[Tokenizing Words and Sentences with NLTK](https://pythonspot.com/tokenizing-words-and-sentences-with-nltk/)

NLTK is one of the leading platforms for working with human language data and Python, the module NLTK is used for natural language processing. NLTK is literally an acronym for Natural Language Toolkit

Install NLTK

Install NLTK with Python 2.x using:

|  |
| --- |
| sudo pip install nltk |

Install NLTK with Python 3.x using:

|  |
| --- |
| sudo pip3 install nltk |

Installation is not complete after these commands. Open python and type:

|  |
| --- |
| **import** nltk  nltk.download() |

A graphical interface will be presented:

Click all and then click download. It will download all the required packages which may take a while, the bar on the bottom shows the progress.

Tokenize words

A sentence or data can be split into words using the method **word\_tokenize():**

|  |
| --- |
| **from** nltk.tokenize **import** sent\_tokenize, word\_tokenize    data = "All work and no play makes jack a dull boy, all work and no play"  **print**(word\_tokenize(data)) |

This will output: