

Измерение информации. Объёмный подход

*Автор: Димитрова Ирина
Сергеевна*

*МБОУ Удомельская средняя
общеобразовательная школа №4
Учитель информатики*

Алфавитный (объёмный) подход к измерению информации

Алфавитный (объёмный) подход к измерению информации позволяет определить количество информации, заключенной в тексте, записанном с помощью некоторого алфавита.

Алфавит – множество используемых символов в языке.

Обычно под алфавитом понимают не только буквы, но и цифры, знаки препинания и пробел.

Мощность алфавита (N) – количество символов, используемых в алфавите.

Например, мощность алфавита из русских букв равна 32 (буква ё обычно не используется).

Если допустить, что все символы алфавита встречаются в тексте с одинаковой частотой (равновероятно), то количество информации, которое несет каждый символ, вычисляется по формуле Хартли: $i = \log_2 N$

N – мощность алфавита. Задаёт связь между количеством возможных событий N и количеством информации: $N=2^i$

Из базового курса информатики известно, что в компьютерах используется двоичное кодирование информации. Для двоичного представления текстов в компьютере чаще всего используется равномерный восьмиразрядный код. С его помощью можно закодировать алфавит из 256 символов, поскольку $2^8=256$.

В стандартную кодовую таблицу (например, ASCII) помещаются все необходимые символы: английские и русские прописные и строчные буквы, цифры, знаки препинания, знаки арифметических операций, всевозможные скобки и пр.

В двоичном коде один двоичный разряд несёт одну единицу информации, которая называется 1 бит.

Например, в 2-символьном алфавите каждый символ «весит» 1 бит ($\log_2 2=1$); в 4-символьном алфавите каждый символ несёт 2 бита информации ($\log_2 4=2$); в 8-символьном – 3 бита ($\log_2 8=3$) и т.

А.

Один символ из алфавита мощностью 256 (28) несёт в тексте 8 битов информации. Такое количество информации называется байтом.

1 байт = 8 битов

Информационный объём текста в памяти компьютера измеряется в байтах. Он равен количеству знаков в записи текста.

Единицы измерения информации

1 байт = 8 бит

1 килобайт (Кб) = 2^{10} байт = 1024 байт

1 мегабайт (Мб) = 2^{10} Кб = 1024 Кб = 2^{20} байт

1 гигабайт (Гб) = 2^{10} Мб = 1024 Мб = 2^{30} байт

1 терабайт (Тб) = 2^{10} Гб = 1024 Гб = 2^{40} байт

1 петабайт (Пб) = 2^{10} Тб = 1024 Тб = 2^{50} байт

1 эксабайт (Эб) = 2^{10} Пб = 1024 Пб = 2^{60} байт

1 зеттабайт (Зб) = 2^{10} Эб = 1024 Эб = 2^{70} байт

1 йоттабайт (Йб) = 2^{10} Зб = 1024 Зб = 2^{80} байт

Единицы измерения количества информации, в названии которых есть приставки «кило», «мега» и т. д., с точки зрения теории измерений не являются корректными, поскольку эти приставки используются в метрической системе мер, в которой в качестве множителей кратных единиц используется коэффициент

