МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

**«БЕЛГОРОДСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ им. В. Г. ШУХОВА» (БГТУ им. В.Г. Шухова)**

Кафедра программного обеспечения вычислительной техники и автоматизированных систем

**Лабораторная работа №1**по дисциплине: «Теория информации»

Выполнил: ст. группы ПВ-211

Медведев Дмитрий Сергеевич

Проверил:

Твердохлеб В.В.

Белгород 2023 г.

Исследование кодирования по методу Хаффмана. Оценка эффективности кода.

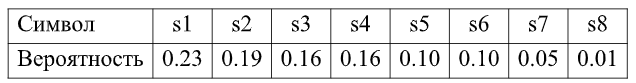
**Цель лабораторной работы:** изучить способ кодирования сообщений по методу Хаффмана. Научиться строить дерево кода Хаффмана и составлять код Хаффмана по таблице вероятности появления символов в пределах алфавита исходного сообщения. Узнать, как вычисляется коэффициент сжатия, значение средней длины образовавшейся кодовой конструкции, величина дисперсии. Научиться выявлять наиболее эффективные из кодовых моделей сообщения по критериям коэффициента сжатия и дисперсии.

**Задания:**

Задание 1.

Построить кодовое представление сообщения, вероятности появления символов в пределах алфавита которого приведены в табл.1.

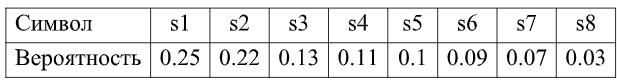
Таблица 1 – Вероятности появления символов в пределах алфавита исходного сообщения



Задание 2.

Построить кодовое представление сообщения, вероятности появления символов в пределах алфавита которого приведены в табл.2.

Таблица 2 – Вероятности появления символов в пределах алфавита исходного сообщения



Задание 3.

Для условий, приведенных в заданиях 1 и 2, выявить возможность построения альтернативных кодовых моделей сообщения. В случае обнаружения таковых, выявить наиболее эффективные из них по критериям K comp и .

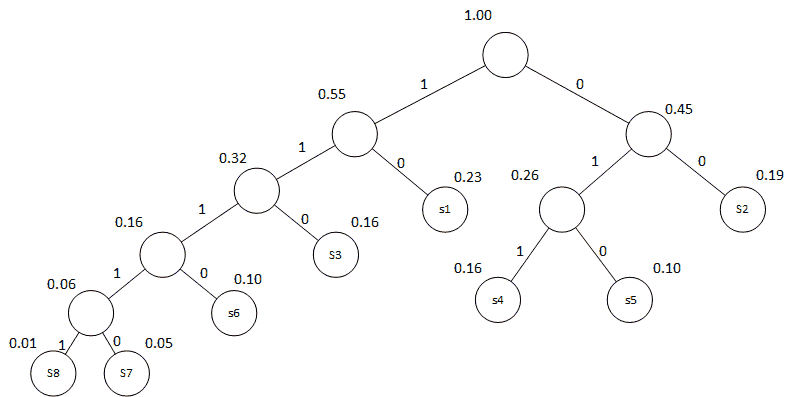
**Выполнение работы:**

Для получения кода по Хаффману воспользуемся методом деревьев по Хаффману.

1. В таблицах символов и их вероятностей каждая пара «символ-вероятность» рассматривается как один узел-лист дерева Хаффмана.
2. На первом шаге построения дерева выбирается пара узлов-листов, которым соответствуют минимальные величины вероятностей. Для них строится новый узел, вес которого будет равно сумме вероятностей входящих в него узлов-листов. Если на этом шаге, и на любом последующем, присутствует возможность выбора нескольких вариантов действий (например, есть более 2 узлов с минимальными и равными между собой вероятностями), можно выбирать любой из вариантов.
3. Новый узел, полученный на шаге 2, будет являться родительским по отношению к формирующим его узлам. С ними он соединяется ребрами, каждому из которых присваивается вес – 0 или 1 (можем назначать произвольно, но лучше в рамках текущего дерева придерживаться какого-то единого правила. Например – первым идет 1, или наоборот).
4. Пункты 2-3 повторяются до тех пор, пока не будет получена единая вершина (корень дерева), вес которой будет равен 1.
5. Двигаясь от корня по направлению к каждому узлу-листу, соответствующему тому или иному символу, считываются веса ребер. Из них формируется цепочка двоичных элементов, которая и будет кодом символа.

Задание 1

Построили дерево Хаффмана.



С его помощью получаем кодировку Хаффмана.

|  |  |
| --- | --- |
| S1 | 10 |
| S2 | 00 |
| S3 | 110 |
| S4 | 011 |
| S5 | 010 |
| S6 | 1110 |
| S7 | 11110 |
| S8 | 11111 |

Посчитаем параметры этого способа кодирования.

Для того, чтобы посчитать коэффициент сжатия допустим, что мы сгенерировали сообщение длинной в 100 символов с заданными вероятностями.

Его вес в 8 битной кодировке B = 8 \* 100 = 800 бит.

Вес при кодировании полученным кодом

Вычислим дисперсию.

 = 2,8

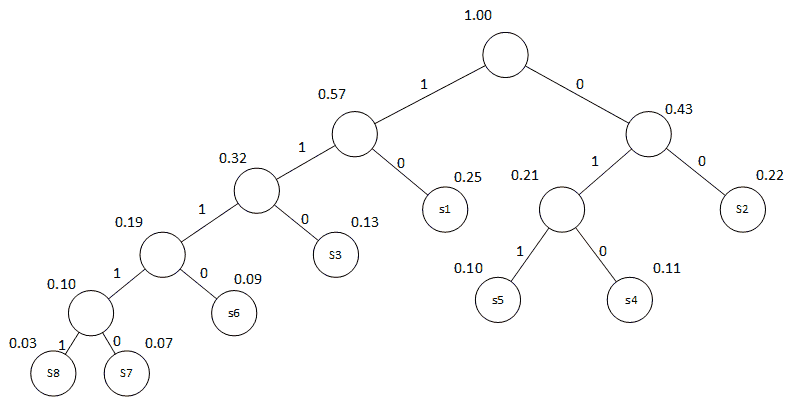


В данном случае есть вариации дерева Хаффмана, но на эффективность кодирования они не влияют, поэтому и рассматривать их не нужно.

Задание 2

В данном случае есть возможность построить два вида деревьев, дающих различные коды и их характеристики.

Рассмотрим первое дерево.



С его помощью получаем кодировку Хаффмана.

|  |  |
| --- | --- |
| S1 | 10 |
| S2 | 00 |
| S3 | 110 |
| S4 | 010 |
| S5 | 011 |
| S6 | 1110 |
| S7 | 11110 |
| S8 | 11111 |

Посчитаем параметры этого способа кодирования.

Для того, чтобы посчитать коэффициент сжатия допустим, что мы сгенерировали сообщение длинной в 100 символов с заданными вероятностями.

Его вес в 8 битной кодировке B = 8 \* 100 = 800 бит.

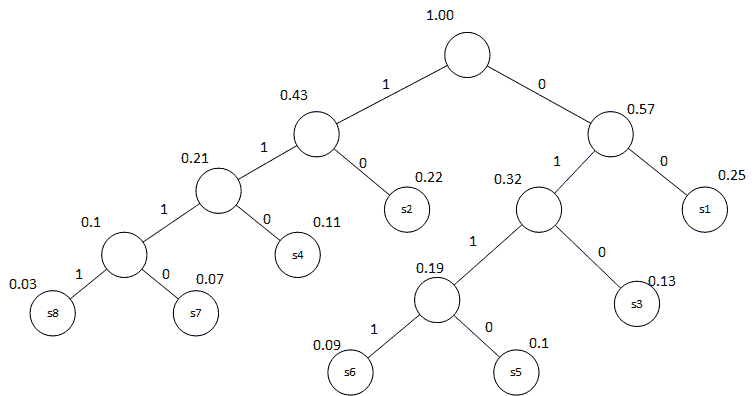
Вес при кодировании полученным кодом

Вычислим дисперсию.

 = 2



Рассмотрим второе дерево.



С его помощью получаем кодировку Хаффмана.

|  |  |
| --- | --- |
| S1 | 00 |
| S2 | 10 |
| S3 | 010 |
| S4 | 110 |
| S5 | 0110 |
| S6 | 0111 |
| S7 | 1110 |
| S8 | 1111 |

Посчитаем параметры этого способа кодирования.

Для того, чтобы посчитать коэффициент сжатия допустим, что мы сгенерировали сообщение длинной в 100 символов с заданными вероятностями.

Его вес в 8 битной кодировке B = 8 \* 100 = 800 бит.

Вес при кодировании полученным кодом

Вычислим дисперсию.

 = 2



В данном случае вторая кодировка превосходит первую по показателю дисперсии, а по коэффициенту сжатия они одинаковы. Вторая кодировка более предпочтительна.

