



Подходы

к измерению информации





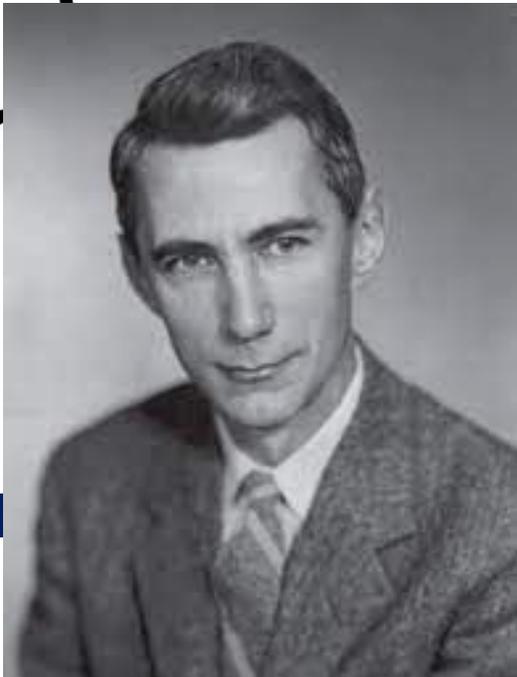
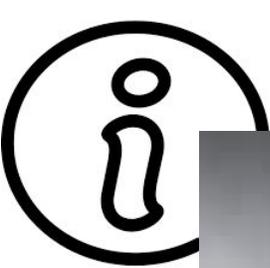
ИЗМЕРЕНИЕ ИНФОРМАЦИИ



**Содержательный подход к
измерению информации**



**Алфавитный подход к
измерению информации**



Клод Элвуд Шеннон
(1916–2001 гг.)

Содержательный подход

Согласно Шеннону, информативность сообщения характеризуется содержащейся в нем полезной информацией — той частью сообщения, которая снимает полностью или уменьшает **неопределенность** какой-либо ситуации.

В 1948 году предложил использовать слово **«бит»** для обозначения наименьшей единицы информации.

Сообщение, уменьшающее неопределенность знаний человека в 2 раза, несёт 1 бит информации



Содержательный подход к измерению информации

$$N = 2^i$$

N – количество возможных событий

i – количество информации

Сообщение, уменьшающее неопределенность знаний человека в 2 раза, несёт 1 бит информации



Задача 1. Монету подбрасывают вверх над ровной поверхностью.

Чему равно количество информации в сообщении о том, что монета упала стороной «орёл» вверх?

Дано:

$$N = 2$$

$$i = ?$$

Ответ: 1 бит.

Решение:

$$N = 2^i$$

$$2 = 2^i$$

$$i = 1 \text{ бит.}$$

Возможные события



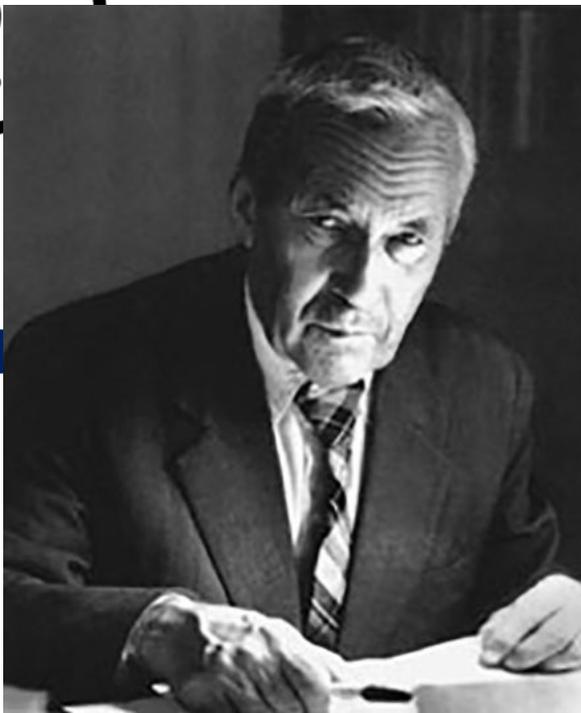
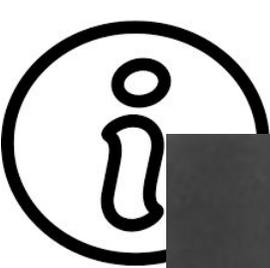
Произошедшее событие





