

Алфавитный подход к измерению информации.

Способ, не связывающий количество информации с содержанием сообщения, называется алфавитным подходом.

Проще всего разобраться в этом на примере текста, написанного на каком-нибудь языке. Для нас удобнее, чтобы это был русский язык.

Всё 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 (промежуток между словами).

Полное число символов алфавита принято называть мощностью алфавита. Будем обозначать эту величину буквой N . Например, мощность алфавита из русских букв и дополнительных символов равна 54.

Представьте себе, что текст к вам поступает последовательно, по одному знаку, словно бумажная ленточка, выползающая из телеграфного аппарата. Предположим, что каждый появляющийся на ленте символ с одинаковой вероятностью может быть любым символом алфавита. В действительности это не совсем так, но для упрощения примем такое предположение.

В каждой очередной позиции текста может появиться любой из N символов. Каждый символ несёт i бит информации; число i можно определить из уравнения:

$$2^i = N.$$

Для $N = 54$, используя таблицу, получаем:

$$i = 5,755 \text{ бит.}$$

Вот сколько информации несёт один символ в русском тексте! А теперь для того, чтобы найти количество информации во всём тексте, нужно посчитать число символов в нём и умножить на i .

Возьмём с книжной полки какую-нибудь книгу и посчитаем количество информации на одной её странице. Пусть страница содержит 50 строк. В каждой строке – 60 символов. Значит, на странице умещается $50 \times 60 = 17265$ бит.

Следовательно, при алфавитном подходе к измерению информации количество информации от содержания не зависит. Количество информации зависит от объёма текста (то есть от числа знаков в тексте) и от мощности алфавита.

Отсюда следует, например, что нельзя сравнивать информационные объёмы текстов, написанных на разных языках, только по объёму. У них отличаются информационные веса одного символа так как мощности алфавитов разных языков – различные.

Но если книги написаны на одном языке, то понятно, что в толстой книге информации больше, чем в тонкой. При этом содержательная сторона книги в расчёт не берётся.

Сформулируем правило, как измерить информацию, используя для этого алфавитный подход.

Количество информации, содержащееся в символьном сообщении, равно $K \times i$, где K – число символов в тексте сообщения а i – информационный вес символа, который находится из уравнения $2^i = N$, где N – мощность используемого алфавита.

Применение алфавитного подхода удобно, прежде всего, при использовании технических средств работы с информацией. В этом случае теряют смысл понятия «новые – старые», «понятные – непонятные» сведения. Алфавитный подход является объективным способом измерения информации в отличие от субъективного, содержательного, подхода.