

Computing the Classics with the Classical Language Toolkit, Part 1

What is the CLTK?

The Classical Language Toolkit (CLTK) is a library of tools for processing and analyzing ancient texts. Many of the tools included automate laborious philological tasks, such as lexical analysis, scansion, and concordance creation, while other tools offer new means of analysis by utilizing modern natural language processing (NLP) techniques. In this series of blog posts, I will show how to harness the CLTK for your own research projects by demonstrating these tools in several mini-projects.

Installation

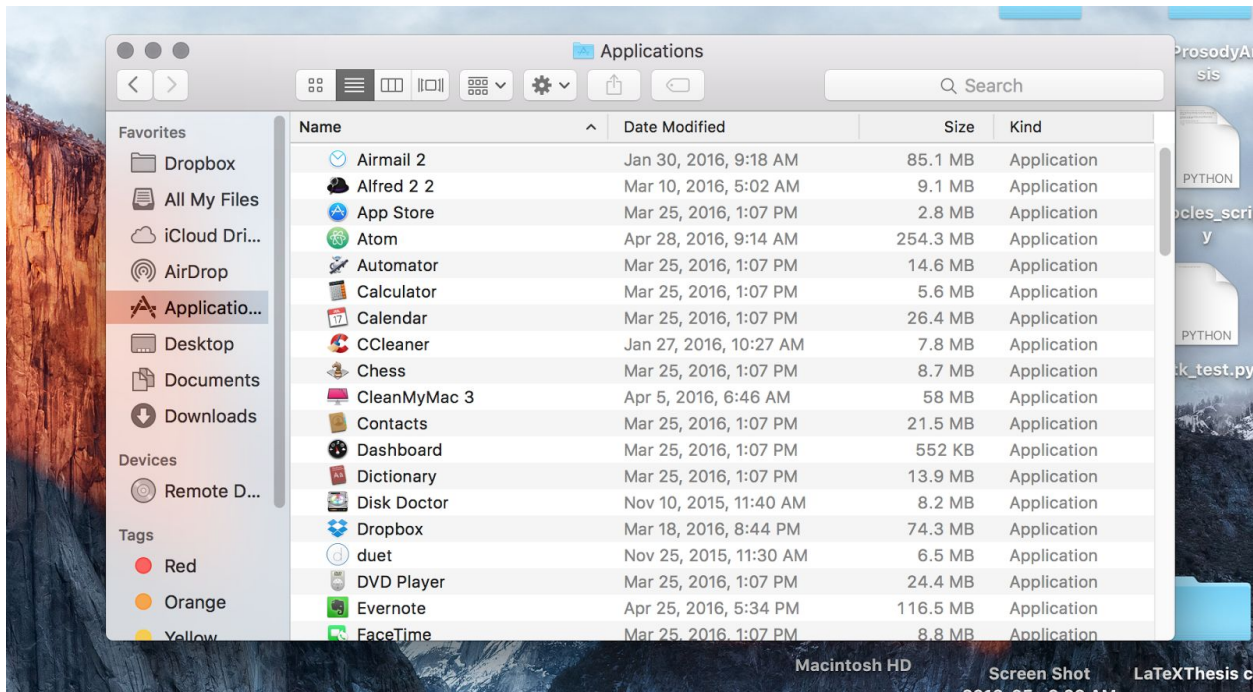
There are a number of software components required for using the `cltk`. Since it is a Python 3 library, meaning that it is written in and requires the Python 3 programming language to function, we will begin by installing Python 3. **NB: The `cltk` only works on POSIX-compliant operating systems (Linux, Mac OSX, FreeBSD, etc.), which excludes Windows operating systems.**

Installing Python 3.5

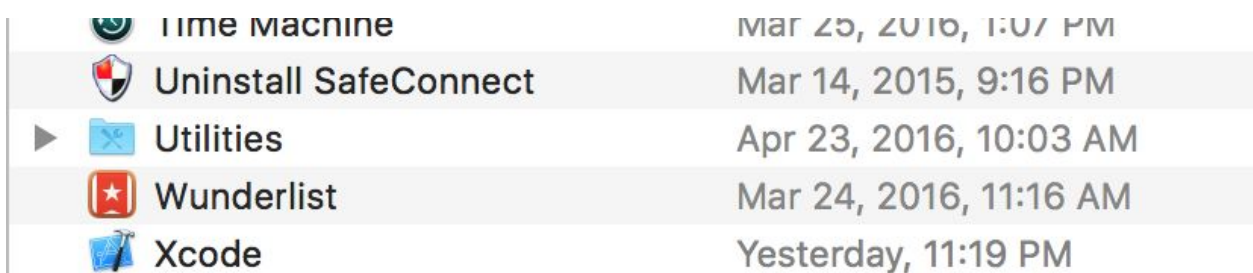
1. If you have a Mac, [click here](#) and download the latest Python 3 release. If you use Linux, use your distribution's package manager. The following instructions are Mac-specific.
2. Once you have downloaded the installation package, open it and follow the installation instructions. If you receive a Gatekeeper warning, you may have to change your downloading privileges. Instructions on how to do this can be found [here](#).
3. After opening the installation package and installing Python, we should ensure that everything was properly installed. To do this, we will need to use the terminal/command line.
4. Open up the Finder application on your Dock (it's the one that looks like a face).



