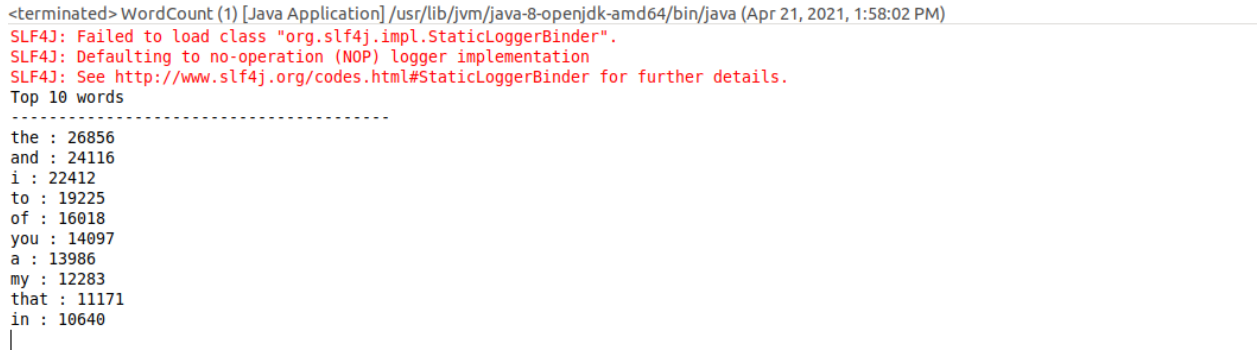
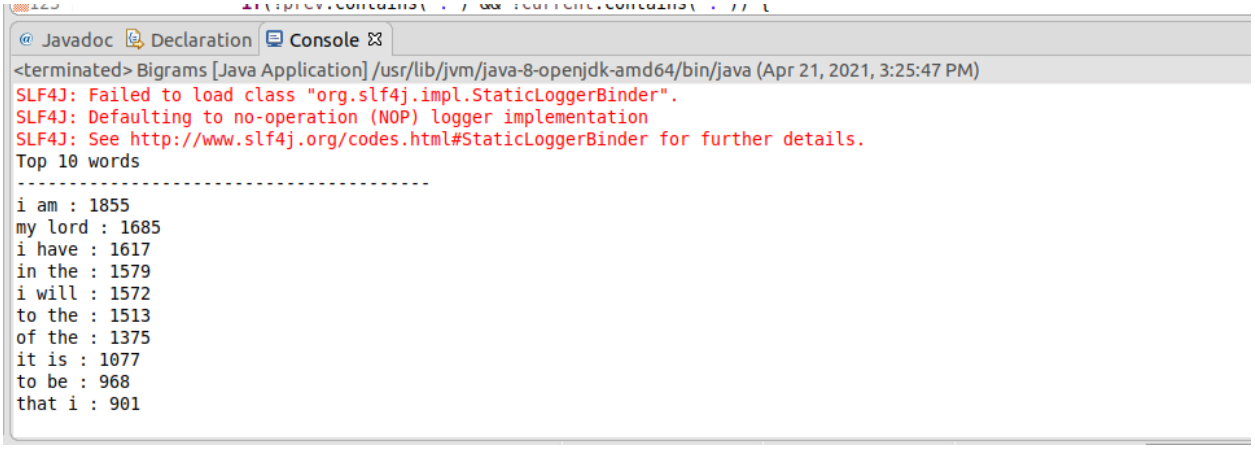
Lab 7

1) The below image is a screenshot obtained from the Exp1 on the Shakespeare dataset. The words were first grouped together using a custom FlatMapFunction and then the counting and sorting were implemented in a sink function.




Results of exp1 bonus are included as a text file called wordcountwindow.txt

2) For this experiment, for bigrams, I replaced words like “i’ll” with “ill” so “i” “ll” will not be included as a bigram. All the question marks and exclamation marks were replaced with “.” for simplicity. It was not clear to me whether the question required to list all the




bigrams or just the top 10. So I have included code for printing the top 10. And also a text file of results if all bigrams are to be printed. Screenshot of exp 2 is pasted below:

The results of exp2 bonus are included under bigramwindow.txt.