```
In [1]: from nltk.tokenize import sent_tokenize, word_tokenize
```

```
In [2]: import nltk
        nltk.download('punkt')
```

```
[nltk_data] Downloading package punkt to
[nltk_data]     C:\Users\sriva\AppData\Roaming\nltk_data...
[nltk_data]   Package punkt is already up-to-date!
Out[2]: True
```

```
In [3]:
text = "Natural language processing (NLP) is a field " + \
       "of computer science, artificial intelligence " + \
       "and computational linguistics concerned with " + \
       "the interactions between computers and human " + \
       "(natural) languages, and, in particular, " + \
       "concerned with programming computers to " + \
       "fruitfully process large natural language " + \
       "corpora. Challenges in natural language " + \
       "processing frequently involve natural " + \
       "language understanding, natural language" + \
       "generation frequently from formal, machine" + \
       "-readable logical forms), connecting language " + \
       "and machine perception, managing human-" + \
       "computer dialog systems, or some combination " + \
       "thereof."

print(sent_tokenize(text))
print(word_tokenize(text))
word = word_tokenize(text)
```

```
['Natural language processing (NLP) is a field of computer science, artificial intelligence and computational l
inguistics concerned with the interactions between computers and human (natural) languages, and, in particular,
concerned with programming computers to fruitfully process large natural language corpora.', 'Challenges in nat
ural language processing frequently involve natural language understanding, natural languagegeneration frequent
ly from formal, machine-readable logical forms), connecting language and machine perception, managing human-com
puter dialog systems, or some combination thereof.']
['Natural', 'language', 'processing', '(', 'NLP', ')', 'is', 'a', 'field', 'of', 'computer', 'science', ',', 'a
rtificial', 'intelligence', 'and', 'computational', 'linguistics', 'concerned', 'with', 'the', 'interactions',
'between', 'computers', 'and', 'human', '(', 'natural', ')', 'languages', ',', 'and', ',', 'in', 'particular',
',', 'concerned', 'with', 'programming', 'computers', 'to', 'fruitfully', 'process', 'large', 'natural', 'langu
age', 'corpora', '.', 'Challenges', 'in', 'natural', 'language', 'processing', 'frequently', 'involve', 'natura
l', 'language', 'understanding', ',', 'natural', 'languagegeneration', 'frequently', 'from', 'formal', ',', 'ma
chine-readable', 'logical', 'forms', ')', ',', 'connecting', 'language', 'and', 'machine', 'perception', ',',
'managing', 'human-computer', 'dialog', 'systems', ',', 'or', 'some', 'combination', 'thereof', '.']
```

```
In [4]: lem = nltk.WordNetLemmatizer()
```

```
In [5]: ps = nltk.PorterStemmer()
```

```
In [17]: wordbank = nltk.tokenize.TreebankWordTokenizer
```

```
In [23]: sen = sent_tokenize(text)
         sen
```

```
Out[23]: ['Natural language processing (NLP) is a field of computer science, artificial intelligence and computational l
inguistics concerned with the interactions between computers and human (natural) languages, and, in particular,
concerned with programming computers to fruitfully process large natural language corpora.',
 'Challenges in natural language processing frequently involve natural language understanding, natural language
generation frequently from formal, machine-readable logical forms), connecting language and machine perception,
managing human-computer dialog systems, or some combination thereof.']
```

```
In [22]: wordbank.tokenize_sents(sen)
```

```
---------------------------------------------------------------------------
TypeError                                 Traceback (most recent call last)
Input In [22], in <cell line: 1>()
----> 1 wordbank.tokenize_sents(sen)

TypeError: TokenizerI.tokenize_sents() missing 1 required positional argument: 'strings'
```

```
In [24]: lis5 = []
         for i in sen:
             lis5.append(wordbank.tokenize(i))
         lis5
```

```
---------------------------------------------------------------------------
TypeError                                 Traceback (most recent call last)
Input In [24], in <cell line: 2>()
      1 lis5 = []
      2 for i in sen:
----> 3     lis5.append(wordbank.tokenize(i))
      4 lis5

TypeError: TreebankWordTokenizer.tokenize() missing 1 required positional argument: 'text'
```

```
In [16]: lis5 = []
         for i in word:
             lis5.append(wordbank.tokenize(word))
         lis5
```

```
---------------------------------------------------------------------------
TypeError                                 Traceback (most recent call last)
Input In [16], in <cell line: 2>()
      1 lis5 = []
      2 for i in word:
----> 3     lis5.append(wordbank.tokenize(word))
      4 lis5

TypeError: TreebankWordTokenizer.tokenize() missing 1 required positional argument: 'text'
```

```
In [6]: lis = []
        for i in word:
            lis.append(ps.stem(i))
        lis
```

```
Out[6]: ['natur',
 'languag',
 'process',
 '(',
 'nlp',
 ')',
 'is',
 'a',
 'field',
 'of',
 'comput',
 'scienc',
 ',',
 'artifici',
 'intellig',
 'and',
 'comput',
 'linguist',
 'concern',
 'with',
 'the',
 'interact',
 'between',
 'comput',
 'and',
 'human',
 '(',
 'natur',
 ')',
 'languag',
 ',',
 'and',
 ',',
 'in',
 'particular',
 ',',
 'concern',
 'with',
 'program',
 'comput',
 'to',
 'fruit',
 'process',
 'larg',
 'natur',
 'languag',
 'corpora',
 '.',
 'challeng',
 'in',
 'natur',
 'languag',
 'process',
 'frequent',
 'involv',
 'natur',
 'languag',
 'understand',
 ',',
 'natur',
 'languagegener',
 'frequent',
 'from',
 'formal',
 ',',
 'machine-read',
 'logic',
 'form',
 ')',
 ',',
 'connect',
 'languag',
 'and',
 'machin',
 'percept',
 ',',
 'manag',
 'human-comput',
 'dialog',
 'system',
 ',',
 'or',
 'some',
 'combin',
 'thereof',
 '.']
```

```
In [7]: lis2 = []
        for i in lis:
            lis2.append(lem.lemmatize(i))
        lis2
```

```
Out[7]: ['natur',
 'languag',
 'process',
 '(',
 'nlp',
 ')',
 'is',
 'a',
 'field',
 'of',
 'comput',
 'scienc',
 ',',
 'artifici',
 'intellig',
 'and',
 'comput',
 'linguist',
 'concern',
 'with',
 'the',
 'interact',
 'between',
 'comput',
 'and',
 'human',
 '(',
 'natur',
 ')',
 'languag',
 ',',
 'and',
 ',',
 'in',
 'particular',
 ',',
 'concern',
 'with',
 'program',
 'comput',
 'to',
 'fruit',
 'process',
 'larg',
 'natur',
 'languag',
 'corpus',
 '.',
 'challeng',
 'in',
 'natur',
 'languag',
 'process',
 'frequent',
 'involv',
 'natur',
 'languag',
 'understand',
 ',',
 'natur',
 'languagegener',
 'frequent',
 'from',
 'formal',
 ',',
 'machine-read',
 'logic',
 'form',
 ')',
 ',',
 'connect',
 'languag',
 'and',
 'machin',
 'percept',
 ',',
 'manag',
 'human-comput',
 'dialog',
 'system',
 ',',
 'or',
 'some',
 'combin',
 'thereof',
 '.']
