Table of Contents

1 Introduction

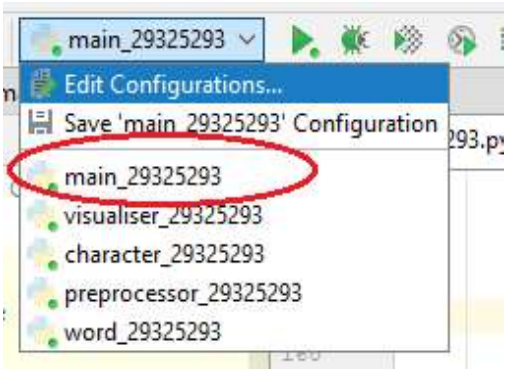
Stylometry is an application of statistical methods for linguistic style analysis. It is often used to attribute authorship to anonymous or disputed works (often in written language) based on the linguistic behaviours or characteristics manifested in the texts. One of the classic authorship attribution studies was whether William Shakespeare wrote all his works, in particular one of his popular plays — Henry VI Trilogy — was highly disputed to have been written or co-authored by Christopher Marlowe.
In this assignment, I have implement the stylometric analyser and represent the character, stop word, punctuation and word length frequency by the visualisation.

2 Steps to run the decoder

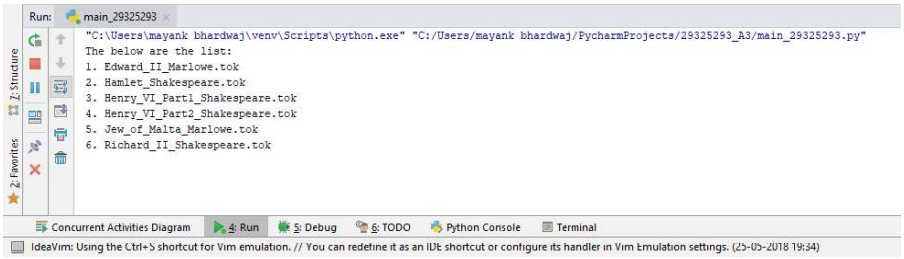
There are four tasks in the assignment which need to run and verify the desired results are given below:

Step 1

1. Go to the PyCharm and open the code file and click the green play button to run the task1 as shown in the below figure.

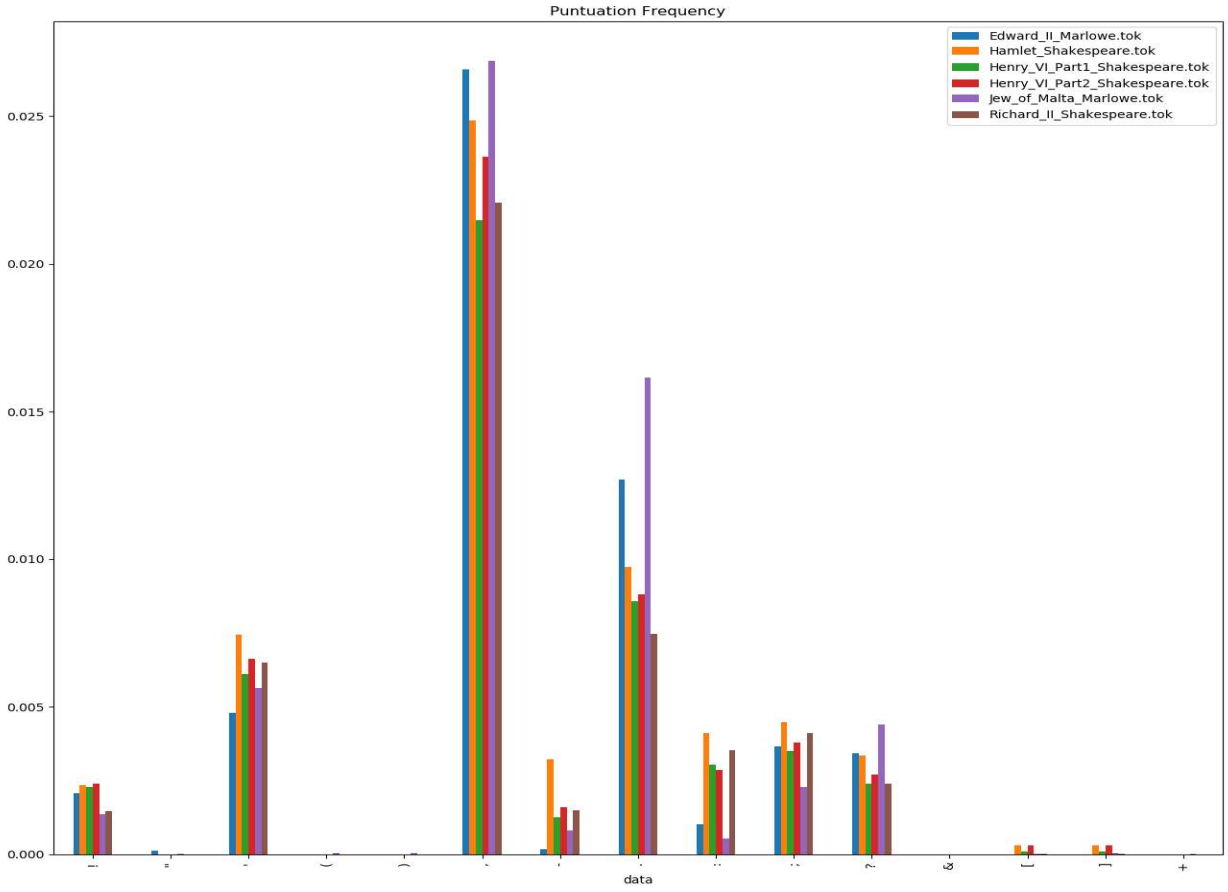


2. After that the list of files got print in the output section showing all the files name, as shown in the below image from the output panel of PyCharm.



Step 2

In this code, I have not received any inputs from the users. All the files must be present in the python file folder. Below are the screenshot for the graph visualization



3 Assumption

There are few assumptions that has to be take care by the user while providing the input. Below are the assumptions made while constructing the Morse code decoder in python:

1. User should not provide any other files.
2. Files names are fixed as given below
    - 'Edward_II_Marlowe.tok'
    - 'Hamlet_Shakespeare.tok'
    - 'Henry_VI_Part1_Shakespeare.tok'
    - 'Henry_VI_Part2_Shakespeare.tok'
    - 'Jew_of_Malta_Marlowe.tok'
    - 'Richard_II_Shakespeare.tok'
3. Words are the tokens encapsulated within space.
4. If the site is not reachable, then it will not print the stop word frequency graph, it will print the rest.