Chapter 11

Anatomizing Lexicon With Natural Language Tokenizer Toolkit 3

Simran Kaur Jolly
Manav Rachna International Institute of Research and Studies, India

Rashmi Agrawal
Manav Rachna International Institute of Research and Studies, India

ABSTRACT

NLTK toolkit is an API platform built with Python language to interact with humans through natural language. The very first version of NLTK was released in 2005 (1.4.3), which was compatible with Python 2.4. The latest version was in September 2017 NLTK (3.2.5), which incorporated features like Arabic stemmers, NIST evaluation, MOSES tokenizer, Stanford segmenter, treebank detokenizer, verbnet, and vader, etc. NLTK was created in 2001 as a part of Computational Linguistic Department at the University of Pennsylvania. Since then it has been tested and developed. The important packages of this system are 1) corpus builder, 2) tokenizer, 3) collocation, 4) tagging, 5) parsing, 6) metrics, and 7) probability distribution system. Toolbox NLTK was built to meet four primary requirements: 1) Simplicity: An substantive framework for building blocks; 2) Consistency: Consistent interface; 3) Extensibility: Which can be easily scaled; and 4) Modularity: All modules are independent of each other.

DOI: 10.4018/978-1-5225-6117-0.ch011
INSTALLING PYTHON AND NLTK

In order to install python 3.4 version go to https://python.org and install the version 3.4. (Figure 1)

After installing python version 3.4 install the NLTK toolkit version 3.0.

1. First we install NLTK(natural language toolkit): pip install nltk.
2. Install Numpy (optional) package if user needs: pip install numpy.
3. While testing the installation of nltk toolkit we can run it on python GUI and write the command:
   >>>import nltk(it will import the whl(wheel files) of nltk and related packages)

There are two existing versions of nlp i.e. python 2.7 and python 3.4 which are very much incompatible with each other. The python 3.y versions are more coherent, more consistent and user friendly GUI is provided. All the instructions written for versions of python 2x may not run in version of python 3y and if they run the output of the code id different in both the versions. Not all the organizations have updated to python 3.y versions and are still relying on python 2.x versions due to the ongoing service and credibility.

INSTALLING PYTHON 3.4 ON WINDOWS SYSTEMS

There are two variants of Python 3.4 for Windows — a 32-bit version and a 64-bit version. Obviously, the 64-bit version requires a 64-bit Windows computer.

Figure 1. Installation of Python 3.4
Related Content

Electronic Records Management - An Old Solution to a New Problem: Governments Providing Usable Information to Stakeholders
[www.igi-global.com/chapter/electronic-records-management---an-old-solution-to-a-new-problem/150264?camid=4v1a](www.igi-global.com/chapter/electronic-records-management---an-old-solution-to-a-new-problem/150264?camid=4v1a)

Swarm Intelligence in Text Document Clustering
[www.igi-global.com/chapter/swarm-intelligence-text-document-clustering/21723?camid=4v1a](www.igi-global.com/chapter/swarm-intelligence-text-document-clustering/21723?camid=4v1a)
Modeling Interpretable Fuzzy Rule-Based Classifiers for Medical Decision Support

The Application of Data Mining for Drought Monitoring and Prediction
Tsegaye Tadesse, Brian Wardlow and Michael J. Hayes (2009). Data Mining Applications for Empowering Knowledge Societies (pp. 278-289). www.igi-global.com/chapter/application-data-mining-drought-monitoring/7557?camid=4v1a