Учреждение образования

«БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ»

Кафедра информационных систем и технологий

Отчет

по дисциплине «Защита информации и надежность информационных систем»

Студент: Шумова Е.И.

ФИТ 3 курс, 1 группа, 1 подгруппа

Преподаватель: Савельева М.Г.

Минск 2022

**Лабораторная работа №8**

**Тема «Сжатие/распаковка данных методом Барроуза-Уилера»**

**Цель:** приобретение практических навыков использования метода Барроуза − Уилера для сжатия/распаковки данных.

**Задачи:**

* Закрепить теоретические знания по алгебраическому описанию и использованию методов сжатия/распаковки (архивации/ разархивации) данных на основе метода Барроуза − Уилера (BurrowsWheeler transform, BWT).
* Разработать приложение для реализации метода Барроуза − Уилера.
* Результаты выполнения лабораторной работы оформить в виде описания разработанного приложения, методики выполнения экспериментов с использованием приложения и результатов эксперимента.

**Теоретические сведения**

Основная цель сжатия – обеспечить более компактное представление данных, вырабатываемых источником, т. е. уменьшить физический объем сообщений, генерируемых источником, и сократить время его передачи (читай – стоимость) по каналам связи. Фундаментальная теорема К. Шеннона о кодировании информации утверждает, что «стоимость кодирования всегда не меньше энтропии источника, хотя может быть сколь угодно близка к ней». Поэтому для любого алгоритма сжатия всегда имеется некоторый предел степени (или эффективности) сжатия, определяемый энтропией входного потока (или сжимаемого сообщения).

Все алгоритмы сжатия преобразуют входной поток данных, минимальной единицей которых является бит, а максимальной – байт или несколько байт. Основными техническими характеристиками процессов сжатия и результатов их работы являются:

* *степень сжатия* (англ. compress rating), или отношение *R* (англ. ratio) объемов исходного (до сжатия, *Vдс*) и результирующего (после сжатия, *Vпс*) потоков данных (сообщений);
* *скорость сжатия* − время, затрачиваемое на сжатие некоторого объема информации входного потока до получения из него эквивалентного выходного потока;
* *качество сжатия* − величина, показывающая, насколько сильно сжат выходной поток при помощи применения к нему повторного сжатия по этому же или иному алгоритму.

Существуют различные подходы к реализации сжатия информации. Они отличаются математической базой, уровнем сложности (простоты) практической реализации, форматом кодируемого потока данных, степенью соответствия сжимаемых и распакованных данных.

По критерию, связанному с характером или форматом данных или степенью соответствия сжимаемых данных распакованным, все методы сжатия разделяют на два класса: обратимое и необратимое сжатие, или иначе: сжатие без потерь и сжатие с частичной потерей информации (англ. lossy compression).

BWT-преобразование (англ. Burrows-Wheeler Transform) – техника сжатия информации (в особенности текстов), основанная на преобразовании, открытом в 1983 г. BWT не сжимает данные в классическом понимании процесса, но преобразует блок данных в формат, исключительно подходящий для сжатия.

BWT оперирует сразу целым блоком данных, который выделяется из входного потока (сообщения). Прямое преобразование (формально – сжатие) выполняется в 4 этапа:

* выделяется блок данных (строка длиной *k* символов некоторого алфавита мощностью *N*), который обозначим символом *М*;
* составляется таблица *W1* размером *k×k* всех циклических сдвигов входной строки *M*: *W1* = (*M*);
* производится лексикографическая (в алфавитном порядке) сортировка строк таблицы *W1*, в результате чего получается таблица *W2* того же размера;
* в качестве выходной строки (обозначим ее BWT(*М*), *z*) выбирается последний столбец (*Мk*) таблицы *W2* преобразования и номер строки *z*, совпадающей с исходной строкой *М*. Как видим, выходная строка (сжатое сообщение) всегда по объему превышает входную.

