

Homework 9

Download three books from Project Gutenberg in English and three in another language of your choice that uses the latin alphabet. Create statistics per language for the percentage of occurrence of

1. all letters,
2. all letters at the beginning of a word,
3. all letters at the end of a word.

Also find the 20 most frequent bi-grams in a word. (A bi-gram is two adjacent letters in a word.)