Natural Lounguage Processing (NLP)

import nltk

nltk.download ( )

note :- Multiline paragraph can be used by pasting b/w """ ---. """

* To download the paragraph into sentences ~~into~~ we have to use the below function

  nltk. Sent_tokenize (Paragraph)

  Returns a list

* To_download the ~~sentences~~ sentences into words

  nltk. word_tokenize (Sentences)

  Returns a list.

* Instead of tokenizing a paragraph based on 'space', we tokenize it based on "." and ",". Therefore, we get all the different sentences consisting the paragraph

```
Str = "I LOVE DIVYA"
words = str. split(" ")
Print (words)
out: ["I", "LOVE", "NLP"]
```

For sentence tokenization just replace Str. split(" ") with str. split(".")

Stemming

→ when Extracting sentences & corpus of sentences

" John does his work inteligently"

" John is an intelligent man"

" John is always working"

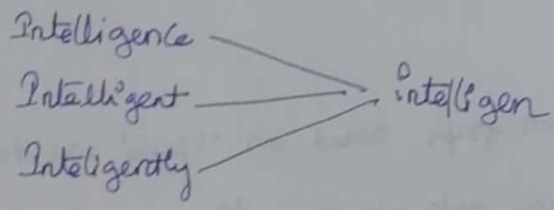
tokenize sentence

Sent 1	sentence 2	Senten 3
John	John	John
does	is	is
his	an	always
work	Intelligent	working
Inteligently	man	

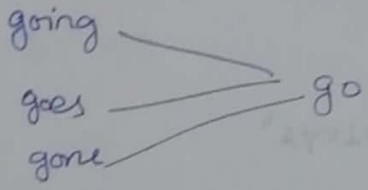
→ We have consider work and working as same

Stemming

"stemming is a process of reducing injected & derived words to their word stem, base & root form"



Intelligence
Intelligent
Inteligently
intelligen



going
goes
gone
go

→ not duplicating the words which gives same meaning

Problem

Produced intermediate representation of the word may not have any meaning

Ex:- intellijan, fina etc
↑
final

Lemmatization

Same as stemming but intermediate representation/root form has a meaning.

intelligence
intelligent → Inetelligent
inteligently

Lemmatization	stemming
* word representations have meaning	* word representation may not have any meaning
* Take more time than stemming	* taken less time
* use when meaning of words is important for analysis Ex:- question answering applicate (chatbot)	* use when meaning of words is not important for analysis Ex:- spam detection, Text classification, Sentiment analysis.

for more (Additional read)

https://nlp.stanford.edu/IR-book/html/htmledition/stemming-and-lemmatization-1.htm

How to implement stemming

Very → Veri
academy → academi
The → Thi

```
  import nltk
  Paragraph = """ ------. """
from, nltk.stem import PorterStemmer
    sentences = nltk.sent_tokenize (Paragramph)
    stemmer = PorterStemmer( )
     for i in range(len(sentences));
        words = nltk.word_tokenize (Sentences[i])
        words = [stemmer.stem(word) for word in words]
        sentences[i] = ' '.join(words)
```

Lemmatization Implementation

```
import nltk
from nltk.stem import WordNetLemmatizer
:
:
:
(Same code as before)
```

Except

```
words = [lemmatizer.lemmatize(word) for word in words]
```