

Подготовила:
Учитель информатики
МБОУ СОШ №2 г. Липецка
Кукина Екатерина Сергеевна

**Двоичное
кодирование
символьной
информации**

17.12.2015

При двоичном кодировании текстовой информации каждому символу ставится в соответствие уникальный десятичный код от 0 до 255 или соответствующий ему двоичный код от 00000000 до 11111111. Так человек различает символы по их начертанию, а компьютер – по их коду.



По формуле, связывающей количество сообщений N и количество информации i , можно вычислить, какое количество информации необходимо, чтобы закодировать каждый знак

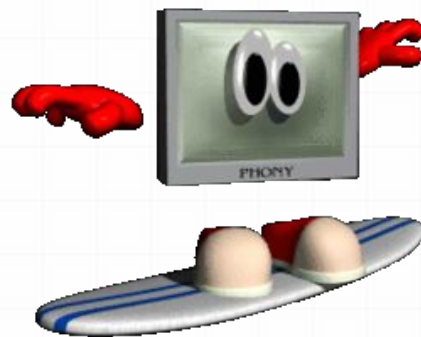
$$0 \quad N = 2^i \rightarrow 256 = 2^i$$

$$2^8 = 2^i \rightarrow i = 8 \text{ битов} = 1 \text{ байт}$$

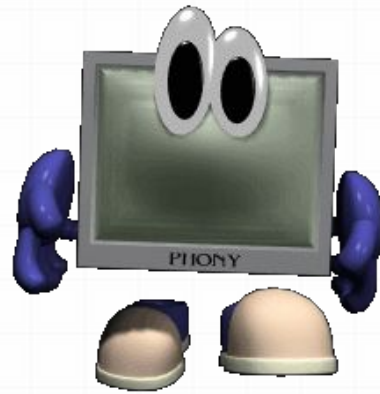
Присвоение символу конкретного двоичного кода – это вопрос соглашения, которое фиксируется в кодовой таблице. Первые 33 кода (с 0 до 32) соответствуют не символам, а операциям (перевод строки, ввод пробела и т.д.). Коды с 33 по 127 являются интернациональными и соответствуют символам латинского алфавита, цифрам, знакам арифметических операций и знакам препинания.

Коды с 128 по 255 являются национальными, т. е. в национальных кодировках одному и тому же коду соответствуют различные символы. Существует 5 однобайтовых кодовых таблиц для русских букв, поэтому тексты созданные в одной кодировке, не будут правильно отображаться в другой.

Хронологически одним из первых стандартов кодирования русских букв на компьютерах был код КОИ - 8 («Код обмена информацией - 8 битный»). Эта кодировка применяется в компьютерах с операционной системой UNIX.



Наиболее распространенная кодировка – это стандартная кириллическая кодировка *Microsoft Windows*, обозначаемая сокращением *CP1251* («CP» означает «Code Page»). Все *Windows* – приложения, работающие с русским языком, поддерживают эту кодировку.



Для работы в среде операционной системы **MS-DOS** используется «альтернативная» кодировка, в терминологии фирмы Microsoft – кодировка **CP 866**.

Фирма Apple разработала для компьютеров
Macintosh свою собственную кодировку русских
букв (*Mac*)



Macintosh

Международная организация по
стандартизации (International Standards
Organization, ISO) утвердила в качестве
стандарта для русского языка еще одну
кодировку под названием **ISO 8859**
– 5.



Стандарты кодировок

- 0 KOI-8 - UNIX
- 0 CP1251 («CP» означает «Code Page») - Microsoft Windows
- 0 CP 866 - MS-DOS
- 0 Mac - Macintosh
- 0 ISO 8859 - 5



Таблица кодировки символов

Двоичный код	Десятичный код	КОИ8	CP1251	CP866	Mac	ISO
0000 0000	0					
.....						
0000 1000	8	Удаление последнего символа (клавиша Backspace)				
.....						
0000 1101	13	Перевод строки (клавиша Enter)				
.....						
0010 0000	32	Пробел				
0010 0001	33	!				
.....						
0101 1010	90	Z				
.....						
0111 1111	127					
.....	128	-	Ъ	А	А	К
.....						
1100 0010	194	Б	В	-	-	Т
.....						
1100 1100	204	Л	М	:	:	Ь
.....						
1101 1101	221	Щ	Э	-	Ё	Н
.....						
1111 1111	225	Ь	я	Нераз. пробел	Нераз. пробел	п

В последнее время появился новый международный стандарт **Unicode**, который отводит на каждый символ не один байт, а два, и поэтому с его помощью можно закодировать не 256 символов, $2^{16} = 65\ 536$ различных символов. Эту кодировку поддерживают редакторы, начиная с MS Office 97.

Задание 1: определите символ по числовому коду.

- 0 Запустите программу БЛОКНОТ
- 0 Нажмите ALT и 0224 (на дополнительной цифровой клавиатуре). Появится символ **а**.
- 0 Повторите эту операцию для числовых кодов от 0225 до 0233. Появятся символы в кодировке (CP 1251 Windows). Запишите их в тетрадь.
- 0 Нажмите ALT и 161 (на дополнительной цифровой клавиатуре). Появится символ **б**.
- 0 Повторите эту операцию для числовых кодов 160, 169, 226. Появятся символы в кодировке (CP 866 MS-DOS). Запишите их в тетрадь.

Задание 2: определите числовой код для символов

0 Определите числовой код, который нужно ввести , удерживая клавишу Alt, чтобы получить символы: ☀, §, \$, ♀

0 Пояснение: данный код содержится в диапазоне от 0 до 50.

СПАСИБО ЗА ВНИМАНИЕ!

