Assignment-2 (NLP)

Name: NITHIN.V

Batch: 013

Roll Number: 2403A52355

```
import nltk
from collections import Counter
import spacy
```

```
# Downloading NLTK neccsassary things
nltk.download('punkt')
nltk.download('averaged_perceptron_tagger')
```

```
[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!
True
```

```
# Load Spacy Model
nlp = spacy.load('en_core_web_sm')
medical_text = """
The patient reported a sudden headache and nausea.
Blood pressure was elevated, and heart rate was irregular.
The doctor prescribed aspirin and advised rest.
Follow-up tests confirmed improved condition after treatment.
"""

# NLTK Tokenisation
sentnce_nltk = nltk.sent_tokenize(medical_text)
print(sentences_nltk)

word_nltk = nltk.word_tokenize(medical_text)
print(words_nltk)
```

```
['\nThe patient presented with acute chest pain and shortness of breath.', 'Electrocardiogram indicated signs of myocardial infarction.', 'Immediate administration of asp
['The', 'patient', 'presented', 'with', 'acute', 'chest', 'pain', 'and', 'shortness', 'of', 'breath', '.', 'Electrocardiogram', 'indicated', 'signs', 'of', 'myocardial',
```

```
# Spacy Tokenisation
doc = nlp(medical_text)
sentences_spacy = list(doc.sents)
words_spacy = [token.text for token in doc]

print(sentences_spacy)
print(words_spacy)
```

```
[
The patient reported a sudden headache and nausea.
, Blood pressure was elevated, and heart rate was irregular.
, The doctor prescribed aspirin and advised rest.
, Follow-up tests confirmed improved condition after treatment.
]
['\n', 'The', 'patient', 'reported', 'a', 'sudden', 'headache', 'and', 'nausea', '.', '\n', 'Blood', 'pressure', 'was', 'elevated', ',', 'and', 'heart', 'rate', 'was', 'i
```

```
# NLTK POS tagging
pos_nltk = nltk.pos_tag(words_nltk)
print(pos_nltk)
```

```
[('The', 'DT'), ('patient', 'NN'), ('presented', 'VBN'), ('with', 'IN'), ('acute', 'JJ'), ('chest', 'NN'), ('pain', 'NN'), ('and', 'CC'), ('shortness', 'NN'), ('of', 'IN'
```

```
# Spacy POS tagging
pos_spacy = [(token.text, token.pos_, token.tag_) for token in doc]
print(pos_spacy)
```

```
[('\n', 'SPACE', '_SP'), ('The', 'DET', 'DT'), ('patient', 'NOUN', 'NN'), ('reported', 'VERB', 'VBD'), ('a', 'DET', 'DT'), ('sudden', 'ADJ', 'JJ'), ('headache', 'NOUN', '
```

```
# From NLTK
nltk_nouns = [w for w, t in pos_nltk if t.startswith("NN")]
nltk_verbs = [w for w, t in pos_nltk if t.startswith("VB")]

# From spaCy
spacy_nouns = [t.text for t in doc if t.pos_ == "NOUN"]
spacy_verbs = [t.text for t in doc if t.pos_ == "VERB"]
```

Toggle Gemini

```
print("\nNLTK Nouns:", nltk_nouns)
print("spaCy Nouns:", spacy_nouns)

print("\nNLTK Verbs:", nltk_verbs)
print("spaCy Verbs:", spacy_verbs)
```

```
NLTK Nouns: ['patient', 'chest', 'pain', 'shortness', 'breath', 'Electrocardiogram', 'signs', 'infarction', 'Immediate', 'administration', 'aspirin', 'nitroglycerin', 'ec
spaCy Nouns: ['patient', 'headache', 'nausea', 'Blood', 'pressure', 'heart', 'rate', 'doctor', 'aspirin', 'rest', 'tests', 'condition', 'treatment']

NLTK Verbs: ['presented', 'indicated', 'was', 'performed', 'showed']
spaCy Verbs: ['reported', 'elevated', 'prescribed', 'advised', 'Follow', 'confirmed']
```