Operating Systems Lab (CS 470):

Lab 2: Create a program to calculate the letter occurrences in a plain text file.

Overview

Calculating occurrences of each character in a text file can be useful to build statistics which can serve to break certain encoding (cryptanalysis). Frequency analysis is based on the fact that, in any given stretch of written language, certain letters and combinations of letters occur with varying frequencies. For instance, given a section of English language, E, T, A and O are the most common, while Z, Q and X are rare. Likewise, TH, ER, ON, and AN are the most common pairs of letters (called bigrams), and SS, EE, TT, and FF are the most common repeats.

Instructions

Write a program which reads its input from a plain text file (containing letters, punctuation marks, new lines and digits) given as input (read the filename from the standard input), and calculates a statistic based on the frequency of each letter. For each letter i, there should be a separate process P_i who is responsible to return the frequency for that particular letter.

Notes

- The text file will be given as input, no hardcoding allowed
- Each letter frequency should be returned by a separate process.
- In the text capital and lower case letter should be considered the same (ex. A = a).
- Punctuation marks such as [, ! ? - .] and digits are to be considered, but not to be mentioned in the stats.
- For each letter (letter pair ex. A-a) a separate process should be lunched considering fork()!

Rubric

Task	Points
Error handling	1
Calculating the letter frequency by separate processes	7
Print the statistics	2