МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ

Учреждение образования «БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ

ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ»

Факультет информационных технологий

Кафедра информационных систем и технологий

Специальность Информационные системы и технологии

**ОТЧЕТ ПО ЛАБОРАТОРНОЙ РАБОТЕ №9 НА ТЕМУ:**

Сжатие/распаковка данных на основе статистических методов

Выполнил студент 3 курса 1 группы

Кашперко Василиса Сергеевна

Минск 2022

**Цель:** приобретение практических навыков использования статистических методов Шеннона−Фано и Хаффмана 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.

**Теоретические сведения**

*Метод Шеннона–Фано:*не является оптимальным (обеспечивает минимальную избыточность) в общем смысле, хотя и дает оптимальные результаты при некоторых распределениях вероятностей. Для одного и того же распределения вероятностей можно построить, вообще говоря, несколько кодов Шеннона – Фано, и все они могут дать различные результаты.

*Метод Хаффмана:*основан на алгоритме оптимального префиксного кодирования алфавита: исходный алгоритм Хаффмана является оптимальным для посимвольного кодирования с известным входным распределением вероятностей, т. е. для отдельного кодирования несвязанных символов в таком потоке данных.Отличается от метода Шеннона–Фано лишь в части кодирования символовисходного алфавита.

В данном случае бинарные коды создаются на основе дерева, ветви которого обозначаются бинарными символами.

**Практическая часть**

Необходимо выполнить прямое и обратное преобразование сообщения, состоящего из собственных фамилии и имени.

Прежде всего необходимо сформировать алфавит. Алфавит сформируем из своих имени и фамилии.

Кашперко Василиса

Включая пробел – 17 символов. Подсчитываем для каждого символа количество их появлений и определяем вероятность их встречи разделяя число появлений на общее число символов (рис. 1). Затем сортируем в порядке убывания вероятности (рис. 2).

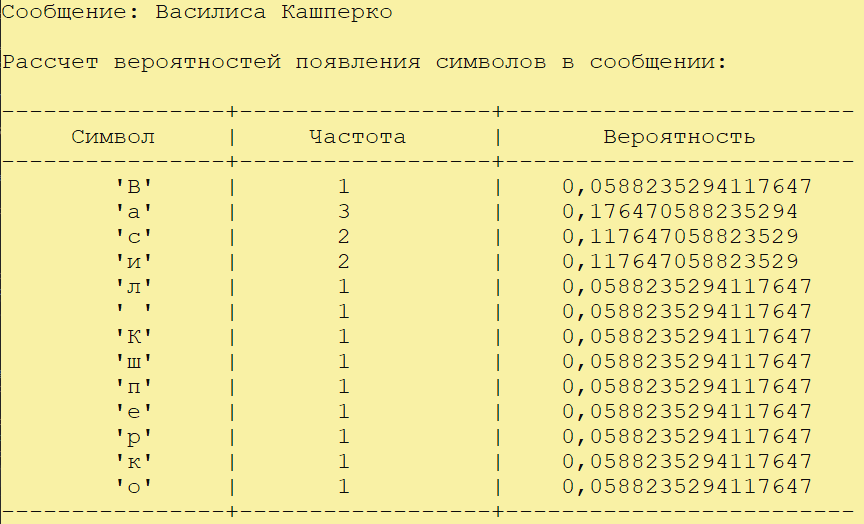


Рисунок 1 – Подсчет повторов и вероятностей символов

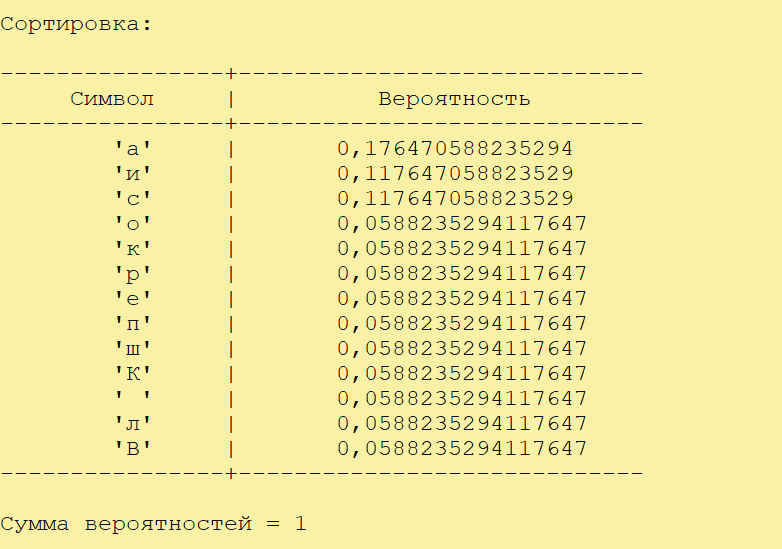


Рисунок 2 - Сортировка

***Метод Шеннона – Фано***

Эту отсортированную таблицу разделим на две части (т.е. два подмножества). Сумма вероятностей в обоих подмножествах должна быть приблизительно равна, и верхняя часть таблицы определяется старшим символом 1, а нижней части – 0. И так продолжаем делить, до тех пор, пока в каждом подмножестве не останется лишь 1 символ. Получим таблицу бинарных кодов (рис. 3).

