

Разбор задач

Задача 1. Световое табло состоит из лампочек. Каждая лампочка может находиться в одном из двух состояний («включено» или «выключено»). Какое наименьшее количество лампочек должно находиться на табло, чтобы с его помощью можно было передать 18 различных сигналов?

Решение:

Т.к. лампочка, может находиться в одном из двух состояний - «включено» или «выключено», то мы можем провести аналогию между лампочками и битами (один бит может принимать значение 0 или 1). Тогда мы можем переформулировать условие задачи в рамках теории кодирования: какое наименьшее количество битов потребуется, чтобы закодировать 18 сигналов?

Ответ на этот вопрос нам даст формула Хартли: $i = \log_2 N$, где i – длина кода, а N – количество кодируемых объектов. В нашем случае $N=18$, следовательно $i = \log_2 18 \approx 5$ бит (лампочек).

Задача 2. Световое табло состоит из лампочек. Каждая лампочка может находиться в одном из двух состояний («включено», «выключено» или «мигает»). Какое наименьшее количество лампочек должно находиться на табло, чтобы с его помощью можно было передать 18 различных сигналов?

Решение:

В отличие от предыдущей задачи лампочка, может находиться в одном из трех состояний. Следовательно, мы должны рассматривать кодирование не в двоичной системе счисления, а в троичной. И тогда в формуле Хартли в качестве основания логарифма мы должны использовать значение 3.

Поэтому нам потребуется $i = \log_3 18 \approx 3$ лампочки.

Задача 3. Сколько существует различных последовательностей из символов «плюс» и «минус», длиной ровно в пять символов?

Решение:

Опять проведем аналогию с двоичным кодированием: один символ «плюс» или «минус» мы можем рассматривать как один бит. Тогда условие данной задачи обратное условию задачи 1: нам известна длина двоичного кода – 5 бит; нас интересует – сколько различных двоичных последовательностей данной длины мы можем получить?

Ответ на этот вопрос нам также даст формула Хартли, записанная в степенной форме: $N = 2^i$. В нашем случае мы получим $N = 2^5 = 32$ различных последовательностей.

Задача 4. Для кодирования секретного сообщения используются 12 специальных значков-символов. При этом символы кодируются одним и тем же минимально возможным количеством бит. Чему равен информационный объем (в байтах) сообщения длиной в 256 символов?

Решение:

Из условия задачи «Для кодирования секретного сообщения используются 12 специальных значков-символов. При этом символы кодируются одним и тем же минимально возможным количеством бит.» следует, что для кодирования используется алфавит, мощность которого равна $N=12$ (все символы алфавита равнозначны). Тогда информационный вес символов одинаков и вычисляется по формуле Хартли $i = \log_2 12 \approx 4$ бита.

Тогда информационный объем сообщения, согласно алфавитному подходу, измеряется по формуле $V = i \cdot k$. Значение i мы уже нашли, а k по условию задачи равно 256. Получаем: $V = 4 \text{ (бита)} \cdot 256 \text{ (символов)} = 1024 \text{ (бита)} = 128 \text{ (байта)}$.

Задача 5. Метеорологическая станция ведет наблюдение за влажностью воздуха. Результатом одного измерения является целое число от 0 до 100 процентов, которое записывается при помощи минимально возможного количества бит. Станция сделала 50 измерений. Определите информационный объем результатов наблюдений в битах.

Решение:

Из условия задачи «Результатом одного измерения является целое число от 0 до 100 процентов, которое записывается при помощи минимально возможного количества бит.» следует, что для записи одного измерения используется «алфавит» из чисел от 0 до 100, мощность которого равна $N=101$ (все символы алфавита равнозначны). Тогда информационный вес одного измерения можно вычислить по формуле Хартли $i=\log_2 101 \approx 7$ бит.

Тогда информационный объем всех измерений, согласно алфавитному подходу, измеряется по формуле $V=i \cdot k$. Значение i мы уже нашли, а k по условию задачи равно 80. Получаем: $V=7$ (бита) $\cdot$ 50 (измерений) = 350 (бит).

Задача 6. Мощность алфавита равна 256. Сколько Кбайт памяти потребуется для сохранения 160 страниц текста, содержащего в среднем 192 символа на каждой странице?

Решение:

Так как указана только мощность алфавита (без вероятностных характеристик символов), то речь идет о равнозначности символов алфавита. Тогда по формуле Хартли информационный вес каждого символа равен $i=\log_2 256=8$ бит.

Так как сообщение занимает 160 страниц по 192 символа на каждой, то общее количество символов равно $k=160 \cdot 192=30720$.

