
Лекция 2

Объемный подход к измерению информации



Объемный (алфавитный) подход к измерению информации

Объемный подход используется для измерения количества информации в тексте, представленном в виде последовательности символов некоторого **алфавита**.

Количество символов в алфавите с помощью которого представлен текст называется ***мощностью алфавита***.

Предполагается, что все символы в тексте встречаются с **равной** вероятностью

Сколько можно составить кодовых комбинаций длиной в 2 символа, используя двоичную систему?

Сколько можно составить кодовых комбинаций длиной в 3 символа, используя двоичную систему?

Количество
возможных
сообщений

Длина кода
сообщения

$$N = a^i$$

0 - 000
1 - 001
2 - 010
3 - 011
4 - 100
5 - 101
6 - 110
7 - 111

$N = 8$
 $a = 2$
 $i = 3$

Количество букв
в алфавите,
который
использовался
для записи
сообщения

Сколько можно составить кодовых комбинаций длиной в 2 символа, используя троичную систему?

Кодовой комбинации какой длины достаточно, чтобы при равномерном кодировании закодировать 18 букв двоичным кодом?

Пример 1

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

Пример 2

Световое табло состоит из трёх светящихся элементов, каждый из которых может светиться одним из пяти различных цветов. Каждая комбинация из трёх цветов кодирует определённый сигнал. Сколько различных сигналов можно передать при помощи табло при условии, что все элементы должны светиться?

Пример 3

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

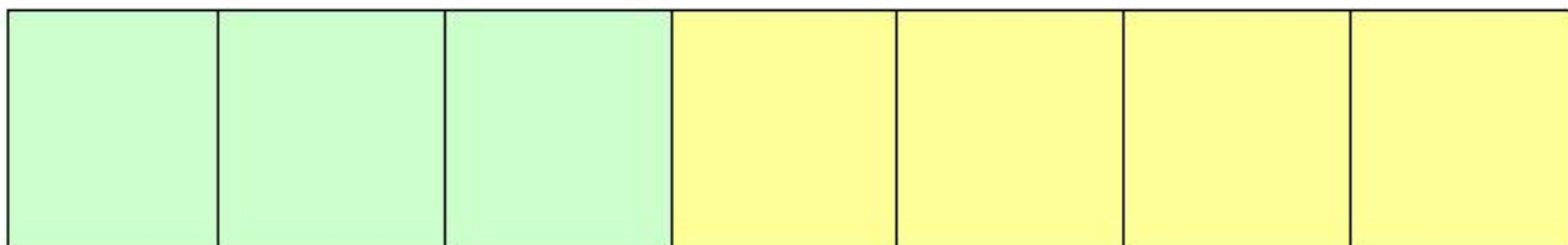
Если слово состоит из i букв, причем есть n_1 вариантов выбора первой буквы, n_2 вариантов выбора второй буквы и т.д., то число возможных слов вычисляется как произведение

$$N = n_1 * n_2 * \dots * n_i$$

a_1 - мощность алфавита подслова 1

i_1 - длина подслова 1

n_1 - количество вариантов подслова 1



a_2 - мощность алфавита подслова 1

i_2 - длина подслова 1

n_2 - количество вариантов подслова 1

Количество вариантов слова, составленного из подслова 1 и подслова 2

$$N = n_1 \cdot n_2$$

Пример 4

На световой панели в ряд расположены 7 лампочек. Каждая из первых двух лампочек может гореть красным, жёлтым или зелёным цветом. Каждая из остальных пяти лампочек может гореть одним из двух цветов - красным или белым. Сколько различных сигналов можно передать с помощью панели (все лампочки должны гореть, порядок цветов имеет значение)?

Пример 5

Сколько слов длины 4, начинающихся с согласной буквы, можно составить из букв Л, Е, Т, О? Каждая буква может входить в слово несколько раз. Слова не обязательно должны быть осмысленными словами русского языка.

--	--	--	--

Если из алфавита мощностью a составляют слова из i_1 букв, i_2 букв и т.д., то число возможных слов вычисляется как сумма

$$N = a^{i_1} + a^{i_2} + \dots$$

Пример 6

Азбука Морзе позволяет кодировать символы для сообщений по радиосвязи, задавая комбинацию точек и тире. Сколько различных символов (цифр, букв, знаков пунктуации и т. д.) можно закодировать, используя код азбуки Морзе длиной **не менее четырёх и не более пяти** сигналов (точек и тире)?

Пример 7

Сколько существует различных символьных последовательностей длины 5 в четырёхбуквенном алфавите {A, C, G, T}, которые содержат ровно две буквы A?

