Output:

['this is a sample text', 'It contains multiple sentences.']

['this', 'is', 'a', 'sample', 'text', '.', 'It', 'contains', 'multiple', 'sentences', '.']

[('this', 'PRON'), ('is', 'AUX'), ('a', 'DET'), ('sample', 'NOUN'), ('text', 'NOUN'), ('.', 'PUNCT'), ('It', 'PRON'), ('contains', 'VERB'), ('multiple', 'ADJ'), ('sentences', 'NOUN'), ('.', 'PUNCT')]

17/4/24

Exp - 10

NLP

- Aim:

  To implement NLP

- Code:

```
import nltk
nltk.download('punkt')
from nltk.tokenize import sent_tokenize
# text = "This is a sample text. It contains
multiple sentences."
sentences = sent_tokenize(text)
print(sentences)

from nltk.tokenize import word_tokenize
tokens = word_tokenize(text) # word tokenization
print(token)

import spacy                    # part-of-speech (POS) tagging
nlp = spacy.load("en_core_web_sm")

text = "This is a sample text. It has multiple
sentences."
doc = nlp(text)
pos_tags = [(token.text, token.pos_) for
token in doc]
print(pos_tags);
```

['This', 'is', 'a', 'sample', 'text', '.', 'It', 'contains', 'multiple', 'sentences', '.']

['sample', 'text', '.', 'contains', 'multiple', 'sentences', '.']

[('This', 'nsubj', 'is'), ('is', 'ROOT', 'is'), ('a', 'det', 'text'), ('text', 'attr', 'is'), ('.', 'punct', 'is'), ('It', 'nsubj', 'contains'), ('contains', 'ROOT', 'contains'), ('multiple', 'amod', 'sentences'), ('sentences', 'dobj', 'contains'), ('.', 'punct', 'contains')]

```
from nltk.stem import WordNetLemmatizer
nltk.download('wordnet')
                                    # lemmatization
lemmatizer = WordNetLemmatizer()
lemmas = [lemmatizer.lemmatize(token) for token in tokens]
print(lemmas)


from nltk.corpus import stopwords # stop word identification
nltk.download('stopwords')
stop_words = set(stopwords.words('english'))
filtered_tokens = [token for token in tokens if token.lower() not in stop_words]
print(filtered_tokens)

dependency_tree = [(token.text, token.dep_, token.head.text) for token in doc]
print(dependency_tree)    # dependency parsing
```

Result:

NLP was successfully implemented