

СРСП №2 - Кодирование и сжатие информации

1. Кодирование информации алгоритмами Хаффмана и Шеннона-Фано

Исходное сообщение: НУРШАРИПОВ ФАРХАТ ЖАНАРБЕКОВИЧ

Частотный анализ:

Символ	Количество	Вероятность
А	5	$5/30 = 0.167$
Р	4	$4/30 = 0.133$
Пробел	2	$2/30 = 0.067$
В	2	$2/30 = 0.067$
И	2	$2/30 = 0.067$
Н	2	$2/30 = 0.067$
О	2	$2/30 = 0.067$
Б	1	$1/30 = 0.033$
Е	1	$1/30 = 0.033$
Ж	1	$1/30 = 0.033$
К	1	$1/30 = 0.033$
П	1	$1/30 = 0.033$
Т	1	$1/30 = 0.033$
У	1	$1/30 = 0.033$
Ф	1	$1/30 = 0.033$
Х	1	$1/30 = 0.033$
Ч	1	$1/30 = 0.033$
Ш	1	$1/30 = 0.033$
Всего	30	1.000

Построение кода Хаффмана

Этапы построения кодового дерева Хаффмана:

Начальное состояние:

- А: 0.1667 (5/30)
- Р: 0.1333 (4/30)

- Пробел: 0.0667 (2/30)
- В: 0.0667 (2/30)
- И: 0.0667 (2/30)
- Н: 0.0667 (2/30)
- О: 0.0667 (2/30)
- Б: 0.0333 (1/30)
- Е: 0.0333 (1/30)
- Ж: 0.0333 (1/30)
- К: 0.0333 (1/30)
- П: 0.0333 (1/30)
- Т: 0.0333 (1/30)
- У: 0.0333 (1/30)
- Ф: 0.0333 (1/30)
- Х: 0.0333 (1/30)
- Ч: 0.0333 (1/30)
- Ш: 0.0333 (1/30)

Шаг 1: Объединяем Б (0.0333) и Е (0.0333) \rightarrow 0.0667

Шаг 2: Объединяем Ж (0.0333) и К (0.0333) \rightarrow 0.0667

Шаг 3: Объединяем П (0.0333) и Т (0.0333) \rightarrow 0.0667

Шаг 4: Объединяем У (0.0333) и Ф (0.0333) \rightarrow 0.0667

Шаг 5: Объединяем Х (0.0333) и Ч (0.0333) \rightarrow 0.0667

Шаг 6: Объединяем Ш (0.0333) и Пробел (0.0667) \rightarrow 0.1000

Шаг 7: Объединяем Б+Е (0.0667) и В (0.0667) \rightarrow 0.1333

Шаг 8: Объединяем Ж+К (0.0667) и И (0.0667) \rightarrow 0.1333

Шаг 9: Объединяем Н (0.0667) и О (0.0667) \rightarrow 0.1333

Шаг 10: Объединяем П+Т (0.0667) и У+Ф (0.0667) \rightarrow 0.1333

Шаг 11: Объединяем Х+Ч (0.0667) и Ш+Пробел (0.1000) \rightarrow 0.1667

Шаг 12: Объединяем Б+Е+В (0.1333) и Ж+К+И (0.1333) \rightarrow 0.2667

Шаг 13: Объединяем Н+О (0.1333) и П+Т+У+Ф (0.1333) → 0.2667

Шаг 14: Объединяем Р (0.1333) и А (0.1667) → 0.3000

Шаг 15: Объединяем Х+Ч+Ш+Пробел (0.1667) и Б+Е+В+Ж+К+И (0.2667) → 0.4333

Шаг 16: Объединяем Н+О+П+Т+У+Ф (0.2667) и Р+А (0.3000) → 0.5667

Шаг 17: Объединяем всё → 1.0000

Построение кода Шеннона-Фано

Этапы построения кода Шеннона-Фано:

Исходная таблица вероятностей (по убыванию):

- А: 0.1667 (5/30)
- Р: 0.1333 (4/30)
- Пробел: 0.0667 (2/30)
- В: 0.0667 (2/30)
- И: 0.0667 (2/30)
- Н: 0.0667 (2/30)
- О: 0.0667 (2/30)
- Б: 0.0333 (1/30)
- Е: 0.0333 (1/30)
- Ж: 0.0333 (1/30)
- К: 0.0333 (1/30)
- П: 0.0333 (1/30)
- Т: 0.0333 (1/30)
- У: 0.0333 (1/30)
- Ф: 0.0333 (1/30)
- Х: 0.0333 (1/30)
- Ч: 0.0333 (1/30)
- Ш: 0.0333 (1/30)

Глубина 1: Разделение на группы:

- Группа 1 (0): ['А', 'Р', ' ', 'В', 'И', 'Н'] - суммарная вероятность 0.5667

- Группа 2 (1): ['О', 'Б', 'Е', 'Ж', 'К', 'П', 'Т', 'У', 'Ф', 'Х', 'Ч', 'Ш'] - суммарная вероятность 0.4333

Глубина 2: Разделение на группы:

- Группа 1.1 (00): ['А', 'Р'] - суммарная вероятность 0.3000
- Группа 1.2 (01): [' ', 'В', 'И', 'Н'] - суммарная вероятность 0.2667
- Группа 2.1 (10): ['О', 'Б', 'Е', 'Ж', 'К', 'П'] - суммарная вероятность 0.2333
- Группа 2.2 (11): ['Т', 'У', 'Ф', 'Х', 'Ч', 'Ш'] - суммарная вероятность 0.2000

Глубина 3: Разделение на группы:

- Группа 1.2.1 (010): [' ', 'В', 'И'] - суммарная вероятность 0.2000
- Группа 1.2.2 (011): ['Н'] - суммарная вероятность 0.0667 ...и так далее

Результаты кодирования

Таблица вероятностей символов:

Символ	Вероятность
А	0.1667 (5/30)
Р	0.1333 (4/30)
Пробел	0.0667 (2/30)
В	0.0667 (2/30)
И	0.0667 (2/30)
Н	0.0667 (2/30)
О	0.0667 (2/30)
Б	0.0333 (1/30)
Е	0.0333 (1/30)
Ж	0.0333 (1/30)
К	0.0333 (1/30)
П	0.0333 (1/30)
Т	0.0333 (1/30)
У	0.0333 (1/30)
Ф	0.0333 (1/30)
Х	0.0333 (1/30)
Ч	0.0333 (1/30)
Ш	0.0333 (1/30)

Коды Хаффмана:

Символ	Код	Длина
А	111	3
Р	100	3
Пробел	1100	4
В	1101	4
И	0110	4
Н	0111	4
О	1010	4
Б	10110	5
Е	10111	5
Ж	0000	4
К	0001	4
П	00110	5
Т	00111	5
У	01010	5
Ф	01011	5
Х	01000	5
Ч	01001	5
Ш	0010	4

Средняя длина кода Хаффмана: 3.9667 бит/символ

Избыточность кода Хаффмана: 0.0118 (1.18%)

Коды Шеннона-Фано:

Символ	Код	Длина
А	000	3
Р	001	3
Пробел	01000	5
В	01001	5
И	0101	4
Н	011	3
О	10000	5
Б	10001	5
Е	1001	4
Ж	10100	5
К	10101	5
П	1011	4
Т	11000	5
У	11001	5
Ф	1101	4
Х	11100	5
Ч	11101	5
Ш	1111	4

Средняя длина кода Шеннона-Фано: 4.0667 бит/символ

Избыточность кода Шеннона-Фано: 0.0361 (3.61%)

Энтропия источника: 3.9199 бит/символ

2. Сокращение избыточности ИКМ сигнала для дибитов

Таблица вероятностей дибитов:

Дибит	Вероятность
00	0.4600 (46/100)
01	0.2200 (22/100)
10	0.2000 (20/100)
11	0.1200 (12/100)

Построение кода Хаффмана для дибитов

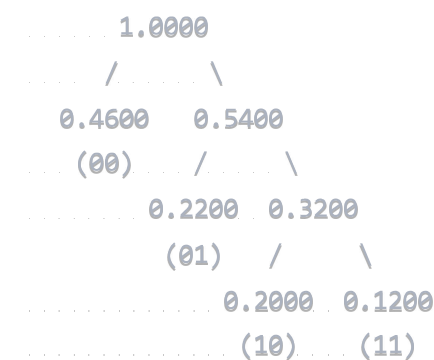
Этапы построения кодового дерева Хаффмана:

Шаг 1: Объединяем 11 (0.1200) и 10 (0.2000) → 0.3200

Шаг 2: Объединяем 01 (0.2200) и (11+10) (0.3200) → 0.5400

Шаг 3: Объединяем 00 (0.4600) и (01+11+10) (0.5400) → 1.0000

Кодовое дерево Хаффмана для дибитов:



Коды Хаффмана для дибитов:

Дибит	Код	Длина
00	0	1
01	10	2
10	111	3
11	110	3

Средняя длина кода Хаффмана: 1.8600 бит/символ

Избыточность кода Хаффмана: 0.0175 (1.75%)

Коды Шеннона-Фано для дибитов:

Дибит	Код	Длина
00	00	2
01	01	2
10	10	2
11	11	2

Средняя длина кода Шеннона-Фано: 2.0000 бит/символ

Избыточность кода Шеннона-Фано: 0.0863 (8.63%)

Энтропия источника: 1.8274 бит/символ

4. Циклическое кодирование дибитовой информации

Исходная дибитовая информация (закодированная алгоритмом Хаффмана):

10,0,0,0,111,0,0,0,110,10,10,110,0,0,111,10,111,0,111,0,111,111,111,0,0,10,0,0,111,110,10,0,0,0



Кодирование циклическим кодом с минимальным кодовым расстоянием $d_0=4$

1. Разбиваем информацию на блоки длиной 11:

- Блок I: 10,0,0,0,111,0,0,0
- Блок II: 110,10,10,110,0
- Блок III: 0,111,10,111,0,1
- Блок IV: 11,0,111,111,11
- Блок V: 1,0,0,10,0,0,111,1
- Блок VI: 10,10,0,0,0,111,0
- Блок VII: 0,0,110,10,10,11
- Блок VIII: 0,0,0,111,10,0,10
- Блок IX: 10,110000000

2. Параметры кода: $n_i=11$; $n_k=4$; $n=15$

3. Выбираем полином для $d_0=4$: $P(x) = (x^4+x+1)(x+1) = x^5+x^4+x^2+1$

4. Представляем блоки в виде многочленов и применяем полиномиальное деление для получения остатка

5. Дополняем каждый блок полученным остатком для формирования кодовых слов

Результаты циклического кодирования:

Блок	Закодированное слово
1	0100011100010010
2	1101010110011101
3	0111101110110100
4	1101111111100000
5	1001000111101010
6	1010000111001110
7	0011010101101010
8	0001111001010011
9	1011000000010000

Общее количество битов до кодирования: 100

Общее количество битов после циклического кодирования: 144

Коэффициент расширения: 1.44

Выводы

1. Сравнение кодов Хаффмана и Шеннона-Фано для текста:

- Код Хаффмана имеет меньшую среднюю длину (3.9667 против 4.0667) и меньшую избыточность (1.18% против 3.61%)
- Оба кода обеспечивают сжатие по сравнению с равномерным кодированием

2. Сравнение кодов Хаффмана и Шеннона-Фано для дибитов:

- Код Хаффмана также показывает лучшие результаты: средняя длина 1.86 против 2.00, избыточность 1.75% против 8.63%
- Код Хаффмана обеспечивает сжатие в ~7% по сравнению с исходной дибитовой последовательностью

3. Циклическое кодирование:

- Обеспечивает возможность обнаружения и исправления ошибок за счет избыточности
- Коэффициент расширения 1.44 означает увеличение объема данных на 44%
- Минимальное кодовое расстояние $d_0=4$ позволяет обнаруживать до 3 ошибок и исправлять 1 ошибку

4. Общие наблюдения:

- Код Хаффмана в обоих случаях показывает лучшую эффективность сжатия
- Циклическое кодирование вносит избыточность, но повышает надежность передачи данных

- Выбор метода кодирования зависит от требований к сжатию и надежности передачи