



计算思维通识教育

Computational Thinking

第1章 信息的存储与表示

主讲人：曹轶臻

联系方式：caoyizhen@cuc.edu.cn

学习程序设计的目标不是把每个人都变成程序员，而是帮助每个人建立起系统化的逻辑思维能力和习惯。

01

信息的存储与表示

信息存储和表示的基础 1-1

数值的表示 1-2

文本的表示 1-3

音频的表示 1-4

图像的表示 1-5

视频的表示 1-6



计算机与网络空间安全学院
School of Computer and Cyber Sciences

计算思维通识教育
Computational Thinking



以4bit存储系统为例

存储范围

十进制

0

1

2

3

.

.

.

14

15

二进制

0000

0001

0010

0011

.

.

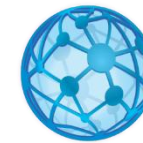
.

1110

1111

二进制的直接表示

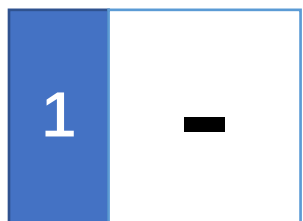
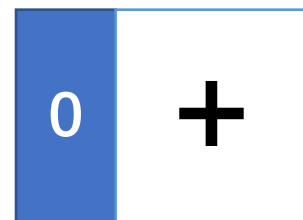
Question: 怎么表示负数?



原码表示法



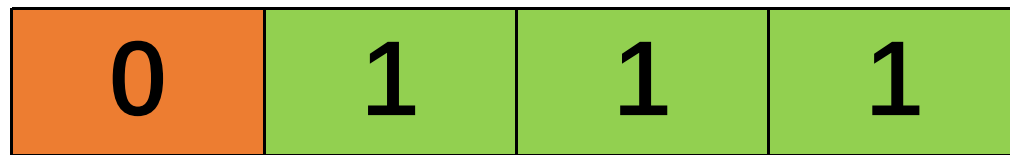
以4bit存储系统为例



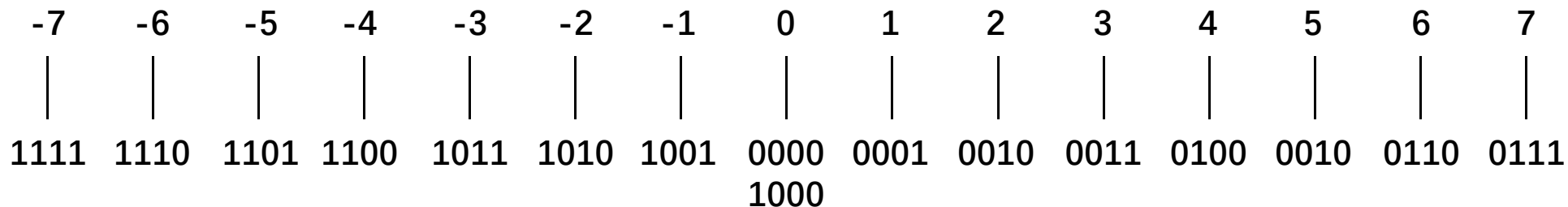


以4bit存储系统为例

最大值 +7



最小值 -7



原码

原码不能直接运算



补码

CPU中只有加法器，通过加法实现减法！

$$2 - 2 = 2 + (-2) = 0$$

BUT

$$\begin{array}{r} (+2) \\ + (-2) \\ \hline 0 \end{array}$$

~~$$\begin{array}{r} 0010 \quad (+2) \\ + 1010 \quad (-2) \\ \hline 1100 \quad (-4) \end{array}$$~~

