

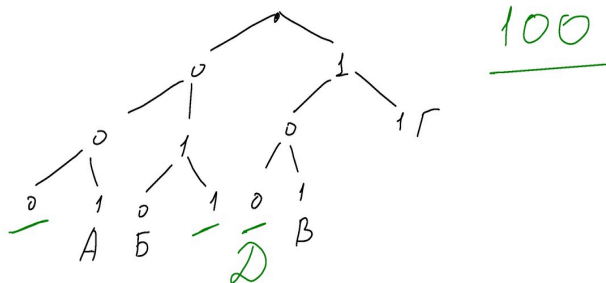
Условие 1

(А. Рогов) По каналу связи передаются сообщения, содержащие только пять букв: А, Б, В, Г, Д. Для передачи используется неравномерный двоичный код, удовлетворяющий условию Фано. Для букв А, Б, В и Г используются кодовые слова 001, 010, 101, 11 соответственно. Укажите кратчайшее кодовое слово для буквы Д, при котором код будет допускать однозначное декодирование. Если таких кодов несколько, укажите код с наибольшим числовым значением.

Решение

Нарисуем двоичное дерево, на котором отмечены коды представленных букв. При этом важно сказать, что при построении дерева делать раздваивание ветвей лишний раз не стоит, так как это сделает схему менее читаемой.

(№ 6699) (А. Рогов) По каналу связи передаются сообщения, содержащие только пять букв: А, Б, В, Г, Д. Для передачи используется неравномерный двоичный код, удовлетворяющий условию Фано. Для букв А, Б, В и Г используются кодовые слова 001, 010, 101, 11 соответственно. Укажите кратчайшее кодовое слово для буквы Д, при котором код будет допускать однозначное декодирование. Если таких кодов несколько, укажите код с наибольшим числовым значением.



Расписав местоположение букв с известными кодами – т.е. А, Б, В, Г, зададимся вопросом, куда разместить букву Д. Прежде всего этот код должен быть самым

коротким из всех. В данной ситуации все оставшиеся кодовые слова одинаковой длины, поэтому, теоретически, Д может быть присвоен любой из этих кодов. Однако в условии сказано, что код для Д должен иметь наиболее числовое значение. Выберем таковой из 000, 011, 100 – код 100 имеет наиболее числовое значение. Значит, он и будет ответом к задаче.

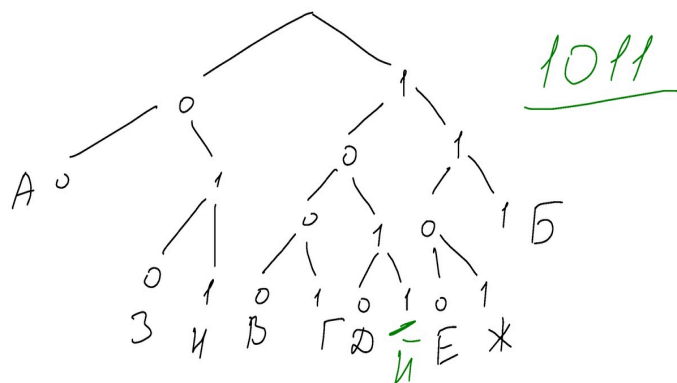
Условие 2

Для кодирования некоторой последовательности, состоящей из букв А, Б, В, Г, Д, Е, Ж, З, И, Й. решили использовать неравномерный двоичный код, удовлетворяющий условию Фано. Для букв А, Б, В, Г, Д, Е, Ж, З, И использовали соответственно кодовые слова 00, 111, 1000, 1001, 1010, 1100, 1101, 010, 011. Укажите кратчайшее возможное кодовое слово для буквы Й, при котором код будет допускать однозначное декодирование. Если таких кодов несколько, укажите код с наименьшим числовым значением.

Решение

Составляем двоичное дерево, вписывая в него коды известных букв.

(№ 1701) Для кодирования некоторой последовательности, состоящей из букв А, Б, В, Г, Д, Е, Ж, З, И, Й. решили использовать неравномерный двоичный код, удовлетворяющий условию Фано. Для букв А, Б, В, Г, Д, Е, Ж, З, И использовали соответственно кодовые слова 00, 111, 1000, 1001, 1010, 1100, 1101, 010, 011. Укажите кратчайшее возможное кодовое слово для буквы Й, при котором код будет допускать однозначное декодирование. Если таких кодов несколько, укажите код с наименьшим числовым значением.



