Text Mining hands-on

03-06-2019

Filip Ilievski

About

The material can be found on GitHub:

https://github.com/cltl/text-mining-ba

Note: next week's material will be updated in the meantime.

Getting started

There is one folder per lab session *X*.

- Text-mining-ba -> lab_sessions -> lab*X*

Content:

- Python Jupyter Notebooks
  - Explanation notebooks (numbered, e.g.,1.1, 1.2, ...)
  - Assignment notebooks
- Text files

Getting started

1. Download/clone the github content from https://github.com/cltl/text-mining-ba
2. Install Anaconda Python 3.7
3. Run Anaconda's jupyterlab environment (will open in your browser)
4. You are ready!

More explanation on how to install and run notebooks:
text-mining-ba/lab_sessions/lab1/Lab1.1-introduction.ipynb

Inside the Notebooks

In general, the notebooks are self-contained, combining:

- Theoretical explanation
- Links to more information
- Installation instructions
- Example code
- Assignments

Lab session 1

Run various text mining functions from two standard packages in Python:

- NLTK
- SpaCy