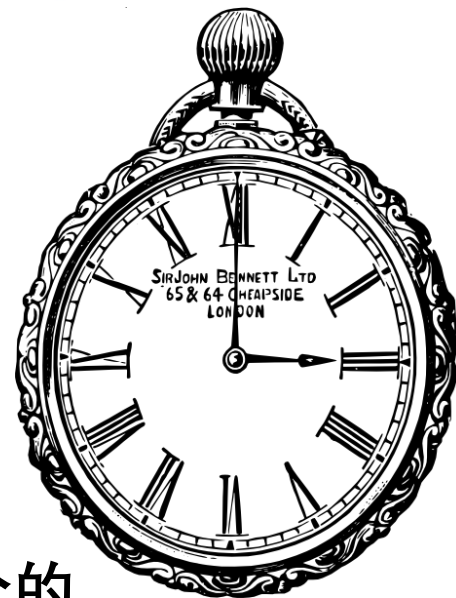
补码

将这个钟拨回12点有两种方式！

$$3 - 3 = 3 + 9 = 12 \text{ (0)}$$

在以12为模的时钟中，-3的补数为9

加一个负数和加这个负数的补数是等价的



4bit

原码		取反		加1		溢出	结果
0011	+	1100	+	0001	=	1	0000
		1101					

-3的补码表示法

补码

正数的补码：

正数



原码

例： +5

0101

补码的最高位也代表符号

负数的补码：

负数



绝对值



原码



取反



加1

例： -5

5

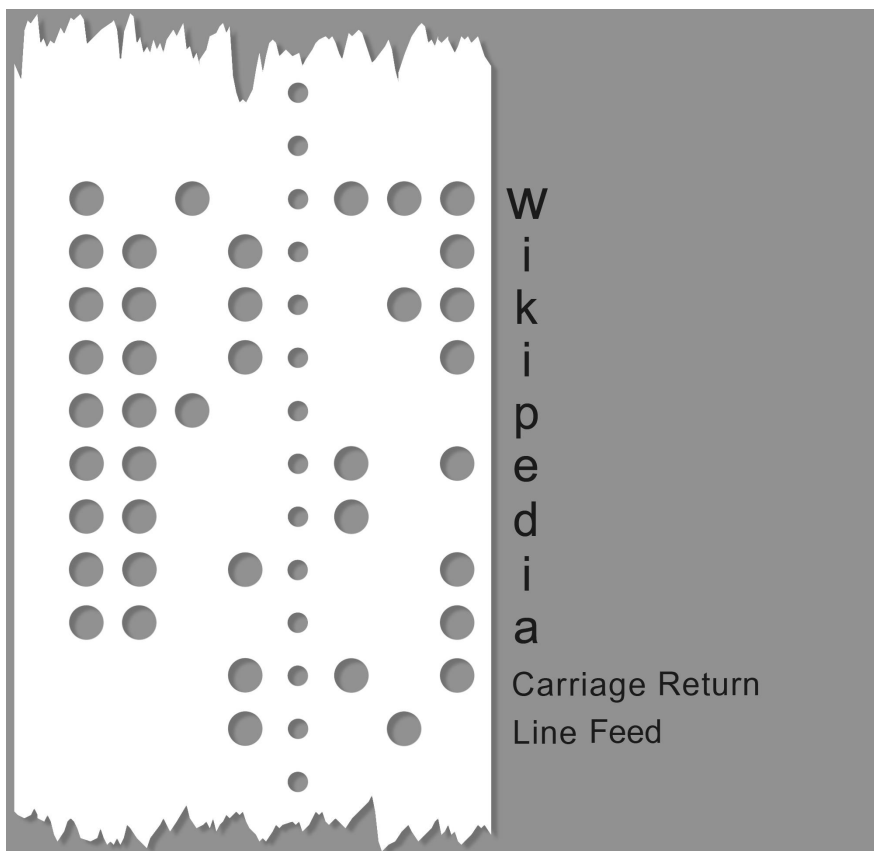
0101

1010

1011

验证： $7-5 = 7+(-5) = 0111 + 1011 = 0010 = 2$

计算机中使用补码进行有符号整数的表示和存储。



1

为什么要进行字符编码？

2

ASCII码

3

扩展ASCII码

4

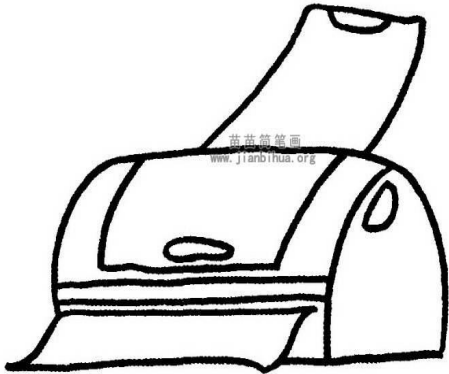
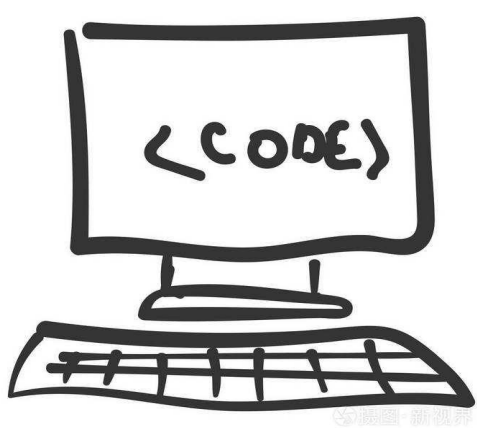
GB2312字符集及编码方式

5

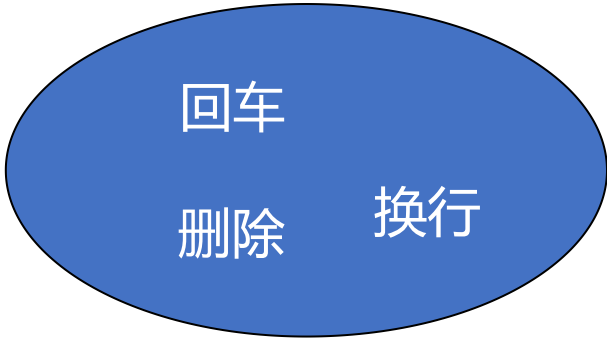
UNICODE编码

1-3-1

为什么要进行字符编码



控制字符



可见字符

码位	解释
0	空字符
1	标题开始
2	正文开始
...	...
10	换行
...	...
13	回车
...	...
31	单元分隔符
32	空格
...	...
48	0
...	...
57	9
...	...
65	A
...	...
90	Z
...	...
97	a
...	...
122	z
...	...
126	~
127	删除

