08 PROGRAMMING FOR CORPUS LINGUISTICS

TU-DA 2019-20

08 NLTK

NLTK (Natural Language Toolkit: http://www.nltk.org)

NLTK book: http://www.nltk.org/book
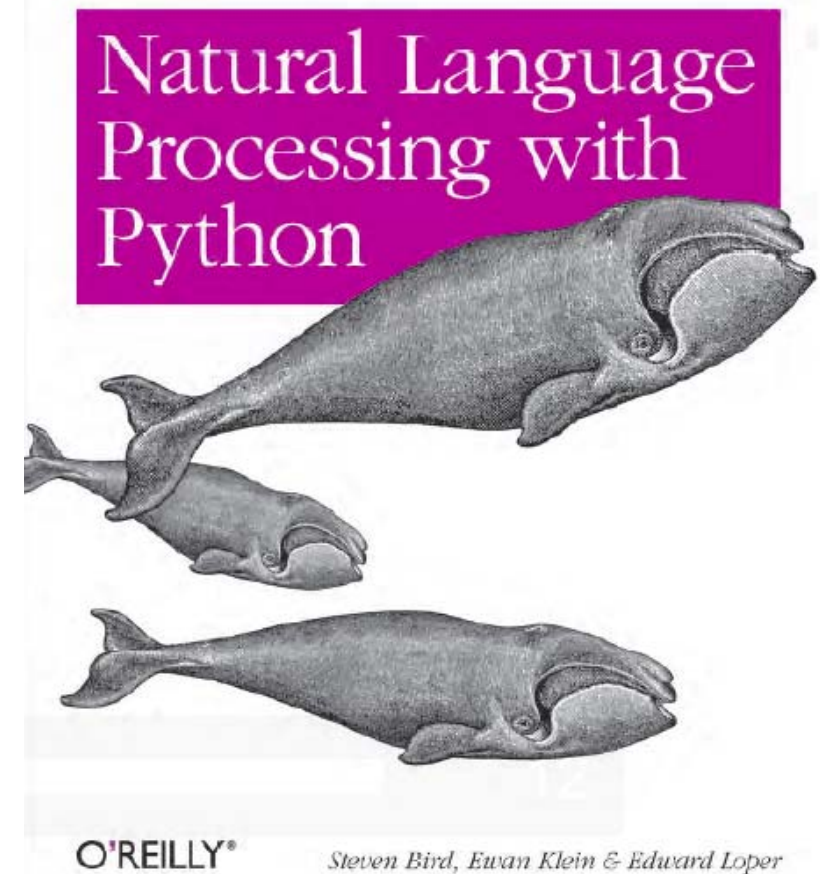
08 NLTK INSTALLATION: IN THE TERMINAL

Installation in the terminal - generic command (please do not type this in before reading on):

```
python -m pip install SomePackage -U
py.exe -m pip install SomePackage -U
```

for Mac OS users:

```
pip3 install SomePackage -U
```

for installation instructions see the Python3 Documentation:

https://www.python.org/doc/

08 NLTK INSTALLATION

Examples:

Windows: `py.exe -m pip install numpy -U`

MaC OS: `pip3 install numpy -U`

repeat the same process to install all packages you want, e.g.
`numpy`, `matplotlib`, `nltk`

08 NLTK INSTALLATION FOR WINDOWS

Open the Powershell with admin priviledges by right-clicking on the Powershell icon and selecting "Run as administrator". Depending on your set-up, you might have to either confirm by clicking "Yes" or you might have to enter an admin password if you are not an admin user on the system your installing to.

At the command prompt, type: `py.exe` or `python.exe` which takes you to the `Interactive Python shell`.

At the prompt which looks like this `>>>` type these three commands, one after the other:

`py.exe -m pip install numpy -U`

`py.exe -m pip install matplotlib -U`

`py.exe -m pip install nltk -U`

In order to test that all has gone to plan, type

`import nltk`

to check that nltk has been installed. If you don't get an error message, you're done! Congratulations!

