POS Tagging using NLTK with Predefined Sentences

In [1]:
```
import nltk
nltk.download('punkt')
nltk.download('averaged_perceptron_tagger_eng')
```

```
[nltk_data] Downloading package punkt to
[nltk_data]     C:\Users\ASUS\AppData\Roaming\nltk_data...
[nltk_data]   Package punkt is already up-to-date!
[nltk_data] Downloading package averaged_perceptron_tagger_eng to
[nltk_data]     C:\Users\ASUS\AppData\Roaming\nltk_data...
[nltk_data]   Package averaged_perceptron_tagger_eng is already up-to-
[nltk_data]       date!
```

Out[1]: True

In [2]:
```
with open("nlp_input.txt", "r") as file:
    text = file.read()
```

In [3]:
```
sentences = nltk.sent_tokenize(text)
```

In [4]:
```
print("Part-Of-Speech Tagging: ")
for sentence in sentences:
    tokens = nltk.word_tokenize(sentence.rstrip('.'))
    pos_tags = nltk.pos_tag(tokens)
    print(f"Sentence: {sentence}")
    print(f"POS Tags: {pos_tags}")
    print('-' * 50)
```

```
Part-Of-Speech Tagging:
Sentence: Natural Language Processing (NLP) is an exciting field of AI.
POS Tags: [('Natural', 'JJ'), ('Language', 'NNP'), ('Processing', 'NNP'), ('(', '('), ('NLP',
'NNP'), (')', ')'), ('is', 'VBZ'), ('an', 'DT'), ('exciting', 'JJ'), ('field', 'NN'), ('of',
'IN'), ('AI', 'NNP')]
--------------------------------------------------
Sentence: It focuses on the interaction between humans and computers using natural language.
POS Tags: [('It', 'PRP'), ('focuses', 'VBZ'), ('on', 'IN'), ('the', 'DT'), ('interaction', 'N
N'), ('between', 'IN'), ('humans', 'NNS'), ('and', 'CC'), ('computers', 'NNS'), ('using', 'VB
G'), ('natural', 'JJ'), ('language', 'NN')]
--------------------------------------------------
Sentence: Tokenization, stemming, and lemmatization are some of its core techniques.
POS Tags: [('Tokenization', 'NN'), (',', ','), ('stemming', 'VBG'), (',', ','), ('and', 'CC'),
('lemmatization', 'NN'), ('are', 'VBP'), ('some', 'DT'), ('of', 'IN'), ('its', 'PRP$'), ('cor
e', 'NN'), ('techniques', 'NNS')]
--------------------------------------------------
Sentence: The quick brown fox jumps over the lazy dog.
POS Tags: [('The', 'DT'), ('quick', 'JJ'), ('brown', 'NN'), ('fox', 'NN'), ('jumps', 'VBZ'),
('over', 'IN'), ('the', 'DT'), ('lazy', 'JJ'), ('dog', 'NN')]
--------------------------------------------------
Sentence: She sells seashells by the seashore.
POS Tags: [('She', 'PRP'), ('sells', 'VBZ'), ('seashells', 'NNS'), ('by', 'IN'), ('the', 'D
T'), ('seashore', 'NN')]
--------------------------------------------------
Sentence: I love programming in Python.
POS Tags: [('I', 'PRP'), ('love', 'VBP'), ('programming', 'VBG'), ('in', 'IN'), ('Python', 'NN
P')]
--------------------------------------------------
```

In [5]:
```
# Tokenization and BoW
from nltk.tokenize import word_tokenize
from sklearn.feature_extraction.text import CountVectorizer
```

```python
import nltk
nltk.download('punkt')

tokens = word_tokenize(text)
print("Tokens:", tokens)

vectorizer = CountVectorizer()
X = vectorizer.fit_transform([text])
print("Vocabulary:", vectorizer.get_feature_names_out())
print("BoW Matrix:", X.toarray())
```

```
Tokens: ['Natural', 'Language', 'Processing', '(', 'NLP', ')', 'is', 'an', 'exciting', 'fiel
d', 'of', 'AI', '.', 'It', 'focuses', 'on', 'the', 'interaction', 'between', 'humans', 'and',
'computers', 'using', 'natural', 'language', '.', 'Tokenization', ',', 'stemming', ',', 'and',
'lemmatization', 'are', 'some', 'of', 'its', 'core', 'techniques', '.', 'The', 'quick', 'brow
n', 'fox', 'jumps', 'over', 'the', 'lazy', 'dog', '.', 'She', 'sells', 'seashells', 'by', 'th
e', 'seashore', '.', 'I', 'love', 'programming', 'in', 'Python', '.']
Vocabulary: ['ai' 'an' 'and' 'are' 'between' 'brown' 'by' 'computers' 'core' 'dog'
 'exciting' 'field' 'focuses' 'fox' 'humans' 'in' 'interaction' 'is' 'it'
 'its' 'jumps' 'language' 'lazy' 'lemmatization' 'love' 'natural' 'nlp'
 'of' 'on' 'over' 'processing' 'programming' 'python' 'quick' 'seashells'
 'seashore' 'sells' 'she' 'some' 'stemming' 'techniques' 'the'
 'tokenization' 'using']
BoW Matrix: [[1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 2 1 2 1 1 1 1 1 1 1 1
  1 1 1 1 1 4 1 1]]
```

```
[nltk_data] Downloading package punkt to
[nltk_data]     C:\Users\ASUS\AppData\Roaming\nltk_data...
[nltk_data]   Package punkt is already up-to-date!
```