Пример 8 (демо 2017)

Вася составляет 5-буквенные слова, в которых встречаются только буквы А, Б, В, Г, причём буква А появляется ровно 1 раз. Каждая из других допустимых букв может встречаться в слове любое количество раз или не встречаться совсем. Словом считается любая допустимая последовательность букв, не обязательно осмысленная. Сколько существует таких слов, которые может написать Вася?

Единицы измерения информации

При $N = 2$ (двоичный алфавит) информационный вес одного символа равен 1 биту.

*С позиции объемного подхода к измерению информации **1 бит** — это информационный вес символа из двоичного алфавита.*

1 байт = 8 бит — это информационный вес символа из алфавита мощностью 256.

1 КБ (килобайт) = 2^{10} байт = 1024 байта,

1 МБ (мегабайт) = 2^{10} КБ = 1024 КБ,

1 ГБ (гигабайт) = 2^{10} МБ = 1024 МБ

1 ТБ (терабайт) = 2^{10} ГБ = 1024 ГБ

1 ПБ (петабайт) = 2^{10} ТБ = 1024 ТБ

1 ЭБ (эксабайт) = 2^{10} ПБ = 1024 ПБ

1 ЗБ (зеттабайт) = 2^{10} ЭБ = 1024 ЭБ

1 ЙБ (йоттабайт) = 2^{10} ЗБ = 1024 ЗБ

1 МБ = ...бит

1 КБ = ...бит

Кодирование текстовых сообщений ДВОИЧНЫМ КОДОМ

При кодировании сообщения с
ПОМОЩЬЮ ДВОИЧНОГО АЛФАВИТА

$$N = 2^i$$



Определение количества информации в сообщении, закодированном двоичным кодом

- 1) Вычисляют количество информации, которое несет каждый символ текста (**информационный вес символа**)

$$i = \log_2 N,$$

N — мощность алфавита

- 2) Общее количество информации (**информационный объем текста**) определяется как сумма информационных весов всех символов, составляющих текст.

$$I = K \cdot i$$

K — количество символов в тексте

-
- Найти информационный объем текста, записанного на языке, алфавит которого содержит 128 символов и 2000 символов в сообщении



Имеется сообщение объемом 2^{23} бит. В мегабайтах объем этого сообщения равен ...

- 64
 - 1
 - 1024
 - 8
-

Объем сообщения, содержащего 4096 символов, равен $\frac{1}{512}$ части Мбайта. Какова мощность алфавита, с помощью которого записано это сообщение?

- 1) 8
 - 2) 16
 - 3) 4096
 - 4) 16384
-

Пример 9

В велокроссе участвуют 119 спортсменов. Специальное устройство регистрирует прохождение каждым из участников промежуточного финиша, записывая его номер с использованием минимально возможного количества бит, одинакового для каждого спортсмена. Каков информационный объем в битах сообщения, записанного устройством, после того как промежуточный финиш прошли 70 велосипедистов?

Пример 10

В некоторой стране автомобильный номер длиной 6 символов составляют из заглавных букв (используются только 33 различных буквы) и десятичных цифр в любом порядке. Каждый такой номер в компьютерной программе записывается минимально возможным и одинаковым целым количеством байтов (при этом используют посимвольное кодирование и все символы кодируются одинаковым и минимально возможным количеством битов). Определите объём памяти в байтах, отводимый этой программой для записи 125 номеров.

Пример 11

В некоторой стране автомобильный номер состоит из 8 символов. Первый символ – одна из 26 латинских букв, остальные семь – десятичные цифры. Пример номера – A1234567. Каждый символ кодируется минимально возможным количеством бит, а каждый номер – одинаковым и минимально возможным целым количеством байт. Определите объем памяти в байтах, необходимый для хранения 30 автомобильных номеров.

Кодирование растровой графики ДВОИЧНЫМ КОДОМ

Количество
возможных
цветов в
изображении

$$N = 2^i$$

Длина кода
цвета 1 точки
(глубина цвета)

Количество букв
в алфавите,
который
использовался
для записи
изображения

Объем памяти для
хранения
изображения

$$I = k \cdot i$$

Количество
точек
(пикселей) в
изображении

Длина кода
цвета 1 точки
(глубина цвета)

Пример 12

Рисунок размером 512 на 256 пикселей занимает в памяти 64 Кбайт (без учёта сжатия). Найдите максимально возможное количество цветов в палитре изображения.

