Sample Sentennces

In [10]:
```
sentence1 = "I will walk 500 miles and I would walk 500 more. Just to be the
            "a thousand miles to fall down at your door!"
sentence2 = "I played the play playfully as the players were playing in the
```

Tokenization

In [11]:
```
from nltk import word_tokenize, sent_tokenize
```

In [12]:
```
print("---------Tokenized Words------------")
print('Tokenized words:', word_tokenize(sentence1))
print("\n")
```

```
---------Tokenized Words------------
Tokenized words: ['I', 'will', 'walk', '500', 'miles', 'and', 'I', 'woul
d', 'walk', '500', 'more', '.', 'Just', 'to', 'be', 'the', 'man', 'who',
'walks', 'a', 'thousand', 'miles', 'to', 'fall', 'down', 'at', 'your', 'do
or', '!']
```

In [13]:
```
print("---------Tokenized Sentences------------")
print('Tokenized sentences:', sent_tokenize(sentence1))
print("\n")
```

```
---------Tokenized Sentences------------
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 doo
r!']
```

POS tagging

In [14]:
```
from nltk import pos_tag
print("---------POS Tagging------------")
token = word_tokenize(sentence1) + word_tokenize(sentence2)
print('POS tagged:', pos_tag(token))
print("\n")
```

```
---------POS Tagging------------
POS tagged: [('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'), ('walk
s', 'VBZ'), ('a', 'DT'), ('thousand', 'NN'), ('miles', 'NNS'), ('to', 'T
O'), ('fall', 'VB'), ('down', 'RP'), ('at', 'IN'), ('your', 'PRP$'), ('doo
r', 'NN'), ('!', '.'), ('I', 'PRP'), ('played', 'VBD'), ('the', 'DT'), ('p
lay', 'NN'), ('playfully', 'RB'), ('as', 'IN'), ('the', 'DT'), ('players',
'NNS'), ('were', 'VBD'), ('playing', 'VBG'), ('in', 'IN'), ('the', 'DT'),
('play', 'NN'), ('with', 'IN'), ('playfullness', 'NN')]
```

Stop-Words Removal

In [15]:
```
print("---------Stop-Words Removal------------")
from nltk.corpus import stopwords
stop_words = set(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("\n")
print("Cleaned version:", cleaned_token)
print("\n")
```

```
---------Stop-Words Removal------------
Unclean version: ['I', 'will', 'walk', '500', 'miles', 'and', 'I', 'woul
d', 'walk', '500', 'more', '.', 'Just', 'to', 'be', 'the', 'man', 'who',
'walks', 'a', 'thousand', 'miles', 'to', 'fall', 'down', 'at', 'your', 'do
or', '!']


Cleaned version: ['I', 'walk', '500', 'miles', 'I', 'would', 'walk', '50
0', '.', 'Just', 'man', 'walks', 'thousand', 'miles', 'fall', 'door', '!']
```

In [16]:
```
print("---------Stemming------------")
from nltk.stem import PorterStemmer
stemmer = PorterStemmer()
token = word_tokenize(sentence2)
stemmed = [stemmer.stem(word) for word in token]
print("Stemmed words:", stemmed)
print("\n")
```

```
---------Stemming------------
Stemmed words: ['i', 'play', 'the', 'play', 'play', 'as', 'the', 'player',
'were', 'play', 'in', 'the', 'play', 'with', 'playful']
```

In [17]:
```
print("---------Lemmatization------------")
from nltk.stem import WordNetLemmatizer
lemmatizer = WordNetLemmatizer()
token = word_tokenize(sentence2)
lemmatized = [lemmatizer.lemmatize(word) for word in token]
print("Lemmatized words:", lemmatized)
print("\n")
```

```
---------Lemmatization------------
Lemmatized words: ['I', 'played', 'the', 'play', 'playfully', 'a', 'the',
'player', 'were', 'playing', 'in', 'the', 'play', 'with', 'playfullness']
```