

Алфавитный (объёмный) подход

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 в информатике 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).

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 $256=2^8$.

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 несет 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 несет 2 бита информации ($\log_2 4 = 2$); в 8-символьном - 3 бита ($\log_2 8 = 3$) и т. д.

Один символ из алфавита мощностью 256 (2^8) несет в тексте 8 битов информации. Такое количество информации называется байтом.

1 байт = 8 битов

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

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 $i=\log^2 N$ ($N=2^i$), 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.

Задача 1

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

Задача 2

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

Задача 3

В велокроссе участвуют 28 спортсменов. Специальное устройство регистрирует прохождение каждым из участников промежуточного финиша, записывая его номер с использованием минимально возможного количества бит, одинакового для каждого спортсмена. Какой объём памяти будет использован устройством, когда все спортсмены прошли промежуточный финиш? (Ответ дайте в битах.)

Задача 4

При регистрации в компьютерной системе каждому пользователю выдаётся пароль, состоящий из 11 символов и содержащий только символы А, Б, В, Г, Д, Е. Каждый такой пароль в компьютерной программе записывается минимально возможным и одинаковым целым количеством байт, при этом используют посимвольное кодирование и все символы кодируются одинаковым и минимально возможным количеством бит. Определите, сколько бит необходимо для хранения 20 паролей.

Домашнее задание

§3