

```
import nltk
```

```
paragraph = """"I have 3 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"""
```

```
nltk.download('stopwords')
```

```
[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\Dell\AppData\Roaming\nltk_data...
[nltk_data] Unzipping corpora\stopwords.zip.
True
```

```
# Cleaning the texts
```

```
import re
from nltk.corpus import stopwords
from nltk.stem.porter import PorterStemmer
from nltk.stem import WordNetLemmatizer
```

```
print(stopwords.words('english'))
```

```
['i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you', "you're", "you've", "you'll", "you'd", 'your', 'yours', 'yourself',
```

```
import nltk
```

```
import nltk
nltk.download('wordnet')
nltk.download('omw-1.4')
nltk.download('punkt')
```

```
[nltk_data] Downloading package wordnet to
[nltk_data] C:\Users\Dell\AppData\Roaming\nltk_data...
[nltk_data] Package wordnet is already up-to-date!
[nltk_data] Downloading package omw-1.4 to
[nltk_data] C:\Users\Dell\AppData\Roaming\nltk_data...
[nltk_data] Package omw-1.4 is already up-to-date!
[nltk_data] Downloading package punkt to
[nltk_data] C:\Users\Dell\AppData\Roaming\nltk_data...
[nltk_data] Unzipping tokenizers\punkt.zip.
True
```

```

ps = PorterStemmer()
wordnet=WordNetLemmatizer()
sentences = nltk.sent_tokenize(paragraph)
# print(sentences)
# print(len(sentences))
corpus = []
for i in range(len(sentences)):
    review = re.sub('[^a-zA-Z]', ' ', sentences[i])
    review = review.lower()
    review = review.split()
#     review = [ps.stem(word) for word in review if not word in set(stopwords.words('english'))]
    review = [wordnet.lemmatize(word) for word in review if not word in set(stopwords.words('english'))]
    review = ' '.join(review)
    corpus.append(review)
print(corpus)
print(len(corpus))

```

['vision india', 'year history people world come invaded u captured land conquered mind', 'alexander onwards greek turk mogul portuguese  
31

```

review1=sentences[0].lower()
for i in review1:
    print(i)

```

i  
h  
a  
v  
e  
  
3  
  
v  
i  
s  
i  
o  
n  
s  
  
f  
o  
r  
  
i  
n  
d  
i  
a  
.

```

test_sent=re.sub('[^a-zA-Z]', ' ',sentences[0])
test_sent

```

'I have visions for India '

```
print(corpus)
```

['vision india', 'year history people world come invaded u captured land conquered mind', 'alexander onwards greek turk mogul portuguese

```

from sklearn.feature_extraction.text import CountVectorizer
cv = CountVectorizer()
X = cv.fit_transform(corpus).toarray()
X

```

```

array([[0, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 1, 1, 0],
       [0, 1, 0, ..., 0, 0, 0],
       ...,
       [0, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0]], dtype=int64)

```

[illegible]

31