AC50002 Programming Languages for Data Engineering

Python Assignment Report

Julia Gratsova 130017011

Program description

The program's aim is to create a file with unique abbreviations corresponding to the given input file and initial letter values and run some tests to check the letter ranking logic.

A simple interface menu was created that prompts the user to enter a file name and then generates a .txt file with abbreviations using the name entered by a user. Second option runs through a number of tests that confirm that letter ranking is performed as asked by the assignment. Third option exits the program. If wrong input is chosen, the program prompts to use any key to return to menu.

The problem was solved via the following steps:
- The given sample text file of letter values is read by a program into a dictionary
- Getter function returns a score for a given letter
- The given sample text file of trees' names is processed according to the instructions (all letters capitalised, apostrophes ignored, any non-letter characters are also ignored and split the names into separate words, trailing spaces removed) and appended into a list
- Additional letter rank logic is applied including E letter exception
- New dictionary created with original name and rank for every letter according to the letter rank logic defined above and given letter values from values.txt

- Abbreviations created into a new dictionary preserving the letter indices ,where each letter gets a score from the dictionary with letter rankings. The total score for each abbreviation is calculated
- Duplicate check is performed by adding all abbreviations to the set (set only contains unique values)
- Prints new .txt file using the names from the original input file and corresponding best scored abbreviations.