После этого замечаем, что свободным остается только один код – 1011. Значит, именно он будет соответствовать Й.

Условие 3

(М. Шагитов) По каналу связи передаются сообщения, содержащие только восемь букв: А, Б, В, Г, Д, Е, Ж, З. Для передачи используется неравномерный двоичный код, удовлетворяющий условию Фано. Для некоторых букв кодовые слова известны: В – 00, Г – 1000, Д – 111, Е – 1001, Ж – 01, З – 110. Какое наименьшее количество двоичных знаков потребуется для кодирования двух оставшихся букв? В ответе запишите суммарную длину кодовых слов для букв А и Б.

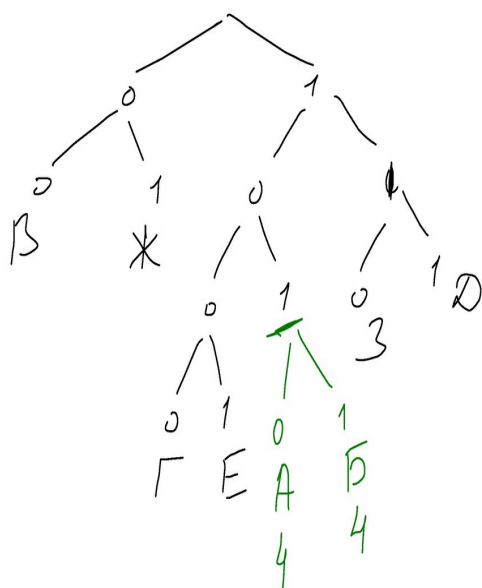
Решение

Нарисуем двоичное дерево, на котором отмечены коды представленных букв.

(№ 6701) (М. Шагитов) По каналу связи передаются сообщения, содержащие только восемь букв: А, Б, В, Г, Д, Е, Ж, З. Для передачи используется неравномерный двоичный код, удовлетворяющий условию Фано. Для некоторых букв кодовые слова известны: В – 00, Г – 1000, Д – 111, Е – 1001, Ж – 01, З – 110. Какое наименьшее количество двоичных знаков потребуется для кодирования двух оставшихся букв? В ответе запишите суммарную длину кодовых слов для букв А и Б.

А: 1010
Б: 1011

$$4+4=8$$



Расписав местоположение букв с известными кодами – т.е. В, Ж, Г, Е, З, Д зададимся вопросом, куда разместить буквы А и Б, т.к. осталось только одно свободное слово, в то время как элементов – два. В таком случае нам снова необходимо сделать «разветвления» свободной ветки на две, чтобы в нее можно было вписать две буквы. Таким образом, сумма длин кодовых слов для А (1010) и для Б (1011) будет равна 8.

Условие 4

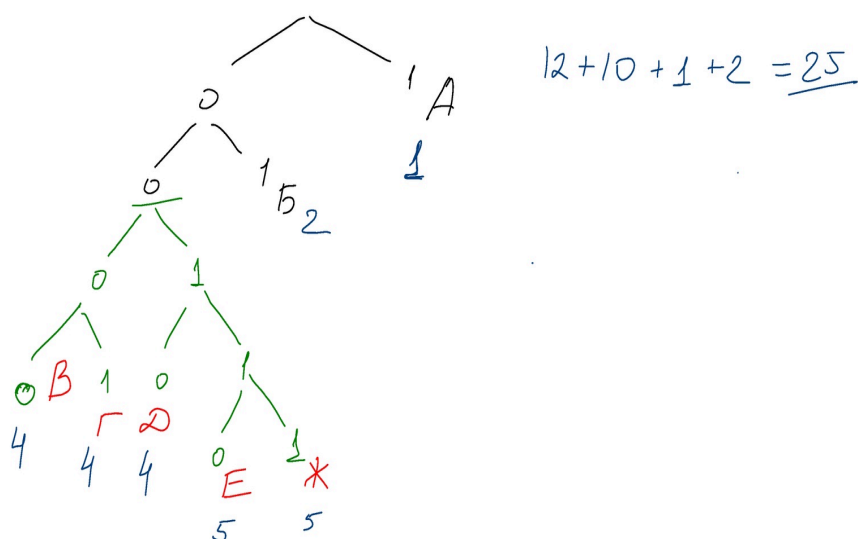
По каналу связи передаются сообщения, содержащие только семь букв: А, Б, В, Г, Д, Е и Ж. Для передачи используется двоичный код, удовлетворяющий условию Фано. Для буквы А используется кодовое слово 1; для буквы Б используется кодовое слово 01. Какова минимальная общая длина кодовых слов для всех семи букв?

