БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ

КАФЕДРА ИНФОРМАЦИОННЫХ СИСТЕМ И ТЕХНОЛОГИЙ

**Лабораторная работа №10**

**«Сжатие/распаковка данных**

**Методом Лемпеля – Зива**»

**Выполнила:**

студентка 3 курса 1 группы

Потапейко Полина Павловна

**Проверил:**

Берников Владислав Олегович

Минск 2021

**Цель:** приобретение практических навыков использования метода Лемпеля – Зива для сжатия/распаковки данных.

**Ход работы**

**Задание.** Выполнить прямое и обратное преобразование произвольного текста длиной несколько килобайт.

Пользователь выбирает файл для преобразования, затем подсчитывается длина этого файла (рис. 1).



Рисунок 1 – Высчитывание длины исходного файла

Затем производится непосредственно сжатие с помощью подстрок, символы считываются до тех пор, пока накапливаемая подстрока входит целиком в одну из фраз словаря. Как только эта подстрока перестанет соответствовать хотя бы одной фразе словаря, генерируется, состоящий из индекса строки в словаре, которая до последнего введенного символа содержала входную строку, и символа, нарушившего совпадение. Затем в словарь добавится введенная подстрока. Если словарь уже заполнен, то из него предварительно удаляют менее всех используемых в сравнении фразу.

Таблица цепочки шагов сжатия представлена на рисунке 2.

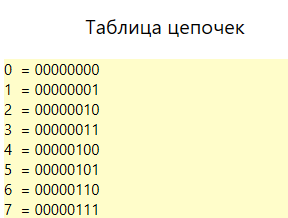


Рисунок 2 – Шаги алгоритма

Также длина сжатого файла представлена на рисунке 3 (эффективность сжатия составляет почти 2,5 раза).



Рисунок 3 – Выходное сообщение

**Вывод:** сжатие данных словарными методами осуществляется за счет замены записей соответствующими кодами из словаря. Классический алгоритм Лемпеля – Зива формулируется так: если в проанализированном (сжатом) ранее выходном потоке уже встречалась подобная последовательность байт, причем запись о ее длине и смещении от текущей позиции короче, чем сама эта последовательность, то в выходной файл записывается ссылка (смещение, длина), а не сама последовательность. Суть метода в том, что упаковщик постоянно хранит некоторое количество последних обработанных символов в буфере. По мере обработки входного потока вновь поступившие символы попадают в конец буфера, сдвигая предшествующие символы и вытесняя самые старые. Размеры этого буфера (скользящего словаря) варьируются.