

AI611μ Word Prediction with N-Grams Model using Python

Project 1: Simple word prediction application

This assessment evaluates the following competencies:

- *AI101 – Understand the N-Grams model* (+1)
- *AI201 – Train an N-Grams model from a given text corpus* (+2)
- *AI501 – Write an application that solves the word prediction problem with N-Grams models* (+2)
- *AI103 – Preprocess a corpus and compute basic statistics on it* (+2)

In this project assessment, you have to write an application that first train a bigram or trigram model based on a given corpus and then can be used to predict the next word of a sentence whose you provide the beginning. Your application simply has to select the word from the vocabulary that maximises the probability of the sentence obtained by adding the word to the provided sentence beginning. To succeed the assessment, you have to:

1. Write and test your application.
2. Explain to your teacher how you wrote the application, how it works and make him/her a demonstration.
3. Answer any questions that your teacher asks you about your program.

For your application, you can use any existing Python library to help you (such as the `nltk`, for example), which is of course recommended, or write all your code by yourself from scratch.