字符集

常见字符集如 ASCII 字符集、ISO-8859-X、GB2312 字符集（简中）、BIG5 字符集（繁中）、GB18030 字符集、Shift-JIS 等等。之前的很长一段时间内，字符集和字符编码区分不是很严格，因为一般一种字符集只对应一种编码方式，但是这些编码方式存在下面的问题：

- 没有一种编码可以覆盖全世界所有国家的字符；
- 各种编码之间也会存在冲突的现象，两种不同编码方式可能使用同一个编码代表不同的字符，亦或用不同的编码代表同一个字符；
- 一个指定的机器（比如我们的服务器）将需要支持许多不同的编码方式，当数据在不同的机器之间传输或者在不同的编码之间转换时，很容易产生乱码问题。

ASCII 码

ASCII ((American Standard Code for Information Interchange): 美国信息交换标准代码）是基于拉丁字母的一套电脑编码系统，主要用于显示现代英语和其他西欧语言。它是最通用的信息交换标准，并等同于国际标准 ISO/IEC 646。ASCII 第一次以规范标准的类型发表是在 1967 年，最后一次更新则是在 1986 年，到目前为止共定义了 128 个字符。

每个 ASCII 码以 1 个字节(Byte)储存，使用指定的 8 位二进制数组合来表示可能的字符。从 0 到数字 127 代表不同的常用符号，例如大写 A 的 ASCII 码是 65，小写 a 则是 97。这套内码加上了许多外文和表格等特殊符号，成为目前常用的内码。

char 类型

char 类型的变量就是用来存储 ASCII 码的，我们可以直接把一个字符赋值给该变量，也可以直接将数码赋值给它，例如：

```
char ca1 = 'a';
char ca2 = 97;
cout<<"ca1:"<<ca1<<" ca2:"<<ca2<<endl;
//输出结果: ca1:a ca2:a
```

常见 ASCII 码的大小规则：0~9 <A~Z <a~z。

1）数字比字母要小。如 “7”<“F”；

2）数字 0 比数字 9 要小，并按 0 到 9 顺序递增。如 “3”<“8” ；

3）字母 A 比字母 Z 要小，并按 A 到 Z 顺序递增。如“A”<“Z” ；

4）同个字母的大写字母比小写字母要小 32。如“A”<“a” 。

5）ASCII 码的值范围为 0-127。

示例：

```
char a = 128; //输出异常
```

6）几个常见字母的 ASCII 码大小： “A”为 65；“a”为 97；“0”为 48

标准表

Bin(二进制)	十进制	缩写/字符	解释
0000 0000	0	NUL(null)	空字符
0000 0001	1	SOH(start of headline)	标题开始
0000 0010	2	STX (start of text)	正文开始
0000 0011	3	ETX (end of text)	正文结束
0000 0100	4	EOT (end of transmission)	传输结束
0000 0101	5	ENQ (enquiry)	请求
0000 0110	6	ACK (acknowledge)	收到通知
0000 0111	7	BEL (bell)	响铃
0000 1000	8	BS (backspace)	退格
0000 1001	9	HT (horizontal tab)	水平制表符
0000 1010	10	LF (NL line feed, new line)	换行键
0000 1011	11	VT (vertical tab)	垂直制表符
0000 1100	12	FF (NP form feed, new page)	换页键
0000 1101	13	CR (carriage return)	回车键
0000 1110	14	SO (shift out)	不用切换
0000 1111	15	SI (shift in)	启用切换
0001 0000	16	DLE (data link escape)	数据链路转义
0001 0001	17	DC1 (device control 1)	设备控制 1
0001 0010	18	DC2 (device control 2)	设备控制 2
0001 0011	19	DC3 (device control 3)	设备控制 3
0001 0100	20	DC4 (device control 4)	设备控制 4
0001 0101	21	NAK (negative acknowledge)	拒绝接收
0001 0110	22	SYN (synchronous idle)	同步空闲
0001 0111	23	ETB (end of trans. block)	结束传输块
0001 1000	24	CAN (cancel)	取消
0001 1001	25	EM (end of medium)	媒介结束
0001 1010	26	SUB (substitute)	代替
0001 1011	27	ESC (escape)	换码(溢出)
0001 1100	28	FS (file separator)	文件分隔符
0001 1101	29	GS (group separator)	分组符
0001 1110	30	RS (record separator)	记录分隔符
0001 1111	31	US (unit separator)	单元分隔符
0010 0000	32	(space)	空格
0010 0001	33	!	叹号

