Software Installations
Seth D Temple
December 14, 2021

This document includes instructions to install necessary programs for text data analysis. The procedure below assumes that the client does not have Python (Anaconda distribution) installed *a priori*. Existing versions of Python, Anaconda, and GitHub may expedite this process.

1. **Install** the individual edition of **Anaconda** (link) appropriate for your operating system. This software is a distribution manager for the Python and R programming languages. Use the most recent version of Python. Further instructions are documented here.
2. Open the software Anaconda Prompt (or some other terminal/console that has conda). **Install** the following Python **libraries** for machine learning **(ML)** and natural language processing **(NLP)**. Many of these may be pre-installed with the distribution. Use the terminal commands below. If you run into issues, you may consider commands of this flavor: pip install [library]. Otherwise, search StackExchange or StackOverflow for help.
    a. conda install numpy pandas matplotlib sklearn nltk
    b. conda install -c conda-forge gensim textblob wordcloud
3. Download **NLTK data** in Python through the terminal/console. In the terminal/console (e.g. Anaconda Prompt), type and Enter the following commands:
    a. python
    b. import nltk, gensim
    c. nltk.download('wordnet')
    d. nltk.download('stopwords')
4. (Optional) You could consider the following downloads of NLTK data from the terminal/console if you run into issues later on.
    a. python -m nltk.downloader popular
    b. python -m textblob.download_corpora
    c. python -m textblob.download_corpora lite
5. Visit the **GitHub repository** and download it (Green button saying 'Code'). Either download as a zipper folder and unzip, or use the GitHub Desktop app. Make sure to have the folder 'coilnlp' with the accompanying code files in a folder where you plan to perform the text data analysis.
6. **Enter coding environment**: (1) use commands 'cd folder/subfolder/subsubfolder/' and 'dir' ('ls') on Windows (Mac/Linux) to move about local folder structure and then 'jupyter notebook' to work in IPython notebooks, or (2) 'python' to work in terminal/console.
7. Advice: when downloading packages and data from the terminal/console, parse the print output to ensure no errors. After importing the 'coilnlp' package in a Python session or IPython notebook, use help(function_name) to see documentation. Please reach out to sdtemple@uw.edu if you run into other issues.