Lab 1

Getting Started with NLTK

Environment Setup

1. Type “Anaconda Navigator” in search bar (Windows Button).
2. Till redirect you to Anaconda Navigator and there “Launch Cmd.exe Prompt”
3. Go to the folder where you want the Application to be
4. You can install NLTK in two ways! 1. From Anaconda prompt 2. From Jupyter notebook 3. even from Google CoLAB

NLTK TOOLKIT

5. Natural language toolkit (NLTK) is the most popular library for natural language processing (NLP).
6. It is written in Python and hence is powerful, easy to use.
7. NLTK also is very easy to learn.

1. From Anaconda Prompt

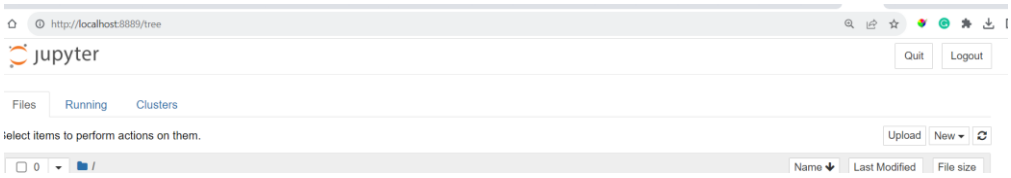
8. Go to the anaconda Prompt in the right path
   (base) D:\ML_DL_NLP>cd NLP

9. Type the command :
   (base) D:\ML_DL_NLP\NLP>**pip install nltk**

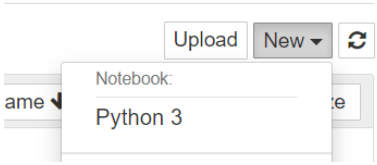
10. Type Jupyter note book , you will be redirected to a browser
    (base) D:\ML_DL_NLP\NLP>**jupyter notebook**

11. Create a new Python file

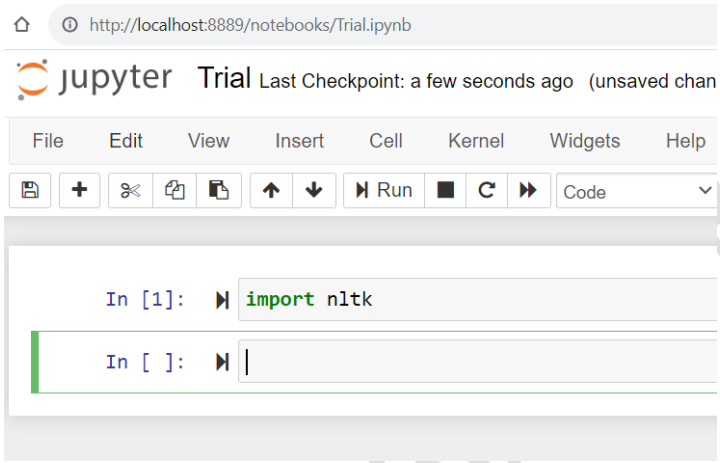
1.



2.



12. Type import nltk



If it doesn't throw any error, you are all set to proceed.

