**Урок по теме:** Представление текстовой информации в компьютере (8 класс)

***Цель урока***:

* сформировать у учащихся представление о том, как в компьютере кодируется текстовая информация.

***Учащиеся должны научиться***:

* кодировать и декодировать символы с помощью таблицы кодов;
* находить информационный объем текстов и сообщений.

***Программно-дидактическое обеспечение***: ПК, таблицы кодов, текстовый редактор, калькулятор.

1. **Постановка целей урока.**
2. Как кодируются символы в компьютере? Почему именно так, а не иначе?
3. Всегда ли разные компьютеры «понимают» друг друга? Почему?
4. Сколько текстов поместится на дискете? А на жестком диске?
5. **Актуализация знаний.**
6. Как в компьютере кодируются символы?
7. Что такое «компьютерный алфавит»? Какова его мощность?
8. Чему равен информационный объем одного символа компьютерного алфавита?
9. Почему иногда текст, состоящий из букв русского алфавита, полученный с другого компьютера, мы видим на своем компьютере в виде "абракадабры"?
10. **Изложение нового материала.**

Компьютеры не самого рождения могли обрабатывать символьную информацию. Лишь с конца 60-х годов они стали использоваться для обработки текстов и в настоящее время большинство пользователей ПК занимаются вводом, редактированием и форматированием текстовой информации.

1. ***Таблица кодирования ASCII.***

А теперь «заглянем» в память компьютера и разберемся, как же представлена в нем текстовая информация.

Текстовая информация состоит из символов: букв, цифр, знаков препинания, скобок и других. Мы уже говорили, что множество всех символов, с помощью которых записывается текст, называется алфавитом, а число символов в алфавите — его мощностью.

Для представления текстовой информации в компьютере используется алфавит мощностью 256 символов. Мы знаем, что один символ такого алфавита несет 8 битов информации: 2 в 8 степени равно 256. 8 битов = 1 байт, следовательно:

Один символ в компьютерном тексте занимает 1 байт памяти.

Как мы выяснили, традиционно для кодирования одного символа используется 8 бит. И, когда люди определились с количеством бит, им осталось договориться о том, каким кодом кодировать тот или иной символ, чтобы не получилось путаницы, т.е. необходимо было выработать стандарт – все коды символов сохранить в специальной таблице кодов. В первые годы развития вычислительной техники таких стандартов не существовало, а сейчас наоборот, их стало очень много, но они противоречивы. Первыми решили эти проблемы в США, в институте стандартизации. Этот институт ввел в действие таблицу кодов ASCII (American Standard Code for Information Interchange – стандартный код информационного обмена США).

Рассмотрим таблицу кодов ASCII.

*Пояснение*: раздать учащимся распечатанную таблицу кодов ASCII.

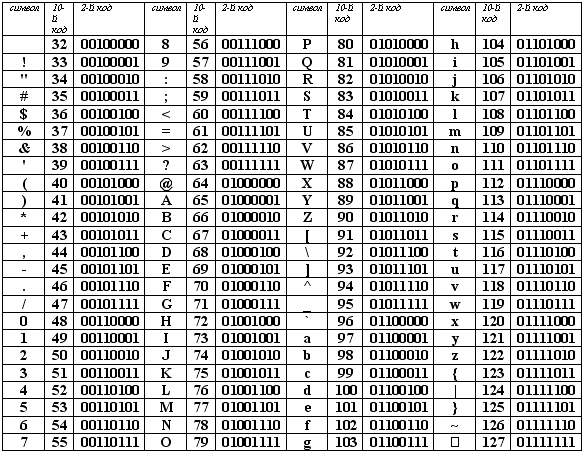
Таблица ASCII разделена на две части. Первая – стандартная – содержит коды от 0 до 127. Вторая – расширенная – содержит символы с кодами от 128 до 255.

Первые 32 кода отданы производителям аппаратных средств и называются они управляющие, т.к. эти коды управляют выводом данных. Им не соответствуют никакие символы.

Коды с 32 по 127 соответствуют символам английского алфавита, знакам препинания, цифрам, арифметическим действиям и некоторым вспомогательным символам.

Коды расширенной части таблицы ASCII отданы под символы национальных алфавитов, символы псевдографики и научные символы.

