

## Практическое занятие № 2

### Изучение методов сжатия (экономного кодирования) цифровых потоков без потерь.

Построить двоичный код Хаффмана и вычислить его скорость для представления фрагмента текста:

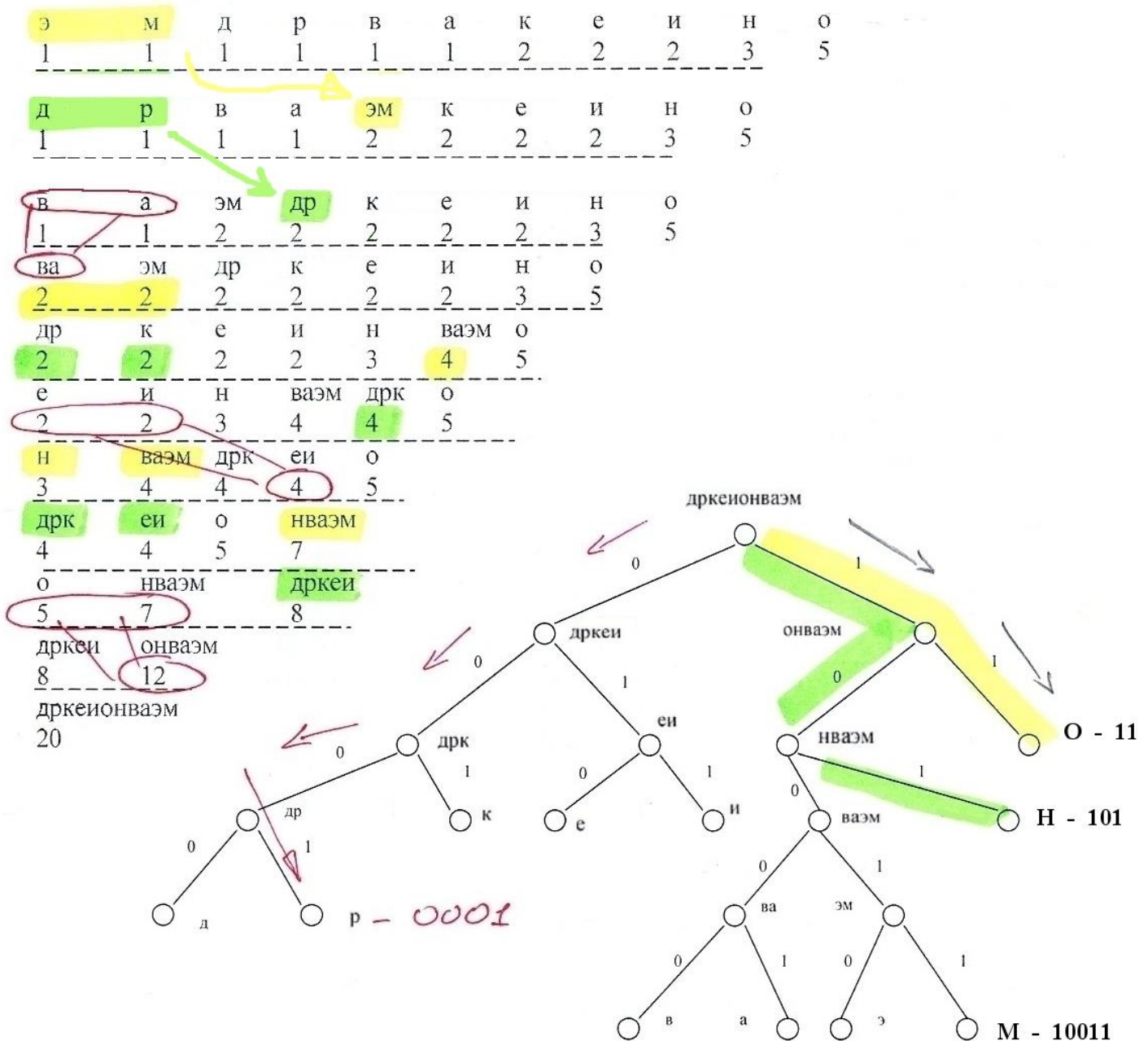
**"экономное кодирование"**

РЕШЕНИЕ:

1. Вычисление вероятностей (частот) появления символов.

э - 1, к - 2, о - 5, н - 3, м - 1, е - 2, д - 1, и - 2, р - 1, в - 1, а - 1.

2. Записываем символы в порядке возрастания частот. Последовательно объединяются два символа с наименьшими вероятностями появления в новый составной символ, вероятность появления которого полагается равной сумме вероятностей составляющих его символов. В конце концов получается дерево, каждый узел которого, имеет суммарную вероятность всех узлов, находящихся ниже него:



### Примечание

Листовые узлы дерева Хаффмана соответствуют символам кодируемого алфавита. Глубина листовых узлов равна длине кода соответствующих символов.

Путь от корня дерева к листовому узлу можно представить в виде битовой строки, в которой "0" соответствует выбору левого поддерева, а "1" - правого. Используя этот механизм, мы без труда можем присвоить коды всем символам кодируемого алфавита. Выпишем, к примеру, коды для всех символов в нашем примере:

Полученный код: о – 11; н – 101; и – 011; е – 010; к – 001; д – 0000; р – 0001; в – 10000; а – 10001; э – 10010; м – 10011.

Длина кода сообщения:  $L = 5 \cdot 2 + 3 \cdot 3 + 2 \cdot 3 + 2 \cdot 3 + 1 \cdot 4 + 1 \cdot 4 + 1 \cdot 5 + 1 \cdot 5 + 1 \cdot 5 + 1 \cdot 5 = 65$  дв.симв.

При кодировании равномерным кодом для представления каждого из 11 символов потребовалось бы 4 дв.симв. Для длины сообщения 20 символов потребовалось бы  $L_1 = 20 \cdot 4 = 80$  дв.симв.

Скорость кода:  $R = L/L_1 = 65/80 = 13/16$ .

$$R = 1/K_{сж} = L/L_1$$

### Примечание

Наряду с коэффициентом сжатия ( $K_{сж}$ ) эффективность сжатия может быть охарактеризована скоростью сжатия ( $R$ ), определяемой как отношение

$$R = \frac{1}{K_{сж}}, \text{ где } K_{сж} = \frac{W_{исх}}{W_{сж}}$$

и измеряемой в "количестве кодовых бит, приходящихся на отсчет данных источника". Система, имеющая большой коэффициент сжатия, обеспечивает меньшую скорость сжатого потока.