

Welcome to Hands-on Week 03

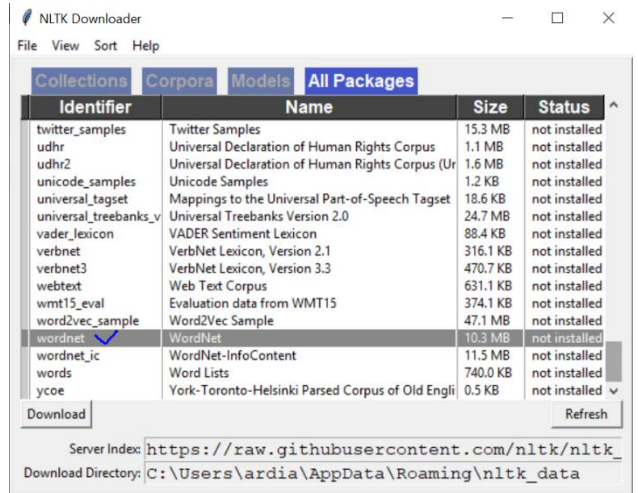
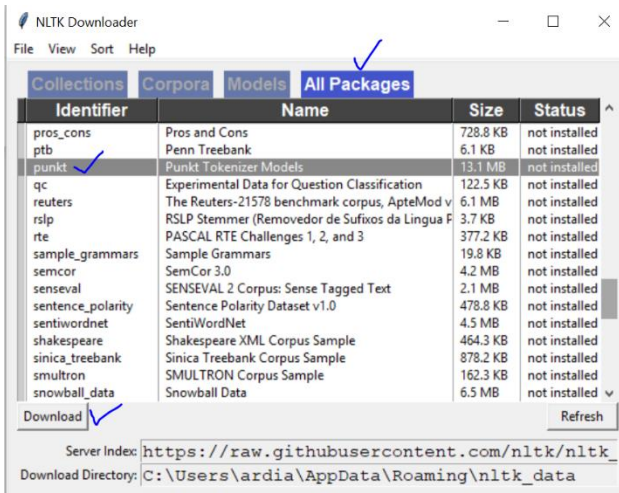
In this week, we will play around text pre-processing, where text is one of data forms that we may encounter other than numerical data, image/video and voice. To do this, we will use NLTK library (*Natural Language Toolkit*) and scikit-learn. Follow this guideline to prepare all the things:

Note: you need to finish this before we start the hands-on class on Wednesday, 4 September 2019.

1. Install NLTK into your python environment (via terminal) by typing `pip install nltk`
2. Enter python environment in the terminal by typing `python`, followed by typing `import nltk` and `nltk.download()` as shown below.

```
(datamining) C:\Users\ardia>python
Python 3.6.9 |Anaconda, Inc.| (default, Jul 30 2019, 14:00:49) [MSC v.1915 64 bit
Type "help", "copyright", "credits" or "license" for more information.
>>> import nltk
>>> nltk.download()
showing info https://raw.githubusercontent.com/nltk/nltk_data/gh-pages/index.xml
```

3. By doing step (2), a new window will appear. Click “All Packages” >> “punkt” >> “Download”, as shown below (*left*). Repeat this step for “wordnet” (*right picture*).



4. Second library we will use is scikit-learn. Install it by typing `pip install scikit-learn`
5. Download “W03_handson_materials.zip”. We will give the password in the class.

~End~