

```
In [4]: from nltk.tokenize import WhitespaceTokenizer
```

```
In [5]: data= "Like a flower in the dessert, I have to grow in the cruelest weather.
```

```
In [11]: print(WhitespaceTokenizer().tokenize(data))
print(len(WhitespaceTokenizer().tokenize(data)))
import nltk

from nltk.tokenize import word_tokenize, sent_tokenize

text1="Hope, is the only thing stronger than fear! #Hope #Amal.M"
text2=" $ student should having the id card around the neck while in the coll

print(word_tokenize(text1))
print(word_tokenize(text2))

['Like', 'a', 'flower', 'in', 'the', 'dessert,', 'I', 'have', 'to', 'grow',
'in', 'the', 'cruelest', 'weather.', 'Holding', 'on', 'to', 'every', 'drop',
'of', 'rain,', 'just', 'to', 'stay', 'alive.', 'But', 'it's', 'not', 'enoug
h', 'to', 'survive,', 'I', 'want', 'to', 'bloom', 'beneath', 'the', 'blazin
g', 'sun.', 'And', 'show', "y'all", 'the', 'colours', 'that', 'live', 'insid
e', 'of', 'me.', 'I', 'want', 'you', 'to', 'see', 'what', 'I', 'can', 'becom
e.', 'Christy', 'Ann', 'Martine']
61
['Hope', ',', 'is', 'the', 'only', 'thing', 'stronger', 'than', 'fear', '!',
'#', 'Hope', '#', 'Amal.M']
['$', 'student', 'should', 'having', 'the', 'id', 'card', 'around', 'the', 'n
eck', 'while', 'in', 'the', 'college', 'campus', '*', '%', '#', '😊😊']
```

```
In [ ]: import nltk

from nltk.tokenize import word_tokenize, sent_tokenize

text1="Hope, is the only thing stronger than fear! #Hope #Amal.M"
text2=" $ student should having the id card around the neck while in the coll

print(word_tokenize(text1))
print(word_tokenize(text2))
```

```
In [ ]: print(word_tokenize(data))
print(len(word_tokenize(data)))
```

```
In [ ]: sent_tokenize(data)
```

```
In [7]: import nltk
from nltk.tokenize import wordpunct_tokenize

print(wordpunct_tokenize(data))
print(len(wordpunct_tokenize(data)))
```

```
['Like', 'a', 'flower', 'in', 'the', 'dessert', ',', 'I', 'have', 'to', 'gro
w', 'in', 'the', 'cruelest', 'weather', '.', 'Holding', 'on', 'to', 'every',
'drop', 'of', 'rain', ',', 'just', 'to', 'stay', 'alive', '.', 'But', 'it',
'', 's', 'not', 'enough', 'to', 'survive', ',', 'I', 'want', 'to', 'bloom',
'beneath', 'the', 'blazing', 'sun', '.', 'And', 'show', 'y', '"', 'all', 'th
```

```
e', 'colours', 'that', 'live', 'inside', 'of', 'me', '.', 'I', 'want', 'you',
'to', 'see', 'what', 'I', 'can', 'become', '.', '"', 'Christy', 'Ann', 'Marti
ne', '"']
75
```

In [8]:

```
import nltk
from nltk.tokenize import TreebankWordTokenizer

print(TreebankWordTokenizer().tokenize(data))
print(len(TreebankWordTokenizer().tokenize(data)))
```

```
[ 'Like', 'a', 'flower', 'in', 'the', 'dessert', ',', 'I', 'have', 'to', 'grow', 'in', 'the', 'cruellest', 'weather.', 'Holding', 'on', 'to', 'every', 'drop', 'of', 'rain', ',', 'just', 'to', 'stay', 'alive.', 'But', 'it's', 'not', 'enough', 'to', 'survive', ',', 'I', 'want', 'to', 'bloom', 'beneath', 'the', 'blazing', 'sun.', 'And', 'show', "y'all", 'the', 'colours', 'that', 'live', 'inside', 'of', 'me.', 'I', 'want', 'you', 'to', 'see', 'what', 'I', 'can', 'become.', 'Christy', 'Ann', 'Martine' ]
```

In [9]:

```
import nltk
from nltk.tokenize import TweetTokenizer

print(TweetTokenizer().tokenize(data))
```

```
[ 'Like', 'a', 'flower', 'in', 'the', 'dessert', ',', 'I', 'have', 'to', 'gro  
w', 'in', 'the', 'cruellest', 'weather', '.', 'Holding', 'on', 'to', 'every',  
'drop', 'of', 'rain', ',', 'just', 'to', 'stay', 'alive', '.', 'But', 'it',  
'', 's', 'not', 'enough', 'to', 'survive', ',', 'I', 'want', 'to', 'bloom',  
'beneath', 'the', 'blazing', 'sun', '.', 'And', 'show', "y'all", 'the', 'colo  
urs', 'that', 'live', 'inside', 'of', 'me', '.', 'I', 'want', 'you', 'to', 's  
ee', 'what', 'I', 'can', 'become', '.', '', 'Christy', 'Ann', 'Martine',  
'']
```

