In [6]:
```
pip install nltk
```

```
Requirement already satisfied: nltk in c:\users\anushka\anaconda3\lib\site-
packages (3.8.1)
Requirement already satisfied: click in c:\users\anushka\anaconda3\lib\site
-packages (from nltk) (8.0.4)
Requirement already satisfied: joblib in c:\users\anushka\anaconda3\lib\sit
e-packages (from nltk) (1.2.0)
Requirement already satisfied: regex>=2021.8.3 in c:\users\anushka\anaconda
3\lib\site-packages (from nltk) (2022.7.9)
Requirement already satisfied: tqdm in c:\users\anushka\anaconda3\lib\site-
packages (from nltk) (4.65.0)
Requirement already satisfied: colorama in c:\users\anushka\anaconda3\lib\s
ite-packages (from click->nltk) (0.4.6)
Note: you may need to restart the kernel to use updated packages.
```

In [1]:
```
text = "It is a truth universally acknowledged, that a single man in pos
text = text.lower()
print(text)
```

```
it is a truth universally acknowledged, that a single man in possession of
a good fortune,must be in want of a wife.
```

In [2]:
```
text = "It is a truth universally acknowledged, that a single man in pos
text = text.upper()
print(text)
```

```
IT IS A TRUTH UNIVERSALLY ACKNOWLEDGED, THAT A SINGLE MAN IN POSSESSION OF
A GOOD FORTUNE,MUST BE IN WANT OF A WIFE.
```

In [3]:
```
import string
print(string.punctuation)
```

```
!"#$%&'()*+,-./:;<=>?@[\]^_`{|}~
```

In [4]:
```
text_p = "".join([char for char in text if char not in string.punctuatio
print(text_p)
```

```
IT IS A TRUTH UNIVERSALLY ACKNOWLEDGED THAT A SINGLE MAN IN POSSESSION OF A
GOOD FORTUNEMUST BE IN WANT OF A WIFE
```

In [5]:
```
import nltk
```

In [6]:
```
nltk.download('punkt')
```

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

Out[6]:
```
True
```

In [17]:
```
from nltk import word_tokenize
from nltk import sent_tokenize
```

In [18]:
```
words = word_tokenize(text_p)
words1 = sent_tokenize(text_p)
print(words)
print(words1)
```

```
['IT', 'IS', 'A', 'TRUTH', 'UNIVERSALLY', 'ACKNOWLEDGED', 'THAT', 'A', 'SIN
GLE', 'MAN', 'IN', 'POSSESSION', 'OF', 'A', 'GOOD', 'FORTUNEMUST', 'BE', 'I
N', 'WANT', 'OF', 'A', 'WIFE']
['IT IS A TRUTH UNIVERSALLY ACKNOWLEDGED THAT A SINGLE MAN IN POSSESSION OF
A GOOD FORTUNEMUST BE IN WANT OF A WIFE']
```

In [14]:
```
nltk.download('stopwords')
```

```
[nltk_data] Downloading package stopwords to
[nltk_data]     C:\Users\Anushka\AppData\Roaming\nltk_data...
[nltk_data]   Unzipping corpora\stopwords.zip.
```

Out[14]:
```
True
```

In [15]:
```
from nltk.corpus import stopwords
```

In [19]:
```
stop_words = stopwords.words('english')
print(stop_words)
```

```
['i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you', "you're
", "you've", "you'll", "you'd", 'your', 'yours', 'yourself', 'yourselves',
'he', 'him', 'his', 'himself', 'she', "she's", 'her', 'hers', 'herself', 'i
t', "it's", 'its', 'itself', 'they', 'them', 'their', 'theirs', 'themselve
s', 'what', 'which', 'who', 'whom', 'this', 'that', "that'll", 'these', 'th
ose', 'am', 'is', 'are', 'was', 'were', 'be', 'been', 'being', 'have', 'ha
s', 'had', 'having', 'do', 'does', 'did', 'doing', 'a', 'an', 'the', 'and',
'but', 'if', 'or', 'because', 'as', 'until', 'while', 'of', 'at', 'by', 'fo
r', 'with', 'about', 'against', 'between', 'into', 'through', 'during', 'be
fore', 'after', 'above', 'below', 'to', 'from', 'up', 'down', 'in', 'out',
'on', 'off', 'over', 'under', 'again', 'further', 'then', 'once', 'here',
'there', 'when', 'where', 'why', 'how', 'all', 'any', 'both', 'each', 'fe
w', 'more', 'most', 'other', 'some', 'such', 'no', 'nor', 'not', 'only', 'o
wn', 'same', 'so', 'than', 'too', 'very', 's', 't', 'can', 'will', 'just',
'don', "don't", 'should', "should've", 'now', 'd', 'll', 'm', 'o', 're', 'v
e', 'y', 'ain', 'aren', "aren't", 'couldn', "couldn't", 'didn', "didn't",
'doesn', "doesn't", 'hadn', "hadn't", 'hasn', "hasn't", 'haven', "haven't",
'isn', "isn't", 'ma', 'mightn', "mightn't", 'mustn', "mustn't", 'needn', "n
eedn't", 'shan', "shan't", 'shouldn', "shouldn't", 'wasn', "wasn't", 'were
n', "weren't", 'won', "won't", 'wouldn', "wouldn't"]
```

In [20]:
```
filtered_words = [word for word in words if word not in stop_words]
print(filtered_words)
```

```
['IT', 'IS', 'A', 'TRUTH', 'UNIVERSALLY', 'ACKNOWLEDGED', 'THAT', 'A', 'SIN
GLE', 'MAN', 'IN', 'POSSESSION', 'OF', 'A', 'GOOD', 'FORTUNEMUST', 'BE', 'I
N', 'WANT', 'OF', 'A', 'WIFE']
```

In [21]:
```
from nltk.stem.porter import PorterStemmer
```

In [22]:
```
porter = PorterStemmer()
```

In [23]:
```
stemmed = [porter.stem(word) for word in filtered_words]
print(stemmed)
```

```
['it', 'is', 'a', 'truth', 'univers', 'acknowledg', 'that', 'a', 'singl',
'man', 'in', 'possess', 'of', 'a', 'good', 'fortunemust', 'be', 'in', 'wan
t', 'of', 'a', 'wife']
```

In [25]:
```
import nltk
nltk.download('averaged_perceptron_tagger')
```

```
[nltk_data] Downloading package averaged_perceptron_tagger to
[nltk_data]     C:\Users\Anushka\AppData\Roaming\nltk_data...
[nltk_data]   Unzipping taggers\averaged_perceptron_tagger.zip.
```

Out[25]:
```
True
```

In [26]:
```
from nltk import pos_tag
pos = pos_tag(filtered_words)
print(pos)
```

```
[('IT', 'NNP'), ('IS', 'VBZ'), ('A', 'NNP'), ('TRUTH', 'NNP'), ('UNIVERSALL
Y', 'NNP'), ('ACKNOWLEDGED', 'NNP'), ('THAT', 'IN'), ('A', 'NNP'), ('SINGL
E', 'NNP'), ('MAN', 'NNP'), ('IN', 'NNP'), ('POSSESSION', 'NNP'), ('OF', 'I
N'), ('A', 'NNP'), ('GOOD', 'NNP'), ('FORTUNEMUST', 'NNP'), ('BE', 'NNP'),
('IN', 'NNP'), ('WANT', 'NNP'), ('OF', 'IN'), ('A', 'NNP'), ('WIFE', 'NN
P')]
```

In [ ]:
```

```

In [ ]:
```

```

In [ ]:
```

```