```

```
In [8]: import string as st
        lis3=[]
        for i in lis2:
            if i not in st.punctuation:
                lis3.append(i)

        lis3
```

```
Out[8]: ['natur',
 'languag',
 'process',
 'nlp',
 'is',
 'a',
 'field',
 'of',
 'comput',
 'scienc',
 'artifici',
 'intellig',
 'and',
 'comput',
 'linguist',
 'concern',
 'with',
 'the',
 'interact',
 'between',
 'comput',
 'and',
 'human',
 'natur',
 'languag',
 'and',
 'in',
 'particular',
 'concern',
 'with',
 'program',
 'comput',
 'to',
 'fruit',
 'process',
 'larg',
 'natur',
 'languag',
 'corpus',
 'challeng',
 'in',
 'natur',
 'languag',
 'process',
 'frequent',
 'involv',
 'natur',
 'languag',
 'understand',
 'natur',
 'languagegener',
 'frequent',
 'from',
 'formal',
 'machine-read',
 'logic',
 'form',
 'connect',
 'languag',
 'and',
 'machin',
 'percept',
 'manag',
 'human-comput',
 'dialog',
 'system',
 'or',
 'some',
 'combin',
 'thereof']
```

```
In [9]: from nltk.tag import DefaultTagger

        tagging = DefaultTagger('NN')
        lis4 =[]
        for i in lis3:
            lis4.append(tagging.tag(i))
```

```
In [ ]: nltk.download()
```

```
In [10]: nltk.download('averaged_perceptron_tagger')
```

```
[nltk_data] Downloading package averaged_perceptron_tagger to
[nltk_data]     C:\Users\sriva\AppData\Roaming\nltk_data...
[nltk_data]   Unzipping taggers\averaged_perceptron_tagger.zip.
Out[10]: True
```

```
In [13]: nltk.pos_tag(lis3)
```

```
Out[13]: [('natur', 'JJ'),
 ('languag', 'NN'),
 ('process', 'NN'),
 ('nlp', 'NN'),
 ('is', 'VBZ'),
 ('a', 'DT'),
 ('field', 'NN'),
 ('of', 'IN'),
 ('comput', 'NN'),
 ('scienc', 'NN'),
 ('artifici', 'NN'),
 ('intellig', 'NN'),
 ('and', 'CC'),
 ('comput', 'NN'),
 ('linguist', 'NN'),
 ('concern', 'NN'),
 ('with', 'IN'),
 ('the', 'DT'),
 ('interact', 'NN'),
 ('between', 'IN'),
 ('comput', 'NN'),
 ('and', 'CC'),
 ('human', 'JJ'),
 ('natur', 'NN'),
 ('languag', 'NN'),
 ('and', 'CC'),
 ('in', 'IN'),
 ('particular', 'JJ'),
 ('concern', 'NN'),
 ('with', 'IN'),
 ('program', 'NN'),
 ('comput', 'NN'),
 ('to', 'TO'),
 ('fruit', 'VB'),
 ('process', 'NN'),
 ('larg', 'NN'),
 ('natur', 'NN'),
 ('languag', 'NN'),
 ('corpus', 'NN'),
 ('challeng', 'NN'),
 ('in', 'IN'),
 ('natur', 'JJ'),
 ('languag', 'JJ'),
 ('process', 'NN'),
 ('frequent', 'JJ'),
 ('involv', 'NN'),
 ('natur', 'JJ'),
 ('languag', 'NN'),
 ('understand', 'NN'),
 ('natur', 'NN'),
 ('languagegener', 'NN'),
 ('frequent', 'NN'),
 ('from', 'IN'),
 ('formal', 'JJ'),
 ('machine-read', 'JJ'),
 ('logic', 'JJ'),
 ('form', 'NN'),
 ('connect', 'NN'),
 ('languag', 'NN'),
 ('and', 'CC'),
 ('machin', 'JJ'),
 ('percept', 'NN'),
 ('manag', 'VBD'),
 ('human-comput', 'JJ'),
 ('dialog', 'NN'),
 ('system', 'NN'),
 ('or', 'CC'),
 ('some', 'DT'),
 ('combin', 'NN'),
 ('thereof', 'NN')]
```

```
In [ ]:
```