ASCII
字符集

128个

33个控制
字符

95个可显示
字符

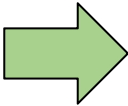
1-3-2 ASCII码

American Standard Code
for Information
Interchange

码位

ASCII
字符集

0	空字符
1	标题开始
2	正文开始
...	...
10	换行
...	...
13	回车
...	...
31	单元分隔符
32	空格
...	...
48	0
...	...
57	9
...	...
65	A
...	...
90	Z
...	...
97	a
...	...
122	z
...	...
126	~
127	删除



0000 0000
0000 0001
0000 0010
...
0000 1010
...
0000 1101
...
0001 1111
0010 0000
...
0011 0000
...
0011 1001
...
0100 0001
...
0101 1010
...
0110 0001
...
0111 1010
...
0111 1110
0111 1111

二进制

ASCII码

1-3-2 ASCII

www.ascii-code.com





ASCII可显示字符 (共95个)

二进制	十进制	十六进制	图形	二进制	十进制	十六进制	图形	二进制	十进制	十六进制	图形
0010 0000	32	20	(space)	0100 0000	64	40	@	0110 0000	96	60	`
0010 0001	33	21	!	0100 0001	65	41	A	0110 0001	97	61	a
0010 0010	34	22	"	0100 0010	66	42	B	0110 0010	98	62	b
0010 0011	35	23	#	0100 0011	67	43	C	0110 0011	99	63	c
0010 0100	36	24	\$	0100 0100	68	44	D	0110 0100	100	64	d
0010 0101	37	25	%	0100 0101	69	45	E	0110 0101	101	65	e
0010 0110	38	26	&	0100 0110	70	46	F	0110 0110	102	66	f
0010 0111	39	27	'	0100 0111	71	47	G	0110 0111	103	67	g
0010 1000	40	28	(0100 1000	72	48	H	0110 1000	104	68	h
0010 1001	41	29)	0100 1001	73	49	I	0110 1001	105	69	i
0010 1010	42	2A	*	0100 1010	74	4A	J	0110 1010	106	6A	j
0010 1011	43	2B	+	0100 1011	75	4B	K	0110 1011	107	6B	k
0010 1100	44	2C	,	0100 1100	76	4C	L	0110 1100	108	6C	l
0010 1101	45	2D	-	0100 1101	77	4D	M	0110 1101	109	6D	m
0010 1110	46	2E	.	0100 1110	78	4E	N	0110 1110	110	6E	n
0010 1111	47	2F	/	0100 1111	79	4F	O	0110 1111	111	6F	o
0011 0000	48	30	0	0101 0000	80	50	P	0111 0000	112	70	p
0011 0001	49	31	1	0101 0001	81	51	Q	0111 0001	113	71	q
0011 0010	50	32	2	0101 0010	82	52	R	0111 0010	114	72	r
0011 0011	51	33	3	0101 0011	83	53	S	0111 0011	115	73	s
0011 0100	52	34	4	0101 0100	84	54	T	0111 0100	116	74	t
0011 0101	53	35	5	0101 0101	85	55	U	0111 0101	117	75	u
0011 0110	54	36	6	0101 0110	86	56	V	0111 0110	118	76	v
0011 0111	55	37	7	0101 0111	87	57	W	0111 0111	119	77	w
0011 1000	56	38	8	0101 1000	88	58	X	0111 1000	120	78	x
0011 1001	57	39	9	0101 1001	89	59	Y	0111 1001	121	79	y
0011 1010	58	3A	:	0101 1010	90	5A	Z	0111 1010	122	7A	z
0011 1011	59	3B	;	0101 1011	91	5B	[0111 1011	123	7B	{
0011 1100	60	3C	<	0101 1100	92	5C	\	0111 1100	124	7C	
0011 1101	61	3D	=	0101 1101	93	5D]	0111 1101	125	7D	}
0011 1110	62	3E	>	0101 1110	94	5E	^	0111 1110	126	7E	~
0011 1111	63	3F	?	0101 1111	95	5F	_				

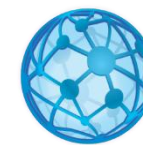
ASCII控制字符 (共33个)