**Стандартная часть таблицы кодов ASCII**

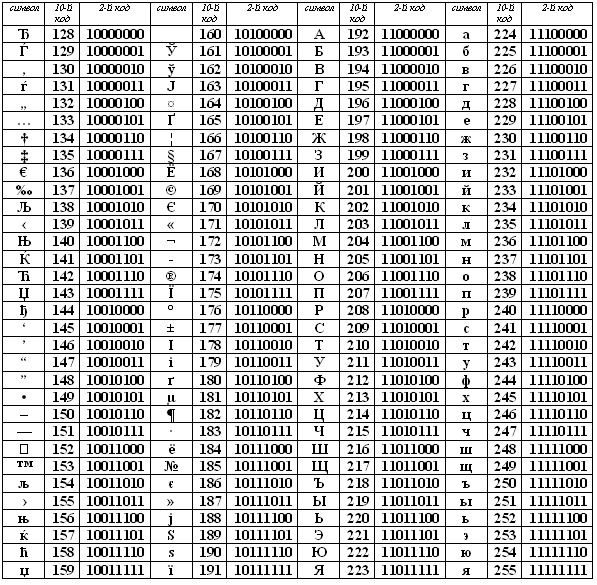


Если вы внимательно посмотрите на обе части таблицы, то увидите, что все буквы расположены в них по алфавиту, а цифры – по возрастанию. Этот принцип последовательного кодирования позволяет определить код символа, не заглядывая в таблицу.

Коды цифр берутся из этой таблицы только при вводе и выводе и если они используются в тексте. Если же они участвуют в вычислениях, то переводятся в двоичную систему счисления.

**Коды национального (русского) алфавита**

**расширенной части таблицы ASCII**

Альтернативные системы кодирования кириллицы.

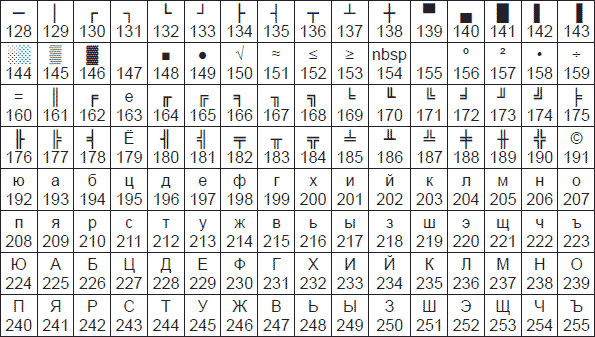
Тексты, созданные в одной кодировке, не будут правильно отображаться в другой. В настоящее время для поддержки букв русского алфавита (кириллицы) существует несколько кодовых таблиц (кодировок), которые используются различными операционными системами, что является существенным недостатком и в ряде случаев приводит к проблемам, связанным с операциями декодирования числовых значений символов.

Для разных типов ЭВМ используются различные кодировки:

В настоящее время существует 5 кодовых таблиц для русских букв: Windows (СР(кодовая страница)1251), MS – DOS (СР(кодовая страница)866), KOИ – 8 (Код обмена информацией, 8-битный) (используется в OS UNIX), Mac (Macintosh), ISO (OS UNIX).

Одним из первых стандартов кодирования кириллицы на компьютерах был стандарт КОИ-8.

***Национальная часть кодовой таблицы стандарта КОИ8-Р***

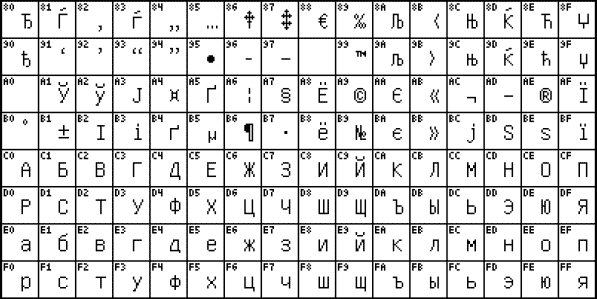


В настоящее время применяется и кодовая таблица, размещенная на странице СР866 стандарта кодирования текстовой информации, которая используется в операционной системе MS DOS или сеансе работы MS DOS для кодирования кириллицы.

***Национальная часть кодовой таблицы СР866***

В настоящее время для кодирования кириллицы наибольшее распространение получила кодовая таблица, размещенная на странице СР1251 соответствующего стандарта, которая используется в операционных системах семейства Windows фирмы Microsoft.

***Национальная часть кодовой таблицы СР1251***



Во всех представленных кодовых таблицах, кроме таблицы стандарта Unicode, для кодирования одного символа отводится 8 двоичных разрядов (8 бит).

