

```
In [1]: # Assignment - A7 | Name : Pratik Pingale | Roll No : 19C0056
```

Sample Sentences

```
In [2]: sentence1 = "I will walk 500 miles and I would walk 500 more. Just to be the man who walks " + \
        "a thousand miles to fall down at your door!"
        sentence2 = "I played the play playfully as the players were playing in the play with playfulness"
```

Tokenization

```
In [3]: from nltk import word_tokenize, sent_tokenize

print('Tokenized words:', word_tokenize(sentence1))
print('\nTokenized sentences:', sent_tokenize(sentence1))
```

Tokenized words: ['I', 'will', 'walk', '500', 'miles', 'and', 'I', 'would', 'walk', '500', 'more', '.', 'Just', 'to', 'be', 'the', 'man', 'who', 'walks', 'a', 'thousand', 'miles', 'to', 'fall', 'down', 'at', 'your', 'door', '!']

Tokenized sentences: ['I will walk 500 miles and I would walk 500 more.', 'Just to be the man who walks a thousand miles to fall down at your door!']

POS Tagging

```
In [4]: from nltk import pos_tag

token = word_tokenize(sentence1) + word_tokenize(sentence2)
tagged = pos_tag(token)

print("Tagging Parts of Speech:", tagged)
```

Tagging Parts of Speech: [('I', 'PRP'), ('will', 'MD'), ('walk', 'VB'), ('500', 'CD'), ('miles', 'NNS'), ('and', 'CC'), ('I', 'PRP'), ('would', 'MD'), ('walk', 'VB'), ('500', 'CD'), ('more', 'JJR'), ('.', '.'), ('Just', 'NNP'), ('to', 'TO'), ('be', 'VB'), ('the', 'DT'), ('man', 'NN'), ('who', 'WP'), ('walks', 'VBZ'), ('a', 'DT'), ('thousand', 'NN'), ('miles', 'NNS'), ('to', 'TO'), ('fall', 'VB'), ('down', 'RP'), ('at', 'IN'), ('your', 'PRP\$'), ('door', 'NN'), ('!', '.'), ('I', 'PRP'), ('played', 'VBD'), ('the', 'DT'), ('play', 'NN'), ('playfully', 'RB'), ('as', 'IN'), ('the', 'DT'), ('players', 'NNS'), ('were', 'VBD'), ('playing', 'VBG'), ('in', 'IN'), ('the', 'DT'), ('play', 'NN'), ('with', 'IN'), ('playfulness', 'NN')]

Stop-Words Removal

```
In [5]: from nltk.corpus import stopwords

stop_words = stopwords.words('english')

token = word_tokenize(sentence1)
cleaned_token = []

for word in token:
    if word not in stop_words:
        cleaned_token.append(word)

print('Unclean version:', token)
print('\nCleaned version:', cleaned_token)
```

Unclean version: ['I', 'will', 'walk', '500', 'miles', 'and', 'I', 'would', 'walk', '500', 'more', '.', 'Just', 'to', 'be', 'the', 'man', 'who', 'walks', 'a', 'thousand', 'miles', 'to', 'fall', 'down', 'at', 'your', 'door', '!']

Cleaned version: ['I', 'walk', '500', 'miles', 'I', 'would', 'walk', '500', '.', 'Just', 'man', 'walks', 'thousand', 'miles', 'fall', 'door', '!']

Stemming

```
In [6]: from nltk.stem import PorterStemmer

stemmer = PorterStemmer()

token = word_tokenize(sentence2)

stemmed = [stemmer.stem(word) for word in token]
print(" ".join(stemmed))
```

i play the play play as the player were play in the play with playful

Lemmatization

```
In [7]: from nltk.stem import WordNetLemmatizer

lemmatizer = WordNetLemmatizer()

token = word_tokenize(sentence2)

lemmatized_output = [lemmatizer.lemmatize(word) for word in token]
print(" ".join(lemmatized_output))
```

I played the play playfully a the player were playing in the play with playfulness