The NLTK Python library

Claudio Greco

September 25, 2017

What is Python?

- Python[1] is a widely used high-level interpreted programming language for general-purpose programming, created by Guido van Rossum and first released in 1991
- It presents a syntax which allows to express complex concepts in a few lines of code while emphasizing code readability
- It supports multiple programming paradigms, including object-oriented, imperative, and functional paradigms
- It presents a large and comprehensive standard library
- Many additional libraries exist to extend its capabilities
- Both the versions 2 and 3 of Python are very popular

---

[1]https://www.python.org/

What is NLTK?

- Natural Language Toolkit (NLTK)[2] is a suite of libraries for building Python programs to work with human language
- It provides easy-to-use interfaces to over 50 corpora and lexical resources, along with libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning
- It is intended to support research and teaching in fields such as Computational Linguistics and Natural Language Processing
- It supports both the versions 2 and 3 of Python

Installing Python

- Python comes pre-installed with Linux and macOS
- On Windows, execute the installer available at the URL:
  `https://www.python.org/downloads/windows/`

Installing NLTK

- ▶ On Windows, execute the installer available at the URL:
  `http://pypi.python.org/pypi/nltk`
- ▶ On Linux:
  1. Install the pip package manager. On Ubuntu you can install it by executing the following command in a system shell:
     ```
     sudo apt-get install python-pip
     ```
  2. Execute the following command in a system shell:
     ```
     sudo pip install -U nltk
     ```

- ▶ On macOS:
  1. Install the pip package manager by executing the following command in a system shell:
     ```
     sudo easy_install pip
     ```
  2. Execute the following command in a system shell:
     ```
     sudo pip install -U nltk
     ```

Reference: `http://www.nltk.org/install.html`

Downloading NLTK data (optional: for tokenizers, etc.)

- Run the following code in a Python environment:
  ```
  import nltk
  nltk.download()
  ```

- Otherwise, run the following command in a system shell:
  ```
  python -m nltk.downloader all
  ```

Reference: http://www.nltk.org/data.html

Do you need help?

Write to me at `claudio.greco@unitn.it`