

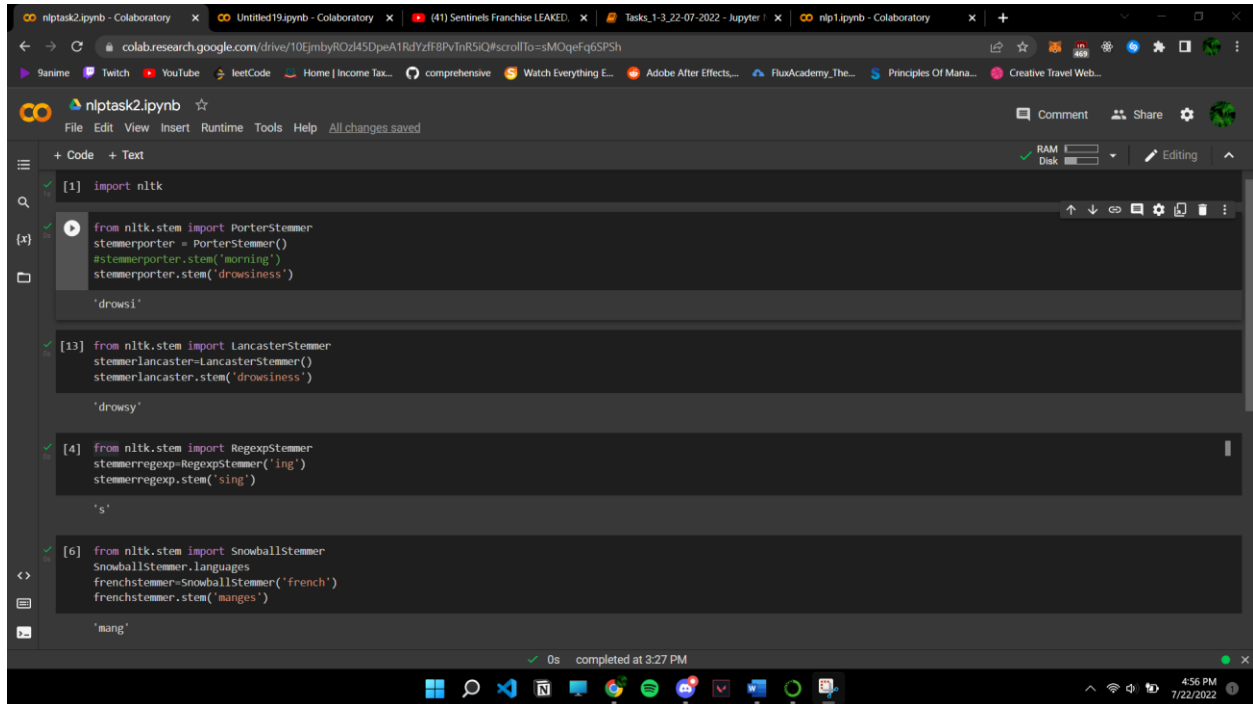
## TASK 5-10

NAME: Mohammad Shahil Hussain

REG:19BCE2447

DATE: 7/22/2022

### #TASK1-STEMMING



The screenshot shows a Jupyter Notebook interface with the title 'nlptask2.ipynb'. The code is as follows:

```
[1] import nltk

from nltk.stem import PorterStemmer
stemmerporter = PorterStemmer()
#stemmerporter.stem('morning')
stemmerporter.stem('drowsiness')

'drowsi'

[13] from nltk.stem import LancasterStemmer
stemmerlancaster=LancasterStemmer()
stemmerlancaster.stem('drowsiness')

'drowsy'

[4] from nltk.stem import RegexpStemmer
stemmerregexp=RegexpStemmer('ing')
stemmerregexp.stem('sing')

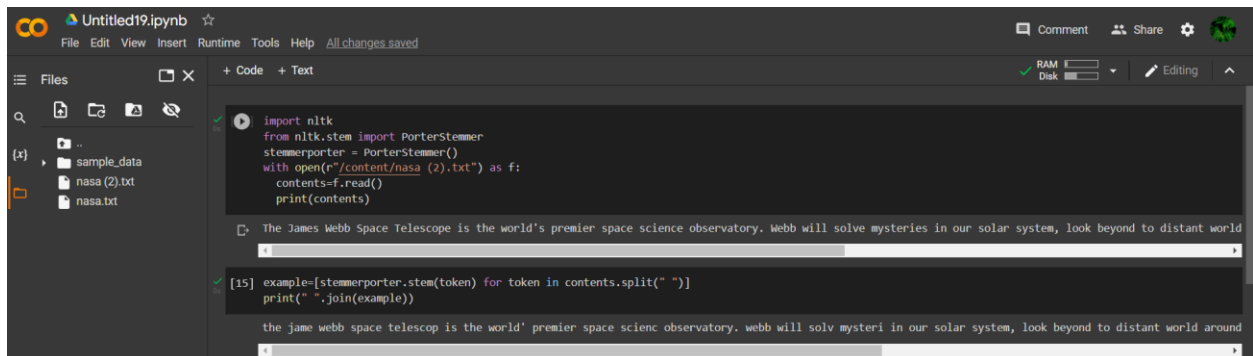
's'

[6] from nltk.stem import SnowballStemmer
SnowballStemmer.Languages
frenchstemmer=SnowballStemmer('french')
frenchstemmer.stem('manges')

'mang'
```

