

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ
ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ
«МИСИС»
ИНСТИТУТ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ И КОМПЬЮТЕРНЫХ НАУК
КАФЕДРА АВТОМАТИЗИРОВАННЫХ СИСТЕМ УПРАВЛЕНИЯ

Лабораторная работа №2
«Принцип кодирования по Хаффману»
по дисциплине
«Информационная безопасность и кодирование»

Выполнила:
студентка группы БИВТ-21-4
Савенко Е. И.

Проверил:
Гончаренко А. Н.

Москва, 2025

Цель работы: Изучить алгоритм Хаффмана и применить его для строки с ФИО, датой рождения и городом.

Задачи:

- Закодируйте по алгоритму Хаффмана строку с вашим именем, отчеством, фамилией, датой и местом рождения (например, «Иванова Наталья Николаевна, 1 января 1990 года, город Тверь»).
- При кодировании не округляйте частоты менее, чем четыре знака после запятой – сокращение точности понижает эффективность кодирования.
- Подсчитайте коэффициент сжатия.
- Наберите эту же строку в редакторе «Блокнот» и сохраните текст в txt-формате (количество байт в файле должно в точности соответствовать числу знаков – букв, цифр и служебных символов – в строке).
- Примените к файлу любой архиватор и сравните его степень сжатия с алгоритмом Хаффмана.

Ход работы:

Входная строка выглядит так:

“САВЕНКО ЕКАТЕРИНА ИГОРЕВНА, 14 АВГУСТА 2003 ГОДА, ГОРОД
МОСКВА”.

Проанализируем частоту встречаемости каждого символа сообщения (таблица 1).

Таблица 1 – Распределение символов сообщения по убыванию частоты

Символ	Повтор
С	3
А	8
В	4
Е	4
Н	3
К	3
О	6
ПРОБЕЛ	8
Т	2
Р	3
И	2
Г	4

№ п/п	Символ	Повтор	Частота
1	А	8	0,1290
2	ПРОБЕЛ	8	0,1290
3	О	6	0,0968
4	В	4	0,0645
5	Е	4	0,0645
6	Г	4	0,0645
7	С	3	0,0484
8	Н	3	0,0484
9	К	3	0,0484
10	Р	3	0,0484
11	Т	2	0,0323
12	И	2	0,0323

ЗАПЯТАЯ	2
1	1
4	1
У	1
2	1
0	2
3	1
Д	2
М	1
ВСЕГО	62 (байт)

13	ЗАПЯТАЯ	2	0,0323
14	0	2	0,0323
15	Д	2	0,0323
16	1	1	0,0161
17	4	1	0,0161
18	У	1	0,0161
19	2	1	0,0161
20	3	1	0,0161
21	М	1	0,0161

Далее были проведены преобразования для определения приоритетности каждого символа (таблица 2).

Таблица 2 – Результаты преобразований

Шаг 1: Исходная таблица частот			
Индекс	Символ	Повтор	Частота
1	А	8	0,129
2	ПРОБЕЛ	8	0,129
3	О	6	0,0968
4	В	4	0,0645
5	Е	4	0,0645
6	Г	4	0,0645
7	С	3	0,0484
8	Н	3	0,0484
9	К	3	0,0484
10	Р	3	0,0484
11	0	2	0,0323
12	Д	2	0,0323
13	Т	2	0,0323
14	ЗАПЯТАЯ	2	0,0323
15	И	2	0,0323
16	1	1	0,0161
17	4	1	0,0161

18	У	1	0,0161
19	2	1	0,0161
20	3	1	0,0161
21	М	1	0,0161

Шаги 2-4: Объединены 1 и 3, М и 2, У и 4			
Индекс	Символ	Повтор	Частота
18, 17	У, 4	1, 1	0,0322
21, 19	М, 2	1, 1	0,0322
14	ЗАПЯТАЯ	2	0,0323
13	Т	2	0,0323
12	Д	2	0,0323
11	0	2	0,0323
15	И	2	0,0323
8	Н	3	0,0484
10	Р	3	0,0484
9	К	3	0,0484
7	С	3	0,0484
6	Г	4	0,0645
5	Е	4	0,0645
4	В	4	0,0645
3	О	6	0,0968
2	ПРОБЕЛ	8	0,129
1	А	8	0,129
16, 20	1, 3	1, 1	0,0322

