

计算机中的信息表示

《计算机应用基础》
长安大学



主要内容



- 数制的概念
- 数制的表示
- 计算机内部的数制
- 数制间的相互转换
- 计算机中的字符编码
- 汉字在计算机中的表示
- 常用汉字编码标准

数制的概念



数制也称为计数制，是用一组固定的符号和统一的规则来表示数值的方法。

按进位的方法进行计数称为进位计数制。

这种以数字符号在各数位上的排列来表示不同数的方式，可以用有限的数字符号代表所有的数值。

一种计数制包含4个基本因素：

(1) 基数。

是指进制中可使用数码的个数。基数为 R ，即可称该数制为 R 进位制，简称 R 进制。例如，十进制的基数是10，二进制的基数是2。

(2) 数码。

一组用来表示数制的符号，符号的个数由该进位制的基数确定。

(3) 数位。

是指数码在一个数中所处的位置。数位不同所代表数的大小也不同。

(4) 位权。

位权是基数的幂，表示数码在不同位置上所代表的倍率值。对于 R 进制数，整数部分第 i 位的位权为 R^{i-1} ，而小数部分第 j 位的位权为 R^{-j} 。

数制的概念



一种计数制包含4个基本因素：

(1) 基数。

是指进制中可使用数码的个数。基数为 R ，即可称该数制为 R 进位制，简称 R 进制。

(2) 数码。

一组用来表示数制的符号，符号的个数由该进位制的基数确定。

(3) 数位。

是指数码在一个数中所处的位置。数位不同所代表数的大小也不同。

(4) 位权。

位权是基数的幂，表示数码在不同位置上所代表的倍率值。对于 R 进制数，整数部分第 i 位的位权为 R^{i-1} ，而小数部分第 j 位的位权为 R^{-j} 。

数制的表示



- 不同进制的基数和位权不同，对一个数的表示方法也不相同。
- 为了区别不同进制数，一般把具体数字用括号括起来，在括号的右下角标上相应表示数制的数字或进制符号

进位数制	二进制	八进制	十进制	十六进制
规则	逢二进一	逢八进一	逢十进一	逢十六进一
基数	$r=2$	$r=8$	$r=10$	$r=16$
数码	0, 1	0, 1, ..., 7	0, 1, ..., 9	0, 1, ..., 9, A, B, C, D, E, F
位权	2^i	8^i	10^i	16^i
进制符号	B	O	D	H

数制的表示



各种进位计数制中位权的值恰好是基数的幂。因此，任何一种 r 进位计数制表示的数 N 都可以写成按其位权展开的多项式之和。

$$N = \sum_{i=n-1}^{-m} D_i \times r^i$$

式中的 r 是基数， i 是数位， r^i 是位权， m 是小数位位数， n 是整数位位数， D_i 为该数位 i 上的数码符号。

按照数制的位权展开式，所有不同基数的数字都可以展开成最常用的十进制数。

$$(1010)_{10} = 1 \times 10^3 + 0 \times 10^2 + 1 \times 10^1 + 0 \times 10^0$$

$$(1010)_2 = 1 \times 2^3 + 0 \times 2^2 + 1 \times 2^1 + 0 \times 2^0 = (10)_{10}$$

$$(1010)_8 = 1 \times 8^3 + 0 \times 8^2 + 1 \times 8^1 + 0 \times 8^0 = (520)_{10}$$

$$(5AEF)_{16} = 5 \times 16^3 + 10 \times 16^2 + 14 \times 16^1 + 15 \times 16^0 = (23279)_{10}$$

计算机内部的数制



计算机采用二进制作为基础数制，一切数值和信息都以二进制代码形式进行存取、处理和传送，计算机的存储器件也是以二进制位作为最小存储单位的。

(1) 易于物理实现。

二进制只有0和1两个数字，可以表示0和1两种状态的电磁器件很多，计算机电路设计在物理上很容易实现，数字装置简单可靠。

(2) 运算规则简单。

二进制数的运算规则简单，两个二进制数和、积运算组合各有三种，有利于简化计算机内部结构，提高运算速度。

(3) 适合逻辑运算。

二进制的0和1正好和逻辑代数的假（false）和真（true）相对应，符合逻辑代数的理论基础，用二进制表示逻辑值可以方便地利用逻辑代数来综合分析逻辑电路。

数制间的相互转换



1、R进制数转换为十进制数。

基数为R的数按照权展开式将各位数字与位权相乘，所得结果相加就是十进制数。

$$(1101101.0101)_B = 1 \times 2^6 + 1 \times 2^5 + 0 \times 2^4 + 1 \times 2^3 + 1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0$$

$$+ 0 \times 2^{-1} + 1 \times 2^{-2} + 0 \times 2^{-3} + 1 \times 2^{-4}$$

$$= (109.3125)_D$$

$$(3506.2)_O = 3 \times 8^3 + 5 \times 8^2 + 0 \times 8^1 + 6 \times 8^0 + 2 \times 8^{-1}$$

