list_2018),3):
    date_wise_text_2018[date_list_2018[ii]] = fetchData(url_list_2018[ii])
```

In [46]:
```
# Generating a dictionary for 2019.
date_wise_text_2019 = dict()
for ii in range(0,len(date_list_2019),3):
    date_wise_text_2019[date_list_2019[ii]] = fetchData(url_list_2019[ii])
```

In [44]:
```
len(date_wise_text_2017['2017/05/04'])
```

```
326
```

Pickling the dictionary.

In [47]:
```
import pickle
```

In [49]:
```
def dumpPickle(myfile, filename):
    pickle.dump(myfile, open(str(filename) , "wb"))  # save it into a file named
```

In [54]:
```
dumpPickle(date_wise_text_2017, 'date_wise_text_2017_hindi.p')
```

In [55]:
```
dumpPickle(date_wise_text_2018, 'date_wise_text_2018_hindi.p')
```

In [56]:
```
dumpPickle(date_wise_text_2019, 'date_wise_text_2019_hindi.p')
```

The End.