

## Подходы к измерению информации

**Подходы к измерению информации** — выделяемые в информатике два подхода к измерению информации, основанные на определении К. Шеннона, применяемое в математической теории информации, и определении А. Н. Колмогорова, применяемое в отраслях информатики, связанных с использованием компьютеров:

1. **Содержательный (вероятностный).** Количество информации можно рассматривать как меру уменьшения неопределённости знания при получении информационных сообщений.
2. **Объёмный (алфавитный).** Это объективный, количественный метод для измерения информации, циркулирующей в информационной технике.

### Содержательный подход

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

По Шеннону, информация — уменьшение неопределенности наших знаний.

Неопределенность некоторого события — это количество возможных исходов данного события.

Так, например, если из колоды карт наугад выбирают карту, то неопределенность равна количеству карт в колоде. При бросании монеты неопределенность равна 2.

Содержательный подход часто называют субъективным, так как разные субъекты информацию об одном и том же предмете оценивают по-разному.

Но если число исходов не зависит от суждений людей (случай бросания кубика или монеты), то информация о наступлении одного из возможных исходов является объективной.

Если сообщение уменьшило неопределенность знаний ровно в два раза, то говорят, что сообщение несет 1 бит информации.

1 бит — объем информации такого сообщения, которое уменьшает неопределенность знания в два раза.

Рассмотрим, как можно подсчитать количество информации в сообщении, используя содержательный подход.

Пусть в некотором сообщении содержатся сведения о том, что произошло одно из  $N$  равновероятных (равновозможных) событий. Тогда количество информации  $i$ , заключенное в этом сообщении, и число событий  $N$  связаны формулой:  $2^i = N$ . Эта формула носит название формулы Хартли. Получена она в 1928 г. американским инженером Р. Хартли.

Если  $N$  равно целой степени двойки (2, 4, 8, 16 и т. д.), то вычисления легко произвести «в уме». В противном случае количество информации становится нецелой величиной, и для решения задачи придется воспользоваться таблицей логарифмов либо определять значение логарифма приблизительно (ближайшее целое число, большее).

Например, если из 256 одинаковых, но разноцветных шаров наугад выбрали один, то сообщение о том, что выбрали красный шар несет 8 бит информации ( $2^8=256$ ).

Для угадывания числа (наверняка) в диапазоне от 0 до 100, если разрешается задавать только двоичные вопросы (с ответом «да» или «нет»), нужно задать 7 вопросов, так как объем информации о загаданном числе больше 6 и меньше 7 ( $2^6 < 100 < 2^7$ ).

## **Алфавитный подход**

Алфавитный подход основан на том, что всякое сообщение можно закодировать с помощью конечной последовательности символов некоторого алфавита.

Двоичный алфавит содержит 2 символа, его мощность равна двум. Сообщения, записанные с помощью символов ASCII, используют алфавит из 256 символов. Сообщения, записанные по системе UNICODE, используют алфавит из 65 536 символов.

С позиций computer science носителями информации являются любые последовательности символов, которые хранятся, передаются и обрабатываются с помощью компьютера. Согласно Колмогорову, информативность последовательности символов не зависит от содержания сообщения, алфавитный подход является объективным, то есть он не зависит от субъекта, воспринимающего сообщение. 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:  $I = K * i$ .

## **Единицы измерения информации**

В информатике используется величина, называемая байтом (byte) и равная 8 битам. И если бит позволяет выбрать один вариант из двух возможных, то байт, соответственно, 1 из 256 ( $2^8$ ). В большинстве современных ЭВМ при кодировании каждому символу соответствует своя последовательность из восьми нулей и единиц, то есть байт.

Наряду с байтами для измерения количества информации используются более крупные единицы

$$1 \text{ Кбайт (один килобайт)} = 2^{10} \text{ байт} = 1024 \text{ байта};$$

$$1 \text{ Мбайт (один мегабайт)} = 2^{10} \text{ Кбайт} = 1024 \text{ Кбайта};$$

$$1 \text{ Гбайт (один гигабайт)} = 2^{10} \text{ Мбайт} = 1024 \text{ Мбайта}.$$

В последнее время в связи с увеличением объёмов обрабатываемой информации входят в употребление такие производные единицы, как

$$1 \text{ Терабайт (Тб)} = 1024 \text{ Гбайта} = 2^{40} \text{ байта}$$

$$1 \text{ Петабайт (Пб)} = 1024 \text{ Тбайта} = 2^{50} \text{ байта}.$$