

```
#Priya More(305C002)
```

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Unzipping tokenizers/punkt.zip.
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Unzipping corpora/stopwords.zip.
[nltk_data] Downloading package wordnet to /root/nltk_data...
[nltk_data] Downloading package averaged_perceptron_tagger to
[nltk_data]   /root/nltk_data...
[nltk_data]   Unzipping taggers/averaged_perceptron_tagger.zip.
True
```

```
import nltk
nltk.download('punkt')
nltk.download('stopwords')
nltk.download('wordnet')
nltk.download('averaged_perceptron_tagger')
```

```
#it display into the small small chunks
text="Good Day Everyone,H0w are you all today?Its fun learning data analysis.hope y
from nltk.tokenize import sent_tokenize
tokenized_text=sent_tokenize(text)
print(tokenized_text)
```

```
➞ ['Good Day Everyone,H0w are you all today?Its fun learning data analysis.hope you all
```

◀ ▶

```
from nltk.tokenize import word_tokenize
tokenized_word=word_tokenize(text)
print(tokenized_word)
```

```
['Good', 'Day', 'Everyone', ',', 'H0w', 'are', 'you', 'all', 'today', '?', 'Its', 'fu
```

◀ ▶

```
#stopword are removed in sentence:(in,are, you)
from nltk.corpus import stopwords
stop_words=set(stopwords.words("english"))
print(stop_words)
```

```
doesn', 'too', 'had', 'himself', 'to', 'off', 'so', 'yourselves', 'wasn', 'each', "it
```

◀ ▶

```

filtered_send=[]
for w in tokenized_word:
    if w not in stop_words:
        filtered_send.append(w)
print(filtered_send)

```

```

today', '?', 'Its', 'fun', 'learning', 'data', 'analysis.hope', 'practicing', 'well']

```

```

#(stemming ,removing waiting,waits,waited)
from nltk.stem import PorterStemmer
e_words=["wait","waiting","waited","waits"]
ps=PorterStemmer()
for w in e_words:
    rootWord=ps.stem(w)
print(rootWord)

```

```

wait

```

```

#(lemmatization)
from nltk.stem import WordNetLemmatizer
wordNetLemmatizer=WordNetLemmatizer()
text="studies studing cries cry"
tokenized_word=nltk.word_tokenize(text)
for w in tokenized_word:
    print((w,wordNetLemmatizer.lemmatize(w)))

```

```

('studies', 'study')
('studing', 'studing')
('cries', 'cry')
('cry', 'cry')

```

```

from nltk.stem import WordNetLemmatizer

wordNetLemmatizer = WordNetLemmatizer()
text = "studies studing cries cry"
tokenized_word = nltk.word_tokenize(text)
for w in tokenized_word:
    print("Original: {}, Lemmatized: {}".format(w, wordNetLemmatizer.lemmatize(w)))

Original: studies, Lemmatized: study
Original: studing, Lemmatized: studing
Original: cries, Lemmatized: cry
Original: cry, Lemmatized: cry

```

```
 #(pos tagging the display the given word is noun,verb ,pronoun)
 #from nltk.tokenize import word_tokenize
 text="the pink swester fits her perfectly"
 tokenized_word=word_tokenize(text)
 for w in tokenized_word:
     print(nltk.pos_tag([w]))
```

```
[('the', 'DT')]
[('pink', 'NN')]
[('swester', 'NN')]
[('fits', 'NNS')]
[('her', 'PRP$')]
[('perfectly', 'RB')]
```

Start coding or [generate](#) with AI.