2. From Jupyter Notebook

13. Navigate to the right path in Anaconda prompt
(base) D:\ML_DL_NLP\NLP>

14. Open Jupyter notebook
(base) D:\ML_DL_NLP\NLP>jupyter notebook

15. Create a new python3 file (look at the steps as above)

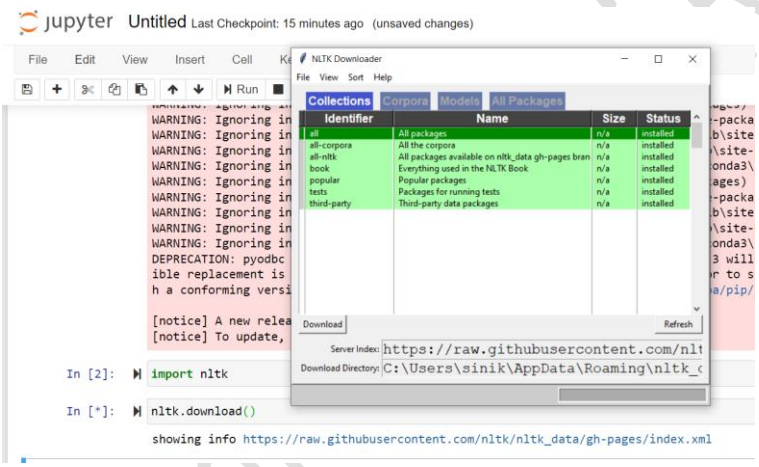
16. Type " pip install nltk "

```
[1]: pip install nltk

Requirement already satisfied: nltk in c:\programdata\anaconda3\lib\site-packages (3.4.5)
Requirement already satisfied: six in c:\programdata\anaconda3\lib\site-packages (from nltk) (1.14.0)
Note: you may need to restart the kernel to use updated packages.

WARNING: Ignoring invalid distribution - (c:\programdata\anaconda3\lib\site-packages)
WARNING: Ignoring invalid distribution -andas (c:\programdata\anaconda3\lib\site-packages)
WARNING: Ignoring invalid distribution -cikit-learn (c:\programdata\anaconda3\lib\site-packages)
WARNING: Ignoring invalid distribution -ensorboard (c:\programdata\anaconda3\lib\site-packages)
WARNING: Ignoring invalid distribution -ensorflow-estimator (c:\programdata\anaconda3\lib\site-packag
WARNING: Ignoring invalid distribution - (c:\programdata\anaconda3\lib\site-packages)
WARNING: Ignoring invalid distribution -andas (c:\programdata\anaconda3\lib\site-packages)
WARNING: Ignoring invalid distribution -cikit-learn (c:\programdata\anaconda3\lib\site-packages)
WARNING: Ignoring invalid distribution -ensorboard (c:\programdata\anaconda3\lib\site-packages)
WARNING: Ignoring invalid distribution -ensorflow-estimator (c:\programdata\anaconda3\lib\site-packag
DEPRECATION: pyodbc 4.0.0-unsupported has a non-standard version number. pip 23.3 will enforce this b
ible replacement is to upgrade to a newer version of pyodbc or contact the author to suggest that the
h a conforming version number. Discussion can be found at https://github.com/pypa/pip/issues/12063

[notice] A new release of pip is available: 23.2.1 -> 23.3.1
[notice] To update, run: python.exe -m pip install --upgrade pip
```

If you are installing for the first time it will show the downloading and installations

17. Type the command " import nltk "

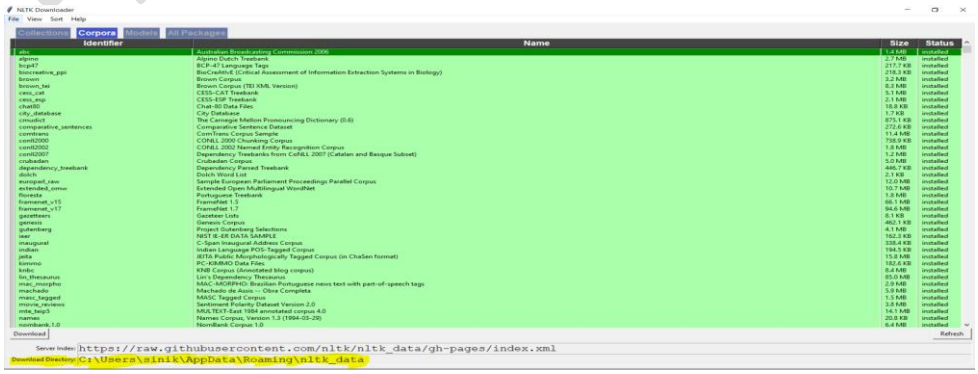
18. Type " nltk.download() "



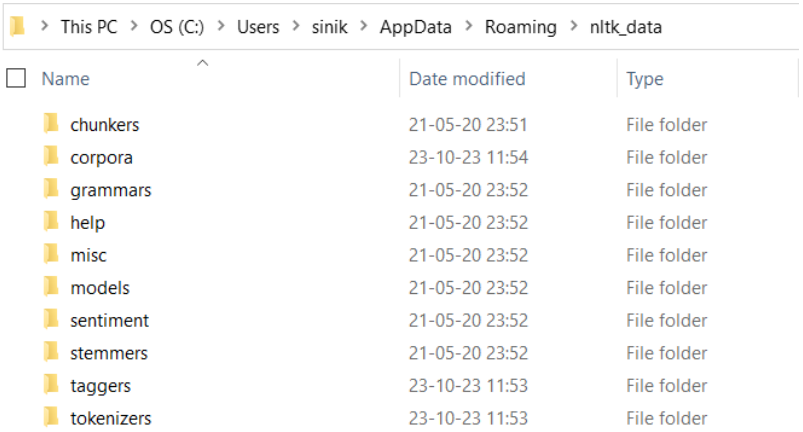
19. A new window will pop up.

