МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ

Учреждение образования «БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ

ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ»

Факультет информационных технологий

Кафедра информационных систем и технологий

Специальность Информационные системы и технологии

**ОТЧЕТ ПО ЛАБОРАТОРНОЙ РАБОТЕ №8 НА ТЕМУ:**

Сжатие/распаковка данных на основе статистических методов

Выполнил студент 3 курса 1 группы

Кашперко Василиса Сергеевна

Минск 2022

**Цель:** приобретение практических навыков использования статистических методов Шеннона−Фано и Хаффмана для сжатия/распаковки данных.

**Теоретические сведения**

До появления уже упоминавшихся работ К. Шеннона кодирование символов алфавита при передаче сообщения по каналам связи осуществлялось одинаковым количеством битов, получаемым по формуле Хартли. Позднее начали появляться способы, кодирующие символы разным числом битов в зависимости от вероятности появления их в тексте, подтверждение чему мы получили при выполнении лабораторной работы № 2. Таким образом, за счет использования для каждого значения байта кодов ASCII (символа алфавита) кода различной длины в соответствии с частотой (вероятностью появления этого символа в сообщении) можно значительно уменьшить общий размер данных.

Эта базовая идея лежит в основе алгоритмов статистических (вероятностных) методов сжатия: Шеннона−Фано и Хаффмана.

Статистические алгоритмы позволяют создавать более короткие коды для часто встречающихся и более длинные – для редко встречающихся символов алфавита или конкретного сообщения.

В первом случае метод считается статическим статистическим, во втором – динамическим статистическим: вероятностные свойства символов подсчитываются для конкретного сообщения или потока данных.

Частота или вероятность появления того или иного символа алфавита в произвольном сообщении, лежащая в основе алгоритмов, дали название этим алгоритмам и соответствующим методам.

Иногда эти методы называют также префиксными.

Таким образом, использование описываемых методов предусматривает создание кодовой таблицы (подобно кодам ASCII или base64). Формально процедура сжатия (прямое преобразование) состоит в подстановке соответствующего бинарного кода вместо символа исходного алфавита и наоборот – при обратном.

Методы относятся к классу «сжатие без потерь». Различие между двумя рассматриваемыми методами состоит лишь в особенностях формирования таблицы бинарных кодов.

При формировании этой таблицы для обоих методов можно воспользоваться статистическими свойствами алфавитов, полученными при выполнении лабораторной работы № 2. преобразовании.

**Ход работы**

Результат выполнения программы приведён на рисунках 1, 2, 3 и 4.

