

Кодирование информации

§ 4. Язык – средство кодирования

Определения

Кодирование — это представление информации в форме, пригодной для её хранения, передачи и автоматической обработки.

Код — это правило, по которому сообщение преобразуется в цепочку знаков.

Язык — это система знаков и правил, используемая для записи и передачи информации.

Естественные языки – сформировались в результате развития общества.

Иероглифы

Египетское письмо	
	рука
	дом
	кобра
	лев
	вода

Иероглифы (Китай)	
日	солнце
月	луна
雨	дождь
山	гора
马	лошадь

Алфавитное письмо

Алфавит — это набор знаков, который используется в языке.

Мощность алфавита — это количество знаков в алфавите.

 Какова мощность русского алфавита? латинского?

АБВГДЕЁЖЗИЙКЛМНОПРСТУФХЦЧШЩЪЫЬЭЮЯ
0123456789 . , ; ? ! - : ... « » ()

МОЩНОСТЬ 56

Какие бывают языки?

<ul style="list-style-type: none">•русский•английский•китайский•шведский•суахили•...	$y = 3 \sin x + 1$ $2H_2 + O_2 = 2H_2O$  <p>1. e2-e4 e7-e5...</p>
---	--

Формальный язык – это язык, в котором однозначно определяется значение каждого слова, а также правила построения предложений и придания им смысла.

Сообщения

Сообщение — это любая последовательность символов некоторого алфавита.

Единицы измерения

 2^{10}

1 **байт** = 8 бит

1 **Кбайт** (килобайт) = 1024 байта

1 **Мбайт** (мегабайт) = 1024 Кбайт

1 **Гбайт** (гигабайт) = 1024 Мбайт

1 **Тбайт** (терабайт) = 1024 Гбайт

1 **байт** = 2^3 бит

1 **Кбайт** = 2^{10} байта = $2^{10} \cdot 2^3$ бит = 2^{13} бит

1 **Мбайт** = 2^{10} Кбайт = $2^{10} \cdot 2^{13}$ бит = 2^{23} бит

Перевод в другие единицы

$$\begin{aligned}2 \text{ Кбайт} &= 2 \times (1 \text{ Кбайт}) = 2 \times 1024 \text{ байт} \\ &= 2048 \text{ байт} \\ &= 2048 \times (1 \text{ байт}) = 2048 \times 8 \text{ бит} \\ &= 16\,384 \text{ бита}\end{aligned}$$

Через степени числа 2:

$$\begin{aligned}2 \text{ Кбайт} &= 2 \times 2^{10} \text{ байт} = 2^{11} \text{ байт} \\ &= 2^{11} \times 2^3 \text{ бит} = 2^{14} \text{ бит.}\end{aligned}$$

Алфавитный подход к измерению информации

Каждый символ некоторого сообщения имеет определённый **информационный вес** – несёт **фиксированное количество информации**.

Все символы одного алфавита имеют один и тот же вес, зависящий от мощности алфавита.

Информационный вес символа двоичного алфавита принят за минимальную единицу измерения информации и называется **1 бит (bit)**».

АЛФАВИТ

– ЭТО ВСЯ СОВОКУПНОСТЬ СИМВОЛОВ, ИСПОЛЬЗУЕМЫХ В НЕКОТОРОМ ЯЗЫКЕ ДЛЯ ПРЕДСТАВЛЕНИЯ ИНФОРМАЦИИ.

МОЩНОСТЬ АЛФАВИТА (N)

– ЭТО ЧИСЛО СИМВОЛОВ В АЛФАВИТЕ

$$N = 2^i$$



N	МОЩНОСТЬ АЛФАВИТА
i	ИНФОРМАЦИОННЫЙ ВЕС СИМВОЛА количество информации в одном символе

Информационный объем сообщения

Информационный объём I сообщения равен произведению количества K символов в сообщении на информационный вес i символа алфавита:

$$I = K \times i$$

K

Число символов
в символьном сообщении

i

Информационный вес
одного символа

Задача 1

Алфавит племени Пульти содержит 8 символов. Каков информационный вес символа этого алфавита?

Решение:

Краткая запись условия задачи

$$\begin{array}{l|l} N=8 & N=2^i \\ \hline i-? & \end{array} \quad \begin{array}{l|l} & 8=2^i \\ & i=3 \text{ бита.} \end{array}$$

Вычисления

Соотношение, связывающее величины i и N

Ответ: $i=3$ бита.

Задача 2

Сообщение, записанное буквами 32-символьного алфавита, содержит 140 символов. Какое количество информации оно несёт?

Решение:

$$\begin{array}{l|l} N = 32, & I = K \times i, \\ \hline K = 140 & N = 2^i \\ I = ? & \end{array}$$

$$32 = 2^i, i = 5 \text{ бит}, I = 140 \times 5 = 700 \text{ (битов)}$$

Ответ: I=700 битов.