**To Generate Word Cloud**

Packages/Libraries

```
import nltk
from nltk import sent_tokenize
from nltk import word_tokenize
```

```
paragraph = """ Natural language processing (NLP) is a subfield of Artificial Intelligence (AI). This is a widely used technology for

Natural Language Processing (NLP) is a subfield of artificial intelligence that deals with the interaction between computers and human

NLP is used in a wide range of applications, including machine translation, sentiment analysis, speech recognition, chatbots, and text
```

```
nltk.download('punkt')
```

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Unzipping tokenizers/punkt.zip.
True
```

**Word Tokenization - Regular Expression Tokenization**

```
words=word_tokenize(paragraph)
```

```
print(len(words))
```

```
384
```

```
words
```

```
 'speech',
 'recognition',
 ',',
 'chatbots',
 ',',
 'and',
 'text',
 'classification',
 '.']
```

**To Remove punctuation marks**

```
#Empty list to store words
words_no_punc = []

#To Remove punctuation marks
for w in words:
  if w.isalpha():
    words_no_punc.append(w.lower())

words_no_punc
```

```
 'the',
 'interaction',
 'between',
 'computers',
 'and',
 'humans',
 'in',
 'natural',
 'language',
 'it',
 'involves',
 'the',
 'use',
 'of',
 'computational',
 'techniques',
 'to',
 'process',
 'and',
 'analyze',
 'natural',
 'language',
 'data',
 'such',
 'as',
 'text',
 'and',
 'speech',
 'with',
 'the',
 'goal',
 'of',
 'understanding',
 'the',
 'meaning',
 'behind',
 'the',
 'language',
 'nlp',
 'is',
 'used',
 'in',
 'a',
 'wide',
 'range',
 'of',
 'applications',
 'including',
 'machine',
 'translation',
 'sentiment',
 'analysis',
 'speech',
 'recognition',
 'chatbots',
 'and',
 'text',
 'classification']
```

```
print(len(words_no_punc))
```

```
340
```

**Stopwords**

Library for stopwords

```
nltk.download('stopwords')
```

```
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Unzipping corpora/stopwords.zip.
True
```

**To list the stopwords in English Language**

```
from nltk.corpus import stopwords
#List stopwords
stopwords = set(stopwords.words('english'))
print(stopwords)
```

```
{'most', 'so', 'have', 'of', 'other', 'herself', 'who', "you'll", "shan't", 'which', 'd', 'this', "wasn't", 'didn', 'about', 're',
```

**To remove stopwords**

```
#Empty list to store words
new_words=[]
for w in words_no_punc:
  if w not in stopwords:
    new_words.append(w)

print(new_words)
```

```
['natural', 'language', 'processing', 'nlp', 'subfield', 'artificial', 'intelligence', 'ai', 'widely', 'used', 'technology', 'pers
```

**To calculate Frequency of words**

```
from nltk.probability import FreqDist


fdist = FreqDist(new_words)
fdist.most_common(10)
```

```
[('nlp', 10),
 ('language', 9),
 ('natural', 7),
 ('machine', 5),
 ('processing', 4),
 ('applications', 4),
 ('humans', 4),
 ('used', 3),
 ('speech', 3),
 ('learning', 3)]
```

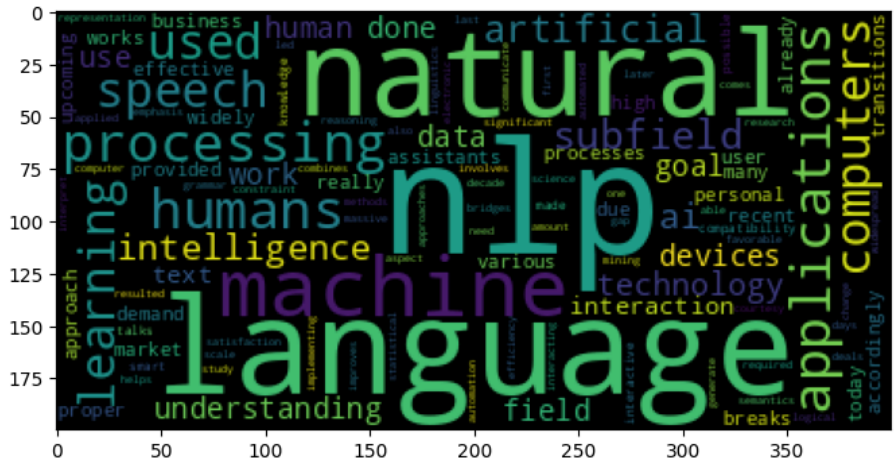
```
#Library
from wordcloud import WordCloud

#Library to plot the wordcloud
import matplotlib.pyplot as plt


#Generating the worcloud
wordcloud = WordCloud().generate_from_frequencies(fdist)

#Plot the wordcloud
plt.figure(figsize = (8,8))
plt.imshow(wordcloud)
```

<matplotlib.image.AxesImage at 0x78a85e1defb0>




```
#Generating the worcloud
wordcloud = WordCloud().generate_from_frequencies(fdist)

#Plot the wordcloud
plt.figure(figsize = (8,8))
plt.imshow(wordcloud)

#To remove axis value
plt.axis("off")
plt.show()
```