

Programming for Computational Linguistics

2022/2023

Final Project

Due 9.03.2023

1 Introduction

Your task for the final project will be to **design, implement, and document an n -gram language model or naive bayes spam filter in python**. You can find detailed specifications describing the theory and requirements for either project on Ilias. Please read them carefully to get an overview and then pick *one* of the projects.

In the coming days/weeks, you will receive a series of documents that provide step-by-step instructions that you can follow. However, you will not need to follow these exactly, or at all, so long as your program aligns with the specification of your project.

2 Grading

Your grade for the project will be determined by the following rubrik, out of a maximum of 100 points. You have to collect **70 points** to pass the class. Be careful and do not assume that you will receive maximum points for the baseline functionality and the report! It is safer to also think about the quality of your code and proper tests.

50 points — Baseline Functionality Your project has to contain all the functionality described in Section Baseline Functionality. Especially, to receive **any points** for baseline functionality, we have to be able to run all steps, the code has to finish without an exception, and the described output has to be created.

File `specification.pdf` specifies the required components of your model (modules, functions, returned types, etc.). Your program has to meet this specification.

15 points — Code Quality Is your code well-structured and readable? Do you use comments to explain what you are doing and why? Are your variable names descriptive? Do you break subtasks nicely into functions?

15 points — Tests Do you use unit tests to test individual components of your program? How much of your code is covered by these tests? Do these tests cover tricky edge-cases that are likely to fail?

20 points — Report The quality of your report. Does it specify what you did? Does it document how to interact with your project?

20 bonus points — Extensions and Analysis For every extension you implement and document, you can receive up to 10 points. To get the full 10 points, your extension must be fully implemented and properly documented in your report. Feel free to spend your time on the extension(s) which are the most interesting to you. Remember to document them well. If you have an idea for an extension, and are unsure if it would be appropriate — ask us!