In [12]:

```
import nltk
from nltk.tokenize import MWETokenizer

print(MWETokenizer().tokenize(text2))
print(MWETokenizer().tokenize(word_tokenize(text2)))

t=MWETokenizer()
t.add_mwe(('id', 'card'))
print(t.tokenize(word_tokenize(text2)))
```

```
[',', '$', ',', 's', 't', 'u', 'd', 'e', 'n', 't', ',', 's', 'h', 'o', 'u',  
[',', 'd', ',', 'h', 'a', 'v', 'i', 'n', 'g', ',', 't', 'h', 'e', ',', 'i',  
[',', 'd', ',', 'c', 'a', 'r', 'd', ',', 'a', 'r', 'o', 'u', 'n', 'd', ',', 't',  
[',', 'h', 'e', ',', 'n', 'e', 'c', 'k', ',', 'w', 'h', 'i', 'l', 'e', ',', 'i',  
[',', 'n', ',', 't', 'h', 'e', ',', 'c', 'o', 'l', 'l', 'e', 'g', 'e', ',', 'c',  
[',', 'a', 'm', 'p', 'u', 's', '*', '%', '#', 😊, 😊]  
['$', 'student', 'should', 'having', 'the', 'id', 'card', 'around', 'the', 'n  
eck', 'while', 'in', 'the', 'college', 'campus', '*', '%', '#', 😊, 😊]  
['$', 'student', 'should', 'having', 'the', 'id_card', 'around', 'the', 'nec  
k', 'while', 'in', 'the', 'college', 'campus', '*', '%', '#', 😊, 😊]
```

In [13]:

```
!pip install -U textblob
```

```
Collecting textblob  
  Downloading textblob-0.17.1-py2.py3-none-any.whl (636 kB)  
    | ██████████ | 636 kB 3.7 MB/s eta 0:00:01
```

```
Requirement already satisfied: nltk>=3.1 in /home/admin1/anaconda3/lib/python3.9/site-packages (from textblob) (3.6.5)
Requirement already satisfied: click in /home/admin1/anaconda3/lib/python3.9/site-packages (from nltk>=3.1->textblob) (8.0.3)
Requirement already satisfied: joblib in /home/admin1/anaconda3/lib/python3.9/site-packages (from nltk>=3.1->textblob) (1.1.0)
Requirement already satisfied: regex>=2021.8.3 in /home/admin1/anaconda3/lib/python3.9/site-packages (from nltk>=3.1->textblob) (2021.8.3)
Requirement already satisfied: tqdm in /home/admin1/anaconda3/lib/python3.9/site-packages (from nltk>=3.1->textblob) (4.62.3)
Installing collected packages: textblob
Successfully installed textblob-0.17.1
```

In [14]:

```
from textblob import TextBlob

text= "But I am glad you'll see me as I am. Above all, I wouldn't want people"

blob_object=TextBlob(text)
text_words=blob_object.words
print(text_words)
print(len(text_words))

['But', 'I', 'am', 'glad', 'you', "'ll", 'see', 'me', 'as', 'I', 'am', 'Above', 'e', 'all', 'I', 'would', "n't", 'want', 'people', 'to', 'think', 'that', 'I', 'want', 'to', 'prove', 'anything']
26
```

In [15]:

```
!python3 -m textblob.download_corpora
```

```
[nltk_data] Downloading package brown to /home/admin1/nltk_data...
[nltk_data]   Package brown is already up-to-date!
[nltk_data] Downloading package punkt to /home/admin1/nltk_data...
[nltk_data]   Package punkt is already up-to-date!
[nltk_data] Downloading package wordnet to /home/admin1/nltk_data...
[nltk_data]   Package wordnet is already up-to-date!
[nltk_data] Downloading package averaged_perceptron_tagger to
[nltk_data]   /home/admin1/nltk_data...
[nltk_data]   Package averaged_perceptron_tagger is already up-to-date!
[nltk_data] Downloading package conll2000 to /home/admin1/nltk_data...
[nltk_data]   Package conll2000 is already up-to-date!
[nltk_data] Downloading package movie_reviews to
[nltk_data]   /home/admin1/nltk_data...
[nltk_data]   Package movie_reviews is already up-to-date!
Finished.
```

