NLP session 02: Preparation for next session

- Read Jurafsky & Martin's Chapter 8: "Sequence Labeling for Parts of Speech and Named Entities"
- Work through sections 5-9 from Chapter 1 of the free interactive course "Advanced NLP with spaCy": https://course.spacy.io/en/
- Do (at least) the outstanding exercises 4 and 5 from class (Are the recipes vegan or vegetarian? and What are the 5 most frequently mentioned ingredients?).
- Programming practice:
  - Define a function that takes two strings as arguments, and returns an integer with the number of occurences of the first string in the second one;
  - Embed your function in a script that takes a link to a Project Gutenberg book (`.txt`) as an argument together with a word to query the frequency of. Make the script print: (i) the title of the book and (ii) the frequency of the word passed as a query.

Optional

There are three topics you could introduce next week. This would count toward your in-class participation grade. If you want to present one of them, announce this on this week's forum on Aula Global. In this way others will know that the topic is already taken.

1. Give an introduction to *Part of Speech Tagging*, based on Jurafsky & Martin's Chapter 8 (from the beginning of the chapter to the end of 8.2)
2. Give an introduction to *Named Entity Tagging*, based on Jurafsky & Martin's Chapter 8 (subsection 8.3, focusing on explaining BIO, IO, and BIOES tagging)
3. Introduce HMM Part-of-Speech Tagging, based on Jurafsky & Martin's Chapter 8 (subsection 8.4). This is a more complex topic and you will need more than a few minutes to explain it. Accordingly, it counts for your entire in-class participation grade.