

Информатика. Кодирование и декодирование информации. Условие Фано.

Условие Фано: никакое кодовое слово не является началом другого кодового слова. Для чего нужно условие Фано? Для однозначного декодирования информации.

Например: буква А имеет кодовое слово 00 и кодирование символов происходит согласно условию Фано, значит, никакая другая буква не может быть закодирована словом, начало которого будет содержать 00, так, например, мы не сможем закодировать букву В кодовыми словами 001, 000.

В задании на условие Фано часто встречается неравномерное кодирование. Что это такое?

Неравномерное кодирование – это когда кодовые слова для разных символов имеют разную длину. Например: А – 0, В – 111, буква А имеет длину в 1 символ, буква В имеет кодовое слово длиной 3 символа.

Тогда логично, что равномерное кодирование – это когда кодовые слова для разных символов имеют одинаковую длину. Например: А – 01, В – 10.

Рассмотрим примеры 4 задания:

Задача

Для 5 букв алфавита заданы их двоичные коды (для некоторых букв – из двух бит, для некоторых – из трех). Эти коды представлены в таблице:

А	Д	К	Н	Р
010	00	111	011	110

Какой набор букв закодирован двоичной строкой 11101100010110?

Ответ

КНДАР

Решение

Заметим, что буквы закодированы согласно условию Фано, поэтому декодировать будем слева на право, так как каждый код не является началом другого кодового слова.

Первое кодовое слово может стоять из сколько угодно символов, но так как согласно таблице максимальная длина – 3 символа, минимальная – 2. Значит, первое кодовое слово может быть либо 111, либо 11, но так как в таблице буква

К-111, то по условию фано не существует такой буквы, которая закодировалась бы кодом 11, поэтому первая буква, однозначно, К. $\underbrace{111}_K 01100010110$

Дальше 01 либо 011, но так как в таблице есть буква Н, двоичный код которой 011, то вторая буква—Н. $\underbrace{111}_K \underbrace{011}_H 00010110$

00 или 000? Очевидно, что 00, так как у нас есть буква Д закодированная 00. $\underbrace{111}_K \underbrace{011}_H \underbrace{00}_D 010110$

Заметим, что и 010 мы тоже можем однозначно декодировать и это буква А. $\underbrace{111}_K \underbrace{011}_H \underbrace{00}_D \underbrace{010}_A 110$

Осталось только 110—это Р. $\underbrace{111}_K \underbrace{011}_H \underbrace{00}_D \underbrace{010}_A \underbrace{110}_R$

Нам удалось декодировать строку, мы победили!

Рассмотрим задачу, где нам нужно самим придумать кодовые слова для букв.

Задача

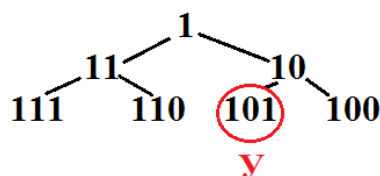
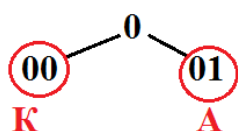
По каналу связи передаются сообщения, содержащие только семь букв: К, У, Р, А, Т, О. Для передачи используется двоичный код, удовлетворяющий условию Фано. Кодовые слова для некоторых букв известны: К — 00, У — 101, А-01. Какое наименьшее кодовое слово может соответствовать букве Р?

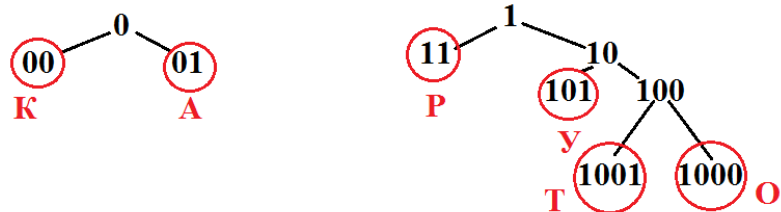
Ответ

11

Решение

В таких задачах нужно составить деревья, "корни" которых будут начинаться с 0 и 1, с каждым шагом мы добавляем либо единицу, либо ноль в конец кодового слова. Составим такое дерево и отметим какие буквы уже закодированы.