The notebook interface includes a menu bar (File, Edit, View, Insert, Runtime, Tools, Help), a toolbar with icons for code, text, and search, and a status bar at the bottom indicating '0s completed at 3:27 PM'.

### #TASK2-PARAGRAHP STEMMING



The screenshot shows a Jupyter Notebook interface with the title 'Untitled19.ipynb'. The code is as follows:

```
[1] import nltk
from nltk.stem import PorterStemmer
stemmerporter = PorterStemmer()
with open(r"/content/nasa (2).txt") as f:
    contents=f.read()
    print(contents)

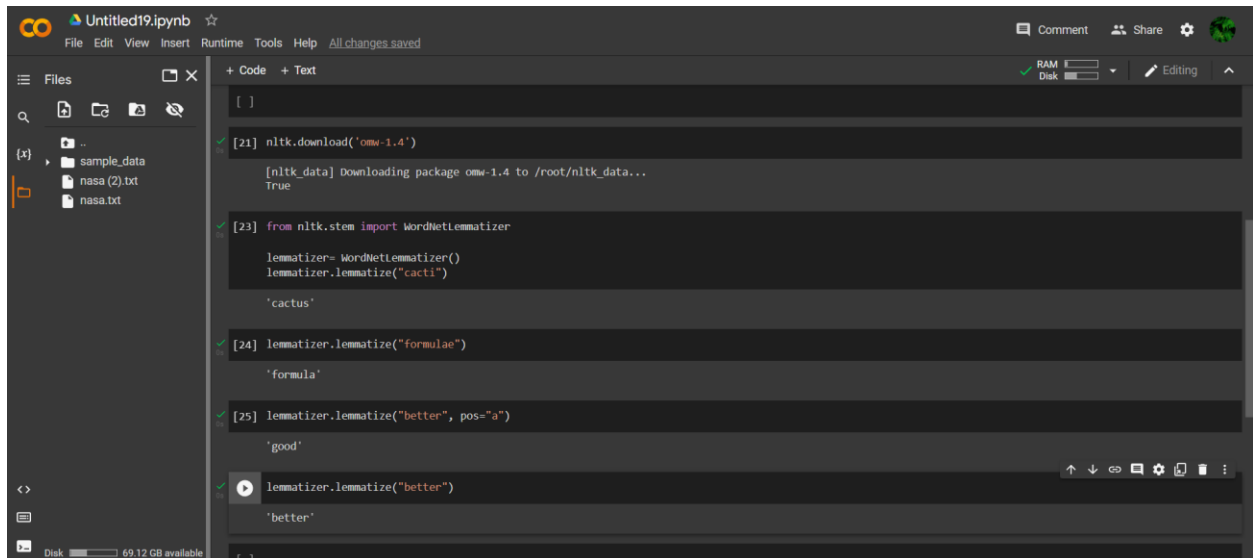
The James Webb Space Telescope is the world's premier space science observatory. Webb will solve mysteries in our solar system, look beyond to distant world

[15] example=[stemmerporter.stem(token) for token in contents.split(" ")]
print(" ".join(example))

the jame webb space telescop is the world' premier space scienc observatory. webb will solv mysteri in our solar system, look beyond to distant world around
```

The notebook interface includes a menu bar (File, Edit, View, Insert, Runtime, Tools, Help), a toolbar with icons for code, text, and search, and a file explorer on the left showing 'sample\_data', 'nasa (2).txt', and 'nasa.txt'. The status bar at the bottom indicates '4:56 PM 7/22/2022'.

