In [ ]:

```
!pip install nltk
```

```
Requirement already satisfied: nltk in /usr/local/lib/python3.10/dist-packages (3.8.1)
Requirement already satisfied: click in /usr/local/lib/python3.10/dist-packages (from nltk) (8.1.7)
Requirement already satisfied: joblib in /usr/local/lib/python3.10/dist-packages (from nltk) (1.4.2)
Requirement already satisfied: regex>=2021.8.3 in /usr/local/lib/python3.10/dist-packages (from nltk) (2024.9.11)
Requirement already satisfied: tqdm in /usr/local/lib/python3.10/dist-packages (from nltk) (4.66.5)
```

In [ ]:

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

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

Out[ ]:

```
True
```

In [ ]:

```
corpus =""" Hello Welcome, to Krish Naik's NLP Tutorials.
Please do watch the entire course! to become expert in NLP.
"""
```

In [ ]:

```
print(corpus)
```

```
 Hello Welcome, to Krish Naik's NLP Tutorials.
Please do watch the entire course! to become expert in NLP.

```

In [ ]:

```
from nltk.tokenize import sent_tokenize, word_tokenize
```

In [ ]:

```
documents = sent_tokenize(corpus)
```

In [ ]:

```
type(documents)
```

Out[ ]:

```
list
```

In [ ]:

```
for sentence in documents:
  print(word_tokenize(sentence))
```

```
 Hello Welcome, to Krish Naik's NLP Tutorials.
Please do watch the entire course!
to become expert in NLP.
```

In [ ]:

```
for sentence in documents:
  print(word_tokenize(sentence))
```

```
['Hello', 'Welcome', ',', 'to', 'Krish', 'Naik', "'s", 'NLP', 'Tutorials', '.']
['Please', 'do', 'watch', 'the', 'entire', 'course', '!']
['to', 'become', 'expert', 'in', 'NLP', '.']
```

In [ ]:

```
from nltk.tokenize import word_tokenize
```

In [ ]:

```
word_tokenize(corpus)
```

Out[ ]:

```
['Hello',
 'Welcome',
 ',',
 'to',
 'Krish',
 'Naik',
 "'s",
 'NLP',
 'Tutorials',
 '.',
 'Please',
 'do',
 'watch',
 'the',
 'entire',
 'course',
 '!',
 'to',
 'become',
 'become',
 'expert',
 'in',
 'NLP',
 '.']
```

In [ ]:

```
from nltk.tokenize import wordpunct_tokenize
```

In [ ]:

```
wordpunct_tokenize(corpus)
```

Out[ ]:

```
['Hello',
 'Welcome',
 ',',
 'to',
 'Krish',
 'Naik',
 "'",
 's',
 'NLP',
 'Tutorials',
 '.',
 'Please',
 'do',
 'watch',
 'the',
 'entire',
 'course',
 '!',
 'to',
 'become',
 'expert',
 'in',
 'NLP',
 '.']
```

In [ ]:

```
from nltk.tokenize import TreebankWordTokenizer
```

In [ ]:

```
tokenizer=TreebankWordTokenizer()

tokenizer.tokenize(corpus)
```

Out[ ]:

```
['Hello',
 'Welcome',
 ',',
 'to',
 'Krish',
 'Naik',
 "'s",
 'NLP',
 'Tutorials.',
 'Please',
 'do',
 'watch',
 'the',
 'entire',
 'course',
 '!',
 'to',
 'become',
 'expert',
 'in',
 'NLP',
 '.']
```

In [ ]:

```
words=["eating","eats","eaten","writing","writes","programming","programs","history","finally","finalized"]
```

In [ ]:

```
from nltk.stem import PorterStemmer
stemming=PorterStemmer()
for word in words:
    print(word+"---->"+stemming.stem(word))
```

```
eating---->eat
eats---->eat
eaten---->eaten
writing---->write
writes---->write
programming---->program
programs---->program
history---->histori
finally---->final
finalized---->final
```

In [ ]:

```
from nltk.stem import RegexpStemmer
```

```
reg_stemmer=RegexpStemmer('$in|s$|e$|able$', min=4)
reg_stemmer.stem('eating')
reg_stemmer.stem('indecisive')
```

Out[ ]:

```
'indecisiv'
```

In [ ]:

```
## Q&A,chatbots,text summarization
from nltk.stem import WordNetLemmatizer
lemmatizer=WordNetLemmatizer()
```