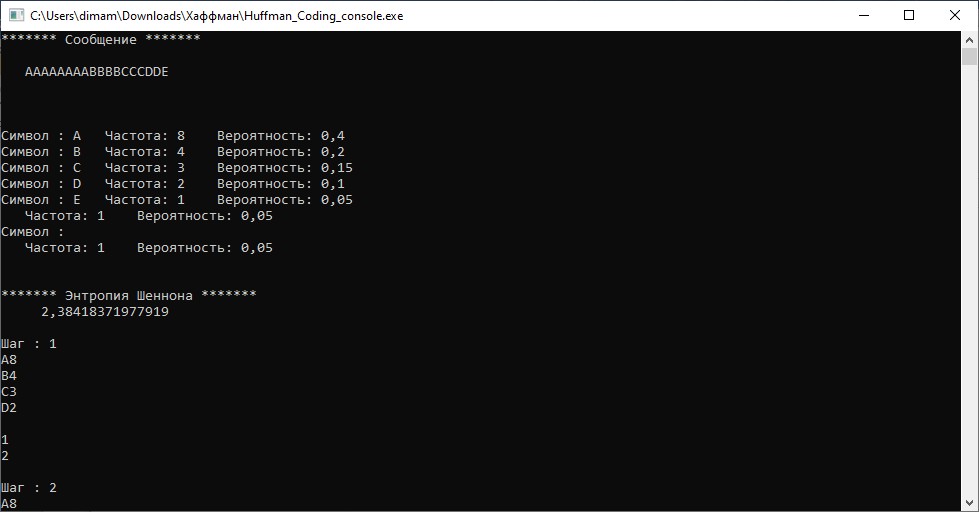
Задание 3

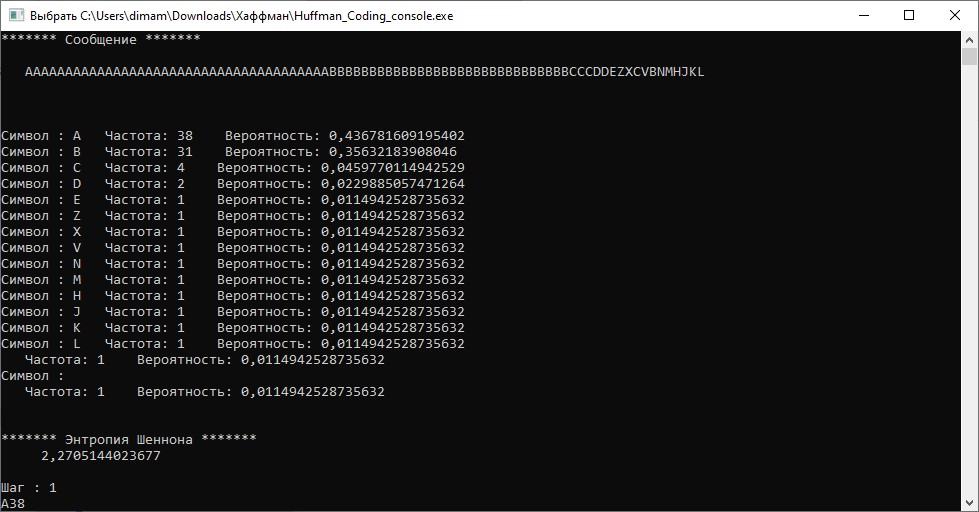
Выполнено в ходе рассмотрения заданий 1 и 2.

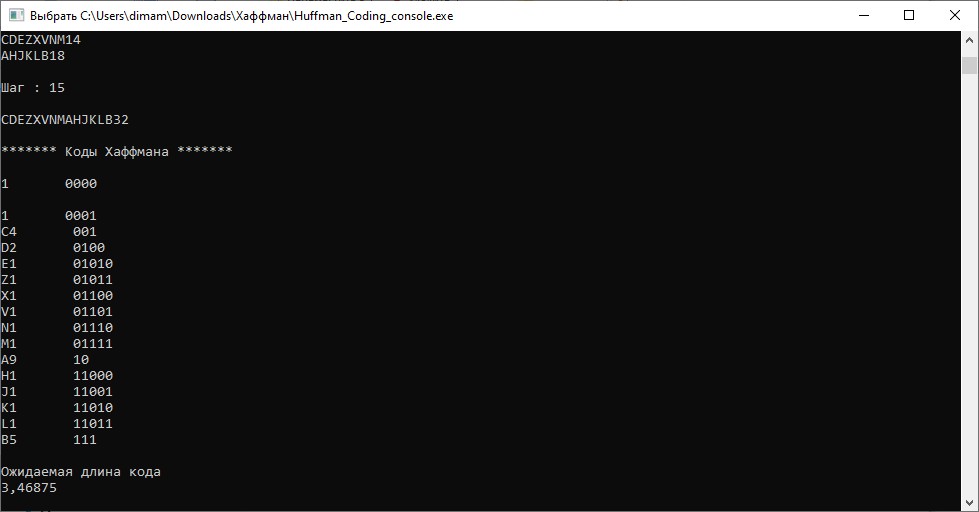
В ходе работы с программой Huffman\_Coding\_console, были замечены следующие зависимости.

Если в сообщение присутствуем много различных символов, энтропия Шеннона и средняя длина кода возрастают.

Если в сообщении будет большая доля символов лишь некоторых типов, то энтропия Шеннона и средняя длина кода принимают низкое значение. Коды Хаффмана хорошо сжимают сообщения, где много одинаковых символов.







**Вывод:** в ходе работы изучен способ кодирования сообщений по методу Хаффмана. Получены навыки составления кода Хаффмана по таблице вероятности появления символов в пределах алфавита исходного сообщения, нахождения коэффициента сжатия, значения средней длины образовавшейся кодовой конструкции, величины дисперсии, выявления наиболее эффективных из кодовых моделей сообщения по критериям коэффициента сжатия и дисперсии.