

Как измерить информацию?

Подходы к определению количества информации

СОДЕРЖАТЕЛЬНЫЙ

Смысл сообщения

АЛФАВИТНЫЙ

**Количество символов
+
Вес одного символа**



ИНФОРМАЦИЯ



Измерение информации: алфавитный подход



Алфавитный подход к измерению информации

Познакомимся с способом измерения информации, который **не связывает количество информации с содержанием сообщения.**

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

Алфавит и его мощность

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**.

Будем обозначать эту величину буквой N . Например, мощность алфавита из заглавных русских букв без буквы ё равна 32.

АБВГДЕЖЗИЙКЛМНОПРСТУФХЦЧШЩЪЭЮЯ

Сколько информации несет один символ в русском языке

Представьте себе, что текст к вам поступает последовательно, по одному знаку, словно бумажная лента, выползающая из телеграфного аппарата. Предположим, что каждый появляющийся на ленте символ с одинаковой вероятностью может быть любым символом алфавита.

В каждой очередной позиции текста может появиться любой из N символов.

Тогда, согласно известной нам формуле $2^I = N$, каждый такой символ несет I бит информации, которое можно определить из решения уравнения: $2^I = 32$.

Получаем: $I = 5$ бит.



Привет! как дела?



АЛФАВИТ – это вся совокупность символов, используемых в некотором языке для представления информации

МОЩНОСТЬ АЛФАВИТА (N) – это число символов в алфавите.

$$2^i = N$$

N

МОЩНОСТЬ АЛФАВИТА
число символов в алфавите (его размер)

i

ИНФОРМАЦИОННЫЙ ВЕС СИМВОЛА
количество информации в одном символе

N

i

I

K

$$I = K \times i$$

K

ЧИСЛО СИМВОЛОВ В СООБЩЕНИИ

I

КОЛИЧЕСТВО ИНФОРМАЦИИ В СООБЩЕНИИ

Достаточный алфавит

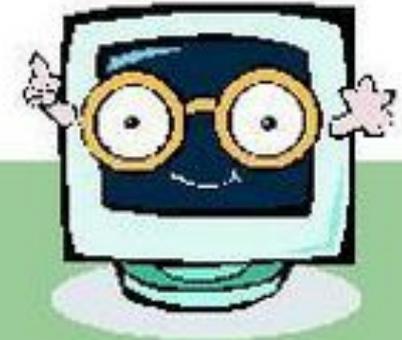
Удобнее всего измерять информацию, когда размер алфавита N равен целой степени двойки. Например, если $N=16$, то каждый символ несет 4 бита информации потому, что $2^4 = 16$. А если $N=32$, то один символ «весит» 5 бит.

Ограничения на максимальный размер алфавита теоретически не существует. Однако есть алфавит, который можно назвать **достаточным**. Это алфавит **мощностью 256 символов**.

Поскольку $256 = 2^8$, то один символ этого алфавита «весит» 8 бит. Причем 8 бит информации — это настолько характерная величина, что ей даже присвоили свое название — байт.

1 байт = 8 бит

Единицы измерения информации



8 бит = 1 байт

1 Кбайт = 2^{10} байт = 1024 байт

1 Мбайт = 2^{10} Кбайт = 1024^2 байт = 1 048 576 байт

1 Гбайт = 2^{10} Мбайт = 1024^3 байт \approx 1 млрд. байт

Количество информации в тексте

Посчитаем количество информации на одной странице книги.



Пусть страница содержит 50 строк. В каждой строке — 60 символов. Значит, на странице уместится $50 \times 60 = 3000$ знаков. Тогда объем информации будет равен: $8 \times 3000 = 24\,000$ бит = 3000 байт = 2,9 Кбайт.

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