Мы можем продолжать это дерево бесконечно, но нам требуется найти НАИ-МЕНЬШЕЕ кодовое слово для буквы Р. Можем ли взять 11? Да! потому что 100 мы можем продолжить дальше и получить 1001, например, для Т и 1000 для О.

Мы победили! Р – 11.

Задачи для самостоятельного решения.

Задача 1

Для 4 букв алфавита заданы их двоичные коды (для некоторых букв — из двух бит, для некоторых — из трех). Эти коды представлены в таблице:

А	М	К	С
101	010	011	000

Какой набор букв закодирован двоичной строкой 010101000011101?

Ответ

МАСКА

Решение

Заметим, что буквы закодированы согласно условию Фано, поэтому декодировать будем слева на право, так как каждый код не является началом другого кодового слова. Так же заметим, что все буквы закодированы равномерно (по три символа), значит разобьем числовую последовательность на тройки и будем декодировать слева направо.

010-это М, 101-это А, затем 000-С, 011-К, 101-А

$\underbrace{010}_M \underbrace{101}_A \underbrace{000}_C \underbrace{011}_K \underbrace{101}_A$

Задача 2

Для 6 букв алфавита заданы их двоичные коды (для некоторых букв — из двух бит, для некоторых — из трех). Эти коды представлены в таблице:

Б	Р	О	В	К	И
10	111	001	011	010	000

Какой набор букв закодирован двоичной строкой 01000010001011111010?

Ответ

КИБОВРК

Решение

Заметим, что буквы закодированы согласно условию Фано, поэтому декодировать будем слева на право, так как каждый код не является началом другого кодового слова. Поэтому будем декодировать слева направо.

010 – К $\underbrace{010}_K$ 0001000101111010

000 – И $\underbrace{010}_K \underbrace{000}_И$ 1000101111010

10 – Б $\underbrace{010}_\text{К} \underbrace{000}_\text{И} \underbrace{10}_\text{Б} 001011111010$
 001 – О $\underbrace{010}_\text{К} \underbrace{000}_\text{И} \underbrace{10}_\text{Б} \underbrace{001}_\text{О} 011111010$
 011 – В $\underbrace{010}_\text{К} \underbrace{000}_\text{И} \underbrace{10}_\text{Б} \underbrace{001}_\text{О} \underbrace{011}_\text{В} 111010$
 111 – Р $\underbrace{010}_\text{К} \underbrace{000}_\text{И} \underbrace{10}_\text{Б} \underbrace{001}_\text{О} \underbrace{011}_\text{В} \underbrace{111}_\text{Р} 010$
 010 – К $\underbrace{010}_\text{К} \underbrace{000}_\text{И} \underbrace{10}_\text{Б} \underbrace{001}_\text{О} \underbrace{011}_\text{В} \underbrace{111}_\text{Р} \underbrace{010}_\text{К}$
 Наш ответ – КИБОВРК.

Задача 3

Для 5 букв алфавита заданы их двоичные коды (для некоторых букв — из двух бит, для некоторых — из трех). Эти коды представлены в таблице:

Т	И	К	Г	Ч	Н	О
001	011	1010	11	100	0001	0101

Какой набор букв закодирован двоичной строкой 0001010111010100101011001010011?

Ответ

НОГОТОЧКИ

Решение

Заметим, что буквы закодированы согласно условию Фано, поэтому декодировать будем слева на право, так как каждый код не является началом другого кодового слова. Поэтому будем декодировать слева направо.