В мире существует примерно 6800 различных языков. Если прочитать текст, напечатанный в Японии на компьютере в России или США, то понять его будет нельзя. Чтобы буквы любой страны можно было читать на любом компьютере, для их кодировки стали использовать 2 байта (16 бит).

N = 2i

2i = 216 = 65536

N = 65536 N – мощность алфавита символов в кодовой таблице Unicode.

i – информационный вес символа

**Основополагающая таблица использования кодового пространства Unicode**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Начало области** | **Конец области** | **Набор символов** | **Начало области** | **Конец области** | **Набор символов** |
| 0000 | 007F | Basic Latin | 2150 | 218F | Number Forms |
| 0080 | 00FF | Latin-1 Supplement | 2190 | 21FF | Arrows |
| 0100 | 017F | Latin Extended-A | 2200 | 22FF | Mathematical Operators |
| 0180 | 024F | Latin Extended-B | 2300 | 23FF | Miscellaneous Technical |
| 0250 | 02AF | IPA Extensions | 2400 | 243F | Control Pictures |
| 02B0 | 02FF | Spacing Modifier Letters | 2440 | 245F | Optical Character Recognition |
| 0300 | 036F | Combining Diacritical Marks | 2460 | 24FF | Enclosed Alphanumerics |
| 0370 | 03FF | Greek | 2500 | 257F | Box Drawing |
| **0400** | **04FF** | **Cyrillic** | 2580 | 259F | Block Elements |
| 0530 | 058F | Armenian | 25A0 | 25FF | Geometric Shapes |
| 0590 | 05FF | Hebrew | 2600 | 26FF | Miscellaneous Symbols |
| 0600 | 06FF | Arabic | 2700 | 27BF | Dingbats |
| 0700 | 074F | Syriac | 2800 | 28FF | Braille Patterns |
| 0780 | 07BF | Thaana | 2E80 | 2EFF | CJK Radicals Supplement |
| 0900 | 097F | Devanagari | 2F00 | 2FDF | Kangxi Radicals |
| 0980 | 09FF | Bengali | 2FF0 | 2FFF | Ideographic Description Characters |
| 0A00 | 0A7F | Gurmukhi | 3000 | 303F | CJK Symbols and Punctuation |
| 0A80 | 0AFF | Gujarati | 3040 | 309F | Hiragana |
| 0B00 | 0B7F | Oriya | 30A0 | 30FF | Katakana |
| 0B80 | 0BFF | Tamil | 3100 | 312F | Bopomofo |
| 0C00 | 0C7F | Telugu | 3130 | 318F | Hangul Compatibility Jamo |
| 0C80 | 0CFF | Kannada | 3190 | 319F | Kanbun |
| 0D00 | 0D7F | Malayalam | 31A0 | 31BF | Bopomofo Extended |
| 0D80 | 0DFF | Sinhala | 3200 | 32FF | Enclosed CJK Letters and Months |
| 0E00 | 0E7F | Thai | 3300 | 33FF | CJK Compatibility |
| 0E80 | 0EFF | Lao | 3400 | 4DB5 | CJK Unified Ideographs Extension A |
| 0F00 | 0FFF | Tibetan | 4E00 | 9FFF | CJK Unified Ideographs |
| 1000 | 109F | Myanmar | A000 | A48F | Yi Syllables |
| 10A0 | 10FF | Georgian | A490 | A4CF | Yi Radicals |
| 1100 | 11FF | Hangul Jamo | AC00 | D7A3 | Hangul Syllables |
| 1200 | 137F | Ethiopic | D800 | DB7F | High Surrogates |
| 13A0 | 13FF | Cherokee | DB80 | DBFF | High Private Use Surrogates |
| 1400 | 167F | Unified Canadian Aboriginal Syllabics | DC00 | DFFF | Low Surrogates |
| 1680 | 169F | Ogham | E000 | F8FF | Private Use |
| 16A0 | 16FF | Runic | F900 | FAFF | CJK Compatibility Ideographs |
| 1780 | 17FF | Khmer | FB00 | FB4F | Alphabetic Presentation Forms |
| 1800 | 18AF | Mongolian | FB50 | FDFF | Arabic Presentation Forms-A |
| 1E00 | 1EFF | Latin Extended Additional | FE20 | FE2F | Combining Half Marks |
| 1F00 | 1FFF | Greek Extended | FE30 | FE4F | CJK Compatibility Forms |
| 2000 | 206F | General Punctuation | FE50 | FE6F | Small Form Variants |
| 2070 | 209F | Superscripts and Subscripts | FE70 | FEFE | Arabic Presentation Forms-B |
| 20A0 | 20CF | Currency Symbols | FEFF | FEFF | Specials |
| 20D0 | 20FF | Combining Marks for Symbols | FF00 | FFEF | Halfwidth and Fullwidth Forms |
| 2100 | 214F | Letterlike Symbols | FFF0 | FFFD | Specials |

Использование Unicode значительно упрощает создание многоязычных документов, публикаций и программных приложений.

