МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ

Учреждение образования

«БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ»

Факультет информационных технологий

Кафедра информационных систем и технологий

Отчет по лабораторной работе №9

«Сжатие/распаковка данных на основе статистических методов»

по дисциплине

«Защита информации и надежность информационных систем»

Студент: Шедько Е.А.

3 курс 3 группа ФИТ

Преподаватель: Берников В. О.

Минск 2021 г.

# Цель

Приобретение практических навыков использования статистических методов Шеннона − Фано и Хаффмана (Shannon-Fano and Huffman coding) для сжатия/распаковки данных.

# Задачи

1. Закрепить теоретические знания по алгебраическому описанию и использованию методов сжатия/распаковки (архивации/разархивации) данных на основе методов Шеннона − Фано и Хаффмана.

2. Разработать приложение для реализации методов Шеннона − Фано и Хаффмана.

3. Результаты выполнения лабораторной работы оформить в виде описания разработанного приложения, методики выполнения экспериментов с использованием приложения и результатов эксперимента.

# Теоретическая часть

Мы неоднократно подчеркивали, что каждый из естественных языков обладает избыточностью. Среди европейских языков белорусский и русский обладают одним из самых высоких уровней избыточности. Об этом можно судить по размерам русского перевода английского текста. Обычно он примерно на 20–30% больше.

До появления уже упоминавшихся работ К. Шеннона кодирование символов алфавита при передаче сообщения по каналам связи осуществлялось одинаковым количеством битов, получаемым по формуле Хартли (см. формулу (2.2)). Позднее начали появляться способы, кодирующие символы разным числом битов в зависимости от вероятности появления их в тексте, подтверждение чему мы получили при выполнении лабораторной работы № 2. Таким образом, за счет использования для каждого значения байта кодов ASCII (символа алфавита) кода различной длины в соответствии с частостью (вероятностью появления этого символа в сообщении) можно значительно уменьшить общий размер данных. Эта базовая идея лежит в основе алгоритмов статистических (вероятностных) методов сжатия: Шеннона − Фано и Хаффмана.

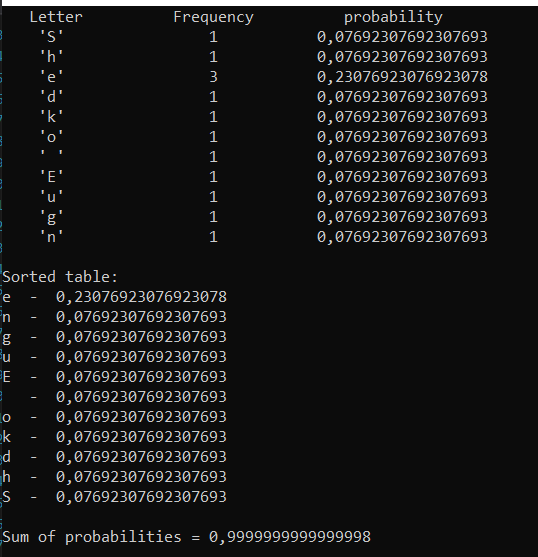
Статистические алгоритмы позволяют создавать более короткие коды для часто встречающихся и более длинные – для редко встречающихся символов алфавита или конкретного сообщения. В первом случае метод считается статическим статистическим, во втором – динамическим статистическим: вероятностные свойства символов подсчитываются для конкретного сообщения или потока данных.

Частота или вероятность появления того или иного символа алфавита в произвольном сообщении, лежащая в основе алгоритмов, дали название этим алгоритмам и соответствующим методам.

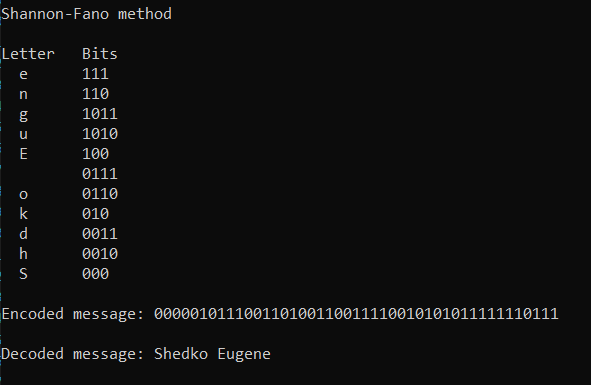
Иногда эти методы называют также префиксными.

# Практическая часть

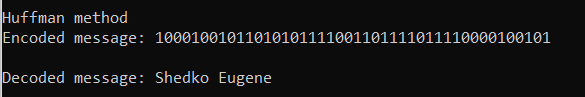
Для реализации методов сжатия Шеннона-Фано и Хаффмана были реализованы соответствующие функции, а также дополнительные функции, необходимые для их работы. Также перед началом работы выводятся вычисленные вероятности, а после – выводится длина кода для каждого метода сжатия, а также для ASCII. Результат выполнения программы представлен на рисунке 1-3.



*Рисунок 1 – Расчет вероятностей*



*Рисунок 2 – Метод Шеннона-Фано*



*Рисунок 3 – Метод Хаффмана*

# Вывод

В данной лабораторной работе мною было изучено сжатие и распаковка данных на основе статистических методов, а именно методов Шеннона-Фано и Хаффмана, а также реализована программа, которая выполняет сжатие и распаковку данных при помощи этих методов и выводит эффективность применения данных методов.