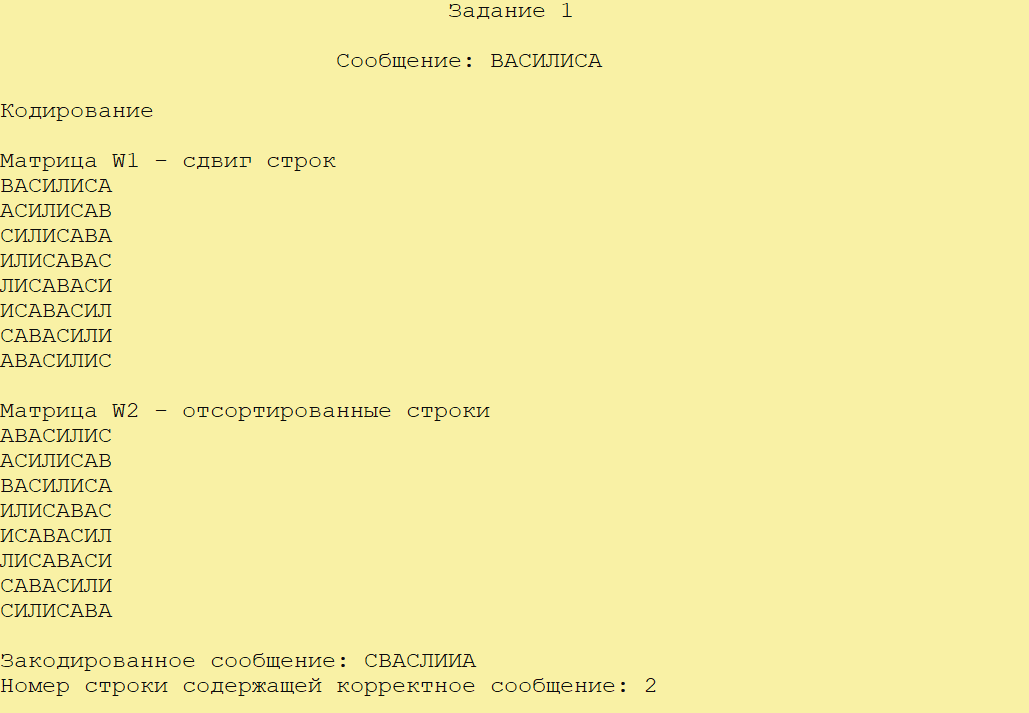


Рисунок 1 – Результат выполнения программы

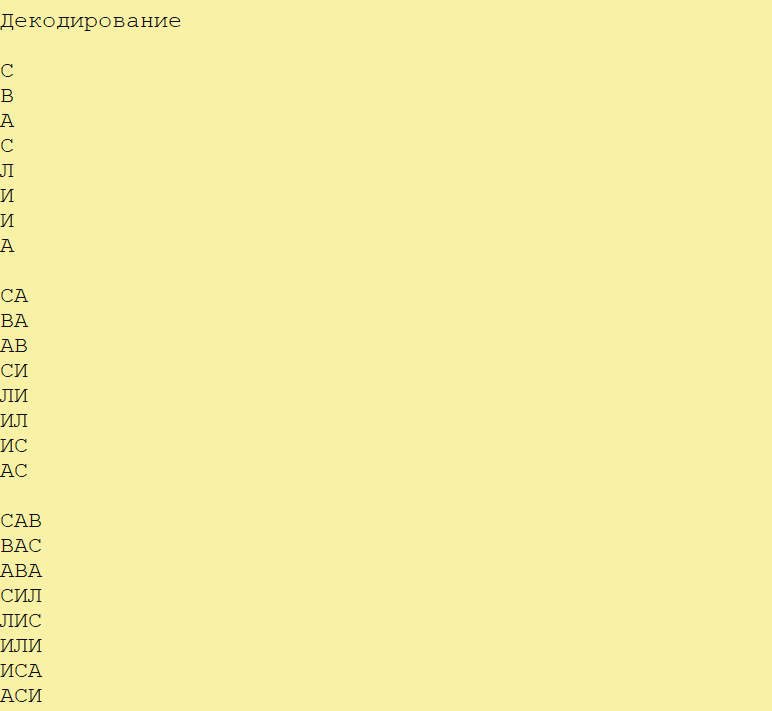


Рисунок 2 – Результат выполнения программы

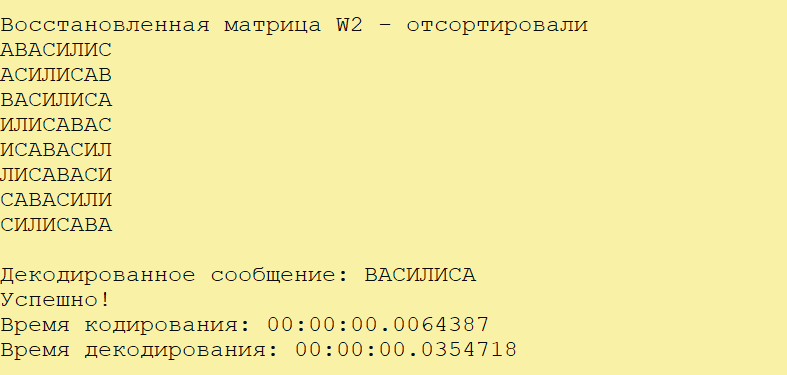


Рисунок 3 – Результат выполнения программы

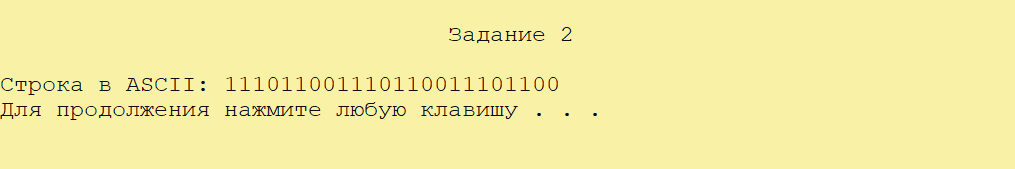


Рисунок 4 – Результат выполнения программы

В реализации программного кода был использован массив строк, что позволяет нам заполнять его данными, необходимыми для кодирования в соответствии с заданием лабораторной работы (имя, фамилия, слово по варианту, комбинация символов ASCII).

Метод не является оптимальным так как при равном распределении вероятностей он может выдавать несколько различных вариантов результатов, однако этот метод дает оптимальное распределение вероятностей и максимальное сжатие.

**Вывод**: в данной работе я изучила и реализовала способ сжатия данных на основе статистических данных – метод Шеннона–Фано. В ходе анализа времени можно сделать вывод: чем больше символов переданы в качестве сообщения, тем дольше происходит сжатие и распаковка. Именно поэтому данный метод не является оптимальным для времени сжатия, ведь если вместо 1 слова мы будем сжимать 20 предложений, это займет очень длительный промежуток времени. Также декодирование происходит дольше из-за того, что нам приходится выстраивать новую матрицу, каждый раз сортируя строки. При кодировании мы только 1 раз сдвигаем и затем сортируем, соответственно, и времени затрачиваем меньше.