0010 0010	34	"	双引号
0010 0011	35	#	井号
0010 0100	36	$	美元符
0010 0101	37	%	百分号
0010 0110	38	&	和号
0010 0111	39	'	闭单引号
0010 1000	40	(	开括号
0010 1001	41	)	闭括号
0010 1010	42	*	星号
0010 1011	43	+	加号
0010 1100	44	,	逗号
0010 1101	45	-	减号/破折号
0010 1110	46	.	句号
0010 1111	47	/	斜杠
0011 0000	48	0	字符 0
0011 0001	49	1	字符 1
0011 0010	50	2	字符 2
0011 0011	51	3	字符 3
0011 0100	52	4	字符 4
0011 0101	53	5	字符 5
0011 0110	54	6	字符 6
0011 0111	55	7	字符 7
0011 1000	56	8	字符 8
0011 1001	57	9	字符 9
0011 1010	58	:	冒号
0011 1011	59	;	分号
0011 1100	60	<	小于
0011 1101	61	=	等号
0011 1110	62	>	大于
0011 1111	63	?	问号
0100 0000	64	@	电子邮件符号
0100 0001	65	A	大写字母 A
0100 0010	66	B	大写字母 B
0100 0011	67	C	大写字母 C
0100 0100	68	D	大写字母 D
0100 0101	69	E	大写字母 E
0100 0110	70	F	大写字母 F

0100 0111	71	G	大写字母 G
0100 1000	72	H	大写字母 H
0100 1001	73	I	大写字母 I
1001010	74	J	大写字母 J
0100 1011	75	K	大写字母 K
0100 1100	76	L	大写字母 L
0100 1101	77	M	大写字母 M
0100 1110	78	N	大写字母 N
0100 1111	79	O	大写字母 O
0101 0000	80	P	大写字母 P
0101 0001	81	Q	大写字母 Q
0101 0010	82	R	大写字母 R
0101 0011	83	S	大写字母 S
0101 0100	84	T	大写字母 T
0101 0101	85	U	大写字母 U
0101 0110	86	V	大写字母 V
0101 0111	87	W	大写字母 W
0101 1000	88	X	大写字母 X
0101 1001	89	Y	大写字母 Y
0101 1010	90	Z	大写字母 Z
0101 1011	91	[	开方括号
0101 1100	92	\	反斜杠
0101 1101	93	]	闭方括号
0101 1110	94	^	脱字符
0101 1111	95	_	下划线
0110 0000	96	`	开单引号
0110 0001	97	a	小写字母 a
0110 0010	98	b	小写字母 b
0110 0011	99	c	小写字母 c
0110 0100	100	d	小写字母 d
0110 0101	101	e	小写字母 e
0110 0110	102	f	小写字母 f
0110 0111	103	g	小写字母 g
0110 1000	104	h	小写字母 h
0110 1001	105	i	小写字母 i
0110 1010	106	j	小写字母 j
0110 1011	107	k	小写字母 k

0110 1100	108	l	小写字母 l
0110 1101	109	m	小写字母 m
0110 1110	110	n	小写字母 n
0110 1111	111	o	小写字母 o
0111 0000	112	p	小写字母 p
0111 0001	113	q	小写字母 q
0111 0010	114	r	小写字母 r
0111 0011	115	s	小写字母 s
0111 0100	116	t	小写字母 t
0111 0101	117	u	小写字母 u
0111 0110	118	v	小写字母 v
0111 0111	119	w	小写字母 w
0111 1000	120	x	小写字母 x
0111 1001	121	y	小写字母 y
0111 1010	122	z	小写字母 z
0111 1011	123	{	开花括号
0111 1100	124 |		垂线
0111 1101	125	}	闭花括号
0111 1110	126	~	波浪号
0111 1111	127	DEL (delete)	删除