二进制	十进制	十六进制	缩写	Unicode 表示法	脱出字符 表示法	名称 / 意义
0000 0000	0	00	NUL	NUL	^@	空字符 (Null)
0000 0001	1	01	SOH	SOH	^A	标题开始
0000 0010	2	02	STX	STX	^B	本文开始
0000 0011	3	03	ETX	ETX	^C	本文结束
0000 0100	4	04	EOT	EOT	^D	传输结束
0000 0101	5	05	ENQ	ENQ	^E	请求
0000 0110	6	06	ACK	ACK	^F	确认回应
0000 0111	7	07	BEL	BEL	^G	响铃
0000 1000	8	08	BS	BS	^H	退格
0000 1001	9	09	HT	HT	^I	水平定位符号
0000 1010	10	0A	LF	LF	^J	换行键
0000 1011	11	0B	VT	VT	^K	垂直定位符号
0000 1100	12	0C	FF	FF	^L	换页键
0000 1101	13	0D	CR	CR	^M	CR (字符)
0000 1110	14	0E	SO	SO	^N	取消变换 (Shift out)
0000 1111	15	0F	SI	SI	^O	启用变换 (Shift in)
0001 0000	16	10	DLE	DLE	^P	跳出数据通讯
0001 0001	17	11	DC1	DC1	^Q	设备控制一 (XON 激活软件速度控制)
0001 0010	18	12	DC2	DC2	^R	设备控制二
0001 0011	19	13	DC3	DC3	^S	设备控制三 (XOFF 停用软件速度控制)
0001 0100	20	14	DC4	DC4	^T	设备控制四
0001 0101	21	15	NAK	NAK	^U	确认失败回应
0001 0110	22	16	SYN	SYN	^V	同步用暂停
0001 0111	23	17	ETB	ETB	^W	区块传输结束
0001 1000	24	18	CAN	CAN	^X	取消
0001 1001	25	19	EM	EM	^Y	连线介质中断
0001 1010	26	1A	SUB	SUB	^Z	替换
0001 1011	27	1B	ESC	ESC	^[退出键
0001 1100	28	1C	FS	FS	^\	文件分割符
0001 1101	29	1D	GS	GS	^]	族群分隔符
0001 1110	30	1E	RS	RS	^^	记录分隔符
0001 1111	31	1F	US	US	^_	单元分隔符\
32~126: ASCII可显示字符						
0111 1111	127	7F	DEL	DEL	^?	Delete字符

部分常见和有趣的控制字符

-  BEL (**7**) : BEL, **响铃**。BEL 用一个可以听得见的声音来吸引人们的注意, 既可以用于计算机, 也可以用于周边设备 (比如打印机) 。
-  BS (**8**) : BackSpace, **退格键**。起初意思是, 在打印机和电传打字机上, 往回移动一格光标, 以起到强调该字符的作用。现代所用的退格键, 不仅仅表示光标往回移动了一格, 同时也删除了移动后该位置的字符。
-  VT (**11**) : Vertical Tab, **垂直制表符**。类似于水平制表符 Tab, 目的是为了减少布局中的工作, 同时也减少了格式化字符时所需要存储字符的空间。VT 控制符用于跳到下一个标记行。
-  DEL (**127**) : Delete, **删除**。127 二进制值为 1111111 (所有7个比特位都是1) 。

1-3-2 ASCII码



计算机与网络空间安全学院
School of Computer and Cyber Sciences

ASCII码

0000 0000

太少了

0111 1111

从0到127
共128个字符

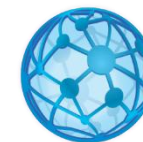
1000 0000

扩展ASCII码

1111 1111

从128扩展到255
新增了128个字符

1-3-3 扩展ASCII码



EASCII码比ASCII码扩充出来的符号包括表格符号、计算符号、希腊字母和特殊的拉丁符号。

- 128到255这一页的字符集被称“扩展字符集”。
- 扩展字符集加上原来的127个符号，共计256符号，一共使用了8比特信息。

www.ascii-code.com

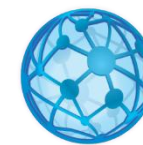
1000 0000

Extended ASCII Codes

128	Ç	144	É	160	á	176	░	192	Ł	208	⋈	224	α	240	≡
129	ü	145	æ	161	í	177	▒	193	±	209	⌚	225	β	241	±
130	é	146	Æ	162	ó	178	▓	194	⌚	210	⌚	226	Γ	242	≥
131	â	147	ô	163	ú	179		195	⌚	211	⋈	227	π	243	≤
132	ä	148	ö	164	ñ	180	⌚	196	—	212	⌚	228	Σ	244	∫
133	à	149	ò	165	Ñ	181	⌚	197	+	213	⌚	229	σ	245	∫
134	â	150	û	166	²	182	⌚	198	⌚	214	⌚	230	μ	246	÷
135	ç	151	ù	167	°	183	⌚	199	⌚	215	⌚	231	τ	247	≈
136	ê	152	ÿ	168	¿	184	⌚	200	⌚	216	⌚	232	Φ	248	°
137	ë	153	Ö	169	⌚	185	⌚	201	⌚	217	⌚	233	⊙	249	⌚
138	è	154	Û	170	⌚	186	⌚	202	⌚	218	⌚	234	Ω	250	⌚
139	ï	155	◊	171	½	187	⌚	203	⌚	219	■	235	δ	251	√
140	î	156	£	172	¼	188	⌚	204	⌚	220	■	236	∞	252	π
141	ì	157	¥	173	⅓	189	⌚	205	=	221	■	237	φ	253	²
142	Ä	158	⌚	174	«	190	⌚	206	⌚	222	■	238	ε	254	■
143	Å	159	f	175	»	191	⌚	207	±	223	■	239	∩	255	

Source: www.LookupTables.com

1-3-3 扩展ASCII码 -- Windows-1252



Windows-1252 是微软为西欧语言（如英语、法语、德语、西班牙语等）创建的一种具体的、标准的字符编码。由于 Windows 操作系统的巨大影响力，Windows-1252 成为了在互联网和西方世界最普及、最常被提及的“扩展 ASCII”编码。其他扩展ASCII标准还有ISO/IEC 8859-1 (Latin-1)等。

^N_U ^S_O ^S_T ^E_T ^E_O ^E_N ^A_C ^B_E ^B_S ^H_T ^L_F ^V_T ^F_R ^S_O ^S_I ^D_L ^D_C₁ ^D_C₂ ^D_C₃ ^D_C₄ ^N_A ^S_Y ^E_T ^C_A ^E_M ^S_U ^E_S ^F_S ^G_S ^R_S ^U_S

! " # \$ % & ' () * + , - . / 0 1 2 3 4 5 6 7 8 9 : ; < = > ?