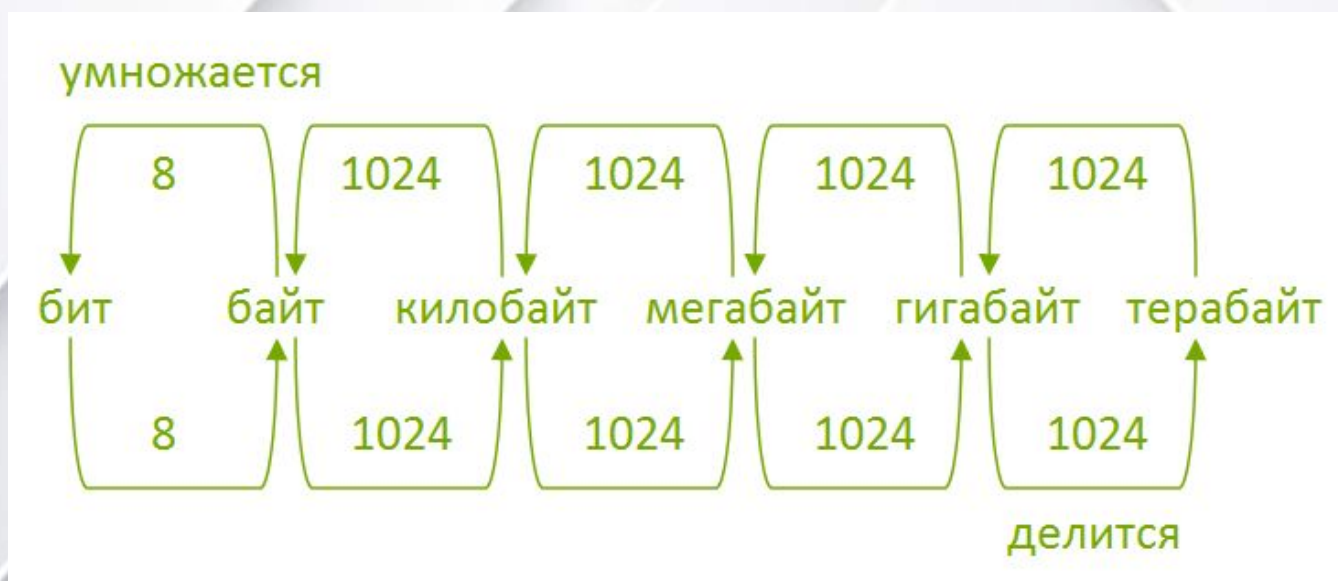
10^n , где $n=3,6,9$ и т. д.

Для устранения этой некорректности Международная электротехническая комиссия, занимающаяся созданием стандартов для отрасли электронных технологий, утвердила ряд новых приставок для единиц измерения количества информации: киби (kibi), меби (mebi), гиби (gibi), теби (tebi), пети (peti), эксби (exbi). Однако пока используются старые обозначения единиц измерения количества информации, и требуется время, чтобы новые названия начали широко применяться.

Название «байт» было придумано в 1956 году В. Бухгольцем при проектировании первого суперкомпьютера. Слово «byte» было получено путем замены второй буквы в созвучном слове «bite», чтобы избежать путаницы с уже имеющимся термином «bit».



Последовательность действий при переводе одних единиц измерения информации в другие приведена на следующей схеме:



Если весь текст состоит из K символов, то при алфавитном подходе объём V содержащейся в нем информации равен:

$$V = K \cdot i$$

где i – информационный вес одного символа в используемом алфавите.

Зная, что $i = \log_2 N$, данную выше формулу можно представить в другом виде:

если количество символов алфавита равно N , а количество символов в записи сообщения – K , то информационный объём V данного сообщения вычисляется по формуле:

$$V = K \cdot \log_2 N$$

При алфавитном подходе к измерению информации информационный объём текста зависит только от размера текста и от мощности алфавита, а не от содержания. Поэтому нельзя сравнивать информационные объёмы текстов, написанных на разных языках, по размеру текста.

Пример:

1. Считая, что каждый символ кодируется одним байтом, оцените информационный объём следующего предложения: Белеет Парус Одинокий В Тумане Моря Голубом!

Решение.

Так как в предложении 44 символа (считая знаки препинания и пробелы), то информационный объём вычисляется по формуле:

$$V = 44 \cdot 1 \text{ байт} = 44 \text{ байта} = 44 \cdot 8 \text{ бит} = 352 \text{ бита}$$

2. Объём сообщения равен 11 Кбайт. Сообщение содержит 11264 символа. Какова мощность алфавита?

Решение.

Выясним, какое количество бит выделено на 1 символ. Для этого переведем объём сообщения в биты:

$$11 \text{ Кбайт} = 11 \cdot 210 \text{ байт} = 11 \cdot 210 \cdot 8 \text{ бит} = 11 \cdot 213 \text{ бит}$$
 и разделим его на число символов.

$$\text{На 1 символ приходится: } (11 \cdot 213) / 11264 = (11 \cdot 213) / (11 \cdot 210) = 23 = 8 \text{ бит.}$$

Мощность алфавита определяем из формулы Хартли: $N = 2^8 = 256$ символов.

Спасибо за внимание