Задачи для тренировок

1. Сообщение о том, что ваш друг живёт на 10 этаже, несёт 4 бита информации. Сколько этажей в доме?
2. В соревновании принимают участие 300 спортсменов. Для реализации базы данных необходимо закодировать номер каждого спортсмена. Какое наименьшее целое количество бит необходимо для кодирования номера спортсмена?



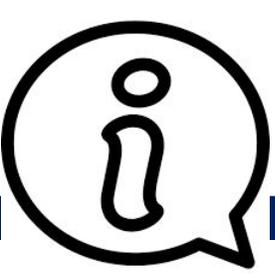
**Андрей Николаевич
Колмогоров
(1903-1987 гг.)**

Алфавитный подход

Количество информации, представленное в сообщении, **не** зависит от содержания сообщения.

Алфавит — упорядоченный набор символов, используемый для кодирования сообщений на некотором языке.

Мощность алфавита — количество символов алфавита.



Алфавитный подход

Чтобы определить объем информации в сообщении при алфавитном подходе, нужно последовательно решить задачи:

1. Определить количество информации (i) в одном символе по формуле

$$2^i = N,$$

где N — мощность алфавита

2. Определить количество символов в сообщении (K)

Вычислить объем информации по формуле: $I = K * i$



Задача 2. Сообщение, записанное буквами 32-символьного алфавита, содержит 180 символов. Какое количество информации оно несёт?

Дано:

$$N = 32$$

$$K = 180$$

$$I = ?$$

Решение:

$$I = K * i$$

$$N = 2^i$$

$$32 = 2^i$$

$$25 = 2^i, \text{ т.о. } i = 5,$$

$$I = 180 * 5 = 900 \text{ бит.}$$

Ответ: $I = 900$ бит.

Итак, информационный вес всего сообщения равен 900 бит.



Задача 3. Определите информационный вес сообщения:
Сегодня хорошая погода

Нужно знать:

1. количество символов в сообщении,
2. информационный вес одного символа
3. мощность алфавита

Решение:

$$I = 22 * 8 = 176 \text{ бита.}$$

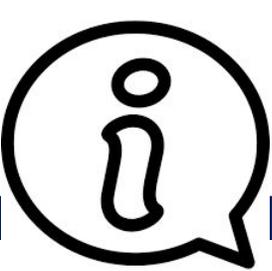
Ответ: сообщение весит 176 бита.

*нужно сосчитать количество символов в этом сообщении и умножить
это число на восемь*



Задачи для тренировок

1. Алфавит состоит из 100 символов. Какое количество информации несет один символ этого алфавита?
2. Сообщение, записанное буквами из 64-символьного алфавита, содержит 20 символов. Какой объем информации оно несет?



Единицы измерения информации

Бит – наименьшая единица измерения объема информации.

| Название | Условное обозначение | Соотношение |
|-----------------|-----------------------------|--|
| Байт | Байт | $1 \text{ байт} = 2^3 \text{ бит} = 8 \text{ бит}$ |
| КилоБайт | Кб | $1 \text{ Кб} = 2^{10} \text{ байт} = 1024 \text{ байт}$ |
| МегаБайт | Мб | $1 \text{ Мб} = 2^{10} \text{ Кб} = 1024 \text{ Кб}$ |
| ГигаБайт | Гб | $1 \text{ Гб} = 2^{10} \text{ Мб} = 1024 \text{ Мб}$ |
| ТераБайт | Тб | $1 \text{ Тб} = 2^{10} \text{ Гб} = 1024 \text{ Гб}$ |



Задача 3. . Получено сообщение, информационный объем которого равен 32 битам. Чему равен этот объем в байтах?

Решение:

Ответ: 4 байта.

1 байт= 8 бит.