## #TASK3-LEMMATIZE



The screenshot shows a Jupyter Notebook titled 'Untitled19.ipynb'. The left sidebar displays a file explorer with a folder named 'sample\_data' containing 'nasa (2).txt' and 'nasa.txt'. The main code area contains the following Python code:

```
[ ]  
[21] nltk.download('omw-1.4')  
[23] from nltk.stem import WordNetLemmatizer  
      lemmatizer= WordNetLemmatizer()  
      lemmatizer.lemmatize('cacti')  
      'cactus'  
[24] lemmatizer.lemmatize("formulae")  
      'formula'  
[25] lemmatizer.lemmatize("better", pos="a")  
      'good'  
[ ] lemmatizer.lemmatize("better")  
      'better'  
[ ]
```

The output of the code is displayed in the right sidebar, showing the lemmatized forms: 'cactus', 'formula', 'good', and 'better'.

## #TASK4-JIEBA

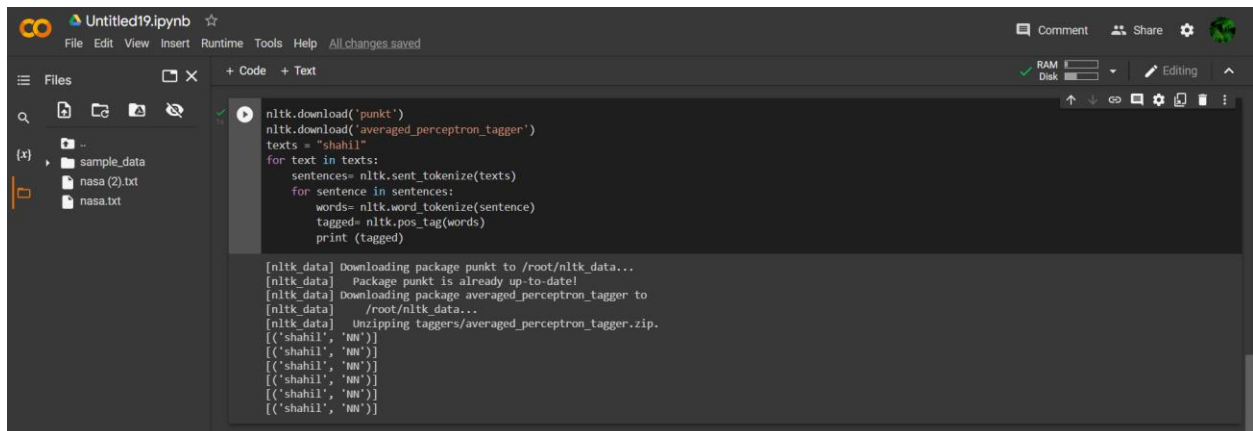


The screenshot shows a Jupyter Notebook titled 'Untitled19.ipynb'. The left sidebar displays a file explorer with a folder named 'sample\_data' containing 'nasa (2).txt' and 'nasa.txt'. The main code area contains the following Python code:

```
import jieba  
seg= jieba.cut("我叫沙希尔") #translation-my name is shahil  
print(" ".join(seg))  
我叫 沙 希尔
```

The output of the code is displayed in the right sidebar, showing the segmented text: '我叫 沙 希尔'.

## #TASK5-TOKEN



The screenshot shows a Jupyter Notebook titled 'Untitled19.ipynb'. The left sidebar displays a file explorer with a folder named 'sample\_data' containing 'nasa (2).txt' and 'nasa.txt'. The main code area contains the following Python code:

```
nltk.download('punkt')  
nltk.download('averaged_perceptron_tagger')  
texts = "shahil"  
for text in texts:  
    sentences= nltk.sent_tokenize(texts)  
    for sentence in sentences:  
        words= nltk.word_tokenize(sentence)  
        tagged= nltk.pos_tag(words)  
        print (tagged)
```

The output of the code is displayed in the right sidebar, showing the tokenized and tagged text:   
[('shahil', 'NN')]  
[('shahil', 'NN')]  
[('shahil', 'NN')]  
[('shahil', 'NN')]  
[('shahil', 'NN')]  
[('shahil', 'NN')]