Name - Amaan Mashooq Nasser
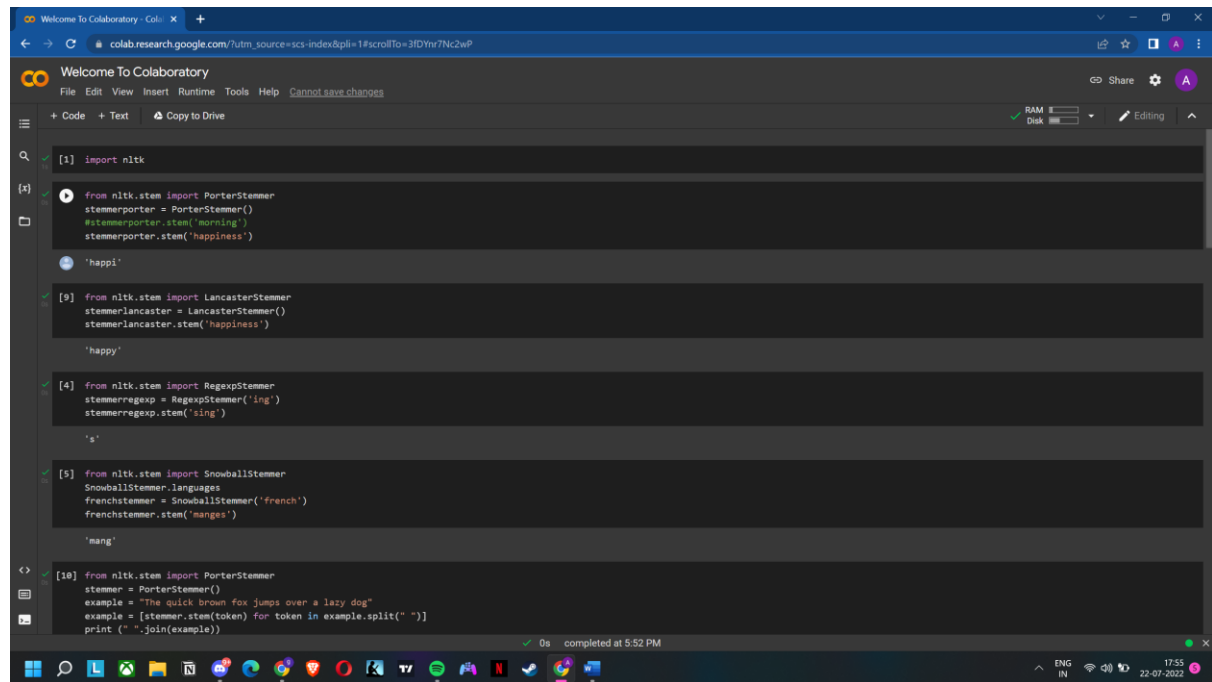# Reg. No. – 19BCE2501
NLP TASKS pt.2(22nd July 2022)

Task 1 – Stemming using different stemmers
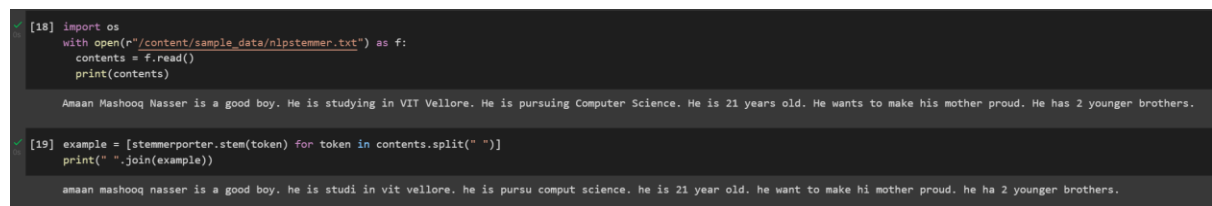

Task 2 – Paragraph Stemming

Task 3 – Lemmatizers

```
[28] import nltk
     nltk.download('omw-1.4')

     [nltk_data] Downloading package omw-1.4 to /root/nltk_data...
     True
```

```
#TASK 3 - Lemmatizer
from nltk.stem import WordNetLemmatizer

lemmatizer = WordNetLemmatizer()
lemmatizer.lemmatize("cacti")
```

```
'cactus'
```

```
lemmatizer.lemmatize("formulae")
```

```
'formula'
```

```
[33] lemmatizer.lemmatize("better", pos="a")
```

```
'good'
```

Task 4 – Jieba Chinese Text Segmentation

```
#TASK 4 - Chinese Segmentation
#pip install jieba
import jieba
seg = jieba.cut("自然語言處理")
#meaning of the chinese text->Natural Language Processing
print(" ".join(seg))
```

```
Building prefix dict from the default dictionary ...
DEBUG:jieba:Building prefix dict from the default dictionary ...
Dumping model to file cache /tmp/jieba.cache
DEBUG:jieba:Dumping model to file cache /tmp/jieba.cache
Loading model cost 0.943 seconds.
DEBUG:jieba:Loading model cost 0.943 seconds.
Prefix dict has been built successfully.
DEBUG:jieba:Prefix dict has been built successfully.
自然 語言 處理
```

Task 5 – Tokens

```
nltk.download('punkt')
nltk.download('averaged_perceptron_tagger')
texts = "AmaanMashooqNasser"
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)
```

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Package punkt is already up-to-date!
[nltk_data] Downloading package averaged_perceptron_tagger to
[nltk_data]     /root/nltk_data...
[nltk_data]   Package averaged_perceptron_tagger is already up-to-
[nltk_data]       date!
[('AmaanMashooqNasser', 'NNP')]
[('AmaanMashooqNasser', 'NNP')]
[('AmaanMashooqNasser', 'NNP')]
[('AmaanMashooqNasser', 'NNP')]
[('AmaanMashooqNasser', 'NNP')]
[('AmaanMashooqNasser', 'NNP')]
[('AmaanMashooqNasser', 'NNP')]
[('AmaanMashooqNasser', 'NNP')]
[('AmaanMashooqNasser', 'NNP')]
[('AmaanMashooqNasser', 'NNP')]
[('AmaanMashooqNasser', 'NNP')]
[('AmaanMashooqNasser', 'NNP')]
[('AmaanMashooqNasser', 'NNP')]
[('AmaanMashooqNasser', 'NNP')]
[('AmaanMashooqNasser', 'NNP')]
[('AmaanMashooqNasser', 'NNP')]
[('AmaanMashooqNasser', 'NNP')]
[('AmaanMashooqNasser', 'NNP')]
```