小提示：

1、char 类型转为 int 时，即将 ASCII 码值赋给 int 型变量。

示例：

```
char a = 'a';
int b = a;
printf("%d",b);  //输出 97
```

2、char 类型与整形相加时，即将当前的 ASCII 码值按整形数值递增，其本质是类型转换规则，后文会讲到。

示例：

```
char a = 'a';
char b = a + 1;
printf("%c",b);   //输出 b
```

3、ASCII 码的值范围为 0-127。

示例：

```
char a = 128; //输出异常
```

4、注意 char 类型的变量赋值为字符时，必须使用单引号 ‘’。

Unicode

多语言软件制造商组成的统一码联盟(The Unicode Consortium)于 1991 年发布的统一码标准（The Unicode Standard)，定义了一个全球统一的通用字符集即 Unicode 字符集解决了上述的问题。统一码标准为每个字符提供一个唯一的编号，旨在支持世界各地的交流，处理和显示现代世界各种语言和技术学科的书面文本。此外，它支持许多书面语言的古典和历史文本，不管是什么平台，设备，应用程序或语言，都不会造成乱码问题， 它已被所有现代软件供应商采用，是目前所有主流操作系统，搜索引擎，浏览器，笔记本电脑和智能手机以及互联网和万维网（URL，HTML，XML，CSS，JSON 等）中表示语言和符号的基础。统一码标准的一个版本由核心规范、Unicode 标准、代码图、Unicode 标准附件以及 Unicode 字符数据库（Unicode Character Database 简写成 UCD）组成，同时也是开发的字符集，在不断的更加和增加新的字符，最新的版本为 Unicode 10.0.0。

Unicode 编码字符集旨在收集全球所有的字符，为每个字符分配唯一的字符编号即代码点。从这一点我们可以看出，Unicode 编码的位数必然是不确定的，且支持增长。

编码方式

UTF-8，UTF-16 ，UTF-32

小提示：打开百度网站：www.baidu.com。右键“查看页面源代码”，我们可以清晰的看到它使用的是“utf-8”字符集。

```
<!-- This Source Code Form is subject to the terms of the Mozilla Public
   - License, v. 2.0. If a copy of the MPL was not distributed with this file,
     You can obtain one at http://mozilla.org/MPL/2.0/. --><!doctype html><html>
<head><meta charset="utf-8"><meta name="defaultLanguage" content="en-US"><meta
name="availableLanguages" content="en-US, zh-CN"><link rel="localization"
href="geckoBrandFtl"><link rel="localization" href="geckoBrandingsFtl"><link
rel="localization" href="geckoNewtabFtl"><link rel="localization"
href="mococnFtl"><title data-l10n-id="newtab-page-title"></title><link
rel="icon" type="image/png" href="chrome://branding/content/icon32.png"><link
rel="stylesheet" href="chrome://browser/content/contentSearchUI.css"><link
href="activity-stream.3b55be4821f1643683c7.css" rel="stylesheet"
integrity="sha512-varGKqbyd6gIhi5LVApAEzuTdg/a1T1xZiZqSBP7xB/icmEzZ8PHrPrhiLTk1v1vbRkvb7zLp1J4ycsuu64qzQ=="
crossorigin="anonymous"></head><body class="activity-stream"><div
id="header-asrouter-container" role="presentation"></div><div id="root"></div>
<div id="footer-asrouter-container" role="presentation"></div><script
type="text/javascript" src="vendors.4f8ed193863d5ac0fbbf.js"
integrity="sha512-O6i0Svnr+wJHwOs2yCye03Ir1ROhPPeagdchp6XpIwMaZniP/bo10ruX0UfUJIkjmaLSygKGyHsGIRLSCF/8eQ=="
crossorigin="anonymous"></script><script type="text/javascript"
src="gecko.8aca2b0a221a3e678341.js"
integrity="sha512-A/NVuJdN9P67bOtt81YZU+ZMmtwm/0rUK1TO1IL72nfINJ4SO+OjyxmVSW0ySWsfrHPua2EA174qcJ4IAMmEPQ=="
crossorigin="anonymous"></script><script type="text/javascript"
src="activity-stream.22623124fb6195e509c5.js"
integrity="sha512-xYr+JeYLd6a3kzJWuWFq5b5ozry6HrNm1Gp+SGq9LbD6GXEF+cNAJNVUG2mfHRWYmWMAOayDuVts4Qv16wzjHg=="
crossorigin="anonymous"></script></body></html>
```

辑航线培优教育，信息学奥赛培训专家。

扫码添加作者获取更多内容。