

Тема урока:

«Алгоритмы сжатия текстовой информации»

Учитель информатики МОУ школа №8

Зайцев А. И.

г. о. Жуковский, 2013

Сжатие - кодирование информации с целью уменьшения ее объема.

Коэффициент сжатия - отношение исходного объема информации к полученному объему в результате сжатия:

$$K_c = \frac{V_0}{V_1}$$

V_0 – исходный объем информации

V_1 – объем сжатой информации

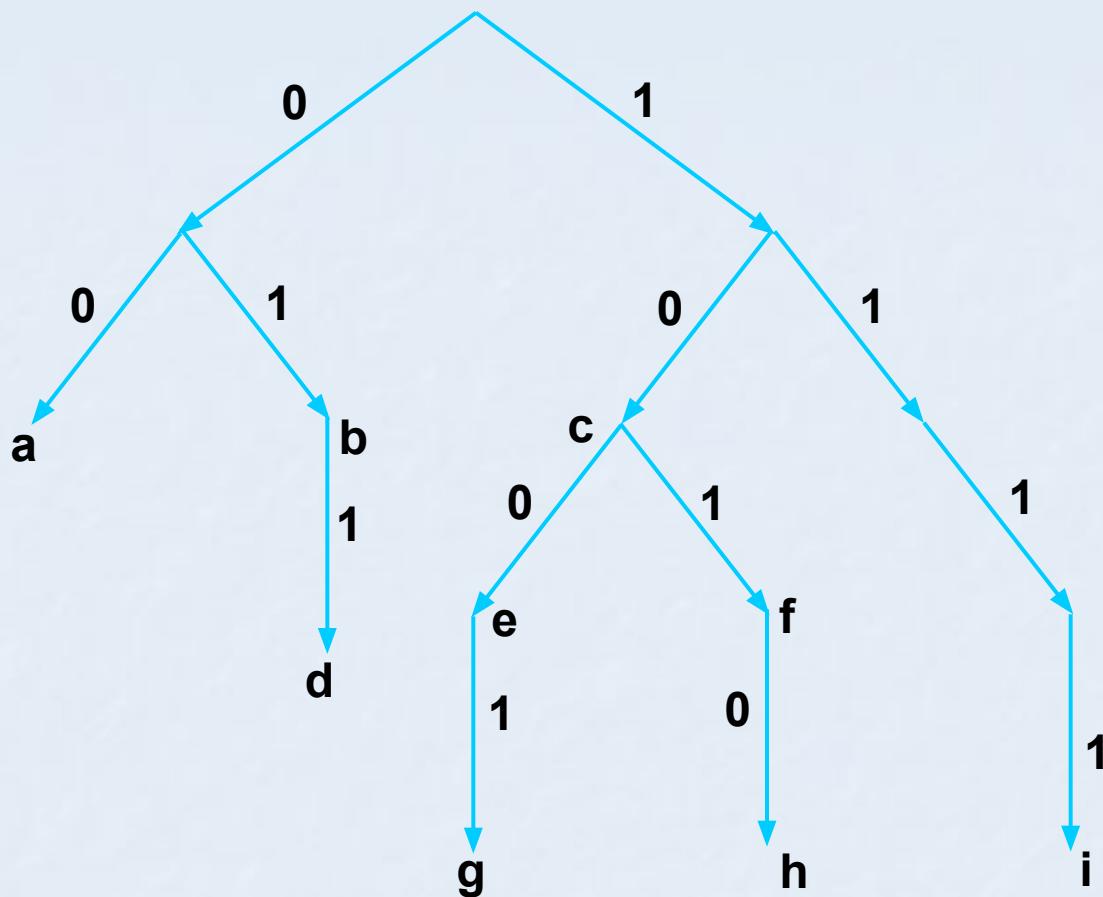
Условие Шеннона-Фано

Никакое кодовое слово не является началом другого кодового слова.

Код, удовлетворяющий условию Шеннона-Фано, называется префиксным кодом.