0001 – Н $\underbrace{0001}_\text{Н} 010111010100101011001010011$
 0101 – О $\underbrace{0001}_\text{Н} \underbrace{0101}_\text{О} 11010100101011001010011$
 11 – Г $\underbrace{0001}_\text{Н} \underbrace{0101}_\text{О} \underbrace{11}_\text{Г} 010100101011001010011$
 0101 – О $\underbrace{0001}_\text{Н} \underbrace{0101}_\text{О} \underbrace{11}_\text{Г} \underbrace{0101}_\text{О} 00101011001010011$
 001 – Т $\underbrace{0001}_\text{Н} \underbrace{0101}_\text{О} \underbrace{11}_\text{Г} \underbrace{0101}_\text{О} \underbrace{001}_\text{Т} 01011001010011$
 0101 – О $\underbrace{0001}_\text{Н} \underbrace{0101}_\text{О} \underbrace{11}_\text{Г} \underbrace{0101}_\text{О} \underbrace{001}_\text{Т} \underbrace{0101}_\text{О} 1001010011$
 100 – Ч $\underbrace{0001}_\text{Н} \underbrace{0101}_\text{О} \underbrace{11}_\text{Г} \underbrace{0101}_\text{О} \underbrace{001}_\text{Т} \underbrace{0101}_\text{О} \underbrace{100}_\text{Ч} 1010011$
 1010 – К $\underbrace{0001}_\text{Н} \underbrace{0101}_\text{О} \underbrace{11}_\text{Г} \underbrace{0101}_\text{О} \underbrace{001}_\text{Т} \underbrace{0101}_\text{О} \underbrace{100}_\text{Ч} \underbrace{1010}_\text{К} 011$

011 – И $\underbrace{0001}_\text{Н} \underbrace{0101}_\text{О} \underbrace{11}_\text{Г} \underbrace{0101}_\text{О} \underbrace{001}_\text{Т} \underbrace{0101}_\text{О} \underbrace{100}_\text{Ч} \underbrace{1010}_\text{К} \underbrace{011}_\text{И}$

Задача 4

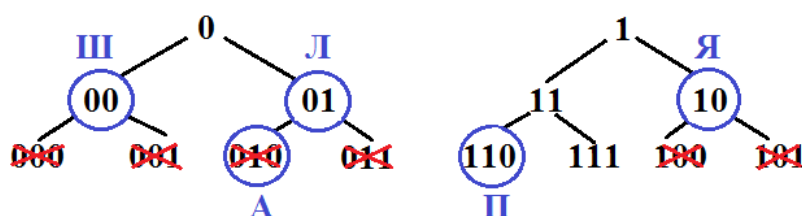
Клон АР составлял задачку для пробника, но закрабил. Необходимо найти ошибку в решении Клона и записать её в ответ. Известно, что задача была составлена на проверку знания прямого условия Фано. Кодировается сообщение с помощью символов Ш, Л, Я, П, А. Для каждой буквы выделено кодовое слово: Ш — 00; Л — 01; Я — 10; П — 110; А — 010. В ответе укажите букву и кодовое слово без пробелов. Например, P110.

Ответ

A010

Решение

Так как буквы закодированы с помощью прямого условия Фано, то составим дерево кодовых слов и отметим на нем закодированные буквы.



Л имеет кодовое слово 01, значит, другие буквы, согласно условию Фано, не могут начинаться с 01, значит 010 и 011 – это запрещенные кодовые слова. Но заметим, что буква А закодирована запретным кодовым словом 010. Значит, ошибка найдена.

Задача 5

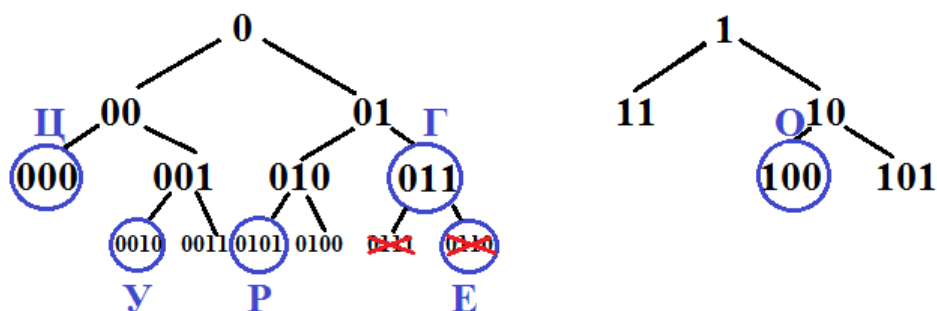
Эльф составлял задачку для пробника, но закрабил. Необходимо найти ошибку в решении Эльфа и записать её в ответ. Известно, что задача была составлена на проверку знания прямого условия Фано. Кодировается сообщение с помощью символов О, Г, У, Р, Е, Ц. Для каждой буквы выделено кодовое слово: О — 100; Г — 011; У — 0010; Р — 0101; Е — 0110; Ц — 000. В ответе укажите букву и кодовое слово без пробелов. Например, P110.

