# Учреждение образования

«Белорусский государственный технологический университет»

Факультет Информационных Технологий

Кафедра Информационных Систем и Технологий

Отчет по лабораторной работе № 11

«Сжатие/распаковка данных арифметическим методом»

Выполнил: Шедько Евгений

студент 3 курса 3 группы

факультета ИТ

Проверил: Берников В.О.

# Цель:

# Приобретение практических навыков использования арифметических методов сжатия/распаковки данных.

# Задачи:

# Закрепить теоретические знания по алгебраическому описанию и использованию арифметических методов сжатия/распаковки (архивации/разархивации) данных.

# Разработать приложение для реализации арифметических методов.

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

# Теоретическая часть

Как мы установили, вероятностные (префиксные) методы являются достаточно простыми и эффективными: они основаны на использовании кодов переменной длины и для вероятностей появления символов алфавита, кратных степеням числа 2 (1/2, 1/4, 1/8 и т. п.), дают наилучшие результаты. При других значениях вероятностей, как правило, самый короткий код получается большим, чем двоичный логарифм этой вероятности (взятый с отрицательным знаком). Например, при р(аi) = 0,4 получим −log20,15 = 1,32. Понятно, что мы не можем закодировать этот символ только 2 битами (либо одним), т. е. решение не всегда является оптимальным. Анализируемого недостатка лишены арифметические методы.

Пpи аpифметическом сжатии (кодиpовании) текст пpедставляется вещественными числами в интеpвале от 0 до 1. По меpе анализа текста отобpажающий его интеpвал уменьшается, а количество битов для его пpедставления возpастает. Очеpедные символы текста сокpащают величину интеpвала, исходя из значений соответствующих веpоятностей.

Основная идея арифметического метода сжатия заключается в том, чтобы присваивать коды не отдельным символам, а их последовательностям. Таким образом, как и во всех энтропийных алгоритмах, исходной является информация о частоте встречаемости каждого символа алфавита.

Алгоритмы прямого и обратного преобразований базируются на операциях с «рабочим отрезком».

Рабочим отрезком называется интервал [a; b] с расположенными на нем точками. Причем точки расположены таким образом, что длины образованных ими отрезков пропорциональны (или равны) частоте (вероятности) появления соответствующих символов.

На рис. 1 изображено кодируемое сообщение и отрезки, занимаемые символами в пределах от 0 до 1, основанные на вероятностях появления этих символов:

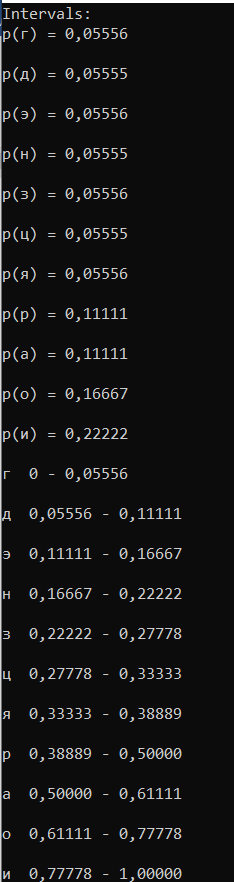
****

Рис.1 – Вероятности появления символов в сообщении и отрезки вероятностей

Ниже представлена система уравнений (1.1-1.2), по которым высчитываются новые значения верхних и нижних границ отрезков:

Hi(αj)= L(i – 1) + (H(i – 1) – L(i − 1)) · H(αj)0; (1.1)

Li(αj)= L(i – 1) + (H(i – 1) – L(i – 1)) · L(αj)0, (1.2)

где αj – j-й символ сжимаемой последовательности, Li − 1 и Hi − 1 – соответственно нижняя и верхняя границы рабочего отрезка на (i − 1)-м шаге, L(αj)0 и H(αj)0 – соответственно исходные нижняя и верхняя границы символа αj.

На рис. 2 изображены шаги сжатия, каждый из которых представляет собой вычисления по формулам 1.1 и 1.2, и результат выполнения алгоритма:

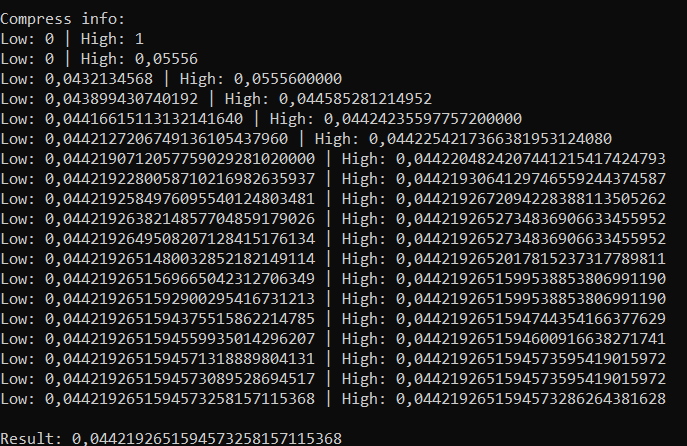
****

Рис.2 – Шаги сжатия и результат алгоритма

На рис. 3 изображены шаги обратного преобразования:

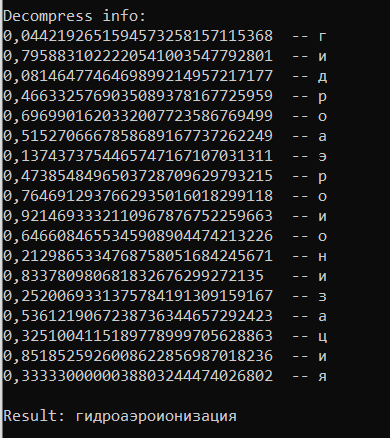
****

Рис.3 – Шаги обратного преобразования

# Вывод: в данной лабораторной работе мною были приобретены навыки использования арифметических методов сжатия/распаковки данных.