Находим информационный вес всего сообщения:

$V=8$ (бит) $\cdot$ 30720 (символов) = 245760 (бит) = 30720 (байт) = 30 (Кбайт).

Замечание. Как видим, при решении нам пришлось оперировать достаточно большими числами. Иногда бывает удобно составить выражение, выделить степень двойки или разложить число на множители, а затем выполнить сокращение. Например, следующим образом:

$$V = 8(\text{бит}) \cdot 30720(\text{символов}) = \frac{8 \cdot 30720}{8} (\text{байт}) = \frac{30 \cdot 1024}{1024} (\text{Кбайт}) = 30 \text{ Кбайт}$$

Задача 7. Объем сообщения – 7,5 Кбайт. Известно, что данное сообщение содержит 7680 символов. Какова мощность алфавита?

Решение:

Чтобы определить мощность алфавита, нужно знать каким количеством бит кодируется каждый символ сообщения, т.е. нужно найти значение i (в битах). Так как нам известны объем сообщения и количество символов в нем, то i мы можем найти по формуле:

$$i = \frac{V}{k} = \frac{7,5(\text{Кбайт})}{7680(\text{символов})} = \frac{7,5 \cdot 8 \cdot 1024(\text{бита})}{7,5 \cdot 1024(\text{символа})} = 8(\text{бит}).$$

Тогда по формуле Хартли, записанной в степенной форме, мы можем вычислить мощность заданного алфавита: $N=2^i=2^8=256$ символов.

Задача 8. В некоторой стране автомобильный номер длиной 7 символов составляется из заглавных букв (всего используется 26 букв) и десятичных цифр в любом порядке. Каждый символ кодируется одинаковым и минимально возможным количеством бит, а каждый номер – одинаковым и минимально возможным количеством байт. Определите объем памяти (в байтах), необходимый для хранения 20 автомобильных номеров.

Решение:

Из условия задачи следует, что для записи номера используется алфавит, состоящий из 26 букв и 10 цифр. Мощность такого алфавита равна 36. Тогда по формуле Хартли информационный вес каждого символа алфавита равен $i=\log_2 36 \approx 6$ бит.

Следовательно, чтобы закодировать 7-ми значный номер автомобиля, нам потребуется $v=6(\text{бит}) \cdot 7(\text{символов в одном номере})=42(\text{бита}) \approx 6(\text{байт})$.

Теперь мы можем вычислить объем памяти, необходимый для хранения 20 автомобильных номеров: $V=6(\text{байт на один номер}) \cdot 20(\text{номеров})=120(\text{байт})$

Задача 9. В базе данных хранятся записи, содержащие информацию о датах. Каждая запись содержит три поля: год (число от 1 до 2100), номер месяца (число от 1 до 12) и номер дня в месяце (число от 1 до 31). Каждое поле записывается отдельно от других полей с помощью минимально возможного числа бит. Определите минимальное количество бит, необходимых для кодирования одной записи.

Решение:

Так как каждая запись состоит из трех полей, записывающихся отдельно друг от друга, то мы можем применить закон аддитивности, чтобы определить минимальное количество бит, необходимых для кодирования одной записи: $i=i_{\text{год}}+i_{\text{месяц}}+i_{\text{день}}$.

Так как год варьирует от 1 до 2100, то $i_{\text{год}}=\log_2 2100 \approx 12(\text{бит})$.

Так как месяц варьирует от 1 до 12, то $i_{\text{месяц}}=\log_2 12 \approx 4(\text{бит})$.

И, так как день варьирует от 1 до 31, то $i_{\text{день}}=\log_2 31 \approx 5(\text{бит})$.

Получаем, что $i=12+4+5=21(\text{бит})$

Задача 10. В корзине лежат 8 черных шаров и 24 белых. Сколько бит информации несет сообщение о том, что достали черный шар?

Решение:

Так как в корзине лежит разное количество черных и белых шаров, при этом черных меньше, то вероятность извлечения черного шара будет меньше, чем белого, а вот количество информации при извлечении черного шара мы получим больше, чем при извлечении белого. Чтобы определить это количество, нам потребуется воспользоваться формулой Шеннона: $i_{\text{черного}} = \log_2 \frac{1}{p_{\text{черного}}}$, где $p_{\text{черного}}$ – вероятность извлечения черного шара

определяется по формуле $p_{\text{черного}} = \frac{k_{\text{черных шаров}}}{n_{\text{общее количество шаров}}}$.