Ответ

E0110

Решение

Так как буквы закодированы с помощью прямого условия Фано, то составим дерево кодовых слов и отметим на нем закодированные буквы.



Г имеет кодовое слово 011, значит, другие буквы, согласно условию Фано, не могут начинаться с 011, значит 0111 и 0110 – это запрещенные кодовые слова. Но заметим, что буква Е закодирована запретным кодовым словом 0110. Значит, ошибка найдена.

Задача 6

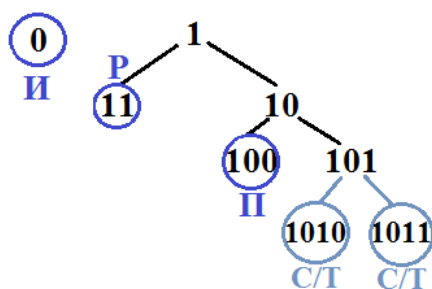
По каналу связи передаются сообщения, которые содержат только следующие буквы: П, Р, И, Т, С; для передачи используется двоичный код, удовлетворяющий условию Фано. Для букв П, И, Р используются такие кодовые слова: П – 100 ; И – 0 ; Р – 11. Укажите кратчайшее кодовое слово для буквы Т, при котором код будет допускать однозначную расшифровку. Если таких кодов несколько, укажите код с наименьшим числовым значением.

Ответ

1010

Решение

Так как буквы закодированы с помощью прямого условия Фано, то составим дерево кодовых слов и отметим на нем закодированные буквы.



Для того чтобы закодировать букву Т нужно продлить дерево от 101, так как в сообщении передаются 5 букв, три из которых уже закодированы, значит, нужно закодировать еще две, поэтому мы не можем закодировать букву Т 101, так как нам не хватит кода для С.

Получается нам нужно выбрать минимальный код для Т, а это 1010.

Задача 7

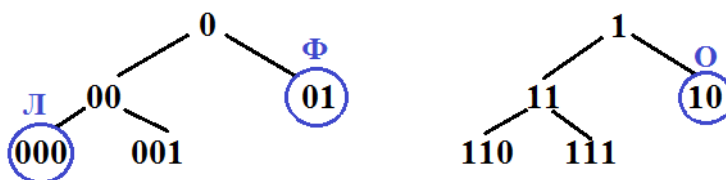
По каналу связи передаются сообщения, которые содержат только следующие буквы: Ф, О, Л, Ъ, Г, А; для передачи используется двоичный код, удовлетворяющий условию Фано. Для букв Ф, О, Л используются такие кодовые слова: Ф – 01 ; О – 10 ; Л – 000. Укажите кратчайшее кодовое слово для буквы А, при котором код будет допускать однозначную расшифровку. Если таких кодов несколько, укажите код с наименьшим числовым значением.

Ответ

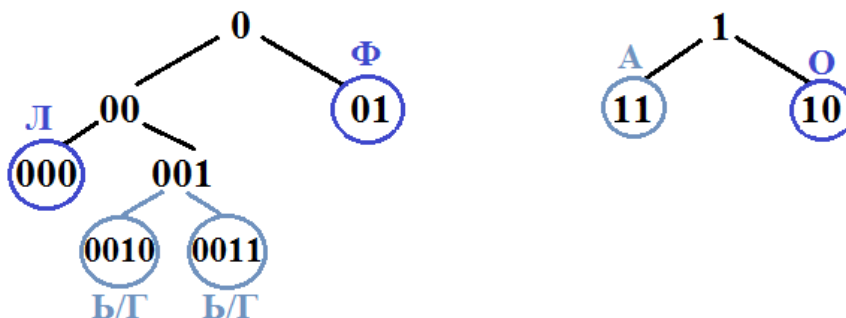
11

Решение

Так как буквы закодированы с помощью прямого условия Фано, то составим дерево кодовых слов и отметим на нем закодированные буквы.



