Федеральное государственное бюджетное образовательное учреждение высшего образования   
«Саратовский государственный технический университет имени Ю. А. Гагарина»

Кафедра прикладных информационных технологий

**Практическая работа по курсу**

**Методы интеллектуального анализа естественного языка**

**на тему:**

**«Построение гистограммы частот вхождения отдельных букв в текстовом файле»**

Выполнил студент группы м-ПИНФ-21

Юфин Вадим Дмитриевич

Проверил: канд. физ.-мат. наук, д.т.н., доцент каф. ПИТ

Файфель Борис Леонидович

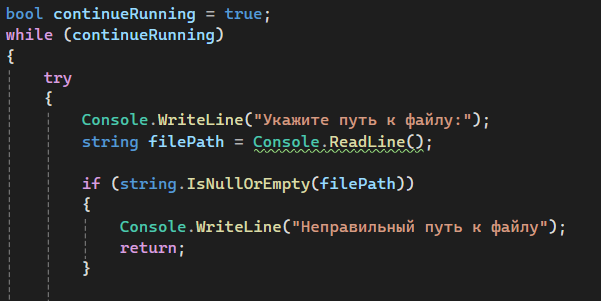
Саратов, 2024

**Задача:** разработать программу, которая на вход принимает текстовый файл, анализирует его содержимое и строит гистограмму частот встречаемости отдельных букв.

**Технология реализация:** для реализации решения был выбран язык программирования C#

**Описание кода:**

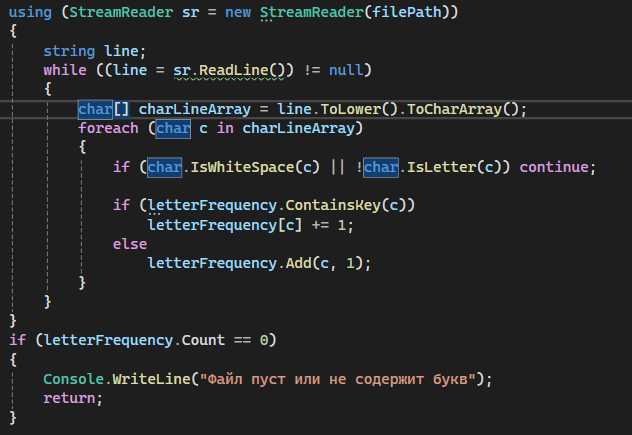
1. **Запрос пути к файлу и проверка на пустой ввод**

****

1. **Инициация словаря для подсчёта букв (ключ – буква, значение – количество встреченных букв)**

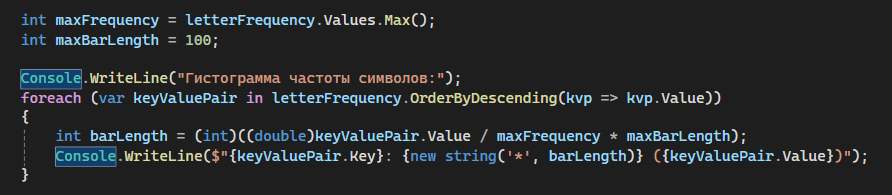
****

1. **Чтение файлов и подсчёт символов**



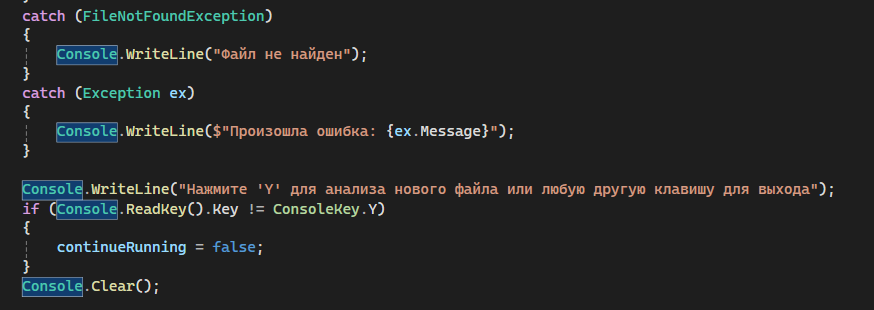
* Используется класс StreamReader для построчного чтения файла;
* Строка приводится в нижний регистр, а далее в массив символов;
* Цикл проходит по всему массиву и пропускает каждый символ, который не является буквой или является пробелом;
* Если в словаре не окажется ни одной пары, то выводится сообщение об отсутствии букв в файле.

1. **Построение гистограммы**



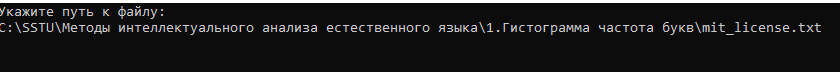
* Находится максимальная частота символа в тексте;
* Каждая частота масштабируется до длины строки 100 с использованием пропорции;
* Символ \* используется для визуального отображения частоты символа;
* Гистограмма отсортирована по убыванию значений (количеству вхождений).