Проводим расчеты: $p_{\text{черного}} = \frac{8}{8+24} = \frac{8}{32} = \frac{1}{4}$. Тогда $i_{\text{черного}} = \log_2 \frac{1}{1/4} = \log_2 4 = 2(\text{бит})$.

Задача 11. В корзине лежат черные и белые шары. Среди них 18 черных шаров. Сообщение о том, что достали белый шар, несет 2 бита информации. Сколько всего шаров в корзине?

Решение:

Так же как и в предыдущей задаче, нам придется использовать формулу Шеннона.

Обозначим количество белых шаров за X, тогда общее количество шаров в корзине X+18.

Нам известно, что сообщение о том, что из корзины достали белый шар, несет 2 бита информации. Следовательно, $i_{\text{белого}} = \log_2 \frac{1}{p_{\text{белого}}} \Rightarrow p_{\text{белого}} = \frac{1}{2^2} = \frac{1}{4}$ (1).

С другой стороны, вероятность того, что из корзины извлекли белый шар, равна:

$$p_{\text{белого}} = \frac{k_{\text{белых шаров}}}{n_{\text{общее количество шаров}}} = \frac{X}{X+18} \quad (2).$$

Приравняем формулы (1) и (2). Получаем: $\frac{X}{X+18} = \frac{1}{4}$. Следовательно, X=6.

Тогда общее количество шаров в корзине равно X+18=6+18=24.

Задача 12. Для 5 букв латинского алфавита заданы их двоичные коды (для некоторых букв - из двух бит, для некоторых - из трех). Эти коды представлены в таблице:

a	b	c	d	e
000	110	01	001	10

Определите, какая последовательность букв закодирована двоичной строкой 1100000100110.

Решение:

Так как ни один код символа не является началом другого кода, то нам дан префиксный код, который подлежит однозначному декодированию.

Проведем декодирование:

1100000100110 → 110 | 0000100110 → b | 0000100110 → b | 000 | 0100110 → ba | 0100110 → ba | 01 | 00110 → bac | 00110 → bac | 001 | 10 → bacde

Задача 13. Для 5 букв латинского алфавита заданы их двоичные коды (для некоторых букв – из двух бит, для некоторых – из трех). Эти коды представлены в таблице:

a	b	c	d	e
100	110	011	01	10

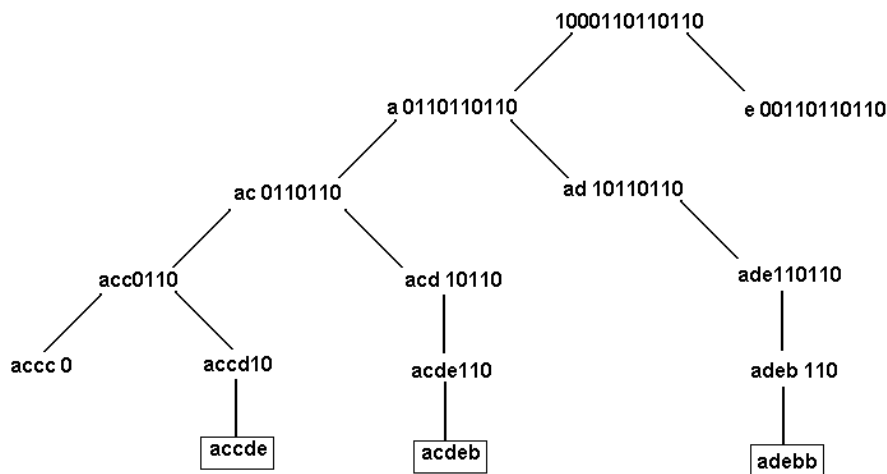
Определите, сколькими способами можно провести успешное декодирование строки 1000110110110

- 1) cbade 2) acdeb 3) acbed 4) bacde

Решение:

В отличие от предыдущей задачи данный код не является префиксным – код буквы d является началом кода буквы c, а код буквы e – началом кода буквы a. Следовательно, однозначно декодировать сообщение невозможно.

Построим возможные варианты декодирования:



Таким образом, успешное декодирование можно провести тремя способами.

Задача 14. Для передачи по каналу связи сообщения, состоящего только из букв А, Б, В, Г, решили использовать неравномерный по длине код: А=1, Б=01, В=001. Как нужно закодировать букву Г, чтобы длина кода была минимальной и допускалось однозначное разбиение кодированного сообщения на буквы?

Решение:

В данной задаче речь идет о построении неравномерного префиксного кода. Следовательно, нам нужно восстановить дерево Хаффмана и определить позицию буквы Г в этом дереве, что поможет определить нам ее код. Но строить дерево Хаффмана мы начнем с корня, а не с листьев.

