1. Input the file data1.txt to a stemmer and lemmatizer of your choice and print both the outputs for all the tokens in the input file. Do not print punctuations in the output.
2. Learn unigram and bi-gram probabilities from the input file data2.txt and print the top 5 unigrams and bi-grams corresponding to highest n-gram probabilities.