In [16]:

```
!pip install -U spacy
```

```
Collecting spacy
  Downloading spacy-3.5.0-cp39-cp39-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (6.6 MB)
    |████████████████████████████████████████| 6.6 MB 2.3 MB/s eta 0:00:01
Collecting preshed<3.1.0,>=3.0.2
  Downloading preshed-3.0.8-cp39-cp39-manylinux_2_5_x86_64.manylinux1_x86_64.manylinux_2_17_x86_64.manylinux2014_x86_64.whl (128 kB)
    |████████████████████████████████████████| 128 kB 69.1 MB/s eta 0:00:01
Requirement already satisfied: numpy>=1.15.0 in /home/admin1/anaconda3/lib/python3.9/site-packages (from spacy) (1.20.3)
Collecting thinc<8.2.0,>=8.1.0
  Downloading thinc-8.1.8-cp39-cp39-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (922 kB)
    |████████████████████████████████████████| 922 kB 12.9 MB/s eta 0:00:01
```

```

Collecting typer<0.8.0,>=0.3.0
  Downloading typer-0.7.0-py3-none-any.whl (38 kB)
Requirement already satisfied: jinja2 in /home/admin1/anaconda3/lib/python3.9/site-packages (from spacy) (2.11.3)
Collecting pathy>=0.10.0
  Downloading pathy-0.10.1-py3-none-any.whl (48 kB)
|████████████████████████████████████████| 48 kB 920 kB/s eta 0:00:01
Collecting spacy-legacy<3.1.0,>=3.0.11
  Downloading spacy_legacy-3.0.12-py2.py3-none-any.whl (29 kB)
Requirement already satisfied: packaging>=20.0 in /home/admin1/anaconda3/lib/python3.9/site-packages (from spacy) (21.0)
Collecting pydantic!=1.8,!=1.8.1,<1.11.0,>=1.7.4
  Downloading pydantic-1.10.5-cp39-cp39-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (3.2 MB)
|████████████████████████████████████████| 3.2 MB 13.6 MB/s eta 0:00:01
Collecting smart-open<7.0.0,>=5.2.1
  Downloading smart_open-6.3.0-py3-none-any.whl (56 kB)
|████████████████████████████████████████| 56 kB 1.1 MB/s s eta 0:00:01
Collecting wasabi<1.2.0,>=0.9.1
  Downloading wasabi-1.1.1-py3-none-any.whl (27 kB)
Requirement already satisfied: setuptools in /home/admin1/anaconda3/lib/python3.9/site-packages (from spacy) (58.0.4)
Collecting srsly<3.0.0,>=2.4.3
  Downloading srsly-2.4.6-cp39-cp39-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (491 kB)
|████████████████████████████████████████| 491 kB 66.8 MB/s eta 0:00:01
Collecting spacy-loggers<2.0.0,>=1.0.0
  Downloading spacy_loggers-1.0.4-py3-none-any.whl (11 kB)
Collecting cymem<2.1.0,>=2.0.2
  Downloading cymem-2.0.7-cp39-cp39-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (35 kB)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in /home/admin1/anaconda3/lib/python3.9/site-packages (from spacy) (2.26.0)
Collecting langcodes<4.0.0,>=3.2.0
  Downloading langcodes-3.3.0-py3-none-any.whl (181 kB)
|████████████████████████████████████████| 181 kB 95.1 MB/s eta 0:00:01
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in /home/admin1/anaconda3/lib/python3.9/site-packages (from spacy) (4.62.3)
Collecting murmurhash<1.1.0,>=0.28.0
  Downloading murmurhash-1.0.9-cp39-cp39-manylinux_2_5_x86_64.manylinux1_x86_64.manylinux_2_17_x86_64.manylinux2014_x86_64.whl (21 kB)
Collecting catalogue<2.1.0,>=2.0.6
  Downloading catalogue-2.0.8-py3-none-any.whl (17 kB)
Requirement already satisfied: pyparsing>=2.0.2 in /home/admin1/anaconda3/lib/python3.9/site-packages (from packaging>=20.0->spacy) (3.0.4)
Collecting typing-extensions>=4.2.0
  Downloading typing_extensions-4.5.0-py3-none-any.whl (27 kB)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in /home/admin1/anaconda3/lib/python3.9/site-packages (from requests<3.0.0,>=2.13.0->spacy) (1.26.7)
Requirement already satisfied: certifi>=2017.4.17 in /home/admin1/anaconda3/lib/python3.9/site-packages (from requests<3.0.0,>=2.13.0->spacy) (2021.10.8)
Requirement already satisfied: idna<4,>=2.5 in /home/admin1/anaconda3/lib/python3.9/site-packages (from requests<3.0.0,>=2.13.0->spacy) (3.2)
Requirement already satisfied: charset-normalizer~=2.0.0 in /home/admin1/anaconda3/lib/python3.9/site-packages (from requests<3.0.0,>=2.13.0->spacy) (2.0.4)
Collecting blis<0.8.0,>=0.7.8
  Downloading blis-0.7.9-cp39-cp39-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (10.2 MB)
|████████████████████████████████████████| 10.2 MB 60.3 MB/s eta 0:00:01
Collecting confection<1.0.0,>=0.0.1
  Downloading confection-0.0.4-py3-none-any.whl (32 kB)
Requirement already satisfied: click<9.0.0,>=7.1.1 in /home/admin1/anaconda3/lib/python3.9/site-packages (from typer<0.8.0,>=0.3.0->spacy) (8.0.3)

