**Лабораторная работа № 8**

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

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

**Теоретические сведения**

Сжатие информации является одним из способов ее кодирования.

В основе сжатия данных, как одна из первопричин, лежит избыточность.

**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).

Основными техническими характеристиками процессов сжатия и результатов их работы являются:

* **степень сжатия, или отношение R** 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.

Степень сжатия R обычно оценивается следующим образом:

Отношение показывает, какую часть объема сообщения (файла) до сжатия занимает сообщение (файл) после сжатия.

Второе отношение выражает основной физический смысл сжатия и показывает степень сжатия.

Что касается третьей из приведенных технических характеристик (качества сжатия), то она показывает, по существу, совместимость данного метода с другими.

Все методы сжатия разделяют на два класса: **обратимое и необратимое** сжатие, или иначе: **сжатие без потерь и сжатие с частичной потерей** информации.

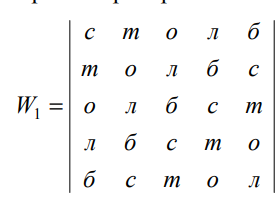
**BWT-преобразование** – техника сжатия информации (в особенности текстов), основанная на преобразовании.

BWT не сжимает данные в классическом понимании процесса, но преобразует блок данных в формат, исключительно подходящий для сжатия.

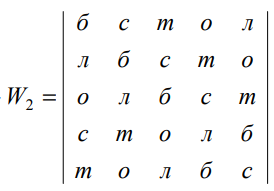
BWT оперирует сразу целым **блоком данных**, который выделяется из входного потока (сообщения).

**Прямое преобразование (формально – сжатие) выполняется в 4 этапа:**

1. выделяется блок данных = М;
2. таблица W1 размером k×k всех циклических сдвигов строки M;



1. производится лексикографическая (в алфавитном порядке) сортировка строк таблицы W1 = W2;

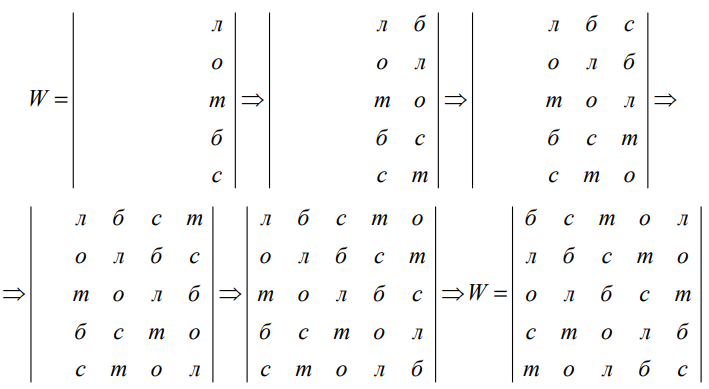


1. в качестве выходной строки (обозначим ее BWT(М), z) выбирается последний столбец (Мk) таблицы W2 преобразования и номер строки z, совпадающей с исходной строкой М.

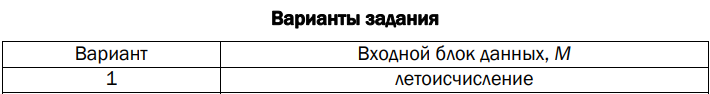
**Обратное преобразование выполняется в 2 этапа:**

1) в крайний справа пустой столбец матрицы записывается последовательность символов Мk;

2) производится лексикографическая сортировка столбцов заполненной части воссоздаваемой матрицы.



**Практическое задание**



1. Разработать авторское приложение в соответствии с целью лабораторной работы. Входной блок данных может иметь произвольную длину.

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 ниже.

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.

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.

4. Результаты оформить в виде отчета по установленным правилам.

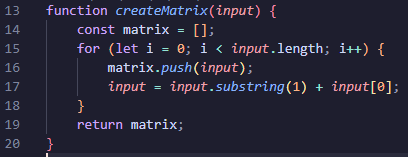


Рисунок 1.1 – Функция создания матрицы W1



Рисунок 1.2 – Функция сортировки матрицы в лексикографическом порядке (по алфавиту)