@ A B C D E F G H I J K L M N O P Q R S T U V W X Y Z [\] ^ _

` a b c d e f g h i j k l m n o p q r s t u v w x y z { | } ~ ^D_E_L

€ • , f „ … † ‡ ^ % Š < Œ • Ž • • ' ' " " • — ~™ Š > œ • ž Ÿ

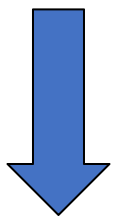
¡ ¢ £ ¤ ¥ ¦ § ¨ © ª « ¬ ® ¯ ° ± ² ³ ´ µ ¶ · ¸ ¹ º » ¼ ½ ¾ ¿

À Á Â Ã Ä Å Æ Ç È É Ê Ë Ì Í Î Ï Ð Ñ Ò Ó Ô Õ Ö × Ø Ù Ú Û Ü Ý Þ ß

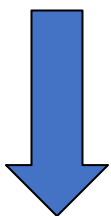
à á â ã ä å æ ç è é ê ë ì í î ï ð ñ ò ó ô õ ö ÷ ø ù ú û ü ý þ ÿ

<https://en.wikipedia.org/wiki/Windows-1252>

当电脑来到中国时



扩展后的ASCII码
(256个字符)



用16位表示一个字符

中文多少个汉字

- 东汉的《说文解字》收字 9353 个
- 宋代《集韵》收字 53525 个
- 清朝《康熙字典》收字 47035 个
- 当代的《汉语大字典》收字 60370 个(2010年版)。
- 1994年中华书局、中国友谊出版公司出版的《中华字海》收字 85568 个，这是目前收字最多的字典。

国家语言文字工作委员会于1988年颁布的《现代汉语常用字表》收录3,500字（2,500个常用字，1,000个次常用字）。

先设计字符集

分区管理的方式，共计94个区，每个区含94个位，共8836个码位。

- 01—09区收录除汉字以外的682个字符
- 10—15区为空白区，没有使用
- 16—55区收录3755个一级汉字，按拼音排序。
- 56—87区收录3008个二级汉字，按部首/比划排序。
- 88—94区为空白区，没有使用。

GB2312字符集

1980年发布，1981年5月1日
实施



01-09区表示除汉字之外的682个字符

01	0	1	2	3	4	5	6	7	8	9
0			,	.	-	~	..	"	々	
1	—	~	≡	…	‘	’	“	”	()
2	<	>	≪	≫	「	」	『	』	【	】
3	【	】	±	×	÷	:	∧	∨	Σ	Π
4	∩	∪	≡	≡	≈	≈	∞	∞	∞	∞
5	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞
6	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞
7	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞
8	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞
9	→	←	↑	↓	≡	≡	≡	≡	≡	≡

02	0	1	2	3	4	5	6	7	8	9
0		i	ii	iii	iv	v	vi	vii	viii	ix
1	x	□	□	□	□	□	1.	2.	3.	
2	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.
3	14.	15.	16.	17.	18.	19.	20.	(1)	(2)	(3)
4	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
5	(14)	(15)	(16)	(17)	(18)	(19)	(20)	①	②	③
6	④	⑤	⑥	⑦	⑧	⑨	⑩	e	□	(-)
7	(-)	(=)	(四)	(五)	(六)	(七)	(八)	(九)	(+)	□
8	□	I	II	III	IV	V	VI	VII	VIII	IX
9	X	XI	XII	□	□					

03	0	1	2	3	4	5	6	7	8	9
0		!	"	#	¥	%	&	'	()
1	*	+	,	-	.	/	0	1	2	3
2	4	5	6	7	8	9	:	;	<	=
3	>	?	@	A	B	C	D	E	F	G
4	H	I	J	K	L	M	N	O	P	Q
5	R	S	T	U	V	W	X	Y	Z	[
6	\]	^	_	`	a	b	c	d	e
7	f	g	h	i	j	k	l	m	n	o
8	p	q	r	s	t	u	v	w	x	y
9	z	{		}						

0388

04	0	1	2	3	4	5	6	7	8	9
0		あ	か	さ	し	ち	に	ぬ	ふ	み
1	お	か	さ	し	ち	に	ぬ	ふ	み	ら
2	ご	さ	だ	な	に	ふ	み	む	る	
3	ぞ	と	ど	び	に	ふ	み	む	る	
4	と	ど	び	に	ふ	み	む	る		
5	と	ど	び	に	ふ	み	む	る		
6	と	ど	び	に	ふ	み	む	る		
7	と	ど	び	に	ふ	み	む	る		
8	と	ど	び	に	ふ	み	む	る		
9	と	ど	び	に	ふ	み	む	る		

