БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ

КАФЕДРА ИНФОРМАЦИОННЫХ СИСТЕМ И ТЕХНОЛОГИЙ

**Лабораторная работа №9**

**«Сжатие/распаковка данных**

**на основе статистических методов**»

**Выполнила:**

студентка 3 курса 1 группы

Потапейко Полина Павловна

**Проверил:**

Берников Владислав Олегович

Минск 2021

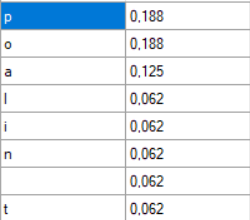
**Цель:** приобретение практических навыков использования статистических методов Шеннона – Фано и Хаффмана для сжатия/распаковки данных.

**Ход работы**

**Задание.** Выполнить прямое и обратное преобразование сообщения, состоящего из собственного имени и фамилии.

Методом кодировки выбран метод Шеннона – Фано.

На первом шаге прямого преобразования подсчитываются вероятности появления символов сообщения, затем сортируются в порядке невозрастания (рис. 1).



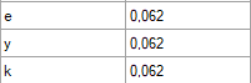
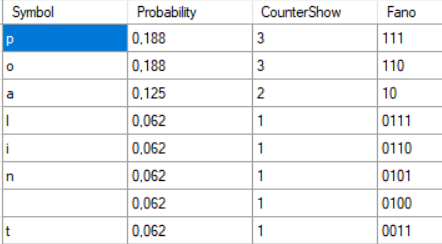


Рисунок 1 – Вероятности

Затем множество символов делится на две группы так, чтобы сумма вероятностей в группах была примерно одинаковой, первому из полученных подмножеств ставится в соответствие бинарная единица, второму бинарный 0. Эта процедура повторяется рекурсивно для каждого из полученных подмножеств, в котором содержится больше одного символа (рис. 2).



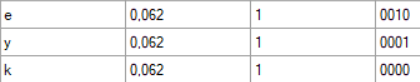


Рисунок 2 – Вычисление кодов

Для обратного преобразования анализируется lmin начальных бинарных символов и осуществляется поиск соответствующего совпадения в таблице, если такое совпадение найдено, то на выходе будет символ исходного алфавита, если нет, то длина анализируемой последовательности увеличивается на 1, снова осуществляется поиск совпадающей бинарной комбинации такой же длины в таблице, и так далее, пока не будет найдена совпадающая последовательность.

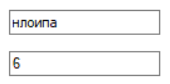


Рисунок 3 – Выходное сообщение

Эффективность сжатия сообщения составляет 57,81% по сравнению с кодом ASCII.

**Вывод:** статистические алгоритмы сжатия данных, примерами которых являются алгоритм Шеннона – Фано и алгоритм Хаффмана, позволяют создавать более короткие коды для часто встречающихся символов алфавита или конкретного сообщения. Эти методы называются также префиксными. Методы относятся к классу «сжатие без потерь». Различие между ними состоит только в особенностях формирования таблицы бинарных кодов.