

Определение количества информации

Как измерить информацию?

Вопрос этот очень непростой.

Ответ на него зависит от того, что понимать под информацией. Но поскольку определять информацию можно по-разному, то и **способы измерения** тоже могут быть разными.



ИНФОРМАЦИЯ

```
graph TD; A[Измерение информации] --- B[Содержательный подход]; A --- C[Алфавитный подход]
```

Измерение
информации

Содержательный
подход

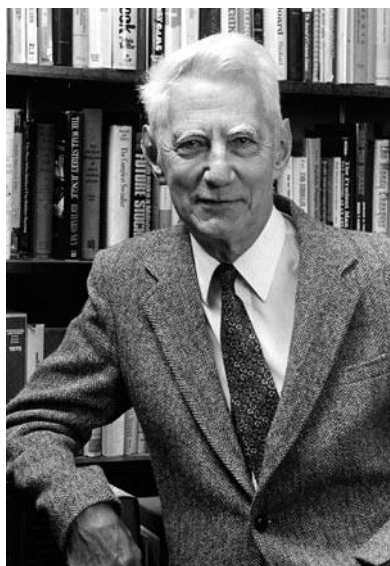
Алфавитный
подход

Содержательный подход к измерению информации.

Для человека **информация** — это знания. Если получение новой информации приводит к расширению знаний, то можно говорить, что такое сообщение **содержит** информацию.

Говорят, что **сообщение информативно если оно пополняет знания** человека.

Например, прогноз погоды на завтра — информативное сообщение, а сообщение о вчерашней погоде неинформативно, т.к. нам это уже известно.



*Основоположником этого подхода является американский учёный Клод Элвуд Шеннон(1916 — 2001). 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 к измерению информации.

Единица измерения информации была определена в науке, которая называется теорией информации. Эта единица носит название «**бит**». Ее определение звучит так:

Сообщение, уменьшающее неопределенность знаний в два раза, несет 1 бит информации.

Неопределенность знаний о некотором событии — это количество возможных результатов события.

Тогда можно записать формулу:

$$2^i = N$$

N - количество событий

i - количество информации одного события

Пример:

На книжном стеллаже восемь полок. Книга может быть поставлена на любую из них. Сколько информации содержит сообщение о том, где находится книга?

Решение:

$$N = 8. \ i - ?$$

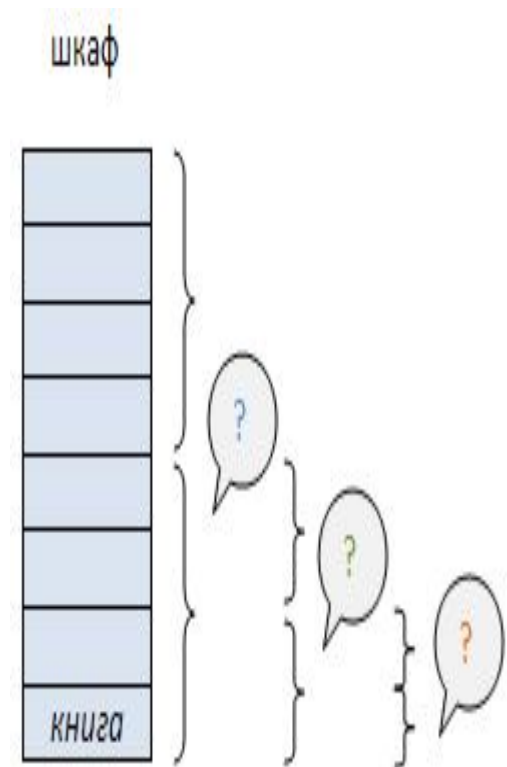
$$2^i = N$$

$$2^i = 8$$

$$2^i = 2^3$$

$$i = 3 \text{ бита}$$

Ответ: сообщение о том, что книга находится на любой из полок равно 3 бита



Алфавитный подход к измерению информации

Алфавитный подход к измерению количества информации основан на подсчете числа символов в сообщении.

При алфавитном подходе к определению количества информации **отвлекаются от содержания** информации и рассматривают информационное сообщение как **последовательность знаков** определенной знаковой системы.

Все множество используемых в языке **символов** будем традиционно называть алфавитом.

Обычно под алфавитом понимают только буквы, но поскольку в тексте могут встречаться знаки препинания, цифры, скобки, то мы их тоже включим в алфавит. В алфавит также следует включить и пробел, т.е. пропуск между словами.

Полное количество символов алфавита принято называть мощностью алфавита.

В формуле

$$2^i = N$$

N - мощность алфавита

i - количество информации одного символа

Алфавитный подход к измерению информации

При алфавитном подходе к измерению информации количество информации зависит не от содержания, а от размера текста и мощности алфавита.

Информационный объем текста (I), содержащего K символов вычисляют по формуле:

$$I = K * i$$

где I - информационный объем текста,
 K - количество символов в тексте,
 i - информационный объем одного символа.



Основоположником этого подхода является Андрей Николаевич Колмогоров, (1903-1987), великий российский ученый-математик.

- **Пример:** Определите информационный объем страницы книги, если для записи текста использовались только заглавные буквы русского алфавита, кроме буквы Ё.

Решение:

$$N = 32$$

$$2^i = N$$

$$2^i = 32$$

$$2^i = 2^5$$

$$i = 5 \text{ бит}$$

На странице 3000 знаков, т.е. $K=3000$,
тогда объем информации $I = K * i$

$$I = 3000 * 5 ,$$

$$I = 15000 \text{ бит.}$$

Ответ: информационный объем страницы книги равен 15000 бит.



ЕДИНИЦЫ ИЗМЕРЕНИЯ

1 байт = 8 бит

- 1 килобайт = 1Кб = 2^{10} байт = 1024 байта;
- 1 мегабайт = 1Мб = 2^{10} Кб = 2^{20} байта;
- 1 гигабайт = 1Гб = 2^{10} Мб = 2^{30} байта;
- 1 Терабайт (Тб) = 2^{10} Гбайта = 2^{40} байта,
- 1 Петабайт (Пб) = 2^{10} Тбайта = 2^{50} байта.

Примеры некоторых алфавитов.

Двоичный алфавит

А что если алфавит состоит только из двух символов 0 и 1?

В этом случае: $N = 2$; $2^i = N$; $2^i = 2$; $i = 1$ бит.

При использовании двоичной системы (алфавит состоит из двух знаков: 0 и 1) каждый двоичный знак несет **1 бит** информации.

Интересно, что сама единица измерения информации «бит» получила свое название от английского сочетания «**binary digit**» - «двоичная цифра».

Достаточный алфавит

Ограничения на максимальный размер алфавита теоретически не существует. Однако есть алфавит, который можно назвать **достаточным**. С ним мы скоро встретимся при работе с компьютером. Это алфавит **мощностью 256 символов**. В алфавит такого размера можно поместить все практически необходимые символы: латинские и русские буквы, цифры, знаки арифметических операций, всевозможные скобки, знаки препинания....

В этом случае: $N = 256$; $2^i = N$; $2^i = 256$; $2^i = 2^8$; $i = 8$ бит.

Один символ этого алфавита «весит» 8 бит или 1 байт, т.к.

1 байт = 8 бит

Скорость передачи информации

Прием-передача информации могут происходить с разной скоростью.

Количество информации, передаваемое за единицу времени, есть скорость передачи информации или скорость информационного потока.

Очевидно, эта скорость выражается в таких единицах, как бит в секунду (бит/с), байт в секунду (байт/с), килобайт в секунду (Кбайт/с) и т.д.