

Алфавитный подход к определению количества информации

Существует и другой способ измерения количества информации – **алфавитный**.

Это - измерение количества информации в тексте (символьном сообщении), составленном из символов некоторого алфавита.

К содержанию текста такая мера информации отношения не имеет.

Поэтому такой подход можно назвать **объективным**, то есть не зависящим от воспринимающего его субъекта.

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

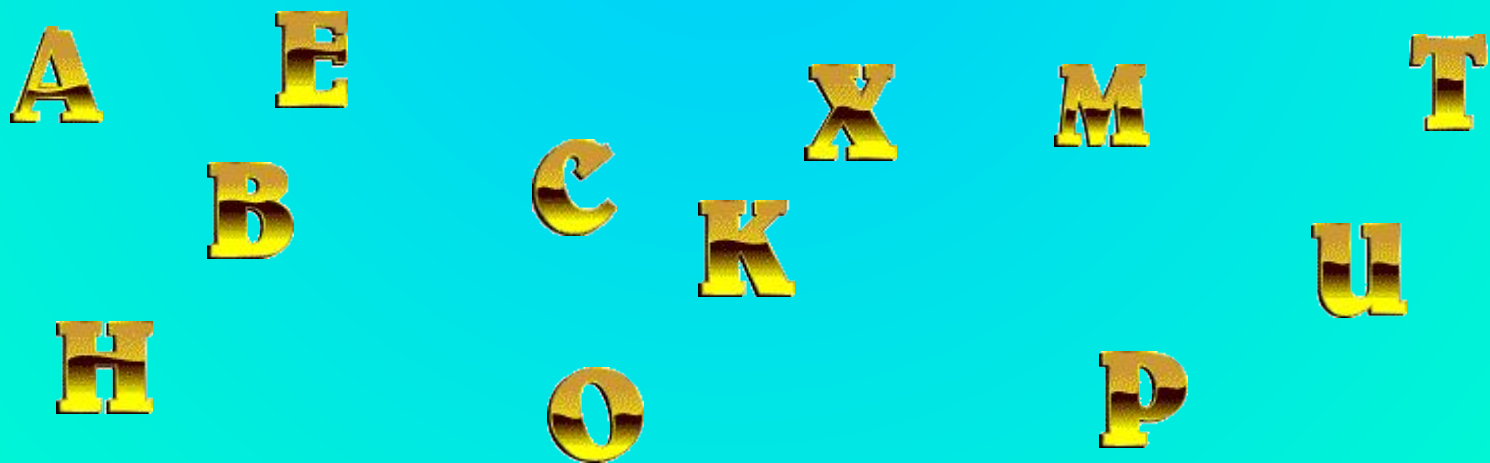
Устройствам нет дела до содержательной стороны сообщений.

Компьютеры, принтеры, модемы работают не с самой информацией а с ее представлением в виде сообщений.

Оценить информационные результаты их работы как **полезные** или **бесполезные** может только человек.

Алфавит

Алфавит – конечное множество символов, используемых для представления информации.



Мощность алфавита

Число символов в алфавите называется **мощностью алфавита**.

Чем меньше знаков в используемом алфавите, тем длиннее сообщение.

Так, например, в алфавите азбуки Морзе всего три знака (точка, тире, пауза), поэтому для кодирования каждой русской или латинской буквы нужно использовать несколько знаков, и текст, закодированный по Морзе, будет намного длиннее, чем при обычной записи.

Пример:

Сигнал **SOS**: 3 знака в латинском алфавите;

11 знаков в алфавите Морзе: **··· пауза – – – пауза ···**.

Количество информации, которое несет в тексте каждый символ (i), вычисляется из уравнения Хартли:

$$2^i = N,$$

где N – мощность алфавита.

Величину i можно назвать **информационным весом символа**.

Отсюда следует, что количество информации во всем тексте (I), состоящем из K символов, равно произведению информационного символа на K :

$$I = i * K.$$

Эту величину можно назвать **информационным объемом текста**.

**Какова минимальная
мощность алфавита, с
помощью которого можно
записывать
(кодировать информацию)?**

Односимвольный алфавит

Сообщение любой длины, использующее **односимвольный алфавит**, содержит **нулевую информацию**.

Доказательство:

Предположим, что используемый алфавит состоит из одного символа, например, «1».

Интуитивно понятно, что сообщить что-либо с помощью единственного символа невозможно.

Но это же доказывается строго с точки зрения алфавитного подхода.

Информационный вес символа в таком алфавите находится из уравнения:

$$2^i = 1$$

Но поскольку

$$1 = 2^0,$$

то отсюда следует, что

$$i = 0 \text{ бит}$$

Минимальная мощность алфавита

Минимальная мощность алфавита, пригодного для передачи информации, равна **2**.

Такой алфавит называется **двоичным алфавитом**.

Информационный вес символа в двоичном алфавите легко определить.

Поскольку

$$2^i = 2,$$

то $i = 1$ бит

Итак, **один символ** двоичного алфавита несет **1 бит информации**.

1 бит – исходная единица измерения информации.

Мощность русского алфавита

Каждая буква русского алфавита
(если считать, что е = ё)
несет информацию 5 бит
($32 = 2^5$).

Компьютерный алфавит

Современный компьютер может обрабатывать числовую, текстовую, графическую, звуковую и видео информацию.

Все эти виды информации в компьютере представлены в двоичном коде, т. е. используется алфавит мощностью два (всего два символа 0 и 1).

Связано это с тем, что удобно представлять информацию в виде последовательности электрических импульсов: импульс отсутствует (0), импульс есть (1).

Такое **кодирование** принято называть **двоичным**, а сами логические последовательности нулей и единиц - **машинным языком**.

Байт

Компьютер для внешнего представления текстов и другой символьной информации использует

алфавит мощностью 256 символов.



Байт вводится как информационный вес символа из алфавита мощностью 256.

Так как

$$256 = 2^8,$$

то **1 байт = 8 бит.**



Производные единицы измерения информации

Килобайт больше байта в 1024 раза, а число $1024 = 2^{10}$.

При приближенных вычислениях можно использовать коэффициент 1000.

1 Мегабайт (1 Мб) = 1024 Кбайт \approx 1000 Кб

1 Гигабайт (1Гб) = 1024 Мбайт \approx 1000 Мб

1 Терабайт (1Тб) = 1024 Гигабайт \approx 1000 Гб

1 Петабайт (Пб) = 1024 Терабайт \approx 1000 Тб

Задачи для закрепления

Задача 1.

Книга, набранная с помощью компьютера, содержит 159 страниц.

На каждой странице – 40 строк.

В каждой строке – 60 символов.

Каков объем информации в книге?

Решение:

Мощность компьютерного алфавита равна 256.

Один символ несет 1 байт информации.

Значит, страница содержит $40 * 60 = 2400$ байт информации.

Объем всей информации в книге (в разных единицах):

$$2400 * 150 = 360\,000 \text{ байт}$$

$$360000/1024 = 351,5625 \text{ Кбайт}$$

$$351,5625/1024 = 0,3433 \text{ Мбайт.}$$