БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ

КАФЕДРА ИНФОРМАЦИОННЫХ СИСТЕМ И ТЕХНОЛОГИЙ

**Лабораторная работа №11**

**«Сжатие/распаковка данных**

**арифметическим методом**»

**Выполнила:**

студентка 3 курса 1 группы

Потапейко Полина Павловна

**Проверил:**

Берников Владислав Олегович

Минск 2021

**Цель:** приобретение практических навыков использования арифметических методов сжатия/распаковки данных.

**Ход работы**

**Задание.** Выполнить прямое и обратное преобразование сообщений («достопримечательность» и «достопримечательностьсорокадневный»).

Высчитываются вероятности появления каждого символа в сообщении (рис. 1).

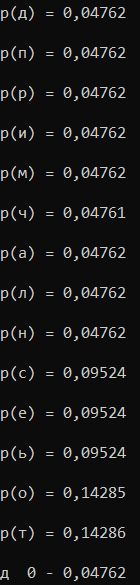


Рисунок 1 – Вероятности появления символов в сообщении

Затем формируются интервалы появлений символов (рис. 2).

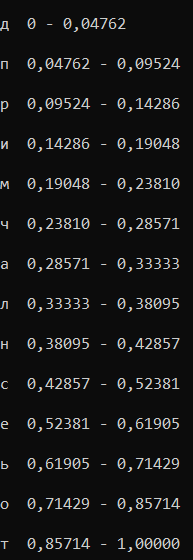


Рисунок 2 – Шаги алгоритма

На основании сформированных интервалов начинается посимвольное сжатие сообщения: берутся верхняя и нижняя границы интервала, в который попадает сжимаемый символ, новый интервал разбивается пропорционально для каждого символа, он становится новым рабочим отрезком; затем снова берутся нижняя и верхняя граница участка для нового символа сообщения, этот участок снова разбивается в соответствии с исходными соотношениями символов и т.д. пока не будет обработан последний символ сообщения (рис. 3).

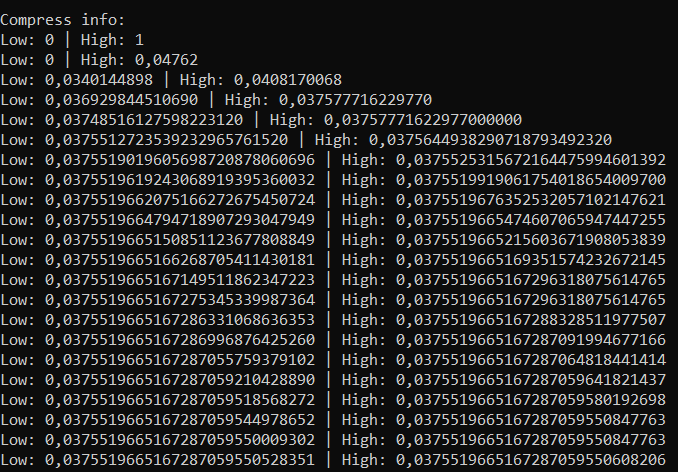


Рисунок 3 – Сжатие

Результатом будет являться значение нижней границы участка последнего символа (рис. 4).



Рисунок 4 – Результат сжатия

Для обратного преобразования (декомпрессии) необходима информация о значении числа, являющегося итогом сжатия сообщения, количестве символов в сжатом сообщении, вероятностных параметрах всех символов исходного сообщения. Как и при сжатии начальный рабочий отрезок разбивается на интервалы, длины которых равны вероятностям появления соответствующих символов. На каждом шаге обратного преобразования выбирается отрезок, в который попадает текущее число. Символ, который соответствует данному отрезку, является очередным символом восстановленного сообщения (рис. 5).

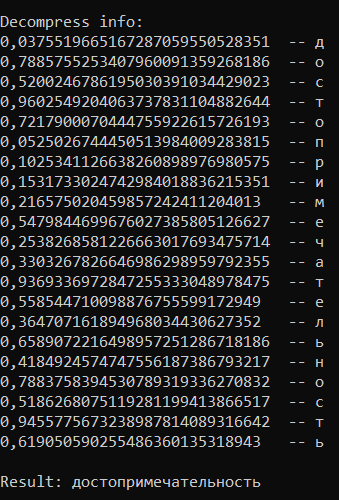


Рисунок 5 – Результат декомпрессии

При сжатии сообщения с большим количеством символов будет постепенно увеличиваться количество знаков после запятой, что в конце концов может сказаться на результате декомпрессии (не хватает вычислительной мощности) (рис. 6).



Рисунок 6 – Результат декомпрессии

**Вывод:** при арифметическом сжатии текст представляется вещественными числами в интервале от 0 до 1. По мере анализа текста отображающий его интервал уменьшается, а количество битов для его представления возрастает. Очередные символы текста сокращают величину интервала, исходя из значений соответствующих вероятностей.