In [ ]:

```
nltk.download('wordnet')
```

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

Out[ ]:

```
True
```

In [ ]:

```
words=["eating","eats","eaten","writing","writes","programming","programs","history","finally","finalized"]
for word in words:
    print(word+"---->"+lemmatizer.lemmatize(word,pos='v'))
```

```
eating---->eat
eats---->eat
eaten---->eat
writing---->write
writes---->write
programming---->program
programs---->program
history---->history
finally---->finally
finalized---->finalize
```

In [ ]:

```
## Speech Of DR APJ Abdul Kalam
paragraph = """I have three visions for India. In 3000 years of our history, people from all over
               the world have come and invaded us, captured our lands, conquered our minds.
               From Alexander onwards, the Greeks, the Turks, the Moguls, the Portuguese, the British,
               the French, the Dutch, all of them came and looted us, took over what was ours.
               Yet we have not done this to any other nation. We have not conquered anyone.
               We have not grabbed their land, their culture,
               their history and tried to enforce our way of life on them.
               Why? Because we respect the freedom of others.That is why my
               first vision is that of freedom. I believe that India got its first vision of
               this in 1857, when we started the War of Independence. It is this freedom that
               we must protect and nurture and build on. If we are not free, no one will respect us.
               My second vision for India’s development. For fifty years we have been a developing nation.
               It is time we see ourselves as a developed nation. We are among the top 5 nations of the world
               in terms of GDP. We have a 10 percent growth rate in most areas. Our poverty levels are falling.
               Our achievements are being globally recognised today. Yet we lack the self-confidence to
               see ourselves as a developed nation, self-reliant and self-assured. Isn’t this incorrect?
               I have a third vision. India must stand up to the world. Because I believe that unless India
               stands up to the world, no one will respect us. Only strength respects strength. We must be
               strong not only as a military power but also as an economic power. Both must go hand-in-hand.
               My good fortune was to have worked with three great minds. Dr. Vikram Sarabhai of the Dept. of
               space, Professor Satish Dhawan, who succeeded him and Dr. Brahm Prakash, father of nuclear material.
               I was lucky to have worked with all three of them closely and consider this the great opportunity of my life.
               I see four milestones in my career"""
from nltk.stem import PorterStemmer
from nltk.corpus import stopwords
import nltk
nltk.download('stopwords')
```

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

Out[ ]:

```
True
```

In [ ]:

```

```

In [ ]:

```
from nltk.stem import PorterStemmer
stemmer=PorterStemmer()
sentences=nltk.sent_tokenize(paragraph)
type(sentences)
```

Out[ ]:

```
list
```

In [ ]:

```
print(sentences)
```

```
['I have three visions for India.', 'In 3000 years of our history, people from all over \n               the world have come and i
nvaded us, captured our lands, conquered our minds.', 'From Alexander onwards, the Greeks, the Turks, the Moguls, the Portuguese,
the British,\n               the French, the Dutch, all of them came and looted us, took over what was ours.', 'Yet we have not do
ne this to any other nation.', 'We have not conquered anyone.', 'We have not grabbed their land, their culture, \n               t
heir history and tried to enforce our way of life on them.', 'Why?', 'Because we respect the freedom of others.That is why my \n
heir history and tried to enforce our way of life on them.', 'Why?', 'Because we respect the freedom of others.That is why my \n
first vision is that of freedom.', 'I believe that India got its first vision of \n                this in 1857, when we started th
e War of Independence.', 'It is this freedom that\n                we must protect and nurture and build on.', 'If we are not free,
no one will respect us.', 'My second vision for India’s development.', 'For fifty years we have been a developing nation.', 'It is
time we see ourselves as a developed nation.', 'We are among the top 5 nations of the world\n                in terms of GDP.', 'We
have a 10 percent growth rate in most areas.', 'Our poverty levels are falling.', 'Our achievements are being globally recognised
today.', 'Yet we lack the self-confidence to\n                see ourselves as a developed nation, self-reliant and self-assured.',
'Isn’t this incorrect?', 'I have a third vision.', 'India must stand up to the world.', 'Because I believe that unless India \n
stands up to the world, no one will respect us.', 'Only strength respects strength.', 'We must be \n                strong not only
as a military power but also as an economic power.', 'Both must go hand-in-hand.', 'My good fortune was to have worked with three
great minds.', 'Dr. Vikram Sarabhai of the Dept.', 'of \n                space, Professor Satish Dhawan, who succeeded him and Dr.
Brahm Prakash, father of nuclear material.', 'I was lucky to have worked with all three of them closely and consider this the great
opportunity of my life.', 'I see four milestones in my career']
```