5. Click on the Applications folder.



6. Open the Utilities folder in the Applications directory.



Then click on the Terminal

	Migration Assistant	Mar 25, 2016, 1:07 PM	2.7 MB
	Script Editor	Mar 25, 2016, 1:07 PM	9.3 MB
	System Information	Mar 25, 2016, 1:07 PM	262.7 MB
	Terminal	Mar 25, 2016, 1:07 PM	8.6 MB
	VoiceOver Utility	Mar 25, 2016, 1:07 PM	9.7 MB
	X11	May 28, 2015, 4:23 PM	7.2 MB

7. Enter `python3 --version` into the terminal and you should receive an output similar to the one below.

```
[Josephs-MacBook-Pro:~ Tyler$ python3 --version
Python 3.5.1
```

8. If the output detailed your version of Python 3, then it was successfully installed. We will now install the CLTK. I recommend making a shortcut to the Terminal from the desktop or dock since we will be using it often.

Installing the cltk

1. We will be working in a virtual environment to install the CLTK and its dependencies. Don't worry! That sounds more complicated than it really is.
2. Set up the virtual environment by entering `pyenv venv`. Then enter `source venv/bin/activate`.

```
[Josephs-MacBook-Pro:~ Tyler$ pyenv venv
[Josephs-MacBook-Pro:~ Tyler$ source venv/bin/activate
(venv) Josephs-MacBook-Pro:~ Tyler$
```

3. Now install the CLTK using pip, a package management system that comes with Python 3. First upgrade pip with the command `pip install --upgrade pip`. Then install the cltk with the command `pip install cltk`. Note that you will see a lot of activity in the terminal if you installed it correctly. **NB: You may have to agree to an Xcode license before installing anything via the terminal. If prompted to do so, enter `sudo xcodebuild license`, hit space to progress through the agreement, then finally enter `agree`. Alternatively, you can download Xcode for free from the App Store and agree to the license within the program itself.**

```
(venv) Josephs-MacBook-Pro:~ Tyler$ pip install cltk
Collecting cltk
  Downloading cltk-0.1.34.tar.gz (292kB)
    100% |#####| 294kB 585kB/s
Collecting fuzzywuzzy (from cltk)
  Downloading fuzzywuzzy-0.10.0.tar.gz
Collecting gitpython (from cltk)
  Downloading GitPython-2.0.2-py2.py3-none-any.whl (425kB)
    100% |#####| 425kB 482kB/s
Collecting nltk (from cltk)
  Downloading nltk-3.2.1.tar.gz (1.1MB)
    100% |#####| 1.1MB 330kB/s
Collecting python-Levenshtein (from cltk)
  Downloading python-Levenshtein-0.12.0.tar.gz (48kB)
    100% |#####| 49kB 454kB/s
Collecting pyuca (from cltk)
  Downloading pyuca-1.1.tar.gz (525kB)
    100% |#####| 528kB 436kB/s
Collecting regex (from cltk)
  Downloading regex-2016.04.25.tar.gz (579kB)
    100% |#####| 581kB 408kB/s
Collecting whoosh (from cltk)
  Downloading Whoosh-2.7.4-py2.py3-none-any.whl (468kB)
    100% |#####| 471kB 539kB/s
Collecting gitdb<=0.6.4 (from gitpython->cltk)
  Downloading gitdb-0.6.4.tar.gz (400kB)
    100% |#####| 401kB 459kB/s
Requirement already satisfied (use --upgrade to upgrade): setuptools in ./venv/lib/python3.5/site-packages (from python-Levenshtein->cltk)
Collecting smmap<=0.8.5 (from gitdb<=0.6.4->gitpython->cltk)
  Downloading smmap-0.9.0.tar.gz
Installing collected packages: fuzzywuzzy, smmap, gitdb, gitpython, nltk, python-Levenshtein, pyuca, regex, whoosh, cltk
Running setup.py install for fuzzywuzzy
Running setup.py install for smmap
Running setup.py install for gitdb
Running setup.py install for nltk
Running setup.py install for python-Levenshtein
Running setup.py install for pyuca
Running setup.py install for regex
Running setup.py install for cltk
Successfully installed cltk-0.1.34 fuzzywuzzy-0.10.0 gitdb-0.6.4 gitpython-2.0.2 nltk-3.2.1 python-Levenshtein-0.12.0 pyuca-1.1 regex-2016.4.25 smmap-0.9.0 whoosh-2.7.4
You are using pip version 7.1.2, however version 8.1.1 is available.
You should consider upgrading via the 'pip install --upgrade pip' command.
(venv) Josephs-MacBook-Pro:~ Tyler$
```

