

```
In [1]: import nltk
```

```
In [2]: from nltk import word_tokenize  
from nltk import sent_tokenize
```

```
In [3]: text='yo what is up.i am good. how are you?'
```

```
In [7]: tokens_words=nltk.word_tokenize(text)  
tokens_sent=nltk.sent_tokenize(text)
```

```
In [6]: tokens_words
```

```
Out[6]: ['yo', 'what', 'is', 'up.i', 'am', 'good', '.', 'how', 'are', 'you', '?']
```

```
In [9]: tokens_sent
```

```
Out[9]: ['yo what is up.i am good.', 'how are you?']
```

Stemming

```
In [26]: from nltk.stem import PorterStemmer  
stem=[]  
for i in tokens_words:  
    ps=PorterStemmer()  
    stem_words=ps.stem(i)  
    stem.append(stem_words)  
print(stem)
```

```
['yo', 'what', 'is', 'up.i', 'am', 'good', '.', 'how', 'are', 'you', '?']
```

Lemmatization

```
In [25]: from nltk.stem import WordNetLemmatizer  
lm=WordNetLemmatizer()  
leme=[]  
for i in stem:  
    lem_word=lm.lemmatize(i)  
    leme.append(lem_word)  
print(leme)
```

```
['yo', 'what', 'is', 'up.i', 'am', 'good', '.', 'how', 'are', 'you', '?']
```

pos

```
In [27]: print(nltk.pos_tag(leme))
```

```
[('yo', 'VB'), ('what', 'WP'), ('is', 'VBZ'), ('up.i', 'JJ'), ('am', 'VBP'), ('good', 'JJ'), ('.', '.'), ('how', 'WRB'), ('are', 'VBP'), ('you', 'PRP'), ('?', '.')] 
```

stop word

```
In [35]: from nltk.corpus import stopwords  
sw=stopwords.words('english')
```

```
In [34]: words=[word for word in text.split() if word.lower() not in sw]  
new=" ".join(words)  
print(new)
```

yo up.i good. you?