

```
import nltk
nltk.download('punkt') #sentence tokenization
from nltk.tokenize import word_tokenize, sent_tokenize
```

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

```
para = "Itachi Uchiha is a pivotal character in Naruto, known for his extraordinary skill and tragic backstory. A prodigy of the Uchiha clan
```

```
#sentence tokenization
sent_tokens = nltk.sent_tokenize(para)
print("Sentence Tokenized : ", sent_tokens)
```

```
#word tokenization
word_tokens = nltk.word_tokenize(para)
print("Word Tokenized : ", word_tokens)
```

➡ Sentence Tokenized : ['Itachi Uchiha is a pivotal character in Naruto, known for his extraordinary skill and tragic backstory.', 'A pro
Word Tokenized : ['Itachi', 'Uchiha', 'is', 'a', 'pivotal', 'character', 'in', 'Naruto', ',', 'known', 'for', 'his', 'extraordinary', '']

```
#stopword removal
from nltk.corpus import stopwords
nltk.download('stopwords')
stop_words = set(stopwords.words('english'))
print("Stop Words : ", stop_words)
nstop = []
for w in word_tokens:
    if w not in stop_words:
        nstop.append(w)
```

```
print("After removing stopwords :", nstop)
```

```
➡ Stop Words : {'so', 'did', 'then', 'wasn't', 'needn't', 'further', 'didn't', 'what', 'once', 'won't', 'i', 'not', 'any', 'very', 'her',  
After removing stopwords: ['Itachi', 'Uchiha', 'pivotal', 'character', 'Naruto', ',', 'known', 'extraordinary', 'skill', 'tragic', 'back  
[nltk_data] Downloading package stopwords to /root/nltk_data...  
[nltk_data] Package stopwords is already up-to-date!
```

```
#stemming
from nltk.stem import PorterStemmer
stemmer = PorterStemmer()
stemmed = []
for w in nstopt:
    print(w, " : ", stemmer.stem(w))
    stemmed.append(stemmer.stem(w))
```

```
print("Stemmed : ", stemmed)
```

```

Itachi : itachi
Uchiha : uchiha
pivotal : pivot
character : charact
Naruto : naruto
, : ,
known : known
extraordinary : extraordinary
skill : skill
tragic : tragic
backstory : backstori
. : .
A : a
prodigy : prodigi
Uchiha : uchiha
clan : clan
, : ,
Itachi : itachi
became : becam
high-ranking : high-rank
ninja : ninja
young : young
age : age
. : .
He : he

```

```

infamous : infamous
slaughtering : slaughter
entire : entire
clan : clan
, : ,
leaving : leaving
younger : younger
brother : brother
, : ,
Sasuke : sasuke
, : ,
alive : alive
. : .
Initially : initially
seen : seen
villain : villain
, : ,
later : later
revealed : revealed
Itachi : itachi
's : 's
actions : action
part : part
secret : secret
mission : mission
prevent : prevent
civil : civil
war : war
protect : protect
village : village
. : .
Stemmed : ['itachi', 'uchiha', 'pivot', 'character', 'naruto', '!', 'known', 'extraordinari', 'skill', 'tragic', 'backstori', '!', 'a', 'prodigi', 'uchiha', 'clan', '!', 'itachi', 'becam', 'high-rank', 'ninja', 'young', 'age', '!', 'he', 'infam', 'slaughter', 'entire', 'clan', '!', 'leaving', 'younger', 'brother', '!', 'sasuke', '!', 'alive']

```

```

#lemmetization
nltk.download('wordnet')
from nltk.stem import WordNetLemmatizer
lemmetizer = WordNetLemmatizer()
lemm = []

```

```

for w in stemmed:
    print(w, ":", lemmetizer.lemmatize(w))
    lemm.append(lemmetizer.lemmatize(w))

```

```
print("After lemmatization: ", lemm)
```

```

↩ pivot : pivot
charact : charact
naruto : naruto
, : ,
known : known
extraordinari : extraordinari
skill : skill
tragic : tragic
backstori : backstori
. : .
a : a
prodigi : prodigi
uchiha : uchiha
clan : clan
, : ,
itachi : itachi
becam : becam
high-rank : high-rank
ninja : ninja
young : young
age : age
. : .
he : he
infam : infam
slaughter : slaughter
entire : entire
clan : clan
, : ,
leaving : leaving
younger : younger
brother : brother
, : ,
sasuke : sasuke
, : ,
alive : alive

```

```
villain : villain
, : ,
later : later
reveal : reveal
itachi : itachi
's : 's
action : action
part : part
secret : secret
mission : mission
prevent : prevent
civil : civil
war : war
protect : protect
villag : villag
. : .
After lemmatization: ['itachi', 'uchiha', 'pivot', 'charact', 'naruto', ',', 'known', 'extraordinari', 'skill', 'tragic', 'backstor']
[nltk_data] Downloading package wordnet to /root/nltk_data...
[nltk_data] Package wordnet is already up-to-date!
```

```
!pip install gensim
from gensim.models import Word2Vec
```

```
Requirement already satisfied: gensim in /usr/local/lib/python3.10/dist-packages (4.3.3)
Requirement already satisfied: numpy<2.0,>=1.18.5 in /usr/local/lib/python3.10/dist-packages (from gensim) (1.26.4)
Requirement already satisfied: scipy<1.14.0,>=1.7.0 in /usr/local/lib/python3.10/dist-packages (from gensim) (1.13.1)
Requirement already satisfied: smart-open>=1.8.1 in /usr/local/lib/python3.10/dist-packages (from gensim) (7.0.4)
Requirement already satisfied: wrapt in /usr/local/lib/python3.10/dist-packages (from smart-open>=1.8.1->gensim) (1.16.0)
```

```
tokenized_sentences = [
    [word for word in nltk.word_tokenize(sentence.lower()) if word not in stop_words]
    for sentence in sent_tokens
]
print("Word Tokenized: ", tokenized_sentences)
model = Word2Vec(sentences=tokenized_sentences, min_count=1)
```

```
Word Tokenized: [['itachi', 'uchiha', 'pivotal', 'character', 'naruto', ',', 'known', 'extraordinary', 'skill', 'tragic', 'backstory',
```

```
model.wv.similar_by_word('itachi')
```

```
[('slaughtering', 0.2119942456483841),
 ('high-ranking', 0.19908729195594788),
 ('tragic', 0.17280089855194092),
 ('part', 0.17130933701992035),
 ('clan', 0.17018885910511017),
 ('character', 0.15279100835323334),
 ('infamous', 0.1486695408821106),
 ('young', 0.14596916735172272),
 ('alive', 0.12109686434268951),
 ('initially', 0.11156687885522842)]
```

```
model.wv.words_closer_than('pivotal', 'naruto')
```

```
<ipython-input-20-c9ce248693f2>:1: DeprecationWarning: Call to deprecated `words_closer_than` (Use closer_than instead).
model.wv.words_closer_than('pivotal', 'naruto')
[,',
'itachi',
'uchiha',
'young',
'ninja',
'high-ranking',
'became',
'prodigy',
'tragic',
'backstory',
'skill',
'extraordinary',
'known',
'character',
'infamous',
'village',
'entire',
'war',
'civil',
'prevent',
'mission',
'secret',
```

```
'later',  
'villain',  
'seen',  
'initially',  
'alive',  
'brother']
```