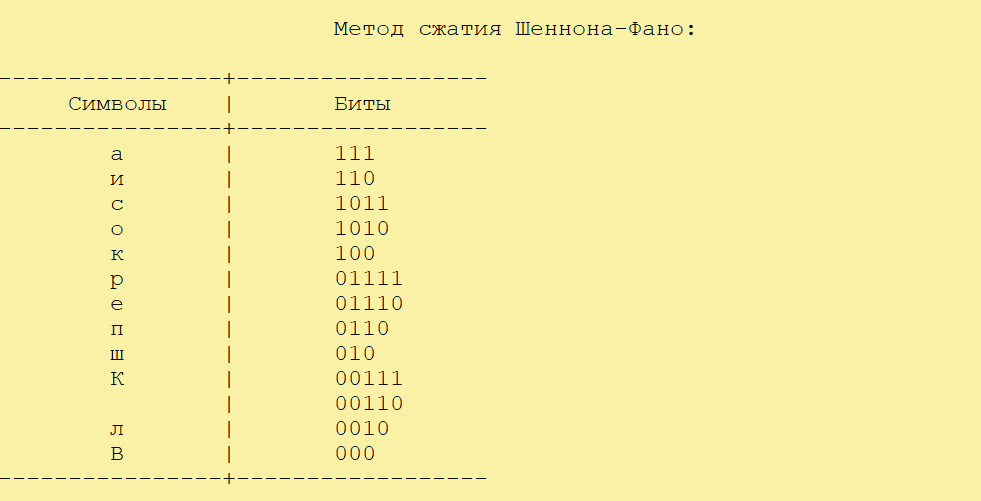


Рисунок 3 – Таблица бинарных кодов для символов

*Алгоритм прямого преобразования (кодирование):* необходимо выполнить одну операцию: заменить символы входного сообщения полученными бинарными кодами. *Алгоритм обратного преобразования (декодирование):* подставляем соответствующий символа вместо бинарного кода исходного алфавита из таблицы бинарных кодов для символов. Закодированное и декодированное сообщение представлено на рисунке 4.

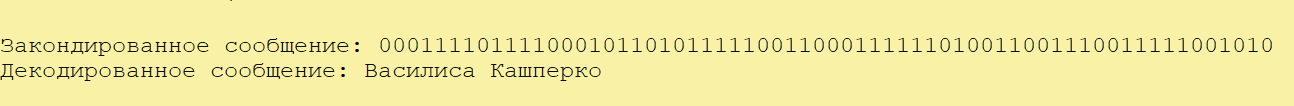


Рисунок 4 – Результаты прямого и обратного преобразований методом

Шеннона-Фано

Также была рассчитана эффективность сжатия сообщения в сравнении с кодами ASCII. Результаты представлены на рисунке 5.

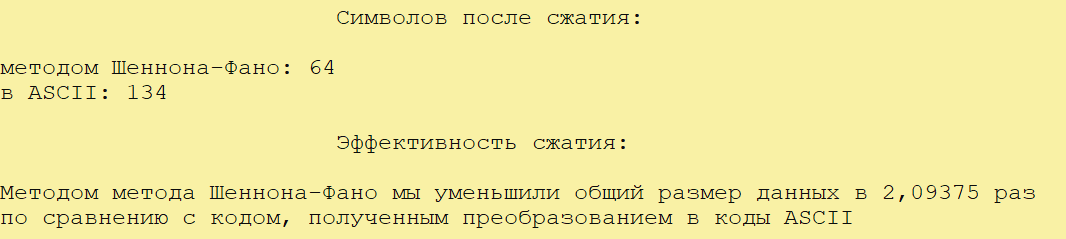


Рисунок 5 – Эффективность сжатия

**Вывод:** в ходе лабораторной работы были изучены методы Шеннона–Фано и Хаффмана. Реализован программно метод Шеннона–Фано. Эффективность сжатия сообщения в сравнении с кодами ASCII можно оценить подсчетом символов в закодированном сообщении методом Шеннона – Фано при помощи простого деления, чтобы узнать, в какое количество раз уменьшилось суммарное количество байт. Было выяснено: можно значительно уменьшить общий размер передаваемых данных посредством сжатия данных методом Шеннона-Фано.