- Let's test the CLTK to ensure that it was installed correctly. We will be using a simple module that translates Beta Code Greek (Greek transliterated into ASCII) into Unicode. We start a new session of Python in the shell by first entering `python3`. Begin the cltk test by importing the tool we need with the code `from cltk.corpus.greek.beta_to_unicode import Replacer`. Next, specify the beta code for translation: `beta = r""""MH=NIN""""`. Then prepare the tool for use with `r = Replacer()`. Finally, translate the beta code: `r.beta_code(beta)`. Your terminal should return `μῆνιν`. If you have any issues with the CLTK, please post questions [here](#), where one of the developers will you out.

```
[Josephs-MacBook-Pro:~ Tyler$ python3
Python 3.5.1 (v3.5.1:37a07cee5969, Dec 5 2015, 21:12:44)
[GCC 4.2.1 (Apple Inc. build 5666) (dot 3)] on darwin
Type "help", "copyright", "credits" or "license" for more information.
[>>> from cltk.corpus.greek.beta_to_unicode import Replacer
[>>> beta = r""""MH=NIN""""
[>>> r = Replacer()
[>>> r.beta_code(beta)
'μῆνιν'
>>>
```

We're almost done! Now we just need to install IPython and Git, both of which will be used for the upcoming mini-projects.

Installing IPython & Git

IPython is a friendly interactive shell, which is useful for casual programming. Git is not Python-specific but very common for software version control.

1. In the same terminal session, install IPython via pip install with the command ``pip install ipython``

```
(venv) Josepchs-MacBook-Pro:~ Tyler$ pip install ipython
Collecting ipython
  Downloading ipython-4.2.0-py3-none-any.whl (736kB)
    100% |#####| 737kB 652kB/s
Collecting pexpect (from ipython)
  Downloading pexpect-4.0.1.tar.gz (143kB)
    100% |#####| 153kB 1.5MB/s
Collecting setuptools>=18.5 (from ipython)
  Downloading setuptools-20.10.1-py2.py3-none-any.whl (509kB)
    100% |#####| 512kB 901kB/s
Collecting decorator (from ipython)
  Downloading decorator-4.0.9-py2.py3-none-any.whl
Collecting gnu readline (from ipython)
  Downloading gnu readline-6.3.3-cp35-cp35m-macosx_10_6_intel.macosx_10_9_intel.macosx_10_9_x86_64.macosx_10_10_intel.macosx_10_10_x86_64.whl (265kB)
    100% |#####| 266kB 967kB/s
Collecting appnope (from ipython)
  Downloading appnope-0.1.0-py2.py3-none-any.whl
Collecting backports.shutil-get-terminal-size (from ipython)
  Downloading backports.shutil-get-terminal-size-1.0.0-py2.py3-none-any.whl
Collecting simplegeneric>0.8 (from ipython)
  Downloading simplegeneric-0.8.1.zip
Collecting traitlets (from ipython)
  Downloading traitlets-4.2.1-py2.py3-none-any.whl (67kB)
    100% |#####| 71kB 139kB/s
Collecting pickleshare (from ipython)
  Downloading pickleshare-0.7.2-py2.py3-none-any.whl
Collecting ptyprocess>=0.5 (from pexpect->ipython)
  Downloading ptyprocess-0.5.1-py2.py3-none-any.whl
Collecting ipython-genutils (from traitlets->ipython)
  Downloading ipython_genutils-0.1.0-py2.py3-none-any.whl
Installing collected packages: ptyprocess, pexpect, setuptools, decorator, gnu readline, appnope, backports.shutil-get-terminal-size, simplegeneric,
Running setup.py install for pexpect ... done
Found existing installation: setuptools 18.2
Uninstalling setuptools-18.2:
Successfully uninstalled setuptools-18.2
Running setup.py install for simplegeneric ... done
Successfully installed appnope-0.1.0 backports.shutil-get-terminal-size-1.0.0 decorator-4.0.9 gnu readline-6.3.3 ipython-4.2.0 ipython-genutils-0.1.0
20.10.1 simplegeneric-0.8.1 traitlets-4.2.1
```

2. That's it for installing IPython! You can test it out by entering ``ipython`` into the terminal. If you're interested, entering the Beta Code example above. You'll notice that it has convenient tab-completions.

```
(venv) Josepchs-MacBook-Pro:~ Tyler$ ipython
Python 3.5.1 (v3.5.1:37a07cee5969, Dec 5 2015, 21:12:44)
Type "copyright", "credits" or "license" for more information.

IPython 4.2.0 -- An enhanced Interactive Python.
?          -> Introduction and overview of IPython's features.
%quickref  -> Quick reference.
help       -> Python's own help system.
object?    -> Details about 'object', use 'object??' for extra details.

In [1]:
```