Решение

Нарисуем кодовое дерево, отмечая на нем буквы А и Б.

После этого видим, что свободная ветка (т.е. свободное кодовое слово) осталось всего одно, однако букв при этом необходимо закодировать еще 5. Значит, эту и последующие ветки необходимо раздваивать до тех пор, пока не поместятся все 5 букв.

(№ 1711) По каналу связи передаются сообщения, содержащие только семь букв: А, Б, В, Г, Д, Е и Ж. Для передачи используется двоичный код, удовлетворяющий условию Фано. Для буквы А используется кодовое слово 1; для буквы Б используется кодовое слово 01. Какова минимальная общая длина кодовых слов для всех семи букв?



Стоит отметить, что размещение букв в данном случае значения не имеет, так как в ответе от нас требует сумму всех кодовых слов, которая здесь составит $12+10+1+2 = 25$.

Условие 5

Все заглавные буквы русского алфавита закодированы неравномерным двоичным кодом, в котором никакое кодовое слово не является началом другого кодового слова. Это условие обеспечивает возможность однозначной расшифровки закодированных сообщений. Кодовые слова для некоторых букв известны: А – 000, Б – 01, В – 100, Г – 11, Д – 001. Укажите возможный код минимальной длины для буквы Я. Если таких кодов несколько, укажите тот из них, который имеет минимальное числовое значение.

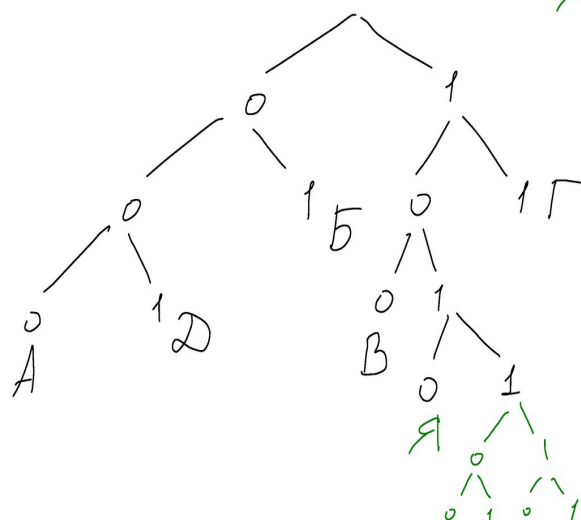
Решение

Перед тем, как приступить к решению задачи, обратим внимание на первую строку, где сказано, что закодировать необходимо не только представленные

буквы, но и те, что содержатся в русском алфавите. Это значит, что для каждой буквы должен остаться свой код.

Нарисуем двоичное дерево, где разместим буквы с известными нам кодами.

(№ 6435) Все заглавные буквы русского алфавита закодированы неравномерным двоичным кодом, в котором никакое кодовое слово не является началом другого кодового слова. Это условие обеспечивает возможность однозначной расшифровки закодированных сообщений. Кодовые слова для некоторых букв известны: А – 000, Б – 01, В – 100, Г – 11, Д – 001. Укажите возможный код минимальной длины для буквы Я. Если таких кодов несколько, укажите тот из них, который имеет минимальное числовое значение.



Я: 1010

После того, как мы расставим все буквы, то у нас останется свободное кодовое слово 101. Можно подумать, что букву Я можно вписать туда, однако же вместе с этим мы отнимем у дерева последнюю возможность к разветвлению и созданию дополнительных кодовых слов; из чего, в свою очередь, следует, что закодировать все буквы русского алфавита не получится.

Значит, прописывать букву Я сразу не стоит – сначала сделает разветвление, получая тем самым код 1010 и 1011. В первый мы впишем букву Я (т.к. требуется наименьшее численное значение), а второй оставим для возможного разветвления и кодирования всех оставшихся букв алфавита.

