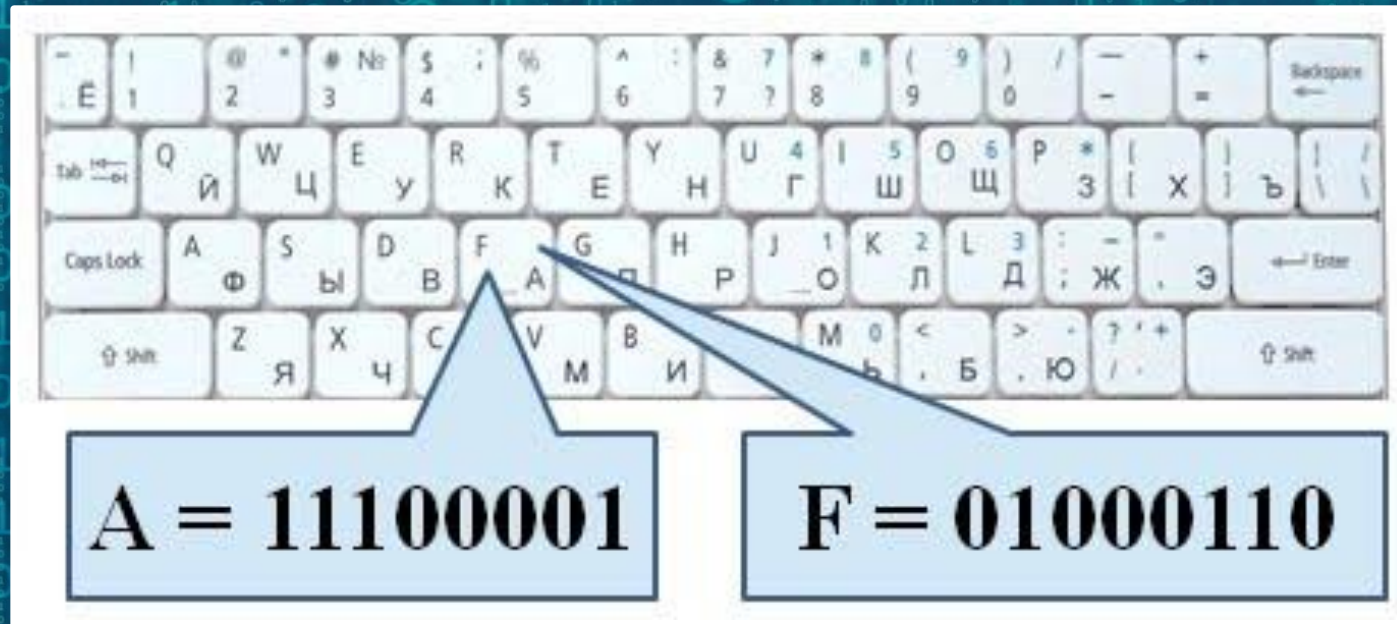


Кодирование Текстовой информации

Выполнил
Вонсович



Кодирование — это процесс преобразования данных из формы, удобной для непосредственного использования, в форму, удобную для передачи, хранения, автоматической переработки и сохранения от несанкционированного доступа.



Основные способы кодирования текстовой информации

Существует несколько основных способов кодирования текстовой информации:

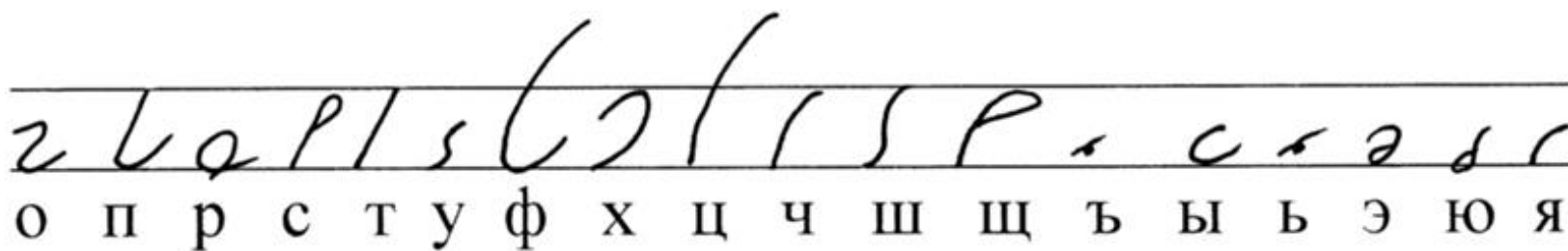
Графический, в котором текстовая информация кодируется путем использования специальных рисунков или знаков;

Символьный, в котором тексты кодируются с использованием символов того же алфавита, на котором написан исходник;

Числовой, в котором текстовая информация кодируется с помощью чисел.



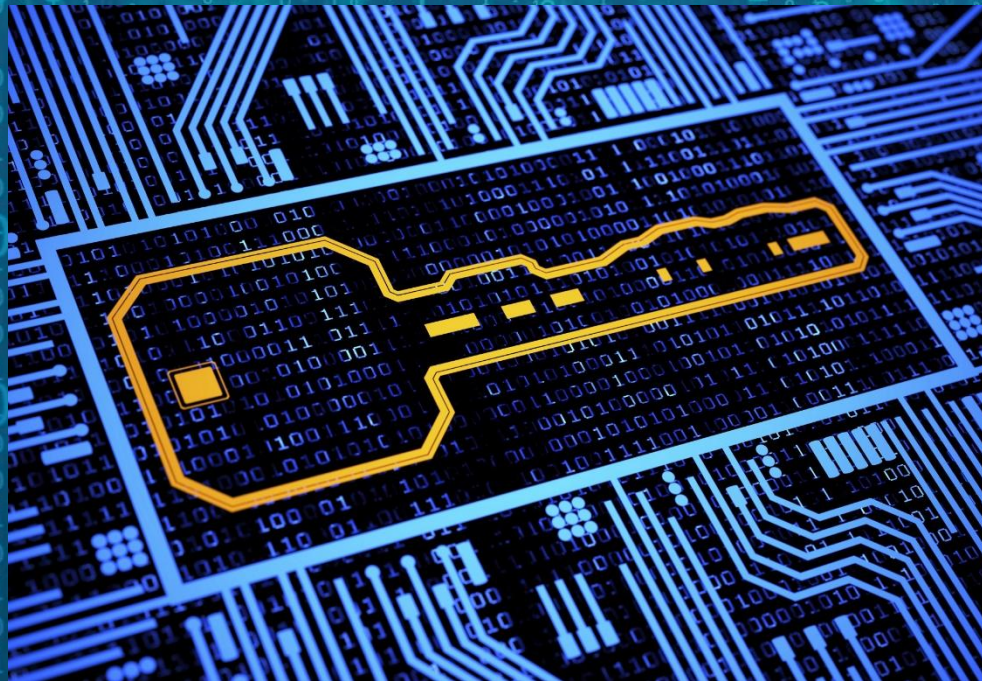
Handwritten shorthand for the Russian alphabet, showing the letters а, б, в, г, д, е, ё, ж, з, и, й, к, л, м, н, н. Each letter is written in a cursive style on a set of three horizontal lines (top, middle, bottom). Below each shorthand symbol is its corresponding Cyrillic letter.



Handwritten shorthand for the Russian alphabet, showing the letters о, п, р, с, т, у, ф, х, ц, ч, ш, щ, ъ, ы, ь, э, ю, я. Each letter is written in a cursive style on a set of three horizontal lines (top, middle, bottom). Below each shorthand symbol is its corresponding Cyrillic letter.

Стенография — это один из способов кодирования текстовой информации с помощью специальных знаков. Она представляет собой быстрый способ записи устной речи. Навыками стенографии могут владеть далеко не все, а лишь немногие специально обученные люди, которых называют стенографистами. Эти люди успевают записывать текст синхронно с речью выступающего человека, что, на наш взгляд, достаточно сложно. Однако для них это не проблема, поскольку в стенограмме целое слово или сочетание букв могут обозначаться одним знаком. Скорость стенографического письма превосходит скорость обычного в 4–7 раз. Расшифровать (декодировать) стенограмму может только сам стенографист.

Шифрован



Шифрование представляет собой процесс превращения открытого текста в зашифрованный, а дешифрование — процесс обратного преобразования, при котором восстанавливается исходный текст. Шифрование — это тоже кодирование, но с засекреченным методом, известным только источнику и адресату. Методами шифрования занимается наука криптография.

Подробнее о числовом способе

В каждом национальном языке имеется свой алфавит, который состоит из определенного набора букв, следующих друг за другом, а значит и имеющих свой порядковый номер. Каждой букве сопоставляется целое положительное число, которое называют кодом символа. Именно этот код и будет хранить память компьютера, а при выводе на экран или бумагу преобразовывать в соответствующий ему символ. Помимо кодов самих символов в памяти компьютера хранится и информация о том, какие именно данные закодированы в конкретной области памяти. Это необходимо для различия представленной информации в памяти компьютера (числа и символы).

Двоичное кодирование текстовой информации

Начиная с конца 60-х годов, компьютеры все больше стали использоваться для обработки текстовой информации и в настоящее время большая часть персональных компьютеров в мире (и наибольшее время) занято обработкой именно текстовой информации.

Традиционно для кодирования одного символа используется количество информации, равное 1 байту, то есть $I = 1 \text{ байт} = 8 \text{ битов}$.

Для кодирования одного символа требуется 1 байт информации.

Если рассматривать символы как возможные события, то по формуле можно вычислить, какое количество различных символов можно закодировать:

$$N = 2^I = 2^8 = 256.$$

Символ	Двоичный код	Десятичный код	Символ	Двоичный код	Десятичный код
A	01000001	65	N	01001110	78
B	01000010	66	O	01001111	79
C	01000011	67	P	01010000	80
D	01000100	68	Q	01010001	81
E	01000101	69	R	01010010	82
F	01000110	70	S	01010011	83
G	01000111	71	T	01010100	84
H	01001000	72	U	01010101	85
I	01001001	73	V	01010110	86
J	01001010	74	W	01010111	87
K	01001011	75	X	01011000	88
L	01001100	76	Y	01011001	89
M	01001101	77	Z	01011010	90

Такое количество символов вполне достаточно для представления текстовой информации, включая прописные и строчные буквы русского и латинского алфавита, цифры, знаки, графические символы и пр.

Кодирование заключается в том, что каждому символу ставится в соответствие уникальный десятичный код от 0 до 255 или соответствующий ему двоичный код от 00000000 до 11111111. Таким образом, человек различает символы по их начертаниям, а компьютер - по их кодам.

При вводе в компьютер текстовой информации происходит ее двоичное кодирование, изображение символа преобразуется в его двоичный код. Пользователь нажимает на клавиатуре клавишу с символом, и в компьютер поступает определенная последовательность из восьми электрических импульсов (двоичный код символа). Код символа хранится в оперативной памяти компьютера, где занимает один байт.

В процессе вывода символа на экран компьютера производится обратный процесс - декодирование, то есть преобразование кода символа в его изображение.

**На сегодня это всё
Была проделана колоссальная
работа**



**Спасибо за
внимание**