08 NLTK INSTALLATION FOR MAC OS

In Mac OS, open your Terminal and type `python3` which takes you to the `Interactive Python shell`.

At the prompt which looks like this `>>>` type these three commands, one after the other:

```
pip3 install numpy -U
```

```
pip3 install numpy -U
```

```
pip3 install numpy -U
```

In order to test that all has gone to plan, type

```
import nltk
```

to check that nltk has been installed. If you don't get an error message, you're done! Congratulations!
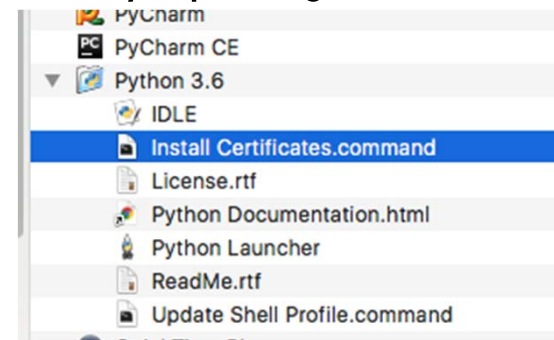
08 NLTK INSTALLATION FOR MAC OS
IMPORTANT EXTRA STEP!!!

CAUTION! Mac OS has a very specific issue with the **Open SSL certificates** that you need to fix before the next step. In order to download and install the SSL certificates to prevent this error, go to your terminal and type the line below. Note that you will have to insert your version number of Python, so if your version of Python is not 3.7 make sure you correct this accordingly:

```
sh "/Applications/Python 3.7/Install Certificates.command"
```

If this does not work for you, you may be able to resolve this by opening **Finder** and double clicking **Install Certificates.command**

PyCharm
PyCharm CE
Python 3.6
IDLE
Install Certificates.command
License.rtf
Python Documentation.html
Python Launcher
ReadMe.rtf
Update Shell Profile.command

reference: https://stackoverflow.com/questions/41348621/ssl-error-downloading-nltk-data

08 NLTK INSTALLATION FOR MAC OS
IMPORTANT EXTRA STEP – WORKAROUND!!!

Workaround – only do this if you did not suceed on the previous slide:

Should you get a PERMISSION DENIED error after running the command line version from the previous slide, please try the following:

In the terminal navigate to the folder

```
Applications/Python\ 3.7
```

and then type

```
sudo ./Install\ Certificates.command
```

08 NLTK_DATA INSTALLATION: WINDOWS AND MAC OS

Installation of additional NLTK packages and data - in the interactive Python shell:

```python
import nltk

nltk.download("all", download_dir="C:\\nltk_data",
halt_on_error=False)
```

This will install the nltk_data directory with all nltk corpora, packages and tools to the specified directory. Note that as this is a GB 3.21download it will take a moment depending on the speed of your network connection.

After this, exit the Interactive Python shell by typing `exit()` and you should be all set to go.

08 NLTK_DATA INSTALLATION: WINDOWS AND MAC OS

Note by the way that if you ever want to install just one individual NLTK package, you can go the same way, but instead of installing `"all"` as in the example on the previous slides, specify the package you want:

```
import nltk

nltk.download("punkt", download_dir="C:\\nltk_data",
halt_on_error=False)
```

08 SOURCES

Eric Matthes. 2016. Python Crash Course. No Starch Press, O'Reilly.

Al Sweigart. 2014. Automate the Boring Stuff with Python: Practical Programming for Total Beginners. No Starch Press, O'Reilly. URL: https://automatetheboringstuff.com/

Steven Bird, Ewan Klein, and Edward Loper. Natural Language Processing with Python – Analyzing Text with the Natural Language Toolkit. URL: http://www.nltk.org/

Stackoverflow SSL error downloading NLTK data: https://stackoverflow.com/questions/41348621/ssl-error-downloading-nltk-data