NAME : GINI CHACKO

ROLL : 8942

CLASS : SE COMPS B

BATCH : B

TOPIC : PYTHON EXPERIMENT 5

Aim: Study Text Processing in Python.

Problem Statement: You will be creating a fragment of text.

1. In this fragment, you need to identify the frequency of articles used (i.e., 'a', 'an', 'the').
2. And you also need to identify dates (which might be expressed in a variety of ways such as '15/11/2012','15/11/12', '15th March 1999','15th March 99' or '20th of March, 1999').

```
from google.colab import drive
drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

```
!pip install datefinder
```

```
Collecting datefinder
  Downloading https://files.pythonhosted.org/packages/0c/4f/29524c9ca35d2ba1a8a3c6c895b9
Requirement already satisfied: pytz in /usr/local/lib/python3.7/dist-packages (from date
Requirement already satisfied: regex>=2017.02.08 in /usr/local/lib/python3.7/dist-packag
Requirement already satisfied: python-dateutil>=2.4.2 in /usr/local/lib/python3.7/dist-p
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.7/dist-packages (from
Installing collected packages: datefinder
Successfully installed datefinder-0.7.1
```

```
import re, datetime, datefinder

month=['january','february','march','april','may','june','july','august',
       'september','october','november','december']

file = open("drive/MyDrive/Colab Notebooks/text_processing.txt")
for line in file:
  txt=''.join(e.lower() for e in line.strip() if e.isalnum() or e=='/'
              or e==' ').split()
  a = print("\nThe frequency of article (a) in the given Paragraph : ",
            sum(e=='a' for e in txt))
  an = print("\nThe frequency of article (an) in the given Paragraph : ",
```

```
            sum(e=='an' for e in txt))
the = print("\nThe frequency of article (the) in the given Paragraph : ",
            sum(e=='the' for e in txt))
date = print("\nThe frequncy of dates found in the given Paragraph : ",
             sum(e in month for e in txt))
matches = datefinder.find_dates(line,strict="True")
print("\nDates found in the given Paragraph : ")
for match in matches:
  print ( match.strftime("%d %B, %Y"))

print("\n")
```

```
The frequency of article (a) in the given Paragraph :  8

The frequency of article (an) in the given Paragraph :  1

The frequency of article (the) in the given Paragraph :  33

The frequncy of dates found in the given Paragraph :  2

Dates found in the given Paragraph :
10 May, 1857
16 December, 1773
```

✓ 0s completed at 10:34 PM