Hands on: Tokenization and Segmentation

3 February 2022
Assignment 2

Due 10 p.m. on Wednesday.

Worksheet 3

Out on Friday
First part due start of class on Tuesday
Today

30 min. Practice using Python to work with strings
45 min. Start Assignment 2 in pairs or individually
Working with Strings in Python
For this part, we’ll use Google Colab (Jupyter notebooks).

These are great for experimenting! But they’re not a great fit for larger programs, including the homework assignments for this course.
Notebook:

colab.research.google.com/drive/19hN8D6ZCJVWr6DREf8zJ_Q8M6MTu84Rj?usp=sharing
Assignment 2
Acknowledgments

This class incorporates material from:

Katrin Erk, University of Texas at Austin