05	0	1	2	3	4	5	6	7	8	9
0		ア	カ	サ	シ	チ	ニ	ヌ	フ	ミ
1	オ	カ	サ	シ	チ	ニ	ヌ	フ	ミ	ム
2	ゴ	サ	ダ	ナ	ビ	ブ	メ	モ	レ	ケ
3	ゾ	タ	ダ	ナ	ビ	ブ	メ	モ	レ	ケ
4	ト	ド	ナ	ビ	ブ	メ	モ	レ	ケ	
5	ヒ	ビ	フ	ブ	メ	モ	レ	ケ		
6	ホ	ボ	マ	ミ	ム	メ	モ	レ	ケ	
7	ユ	ヨ	ラ	リ	ル	レ	ケ			
8	キ	ク	シ	ス	セ	ソ	デ	パ	ホ	ユ
9	□	□	□	□	□	□	□	□	□	□

06	0	1	2	3	4	5	6	7	8	9
0		A	B	Γ	Δ	E	Z	H	Θ	I
1	K	Λ	M	N	Ξ	O	Π	P	Σ	T
2	τ	Φ	X	Ψ	Ω	□	□	□	□	□
3	□	□	□	α	β	γ	δ	ε	ζ	η
4	θ	ι	κ	λ	μ	ν	ξ	ο	π	ρ
5	σ	τ	υ	φ	χ	ψ	ω	□	□	□
6	:	;	!	?	()	□	□	□	□
7	《	》	「	」	『	』	【	】	□	□
8	^	~		!	!	!	□	□	□	□
9	□	□	□	□	□	□	□	□	□	□

07	0	1	2	3	4	5	6	7	8	9
0		A	B	В	Г	Д	Е	Ё	Ж	З
1	И	Й	К	Л	М	Н	О	П	Р	С
2	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы
3	Ь	Э	Ю	Я	□	□	□	□	□	□
4	□	□	□	□	□	□	□	□	□	а
5	б	в	г	д	е	ё	ж	з	и	й
6	к	л	м	н	о	п	р	с	т	у
7	ф	х	ц	ч	ш	щ	ъ	ы	ь	э
8	ю	я	□	□	□	□	□	□	□	□
9	□	□	□	□	□	□	□	□	□	□

08	0	1	2	3	4	5	6	7	8	9
0		ā	á	ǎ	à	ē	é	ě	è	ì
1	í	ï	ï	ō	ó	õ	ò	ú	û	
2	ù	ū	ū	ü	ü	ê	á	ñ		
3	ñ	à	g	□	□	□	□	□	□	□
4	□	□	□	□	□	□	□	□	□	□
5	□	□	□	□	□	□	□	□	□	□
6	□	□	□	□	□	□	□	□	□	□
7	□	□	□	□	□	□	□	□	□	□
8	□	□	□	□	□	□	□	□	□	□
9	□	□	□	□	□	□	□	□	□	□

09	0	1	2	3	4	5	6	7	8	9
0		□	□	□	—	—			—	—
1	!	!	!	!	!	!	!	!	!	!
2	□	□	□	□	□	□	□	□	□	□
3	□	□	□	□	□	□	□	□	□	□
4	□	□	□	□	□	□	□	□	□	□
5	□	□	□	□	□	□	□	□	□	□
6	□	□	□	□	□	□	□	□	□	□
7	□	□	□	□	□	□	□	□	□	□
8	□	□	□	□	□	□	□	□	□	□
9	□	□	□	□	□	□	□	□	□	□



16-55区收录了3755个一级汉字，按拼音排序

16	0	1	2	3	4	5	6	7	8	9
0	藹	啊	阿	埃	挨	哎	唉	哀	皑	癌
1	按	矮	艾	碍	爱	隘	鞍	氨	安	俺
2	熬	暗	岸	胺	案	肮	昂	盎	凹	敖
3	叭	翱	袄	胺	奥	懊	澳	芭	捌	扒
4	吧	吧	芭	八	疤	巴	拔	跋	靶	把
5	耙	坝	霸	罢	爸	白	柏	百	摆	佰
6	版	拜	裨	伴	班	搬	办	般	颁	帮
7	梆	膀	拌	绑	瓣	半	蚌	镑	邦	谤
8	苞									
9										

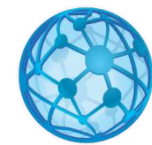
1692

17	0	1	2	3	4	5	6	7	8	9
0	豹	薄	雹	保	堡	饱	宝	抱	报	暴
1	贝	鲍	爆	杯	碑	悲	卑	北	辈	背
2	比	笨	倍	狈	备	惫	焙	被	奔	苯
3	币	鄙	崩	绷	甬	泵	蔽	迸	逼	鼻
4	避	庇	笔	彼	碧	蓖	必	毕	毙	臂
5	辨	陛	痹	闭	敝	弊	扁	譬	壁	卞
6	别	辩	鞭	边	编	贬	膘	变	壁	憋
7	柄	丙	彬	遍	标	彪	宾	兵		冰
8			秉	斌	濒	滨				
9			饼	斌	饼					

.....

