

**Представление информации,
языки,
кодирование**

Язы

К
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

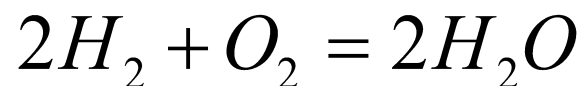
языки?

Естественные

- русский
- английский
- китайский
- шведский
- суахили
- ...

Формальные

$$y = 3 \sin x + 1$$



1. e2-e4 e7-e5...

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

Естественные и формальные языки

Естественн

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


Формальны

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

Кодирование – процесс представления информации, удобный для ее хранения и (или) передачи. Правило такого преобразования называется кодом.

Запись текста на естественном языке можно рассматривать как способ кодирования речи с помощью графических элементов (букв, иероглифов).

Текст:

- в России: **Привет, Вася!**
- передача за рубеж (транслит): **Privet, Vasya!**
- Windows-1251: **CFF0E8E2E52C20C2E0F1FF21**
- стенография:
- шифрование: **F  1a"**

Числа:

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

Чтение текста - это обратный по отношению к письму процесс, при котором письменный текст преобразуется в устную речь. Чтение можно

назвать декодированием письменного текста

Декодирование – процесс обратный кодированию.



Цели и способы кодирования

Для кодирования одной и той же информации могут быть использованы разные способы; их выбор зависит от цели кодирования, условий, имеющихся средств. Выбор способа кодирования информации связан с предполагаемым способом ее обработки.

Цели и способы кодирования

Шифрование	Засекречивание информации
Стенография	Быстрый способ записи
Телеграфный код	Передача по техническим каналам связи
Системы счисления	Выполнение математических вычислений

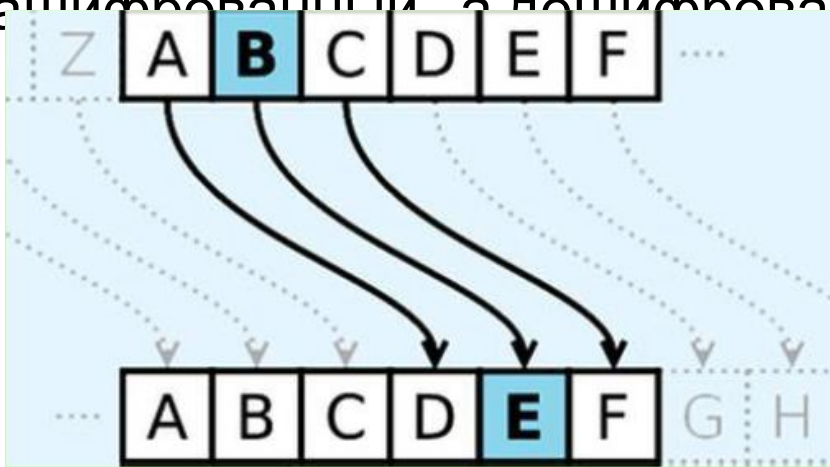
Стенография

Стенография – это быстрый способ записи устной речи.

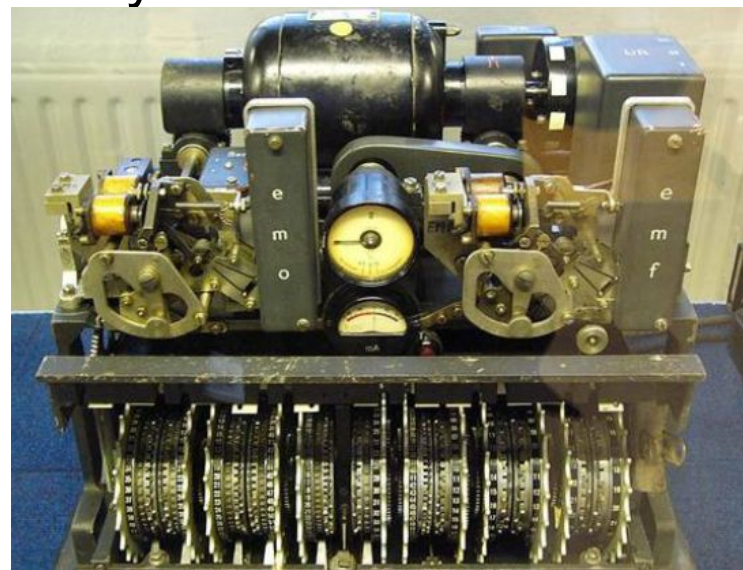


Шифрование

Шифрование – это процесс превращения открытого текста в зашифрованный, а дешифрование – обратный процесс.



Методами шифрования занимается наука **криптография**.



Телеграфный код

Первым техническим средством передачи информации на расстояние стал телеграф, изобретенный в 1837 году американцем Сэмюэлем Морзе.

Телеграфное сообщение - это последовательность электрических сигналов передаваемых от одного телеграфного провода к другому!