You may install the required or all the packages

20. In the window, it will show the path where the nltk downloader is downloading all the packages



21. In the folder , you will see all the following details required to use for NLP Programming



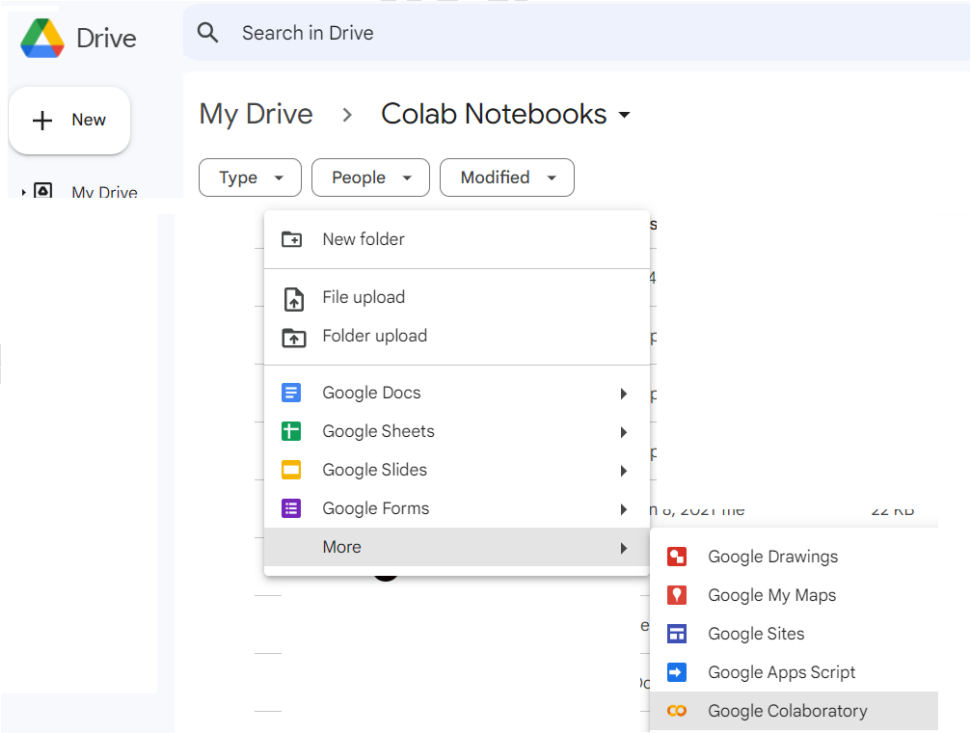
22. Now you are all set to proceed using Jupyter notebook

3. From Google CoLAB or Google CoLaboratory

23. Go to the Google Drive, where you have sufficient space ☺

24. Go to

25. Right click and create a new Google CoLaboratory



26. A new Online Google Jupyter notebook will be open

27. Rename the Notebook GS_NLTK

28. Extension of the google CoLAB will be .ipynb

29. Type the command

!pip install nltk

```
#This Notebook will help you to learn how to install NLTK Pckage and get started with NLP Applications!
```

```
!pip install nltk
```

```
Requirement already satisfied: nltk in c:\programdata\anaconda3\lib\site-packages (3.4.5)
Requirement already satisfied: six in c:\programdata\anaconda3\lib\site-packages (from nltk) (1.14.0)
```

30. Type import nltk

31. Type the command nltk.download(all)

```
In [ ]: import nltk
```

```
In [8]: nltk.download('all')
```

```
[nltk_data] Downloading collection 'all'
[nltk_data]    |
[nltk_data]    | Downloading package abc to
[nltk_data]    |     C:\Users\sinik\AppData\Roaming\nltk_data...
[nltk_data]    |   Package abc is already up-to-date!
[nltk_data]    | Downloading package alpino to
[nltk_data]    |     C:\Users\sinik\AppData\Roaming\nltk_data...
[nltk_data]    |   Package alpino is already up-to-date!
```

32. Inorder to test,

Type the command

nltk.download('wordnet')

nltk.download('stopwords')

```
In [11]: nltk.download('wordnet')
```

```
[nltk_data] Downloading package wordnet to
[nltk_data]     C:\Users\sinik\AppData\Roaming\nltk_data...
[nltk_data]   Package wordnet is already up-to-date!
```

```
Out[11]: True
```

```
In [12]: nltk.download('stopwords')
```

```
[nltk_data] Downloading package stopwords to
[nltk_data]     C:\Users\sinik\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
```

```
Out[12]: True
```

33. Try the below code

In [13]:
```
import nltk

def remove_stopwords(text):

  stop_words = nltk.corpus.stopwords.words('english')

  filtered_text = [word for word in text.split() if word not in stop_words]

  return filtered_text

text = "This is a sentence with some stopwords."

filtered_text = remove_stopwords(text)

print(filtered_text)
```

```
['This', 'sentence', 'stopwords.']
```

In this lab, we have seen how to install and use NLTK package in all the three ways!

****