

ДВОИЧНОЕ КОДИРОВАНИЕ ТЕКСТОВОЙ ИНФОРМАЦИИ

-
- Традиционно для кодирования одного символа используется количество информации, равное 1 байту, то есть

$$I = 1 \text{ байт} = 8 \text{ битов.}$$

- Если рассматривать символы как возможные события, то можно вычислить какое количество символов можно закодировать:

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

-
- Каждому символу ставится в соответствие уникальный десятичный код от 0 до 255 или соответствующий ему двоичный код от 00000000 до 11111111.
 - Человек различает символы по их начертаниям, а компьютер – по их кодам.

КОДИРОВАНИЕ - ДЕКОДИРОВАНИЕ

- При вводе в компьютер текстовой информации происходит ее двоичное кодирование. Код символа хранится в оперативной памяти и занимает **1 байт**.
- В процессе вывода символа на экран компьютера происходит декодирование, т.е. преобразование кода символа в его изображение.

КОДОВАЯ ТАБЛИЦА

Коды	Содержимое
0 - 32	Соответствуют операциям (перевод строки, разрыв страницы, ввод пробела и т.д.)
33 - 127	Интернациональные коды, соответствуют символам латинского алфавита, цифрам, знакам арифметических операций и знакам препинания.
128 - 255	Национальные коды – в национальных кодировках одному и тому же коду соответствуют разные символы.

КОДОВЫЕ ТАБЛИЦЫ

- Кодовые таблицы для русских букв: КОИ-8, CP1251 (Windows), CP866, Mac, ISO.

Двоичный код	Десятичный код	КОИ8	CP1251	CP866	Mac	ISO
00001101	13	Перевод строки (клавиша Enter)				
.....						
00100000	32	Пробел				
.....						
01011010	90	Z				
.....						
11000010	194	б	В	-	-	Т

Закодируем слово «ЛУНА»

десятичными числами

Л У Н А

203 211 205 192

двоичной последовательностью

Л У Н А

11001011 11010011 11001101 10000000

Попробуем расшифровать последовательность

1101100100101100100011100011111100000

Т и г р

СТАНДАРТ UNICODE

- Этот стандарт отводит на каждый символ 2 байта, поэтому можно закодировать
- $N = 2^l = 2^{15} = 65\,536$ СИМВОЛОВ
- Эту кодировку поддерживают последние версии платформы Microsoft Windows&Office (начиная с 1997 года).