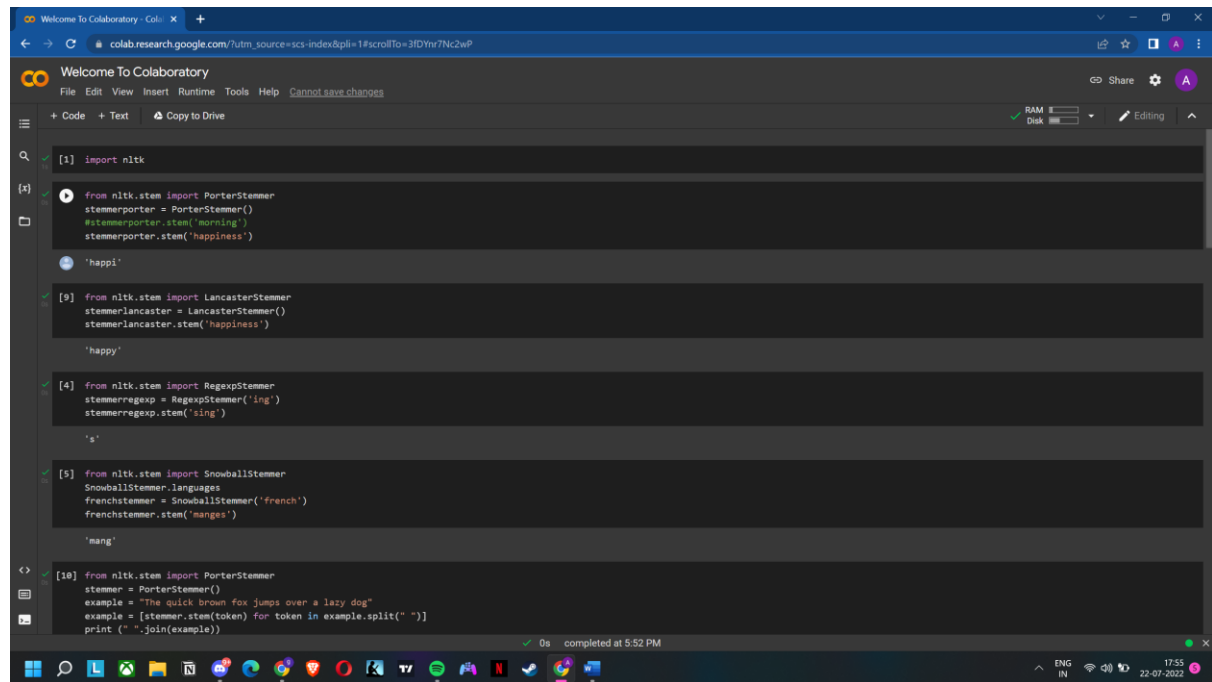


Name - Amaan Mashooq Nasser

Reg. No. – 19BCE2501

NLP TASKS pt.2(22nd July 2022)

Task 1 – Stemming using different stemmers

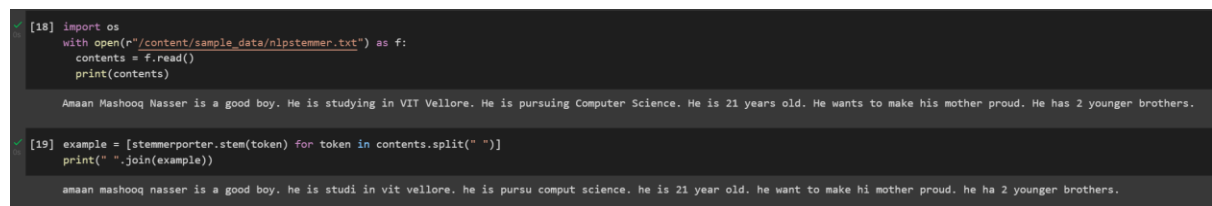


```

Welcome To Colaboratory
File Edit View Insert Runtime Tools Help Cannot save changes
+ Code + Text Copy to Drive
[1] import nltk
[2] from nltk.stem import PorterStemmer
stemmerporter = PorterStemmer()
stemmerporter.stem('morning')
stemmerporter.stem('happiness')
'happi'
[3] from nltk.stem import LancasterStemmer
stemmerlancaster = LancasterStemmer()
stemmerlancaster.stem('happiness')
'happy'
[4] from nltk.stem import RegexpStemmer
stemmerregexp = RegexpStemmer('ing')
stemmerregexp.stem('sing')
's'
[5] from nltk.stem import SnowballStemmer
SnowballStemmer.languages
frenchstemmer = SnowballStemmer('french')
frenchstemmer.stem('manges')
'mang'
[10] from nltk.stem import PorterStemmer
stemmer = PorterStemmer()
example = "The quick brown fox jumps over a lazy dog"
example = [stemmer.stem(token) for token in example.split(" ")]
print(" ".join(example))
0s completed at 5:52 PM

```

Task 2 – Paragraph Stemming



```

[18] import os
with open(r"/content/sample_data/nlpstemmer.txt") as f:
    contents = f.read()
    print(contents)
Amaan Mashooq Nasser is a good boy. He is studying in VIT Vellore. He is pursuing Computer Science. He is 21 years old. He wants to make his mother proud. He has 2 younger brothers.
[19] example = [stemmerporter.stem(token) for token in contents.split(" ")]
print(" ".join(example))
amaan mashooq nasser is a good boy. he is studi in vit vellore. he is pursu comput science. he is 21 year old. he want to make hi mother proud. he ha 2 younger brothers.

```

Task 3 – Lemmatizers

```
✓ 0s [28] import nltk
      nltk.download('omw-1.4')

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

✓ 0s #TASK 3 - Lemmatizer
      from nltk.stem import WordNetLemmatizer

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

      'cactus'

✓ 0s lemmatizer.lemmatize("formulae")

      'formula'

✓ 0s [33] lemmatizer.lemmatize("better", pos="a")

      'good'
```

Task 4 – Jieba Chinese Text Segmentation

```
✓ 1s #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)

```

[illegible]