***Рассмотрим примеры.***

1. Представьте в форме шестнадцатеричного кода слово «ЭВМ» во всех пяти кодировках. Воспользуемся компьютерным калькулятором для перевода чисел из десятичной в шестнадцатеричную систему счисления.

Последовательности десятичных кодов слова «ЭВМ» в различных кодировках составляем на основе кодировочных таблиц:

КОИ8-Р: 252 247 237

СР1251: 221 194 204

СР866: 157 130 140

Мас: 157 130 140

ISO: 205 178 188

Переводим с помощью калькулятора последовательности кодов из десятичной системы в шестнадцатеричную:

КОИ8-Р: FC F7 ED

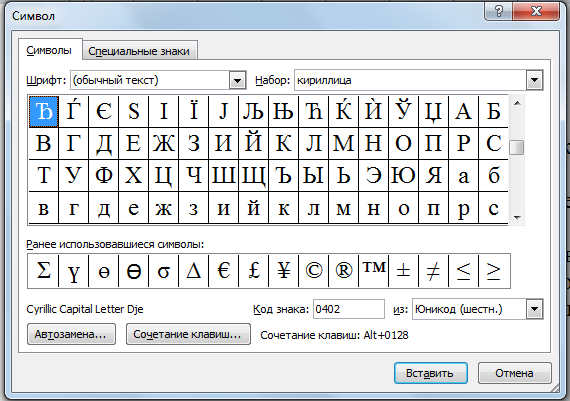
СР1251: DD C2 CC

СР866: 9D 82 8C

Мас: 9D 82 8C

ISO: CD B2 BC

1. Определить числовой код символа в кодировке Unicode с помощью текстового редактора Microsoft Word.
2. В операционной системе Windows запустить текстовый редактор Microsoft Word.
3. В текстовом редакторе Microsoft Word ввести команду [*Вставка-Символ…*]. На экране появится диалоговое окно *Символ.* Центральную часть диалогового окна занимает фрагмент таблицы символов.



1. Для определения числового кола знака кириллицы с помощью раскрывающегося списка *Набор*: выбрать пункт *кириллица*.
2. Для определения шестнадцатеричного числового кода символа в кодировке *Unicode* с помощью раскрывающегося списка *из*: выбрать тип кодировки *Юникод* (шестн.).
3. В таблице символов выбрать символ Э. В текстовом поле *код* *знака* : появится его шестнадцатеричный числовой код (в данном случае 042D).
4. **Закрепление изученного материала.**

*Решение задач.*

1. **Используем кодировочные таблицы**

**№1**

*Закодируйте с помощью кодировочной таблицы ASCII и представьте в шестнадцатеричной системе счисления следующие тексты:*

1. Password;
2. Windows;
3. Информация;
4. Paint.

***Решение*:**

1. Найдите в кодовой таблице ASCII коды соответствующих символов (в десятичной системе счисления)
2. Password → 80 97 115 115 119 111 114 100.
3. Windows → 87 105 110 100 111 119 115.
4. Информация → 200 237 244 238 240 236 224 246 232 255
5. Paint → 80 97 105 110 116
6. Переведите коды с помощью калькулятора в шестнадцатеричную систему счисления.
7. 80 97 115 115 119 111 114 100 → 50 61 73 73 77 6F 72 64
8. 87 105 110 100 111 119 115 → 57 69 6E 64 6F 77 73
9. 200 237 244 238 240 236 224 246 232 255→ C8 ED F4 EE F0 EC E0 F6 E8 FF
10. 80 97 105 110 116 → 50 61 69 6E 74

***Ответ*:**

1. 50 61 73 73 77 6F 72 64
2. 57 69 6E 64 6F 77 73
3. C8 ED F4 EE F0 EC E0 F6 E8 FF
4. 50 61 69 6E 74

**№2**

*Декодируйте с помощью кодировочной таблицы ASCII следующие тексты, заданные шестнадцатеричным кодом:*

1. 54 6F 72 6E 61 64 6F; (***Tornado***)
2. 49 20 6С 6F 76 65 20 79 6F 75; (***I love you***)
3. 32 2A 78 2B 79 3D 30. (***2+x+y=0***)
4. **Не используем кодировочные таблицы**

