

Text Analysis with NLTK

Prework

Python: <https://programminghistorian.org/en/lessons/introduction-and-installation>

Jupyter Notebooks: <https://glam-workbench.github.io/getting-started/>

Noteable: <https://www.ed.ac.uk/information-services/learning-technology/noteable/accessing-noteable>

Noteable User Guide: https://noteable.edina.ac.uk/user_guide/#hide_ge_7

Using Jupyter Notebooks and Noteable:

<https://github.com/edina/Exemplars2020/blob/master/TeachingDocs/Tutorials/UsingNoteableBeginner.ipynb>

Week 1: The Natural Language Toolkit (NLTK)

Helpful Resources

- *Natural Language Processing with Python – Analyzing Text with the Natural Language Toolkit*, 3rd Edition, by Steven Bird, Ewan Klein and Edward Loper (2019): <https://www.nltk.org/book/>
- NLTK Documentation: <https://www.nltk.org>
- NLTK Demos (no coding required!): <http://text-processing.com/demo/>

Assignment

1. **Complete W3 Schools' Python RegEx (regular expression) tutorial**
https://www.w3schools.com/python/python_regex.asp
Including the "Try it yourself" buttons!
2. **Watch the video "Understand NLP: NLTK" in section 10 of the LinkedIn Learning course, "Data Science Foundations: Python Scientific Stack"**
<https://www.linkedin.com/learning/data-science-foundations-python-scientific-stack/understand-nlp-nltk?u=50251009>
All University students have access to LinkedIn Learning; you can search for it within the MyEd portal and connect your LinkedIn Learning account to your University account.
3. **Read "Why I Dig: Feminist Approaches to Text Analysis" by Lisa Marie Rhody**
<https://dhdebates.gc.cuny.edu/read/untitled/section/508c8664-15c8-4262-a72a-e49299873d11>
4. **Watch sections 2-5 in the LinkedIn Learning course, "Processing Text with Python Essential Training"**
<https://www.linkedin.com/learning/processing-text-with-python-essential-training/reading-raw-files?u=50251009>
All University students have access to LinkedIn Learning; you can search for it within the MyEd portal and connect your LinkedIn Learning account to your University account.
5. Note how the article from #3 and the video from #4 talk about **stop words**. What differences do you see in the value they each put on stop words?
6. **Complete #4-13 ("Tokenising Text" through "Part-of-Speech Tagging Text") in the *Library Carpentry: Text and Data Mining***
<http://librarycarpentry.org/lc-tdm/>
Complete the exercises in your own Jupyter Notebook!

Go Further

Read: the NLTK book, Chapter 1, *Language Processing with Python*:
<https://www.nltk.org/book/ch01.html>

Read: *Word Embeddings: A Very Short Introduction* by Anouk Lang:
<https://aelang.github.io/word-embeddings>