

# 2019-02-07 Lab 4

## Contents

- [Exercise: Create N-Gram Language Models](#)
- [Exercise: Annotate Corpus with NLTK](#)
- [Homework](#)

## Exercise: Create N-Gram Language Models

- ☐ Follow along review of [Colab Jupyter Notebook on LMs](#).

## Exercise: Annotate Corpus with NLTK

- ☐ Follow along review of [Colab Jupyter Notebooks on NLTK Annotation](#).

## Homework

- ☐ Create an F2 level version of Austen's *Persuasion*.
  - ☐ You may build on the code and source text used in [Lab 2](#) to create your initial tokens table.
  - ☐ Be sure to extract a vocabulary table from the tokens table before proceeding to the next step.
- ☐ Use NLTK's annotation tools to create an F3 version of the text by adding POS, Porter stems, and stopwords.
- ☐ Use a single Spyder or Jupyter file to do your work and upload these files along with your source text file to Collab in the appropriate Assignment.
- ☐ To be clear, your code should produce the following data frames:
  - ☐ A token table with an OHCO index and appropriate features.
  - ☐ A vocabulary table with appropriate features.
  - ☐ The two tables should share a `term_id` series.
- ☐ To receive full credit, your code should be runnable and produce the dataframes described.