```

Requirement already satisfied: MarkupSafe>=0.23 in /home/admin1/anaconda3/lib/python3.9/site-packages (from jinja2->spacy) (1.1.1)
 Installing collected packages: typing-extensions, catalogue, srsly, pydantic, murmurhash, cymem, wasabi, typer, smart-open, preshed, confection, blis, thinc, spacy-loggers, spacy-legacy, pathy, langcodes, spacy
 Attempting uninstall: typing-extensions
 Found existing installation: typing-extensions 3.10.0.2
 Uninstalling typing-extensions-3.10.0.2:
 Successfully uninstalled typing-extensions-3.10.0.2
 Successfully installed blis-0.7.9 catalogue-2.0.8 confection-0.0.4 cymem-2.0.7 langcodes-3.3.0 murmurhash-1.0.9 pathy-0.10.1 preshed-3.0.8 pydantic-1.10.5 smart-open-6.3.0 spacy-3.5.0 spacy-legacy-3.0.12 spacy-loggers-1.0.4 srsly-2.4.6 thinc-8.1.8 typer-0.7.0 typing-extensions-4.5.0 wasabi-1.1.1

In [17]:

```
from nltk.stem import PorterStemmer

ps = PorterStemmer()

# choose some words to be stemmed
words = ["program", "programs", "programmer", "programming", "programmers"]

for w in words:
    print(w, " : ", ps.stem(w))
```

```
program : program
programs : program
programmer : programm
programming : program
programmers : programm
```

In [18]:

```
from nltk.stem import PorterStemmer
from nltk.tokenize import word_tokenize

ps = PorterStemmer()

sentence = "Programmers program with programming languages"
tokens = word_tokenize(sentence)

for w in tokens:
    print(w, " : ", ps.stem(w))
```

```
Programmers : programm
program : program
with : with
programming : program
languages : languag
```

In [19]:

```
import nltk
from nltk.stem.snowball import SnowballStemmer

snow_stemmer = SnowballStemmer(language='english')

words = ['cared', 'university', 'fairly', 'easily', 'singing',
         'sings', 'sung', 'singer', 'sportingly']

stem_words = []
for w in words:
    x = snow_stemmer.stem(w)
    stem_words.append(x)
```

```
for e1,e2 in zip(words,stem_words):
    print(e1+' ----> '+e2)
```

```
cared ----> care
university ----> univers
fairly ----> fair
easily ----> easili
singing ----> sing
sings ----> sing
sung ----> sung
singer ----> singer
sportingly ----> sport
```

In [20]:

```
import nltk
from nltk.corpus import stopwords
nltk.download('stopwords')
from nltk.tokenize import word_tokenize
document = "In Python, you do not need to end a statement with a semicolon."
tokens = word_tokenize(document)
filtered_text = [t for t in tokens if not t in stopwords.words("english")]
print(" ".join(filtered_text))
```

In Python , need end statement semicolon .

```
[nltk_data] Downloading package stopwords to /home/admin1/nltk_data...
[nltk_data] Package stopwords is already up-to-date!
```

In [22]:

```
import nltk
from nltk import word_tokenize, pos_tag
sentence1 = "Keep the book on top shelf"
print (pos_tag(word_tokenize(sentence1)))
```

```
[('Keep', 'VB'), ('the', 'DT'), ('book', 'NN'), ('on', 'IN'), ('top', 'JJ'),
('shelf', 'NN')]
```

In [28]:

```
from nltk.stem import WordNetLemmatizer
lemmatizer = WordNetLemmatizer()
print(lemmatizer.lemmatize("vary", 'v'))
print(lemmatizer.lemmatize("varied", 'v'))
print(lemmatizer.lemmatize("varies", 'v'))
print(lemmatizer.lemmatize("varying", 'v'))
```

```
vary
vary
vary
vary
```