1. **Обработка исключений и цикл для повторного анализа**

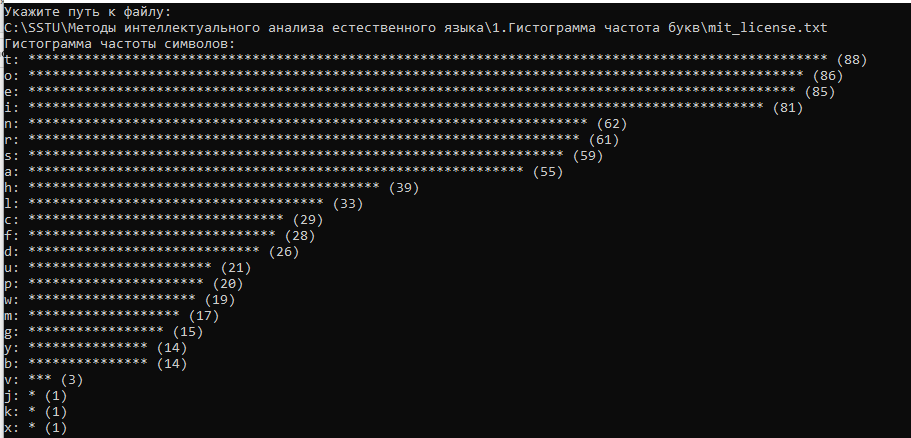


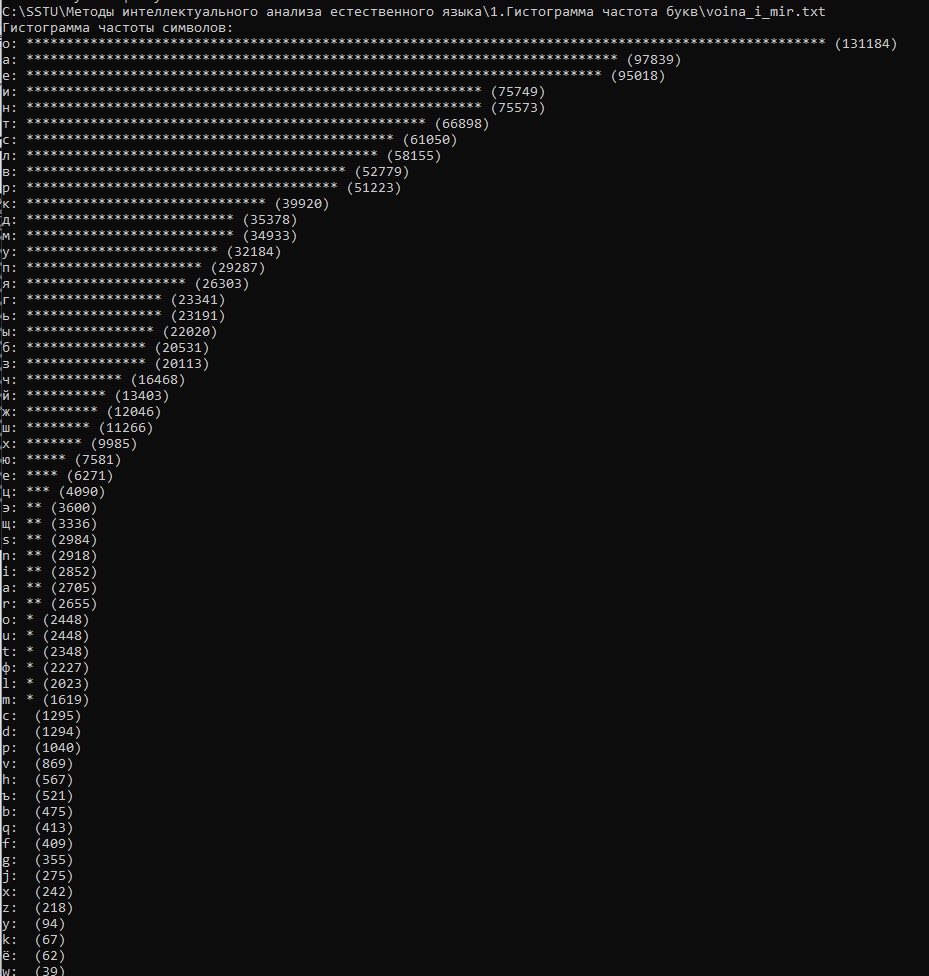
**Результат работы программы:**

1. **Пользователь вводит путь к файлу**

****

1. **Демонстрация гистограммы**

****

****