54	0	1	2	3	4	5	6	7	8	9
0	知	帧	症	郑	证	芝	枝	支	歧	蜘
1	执	肢	脂	汁	之	织	职	直	植	殖
2	志	值	侄	址	指	止	趾	只	旨	纸
3	秩	挚	掷	至	致	置	治	峙	制	智
4	忠	稚	质	炙	痔	滞	重	室	中	盅
5	周	钟	衷	终	种	肿	肘	仲	众	舟
6	宙	州	洲	诒	粥	轴	朱	帚	咒	皱
7	逐	蛀	烛	铸	株	瞩	嘱	猪	著	柱
8										
9										

55	0	1	2	3	4	5	6	7	8	9
0	转	住	注	祝	驻	抓	爪	拽	专	砖
1	状	撰	赚	篆	桩	庄	装	妆	撞	壮
2	拙	椎	锥	追	赘	坠	缀	敦	准	捉
3	兹	卓	桌	琢	茁	坠	啄	着	灼	浊
4	滓	咨	资	琢	茁	姿	孜	紫	仔	籽
5	总	子	自	姿	滋	渍	字	踪	宗	综
6	祖	纵	邹	走	字	走	奏	踪	卒	族
7	尊	遵	阻	组	佐	左	揍	醉	最	罪
8							纂	作	坐	座
9										



56-87区收录了3008个二级汉字，按部首/笔画排序

56	0	1	2	3	4	5	6	7	8	9
0	𠂇	𠂈	𠂉	𠂊	𠂋	𠂌	𠂍	𠂎	𠂏	𠂐
1	𠂑	𠂒	𠂓	𠂔	𠂕	𠂖	𠂗	𠂘	𠂙	𠂚
2	𠂛	𠂜	𠂝	𠂞	𠂟	𠂠	𠂡	𠂢	𠂣	𠂤
3	𠂥	𠂦	𠂧	𠂨	𠂩	𠂪	𠂫	𠂬	𠂭	𠂮
4	𠂯	𠂰	𠂱	𠂲	𠂳	𠂴	𠂵	𠂶	𠂷	𠂸
5	𠂹	𠂺	𠂻	𠂼	𠂽	𠂾	𠂿	𠃀	𠃁	𠃂
6	𠃃	𠃄	𠃅	𠃆	𠃇	𠃈	𠃉	𠃊	𠃋	𠃌
7	𠃍	𠃎	𠃏	𠃐	𠃑	𠃒	𠃓	𠃔	𠃕	𠃖
8	𠃗	𠃘	𠃙	𠃚	𠃛	𠃜	𠃝	𠃞	𠃟	𠃠
9	𠃡	𠃢	𠃣	𠃤	𠃥	𠃦	𠃧	𠃨	𠃩	𠃪

57	0	1	2	3	4	5	6	7	8	9
0	𠃫	𠃬	𠃭	𠃮	𠃯	𠃰	𠃱	𠃲	𠃳	𠃴
1	𠃵	𠃶	𠃷	𠃸	𠃹	𠃺	𠃻	𠃼	𠃽	𠃾
2	𠃿	𠄀	𠄁	𠄂	𠄃	𠄄	𠄅	𠄆	𠄇	𠄈
3	𠄉	𠄊	𠄋	𠄌	𠄍	𠄎	𠄏	𠄐	𠄑	𠄒
4	𠄓	𠄔	𠄕	𠄖	𠄗	𠄘	𠄙	𠄚	𠄛	𠄜
5	𠄝	𠄞	𠄟	𠄠	𠄡	𠄢	𠄣	𠄤	𠄥	𠄦
6	𠄧	𠄨	𠄩	𠄪	𠄫	𠄬	𠄭	𠄮	𠄯	𠄰
7	𠄱	𠄲	𠄳	𠄴	𠄵	𠄶	𠄷	𠄸	𠄹	𠄺
8	𠄻	𠄼	𠄽	𠄾	𠄿	𠅀	𠅁	𠅂	𠅃	𠅄
9	𠅅	𠅆	𠅇	𠅈	𠅉	𠅊	𠅋	𠅌	𠅍	𠅎

.....

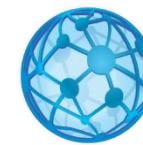
.....

86	0	1	2	3	4	5	6	7	8	9
0	𠅏	𠅐	𠅑	𠅒	𠅓	𠅔	𠅕	𠅖	𠅗	𠅘
1	𠅙	𠅚	𠅛	𠅜	𠅝	𠅞	𠅟	𠅠	𠅡	𠅢
2	𠅣	𠅤	𠅥	𠅦	𠅧	𠅨	𠅩	𠅪	𠅫	𠅬
3	𠅭	𠅮	𠅯	𠅰	𠅱	𠅲	𠅳	𠅴	𠅵	𠅶
4	𠅷	𠅸	𠅹	𠅺	𠅻	𠅼	𠅽	𠅾	𠅿	𠆀
5	𠆁	𠆂	𠆃	𠆄	𠆅	𠆆	𠆇	𠆈	𠆉	𠆊
6	𠆋	𠆌	𠆍	𠆎	𠆏	𠆐	𠆑	𠆒	𠆓	𠆔
7	𠆕	𠆖	𠆗	𠆘	𠆙	𠆚	𠆛	𠆜	𠆝	𠆞
8	𠆟	𠆠	𠆡	𠆢	𠆣	𠆤	𠆥	𠆦	𠆧	𠆨
9	𠆩	𠆪	𠆫	𠆬	𠆭	𠆮	𠆯	𠆰	𠆱	𠆲

