Двоичное кодирование символьной информации

При **двоичном кодировании текстовой информации** каждому символу ставится в соответствие своя уникальная последовательность из восьми нулей и единиц, свой уникальный код
от 00000000 до 11111111
(десятичный код от 0 до 255)

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 33 до 127 являются интернациональными и соответствуют символам латинского алфавита, цифрам, знакам арифметических операций и знакам препинания.

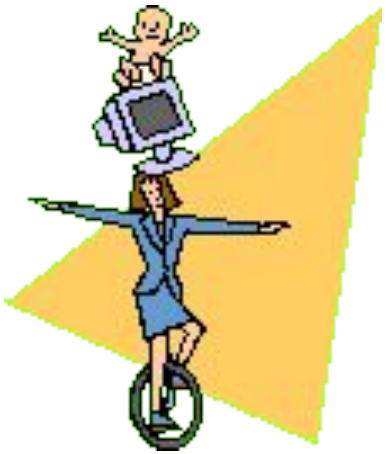
Коды с 128 по 255 являются национальными, т.е. в национальных кодировках одному и тому же коду соответствуют различные символы. К сожалению, в настоящее время существует 5 различных кодовых таблиц для русских букв, поэтому тексты созданные в одной кодировке, не будут правильно отображаться в другой.

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

Наиболее распространенная кодировка – это стандартная кириллистическая 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)



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:

1. **КОИ-8 - UNIX**
2. **CP1251 («CP» означает «Code Page») - Microsoft Windows**
3. **CP 866 - MS-DOS**
4. **Mac - Macintosh**
5. **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 поддерживает платформа Microsoft Windows&Office.