$32:8=4$

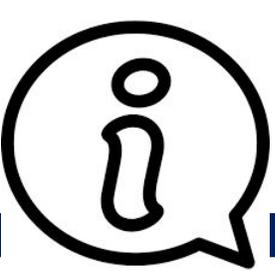
Задачи для тренировок

1. Компьютер имеет оперативную память 512 Мб. Количество соответствующих этой величине бит больше:

- 1) 10 000 000 000бит
- 2) 8 000 000 000бит
- 3) 6 000 000 000бит
- 4) 4 000 000 000бит

2. Считая, что каждый символ кодируется одним байтом, оцените информационный объем следующего предложения:

Мой дядя самых честных правил



Информационные объекты различных видов

Информационный объект – это обобщающее понятие, описывающее различные виды объектов; это предметы, процессы, явления материального и нематериального свойства, рассматриваемые с точки зрения их информационных свойств.

Информационный объект:

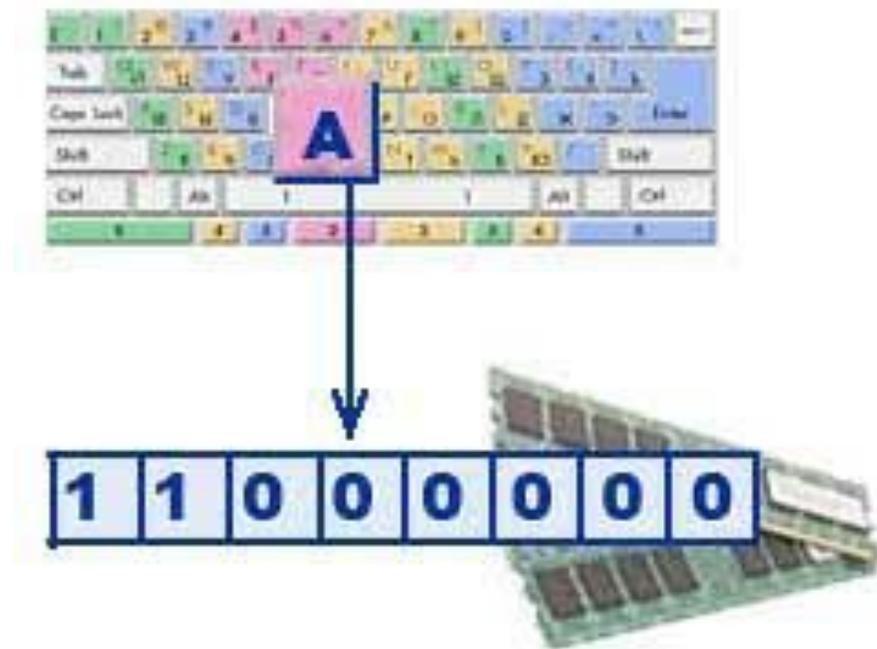
1. Обладает определенными потребительскими качествами (т.е. он нужен пользователю)
2. Допускает хранение на цифровых носителях
3. Допускает над ними определенные действия путем использования аппаратных и программных средств компьютера



Дискретное (цифровое) представление текстовой информации

Текстовая информация дискретна – состоит из отдельных знаков.

Для обработки текстовой информации на компьютере необходимо представить ее в двоичной знаковой системе. Каждому знаку необходимо поставить в соответствие уникальный 8-битовый двоичный код, значения которого находятся в интервале от 00000000 до 11111111 (в десятичном коде от 0 до 255).





Дискретное (цифровое) представление графической информации

Изображение на экране монитора составляется из отдельных точек – пикселей.

Пиксель – минимальный участок изображения, которому независимым образом можно задать цвет.

Количество цветов N в палитре и количество информации I , необходимое для кодирования цвета каждой точки вычисляется по формуле:

$$N = 2^i$$



Задача 4. Наиболее распространенными значениями глубины цвета при кодировании цветных изображений являются 4, 8, 16 или 24 бита на точку. Можно определить количество цветов в 24-битной палитре: $N = 2^i = 2^{24} = 1677721$ бит.