В итоге у нас получается, что есть 3 свободных кодовых слова, значит, выберем наименьший из них 001 и все. НО нет! Это будет ошибкой. Почему? Потому что мы должны выбрать **НАИМЕНЬШИЙ** код для буквы А, а про другие буквы нам ничего не сказано, значит мы можем дать им любое кодовое слово. Заметим, что 11 – это наименьшее возможное кодовое слово! Значит для А выберем 11, а для Ъ, Г продолжим 001.



Задача 8

По каналу связи передаются сообщения, которые содержат только следующие буквы: Б, Р, О, В, К, И; для передачи используется двоичный код, удовлетворяющий условию Фано. Для букв Б, Р, О используются такие кодовые слова: Б –

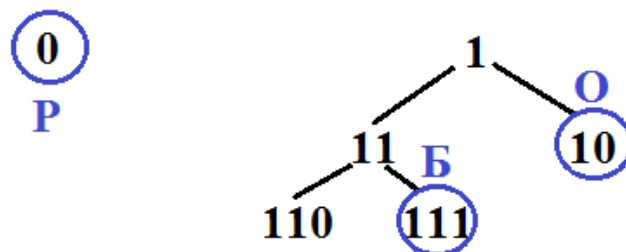
111 ; Р – 0 ; О – 10. Укажите кратчайшее кодовое слово для буквы К, при котором код будет допускать однозначную расшифровку. Если таких кодов несколько, укажите код с наименьшим числовым значением.

Ответ

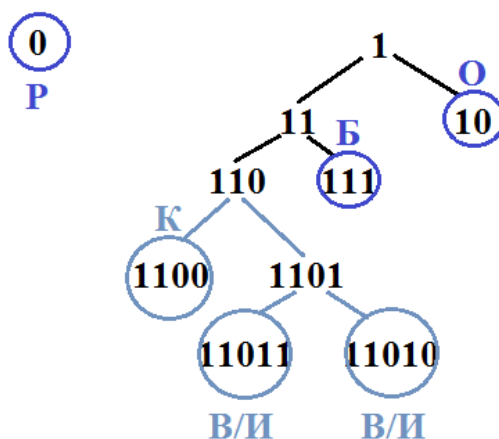
1100

Решение

Так как буквы закодированы с помощью прямого условия Фано, то составим дерево кодовых слов и отметим на нем закодированные буквы.



Заметим, что свободное кодовое слово только 110. Можем ли мы им закодировать букву К? Конечно нет. Потому что мы не сможем закодировать буквы В,И по условию Фано. Значит нам нужно продолжить 110. Получим 1100 и 1101. Мы должны выбрать **НАИМЕНЬШИЙ** код для буквы К, а про другие буквы нам ничего не сказано, значит мы можем дать им любое кодовое слово. Заметим, что 1100 – это наименьшее возможное кодовое слово! А что с буквами В,И? Просто продлим дальше 1101.



Задача 9

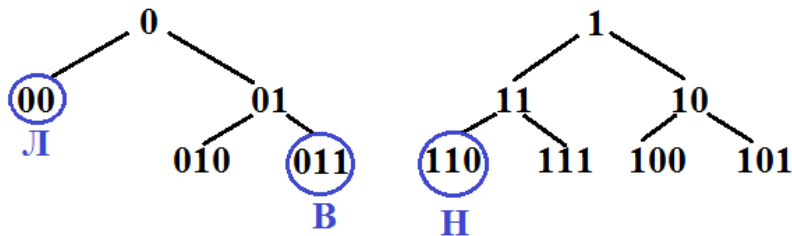
По каналу связи передаются сообщения, содержащие только пять букв: Л, А, В, Н, Д. Для передачи используется двоичный код, удовлетворяющий условию Фано. Кодовые слова для некоторых букв известны: Л – 00 ; В – 011 ; Н – 110. Какое наименьшее количество двоичных знаков потребуется для кодирования слова ЛАВАНДА?