Первый телеграф Морзе,
изобретенный в 1837 году



Сэмюэль Морзе
(1791-1892), США

Технические обстоятельства привели Морзе

к идее использования двух видов сигналов короткого и длинного. Каждая буква алфавита кодируется последовательностью коротких сигналов (точек) и длинных сигналов (тире). Буквы отделяются друг от друга паузами – отсутствием сигналов

А	• —	Р	• — •
Б	— • •	С	• •
В	• — —	Т	—
Г	— — •	У	• • —
Д	— • •	Ф	• • — •
Е	•	Х	• • •
Ж	• • • —	Ц	— • — •
З	— — • •	Ч	— — — •
И	• •	Ш	— — — —
Й	• — — —	Щ	— — • —
К	— • —	Ъ	• — — • •
Л	• • • •	Ы	— — — —
М	— — —	Ь	— • • —
Н	— •	Э	• • • • •
О	— — —	Ю	• • — — • •
П	• — — •	Я	• — • •

Кодовая таблица азбуки Морзе



Код Морзе является:

- **неравномерным кодом** (переменная длина кода разных букв)
- **троичным кодом** (точка, тире, пауза).

Телеграфный



Равномерный телеграфный код был изобретен французом Жаном Морисом Бодо в конце 19 века. В нем использовалось всего **два** вида отличающихся друг от друга электрических сигналов — **длинный** и **короткий**. Длина кодов всех символов алфавита одинакова и равна 5, в этом случае не возникает проблемы отделения букв друг от друга.

Оригинальный код Бодо

Управляющие символы	
о . . .	пробел, перейти к таблице букв
. о . . .	пробел, перейти к таблице цифр
о о . . .	удалить последний знак

таблица букв		таблица цифр	
.. о..	А	оо о..	К
.. оо.	É	оо оо.	L
.. .о.	Е	оо .о.	M
.. .оо	I	оо .оо	N
.. ооо	O	оо ооо	P
.. о.о	U	оо о.о	Q
.. ..о	Y	оо ..о	R
.о ..о	В	о. ..о	S
.о о.о	С	о. о.о	T
.о ооо	D	о. ооо	V
.о .оо	F	о. .оо	W
.о .о.	G	о. .о.	X
.о оо.	Н	о. оо.	Z
.о о..	J	о. о..	—
.. о..	1	о. о..	.
.. .о.	2	о. .о.	°/
.. ..о	3	о. ..о	7/
.. о.о	4	о. о.о	2/
.. ооо	5	о. ооо	'
.. оо.	1/	о. оо.	:
.. ..о	3/	о. ..о	?
.о о..	6	оо о..	(
.о .о.	7	оо .о.)
.о ..о	8	оо ..о	-
.о о.о	9	оо о.о	/
.о ооо	0	оо ооо	+
.о оо.	4/	оо оо.	=
.о .оо	5/	оо .оо	£

Код Бодо - это первый в истории техники способ двоичного кодирования информации. Благодаря идее Бодо удалось автоматизировать процесс передачи и печати букв. Был создан клавишный телеграфный аппарат. Нажатие клавиши с определенной буквой вырабатывает соответствующий пятиимпульсный сигнал, который передается по линии связи. Принимающий аппарат под

