

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

тема: «Исследование кодирования по методу Хаффмана. Оценка
эффективности кода»

Выполнил: ст. группы ПВ-233

Ситников Алексей Павлович

Проверил: Твердохлеб Виталий

Викторович

Белгород 2025 г.

Цель работы: изучить метод Хаффмана и научиться его применять.

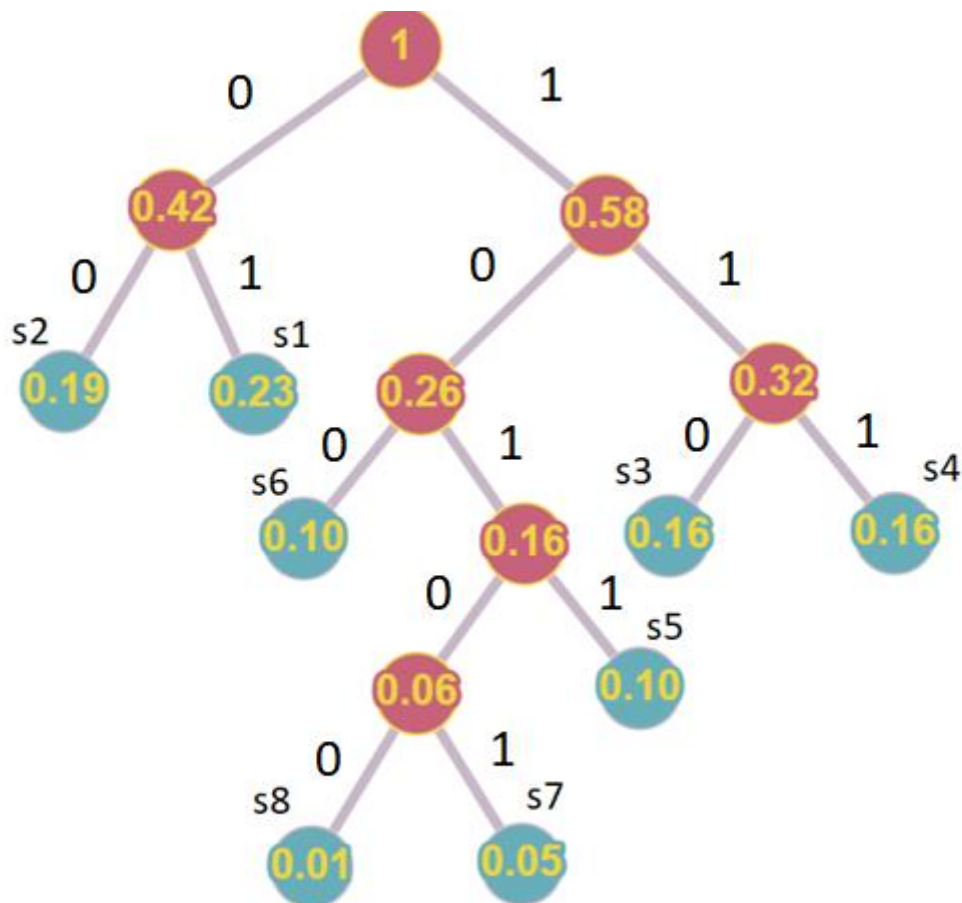
Задание 1. Построить кодовое представление сообщения, вероятности появления символов в пределах алфавита которого приведены в табл.1.

Таблица 1 – Вероятности появления символов в пределах алфавита исходного сообщения

Символ	s1	s2	s3	s4	s5	s6	s7	s8
Вероятность	0.23	0.19	0.16	0.16	0.10	0.10	0.05	0.01

Расположим вероятности по возрастанию:

Символ	s8	s7	s5	s6	s3	s4	s2	s1
Вероятность	0.01	0.05	0.10	0.10	0.16	0.16	0.19	0.23



s1 = 01
s2 = 00
s3 = 110
s4 = 111
s5 = 1011

s6 = 100
s7 = 10101
s8 = 10100

Пример предложения с данным алфавитом:

s1s2s2s4s4s5s1s1s7s8s8s4s3s1s2s2s3 кодируется в:
010000111111101101011010110100101001111100100001110

Найдём В: $B = 17 * 8 = 136$

Найдём В': $B' = 4 * 2 + 4 * 2 + 2 * 3 + 3 * 3 + 4 + 5 + 2 * 5 = 50$

Коэффициент сжатия $K_{\text{comp}} = 136/50 = 2,72$

Найдём $\lambda_{\text{ср}} = 2 * 0,23 + 2 * 0,19 + 3 * 0,16 + 3 * 0,16 + 3 * 0,1 + 4 * 0,1 + 5 * 0,05 + 5 * 0,01 = 2,4$