Ответ

17

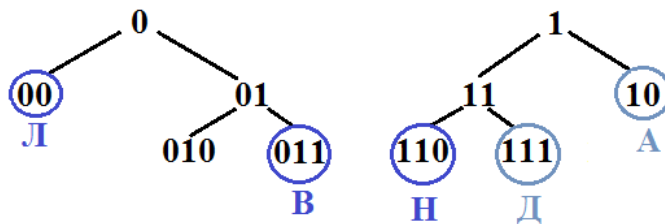
Решение

Так как буквы закодированы с помощью прямого условия Фано, то составим дерево кодовых слов и отметим на нем закодированные буквы.



В итоге задача сводится к тому, что нам нужно додуматься какая буква должна иметь наименьшую длину кодового слова. В нашем случае наименьшую длину должна иметь буква А, так как она встречается в слове три раза.

У нас есть 4 свободных узла, если мы не берем 10 (010, 100, 101, 111) и 3, если мы берем 10 (10, 010, 111). Логично, если нам нужно наименьшее количество знаков, то нам подходит второй вариант, так как для А мы выделим 10, а для Д любое кодовое слово из оставшихся (010, 111), например, 111.



Тогда имеем: $2 + 2 + 3 + 2 + 3 + 3 + 2 = 17$

Задача 10

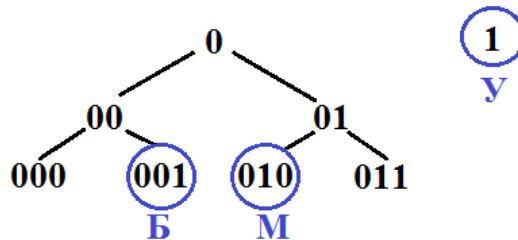
По каналу связи передаются сообщения, содержащие только шесть букв: Б, А, У, Е, О, М. Для передачи используется двоичный код, удовлетворяющий условию Фано. Кодовые слова для некоторых букв известны: У – 1 ; Б – 001 ; М – 010. Какое наименьшее количество двоичных знаков потребуется для кодирования слова БАОБАБ?

Ответ

19

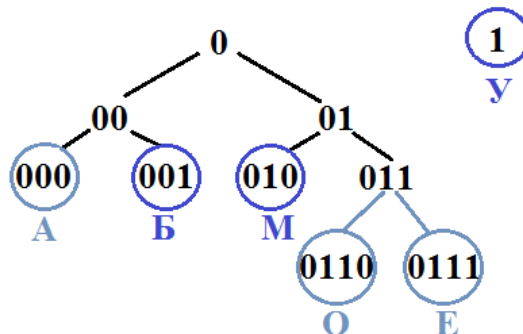
Решение

Так как буквы закодированы с помощью прямого условия Фано, то составим дерево кодовых слов и отметим на нем закодированные буквы.



В итоге задача сводится к тому, что нам нужно додуматься какая буква должна иметь наименьшую длину кодового слова. В нашем случае наименьшую длину должна иметь буква А, так как она встречается в слове два раза.

У нас есть два свободных узла (000, 010), но нам этого не хватит, так как нужно закодировать еще 3 буквы (А, Е, О), поэтому нужно продлить любой из узлов. Логично, если нам нужно наименьшее количество знаков, то нам нужно взять одно из свободных кодов длиной 3 (000 или 011), а другой продолжить. Допустим, что А – 000, поэтому продолжим 011.



Пусть О будет иметь кодовое слово 0110, а Е – 0111.

Тогда имеем: $3 + 3 + 4 + 3 + 3 + 3 = 19$

Задача 11

Для кодирования букв М, О, Щ, Н решили использовать двоичное представление чисел 4, 5, 6 и 7 соответственно (с сохранением одного незначащего нуля в случае одноразрядного представления). Закодируйте последовательность букв МОЩНО таким способом и результат запишите шестнадцатеричным кодом.

Ответ

4BBD

Решение

Закодируем каждую букву в двоичной системе счисления:

М – 100, О – 101, Щ – 110, Н – 111.

Тогда слово МОЩНО – 100101110111101_2 .

Переведем закодированное слово в 16СС:

$$100101110111101_2 = 100|1011|1011|1101_2 = 4BBD_{16}$$

Задача 12

Для кодирования букв М, С, А, Р, К, Е, И, Ш решили использовать двоичное представление чисел 0, 1, 2, 3, 4, 5, 6 и 7 соответственно (с сохранением одного незначащего нуля в случае одноразрядного представления). Закодируйте последовательность букв СМЕШАРИКИ таким способом и результат запишите шестнадцатеричным кодом.

