

Assignment-2 (NLP)

Name: NITHIN.V

Batch: 013

Roll Number: 2403A52355

```
import nltk
from collections import Counter
import spacy

# Downloading NLTK necessary 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
sentence_nltk = nltk.sent_tokenize(medical_text)
print(sentences_nltk)

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

['\n', 'The patient presented with acute chest pain and shortness of breath.', 'Electrocardiogram indicated signs of myocardial infarction.', 'Immediate administration of aspirin and rest was advised. Follow-up tests confirmed improved condition after treatment.', '\n']
['The', 'patient', 'presented', 'with', 'acute', 'chest', 'pain', 'and', 'shortness', 'of', 'breath', '.', 'Electrocardiogram', 'indicated', 'signs', 'of', 'myocardial', 'infarction', '.', 'Immediate', 'administration', 'of', 'aspirin', 'and', 'rest', 'was', 'advised', '.', 'Follow-up', 'tests', 'confirmed', 'improved', 'condition', 'after', 'treatment', '.']
```

```
# 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', 'irregular', '.', '\n', 'The', 'doctor', 'prescribed', 'aspirin', 'and', 'advised', 'rest', '.', '\n', 'The', 'doctor', 'prescribed', 'aspirin', 'and', 'advised', 'rest', '.', '\n', 'Follow-up', 'tests', 'confirmed', 'improved', 'condition', 'after', 'treatment', '.']

# 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'), ('breath', 'NN'), ('.', '.'), ('Electrocardiogram', 'NN'), ('indicated', 'VBD'), ('signs', 'NNS'), ('of', 'IN'), ('myocardial', 'NN'), ('infarction', 'NN'), ('.', '.'), ('Immediate', 'JJ'), ('administration', 'NN'), ('of', 'IN'), ('aspirin', 'NN'), ('and', 'CC'), ('rest', 'NN'), ('was', 'VBD'), ('advised', 'VBN'), ('.', '.'), ('Follow-up', 'NN'), ('tests', 'NNS'), ('confirmed', 'VBN'), ('improved', 'JJ'), ('condition', 'NN'), ('after', 'IN'), ('treatment', 'NN'), ('.', '.')]

# 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', 'NN'), ('and', 'CC', 'CC'), ('nausea', 'NOUN', 'NN'), ('.', 'PUNCT', 'PUNCT'), ('\n', 'SPACE', '_SP'), ('Blood', 'NOUN', 'NN'), ('pressure', 'NOUN', 'NN'), ('was', 'AUX', 'VBD'), ('elevated', 'ADJ', 'JJ'), ('and', 'CC', 'CC'), ('heart', 'NOUN', 'NN'), ('rate', 'NOUN', 'NN'), ('was', 'AUX', 'VBD'), ('irregular', 'ADJ', 'JJ'), ('.', 'PUNCT', 'PUNCT'), ('\n', 'SPACE', '_SP'), ('The', 'DET', 'DT'), ('doctor', 'NOUN', 'NN'), ('prescribed', 'VERB', 'VBD'), ('aspirin', 'NOUN', 'NN'), ('and', 'CC', 'CC'), ('advised', 'VERB', 'VBD'), ('rest', 'NOUN', 'NN'), ('.', 'PUNCT', 'PUNCT'), ('\n', 'SPACE', '_SP'), ('The', 'DET', 'DT'), ('doctor', 'NOUN', 'NN'), ('prescribed', 'VERB', 'VBD'), ('aspirin', 'NOUN', 'NN'), ('and', 'CC', 'CC'), ('advised', 'VERB', 'VBD'), ('rest', 'NOUN', 'NN'), ('.', 'PUNCT', 'PUNCT'), ('\n', 'SPACE', '_SP'), ('Follow-up', 'NOUN', 'NN'), ('tests', 'NOUN', 'NN'), ('confirmed', 'VERB', 'VBD'), ('improved', 'ADJ', 'JJ'), ('condition', 'NOUN', 'NN'), ('after', 'IN', 'IN'), ('treatment', 'NOUN', 'NN'), ('.', 'PUNCT', 'PUNCT')]

# 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"]

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', 'ect']
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']
```