**№1**

*Буква «I »в таблице кодировки символов имеет десятичный код 105. что зашифровано последовательностью десятичных кодов: 108 105 110 107?*

***Решение*:**

Учитываем принцип последовательности кодирования и порядок букв в латинском алфавите и, можно, не обращаться к таблице кодировки символов.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Десятичный код** | *105* | *106* | *107* | *108* | *109* | *110* |
| **Латинская буква** | *i* | *j* | *k* | *l* | *m* | *n* |

***Ответ*:** Закодировано слово **«link»**

**№2**

*Десятичный код (номер) буквы «е» в таблице кодировки символов ASCII равен 101. Какая последовательность десятичных кодов будет соответствовать слову:*

*1) file; 2) help?*

***Решение*:**

Учитываем принцип последовательности кодирования и порядок букв в латинском алфавите:

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Десятичный код** | 101 | 102 | 103 | 104 | 105 | 106 | 107 | 108 | 109 | 110 | 111 | 112 |
| **Латинская буква** | **e** | **f** | **g** | **h** | **i** | **j** | **k** | **l** | **m** | **n** | **o** | **p** |

***Ответ*:**

* 1. 102 105 108 101
  2. 104 101 108 112

1. **Используем ПО (текстовый редактор Блокнот).**

Блокнот позволяет работать с текстами в кодировках ANSI и Unicode, а также выполнять преобразование из одного формата в другой. Для этого при сохранении документа выберите нужную кодировку в соответствующем поле.

**№1**

*Перейдите от двоичного кода к десятичному и декодируйте следующие тексты:*

а) 01010101 01110000 0100000 00100110 00100000 01000100 1101111 01110111 01101110;  
б) 01001001 01000010 01001101;  
в) 01000101 01101110 01110100 01100101 01110010

***Решение:***

1. Переведите коды из двоичной системы счисления в десятичную.  
а) 01010101 01110000 00100000 00100110 00100000 01000100 1101111 01110111 01101110 → 85 112 32 38 32 68 111 119 110   
б) 01001001 01000010 01001101 → 73 66 77   
в) 01000101 01101110 01110100 01100101 01110010 → 69 110 116 101 114  
2. Запустите текстовый редактор Блокнот.  
3. Включить клавишу Num Lock. Удерживая клавишу Alt, набрать код символа на цифровой клавиатуре. Отпустить клавишу Alt, на экране появится соответствующая буква.   
а) 85 112 32 26 32 68 111 119 110 → Up & Down;  
б) 73 66 77 → IBM;  
в) 69 110 116 101 114 → Enter

***Ответ*:** Up & Down; IBM; Enter

**№2.**

*Декодируйте следующие тексты, заданные десятичным кодом:*

1. 087 111 114 100;
2. 068 079 083;
3. 080 097 105 110 116 098 114 117 115 104.

***Решение:***

Запустите текстовый редактор Блокнот. Включить клавишу Num Lock. Удерживая клавишу Alt, набрать код символа на цифровой клавиатуре. Отпустить клавишу Alt, на экране появится соответствующая буква.

1. 087 111 114 100 → Word;
2. 068 079 083 → DOS;
3. 080 097 105 110 116 098 114 117 115 104 → Paintbrush.

***Ответ***: Word; DOS; Paintbrush.

1. **Итоги урока**

Оцениваем работу класса. Отмечаем отличившихся учащихся.

**Домашнее задание**

*Уровень знания:*

1. Как кодируется текстовая информация?
2. Где можно найти коды символов?
3. Почему существует несколько таблиц кодов? Чем они отличаются друг от друга?

*Уровень понимания:*

1. Заполните таблицу:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Двоичный код** | **Десятичный код** | **Символы** | | | | |
| **ASCII** | **KOИ-8** | **Windows 1251** | **ISO** | **Unicode** |
| 10000001 |  |  |  |  |  |  |
| 11000101 |  |  |  |  |  |  |
| 11100010 |  |  |  |  |  |  |
| 11101011 |  |  |  |  |  |  |
| 11110000 |  |  |  |  |  |  |
| 11111000 |  |  |  |  |  |  |
| 11111101 |  |  |  |  |  |  |
| 11111101 |  |  |  |  |  |  |

1. Какие последовательности букв будут записаны в кодировках Windows (СР) 1251 и КОИ-8 и соответствовать слову «СИМВОЛ», записанному в кодировке ASCII.