КРАТКАЯ ИСТОРИЯ ASCII

ASCII (американский стандартный код для информации обмен), стандартный код обмена информацией США. В 1960-х годах американцы разработали этот код персонажа, и английские персонажи сделали единые положения, в том числе 128 символов. Для этой реализации кодирования вам нужно только использовать один байт для представления, и используются только 7 битов. Фигура:

ASCII 字符代码表 一

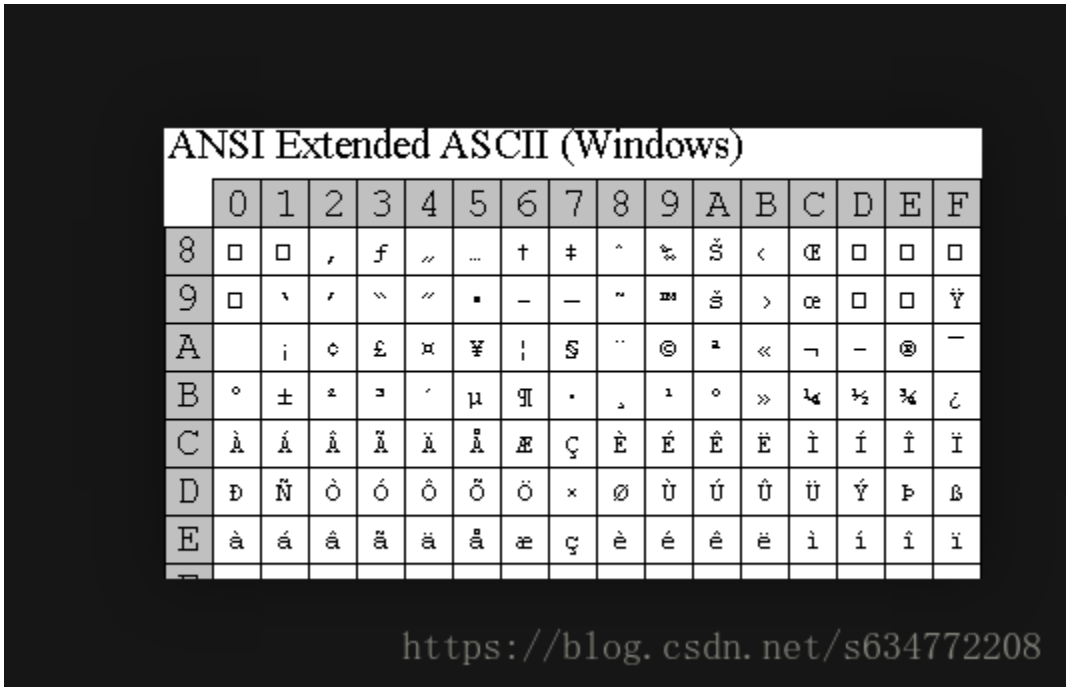
| 高四位 / 低四位 | | ASCII非打印控制字符 | | | | | | | | | | | ASCII 打印字符 | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 0000 | | | | | 0001 | | | | | | 0010 | | 0011 | | 0100 | | 0101 | | 0110 | | 0111 | | | | | |
| | | 0 | | | | | 1 | | | | | | 2 | | 3 | | 4 | | 5 | | 6 | | 7 | | | | | |
| | | +进制 | 字符 | ctrl | 代码 | 字符解释 | +进制 | 字符 | ctrl | 代码 | 字符解释 | | +进制 | 字符 | +进制 | 字符 | +进制 | 字符 | +进制 | 字符 | +进制 | 字符 | +进制 | 字符 | ctrl | | | |
| 0000 | 0 | 0 | BLANK NULL | ^@ | NUL | 空 | 16 | ► | ^P | DLE | 数据链路转意 | | 32 | | 48 | 0 | 64 | @ | 80 | P | 96 | ` | 112 | p | | | | |
| 0001 | 1 | 1 | ☺ | ^A | SOH | 头标开始 | 17 | ◄ | ^Q | DC1 | 设备控制 1 | | 33 | ! | 49 | 1 | 65 | A | 81 | Q | 97 | a | 113 | q | | | | |
| 0010 | 2 | 2 | ☻ | ^B | STX | 正文开始 | 18 | ↕ | ^R | DC2 | 设备控制 2 | | 34 | " | 50 | 2 | 66 | B | 82 | R | 98 | b | 114 | r | | | | |
| 0011 | 3 | 3 | ♥ | ^C | ETX | 正文结束 | 19 | ‼ | ^S | DC3 | 设备控制 3 | | 35 | # | 51 | 3 | 67 | C | 83 | S | 99 | c | 115 | s | | | | |
| 0100 | 4 | 4 | ♦ | ^D | EOT | 传输结束 | 20 | ¶ | ^T | DC4 | 设备控制 4 | | 36 | $ | 52 | 4 | 68 | D | 84 | T | 100 | d | 116 | t | | | | |
| 0101 | 5 | 5 | ♣ | ^E | ENQ | 查询 | 21 | § | ^U | NAK | 反确认 | | 37 | % | 53 | 5 | 69 | E | 85 | U | 101 | e | 117 | u | | | | |
| 0110 | 6 | 6 | ♠ | ^F | ACK | 确认 | 22 | ■ | ^V | SYN | 同步空闲 | | 38 | & | 54 | 6 | 70 | F | 86 | V | 102 | f | 118 | v | | | | |
| 0111 | 7 | 7 | • | ^G | BEL | 震铃 | 23 | ↨ | ^W | ETB | 传输块结束 | | 39 | ' | 55 | 7 | 71 | G | 87 | W | 103 | g | 119 | w | | | | |
| 1000 | 8 | 8 | ◘ | ^H | BS | 退格 | 24 | ↑ | ^X | CAN | 取消 | | 40 | ( | 56 | 8 | 72 | H | 88 | X | 104 | h | 120 | x | | | | |
| 1001 | 9 | 9 | ○ | ^I | TAB | 水平制表符 | 25 | ↓ | ^Y | EM | 媒体结束 | | 41 | ) | 57 | 9 | 73 | I | 89 | Y | 105 | i | 121 | y | | | | |
| 1010 | A | 10 | ◙ | ^J | LF | 换行/新行 | 26 | → | ^Z | SUB | 替换 | | 42 | * | 58 | : | 74 | J | 90 | Z | 106 | j | 122 | z | | | | |
| 1011 | B | 11 | ♂ | ^K | VT | 竖直制表符 | 27 | ← | ^[ | ESC | 转意 | | 43 | + | 59 | ; | 75 | K | 91 | [ | 107 | k | 123 | { | | | | |
| 1100 | C | 12 | ♀ | ^L | FF | 换页/新页 | 28 | ∟ | ^\ | FS | 文件分隔符 | | 44 | , | 60 | < | 76 | L | 92 | \ | 108 | l | 124 | \| | | | | |
| 1101 | D | 13 | ♪ | ^M | CR | 回车 | 29 | ↔ | ^] | GS | 组分隔符 | | 45 | - | 61 | = | 77 | M | 93 | ] | 109 | m | 125 | } | | | | |
| 1110 | E | 14 | ♫ | ^N | SO | 移出 | 30 | ▲ | ^6 | RS | 记录分隔符 | | 46 | . | 62 | > | 78 | N | 94 | ^ | 110 | n | 126 | ~ | | | | |
| 1111 | F | 15 | ☼ | ^O | SI | 移入 | 31 | ▼ | ^- | US | 单元分隔符 | | 47 | / | 63 | ? | 79 | O | 95 | _ | 111 | o | 127 | Δ | ^Back space | | | |