Шаг 5-8: Объединены 0 и И; Т и Д; М, 2 и ЗАПЯТАЯ; У, 4 и 1, 3; 18, 17, 16, 20			
Индекс	Символ	Повтор	Частота
10	Р	3	0,0484
7	С	3	0,0484
8	Н	3	0,0484
9	К	3	0,0484
18, 17, 16, 20	У, 4, 1, 3	1, 1, 1, 1	0,0644

6	Г	4	0,0645
5	Е	4	0,0645
4	В	4	0,0645
21, 19, 14	М, 2, ЗАПЯТАЯ	1, 1, 2	0,0645
13, 12	Т, Д	2, 2	0,0646
3	О	6	0,0968
2	ПРОБЕЛ	8	0,129
1	А	8	0,129
11, 15	0, И	2, 2	0,0646

Шаг 9-10: Объединены Н и К, Р и С			
Индекс	Символ	Повтор	Частота
18, 17, 16, 20	У, 4, 1, 3	1, 1, 1, 1	0,0644
6	Г	4	0,0645
5	Е	4	0,0645
4	В	4	0,0645
21, 19, 14	М, 2, ЗАПЯТАЯ	1, 1, 2	0,0645
13, 12	Т, Д	2, 2	0,0646
11, 15	0, И	2, 2	0,0646
3	О	6	0,0968
10, 7	Р, С	3, 3	0,0968
2	ПРОБЕЛ	8	0,129
1	А	8	0,129
8, 9	Н, К	3, 3	0,0968

Шаг 12-14: Объединены 0, И и О; У, 4, 1, 3 и Г; М, 2, ЗАПЯТАЯ и Т, Д; Е и В			
Индекс	Символ	Повтор	Частота
10, 7	Р, С	3, 3	0,0968
8, 9	Н, К	3, 3	0,0968
18, 17, 16, 20, 6	У, 4, 1, 3, Г	1, 1, 1, 1, 4	0,1289
2	ПРОБЕЛ	8	0,129
1	А	8	0,129
5, 4	Е, В	4, 4	0,129

21, 19, 14, 13, 12	М, 2, ЗАПЯТАЯ, Т, Д	1, 1, 2, 2, 2	0,1291
11, 15, 3	0, И, О	2, 2, 6	0,1614

Шаг 15-18: Объединены М, 2, ЗАПЯТАЯ, Т, Д и 0, И, О; А и Е, В; У, 4, 1, 3, Г и ПРОБЕЛ; Р, С и Н, К

Индекс	Символ	Повтор	Частота
10, 7, 8, 9	Р, С, Н, К	3, 3, 3, 3	0,1936
18, 17, 16, 20, 6, 2	У, 4, 1, 3, Г, ПРОБЕЛ	1, 1, 1, 1, 4, 8	0,2579
1, 5, 4	А, Е, В	8, 4, 4	0,258
21, 19, 14, 13, 12, 11, 15, 3	М, 2, ЗАПЯТАЯ, Т, Д, 0, И, О	1, 1, 2, 2, 2, 2, 2, 6	0,2905

Шаг 19-20: Объединены А, Е, В и М, 2, ЗАПЯТАЯ, Т, Д, 0, И, О; Р, С, Н, К и У, 4, 1, 3, Г, ПРОБЕЛ

Индекс	Символ	Повтор	Частота
10, 7, 8, 9, 18, 17, 16, 20, 6, 2	Р, С, Н, К, У, 4, 1, 3, Г, ПРОБЕЛ	3, 3, 3, 3, 1, 1, 1, 1, 4, 8	0,4515
1, 5, 4, 21, 19, 14, 13, 12, 11, 15, 3	А, Е, В, М, 2, ЗАПЯТАЯ, Т, Д, 0, И, О	8, 4, 4, 1, 1, 2, 2, 2, 2, 2, 6	0,5485

Итоговая таблица для построения кодов Хаффмана:

Индекс	Символ	Повтор	Частота
1, 5, 4, 21, 19, 14, 13, 12, 11, 15, 3	А, Е, В, М, 2, ЗАПЯТАЯ, Т, Д, 0, И, О	8, 4, 4, 1, 1, 2, 2, 2, 2, 2, 6	0,5485
10, 7, 8, 9, 18, 17, 16, 20, 6, 2	Р, С, Н, К, У, 4, 1, 3, Г, ПРОБЕЛ	3, 3, 3, 3, 1, 1, 1, 1, 4, 8	0,4515

Из полученных таблиц последовательностей, было построено дерево кодирования (рисунок 1).

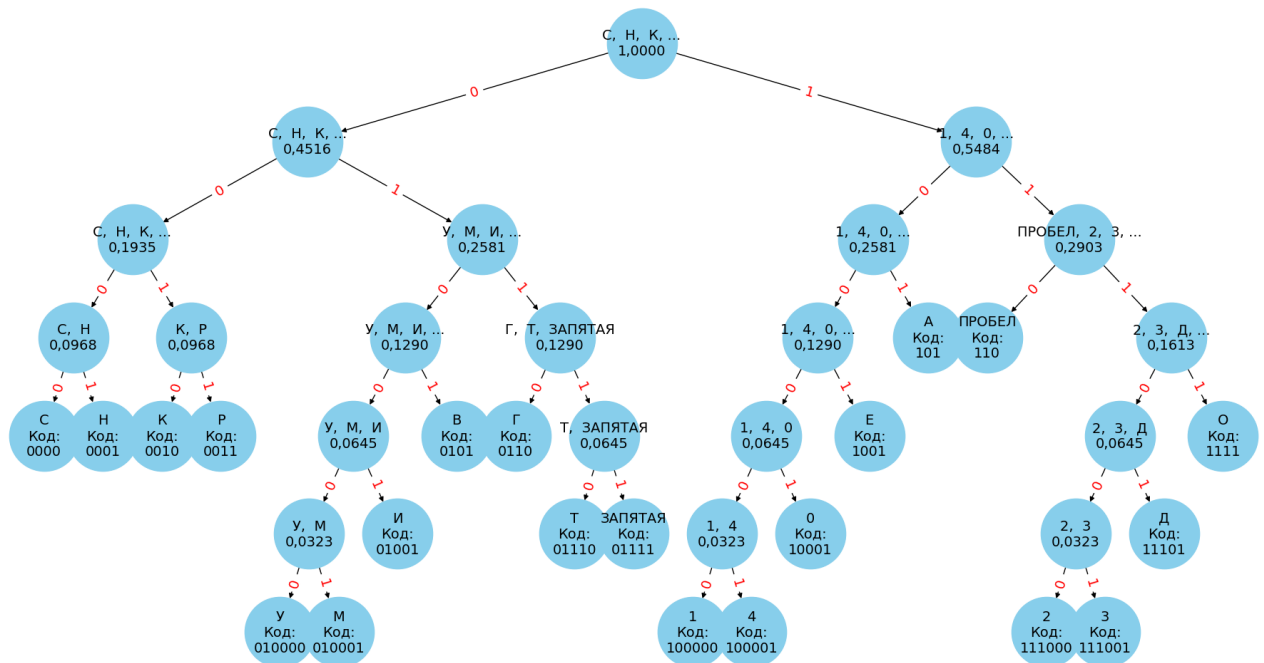


Рисунок 1 - Дерево кодирования

По результатам данных операций можно составить таблицу кодирования (таблица 3).

Таблица 3 – Таблица кодирования

№ п/п	Символ	Код
1	А	101
2	ПРОБЕЛ	110
3	С	0000
4	Н	0001
5	К	0010
6	Р	0011
7	В	0101
8	Г	0110
9	Е	1001
10	О	1111
11	И	01001
12	Т	01110

13	ЗАПЯТАЯ	01111
14	0	10001
15	Д	11101
16	У	010000
17	М	010001
18	1	100000
19	4	100001
20	2	111000
21	3	111001

В итоге закодированная строка будет выглядеть так:

```
000010101011001000100101111110100100101010111010010011010010001101110010
01011011110011100101010001101011111101000001000011101010101011001000000000111
010111011100010001100011110011001101111111011010111111001101111001111111101
1100100011111000000100101101
```

Подсчитаем полученную степень сжатия, умножая длину кода на сумму символов, кодируемых кодом данной длины:

$$3(8 + 8) + 4(6 + 4 + 4 + 4 + 3 + 3 + 3 + 3) + 5(2 + 2 + 2 + 2 + 2) + 6(1 + 1 + 1 + 1 + 1 + 1) = 254 \text{ бит} = 31,75 \text{ байт}$$

Таким образом, имеем коэффициент сжатия $K = \frac{62}{32} = 1,94$

Теперь сравним результаты сжатия с работой архиватора (рисунок 2, 3).

