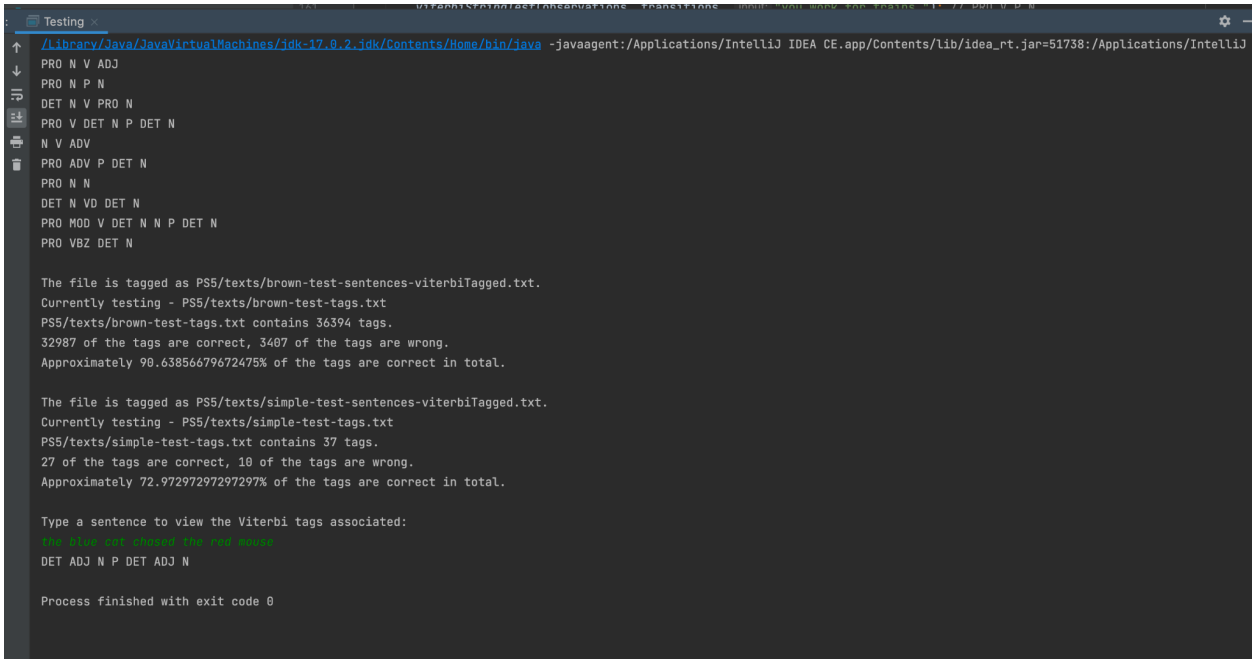
Andrew Cheng



Above is a screenshot of the viterbi tagging program running in the terminal. The first section are the static test cases that use hard-coded test sentences. Sentences such as "your work is beautiful " or "this cave is my night " were fully accurate based on the taggings from the sample texts. However, some sentences like "you work for trains " mistakenly tagged the work word as N instead of V. This could likely be because the training data provided which is used in the hidden markov model class causes certain words to have a significantly lower probability, which causes it to be identified incorrectly. Next were the files that were tested with the viterbi tagging, which ended up with 90% accuracy for the brown test files and 72% for the simple test files. The likely reason for a lower on the simple test files were just the smaller amount of data which led the predictions of the markov model to be more inaccurate. Finally, there was a console input test for the viterbi tagging for the statement “the blue cat chased the red mouse”. In this case, the tagging was correct as the is det, blue is an adjective, cat is a noun, chased is a verb, the is a determinant, red is an adjective, and mouse is a noun. As a result, this passed the test. The screenshot below shows when the unseen penalty was increased to -10 instead of -100. This

represents how it causes it to become more inaccurate likely because a higher penalty allows more ambiguity.