注：表中的ASCII字符可以用:ALT + "小键盘上的数字键" 输入

**Проблема появляется:**С развитием расчета этого набора стран с большим количеством персонажей, таких как европейские страны, нет нового кодирования персонажей.

EASCII

EASCII (расширенный американский стандартный код для информации обмен), расширяет стандартный код обмена информацией США. Чтобы пробиться через 127 символов под ASCII, некоторые европейские страны расширили ASCII, используя 8-й цифры байта, которые могут представлять 256 символов, как показано:




**Проблема появляется:** 256 символов для Китая, Япония, Южная Корея и т. Д., Намного меньше, поэтому есть еще один набор кодировков.

ANSI (Институт американского национального стандарта), американские национальные стандарты общества. Он отвечает за развитие некоммерческих организаций в национальных стандартах США. Для того, чтобы поддержать больше языков, разные страны и регионы разработали различные стандарты

и были признаны ASNI, и мир обычно используется этим кодированием при выражении соответствующего национального текста. Это называется кодировкой ANSI. Этот тип кода реализован, например, Китай**GB2312**Традиционный китайский**BIG5**, Япония**JIS**И многое другое. Один байт может представлять 256 символов, поэтому эта кодировка может быть представлена только несколькими байтами (Примечание: после обсуждения конкретной реализации).

**Проблема появляется:**Разные страны имеют различные реализации кодирования ANSI, взаимную совместимость. Чтобы отобразить текст, вы должны знать, что он кодируется, и он будет искажен. Так что есть более научный код.

UNICODE、UCS

- Unified Code Alliance, они состояли в производителях программного обеспечения, такие как Xerox, Apple и разработаны стандарт Unicode.**Текущая реализация Unicode: UTF-8, UTF-16, UTF-32.**
- Международная организация стандартизации (ISO) создает рабочую группу ISO / IEC JTC1 / SC2 / WG2 в 1984 году, попыталась разработать «универсальный набор символов, называемый UCS), и, наконец, разработка стандартов ISO 10646.**Текущая реализация UCS: UCS-2, UCS-4.**

До 1991 года участники двух проектов поняли, что мир не нуждается в двух несовместимых наборах персонажей. В результате они начинают объединить

результаты работы обеих сторон и работать вместе для одной таблицы кодирования. Начиная с Unicode 2.0, Unicode использует одну и ту же библиотеку шрифтов и код Word, поскольку ISO 10646-1; ISO также обещает, что ISO 10646 не будет назначать значение UCS-4, превышающих U + 10FFFF, чтобы оба они последовательны. Два проекта все еще существуют независимо, и публикуют свои соответствующие стандарты самостоятельно. Однако из-за названия Unicode он ширина.