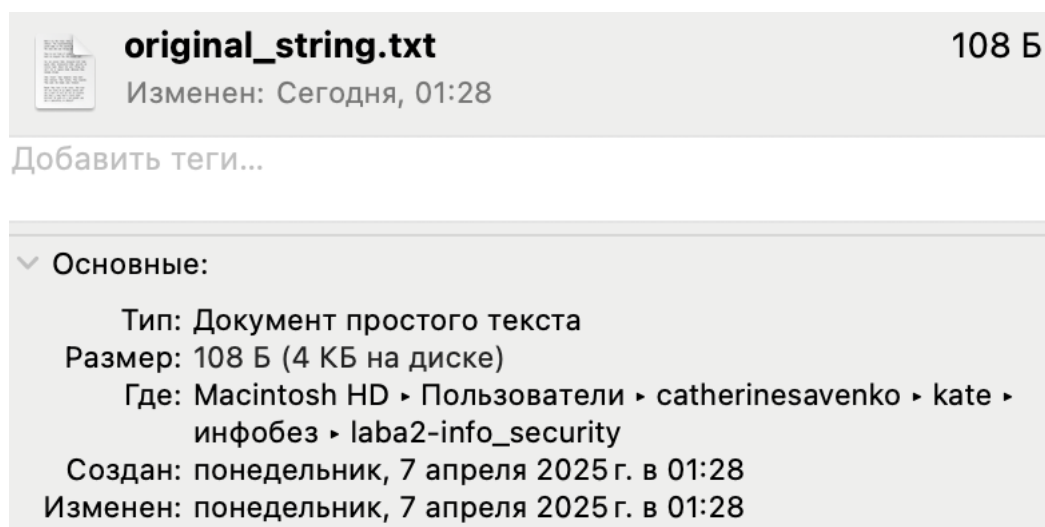


Рисунок 2 - Сохраненный файл на ПК

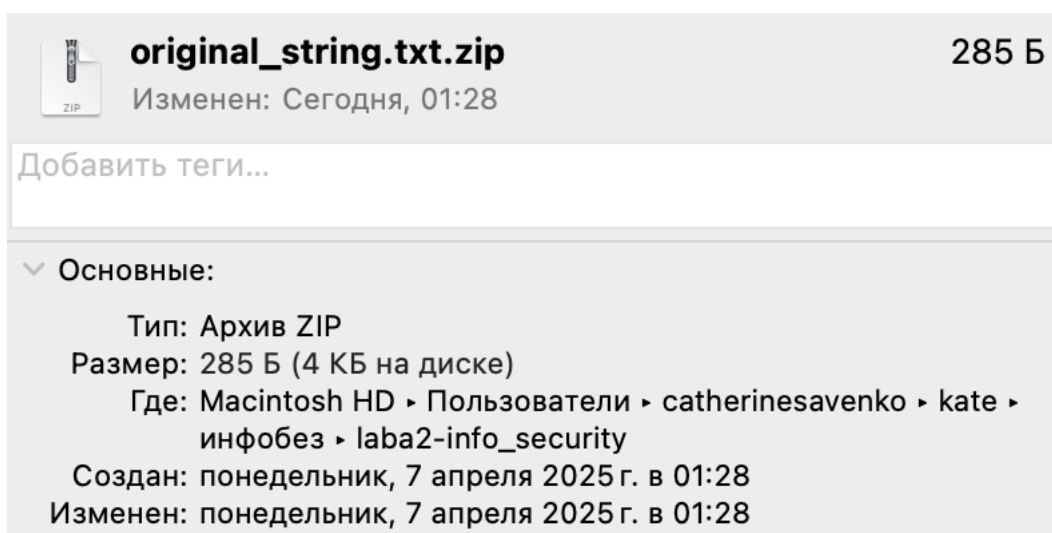


Рисунок 3 - Результат архивации на ПК

Таким образом, имеем коэффициент сжатия $K = \frac{108}{285} = 0,378947$, что связано с добавлением служебной информации после сжатия: метаданные, структура архива.

Листинг: <https://github.com/darcysoul/haffman/>