

# Измерение информации



# Подходы измерения информации

В информатике используют различные подходы к измерению информации :

- содержательный подход
- алфавитный подход
- вероятностный подход

# Содержательный подход к измерению информации

В содержательном подходе количество информации, заключенное в сообщение, определяется объемом знаний, который это сообщение несет получающему его человеку.

Если сообщение не информативно, то количество информации с точки зрения человека = 0

В содержательном подходе используется формула Хартли

# Вероятностный подход к измерению информации

Вероятностью случайного события ( $p$ ) называется отношение числа благоприятствующих событию исходов ( $m$ ) к общему числу исходов ( $n$ ):  $p=m:n$

Вероятностный подход предполагает, что возможные события имеют различные вероятности реализации.

Первая формула вероятностного подхода к измерению количества информации была предложена в 1928 г. **Ральфом Хартли**, вторая формула - в 1948 году - **Клодом Шенноном**.

$$I = \log_2 N$$

$$N = 2^I$$

$I$  – количество информации;

$N$  – количество возможных событий;

*Все события – равновероятны.*

$$I = -\sum_{i=1}^N p_i \log_2 p_i$$

$I$  – количество информации;

$N$  – количество возможных событий;

$p_i$  - вероятность  $i$ -го события.

# Пример

*Шахматная доска состоит из 64 полей: 8 столбцов на 8 строк. Какое количество бит несет сообщение о выборе одного шахматного поля?*

*Решение.*

*Поскольку выбор любой из 64 клеток равновероятен, то количество бит находится из формулы:*

*$2^i=64$ ,  $i=\log_2 64=6$ , так как  $2^6=64$ . Следовательно,  $i=6$  бит.*

# Алфавитный подход к измерению информации

**Алфавитный подход** позволяет измерять количество информации в тексте (символьном сообщении), составленном из символов некоторого алфавита.

Алфавитный подход удобен при подсчете количества информации, хранимого, передаваемого и обрабатываемого техническими устройствами.

**Алфавит** — это набор букв, знаков, цифр, скобок и т.д. Количество символов в алфавите называется его **мощностью**

При алфавитном подходе считается, что каждый символ текста имеет определенный **информационный вес**. Информационный вес символа зависит от мощности алфавита.

# Пример

*Считая, что каждый символ кодируется одним байтом, оцените информационный объем следующего предложения: **Белеет Парус Одинокий В Тумане Моря Голубом!***

**Решение.**

Так как в предложении 44 символа (считая знаки препинания и пробелы), то информационный объем вычисляется по формуле:

$$V=44 \cdot 1 \text{ байт}=44 \text{ байта}=44 \cdot 8 \text{ бит}=352 \text{ бита}$$

# Пример

*В беспроигрышной лотерее разыгрывается 3 книги, 2 альбома, 10 наборов маркеров, 10 блокнотов. Какова вероятность выиграть книгу?*

*Решение*

*Общее число исходов  $2+3+10+10=25$ ; число благоприятствующих исходу событий равно 3. Вероятность выигрыша книги вычисляется по формуле:  $p=3:25=0,12$*



# Единицы измерения информации

**Бит** - наименьшая единица измерения, которую ввел американский инженер и математик Клод Шеннон

Бит может принимать одно из двух значений - 0 или 1

**Байт** - это единица измерения информации

Умножаем

1 байт = 8 бит

1 килобайт = 1024 байт

1 мегабайт = 1024 килобайт

1 гигабайт = 1024 мегабайт

1 терабайт = 1024 гигабайт

1 петабайт = 1024 терабайт

1 эксабайт = 1024 петабайт

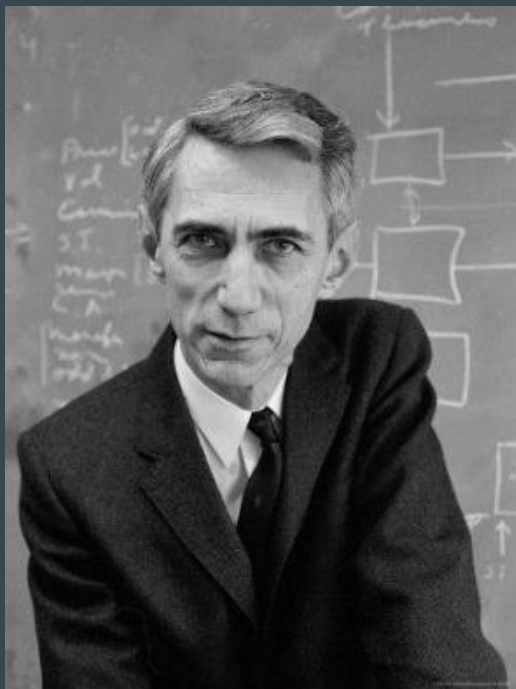
1 зеттабайт = 1024 эксабайт

1 йоттабайт = 1024 зеттабайт

Делим

Название	Символ	Степень
байт	Б	$2^0$
килобайт	кБ	$2^{10}$
мегабайт	МБ	$2^{20}$
гигабайт	ГБ	$2^{30}$
терабайт	ТБ	$2^{40}$
петабайт	ПБ	$2^{50}$
эксабайт	ЭБ	$2^{60}$
зеттабайт	ЗБ	$2^{70}$
йоттабайт	ЙБ	$2^{80}$

# Формула Шеннона



**К.Шеннон** в 1948 году предложил формулу для вычисления информации в случае различных вероятностей событий

$$I = -\sum_{i=1}^N p_i \log_2 p_i,$$

$I$  - количество информации

$N$  - количество возможных событий

$p_i$  - вероятность  $i$ -го события.

|

## Например:

*Пусть при бросании несимметричной четырехгранной пирамидки вероятности отдельных событий будут равны:*

$$P_1 = 1/2, p_2 = 1/4, p_3 = 1/8, p_4 = 1/8.$$

Тогда количество информации, которое мы получим после реализации одного из них, можно рассчитать по формуле.

$$I = -(1/2 \log_2 1/2 + 1/4 \log_2 1/4 + 1/8 \log_2 1/8 + 1/8 \log_2 1/8) = (1/2 + 2/4 + 3/8 + 3/8) \text{ битов} = 14/8 \text{ битов} = 1,75 \text{ бита.}$$