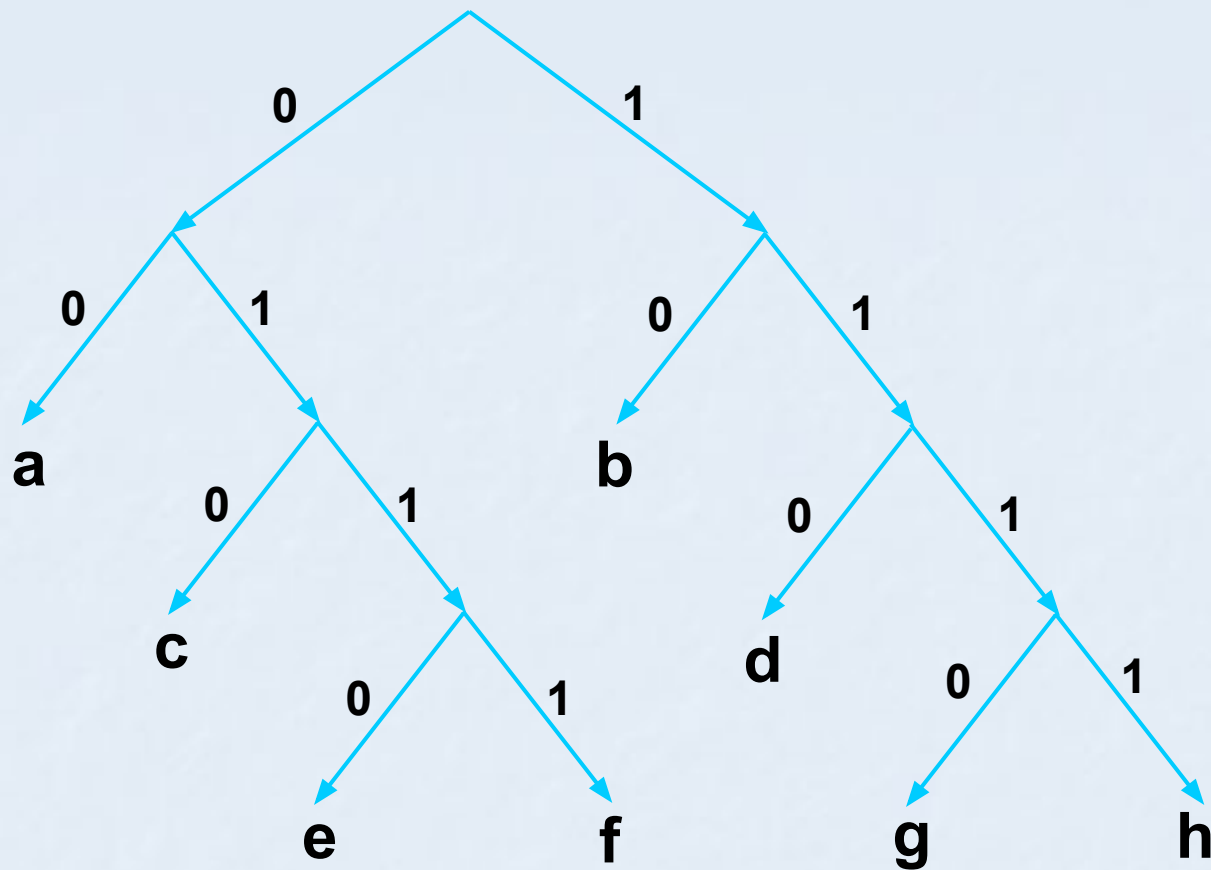
Каждому набору кодовых слов можно сопоставить ориентированный граф, определяющий этот код.

a	00
b	01
c	10
d	011
e	100
f	101
g	1001
h	1010
i	1111



Данный код не является префиксным

a	00
b	10
c	010
d	110
e	0110
f	0111
g	1110
h	1111



Данный код – префиксный, т.к. кодируемые символы располагаются в вершинах, из которых не выходят новые дуги.

Алгоритм Хаффмана построения префиксного кода

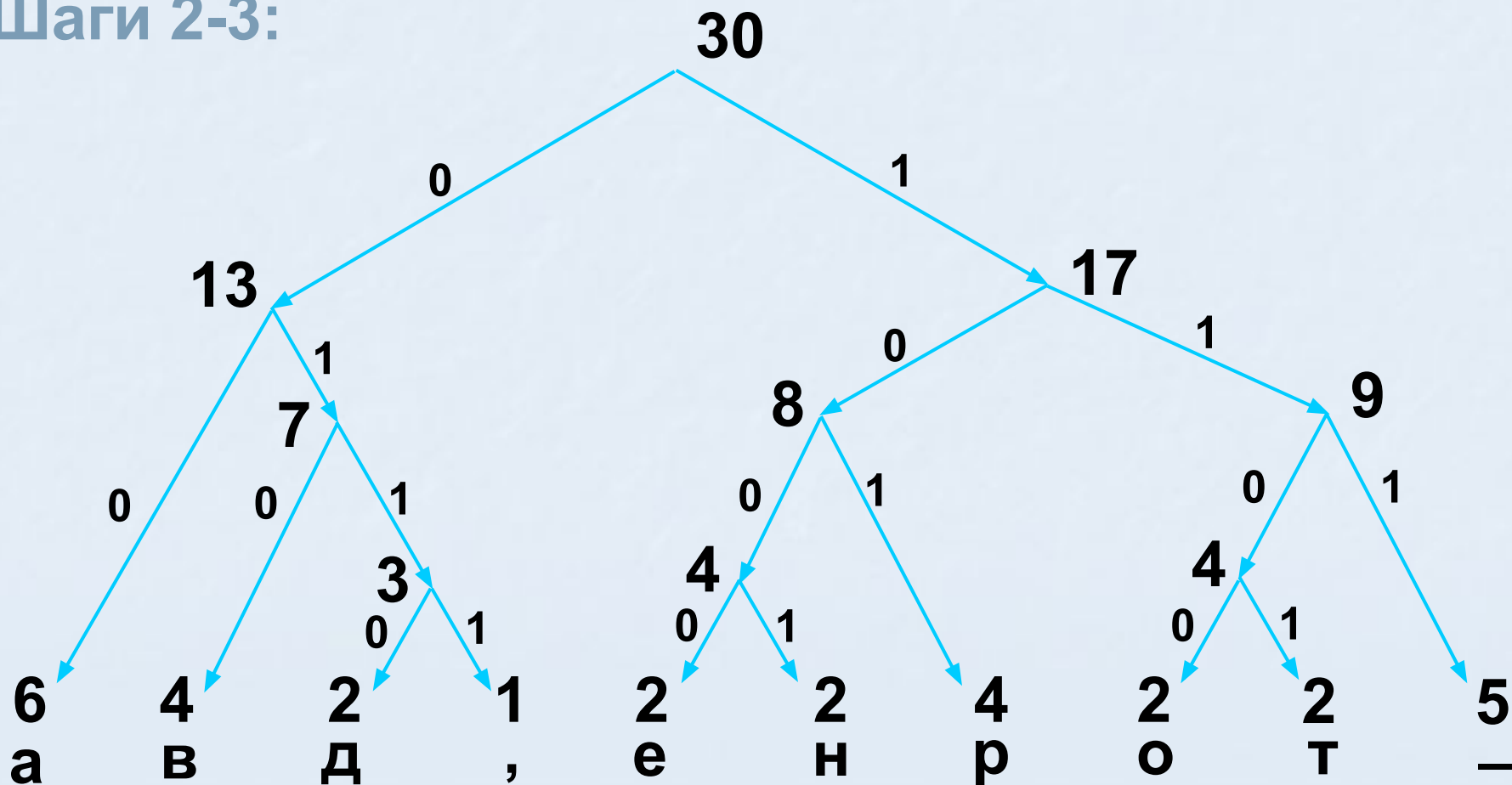
1. Все символы кодируемой информации образуют вершины-листья. Каждой вершине приписывается вес, равный количеству вхождений данного символа в сообщение.
2. Среди вершин, которым приписаны веса, выбираются две с наименьшими весами (если таких несколько, любые из них).
3. Создается следующая вершина графа, из которой выходят две дуги к выбранным на предыдущем шаге вершинам; одна дуга помечается символом 0, другая – символом 1. Созданной вершине приписывается вес, равный сумме весов выбранных вершин, а веса этих двух вершин стираются.
4. К вершинам, которым приписаны веса, применяются шаги 2 и 3 до тех пор, пока не останется одна вершина с весом, равным сумме весов исходных символов.

