Лабораторная работа №1 «Энтропия текста»

Необходимо подсчитать энтропию текста, содержащегося в файле

Входные данные:

Файл с текстом (можете сами себе поставить ограничения на разрешение файла)

В данных файлах содержаться различные буквы (латиница и кириллица), знаки препинаний, цифры. Символы `@#$^&\*{}[]< > <= >= / \ | = + ` можете удалять. Не забывайте, что пробел тоже символ

Выходные данные:

1. Гистограмма появления всех символов, которые встретились в тексте, возможны три варианта построения гистограммы:

а. полная гистограмма всех символов на одной оси

б. 3-4 разных гистограмм, разбитых по смыслу (отдельно кириллица, латиница, символы)

в. полная таблица символ-частота (вероятность) а гистограмма – первые 10-20 часто встречаемых символов

2. Энтропия текста

Язык: любой, кроме Паскаля, Делфи, Бейсика и подобных

Интерфейс – нужен, конкретных требований к нему на первой лабе не предъявляю