

```
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) (8.1.6)
Requirement already satisfied: joblib in /usr/local/lib/python3.10/dist-packages (from nltk) (1.3.2)
Requirement already satisfied: regex<=2021.8.3 in /usr/local/lib/python3.10/dist-packages (from nltk) (2023.6.3)
Requirement already satisfied: tqdm in /usr/local/lib/python3.10/dist-packages (from nltk) (4.66.1)
```

```
import nltk
```

```
nltk.download()
```

```

📁 NLTK Downloader
-----
d) Download  l) List  u) Update  c) Config  h) Help  q) Quit
-----
Downloader> d

Download which package (l=list; x=cancel)?
Identifier> punkt
Downloading package punkt to /root/nltk_data...
Unzipping tokenizers/punkt.zip.

-----
d) Download  l) List  u) Update  c) Config  h) Help  q) Quit
-----
Downloader> q
True

```

```
from nltk.tokenize import sent_tokenize
```

```
text = " Myself Chirag thanth from Vidyavardhinis college of Enginnering . The college is affiliated to University of Mumbai . "
```

```
text
```

```
' Myself Chirag thanth from Vidyavardhinis college of Enginnering . The college is affiliated to University of Mumbai . '
```

```
sentences = sent_tokenize (text)
```

```
sentences
```

```
[' Myself Chirag thanth from Vidyavardhinis college of Enginnering .',
 'The college is affiliated to University of Mumbai .']
```

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

```
words = word_tokenize (text)
```

```
words
```

```
['Myself',
 'Chirag',
 'thanth',
 'from',
 'Vidyavardhinis',
 'college',
 'of',
 'Enginnering',
 '.',
 'The',
 'college',
 'is',
 'affiliated',
 'to',
 'University',
 'of',
 'Mumbai',
 '.']
```

```
for w in words:
    print (w)
```

```
Myself
Chirag
thanth
```

```

from
Vidyavardhinis
college
of
Enginnering
.
The
college
is
affiliated
to
University
of
Mumbai
.

```

```
sent_tokenize (text)
```

```

[' Myself Chirag thanth from Vidyavardhinis college of Enginnering .',
 'The college is affiliated to University of Mumbai .']

```

```
[word_tokenize (text) for t in sent_tokenize(text)]
```

```

[['Myself',
 'Chirag',
 'thanth',
 'from',
 'Vidyavardhinis',
 'college',
 'of',
 'Enginnering',
 '.',
 'The',
 'college',
 'is',
 'affiliated',
 'to',
 'University',
 'of',
 'Mumbai',
 '.'],
 ['Myself',
 'Chirag',
 'thanth',
 'from',
 'Vidyavardhinis',
 'college',
 'of',
 'Enginnering',
 '.',
 'The',
 'college',
 'is',
 'affiliated',
 'to',
 'University',
 'of',
 'Mumbai',
 '.']]

```

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

```
wordpunct_tokenize (text)
```

```

['Myself',
 'Chirag',
 'thanth',
 'from',
 'Vidyavardhinis',
 'college',
 'of',
 'Enginnering',
 '.',
 'The',
 'college',
 'is',
 'affiliated',
 'to',
 'University',
 'of',
 'Mumbai',
 '.']

```

```
text.lower()
```

```
' myself chirag thanth from vidyavardhinis college of enginnering . the college is affiliated to university of mumbai . '
```

```
text.upper()
```

```
' MYSELF CHIRAG THANTH FROM VIDYAVARDHINIS COLLEGE OF ENGINEERING . THE COLLEGE IS AFFILIATED TO UNIVERSITY OF MUMBAI . '
```