

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

```
import nltk
```

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

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

```
sent = "I will walk 500 miles and I would walk 500 more , just to be the man who walks a t
print(word_tokenize(sent))
print(sent_tokenize(sent))
```

```
['I', 'will', 'walk', '500', 'miles', 'and', 'I', 'would', 'walk', '500', 'more', ', ',
['I will walk 500 miles and I would walk 500 more , just to be the man who walks a t
```

```
from nltk.corpus import stopwords
```

```
nltk.download("stopwords")
```

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

```
stopwords=stopwords.words('english')
```

```
print(stopwords)
```

```
['i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you', "you're", "you\'
```

```
tokens=word_tokenize(sent)
```

```
cleaned_token=[]
```

```
for word in tokens:
```

```
    if word not in stopwords:
```

```
        cleaned_token.append(word)
```

```
print(cleaned_token)
```

```
['I', 'walk', '500', 'miles', 'I', 'would', 'walk', '500', ', ', 'man', 'walks', 'thou
```

```

from nltk.stem import PorterStemmer
stemmer=PorterStemmer()

sent2 = "I played the play playfully as the players were playing in the play with playfull
tokens2 = word_tokenize(sent2)
stemmed = ""

for words in tokens2:

    stemmed += stemmer.stem(words) + " "

print(stemmed)

    i play the play play as the player were play in the play with playful

from nltk.stem.snowball import SnowballStemmer

snow_stemmer = SnowballStemmer(language='english')

sent3 = "I played the play playfully as the players were playing in the play with playfull
tokens3 = word_tokenize(sent3)

stem_words = []

for words in tokens3:
    stem_words.append(snow_stemmer.stem(words))

print(stem_words)

for e1,e2 in zip(tokens3,stem_words):
    print(e1 + " ---> " + e2)

    ['i', 'play', 'the', 'play', 'play', 'as', 'the', 'player', 'were', 'play', 'in', 'th
    I ---> i
    played ---> play
    the ---> the
    play ---> play
    playfully ---> play
    as ---> as
    the ---> the
    players ---> player
    were ---> were
    playing ---> play
    in ---> in
    the ---> the
    play ---> play
    with ---> with
    playfullness ---> playful

```

▼ Tagging Parts of Speech(POS)

```

from nltk import pos_tag
nltk.download('averaged_perceptron_tagger')

[nltk_data] Downloading package averaged_perceptron_tagger to
[nltk_data] /root/nltk_data...
[nltk_data] Unzipping taggers/averaged_perceptron_tagger.zip.
True

tagged=pos_tag(cleaned_token)
print(tagged)

[('I', 'PRP'), ('walk', 'VBP'), ('500', 'CD'), ('miles', 'NNS'), ('I', 'PRP'), ('wou]

```

▼ Lemmatization

```

nltk.download('wordnet')

[nltk_data] Downloading package wordnet to /root/nltk_data...
True

from nltk.stem import WordNetLemmatizer
obj=WordNetLemmatizer()

for words in cleaned_token:
    print(words + "--->" + obj.lemmatize(words))

I--->I
walk--->walk
500--->500
miles--->mile
I--->I
would--->would
walk--->walk
500--->500
, --->,
man--->man
walks--->walk
thousand--->thousand
miles--->mile
fall--->fall
door--->door

```

▼ Sample example of Lemmatization

```

list1 = ["kites" , "babies" , "dogs" , "flying" , "smiling" , "driving" , "died" ,

for words in list1:

```

```
print(words + " ---> " + obj.lemmatize(words))
```

```
↳ kites ---> kite  
babies ---> baby  
dogs ---> dog  
flying ---> flying  
smiling ---> smiling  
driving ---> driving  
died ---> died  
tried ---> tried  
feet ---> foot
```

✓ 0s completed at 10:42 PM