Условие 6

По каналу связи передаются сообщения, содержащие только семь букв: А, Б, Г, И, М, Р, Я. Для передачи используется двоичный код, удовлетворяющий

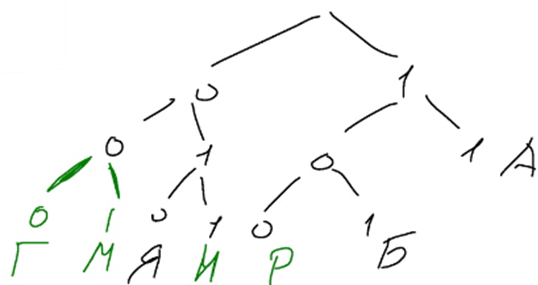
условию Фано. Кодовые слова для некоторых букв известны: А – 11, Б – 101, Я – 010. Какое наименьшее количество двоичных знаков потребуется для кодирования слова ГРАММ?

Решение

Для решения подобного типа задач важно понимать, что здесь у букв появляться приоритет. Например, в слове ГРАММ присутствует две буквы ММ. Значит, чтобы сумма двоичных знаков для кодирования слова ГРАММ была наименьшей, длины кодовых слов для М должны быть короче, чем остальные.

Также мы видим, что у буквы Г перед Я тоже, например, есть свой приоритет. Так как буква Г используется 1 раз, а буква Я не используется вовсе.

Таким образом: чем чаще используется буква в кодируемом слове, тем короче должен быть её код. При этом нередко возникает неоднозначные ситуации, когда становится не сразу понятно, как рациональное расставить кодовые слова для букв. В связи с этим необходимо сделать две модели и сравнить длину кодируемого слова в обоих вариантах деревьев.



В таком случае буквы из слова ГРАММ имеют следующие коды:

Г – 000

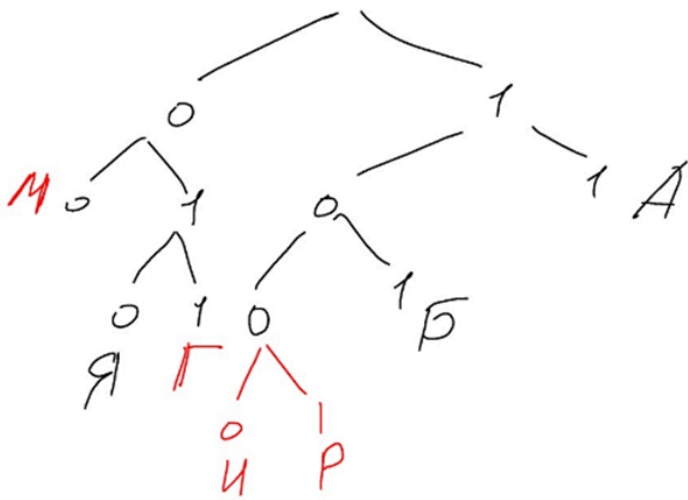
Р – 100

А – 11

М – 001

Значит, сумма двоичный знаком для кодирования равна $3+3+2+3*2=14$

Однако не остановимся на этом и рассмотрим другой вариант кодового дерева.



В таком случае буквы из слова ГРАММ имеют следующие коды:

Г – 011

Р – 1001

А – 11

М – 00

Значит, сумма двоичный знаком для кодирования равна $3+4+2+2*2=13$

Таким образом видно, что второй вариант кодового дерева оказался наиболее подходящим, так как сумма длин кодовых слов оказалась наименьшей. Ответ на задачу, следовательно, 13.

Условие 7

(И. Карпачев) По каналу связи передаются сообщения, содержащие только буквы из набора: Б, О, Р, Т, Ф, Я. Для передачи используется двоичный код, удовлетворяющий условию Фано. Кодовые слова для некоторых букв известны: Р – 01, Ф – 110. Для оставшихся букв Б, О, Т, Я кодовые слова неизвестны. Какое количество двоичных знаков потребуется для кодирования слова ФОТОРОБОТ, если известно, что оно закодировано минимально возможным количеством двоичных знаков?