Задача 5. Какой минимальный объём памяти (в Кбайт) нужно зарезервировать, чтобы можно было сохранить любое растровое изображение размером 128×128 пикселей при условии, что в изображении могут использоваться 256 различных цветов? В ответе запишите только целое число, единицу измерения писать не нужно.

Решение:

Ответ: 16 Кбайт.

Один пиксель кодируется 8 битами памяти, так как $2^8 = 256$.

Всего $128 * 128 = 2^7 \cdot 2^7 = 2^{14}$ пикселей.

Тогда объём памяти, занимаемый изображением $2^{14} * 8 = 2^{17}$ бит = 2^{14} байт = 2^4 Кбайт = 16 Кбайт.



Дискретное (цифровое) представление звуковой информации

Частота дискретизации звука – это количество измерений громкости звука на одну секунду.

Глубина кодирования звука – это количество информации, которое необходимо для кодирования дискретных уровней громкости цифрового звука.

Если известна глубина кодирования, то количество уровней громкости цифрового звука можно рассчитать по формуле:

$$N = 2^i$$



Задача 6. Производится звукозапись музыкального фрагмента в формате стерео (двухканальная запись) с частотой дискретизации 32 кГц и 32-битным разрешением.

Результаты записываются в файл, сжатие данных не производится; размер полученного файла 40 Мбайт. Затем производится повторная запись этого же фрагмента в формате моно (одноканальная запись) с частотой дискретизации 16 кГц и 16-битным разрешением. Сжатие данных не производилось.

Укажите размер файла в Мбайт, полученного при повторной записи.

В ответе запишите только целое число, единицу измерения писать не нужно.

Решение:

Ответ: 5 Мбайт

40 (Мбайтах) = 2 (количество каналов) · 32000 (частота) · 4 (разрешение в байтах) · X (время в секундах)

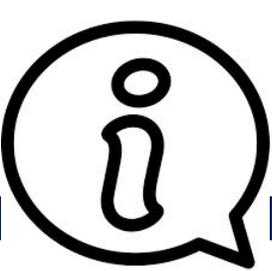
V (Мбайтах) = 1 (количество каналов) · 16000 (частота) · 2 (разрешение в байтах) · X (время в секундах)

Тогда, $v = 40 / (2 \cdot 2 \cdot 2) = 5$ Мбайт.



Дискретное (цифровое) представление видеоинформации

Видеоинформация – это сочетание звуковой и графической информации. Кроме того, для создания на экране эффекта движения используется дискретная технология быстрой смены статистических картинок.



Решение задач

1. Валя шифрует русские слова, записывая вместо каждой буквы её код.

А В Д О Р У

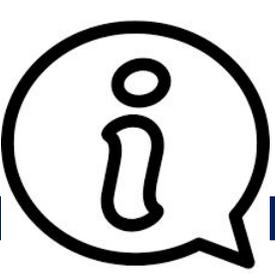
01 011100 111010 001

Некоторые цепочки можно расшифровать не одним способом. Например, 00101001 может означать не только УРА, но и УАУ.

Даны три кодовые цепочки:

11101001 010111011 01001010

Найдите среди них ту, которая имеет только одну расшифровку и запишите в ответе расшифрованное слово.



Решение задач

2. В кодировке КОИ-8 каждый символ кодируется 8 битами. Паша написал текст (в нём нет лишних пробелов):

«Аки, Бали, Банда, Сибуян, Камотес, Лабрадор, Линкольна — моря».

Ученик вычеркнул из списка название одного из морей. Заодно он вычеркнул ставшие лишними запятые и пробелы — два пробела не должны идти подряд.

При этом размер нового предложения в данной кодировке оказался на 7 байтов меньше, чем размер исходного предложения. Напишите в ответе вычеркнутое название моря.



Решение задач

3. Статья, набранная на компьютере, содержит 48 страниц, на каждой странице 40 строк, в каждой строке 40 символов. Определите размер статьи в кодировке КОИ-8, в которой каждый символ кодируется 8 битами.

4. В одной из кодировок Unicode каждый символ кодируется 16 битами.

Определите размер следующего предложения в данной кодировке:

Слух обо мне пройдёт по всей Руси великой.

Спасибо за внимание!