Пример. Построить код Хаффмана для фразы:
на дворе трава, на траве дрова

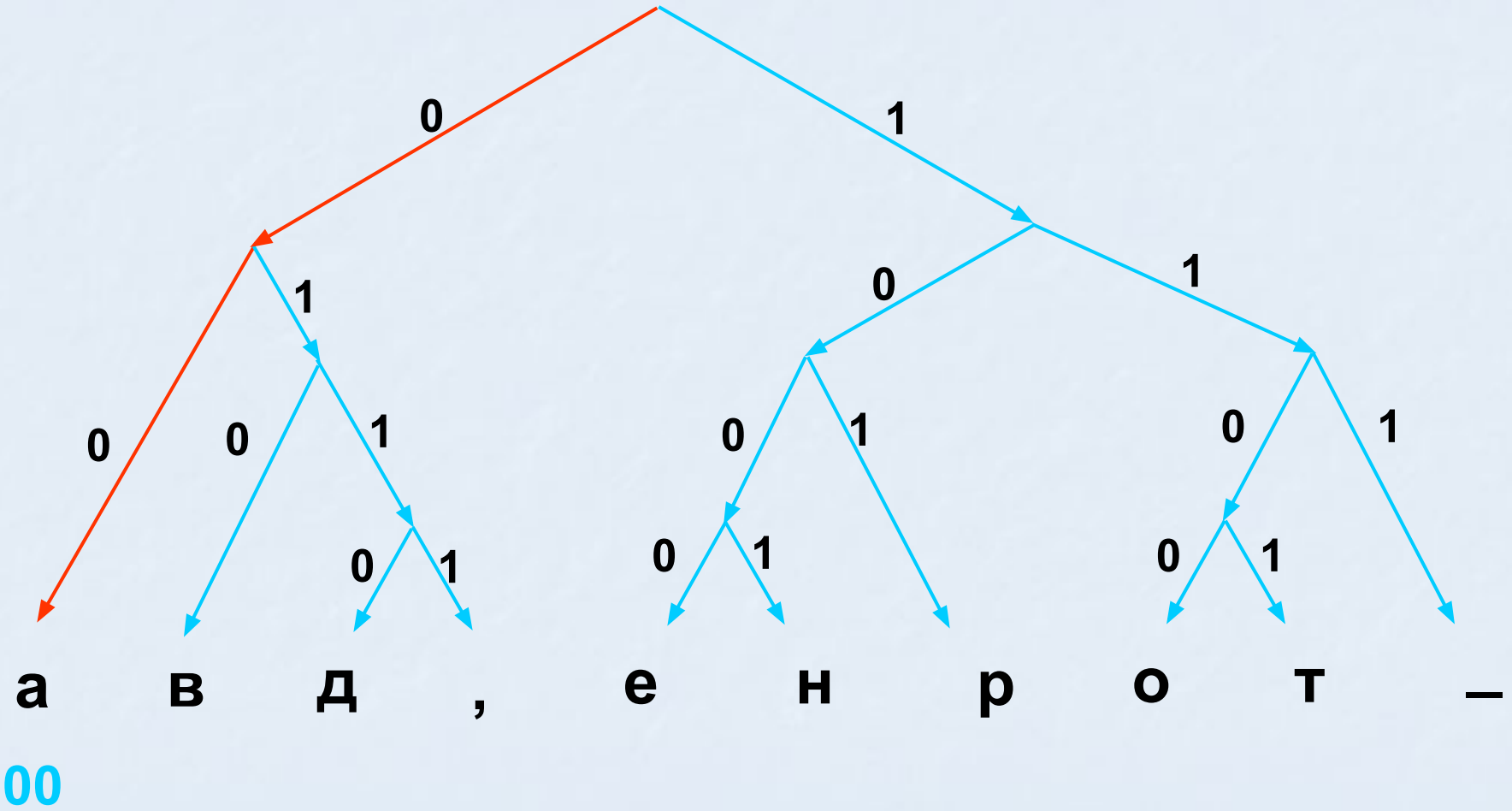
Шаг 1:

а	в	д	,	е	н	р	о	т	—
6	4	2	1	2	2	4	2	2	5

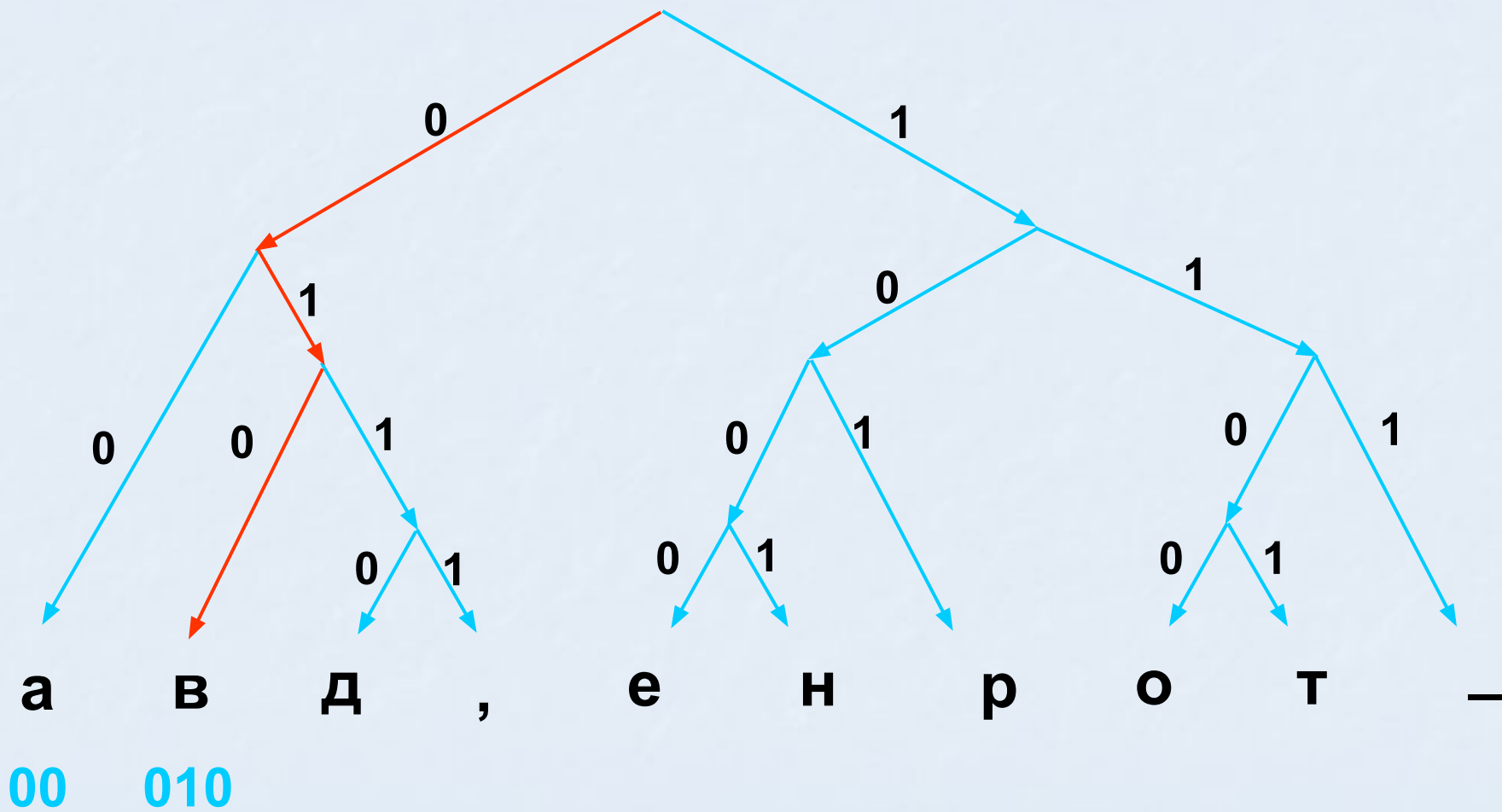
Шаги 2-3:



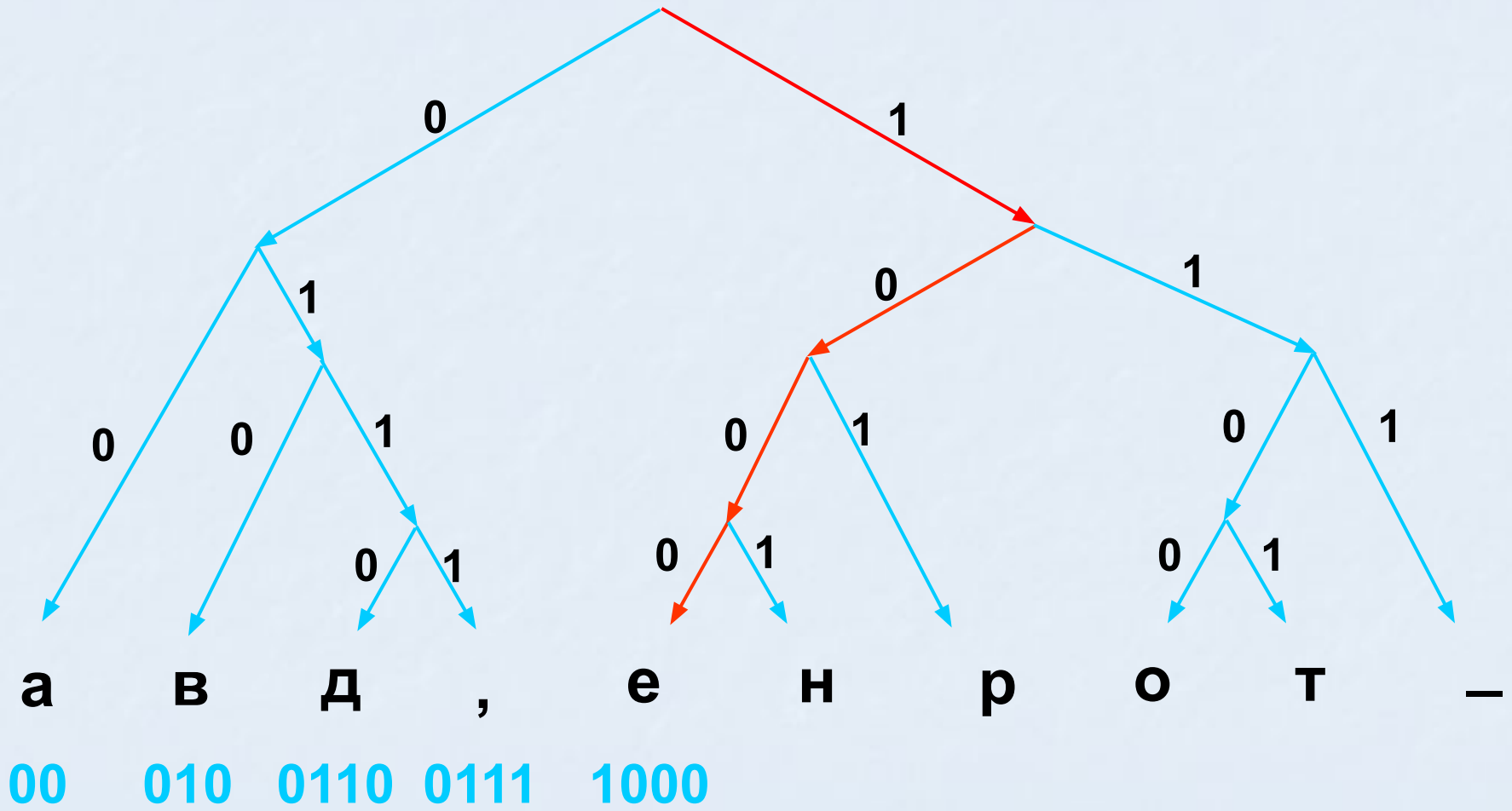
Шаг 4:



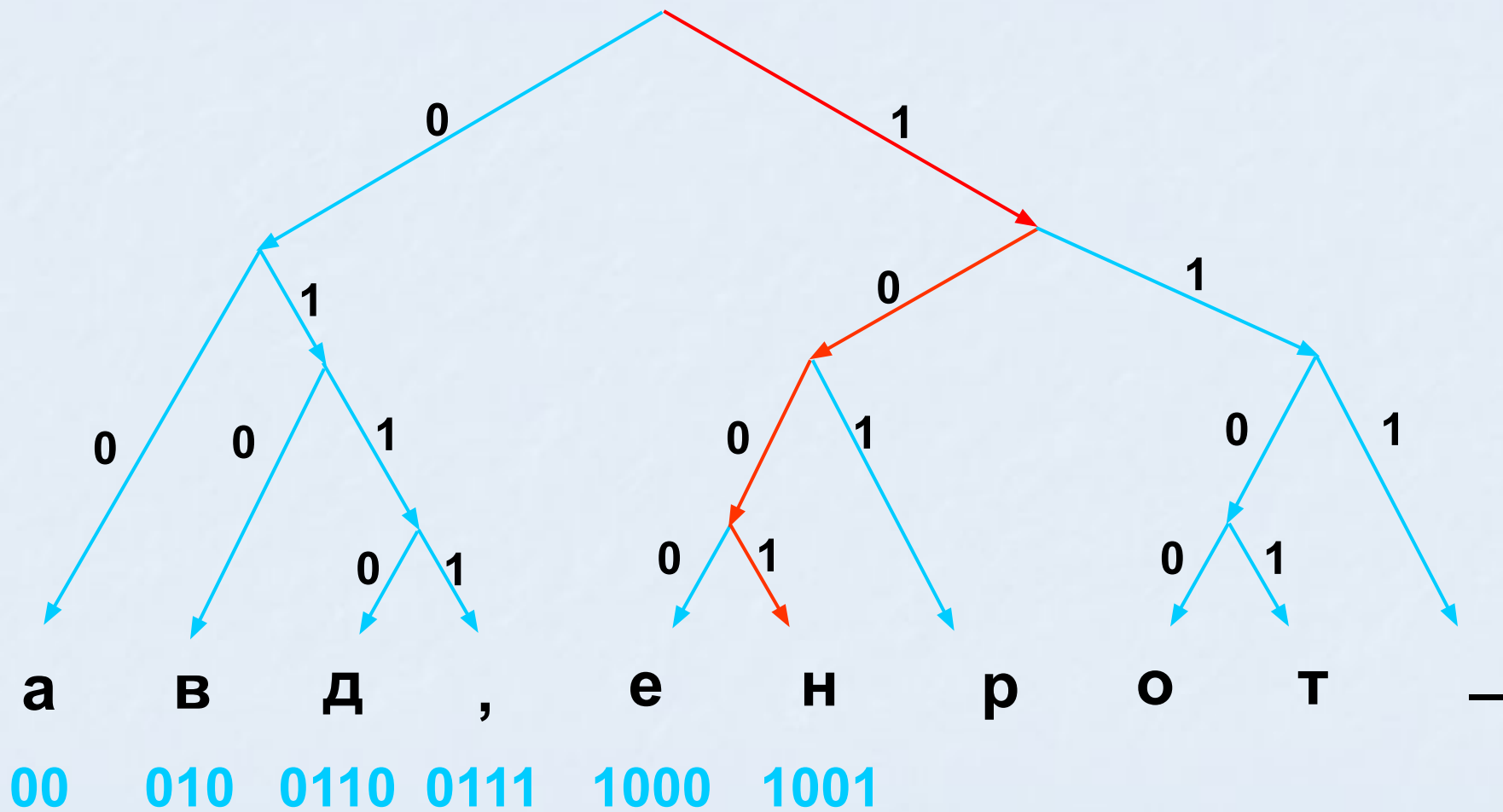
Шаг 4:



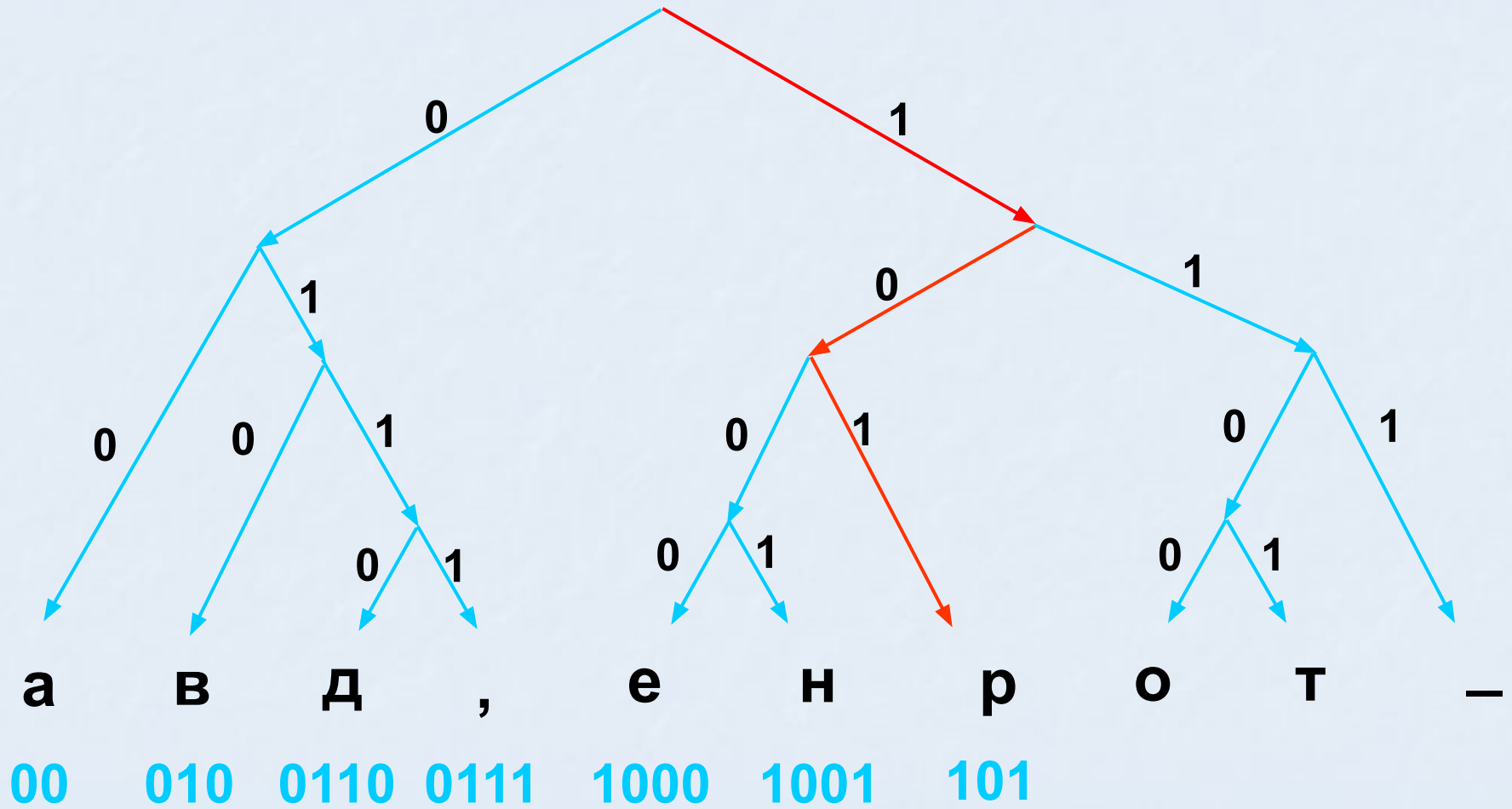
Шаг 4:



