

# Кодирование информации

- § 5. Язык и алфавит
- § 6. Кодирование
- § 7. Дискретность
- § 8. Алфавитный подход к измерению количества информации

# Кодирование информации

## § 5. Язык и алфавит

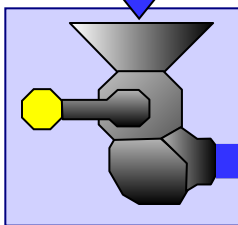
# Что такое кодирование?

**Кодирование** – это запись информации с помощью некоторой знаковой системы (языка).



**Зачем кодируют информацию?**

кодирование



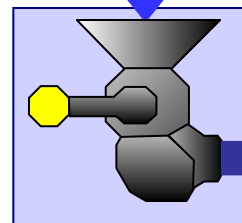
данные (код)

10101001010

передача

Информация передается, обрабатывается и хранится в виде кодов.

борьба с помехами  
(специальные способы кодирования)



данные (код)

11111100010

передача

обработка



хранение

# Что такое кодирование?

**Кодирование** — это представление информации в форме, удобной для её хранения, передачи и обработки. Правило такого преобразования называется **КОДОМ**.

## Текст:

- в России: **Привет, Вася!**
- передача за рубеж (*транслит*): **Privet, Vasya!**
- Windows-1251: **CFF0F8F2E52C20C2E0F1FF21**
- стенография:
- шифрование: **Рсйгжу-!Гбта”**

## Числа:

- для вычислений: **25**
- прописью: **двадцать пять**
- римская система: **XXV**



Как зашифровано?



Зачем?

# Код Морзе

А	•—	О	---	Э	••—••
Б	—•••	П	•—••	Ю	••—
В	•—	Р	•—•	Я	•—•—
Г	—••	С	•••		
Д	—••	Т	—	1	•—
Е	•	У	••—	2	••—
Ж	•••—	Ф	••—•	3	•••—
З	—•••	Х	••••	4	••••—
И	••	Ц	—•—•	5	•••••
Й	•—	Ч	—•••	6	—••••
К	—•—	Ш	—•••	7	—••••
Л	•—••	Щ	—••—	8	—••••
М	—	Ь	—••—	9	—••••
Н	—•	Ы	—•—	0	—••••



Самюэль Морзе  
(1791–1872)



Код неравномерный,  
нужен разделитель!

•— •— •• •— ВАСЯ  
•—•— ВА, АК, ПТ, ЕМЕТ?

# Язык


**Язык** — это система знаков, используемая для хранения, передачи и обработки информации.

Иероглифы:

Египетское письмо	
	рука
	дом
	кобра
	лев
	вода

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

## Какие бывают языки?

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

# Естественные и формальные языки

## *Естественные*

- результат развития общества
- для общения в быту
- значения слов зависят от контекста
- есть синонимы
- есть омонимы
- нет строгих правил образования предложений
- есть исключения

## *Формальные*

- созданы людьми
- в специальных областях знаний
- значения слов не зависят от контекста
- синонимов нет
- омонимов нет
- правила образования предложений строго определены
- нет исключений



# Алфавитное письмо

---

**Алфавит** — это набор знаков, который используется в языке.

**Мощность алфавита** — это количество знаков в алфавите.

А Б В Г Д Е Ё Ж З И Й К Л М Н О П Р С Т У Ф Х Ц Ч Ш Щ Ъ Ы Ь Э Ю Я  
0 1 2 3 4 5 6 7 8 9 . , ; ? ! - : ... « » ( )

мощность 56

**Слово** — это последовательность символов алфавита, которая используется как самостоятельная единица и имеет определённое значение.

# Сообщения

**Сообщение** — это любая последовательность символов некоторого алфавита.

**Пример:** алфавит @ # \$ %.

Сообщения длины 1: @ # \$ %.

всего 4

Сообщения длины 2:

@@	@#	@\$	@%
#@	##	#\$	#%
\$@	\$#	\$\$	\$%
%@	%#	\$\$	%%

всего 16



Сколько сообщений длины  $L$  ?

# Двоичное кодирование

**Двоичное кодирование** — это кодирование с помощью двух знаков.

Равномерный код:

А	Б	В	Г
00	01	10	11

АБАВГБ → 000100101101

Количество сообщений длиной  $I$  битов:  $N = 2^I$

**Пример.** Нужно закодировать номер спортсмена от 1 до 200. Сколько битов потребуется?

$$2^7 < 200 \leq 2^8 = 256$$

8 битов