

Измерение информации Алфавитный подход

Измерение информации

Алфавитный (объемный) подход

Алфавитный подход к измерению информации применяется в цифровых (компьютерных) системах хранения и передачи информации.

Множество символов, используемых при записи информации, называется **алфавитом**.

Полное количество символов в алфавите называется **мощностью (размером) алфавита**.

Если допустить, что все символы алфавита встречаются в тексте с одинаковой частотой (равновероятно), то **количество информации**, которое несет **каждый символ**, вычисляется по формуле:

$$i = \log_2 N, \text{ т.е. } 2^i = N$$

где

N – мощность алфавита;

i – количество информации одного символа.

Если весь текст состоит из **К символов**, то при **алфавитном** подходе размер содержащейся в нем информации равен:

$$I = K \cdot i$$

где:

i – информационный вес (объем) одного символа в используемом алфавите;

K – количество символов в тексте;

I – объем информации.

Один символ из алфавита мощностью 256 (2^8) несет в тексте 8 бит информации, называемых байтом
(1 байт = 8 битов).

ASCII (*American Standard Code for Information Interchange*) - американский стандартный код для обмена информацией. На каждый символ отводится 8 бит информации; размер алфавита 256 символов.

КОИ8 – восьмибитовая ASCII-совместимая кодовая страница, разработанная для кодирования букв кириллических алфавитов.

Unicode - стандарт кодирования символов, позволяющий представить знаки практически всех письменных языков. На каждый символ отводится 2 байта; размер алфавита 65536 знаков.

Для измерения информации используются и более крупные единицы:

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

1 Мбайт (мегабайт) = 2^{20} байт = 1024 Кбайта

1 Гбайт (гигабайт) = 2^{30} байт = 1024 Мбайта

Задачи

1. Шахматная доска состоит из 8 столбцов и 8 строк. Какое минимальное количество бит потребуется для кодирования координат одного шахматного поля?

- 1) 4 2) 5 3) 6 4) 7

2. Два текста содержат одинаковое количество символов. Первый текст составлен в алфавите мощностью 16 символов, а второй текст – в алфавите из 256 символов. Во сколько раз количество информации во втором тексте больше, чем в первом?

- 1) 12 2) 2 3) 24 4) 4

Задачи

3. Автоматическое устройство осуществило перекодировку информационного сообщения на русском языке, первоначально записанного в 16-битном коде Unicode, в 8-битную кодировку КОИ-8. При этом информационное сообщение уменьшилось на 320 бит. Какова длина сообщения в символах?

- 1) 20 2) 40 3) 80 4) 320

4. Объем сообщения, содержащего 4096 символов, равен $1/512$ части Мбайта. Какова мощность алфавита, с помощью которого записано это сообщение?

- 1) 8 2) 16 3) 4096 4) 16384

Задачи

5. Для передачи секретного сообщения используется код, состоящий из десятичных цифр. При этом все цифры кодируются одним и тем же (минимально возможным) количеством бит. Определите информационный объем сообщения длиной 150 символов.

1) 600 бит 2) 750 бит 3) 1200 бит 4) 60 байт

6. Мощность алфавита равна 256. Сколько Кбайт памяти потребуется для сохранения 160 страниц текста, содержащего в среднем 192 символа на каждой странице?

1) 10 2) 20 3) 30 4) 40

Задачи

7. Автоматическое устройство осуществило перекодировку информационного сообщения на русском языке, первоначально записанного в 16-битном коде Unicode, в 8-битную кодировку КОИ-8. При этом информационное сообщение уменьшилось на 480 бит. Какова длина сообщения в символах?

- 1) 30 2) 60 3) 120 4) 480

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

- 1) 64 2) 128 3) 256 4) 512