In [ ]:

```
for i in range(len(sentences)):
    words=nltk.word_tokenize(sentences[i])
    words=[stemmer.stem(word) for word in words if word not in set(stopwords.words('english'))]
    sentences[i]=' '.join(words)# converting all the list of words into sentences
```

In [ ]:

```
sentences
```

Out[ ]:

```
['i three vision india .',
 'in 3000 year histori , peopl world come invad us , captur land , conquer mind .',
 'from alexand onward , greek , turk , mogul , portugues , british , french , dutch , came loot us , took .',
 'yet done nation .',
 'we conquer anyon .',
 'we grab land , cultur , histori tri enforc way life .',
 'whi ?',
 'becaus respect freedom others.that first vision freedom .',
 'i believ india got first vision 1857 , start war independ .',
 'it freedom must protect nurtur build .',
 'if free , one respect us .',
 'my second vision india ’ develop .',
 'for fifti year develop nation .',
 'it time see develop nation .',
 'we among top 5 nation world term gdp .',
 'we 10 percent growth rate area .',
 'our poverti level fall .',
 'our achiev global recognis today .',
 'yet lack self-confid see develop nation , self-reli self-assur .',
 'isn ’ incorrect ?',
 'i third vision .',
 'india must stand world .',
 'becaus i believ unless india stand world , one respect us .',
 'onli strength respect strength .',
 'we must strong militari power also econom power .',
 'both must go hand-in-hand .',
 'my good fortun work three great mind .',
 'dr. vikram sarabhai dept .',
 'space , professor satish dhawan , succeed dr. brahm prakash , father nuclear materi .',
 'i lucki work three close consid great opportun life .',
 'i see four mileston career']
```

In [ ]:

```
## Speech Of DR APJ Abdul Kalam
paragraph = """I have three visions for India. In 3000 years of our history, people from all over
               the world have come and invaded us, captured our lands, conquered our minds.
               From Alexander onwards, the Greeks, the Turks, the Moguls, the Portuguese, the British,
               the French, the Dutch, all of them came and looted us, took over what was ours.
               Yet we have not done this to any other nation. We have not conquered anyone.
               We have not grabbed their land, their culture,
               their history and tried to enforce our way of life on them.
               Why? Because we respect the freedom of others.That is why my
               first vision is that of freedom. I believe that India got its first vision of
               this in 1857, when we started the War of Independence. It is this freedom that
               we must protect and nurture and build on. If we are not free, no one will respect us.
               My second vision for India’s development. For fifty years we have been a developing nation.
               It is time we see ourselves as a developed nation. We are among the top 5 nations of the world
               in terms of GDP. We have a 10 percent growth rate in most areas. Our poverty levels are falling.
               Our achievements are being globally recognised today. Yet we lack the self-confidence to
               see ourselves as a developed nation, self-reliant and self-assured. Isn’t this incorrect?
               I have a third vision. India must stand up to the world. Because I believe that unless India
               stands up to the world, no one will respect us. Only strength respects strength. We must be
               strong not only as a military power but also as an economic power. Both must go hand-in-hand.
               My good fortune was to have worked with three great minds. Dr. Vikram Sarabhai of the Dept. of
               space, Professor Satish Dhawan, who succeeded him and Dr. Brahm Prakash, father of nuclear material.
               I was lucky to have worked with all three of them closely and consider this the great opportunity of my life.
               I see four milestones in my career"""
from nltk.corpus import stopwords
sentences=nltk.sent_tokenize(paragraph)
```