Величина дисперсии:

$\delta = 0,23 * (2 - 2,4)^2 + 0,19 * (2 - 2,4)^2 + 0,16 * (3 - 2,4)^2 + 0,16 * (3 - 2,4)^2 + 0,1 * (4 - 2,4)^2 + 0,1 * (3 - 2,4)^2 + 0,05 * (5 - 2,4)^2 + 0,01 * (5 - 2,4)^2 = 0,57$

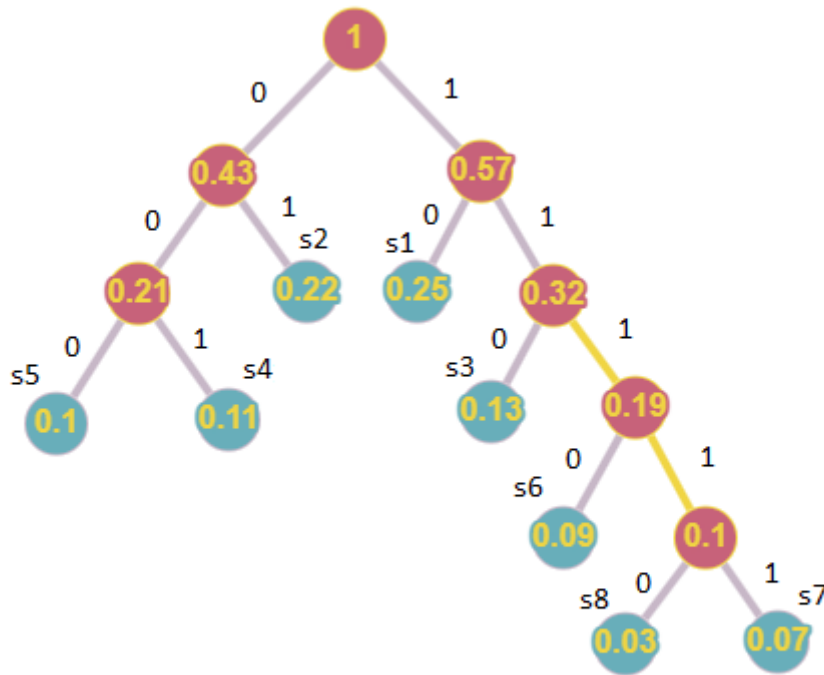
Задание 2. Построить кодовое представление сообщения, вероятности появления символов в пределах алфавита которого приведены в табл.2.

Таблица 2 – Вероятности появления символов в пределах алфавита исходного сообщения

Символ	s1	s2	s3	s4	s5	s6	s7	s8
Вероятность	0.25	0.22	0.13	0.11	0.1	0.09	0.07	0.03

Расположим вероятности по возрастанию:

Символ	s8	s7	s6	s5	s4	s3	s2	s1
Вероятность	0.03	0.07	0.09	0.1	0.11	0.13	0.22	0.25



s1 = 10

s2 = 01

s3 = 110

s4 = 001

s5 = 000

s6 = 1110

s7 = 11111

s8 = 11110

Пример предложения с данным алфавитом:

s1s2s2s4s4s5s1s1s7s8s8s4s3s1s2s2s3 кодируется в:

100101001001000101011111111011110001110100101110

Найдём В: $B = 17 * 8 = 136$

Найдём В': $B' = 4 * 2 + 4 * 2 + 2 * 3 + 3 * 3 + 3 + 5 + 2 * 5 = 49$

Коэффициент сжатия $K_{comp} = 136/49 \approx 2,78$

Найдём $\lambda_{cp} = 2 * 0,25 + 2 * 0,22 + 3 * 0,13 + 3 * 0,11 + 3 * 0,1 + 4 * 0,09 + 5 * 0,07 + 5 * 0,03 = 2,82$

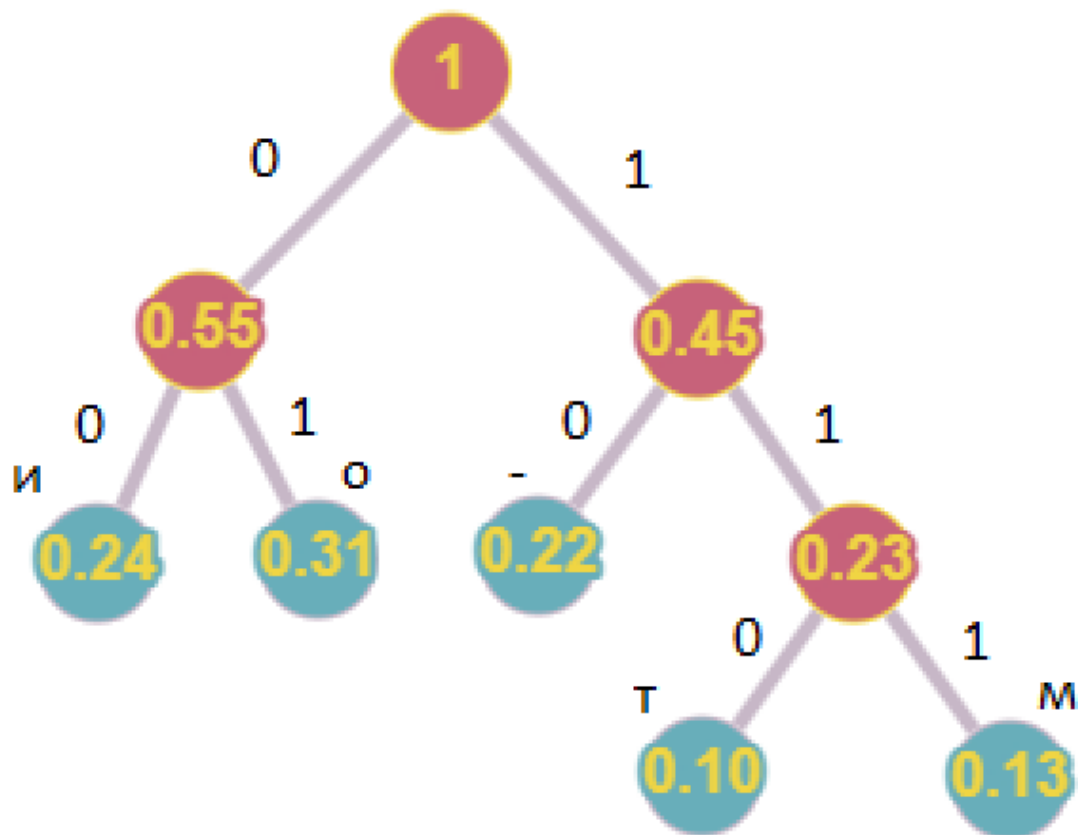
Величина дисперсии:

$$\delta = 0,25 * (2 - 2,82)^2 + 0,22 * (2 - 2,82)^2 + 0,13 * (3 - 2,82)^2 + 0,11 * (3 - 2,82)^2 + 0,1 * (3 - 2,82)^2 + 0,09 * (4 - 2,82)^2 + 0,07 * (5 - 2,82)^2 + 0,03 * (5 - 2,82)^2 = 0,93$$

Задание 3. Построить кодовое представление сообщения:

оитомии о ими оооитми о о о ооимтмотиимотоим оои тоо и и м оио и омтоо тоимо т и

Символ	о	и	т	м	–
Вероятность	0.31	0.24	0.1	0.13	0.22



и = 00
о = 01
- = 10
т = 110
м = 111

0100110011110000100110001110010001100100101010011011001000110011001
10010100001111100111100001110111001001111001010010110010110001000101
111001000110001001111110010110110010011101101101000

Задание 4. Для условий, приведенных в заданиях 1 и 2 и 3, выявить возможность построения альтернативных кодовых моделей сообщения. В случае обнаружения таковых выявить наиболее эффективные из них по критериям K_{comp} и δ .

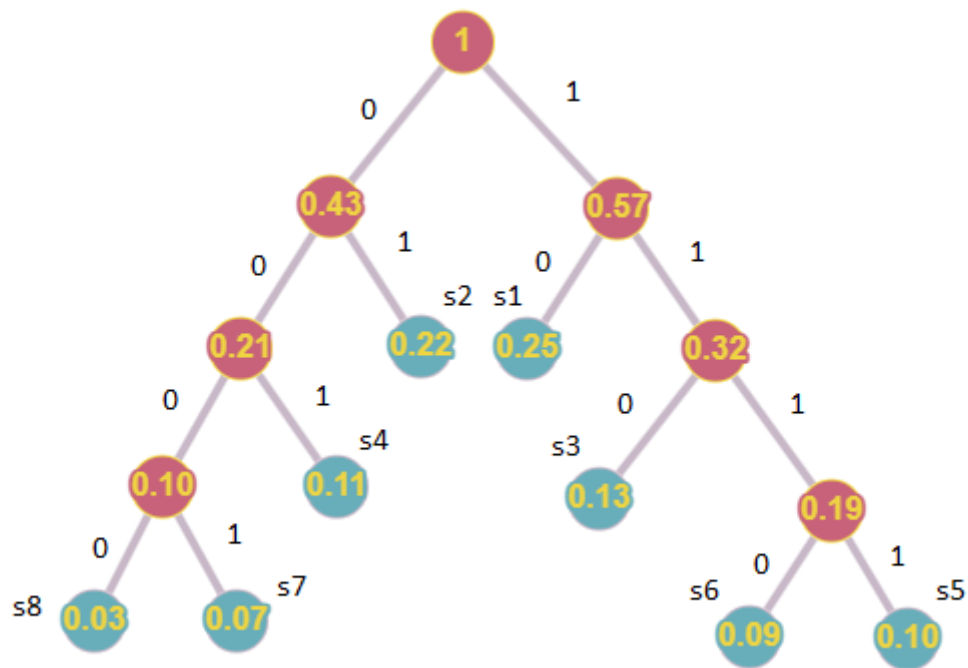
Рассмотрим задание 1.

После второго склеивания получим три одинаковые вероятности, но при замене мест будет меняться только код символов, но не размер, перестановка не даст разницы.

Рассмотрим задание 2.

После первого склеивания получаем две одинаковые вероятности 0.1, выбор, какую вероятность разместить раньше, повлияет на итог.

В задание выше первым взяли полученную на предыдущем этапе вероятность, но если взять из таблицы вероятность символа получится такое дерево:



Получим такие коды:

$s_1 = 10$

$s_2 = 01$

$s_3 = 110$

$s_4 = 001$

$s_5 = 1111$

$s_6 = 1110$

$s_7 = 0001$

$s_8 = 0000$

Высота дерева уменьшилась на 1.

Пример предложения с данным алфавитом:

$s_1 s_2 s_2 s_4 s_4 s_5 s_1 s_1 s_7 s_8 s_8 s_4 s_3 s_1 s_2 s_2 s_3$ кодируется в:

10010100100111111010000100000000001110100101110

Найдём B : $B = 17 * 8 = 136$

Найдём B' : $B' = 4 * 2 + 4 * 2 + 2 * 3 + 3 * 3 + 4 + 4 + 2 * 4 = 47$

Коэффициент сжатия $K_{\text{comp}} = 136/47 \approx 2,89$

Найдём $\lambda_{\text{ср}} = 2 * 0,25 + 2 * 0,22 + 3 * 0,13 + 3 * 0,11 + 4 * 0,1 + 4 * 0,09 + 4 * 0,07 + 4 * 0,03 = 2,82$

Величина дисперсии:

$$\delta = 0,25 * (2 - 2,82)^2 + 0,22 * (2 - 2,82)^2 + 0,13 * (3 - 2,82)^2 + 0,11 * (3 - 2,82)^2 + 0,1 * (4 - 2,82)^2 + 0,09 * (4 - 2,82)^2 + 0,07 * (4 - 2,82)^2 + 0,03 * (4 - 2,82)^2 = 0,72$$

Коэффициент сжатия был 2,78; а стал 2,89. Дисперсия была 0,93; а стала 0,72. Чем больше коэффициент сжатия, тем лучше, чем меньше дисперсия, тем лучше, значит второй вариант по обоим параметрам обходит первый вариант.

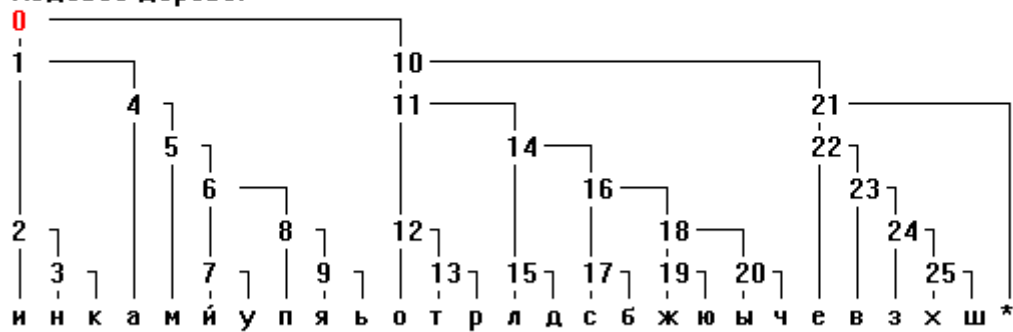
В третьем задании дерево, построенное по методу Хаффмана, имеет один вариант.

Приложение Хаффман:

Инструкции автомата:

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
0	1	2	и	н	а	м	7	й	п	я	11	12	о	т	15	л	17	с	19	ж	ы	22	е	в	з	х
1	10	4	3	к	5	6	8	у	9	ь	21	14	13	р	16	д	18	б	20	ю	ч	*	23	24	25	ш

Кодовое дерево:

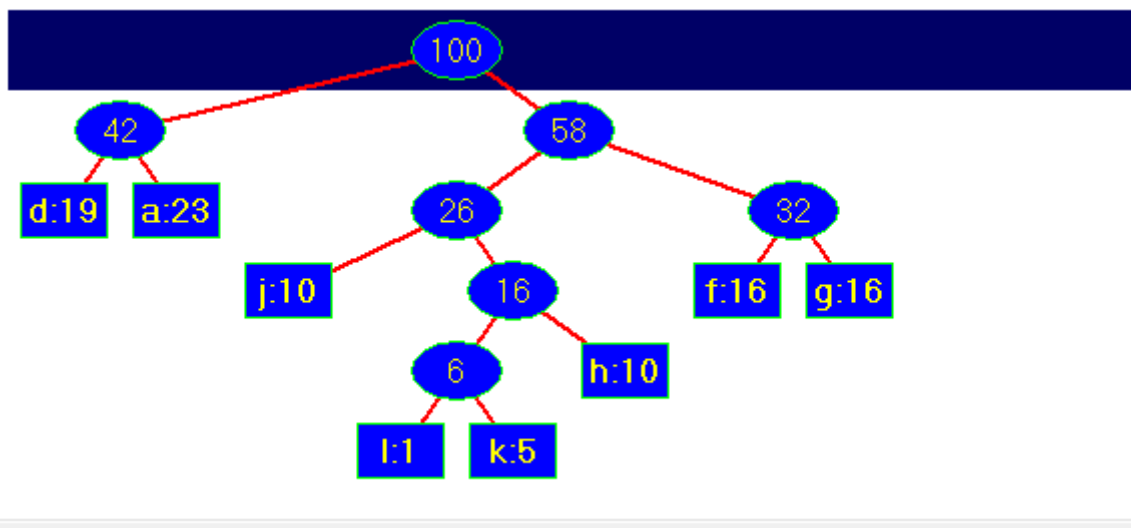


Входная строка: 0111101001100011010110010010

Выходная строка: привет

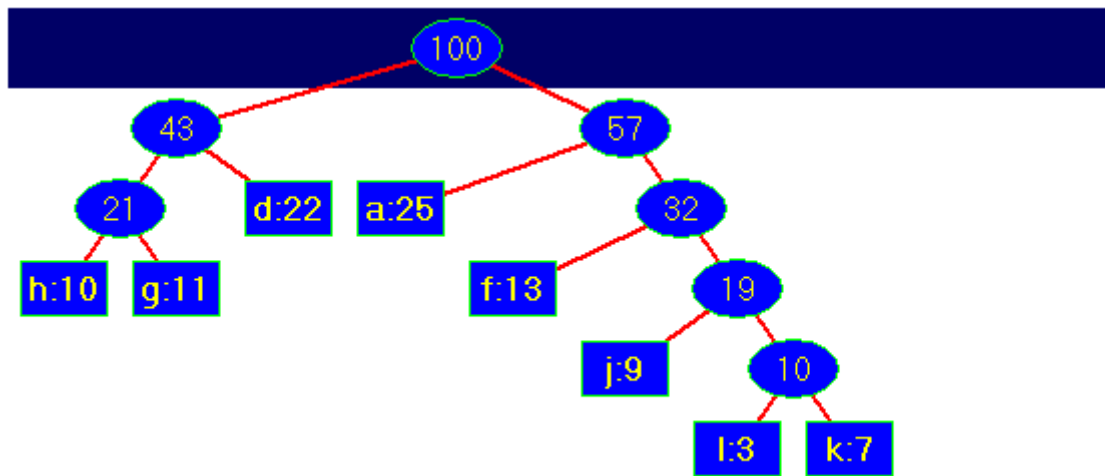
Приложение greedy2.exe:

Входные данные для задания №1, получил такое дерево:



Совпало с вычисленным вручную.

Входные данные для задания №1, получил такое дерево:



Совпало с вычисленным вручную.

Вывод: после проделанной работы я изучил метод Хаффмана и научился его применять, изучил критерии эффективности кодирования.