Пример 13

Демо 2017

Для хранения произвольного растрового изображения размером 1024×1024 пикселей отведено 512 Кбайт памяти, при этом для каждого пикселя хранится двоичное число – код цвета этого пикселя. Для каждого пикселя для хранения кода выделено одинаковое количество бит. Сжатие данных не производится.

Какое максимальное количество цветов можно использовать в изображении?

Пример 14

- Какой минимальный объём памяти (в Кбайт) нужно зарезервировать, чтобы можно было сохранить любое растровое изображение размером 64×64 пикселей при условии, что в изображении могут использоваться 256 различных цветов? В ответе запишите только целое число, единицу измерения писать не нужно.
-

Кодирование звуковой информации ДВОИЧНЫМ КОДОМ

Дискретизация – представление аналогового сигнала в виде набора чисел, т.е. запись в память только значений сигнала в отдельных точках, взятых с некоторым шагом T по времени.

Число T называется **интервалом дискретизации**, а обратная ему величина **$f=1/T$ – частотой дискретизации**

Частота дискретизации определяет количество отсчетов, запоминаемых за 1 секунду; 1 Гц (один герц) – это один отсчет в секунду, а 8 кГц – это 8000 отсчетов в секунду.

Разрядность кодирования (i) (глубина кодирования, битрейд) – это число битов, используемое для кодирования одного отсчета

Если хранения одного отсчета отводится $i=3$ бита, то код каждого отсчета это целое число от 0 до 7 (от 000 до 111).

Диапазон возможных значений сигнала делится на 8 полос (уровней дискретизации), каждой из которых присваивается номер (код).

Все отсчеты, попавшие в одну полосу, получают одинаковый код.

Уровни дискретизации (N) – количество разных звуков, которые используются для записи мелодии

Количество
разных звуков,
которые можно
использовать
при записи
(уровни
дискретизации)

$$N = 2^i$$

Длина кода
одного отсчета
(глубина
кодирования,
битрейт)

Количество букв
в алфавите,
который
использовался
для записи
изображения

Объем памяти для хранения звукового файла

$$I = f \cdot i \cdot t$$

Количество отсчетов в секунду
(частота дискретизации)

Длина кода одного отсчета
(битрейт)

Количество секунд в записи

При $f = 8$ кГц, глубине кодирования 16 бит на отсчёт и длительности звука 128 секунд требуется

$$I = 8000 \cdot 16 \cdot 128 = 16384000 \text{ бит}$$

$$I = 8000 \cdot 16 \cdot 128 / 8 = 2048000 \text{ байт}$$

$$I = 8000 \cdot 16 \cdot 128 / 8 / 1024 = 2000 \text{ Кбайт}$$

$$I = 8000 \cdot 16 \cdot 128 / 8 / 1024 / 1024 \approx 1,95 \text{ Мбайт}$$

Для упрощения ручных расчетов можно использовать приближённые равенства

$$1 \text{ мин} = 60 \text{ сек} \approx 64 \text{ сек} = 2^6 \text{ сек}$$

$$1000 \approx 1024 = 2^{10}$$

нужно помнить, что

$$1 \text{ Мбайт} = 2^{20} \text{ байт} = 2^{23} \text{ бит},$$

$$1 \text{ Кбайт} = 2^{10} \text{ байт} = 2^{13} \text{ бит}$$

при двухканальной записи (стерео) объем памяти, необходимый для хранения данных одного канала, умножается на 2

Пример 15

Производится одноканальная (моно) звукозапись с частотой дискретизации 48 кГц и глубиной кодирования 16 бит. Запись длится 2 минуты 8 секнд, ее результаты записываются в файл, сжатие данных не производится. Какое из приведенных ниже чисел наиболее близко к размеру полученного файла, выраженному в мегабайтах?

- 1) 11 2) 12 3) 13 4) 20

Пример 16

Производится одноканальная (моно) звукозапись с частотой дискретизации 128 Гц. При записи использовались 64 уровня дискретизации. Запись длится 4 минут 16 секунд, её результаты записываются в файл, причём каждый сигнал кодируется минимально возможным и одинаковым количеством битов. Какое из приведённых ниже чисел наиболее близко к размеру полученного файла, выраженному в килобайтах?

- 1) 24 2) 36 3) 128 4) 384

Пример 17

Проводилась одноканальная (моно) звукозапись с частотой дискретизации 16 кГц и 24-битным разрешением. В результате был получен файл размером 3 Мбайт, сжатие данных не производилось. Какая из приведенных ниже величин наиболее близка к времени, в течение которого проводилась запись?

- 1) 30 сек 2) 60 сек 3) 90 сек 4) 120 сек