In [ ]:

```
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.
```

Out[ ]:

```
True
```

In [ ]:

```python
## We will find the Pos Tag

for i in range(len(sentences)):
    words=nltk.word_tokenize(sentences[i])
    words=[word for word in words if word not in set(stopwords.words('english'))]
    #sentences[i]=' '.join(words)# converting all the list of words into sentences
    pos_tag=nltk.pos_tag(words)
    print(pos_tag)
```

```
[('I', 'PRP'), ('three', 'CD'), ('visions', 'NNS'), ('India', 'NNP'), ('.', '.')]
[('In', 'IN'), ('3000', 'CD'), ('years', 'NNS'), ('history', 'NN'), (',', ','), ('people', 'NNS'), ('world', 'NN'), ('come', 'VBP'
), ('invaded', 'VBN'), ('us', 'PRP'), (',', ','), ('captured', 'VBD'), ('lands', 'NNS'), (',', ','), ('conquered', 'VBD'), ('minds
', 'NNS'), ('.', '.')]
[('From', 'IN'), ('Alexander', 'NNP'), ('onwards', 'NNS'), (',', ','), ('Greeks', 'NNP'), (',', ','), ('Turks', 'NNP'), (',', ',')
, ('Moguls', 'NNP'), (',', ','), ('Portuguese', 'NNP'), (',', ','), ('British', 'NNP'), (',', ','), ('French', 'NNP'), (',', ','),
('Dutch', 'NNP'), (',', ','), ('came', 'VBD'), ('looted', 'JJ'), ('us', 'PRP'), (',', ','), ('took', 'VBD'), ('.', '.')]
[('Yet', 'RB'), ('done', 'VBN'), ('nation', 'NN'), ('.', '.')]
[('We', 'PRP'), ('conquered', 'VBD'), ('anyone', 'NN'), ('.', '.')]
[('We', 'PRP'), ('grabbed', 'VBD'), ('land', 'NN'), (',', ','), ('culture', 'NN'), (',', ','), ('history', 'NN'), ('tried', 'VBD')
, ('enforce', 'JJ'), ('way', 'NN'), ('life', 'NN'), ('.', '.')]
[('Why', 'WRB'), ('?', '.')]
[('Because', 'IN'), ('respect', 'NN'), ('freedom', 'NN'), ('others.That', 'IN'), ('first', 'JJ'), ('vision', 'NN'), ('freedom', 'N
N'), ('.', '.')]
[('I', 'PRP'), ('believe', 'VBP'), ('India', 'NNP'), ('got', 'VBD'), ('first', 'JJ'), ('vision', 'NN'), ('1857', 'CD'), (',', ',')
, ('started', 'VBD'), ('War', 'NNP'), ('Independence', 'NNP'), ('.', '.')]
[('It', 'PRP'), ('freedom', 'NN'), ('must', 'MD'), ('protect', 'VB'), ('nurture', 'NN'), ('build', 'NN'), ('.', '.')]
[('If', 'IN'), ('free', 'JJ'), (',', ','), ('one', 'CD'), ('respect', 'NN'), ('us', 'PRP'), ('.', '.')]
[('My', 'PRP$'), ('second', 'JJ'), ('vision', 'NN'), ('India', 'NNP'), ('’', 'NNP'), ('development', 'NN'), ('.', '.')]
[('For', 'IN'), ('fifty', 'JJ'), ('years', 'NNS'), ('developing', 'VBG'), ('nation', 'NN'), ('.', '.')]
[('It', 'PRP'), ('time', 'NN'), ('see', 'VB'), ('developed', 'JJ'), ('nation', 'NN'), ('.', '.')]
[('We', 'PRP'), ('among', 'IN'), ('top', 'JJ'), ('5', 'CD'), ('nations', 'NNS'), ('world', 'NN'), ('terms', 'NNS'), ('GDP', 'NNP')
, ('.', '.')]
[('We', 'PRP'), ('10', 'CD'), ('percent', 'JJ'), ('growth', 'NN'), ('rate', 'NN'), ('areas', 'NNS'), ('.', '.')]
[('Our', 'PRP$'), ('poverty', 'NN'), ('levels', 'NNS'), ('falling', 'VBG'), ('.', '.')]
[('Our', 'PRP$'), ('achievements', 'NNS'), ('globally', 'RB'), ('recognised', 'VBN'), ('today', 'NN'), ('.', '.')]
[('Yet', 'RB'), ('lack', 'JJ'), ('self-confidence', 'NN'), ('see', 'NN'), ('developed', 'JJ'), ('nation', 'NN'), (',', ','), ('sel
f-reliant', 'JJ'), ('self-assured', 'JJ'), ('.', '.')]
[('Isn', 'NNP'), ('’', 'NNP'), ('incorrect', 'NN'), ('?', '.')]
[('I', 'PRP'), ('third', 'JJ'), ('vision', 'NN'), ('.', '.')]
[('India', 'NNP'), ('must', 'MD'), ('stand', 'VB'), ('world', 'NN'), ('.', '.')]
[('Because', 'IN'), ('I', 'PRP'), ('believe', 'VBP'), ('unless', 'IN'), ('India', 'NNP'), ('stands', 'VBZ'), ('world', 'NN'), (','
, ','), ('one', 'CD'), ('respect', 'NN'), ('us', 'PRP'), ('.', '.')]
[('Only', 'RB'), ('strength', 'NN'), ('respects', 'NNS'), ('strength', 'NN'), ('.', '.')]
[('We', 'PRP'), ('must', 'MD'), ('strong', 'JJ'), ('military', 'JJ'), ('power', 'NN'), ('also', 'RB'), ('economic', 'JJ'), ('power
', 'NN'), ('.', '.')]
[('Both', 'DT'), ('must', 'MD'), ('go', 'VB'), ('hand-in-hand', 'NN'), ('.', '.')]
[('My', 'PRP$'), ('good', 'JJ'), ('fortune', 'NN'), ('worked', 'VBD'), ('three', 'CD'), ('great', 'JJ'), ('minds', 'NNS'), ('.', '
.')]
[('Dr.', 'NNP'), ('Vikram', 'NNP'), ('Sarabhai', 'NNP'), ('Dept', 'NNP'), ('.', '.')]
[('space', 'NN'), (',', ','), ('Professor', 'NNP'), ('Satish', 'NNP'), ('Dhawan', 'NNP'), (',', ','), ('succeeded', 'VBD'), ('Dr.'
, 'NNP'), ('Brahm', 'NNP'), ('Prakash', 'NNP'), (',', ','), ('father', 'RB'), ('nuclear', 'JJ'), ('material', 'NN'), ('.', '.')]
```

```
[('I', 'PRP'), ('lucky', 'VBP'), ('worked', 'VBD'), ('three', 'CD'), ('closely', 'RB'), ('consider', 'VBP'), ('great', 'JJ'), ('op
portunity', 'NN'), ('life', 'NN'), ('.', '.')]
[('I', 'PRP'), ('see', 'VBP'), ('four', 'CD'), ('milestones', 'NNS'), ('career', 'NN')]
```

In [ ]:

```
"Taj Mahal is a beautiful Monument".split()
```

Out[ ]:

```
['Taj', 'Mahal', 'is', 'a', 'beautiful', 'Monument']
```

In [ ]:

```
print(nltk.pos_tag("Taj Mahal is a beautiful Monument".split()))
```

```
[('Taj', 'NNP'), ('Mahal', 'NNP'), ('is', 'VBZ'), ('a', 'DT'), ('beautiful', 'JJ'), ('Monument', 'NN')]
```

In [ ]:

```
sentence="The Eiffel Tower was built from 1887 to 1889 by Gustave Eiffel, whose company specialized in building metal frameworks a
nd structures."
import nltk
words=nltk.word_tokenize(sentence)
tag_elements=nltk.pos_tag(words)
```

In [ ]:

```
nltk.download('maxent_ne_chunker')
```

```
[nltk_data] Downloading package maxent_ne_chunker to
[nltk_data]     /root/nltk_data...
[nltk_data]   Unzipping chunkers/maxent_ne_chunker.zip.
```

Out[ ]:

```
True
```

In [ ]:

```
nltk.download('words')
```

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

Out[ ]:

```
True
```

In [ ]:

```
nltk.ne_chunk(tag_elements).pprint()
```

```
(S
  The/DT
  (ORGANIZATION Eiffel/NNP Tower/NNP)
  was/VBD
  built/VBN
  from/IN
  1887/CD
  to/TO
  1889/CD
  by/IN
  (PERSON Gustave/NNP Eiffel/NNP)
  ,/,
  whose/WP$
  company/NN
  specialized/VBD
  in/IN
  building/NN
  metal/NN
  frameworks/NNS
  and/CC
  structures/NNS
  ./.)
```

In [ ]:

```

```

In [ ]:

```

```

In [ ]:

```

```

In [ ]:

```

```

In [ ]:

```

```

In [ ]:

```

```

In [ ]:

In [ ]:

In [ ]: