

NLTK

<https://pythonprogramming.net/tokenizing-words-sentences-nltk-tutorial/>
(<https://pythonprogramming.net/tokenizing-words-sentences-nltk-tutorial/>).

In []:

```
import sys
!{sys.executable} -m pip install nltk
```

In [5]:

```
import nltk
nltk.download()
from nltk.tokenize import sent_tokenize, word_tokenize

EXAMPLE_TEXT = "Hello Mr. Smith, how are you doing today? The weather is great, and Python is awesome. The sky is pinkish-blue. You shouldn't eat cardboard."

print(sent_tokenize(EXAMPLE_TEXT))
print(word_tokenize(EXAMPLE_TEXT))
```

NLTK Downloader

d) Download l) List u) Update c) Config h) Help q) Quit

Download which package (l=list; x=cancel)?

Downloading collection 'all'

```
Downloading package abc to /home/jupyterlab/nltk_data...
  Unzipping corpora/abc.zip.
Downloading package alpino to /home/jupyterlab/nltk_data...
  Unzipping corpora/alpino.zip.
Downloading package biocreative_ppi to
  /home/jupyterlab/nltk_data...
  Unzipping corpora/biocreative_ppi.zip.
Downloading package brown to /home/jupyterlab/nltk_data...
  Unzipping corpora/brown.zip.
Downloading package brown_tei to
  /home/jupyterlab/nltk_data...
  Unzipping corpora/brown_tei.zip.
Downloading package cess_cat to /home/jupyterlab/nltk_data...
  Unzipping corpora/cess_cat.zip.
Downloading package cess_esp to /home/jupyterlab/nltk_data...
  Unzipping corpora/cess_esp.zip.
Downloading package chat80 to /home/jupyterlab/nltk_data...
  Unzipping corpora/chat80.zip.
Downloading package city_database to
  /home/jupyterlab/nltk_data...
  Unzipping corpora/city_database.zip.
Downloading package cmudict to /home/jupyterlab/nltk_data...
  Unzipping corpora/cmudict.zip.
Downloading package comparative_sentences to
  /home/jupyterlab/nltk_data...
  Unzipping corpora/comparative_sentences.zip.
Downloading package comtrans to /home/jupyterlab/nltk_data...
Downloading package conll2000 to
  /home/jupyterlab/nltk_data...
  Unzipping corpora/conll2000.zip.
Downloading package conll2002 to
  /home/jupyterlab/nltk_data...
  Unzipping corpora/conll2002.zip.
Downloading package conll2007 to
  /home/jupyterlab/nltk_data...
Downloading package crubadan to /home/jupyterlab/nltk_data...
  Unzipping corpora/crubadan.zip.
Downloading package dependency_treebank to
  /home/jupyterlab/nltk_data...
  Unzipping corpora/dependency_treebank.zip.
Downloading package dolch to /home/jupyterlab/nltk_data...
  Unzipping corpora/dolch.zip.
Downloading package europarl_raw to
  /home/jupyterlab/nltk_data...
  Unzipping corpora/europarl_raw.zip.
Downloading package floresta to /home/jupyterlab/nltk_data...
  Unzipping corpora/floresta.zip.
Downloading package framenet_v15 to
  /home/jupyterlab/nltk_data...
  Unzipping corpora/framenet_v15.zip.
Downloading package framenet_v17 to
  /home/jupyterlab/nltk_data...
  Unzipping corpora/framenet_v17.zip.
Downloading package gazetteers to
  /home/jupyterlab/nltk_data...
```

```
| Unzipping corpora/gazetteers.zip.  
| Downloading package genesis to /home/jupyterlab/nltk_data...  
| Unzipping corpora/genesis.zip.  
| Downloading package gutenber to  
| /home/jupyterlab/nltk_data...  
| Unzipping corpora/gutenberg.zip.  
| Downloading package ieer to /home/jupyterlab/nltk_data...  
| Unzipping corpora/ieer.zip.  
| Downloading package inaugural to  
| /home/jupyterlab/nltk_data...  
| Unzipping corpora/inaugural.zip.  
| Downloading package indian to /home/jupyterlab/nltk_data...  
| Unzipping corpora/indian.zip.  
| Downloading package jeita to /home/jupyterlab/nltk_data...  
| Downloading package kimmo to /home/jupyterlab/nltk_data...  
| Unzipping corpora/kimmo.zip.  
| Downloading package knbc to /home/jupyterlab/nltk_data...  
| Downloading package lin_thesaurus to  
| /home/jupyterlab/nltk_data...  
| Unzipping corpora/lin_thesaurus.zip.  
| Downloading package mac_morpho to  
| /home/jupyterlab/nltk_data...  
| Unzipping corpora/mac_morpho.zip.  
| Downloading package machado to /home/jupyterlab/nltk_data...  
| Downloading package masc_tagged to  
| /home/jupyterlab/nltk_data...  
| Downloading package moses_sample to  
| /home/jupyterlab/nltk_data...  
| Unzipping models/moses_sample.zip.  
| Downloading package movie_reviews to  
| /home/jupyterlab/nltk_data...  
| Unzipping corpora/movie_reviews.zip.  
| Downloading package names to /home/jupyterlab/nltk_data...  
| Unzipping corpora/names.zip.  
| Downloading package nombank.1.0 to  
| /home/jupyterlab/nltk_data...  
| Downloading package nps_chat to /home/jupyterlab/nltk_data...  
| Unzipping corpora/nps_chat.zip.  
| Downloading package omw to /home/jupyterlab/nltk_data...  
| Unzipping corpora/omw.zip.  
| Downloading package opinion_lexicon to  
| /home/jupyterlab/nltk_data...  
| Unzipping corpora/opinion_lexicon.zip.  
| Downloading package paradigms to  
| /home/jupyterlab/nltk_data...  
| Unzipping corpora/paradigms.zip.  
| Downloading package pil to /home/jupyterlab/nltk_data...  
| Unzipping corpora/pil.zip.  
| Downloading package pl196x to /home/jupyterlab/nltk_data...  
| Unzipping corpora/pl196x.zip.  
| Downloading package ppattach to /home/jupyterlab/nltk_data...  
| Unzipping corpora/ppattach.zip.  
| Downloading package problem_reports to  
| /home/jupyterlab/nltk_data...  
| Unzipping corpora/problem_reports.zip.  
| Downloading package propbank to /home/jupyterlab/nltk_data...  
| Downloading package ptb to /home/jupyterlab/nltk_data...
```

```
Unzipping corpora/ptb.zip.
Downloading package product_reviews_1 to
/home/jupyterlab/nltk_data...
Unzipping corpora/product_reviews_1.zip.
Downloading package product_reviews_2 to
/home/jupyterlab/nltk_data...
Unzipping corpora/product_reviews_2.zip.
Downloading package pros_cons to
/home/jupyterlab/nltk_data...
Unzipping corpora/pros_cons.zip.
Downloading package qc to /home/jupyterlab/nltk_data...
Unzipping corpora/qc.zip.
Downloading package reuters to /home/jupyterlab/nltk_data...
Downloading package rte to /home/jupyterlab/nltk_data...
Unzipping corpora/rte.zip.
Downloading package semcor to /home/jupyterlab/nltk_data...
Downloading package senseval to /home/jupyterlab/nltk_data...
Unzipping corpora/senseval.zip.
Downloading package sentiwordnet to
/home/jupyterlab/nltk_data...
Unzipping corpora/sentiwordnet.zip.
Downloading package sentence_polarity to
/home/jupyterlab/nltk_data...
Unzipping corpora/sentence_polarity.zip.
Downloading package shakespeare to
/home/jupyterlab/nltk_data...
Unzipping corpora/shakespeare.zip.
Downloading package sinica_treebank to
/home/jupyterlab/nltk_data...
Unzipping corpora/sinica_treebank.zip.
Downloading package smultron to /home/jupyterlab/nltk_data...
Unzipping corpora/smultron.zip.
Downloading package state_union to
/home/jupyterlab/nltk_data...
Unzipping corpora/state_union.zip.
Downloading package stopwords to
/home/jupyterlab/nltk_data...
Unzipping corpora/stopwords.zip.
Downloading package subjectivity to
/home/jupyterlab/nltk_data...
Unzipping corpora/subjectivity.zip.
Downloading package swadesh to /home/jupyterlab/nltk_data...
Unzipping corpora/swadesh.zip.
Downloading package switchboard to
/home/jupyterlab/nltk_data...
Unzipping corpora/switchboard.zip.
Downloading package timit to /home/jupyterlab/nltk_data...
Unzipping corpora/timit.zip.
Downloading package toolbox to /home/jupyterlab/nltk_data...
Unzipping corpora/toolbox.zip.
Downloading package treebank to /home/jupyterlab/nltk_data...
Unzipping corpora/treebank.zip.
Downloading package twitter_samples to
/home/jupyterlab/nltk_data...
Unzipping corpora/twitter_samples.zip.
Downloading package udhr to /home/jupyterlab/nltk_data...
Unzipping corpora/udhr.zip.
```

```
Downloading package udhr2 to /home/jupyterlab/nltk_data...
Unzipping corpora/udhr2.zip.
Downloading package unicode_samples to
/home/jupyterlab/nltk_data...
Unzipping corpora/unicode_samples.zip.
Downloading package universal_treebanks_v20 to
/home/jupyterlab/nltk_data...
Downloading package verbnet to /home/jupyterlab/nltk_data...
Unzipping corpora/verbnet.zip.
Downloading package verbnet3 to /home/jupyterlab/nltk_data...
Unzipping corpora/verbnet3.zip.
Downloading package webtext to /home/jupyterlab/nltk_data...
Unzipping corpora/webtext.zip.
Downloading package wordnet to /home/jupyterlab/nltk_data...
Unzipping corpora/wordnet.zip.
Downloading package wordnet_ic to
/home/jupyterlab/nltk_data...
Unzipping corpora/wordnet_ic.zip.
Downloading package words to /home/jupyterlab/nltk_data...
Unzipping corpora/words.zip.
Downloading package ycoe to /home/jupyterlab/nltk_data...
Unzipping corpora/ycoe.zip.
Downloading package rslp to /home/jupyterlab/nltk_data...
Unzipping stemmers/rslp.zip.
Downloading package maxent_treebank_pos_tagger to
/home/jupyterlab/nltk_data...
Unzipping taggers/maxent_treebank_pos_tagger.zip.
Downloading package universal_tagset to
/home/jupyterlab/nltk_data...
Unzipping taggers/universal_tagset.zip.
Downloading package maxent_ne_chunker to
/home/jupyterlab/nltk_data...
Unzipping chunkers/maxent_ne_chunker.zip.
Downloading package punkt to /home/jupyterlab/nltk_data...
Unzipping tokenizers/punkt.zip.
Downloading package book_grammars to
/home/jupyterlab/nltk_data...
Unzipping grammars/book_grammars.zip.
Downloading package sample_grammars to
/home/jupyterlab/nltk_data...
Unzipping grammars/sample_grammars.zip.
Downloading package spanish_grammars to
/home/jupyterlab/nltk_data...
Unzipping grammars/spanish_grammars.zip.
Downloading package basque_grammars to
/home/jupyterlab/nltk_data...
Unzipping grammars/basque_grammars.zip.
Downloading package large_grammars to
/home/jupyterlab/nltk_data...
Unzipping grammars/large_grammars.zip.
Downloading package tagsets to /home/jupyterlab/nltk_data...
Unzipping help/tagsets.zip.
Downloading package snowball_data to
/home/jupyterlab/nltk_data...
Downloading package bllip_wsj_no_aux to
/home/jupyterlab/nltk_data...
Unzipping models/bllip_wsj_no_aux.zip.
```

```

| Downloading package word2vec_sample to
|   /home/jupyterlab/nltk_data...
|   Unzipping models/word2vec_sample.zip.
| Downloading package panlex_swadesh to
|   /home/jupyterlab/nltk_data...
| Downloading package mte_teip5 to
|   /home/jupyterlab/nltk_data...
|   Unzipping corpora/mte_teip5.zip.
| Downloading package averaged_perceptron_tagger to
|   /home/jupyterlab/nltk_data...
|   Unzipping taggers/averaged_perceptron_tagger.zip.
| Downloading package averaged_perceptron_tagger_ru to
|   /home/jupyterlab/nltk_data...
|   Unzipping taggers/averaged_perceptron_tagger_ru.zip.
| Downloading package perluniprops to
|   /home/jupyterlab/nltk_data...
|   Unzipping misc/perluniprops.zip.
| Downloading package nonbreaking_prefixes to
|   /home/jupyterlab/nltk_data...
|   Unzipping corpora/nonbreaking_prefixes.zip.
| Downloading package vader_lexicon to
|   /home/jupyterlab/nltk_data...
| Downloading package porter_test to
|   /home/jupyterlab/nltk_data...
|   Unzipping stemmers/porter_test.zip.
| Downloading package wmt15_eval to
|   /home/jupyterlab/nltk_data...
|   Unzipping models/wmt15_eval.zip.
| Downloading package mwa_ppdb to /home/jupyterlab/nltk_data...
|   Unzipping misc/mwa_ppdb.zip.

```

Done downloading collection all

```

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

```

```

['Hello Mr. Smith, how are you doing today?', 'The weather is great, and Pyth
on is awesome.', 'The sky is pinkish-blue.', "You shouldn't eat cardboard."]
['Hello', 'Mr.', 'Smith', ',', 'how', 'are', 'you', 'doing', 'today', '?', 'T
he', 'weather', 'is', 'great', ',', 'and', 'Python', 'is', 'awesome', '.', 'T
he', 'sky', 'is', 'pinkish-blue', '.', 'You', 'should', "n't", 'eat', 'cardbo
ard', '.']

```

In []: