

## Тема 4.1. Основные понятия теории кодирования.

### Оптимальные коды

**План:** Коды. Алфавитное кодирование. Разделимые коды. Оптимальные коды и их свойства. Алгоритмы Фано и Хаффмана построения оптимальных кодов.

### Задачи с решением

**Пример 1:** Закодировать по Фано сообщения, имеющие следующие вероятности:

символ	1	2	3	4	5	6	7
вероятность	0,4	0,2	0,1	0,1	0,1	0,05	0,05

Проверим выполнимость необходимого условия:

$$0,4 + 0,2 + 0,1 + 0,1 + 0,1 + 0,1 + 0,05 + 0,05 = 1.$$

Расположим элементы в порядке убывания вероятностей. Затем будем последовательно делить, не меняя порядка, все элементы на две группы, максимально близкие по суммарной вероятности (т.е. модуль разности сумм вероятностей первой и второй группы должен быть минимальных из всех возможных разбиений на группы). Для «верхней» группы будем ставить значение 0, «нижней» - 1:

Символ	Вероятность	Шаг 1	Шаг 2	Шаг 3	Шаг 4	Полученный код
1	0,4	0				0
2	0,2	1	0	0		100
3	0,1			1		101
4	0,1		1	0	0	1100
5	0,1				1	1101
6	0,05			1	0	1110
7	0,05				1	1111

Найдем стоимость кода (средняя длина кодового слова). Он является критерием степени оптимальности кодирования. Вычислим ее в нашем случае.

$$l = \sum_{i=1}^7 l_i \cdot p_i = 1 \cdot 0,4 + 3 \cdot 0,2 + 3 \cdot 0,1 + 4 \cdot (0,1 \cdot 2 + 0,05 \cdot 2) = 2,5.$$

**Пример 2:** Закодировать по Хаффману сообщения, имеющие следующие вероятности:

символ	1	2	3	4	5	6	7
вероятность	0,4	0,2	0,1	0,1	0,1	0,05	0,05

Решение.

сообщения	p	p <sub>1</sub>	p <sub>2</sub>	p <sub>3</sub>	p <sub>4</sub>	p <sub>5</sub>
1	0,4	0,4	0,4	0,4	0,4	0,6
2	0,2	0,2	0,2	0,2	0,4	0,4
3	0,1	0,1	0,2	0,2	0,2	
4	0,1	0,1	0,1	0,2		
5	0,1	0,1	0,1			
6	0,05	0,1				
7	0,05					

Вторым шагом производим кодирование, «проходя» по таблице справа налево (обычно это продлевается в одной таблице):

	A	A <sub>1</sub>	A <sub>2</sub>	A <sub>3</sub>	A <sub>4</sub>	A <sub>5</sub>
1	0,4	0,4	0,4	0,4	0,4	0,6 0
2	0,2	0,2	0,2	0,2	0,4 00	0,4 1
3	0,1	0,1	0,2	0,2 000	0,2 01	
4	0,1	0,1	0,1 0010	0,2 001		
5	0,1	0,1 0000	0,1 0011			
6	0,05 00010	0,1 0001				
7	0,05 00011					

Найдем стоимость кода (средняя длина кодового слова). Он является критерием степени оптимальности кодирования.

$$l = \sum_{i=1}^7 l_i \cdot p_i = 1 \cdot 0,4 + 2 \cdot 0,2 + 4 \cdot (0,1 \cdot 3) + 5 \cdot (0,05 \cdot 2) = 2,5.$$

### Задачи для самостоятельного решения

1. Построить код Фано и Хаффмана для списка сообщений с заданным распределением частот. Определить стоимость кода. (\* - правильная вероятность символа, т.е. такая вероятность, что сумма всех вероятностей равна 1):

1.1.

S	T	U	V	W	X	Y	Z
0,10	0,02	0,22	0,15	*	0,15	0,1	0,1

1.2.

S	T	U	V	W	X	Y	Z
0,15	0,12	0,22	0,15	*	0,05	0,1	0,1

1.3.

S	T	U	V	W	X	Y	Z
0,11	0,12	0,14	0,15	*	0,15	0,1	0,1

1.4.

S	T	U	V	W	X	Y	Z
0,15	0,06	0,21	0,05	*	0,15	0,1	0,1

1.5.

S	T	U	V	W	X	Y	Z
0,15	0,12	0,21	0,25	*	0,05	0,01	0,1

1.6.

S	T	U	V	W	X	Y	Z
0,15	0,02	0,02	0,25	*	0,15	0,01	0,1

1.7.

S	T	U	V	W	X	Y	Z
0,11	0,02	0,12	0,11	*	0,05	0,1	0,1

1.8.

S	T	U	V	W	X	Y	Z
0,13	0,12	0,02	0,15	*	0,05	0,1	0,1

1.9.

S	T	U	V	W	X	Y	Z
0,15	0,02	0,23	0,15	*	0,15	0,01	0,1

1.10.

S	T	U	V	W	X	Y	Z
0,15	0,02	0,2	0,15	*	0,05	0,1	0,08