БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ

КАФЕДРА ИНФОРМАЦИОННЫХ СИСТЕМ И ТЕХНОЛОГИЙ

**Лабораторная работа №3**

**«Элементы теории информации.**

**Информативность данных в различных кодировках**»

**Выполнила:**

студентка 3 курса 1 группы

Потапейко Полина Павловна

**Проверил:**

Берников Владислав Олегович

Минск 2021

**Цель:** приобретение практических навыков трансформации данных и сопоставление энтропийных свойств используемых при этом алфавитов.

**Ход работы**

**Задание 1.** Создать приложение для конвертации документа на латинице в документ формата base64.

В качестве исходного текста взят произвольный электронный текстовый документ на английском языке (рисунок 1).

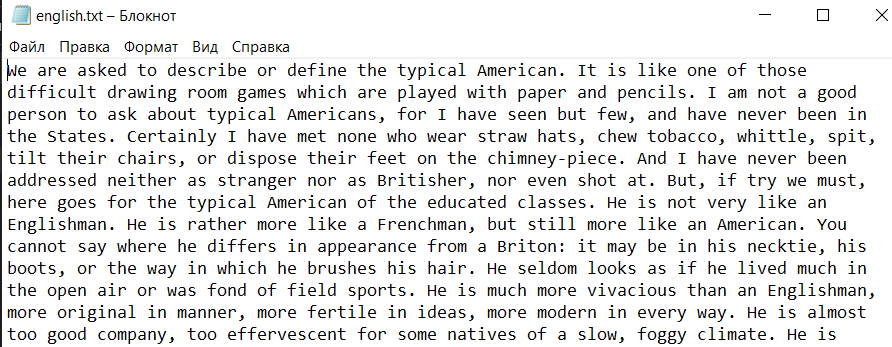


Рисунок 1 – Часть исходного текста

Для конвертации документа в формат base64 напишем приложение, используя специальные функции, поддерживаемые выбранным языком программирования. Код представлен на рисунке 2.

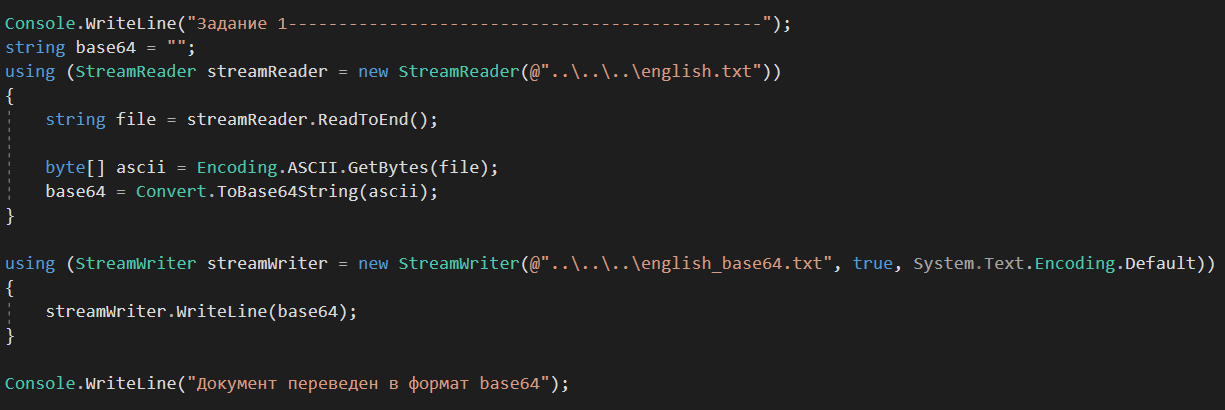


Рисунок 2 – Код приложения для конвертации

Часть конвертированного текста представлена на рисунке 3.



Рисунок 3 – Часть конвертированного текста

**Задание 2.** Получить распределение частотных свойств алфавитов по исходному и конвертированному документам. Вычислить энтропию Хартли и Шеннона, а также избыточность алфавитов.

Воспользуемся приложением, разработанным в ходе лабораторной работы 2 и найдем количество вхождений и вероятности появления каждой буквы английского алфавита в исходном документе.

Для расчета энтропии Шеннона использовалась формула:

где – вероятность появления символа алфавита.

Для расчета энтропии Хартли использовалась формула:

где N – число символов алфавита.

Для расчета избыточности алфавита использовалась формула:

На рисунке 4 представлены полученные значения.

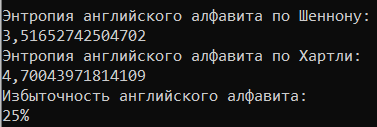


Рисунок 4 – Рассчитанные значения энтропий Шеннона и Хартли, а также избыточность алфавита для исходного текста

Аналогично рассчитываем частоту и вероятности появления символов алфавита base64, а также энтропии Шеннона и Хартли и избыточность алфавита (рисунок 5).

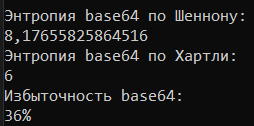


Рисунок 5 – Рассчитанные значения энтропий Шеннона и Хартли, а также избыточность алфавита для исходного текста

**Задание 3.** Написать функцию, которая вычисляет XOR двух буферов (имя и фамилия).

Функция принимает два параметра: буфер с именем и буфер с фамилией. Происходит перевод буферов в код ASCII и код base64. При неравном количестве битов буферов меньший буфер дополняется нулями до длины большего буфера. Затем выполняется операция исключающего ИЛИ этих буферов. Результат работы функции представлен на рисунке 6.

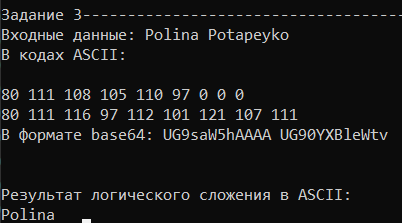


Рисунок 6 – Количество информации в сообщении (в исходном коде)

**Вывод:** для характеристики источника сообщения с различным алфавитом используется понятие избыточности источника сообщений или избыточности алфавита. Для кодирования бинарных данных и решения проблемы обнуления старших битов в сообщении используется алгоритм base64.