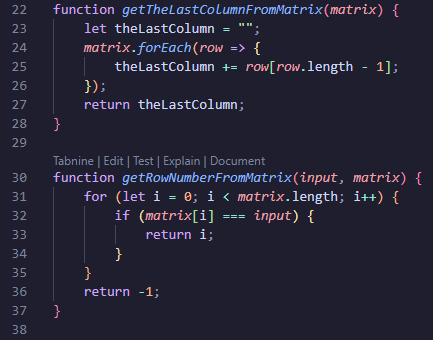


Рисунок 1.3 – Функции получения Mk и z

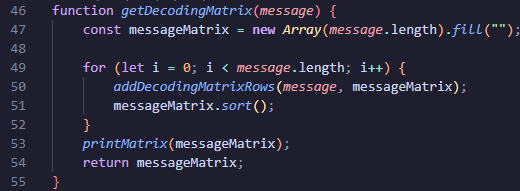


Рисунок 1.4 – Функция получения матрицы W2

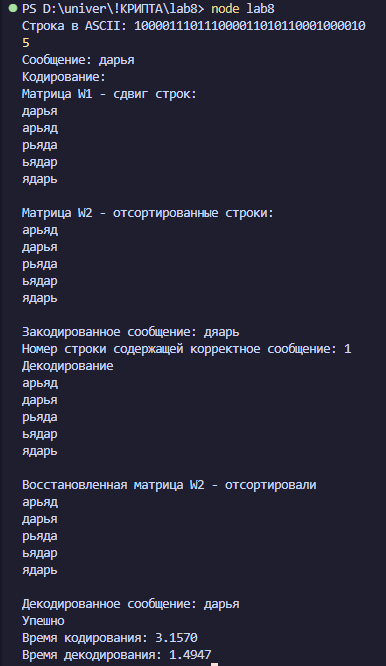


Рисунок 1.5 – Результат работы для пункта 1.а

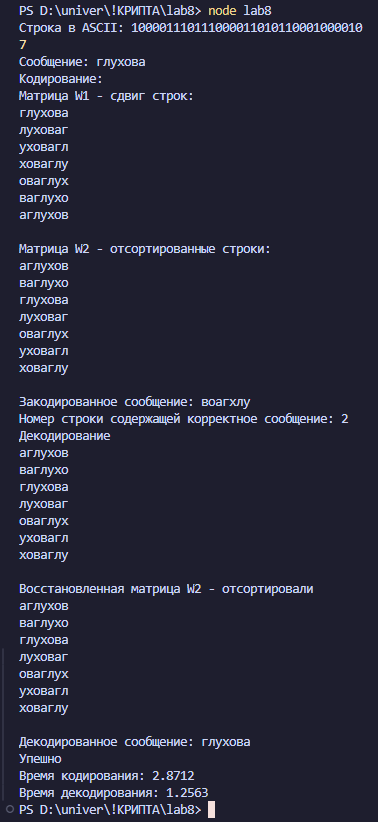
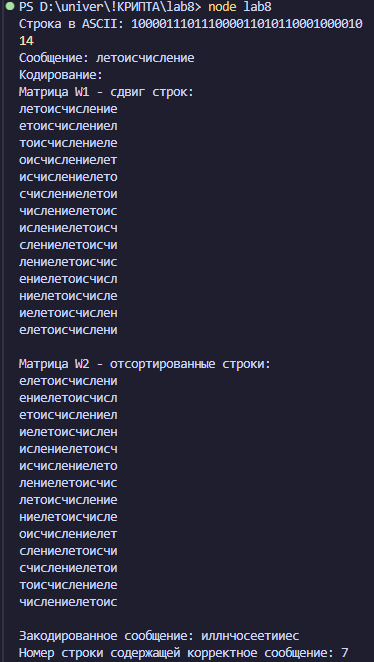


Рисунок 1.6 – Результат работы для пункта 1.б



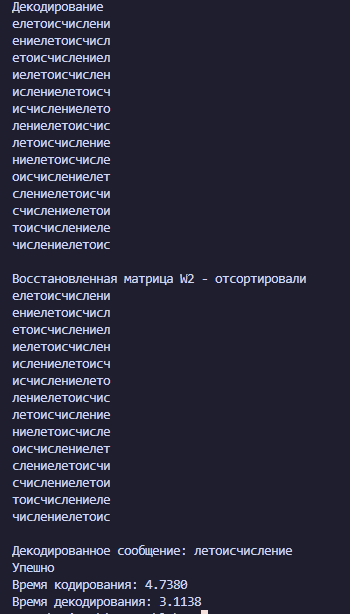
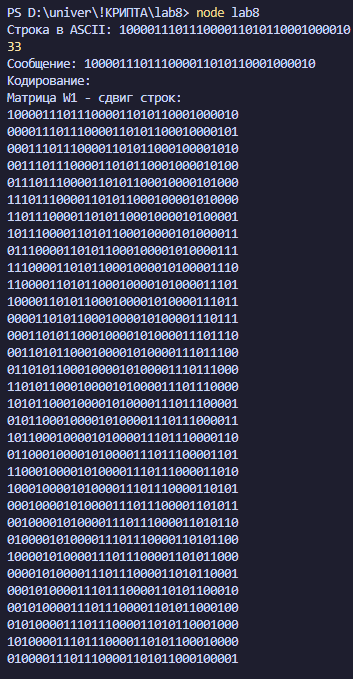
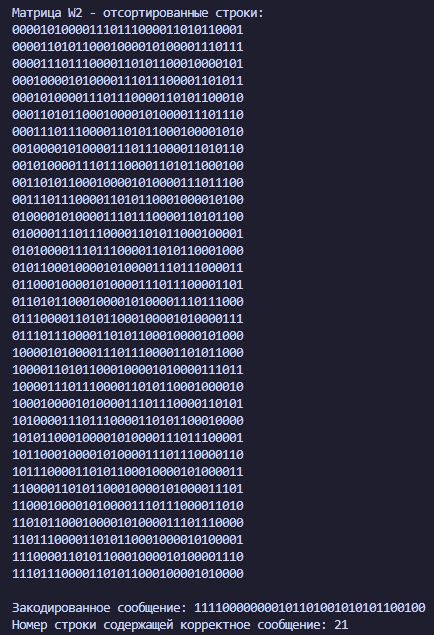


Рисунок 1.7 – Результат работы для пункта 1.в







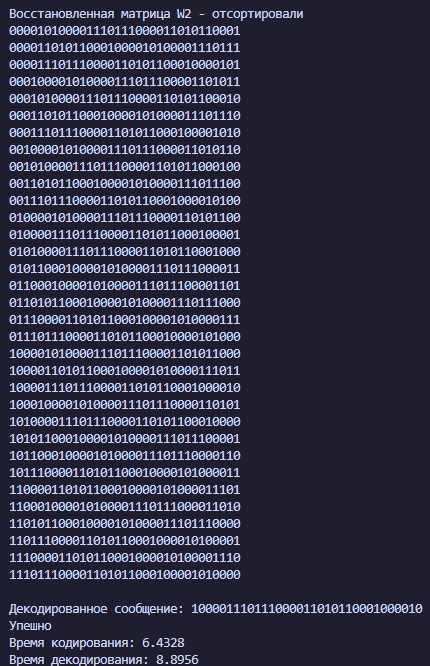


Рисунок 1.8 – Результат работы для пункта 3

**Вывод**

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