8609

87	0	1	2	3	4	5	6	7	8	9
0	𠆳	𠆴	𠆵	𠆶	𠆷	𠆸	𠆹	𠆺	𠆻	𠆼
1	𠆽	𠆾	𠆿	𠇀	𠇁	𠇂	𠇃	𠇄	𠇅	𠇆
2	𠇇	𠇈	𠇉	𠇊	𠇋	𠇌	𠇍	𠇎	𠇏	𠇐
3	𠇑	𠇒	𠇓	𠇔	𠇕	𠇖	𠇗	𠇘	𠇙	𠇚
4	𠇛	𠇜	𠇝	𠇞	𠇟	𠇠	𠇡	𠇢	𠇣	𠇤
5	𠇥	𠇦	𠇧	𠇨	𠇩	𠇪	𠇫	𠇬	𠇭	𠇮
6	𠇯	𠇰	𠇱	𠇲	𠇳	𠇴	𠇵	𠇶	𠇷	𠇸
7	𠇹	𠇺	𠇻	𠇼	𠇽	𠇾	𠇿	𠈀	𠈁	𠈂
8	𠈃	𠈄	𠈅	𠈆	𠈇	𠈈	𠈉	𠈊	𠈋	𠈌
9	𠈍	𠈎	𠈏	𠈐	𠈑	𠈒	𠈓	𠈔	𠈕	𠈖

1-3-4 GB2312编码



有了每个字符的码位，下一步该如何进行存储呢？



ASCII码是直接将ASCII字符集转化位二进制进行存储

GB2312字符集的编码存储方式

举例：8609

$$\begin{array}{r} \text{8609} \\ + 0xA0 \\ \hline 0xF6 \end{array} \quad \begin{array}{r} 09 \\ + 0xA0 \\ \hline 0xA9 \end{array}$$

0xF6A9

Python 123

Python 3.9

海龟绘图

C

C++

新建

下载

运行 (Ctrl + Enter)

🔍

🔍

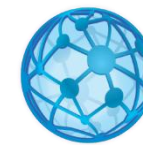
```
1 print("雯".encode('gb2312'))
```

```
2
```

输出

b'\xf6\xa9'

1-3-4 GB2312编码



01

GB2312

1980年

Complete Character List for
GB2312

高位和低位都大于127, 6763个汉字

02

GBK

1995年

不再规定低位大于127, 新增近20000个汉字和符号

03

GB18030

2000年第一版, 2005年第二版, 2022年第三版

收入汉字70000余个, 新增国内少数民族字符, 支持中日韩和繁体字以及emoji等字符, 最多可定义161万个字符

越南VSCII码

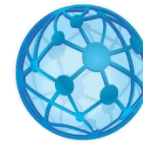
日文Shift JIS码

印度文字信息
交换码

繁体中文码Big5

西欧标准码
Latin-1

如何统一不同国家，不同地区采用的不同的编码方式？

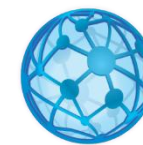


UNICODE (Universal Multiple-Octet Coded Character Set, 通用编码字符集, 简称 UCS)

- **ISO** (国际标准化组织) 重新制定了一个包括了地球上所有文化、所有字母和符号的编码。
- UNICODE 在制订时没有考虑与任何一种现有的编码方案保持兼容, 没有一种简单的方法, 可以把文本内容从UNICODE编码和另一种编码进行转换, 这种转换必须通过查表来进行。
- UNICODE 总共可以组合出65535不同的字符 (通用编码字符集 UCS-2), 大概已经可以覆盖世界上所有文化的符号。

Unicode查询

1-3-5 UNICODE编码



计算机与网络空间安全学院
School of Computer and Cyber Sciences

UCS-2

每个字符都由2字节（16位）Unicode代码表示，可表示65535个字符

(Unicode 0x044B) 以2字节序列存储：0x04 0x4B

UCS-4

采用 4 个字节编码，可以编码20多亿个字符

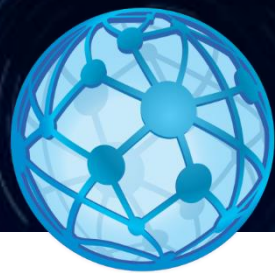
每个字符都需要多个固定字节来编码，造成资源浪费

UTF-8

UTF-8(Unicode Transformation Format 8-bit)是一种可变宽度的编码格式。用1~4个字节表示一个Unicode字符编码。它可以表示Unicode字符集中的所有字符。

本次课程结束， 请继续加油！

Computational Thinking



计算机与网络空间安全学院

School of Computer and Cyber Sciences

计算思维通识教育 Computational Thinking