Решение

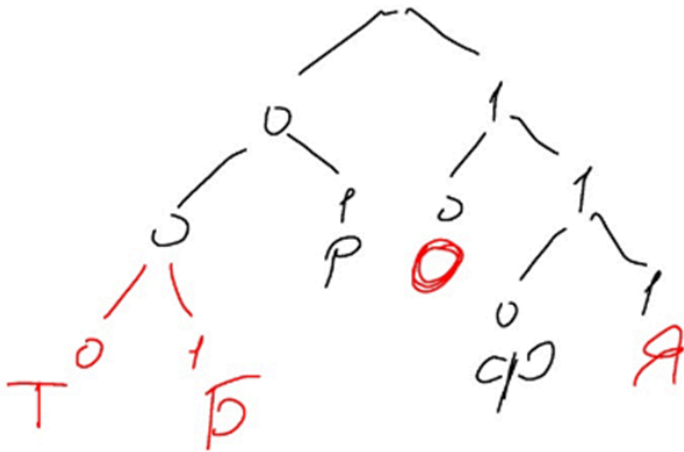
Как и в предыдущей задаче, обращаем внимание на частоту встречаемых букв в кодируемом слове ФОТОРОБОТ.

О – 4 раза

Т – 2 раза

Ф, Р, Б – 3 раза

Отсюда следует, что стоит постараться взять длины кодовых слов для часто встречающихся букв как можно короче. Снова нарисуем два варианта кодовых деревьев, чтобы найти оптимальное расположение букв и, следовательно, наименьшую длину закодированной последовательности.



Для первого варианта дерева буквы имеют следующий код:

О – 01

Т – 000

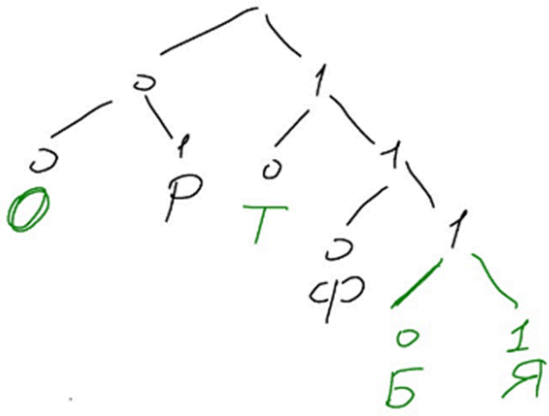
Б – 001

Ф – 110

Р – 01

Значит, суммарная длина закодированного слово ФОТОРОБОТ будет
 $2*4+3*2+3+3+2 = 23$

Далее рассмотрим второй вариант двоичного дерева:



Для первого варианта дерева буквы имеют следующий код:

О – 00

Т – 01

Б – 1110

Ф – 110

Р – 01

Значит, суммарная длина закодированного слово ФОТОРОБОТ будет

$$3+2+2+2+2+2+4+2+2 = 21$$

Таким образом, второй вариант дерева является более оптимальным и наименьшая длина равна 21.

Условие 8

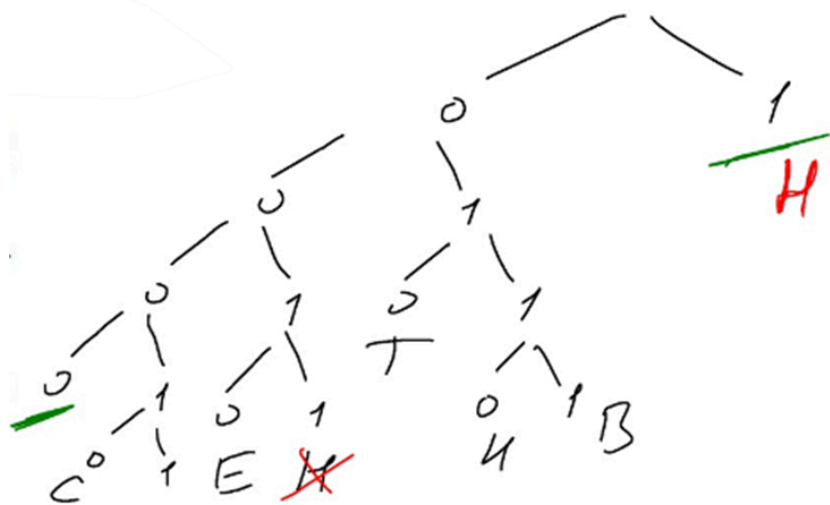
(Е. Джобс) По каналу связи передаются сообщения, содержащие только шесть букв: Т, Е, Н, С, И, В. Для передачи используется двоичный код, допускающий однозначное декодирование. Кодовые слова для букв известны: Т – 010, Е – 0100, Н – 1100, С – 01000, И – 0110, В – 1110. Как можно сократить код для буквы Н, чтобы сохранялось свойство однозначности декодирования? Если таких кодов несколько, в качестве ответа указать код наименьшей длины.

Решение

Перед тем, как решать данную задачу, обратим внимание на то, что прямое условие Фано нарушено. Код буквы Т начинается так, как у буквы Е и С. В таком случае необходимо воспользоваться обратным условие Фано. Что такое обратное условие? Давай смотреть.

1) Если прямое условие не соблюдено, то необходимо отзеркалить все коды. То есть: Т – 010, Е – 0010, Н – 0011, С – 00010, И – 0110, В – 0111.

2) Далее, уже для отзеркаленных кодов, строим двоичной дерево:



По условию задачи нам необходимо переставить Н так, чтобы ее код был как можно короче. Переставляем ее к незанятому коду с наименьшей длиной – т.е. на 1.

3) После этого отзеркаливаем обратно необходимое кодовое слово. Для Н, конечно, отзеркаленная 1 так и останется 1, что и будет ответом к задаче; но, например, у Н был бы код в дереве 011, тогда при отзеркаливании он имел бы вид 110, что и пошло бы в ответ.

Условие 9

(М. Дунаев) Для кодирования сообщений, состоящих только из букв Д, Е, О, М использовали неравномерный двоичный код, удовлетворяющий условию Фано. Известно, что при этом слово ДЕД кодируется с помощью последовательности 1010101. Найдите минимальную длину кодовой последовательности для слова ДОМ.

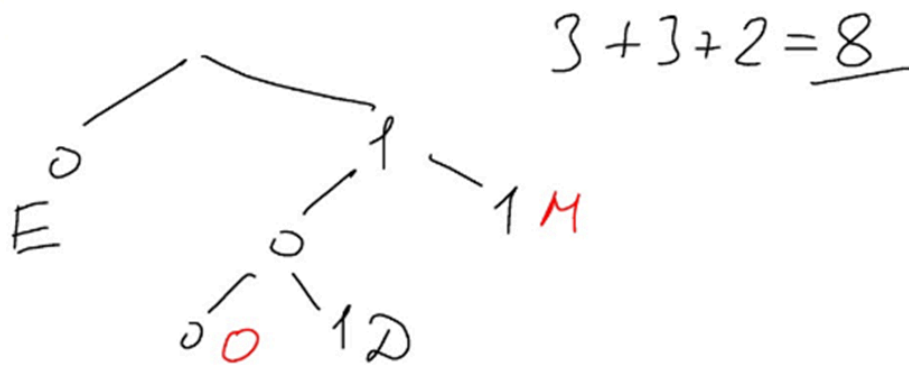
Решение

Попробуем определить, какой код соответствует буквам Д и Е, исходя из данных условий. Кодовую последовательность 1010101 можно разбить на подходящие кодовые слова двумя способами: 101 0 101 и 1 01010 1.

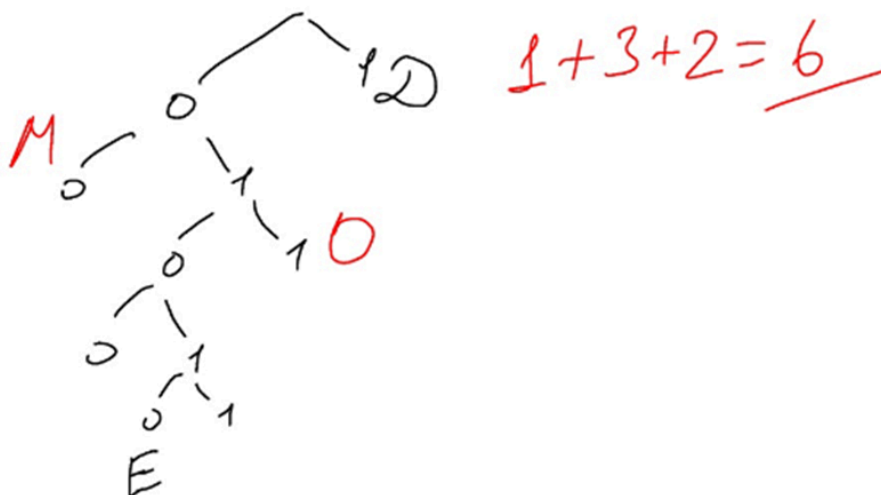
В первом случае, букве Д соответствует код 101, а Е – 0. Во втором случае, букве Д соответствует код 1, а букве Е – 01010.

Нарисуем кодовые деревья для двух случаев, чтобы найти минимально возможные кодовые слова для букв О и М.

В первом случае суммарная длина кодовой последовательности для слова ДОМ составит 8:



Во втором случае – 6:



Таким образом, наиболее оптимальным разделением было 1 01010 1, а подходящие коды для М и О – 00 и 011 соответственно. Таким образом, ответ на задачу – 6.

Условие 10

По каналу связи передаются сообщения, содержащие только буквы из набора: Ф, А, Р, О, Н. Для передачи используется двоичный код, удовлетворяющий прямому условию Фано. Это условие обеспечивает возможность однозначной расшифровки закодированных сообщений. Известен код для буквы А – 1. Также известно, что для слова АНАФОРА использовались 16 двоичных знаков, а для

слова ФАРФОР – 15. Укажите код, использующийся для буквы Н. Если кодов одинаковой длины несколько – выберите имеющий наименьшее числовое значение.

Решение

Составим некоторые математические элементарные уравнения, которые позволят сделать выводы о сумме кодов каких-то букв. С учетом того, что мы знаем длину кодовых последовательностей двух слов и длину кода для буквы А, мы получим следующие уравнения:

$$3 \cdot \overset{1}{A} + H + \Phi + O + P = 16$$

$$\underbrace{H + \Phi + O + P}_{\substack{4 \quad 9}} = 13$$

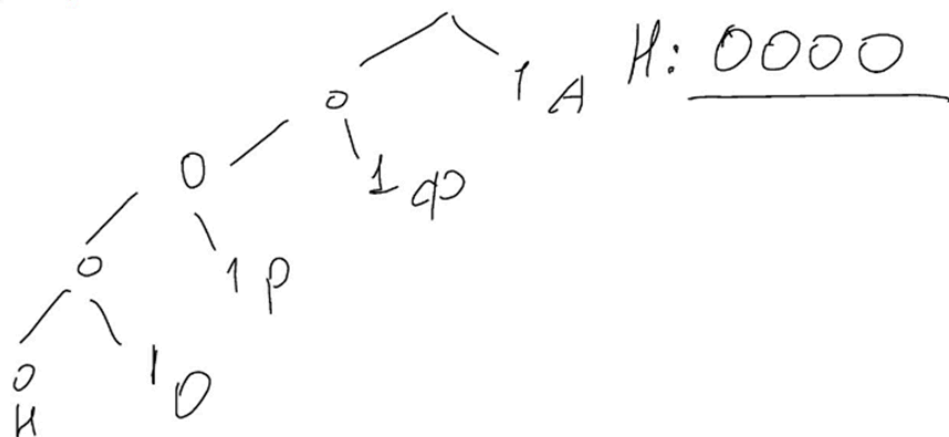
$$2\Phi + 2P + \bar{A} + O = 15$$

$$\underbrace{2\Phi + 2P}_{\substack{\Phi + P = 5}} + \underbrace{O}_{4} = 14$$

Далее мы начинаем предполагать, чему может быть равен код для буквы О. Почему для О, а не для Н? Потому что касательно буквы О можно сделать вывод, что длина ее кодового слова – четная (исходя из второго уравнения). Значит, перебрать необходимо только четные варианты: 2, 4, 6 и т.д.

Предположим, что код для О имеет длину, равную 4. Тогда сумма 2Φ и $2P$ из второго уравнения равна 10, а, значит, сумма Φ и P равна 5. В это же время из первого уравнения видно, что код для буквы Н равняется 4.

Попробуем составить такое двоичное дерево, которое будет удовлетворять всем заданным условиям.



Как видим, такое дерево построить возможно. Значит, наименьшая длина кодового слова для Н равна 4.

P.S: А такое дерево, удовлетворяющее условие про суммы длин букв – единственно? Нет! Был еще вариант составления и другого дерева, однако в нем длина кода для Н не была бы численно минимальной:

