

## Работа №1

### Алгоритм оптимального кодирования.

#### Метод Шеннона-Фано

Если вероятности появления букв алфавита различные, то равномерный код (соответствие одного символа одной букве) не является оптимальным, т.к. энтропия равновесного, хаотического состояния (с одинаковой вероятностью элементов  $P = P_i = 1/N$ ) всегда больше энтропии неравновесного, упорядоченного (через некоторые условия) состояния. Поэтому применяются неравномерные коды, в которых буквы, имеющие наибольшую вероятность, кодируются наиболее короткими кодовыми последовательностями, а более длинные комбинации присваиваются редким буквам.

Воспользуемся методом Шеннона-Фано. Буквы (или любые сигналы, сообщения, подлежащие кодированию) исходного алфавита записываются в порядке убывающей вероятности. Упорядоченное таким образом множество букв разбивается на два подмножества так, чтобы суммарные вероятности этих подмножеств были примерно равны. Затем каждое подмножество снова разбивается на два подмножества с соблюдением того же условия равенства вероятностей. Такое разбиение продолжается до тех пор, пока в подмножествах не окажется только по одной букве кодируемого алфавита. При каждом разбиении буквам верхнего подмножества присваивается кодовый элемент 1, а буквам нижнего подмножества - 0 (подумайте, что будет, если поменять местами 1 и 0).

Полученный таким способом неравномерный код не является однозначно декодируемым (расшифруемым). Чтобы снять эту проблему можно использовать разделительный символ между буквами алфавита, или, запретить такие кодовые комбинации, начальные части которых уже использованы в качестве самостоятельной комбинации. Например, если 101 означает код какой-то буквы, то нельзя использовать комбинации 1, 10 или 10101.

#### Пример

Дана последовательность букв (сигналов)  $x_i$  ( $i = 1, \dots, 8$ ) с убывающей вероятностью их появления  $P_i$  (см. табл.1). По вышеописанному алгоритму определены коды с длиной  $m_i$ .

Таблица 1.

Буква $x_i$	Вероятность $P_i$	Код	Длина $m_i$
$x_1$	0,25	11	2
$x_2$	0,25	10	2
$x_3$	0,15	011	3
$x_4$	0,15	010	3
$x_5$	0,05	0011	4
$x_6$	0,05	0010	4
$x_7$	0,05	0001	4
$x_8$	0,05	0000	4

## Задания

1 Построить оптимальное кодирование для последовательности из 9 членов ( $i=1, \dots, 9$ ,  $x_9 = 0,03$ )

2 Найти коды букв русского алфавита:

Буквы	А	Б	В	Г	Д	Е	Ж	З	И	Й	К	Л	М	Н	О	П
Вероятность	0,062	0,014	0,038	0,013	0,025	0,072	0,007	0,016	0,062	0,010	0,028	0,035	0,026	0,053	0,090	0,023
Буквы	Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Э	Ю	Я	пробел
Вероятность	0,040	0,045	0,053	0,021	0,002	0,009	0,004	0,012	0,006	0,003	0,014	0,016	0,003	0,006	0,018	0,174