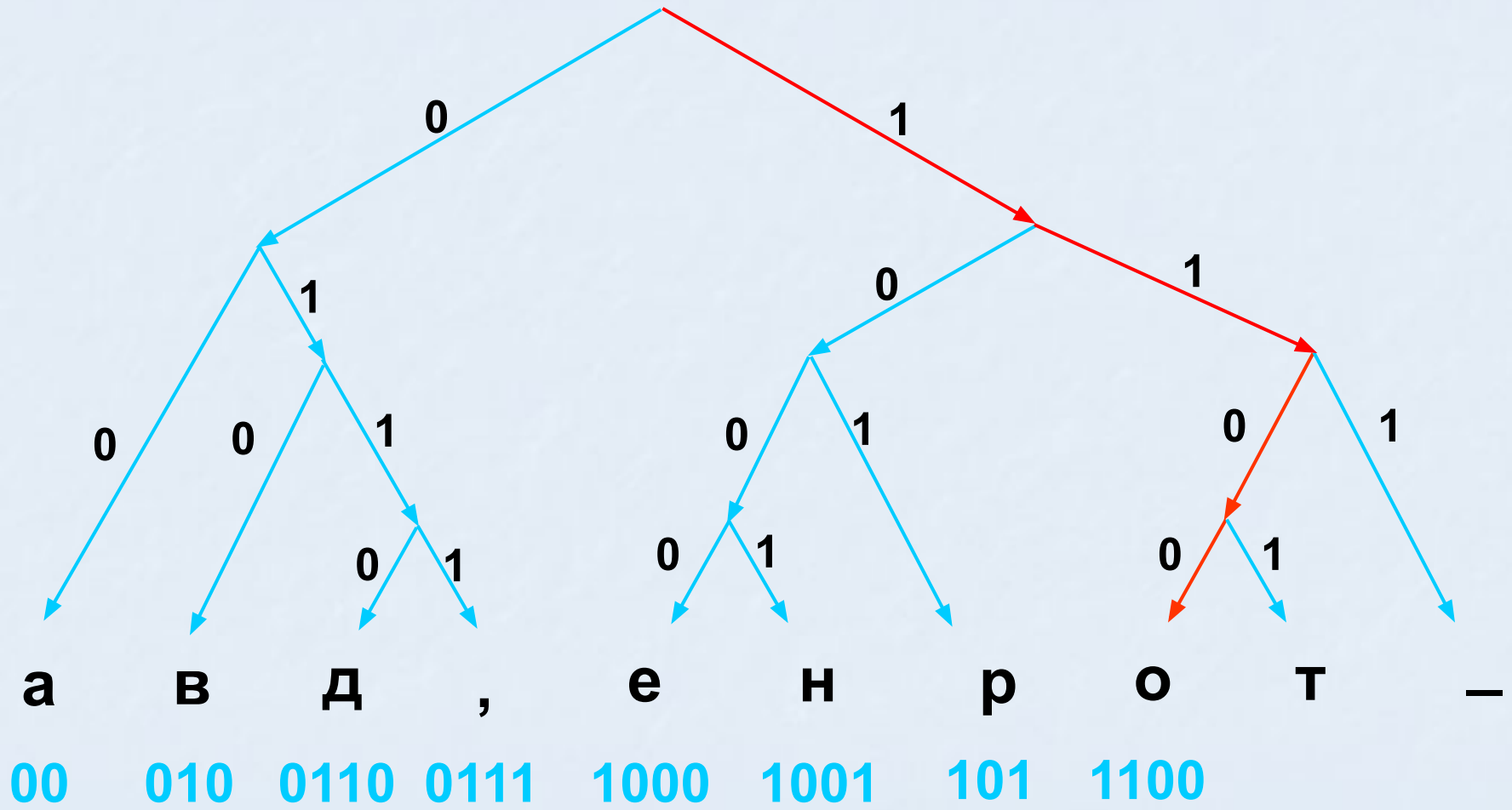
Шаг 4:



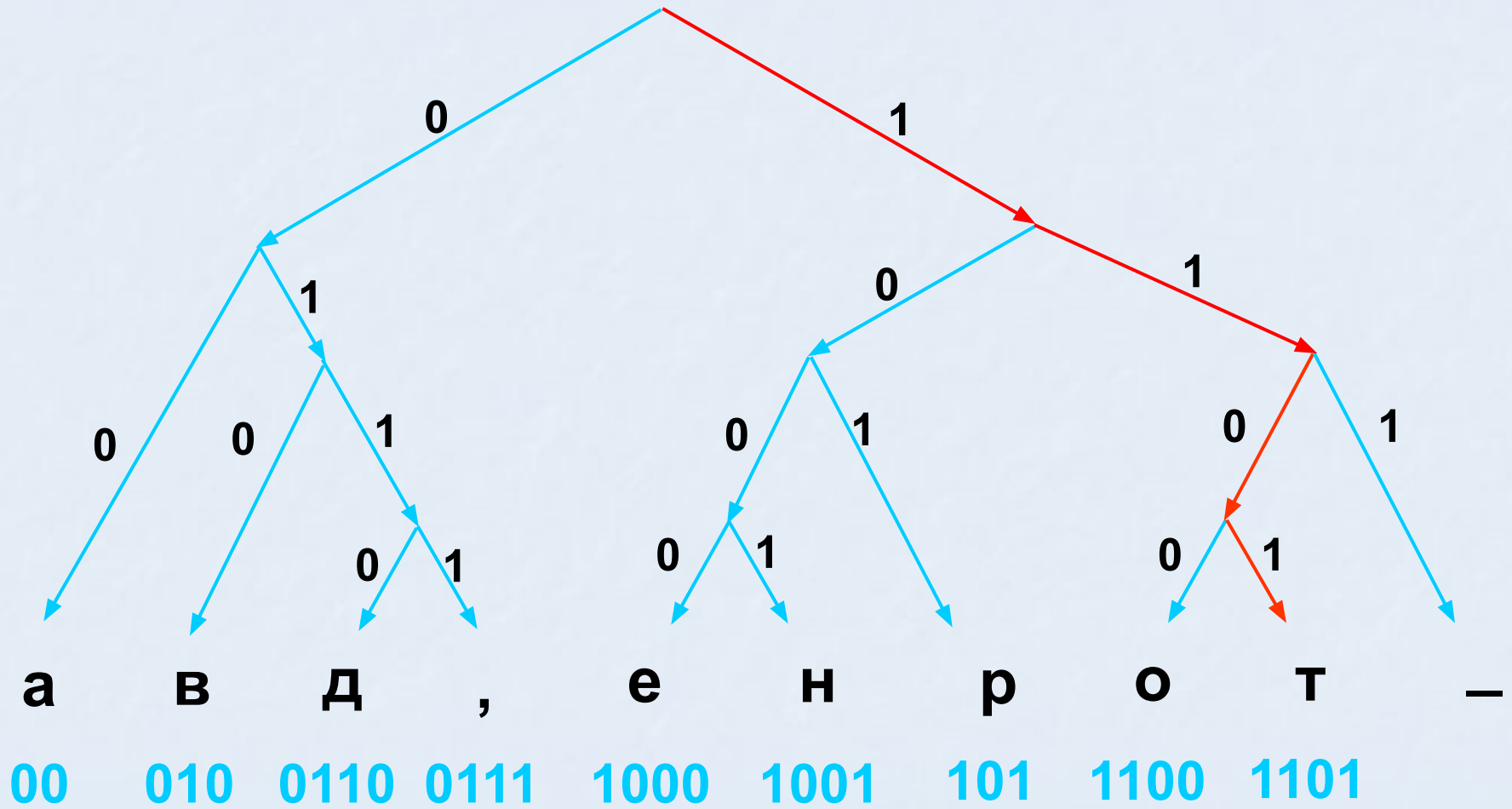
Шаг 4:



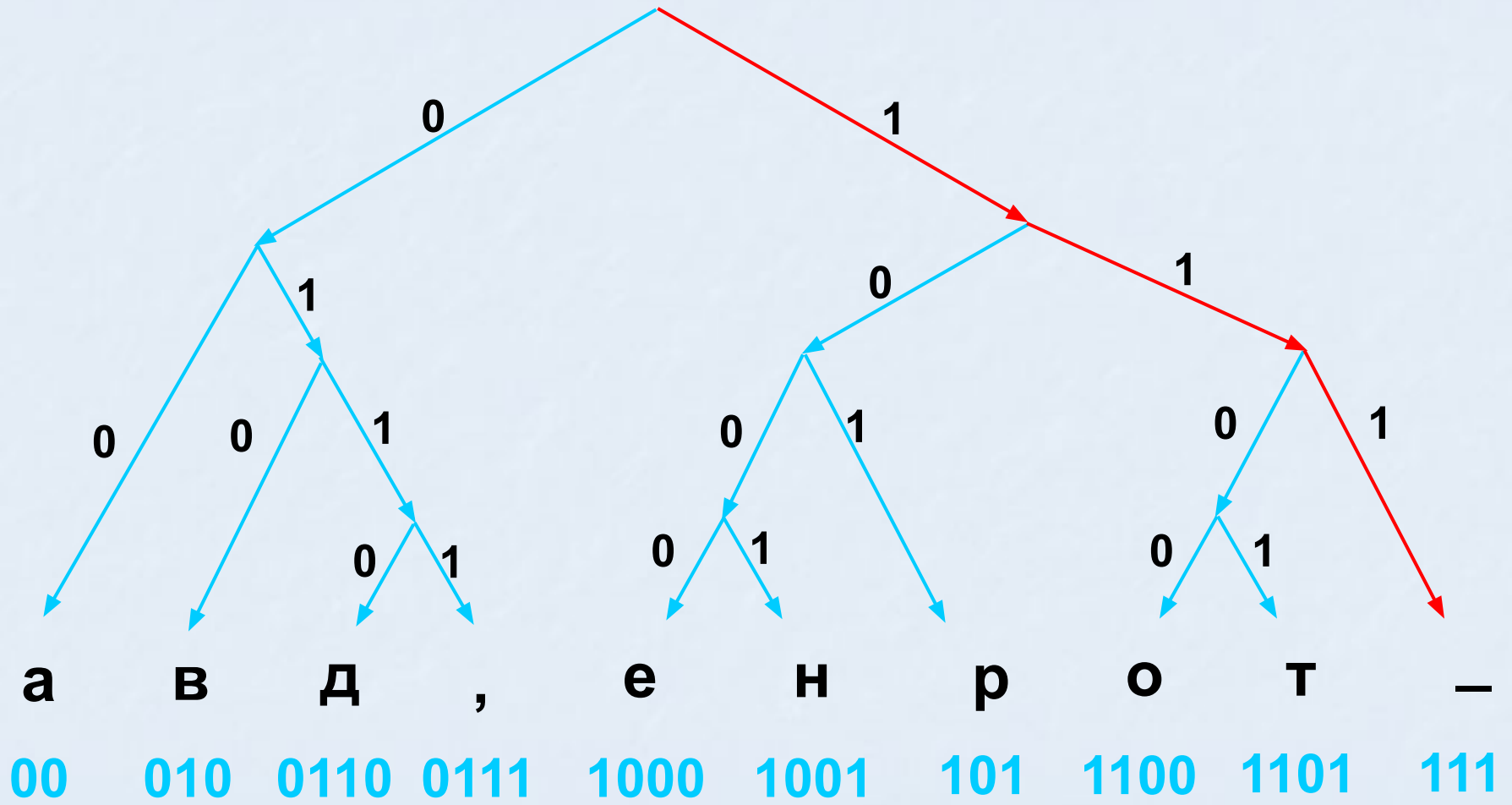
Шаг 4:



Шаг 4:



Шаг 4:



Вопросы

1. За счет чего достигается эффект сжатия данных при их упаковке?
2. Какой код называется префиксным?

Задание

а) Постройте код Хаффмана для фразы:

КАРЛ У КЛАРЫ УКРАЛ КОРАЛЛЫ, А КЛАРА У КАРЛА
УКРАЛА КЛАРНЕТ

б) Определите коэффициент сжатия для данной фразы, если каждый символ кодируется кодом ASCII и равномерным кодом.