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

```
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
[nltk_data] Downloading package wordnet to /root/nltk_data...
[nltk_data]   Package wordnet is already up-to-date!
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Package punkt is already up-to-date!
[nltk_data] Downloading package omw-1.4 to /root/nltk_data...
[nltk_data]   Package omw-1.4 is already up-to-date!
[nltk_data] Downloading package gutenberg to /root/nltk_data...
[nltk_data]   Package gutenberg is already up-to-date!
[nltk_data] Downloading package genesis to /root/nltk_data...
[nltk_data]   Unzipping corpora/genesis.zip.
True
```

```
import re
import sys
from collections import Counter, defaultdict, namedtuple
from functools import reduce
from math import log

from nltk.collocations import BigramCollocationFinder
from nltk.lm import MLE
from nltk.lm.preprocessing import padded_everygram_pipeline
from nltk.metrics import BigramAssocMeasures, f_measure
from nltk.probability import ConditionalFreqDist as CFD
from nltk.probability import FreqDist
from nltk.tokenize import sent_tokenize
from nltk.util import LazyConcatenation, tokenwrap
```

```
from nltk.book import *
```

```
*** Introductory Examples for the NLTK Book ***
Loading text1, ..., text9 and sent1, ..., sent9
Type the name of the text or sentence to view it.
Type: 'texts()' or 'sents()' to list the materials.
text1: Moby Dick by Herman Melville 1851
text2: Sense and Sensibility by Jane Austen 1811
text3: The Book of Genesis
text4: Inaugural Address Corpus
text5: Chat Corpus
text6: Monty Python and the Holy Grail
```

```
text7: Wall Street Journal
text8: Personals Corpus
text9: The Man Who Was Thursday by G . K . Chesterton 1908
```

nltk download

this downloads all the relevant/necessary items from nltk

```
nltk.download()
```

```
    Unzipping corpora/state_union.zip.
Downloading package stopwords to /root/nltk_data...
  Package stopwords is already up-to-date!
Downloading package subjectivity to /root/nltk_data...
  Unzipping corpora/subjectivity.zip.
Downloading package swadesh to /root/nltk_data...
  Unzipping corpora/swadesh.zip.
Downloading package switchboard to /root/nltk_data...
  Unzipping corpora/switchboard.zip.
Downloading package tagsets to /root/nltk_data...
  Unzipping help/tagsets.zip.
Downloading package timit to /root/nltk_data...
  Unzipping corpora/timit.zip.
Downloading package toolbox to /root/nltk_data...
  Unzipping corpora/toolbox.zip.
Downloading package treebank to /root/nltk_data...
  Unzipping corpora/treebank.zip.
Downloading package twitter_samples to /root/nltk_data...
  Unzipping corpora/twitter_samples.zip.
Downloading package udhr to /root/nltk_data...
  Unzipping corpora/udhr.zip.
Downloading package udhr2 to /root/nltk_data...
  Unzipping corpora/udhr2.zip.
Downloading package unicode_samples to /root/nltk_data...
  Unzipping corpora/unicode_samples.zip.
Downloading package universal_tagset to /root/nltk_data...
  Unzipping taggers/universal_tagset.zip.
Downloading package universal_treebanks_v20 to
    /root/nltk_data...
Downloading package vader_lexicon to /root/nltk_data...
Downloading package verbnet to /root/nltk_data...
  Unzipping corpora/verbnet.zip.
Downloading package verbnet3 to /root/nltk_data...
  Unzipping corpora/verbnet3.zip.
Downloading package webtext to /root/nltk_data...
  Unzipping corpora/webtext.zip.
Downloading package wmt15_eval to /root/nltk_data...
  Unzipping models/wmt15_eval.zip.
Downloading package word2vec_sample to /root/nltk_data...
  Unzipping models/word2vec_sample.zip.
Downloading package wordnet to /root/nltk_data...
  Package wordnet is already up-to-date!
Downloading package wordnet2021 to /root/nltk_data...
```

```
    | Downloading package wordnet31 to /root/nltk_data...
    | Downloading package wordnet_ic to /root/nltk_data...
    |   Unzipping corpora/wordnet_ic.zip.
    | Downloading package words to /root/nltk_data...
    |   Unzipping corpora/words.zip.
    | Downloading package ycoe to /root/nltk_data...
    |   Unzipping corpora/ycoe.zip.
    |
  Done downloading collection all-nltk

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

tokens

the tokens method breaks apart a string into a list it has two options to separate by words or into sentences

```
#tokenize and list the first 20 words in text 1
from nltk.tokenize import word_tokenize


texter = str(text1)
tokens = word_tokenize(texter)
print(tokens[:20])
```

```
['<', 'Text', ':', 'Moby', 'Dick', 'by', 'Herman', 'Melville', '1851', '>']
```

concordance finds the occurences of the word sea, parameters allow to check for a certain number (in this case 5)

```
from nltk.corpus import gutenberg
from nltk.text import Text
concord = text1.concordance("sea", width = 20, lines = 5)
```

```
Displaying 5 of 455 matches:
 in the sea ." -- I
 Indian Sea breedet
 on the sea , when
 of the sea , appea
ing the sea before
```

```
"""
Nltk text count determines the number of times a word appears in a text
The python count function determines the number of elements in a list or other data type
```

```
The python count function determines the number of elements in a list or other data type
"""


text1.count("sea")
```

```
433
```

```
# text is paragraph 1 p.49 of Exploring NLP with Python Building Understanding Through Code b
raw_text = "Linguistics and NLP are closely bound together. In fact, NLP is sometimes called
tokens = word_tokenize(raw_text) #convert the text into tokens (turning it into a list of wor
tokens[:10]
```

```
['Linguistics',
 'and',
 'NLP',
 'are',
 'closely',
 'bound',
 'together',
 '.',
 'In',
 'fact']
```

```
#convert the raw text into sentence tokens
sents = sent_tokenize(raw_text)
sents
```

```
['Linguistics and NLP are closely bound together.',
 'In fact, NLP is sometimes called Computational Linguistics.',
 'Linguistics is the study of human language, and is a fascinating field of study.',
 'Many universities offer advanced degrees in Linguistics.',
 'The goal of this chapter is more modest: To familiarize the reader with terminology
and concepts that are frequently used in NLP.']
```

9. Stem

Stem sometim univers linguist thi terminolog

9.

Stem	Lemma
sometim	sometimes
univers	university
linguist	Linguistics
thi	this
terminolog	terminology

```
import nltk.stem


stemmer = nltk.PorterStemmer()
stemmed = [stemmer.stem(token) for token in tokens]
stemmed
```

```
['linguist',
 'and',
 'nlp',
 'are',
 'close',
 'bound',
 'togeth',
 '.',
 'in',
 'fact',
 ',',
 'nlp',
 'is',
 'sometim',
 'call',
 'comput',
 'linguist',
 '.',
 'linguist',
 'is',
 'the',
 'studi',
 'of',
 'human',
 'languag',
 ',',
 'and',
 'is',
 'a',
 'fascin',
 'field',
```

```
 'of',
 'studi',
 '.',
 'mani',
 'univers',
 'offer',
 'advanc',
 'degre',
 'in',
 'linguist',
 '.',
 'the',
 'goal',
 'of',
 'thi',
 'chapter',
 'is',
 'more',
 'modest',
 ':',
 'to',
 'familiar',
 'the',
 'reader',
 'with',
 'terminolog',
 'and'.
```

```
from nltk.stem.wordnet import WordNetLemmatizer
lemmer = WordNetLemmatizer()
lemma = [lemmer.lemmatize(token) for token in tokens]
lemma
```

```
['Linguistics',
 'and',
 'NLP',
 'are',
 'closely',
 'bound',
 'together',
 '.',
 'In',
 'fact',
 ',',
 'NLP',
 'is',
 'sometimes',
 'called',
 'Computational',
 'Linguistics',
 '.',
 'Linguistics',
 'is',
 'the',
 'study',
 'of',
```

```
        'human',
        'language',
        ',',
        'and',
        'is',
        'a',
        'fascinating',
        'field',
        'of',
        'study',
        '.',
        'Many',
        'university',
        'offer',
        'advanced',
        'degree',
        'in',
        'Linguistics',
        '.',
        'The',
        'goal',
        'of',
        'this',
        'chapter',
        'is',
        'more',
        'modest',
        ':',
        'To',
        'familiarize',
        'the',
        'reader',
        'with',
        'terminology',
        'and',
```

```
"""
The nltk library simplifies a lot of tasks that seem to be commonly used in language processi
I find this to be very useful because it is much easier to process meaning when it can be bro
The nltk code has good documentation which makes it easy to use and understand.
I previously made a scrabble game in one of my much earlier python classes.
I would love to make the game more complex by implementing some of the functions in the nltk

"""
```

Colab paid products - Cancel contracts here

✓ 0s completed at 5:40 PM