Ответ

25F7A6

Решение

Закодируем каждую букву в двоичной системе счисления:

С – 01, М – 00, Е – 101, Ш – 111, А – 10, Р – 11, И – 110, К – 100.

Тогда слово СМЕШАРИКИ – $01001011111011110100110_2$.

Переведем закодированное слово в 16СС:

$$01001011111011110100110_2 = 010|0101|1111|0111|1010|0110_2 = 25F7A6_{16}$$

Задача 13

Для кодирования букв А, Р, О, Я, И, С, Н, Т, Е решили использовать двоичное представление чисел 0, 1, 2, 3, 4, 5, 6, 7 и 8 соответственно (с сохранением одного незначащего нуля в случае одноразрядного представления). Закодируйте последовательность букв РАССТОЯНИЕ таким способом и результат запишите шестнадцатеричным кодом.

Ответ

25BEF48

Решение

Закодируем каждую букву в двоичной системе счисления:

Р – 01, А – 00, С – 101, Т – 111, О – 10, Я – 11, Н – 110, И – 100, Е – 1000.

Тогда слово РАССТОЯНИЕ – $010010110111110111101001000_2$.

Переведем закодированное слово в 16СС:

$$010010110111110111101001000_2 = 010|0101|1011|1110|1111|0100|1000_2 = 25BEF48_{16}$$

Задача 14

Для кодирования букв Р, Д, П, И, Я, С, Е решили использовать двоичное представление чисел 0, 1, 2, 3, 4, 5 и 6 соответственно (с сохранением одного незначащего нуля в случае одноразрядного представления). Закодируйте последовательность букв ДЕПРЕССИЯ таким способом и результат запишите восьмеричным кодом.

Ответ

16432274

Решение

Закодируем каждую букву в двоичной системе счисления:

Д – 01, Е – 110, П – 10, Р – 00, С – 101, И – 11, Я – 100.

Тогда слово ДЕПРЕССИЯ – $01110100011010110111100_2$.

Переведем закодированное слово в 8СС:

$$01110100011010110111100_2 = 01|110|100|011|010|110|111|100_2 = 16432674_8$$

Задача 15

Для кодирования букв К, Ф, З, Р, О, Е, Э, Л, Т решили использовать двоичное представление чисел 0, 1, 2, 3, 4, 5, 6, 7 и 8 соответственно (с сохранением одного незначащего нуля в случае одноразрядного представления). Закодируйте последовательность букв ЭЛЕКТРОФОРЕЗ таким способом и результат запишите восьмеричным кодом.

Ответ

33644343166

Решение

Закодируем каждую букву в двоичной системе счисления:

Э – 110, Л – 111, Е – 101, К – 00, Т – 1000, Р – 11, О – 100, Ф – 01, З – 10.

Тогда слово ЭЛЕКТРОФОРЕЗ – $11011110100100011100011001110110_2$.

Переведем закодированное слово в 8СС:

$$11|011|110|100|100|011|100|011|001|110|110_2 = 33644343166_8$$

Задача 16

Для кодирования букв Ф, И, Т, Ц, Н, Я, Р, Е решили использовать двоичное представление чисел 3, 4, 5, 6, 7, 8, 9 и 10 соответственно (с сохранением одного незначащего нуля в случае одноразрядного представления). Закодируйте последовательность букв ИНТЕРФЕРЕНЦИЯ таким способом и результат запишите шестнадцатеричным кодом.

Ответ

9ED4F535F48

Решение

Закодируем каждую букву в двоичной системе счисления:

И – 100, Н – 111, Т – 101, Е – 1010, Р – 1001, Ф – 11, Ц – 110, Я – 1000

Тогда слово ИНТЕРФЕРЕНЦИЯ –

$10011110110101001111010100110101111101001000_2$.

Переведем закодированное слово в 16СС:

$1001|1110|1101|0100|1111|0101|0011|0101|1111|0100|1000_2 = 9ED4F535F48_{16}$