Учреждение образования

«БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ»

Защита информации и надёжность информационных систем

Студент: Скалкович С.Л.

ФИТ 3 курс 2 группа

Преподаватель: Нистюк О.А.

**Лабораторная работа №7**

**СЖАТИЕ/РАСПАКОВКА ДАННЫХ МЕТОДОМ БАРРОУЗА − УИЛЕРА**

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

Задачи:

1. Закрепить теоретические знания по алгебраическому описанию и использованию методов сжатия/распаковки (архивации/ разархивации) данных на основе метода Барроуза − Уилера (BurrowsWheeler transform, BWT).

2. Разработать приложение для реализации метода Барроуза − Уилера.

1. Теоретические сведения

Сжатие информации является одним из способов ее кодирования. Развитие методов сжатия данных имеют длинную историю, которая началась задолго до появления компьютеров и компьютерных сетей.

В основе сжатия данных, как одна из первопричин, лежит избыточность, что являлось предметом исследования в лабораторной работе № 3.

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.**

**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:**

**1) выделяется блок данных (строка длиной k символов некоторого алфавита мощностью N), который обозначим символом М;**

**2) составляется таблица W1 размером k×k всех циклических сдвигов входной строки M: W1 = (M);**

**3) производится лексикографическая (в алфавитном порядке) сортировка строк таблицы W1, в результате чего получается таблица W2 того же размера;**

**4) 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,**

1. Практическая часть

**Задание 1.**

**В соответствии с вариантом №12 необходимо разработать авторское приложение в соответствии с целью лабораторной работы. Входной блок данных может иметь произвольную длину.**

****Задание 2.****

**С помощью приложения выполнить прямое и обратное преобразования 3 отдельных блоков данных, состоящих: а) из собственного имени (можно краткий вариант записи); б) собственной фамилии; в) варианта в соответствии с таблицей. Можно использовать любой из известных методов сортировки символов массива. Выполнить качественный сравнительный анализ длительности процессов прямого и обратного преобразований в зависимости от длины блока данных.**

**Задание 3.**

**Перевести первые 3 символа из блока данных, указанного в варианте таблицы, в бинарную последовательность в соответствии с кодами ASCII. Выполнить прямое и обратное преобразование. Оценить время прямого и обратного преобразований.**

**Оценка времени отображена на рисунке 1.**

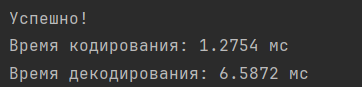
****

Рисунок 1 – Оценка времени прямых и обратных преобразований

Процесс кодирования и декодирования отображен на рисунках 2 - 6.

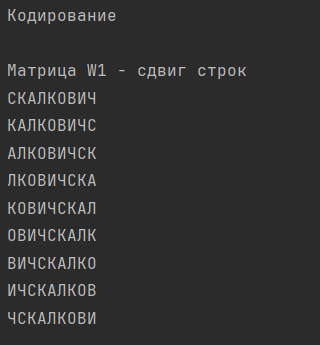


Рисунок 2 – Сдвиг строк

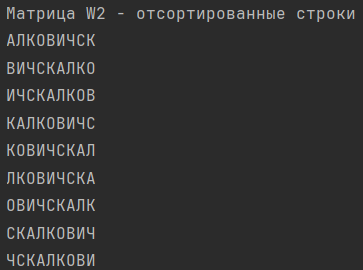


Рисунок 3 – Сортировка и получение закодированного сообщения и номера строки

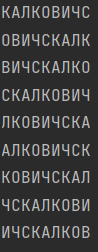


Рисунок 4 – Построение восстановленной матрицы

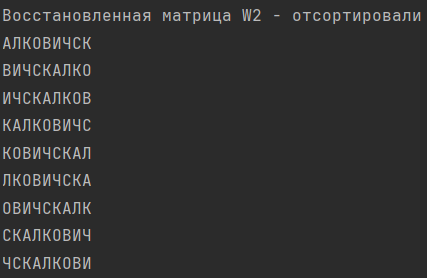


Рисунок 5 – Сортировка восстановленной матрицы

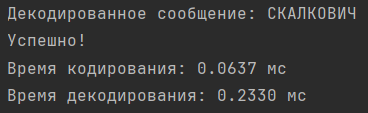


Рисунок 6 – Получение декодированного сообщения и анализ времени

**Вывод**

В результате данной лабораторной работы было разработано приложение для с использованием метода Барроуза−Уилера для сжатия/распаковки данных, а также получены теоретические сведения о сжатии и распаковке данных.