Итак, входной для обратного преобразования является информация вида BWT(*М*), *i*. Это преобразование заключается в выполнении *k* одинаковых шагов, каждый из которых состоит из 2 операций, с целью воссоздания матрицы *W2*:

* в крайний справа пустой столбец матрицы записывается последовательность символов *Мk*;
* производится лексикографическая сортировка столбцов заполненной части воссоздаваемой матрицы.

**Практическая часть**

**Вариант 14**

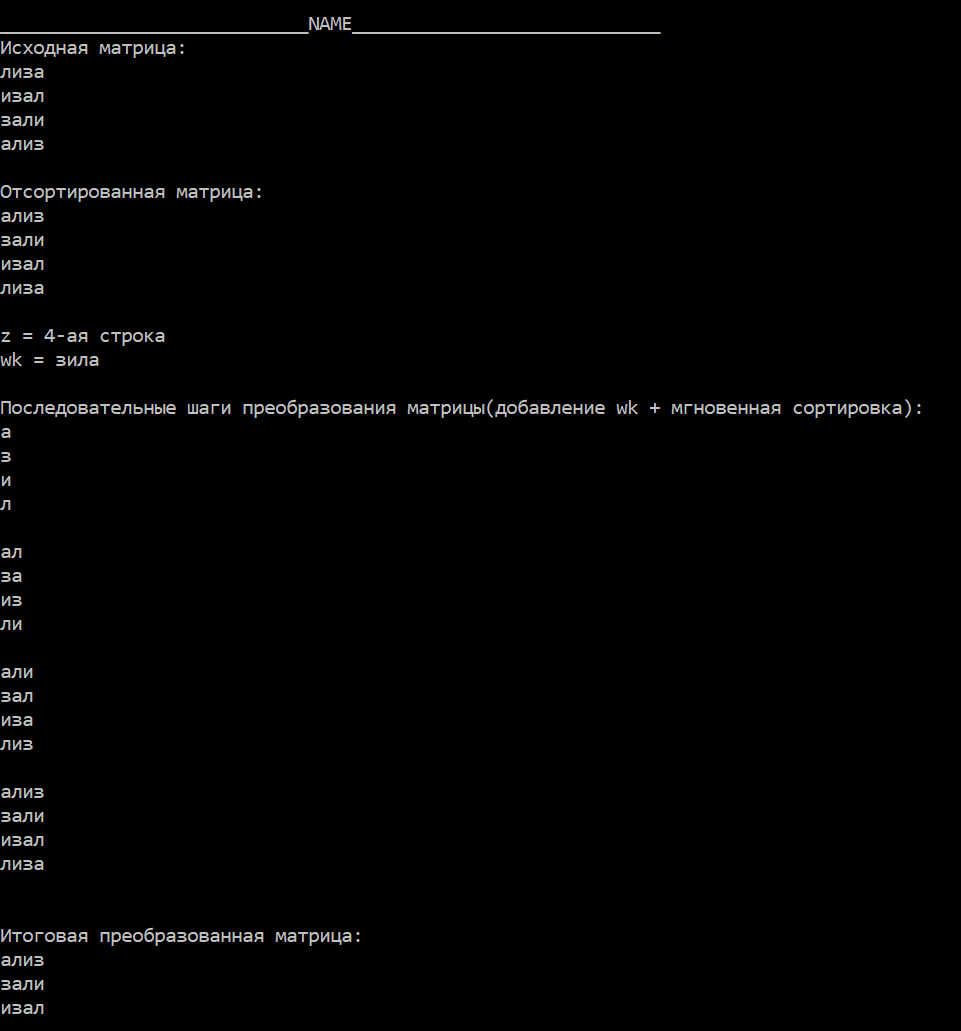


Рисунок 1 – преобразования имени

Для фамилии, строки по варианту и первым трем символам строки по варианту в кодах ASCII алгоритм процесса аналогичный. Результаты вышеперечисленных строк более объемные, т.к. в них больше символов.

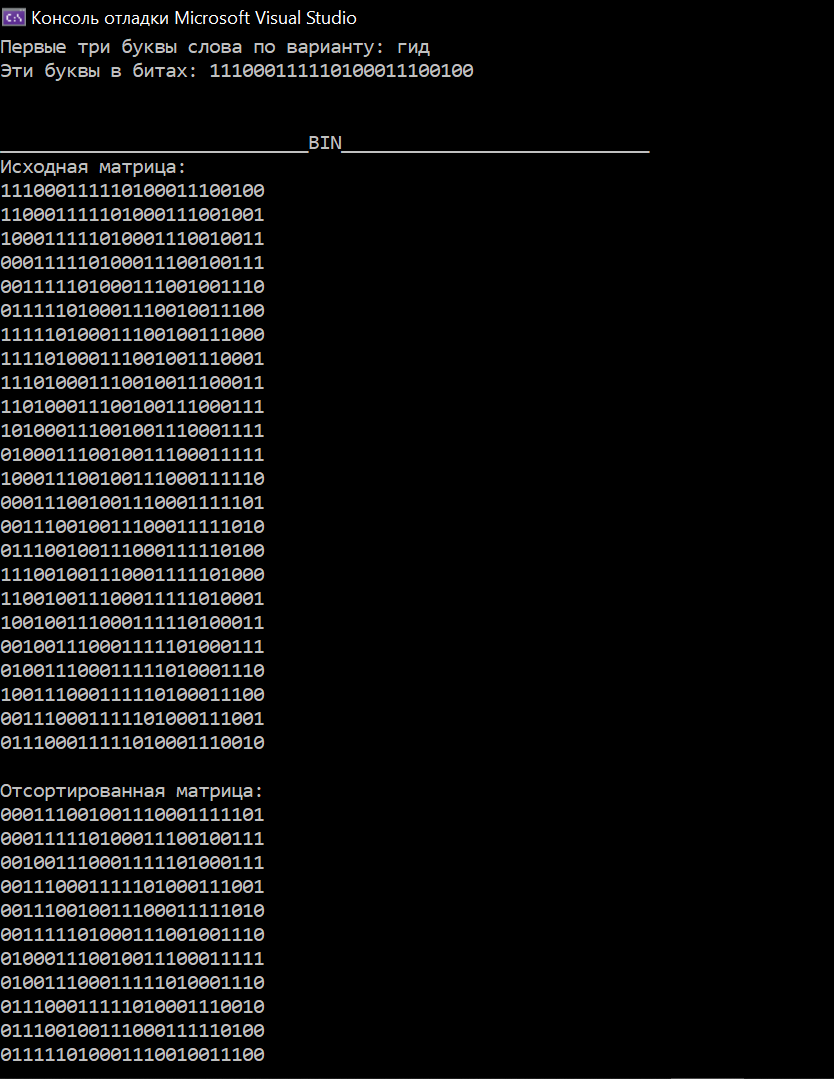


Рисунок 2 – преобразование бинарного представления первых трех букв слова по варианту

Т.к. количество шагов обратного преобразования велико, большая их часть будет пропущена.

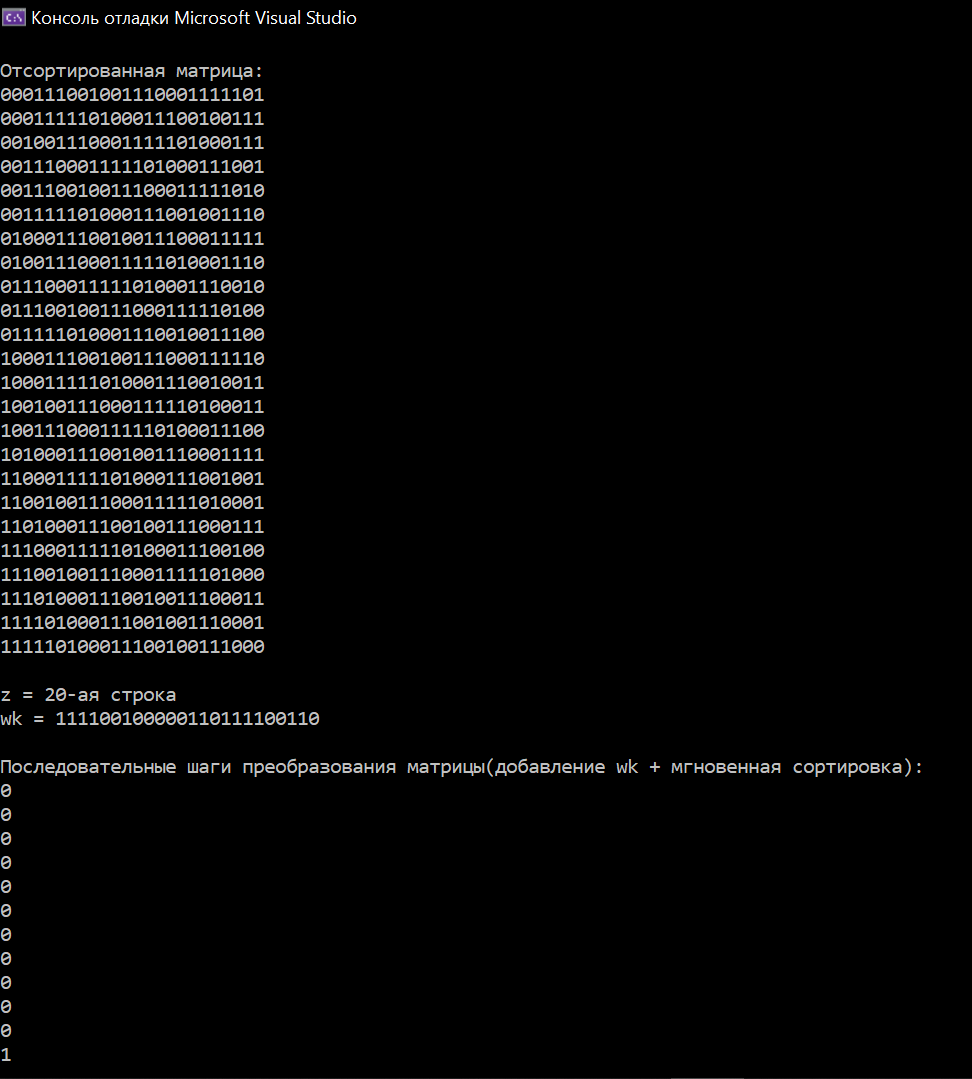


Рисунок 3 – преобразование бинарного представления первых трех букв слова по варианту

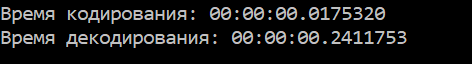


Рисунок 4 – сравнение времени кодирования и декодирования для сообщения длиной 24 символа

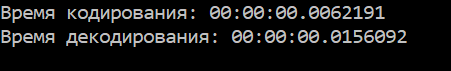


Рисунок 5 – сравнение времени кодирования и декодирования для сообщения длиной 4 символа

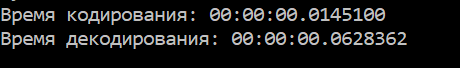


Рисунок 6 – сравнение времени кодирования и декодирования для сообщения длиной 6 символа

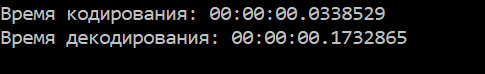


Рисунок 7 – сравнение времени кодирования и декодирования для сообщения длиной 18 символа

По рисункам 4-7 можно сделать вывод, что для декодирования сообщения методом Барроуза-Уилера уходит больше времени, чем на его кодирование. Чем больше символов в сообщении, тем дольше будет происходить процесс кодирования-декодирования.

**Вывод:** в ходе лабораторной работы были приобретены практические навыки использования метода Барроуза − Уилера для сжатия/распаковки данных.