печатает

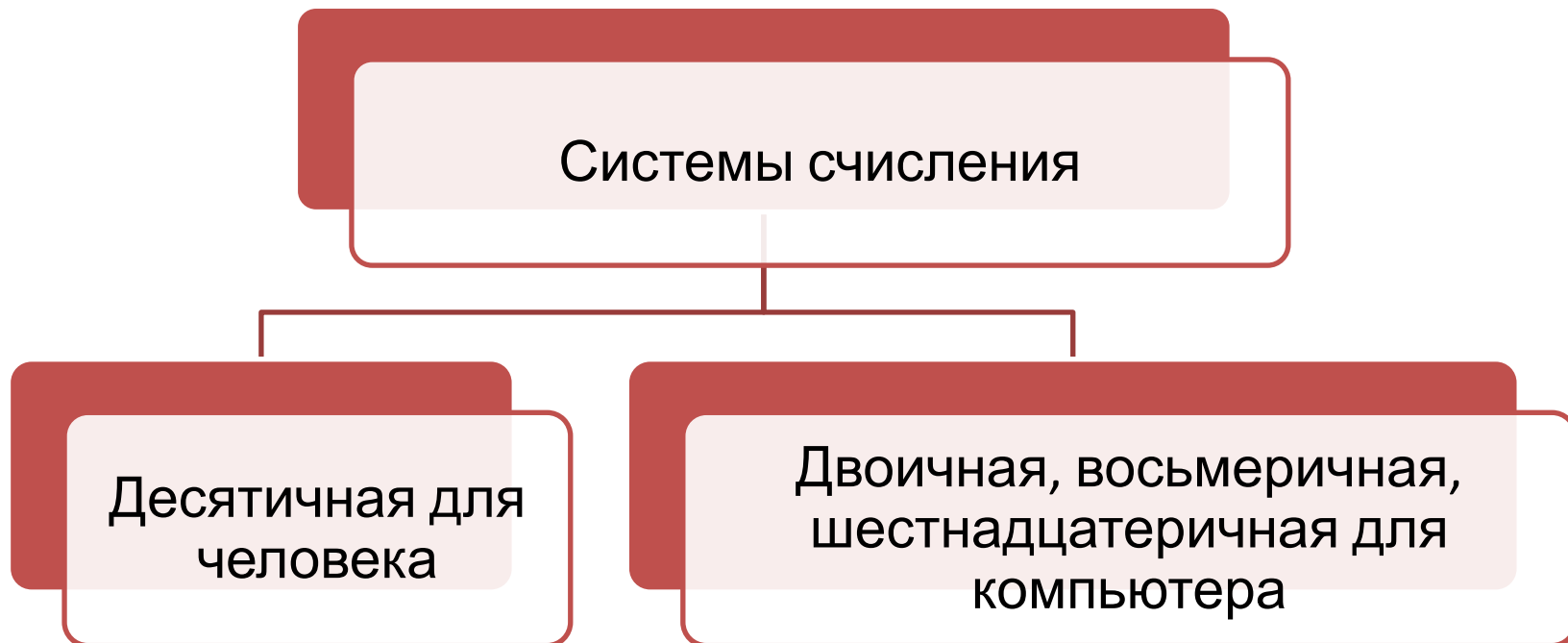


Перфорированная лента



Аппарат Бодо-дуплекс.
Для печати на пяти клавишах
используются два пальца левой руки и
три – правой

Выполнение математических вычислений



Алфавитное подход к измерению количества информации

Алфавитный подход позволяет определить

количество

Алфавит — это набор знаков, который используется в информации, заключенной в тексте.
языке.

Мощность алфавита — это количество знаков в алфавите.

АБВГДЕЁЖЗИЙКЛМНОПРСТУФХЦЧШЩЪЫЬЭЮЯ

0123456789 . , ; ? ! - : ... « » ()

МОЩНОСТЬ 56

$$2^i = N$$

N — количество символов в алфавите (мощность алфавита)

i — количество информации, содержащейся в одном символе алфавита

Алфавитный подход

- 1) определяем мощность алфавита N ;
- 2) определяем количество битов информации i , приходящихся на один символ, — информационную ёмкость (объём) символа:

N , символов	2	4	8	16	32	64	128	256	512	1024
i , битов информации	1	2	3	4	5	6	7	8	9	10

- 3) количество информации в сообщении:

$$I = L \cdot i$$

где L — количество символов в сообщении.

Алфавитный подход

- каждый символ несёт одинаковое количество информации
- частота появления разных символов (и сочетаний символов) не учитывается
- количество информации определяется только длиной сообщения и мощностью алфавита
- смысл сообщения не учитывается

Решение задач

Задача 1. Определить количество информации в 10 страницах текста

(на каждой странице 32 строки по 64 символа) при использовании алфавита из 256 символов.

Решение:

1) информационная ёмкость символа:

$$256 = 2^8 \Rightarrow i = 8 \text{ бит} = 1 \text{ байт}$$

2) количество символов на странице:

$$32 \cdot 64 = 25 \cdot 26 = 211$$

3) общее количество символов:

$$L = 10 \cdot 211$$

4) информационный объём сообщения:

$$I = L \cdot i = 10 \cdot 211 \cdot 1 \text{ байтов} = 20 \text{ Кбайт}$$

Задачи для самостоятельного решения

1. Алфавит племени Мульти состоит из 8 букв. Какое количество информации несет 1 буква этого алфавита?
2. Информационный объем одного символа некоторого сообщения из алфавита племени Пульти равен 6 битам. Сколько символов входит в алфавит этого племени, с помощью которого пультияне составили это сообщение?
3. Сообщение, записанное буквами из 128 – символьного алфавита, содержит 30 символов. Какой объем информации оно несет?
4. Сообщение, составленное с помощью 32 – символьного алфавита, содержит 80 символов. Другое сообщение составлено с использованием 64 – символьного алфавита и содержит 70 символов. Сравните объемы информации, содержащейся в сообщениях.
5. Информационное сообщение объемом 1 Кбайта содержит 4096

6. Сколько килобайтов составляет сообщение из 512 символов 16 –

символьного алфавита?

7. Для записи текста использовался 256 – символьный алфавит.

Каждая страница содержит 30 строк по 70 символов в строке.

Какой объем информации содержат 5 страниц текста?

8. Сообщение занимает 3 страницы по 25 строк. В каждой строке

записано по 60 символов. Сколько символов в

использованном алфавите, если все сообщение содержит 1125 байтов?

9. Сколько символов содержит сообщение, записанное с помощью 16-символьного алфавита, если его объем составляет 1/64 часть Мегабайта?

10. Пользователь компьютера, хорошо владеющий навыками ввода информации с клавиатуры, может вводить в минуту 100 знаков. Мощность алфавита, используемого в компьютере, равна 256. какое количество информации в байтах может ввести пользователь в компьютер за 1