In [17]:
```
import nltk
```

Removing Stopwords

In [18]:
```
from nltk.tokenize import word_tokenize
text="""Hello Mr. Smith, how are you doing today? The weather is great, and city
The sky is pinkish-blue. You shouldn't eat cardboard"""
tokenized_sent=word_tokenize(text)

filtered_sent=[]
for w in tokenized_sent:
    if w not in stop_words:
        filtered_sent.append(w)
print("Tokenized Sentence:",tokenized_sent)
print("Filterd Sentence:",filtered_sent)
```

```
Tokenized Sentence: ['Hello', 'Mr.', 'Smith', ',', 'how', 'are', 'you', 'doin
g', 'today', '?', 'The', 'weather', 'is', 'great', ',', 'and', 'city', 'is', 'a
wesome', '.', 'The', 'sky', 'is', 'pinkish-blue', '.', 'You', 'should', "n't",
'eat', 'cardboard']
Filterd Sentence: ['Hello', 'Mr.', 'Smith', ',', 'how', 'are', 'you', 'doing',
'today', '?', 'The', 'weather', 'is', 'great', ',', 'and', 'city', 'is', 'aweso
me', '.', 'The', 'sky', 'is', 'pinkish-blue', '.', 'You', 'should', "n't", 'ea
t', 'cardboard']
```

Lexicon Normalization

Stemming

In [19]:
```
# Stemming
from nltk.stem import PorterStemmer
from nltk.tokenize import sent_tokenize, word_tokenize

ps = PorterStemmer()

stemmed_words=[]
for w in filtered_sent:
    stemmed_words.append(ps.stem(w))

print("Filtered Sentence:",filtered_sent)
print("Stemmed Sentence:",stemmed_words)
```

```
Filtered Sentence: ['Hello', 'Mr.', 'Smith', ',', 'how', 'are', 'you', 'doing',
'today', '?', 'The', 'weather', 'is', 'great', ',', 'and', 'city', 'is', 'aweso
me', '.', 'The', 'sky', 'is', 'pinkish-blue', '.', 'You', 'should', "n't", 'ea
t', 'cardboard']
Stemmed Sentence: ['hello', 'mr.', 'smith', ',', 'how', 'are', 'you', 'do', 'to
day', '?', 'the', 'weather', 'is', 'great', ',', 'and', 'citi', 'is', 'awesom',
'.', 'the', 'sky', 'is', 'pinkish-blu', '.', 'you', 'should', "n't", 'eat', 'ca
rdboard']
```

Lemmatization

In [20]:
```
#Lexicon Normalization
#performing stemming and Lemmatization

from nltk.stem.wordnet import WordNetLemmatizer
lem = WordNetLemmatizer()

from nltk.stem.porter import PorterStemmer
stem = PorterStemmer()

word = "Lemmatization"
print("Lemmatized Word:",lem.lemmatize(word,"v"))
print("Stemmed Word:",stem.stem(word))
```

```
Lemmatized Word: Lemmatization
Stemmed Word: lemmat
```

POS Tagging

In [21]:
```
sent = "Albert Einstein was born in Ulm, Germany in 1879."
```

In [22]:
```
tokens=nltk.word_tokenize(sent)
print(tokens)
```

```
['Albert', 'Einstein', 'was', 'born', 'in', 'Ulm', ',', 'Germany', 'in', '187
9', '.']
```

In [25]:
```
nltk.pos_tag(tokens)
```

Out[25]:
```
[('Albert', 'NNP'),
 ('Einstein', 'NNP'),
 ('was', 'VBD'),
 ('born', 'VBN'),
 ('in', 'IN'),
 ('Ulm', 'NNP'),
 (',', ','),
 ('Germany', 'NNP'),
 ('in', 'IN'),
 ('1879', 'CD'),
 ('.', '.')]
```

NNP:Proper noun singular

VBD : Verb past tense

CD:Cardinal number

VBN:Verb past participle

Sentiment Analysis

Text Classification

In [ ]: