Stemming

In [ ]:

```
import nltk
import pandas as pd
from nltk.stem import PorterStemmer

porter_stemmer=PorterStemmer()
```

In [ ]:

```
words=["connect","connected","connection","connections","connects"]
stemmed_words=[porter_stemmer.stem(word=word) for word in words]

stemdf= pd.DataFrame({'original_word': words,'stemmed_word': stemmed_word
stemdf
```

Out[ ]:

	original_word	stemmed_word
0	connect	connect
1	connected	connect
2	connection	connect
3	connections	connect
4	connects	connect

In [ ]:

```
words=["trouble","troubled","troubles","troublemsome"]
stemmed_words=[porter_stemmer.stem(word=word) for word in words]

stemdf= pd.DataFrame({'original_word': words,'stemmed_word': stemmed_word
stemdf
```

Out[ ]:

	original_word	stemmed_word
0	trouble	troubl
1	troubled	troubl
2	troubles	troubl
3	troublemsome	troublemsom

Lemmatization

In [ ]:

```
from nltk.stem import WordNetLemmatizer
nltk.download('wordnet')

lemmatizer = WordNetLemmatizer()
```

```
[nltk_data] Downloading package wordnet to /Users/kavgan/nltk_data...
[nltk_data]   Package wordnet is already up-to-date!
```

In [ ]:

```
words=["trouble","troubling","troubled","troubles",]
lemmatized_words=[lemmatizer.lemmatize(word=word,pos='v') for word in wor
lemmatizeddf= pd.DataFrame({'original_word': words,'lemmatized_word': lem
lemmatizeddf=lemmatizeddf[['original_word','lemmatized_word']]
lemmatizeddf
```

Out[ ]:

	original_word	lemmatized_word
0	trouble	trouble
1	troubling	trouble
2	troubled	trouble
3	troubles	trouble

In [ ]:

```
words=["goose","geese"]
lemmatized_words=[lemmatizer.lemmatize(word=word,pos='n') for word in wor
lemmatizeddf= pd.DataFrame({'original_word': words,'lemmatized_word': lem
lemmatizeddf=lemmatizeddf[['original_word','lemmatized_word']]
lemmatizeddf
```

Out[ ]:

	original_word	lemmatized_word
0	goose	goose
1	geese	goose

Stop Word Removal

In [9]:

```
stopwords=['this','that','and','a','we','it','to','is','of','up','need']
text="this is a text full of content and we need to clean it up"
```

In [ ]:

```
words=text.split(" ")
shortlisted_words=[]

for w in words:
    if w not in stopwords:
        shortlisted_words.append(w)
    else:
        shortlisted_words.append("W")

print("original sentence = ",text)
print("sentence with stop words removed= ",' '.join(shortlisted_words))
```

```
original sentence =  this is a text full of content and we need to clean i
t up
sentence with stop words removed=  W W W text full W content W W W W clean
W W
```