$$= (1862.25)_D$$

$$(0.2A)_H = 2 \times 16^{-1} + 10 \times 16^{-2}$$

$$= (0.1640625)_D$$

数制间的相互转换

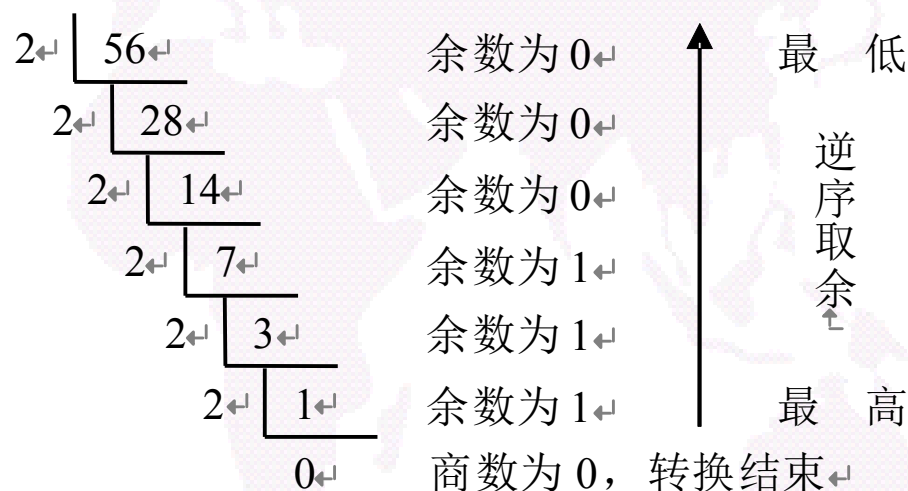


2、十进制数转换为R进制数

(1) 整数部分的转换——除R取余法

把一个十进制的整数不断除以所需要的基数R，直到商数为0。取每次得到的余数逆序排列，这种方法称为除R取余法。

$$(56)_{10} = (111000)_2$$



数制间的相互转换



(2) 小数部分的转换——乘R取整法

将十进制小数部分不断地乘以R，直到小数部分为0，或达到所要求的精度为止（小数部分可能永不0）。取每次得到的整数顺序排列，这种方法称为乘R取整法。

$$(0.6875)_{10} = (0.1011)_2$$

0.6875		
× 2		
1.375	整数为 1	最高位
× 2		
0.750	整数为 0	顺序取整
× 2		
1.500	整数为 1	
× 2		
1.000	整数为 1	最低位
	小数部分为 0，转换结束	

数制间的相互转换



3、非十进制数间的转换

将二进制数转换成八进制数。

010 100 101 . 010 111 010

2 4 5 . 2 7 2

将二进制转换成十六进制数。

0001 1111 1100 0111.1001 0101 1000

1 F C 7 . 9 5 8

二进制	八进制	二进制	十六进制	二进制	十六进制
000	0	0000	0	1000	8
001	1	0001	1	1001	9
010	2	0010	2	1010	A
011	3	0011	3	1011	B
100	4	0100	4	1100	C
101	5	0101	5	1101	D
110	6	0110	6	1110	E
111	7	0111	7	1111	F

计算机中的字符编码

- ASCII码（American Standard Code for Information Interchange，美国标准信息交换码）
- 是国际标准化组织确定的，国际上通用的字符编码标准。

	000	001	010	011	100	101	110	111
0000	NUL	DLE		0	@	P	`	p
0001	SOH	DC1	!	1	A	Q	a	q
0010	STX	DC2	"	2	B	R	b	r
0011	ETX	DC3	#	3	C	S	c	s
0100	DOT	DC4	\$	4	D	T	d	t
0101	ENG	NAK	%	5	E	U	e	u
0110	ACK	SYN	&	6	F	V	f	v
0111	BEL	ETB	'	7	G	W	g	w
1000	BS	CAN	(	8	H	X	h	x
1001	HT	EM	)	9	I	Y	i	y
1010	LF	SUB	*	:	J	Z	j	z
1011	VT	ESC	+	;	K	[	k	{
1100	FF	FS	,	<	L	\	l	
1101	CR	GS	-	=	M	]	m	}
1110	SO	RS	.	>	N	^	n	~
1111	SI	US	/	?	O	_	o	DEL

汉字在计算机中的表示



- 在计算机系统中使用汉字，需要为每一个汉字分配一个唯一的二进制编码。
- 汉字结构复杂，字型繁多，其编码是一个复杂的过程，需要在汉字处理的不同环节，如输入、编辑、存储、输出和交换等过程中使用不同的编码。
- 汉字编码主要包括用于信息交换的国标码、用于汉字输入的输入码、用于在计算机内部存储处理汉字的机内码和用于显示、打印汉字的字形码。

常用汉字编码标准



1、GB2312

GB2312编码是中华人民共和国国家汉字信息交换用汉字编码字符集，由国家标准总局于1980年发布，通行于中国大陆，新加坡等地也使用此编码，标准号是GB2312-1980。

- 在GB2312的字符集中共收集了7445个常用汉字和图形符号，放置于94区×94位的大表格里。
- 682个数字和符号包括标点符号、数学符号、数字序号、全角数字字母、日文平假名、片假名、希腊字母、俄文字母、图形符号、制表符、拼音和注音符号等，放置于01~09区。
- 6763个常用汉字分为两级，将其中使用频度高的常用汉字（3755个）作为第一级汉字，放置于16~55区，按照拼音字母和笔形顺序排列；较不常用的汉字（3008个）作为第二级汉字，放置于56~87区，按照部首和笔画顺序排列。

常用汉字编码标准



2、Big5

大五码（Big5），又称为五大码，是使用繁体中文字社群中最常用的计算机汉字字符集标准，共收录13060个中文字，包括繁体汉字、标点符号、希腊字母及特殊符号。Big5常用于中国台湾、中国香港和中国澳门等使用繁体中文的地区。

3、GBK

- GBK是《汉字内码扩展规范》，GBK编码标准兼容GB2312，是对GB2312的扩展，同时在字汇一级支持ISO/IEC10646-1和GB13000-1的全部中、日、韩（CJK）汉字，共收录汉字20902个、符号883个，并提供1894个造字码位，融简、繁体字于一库。
- GBK并非国家标准，只是由国家技术监督局标准化司、电子工业部科技与质量监督司公布为“技术规范指导性文件”，原始的国家标准GB13000《信息技术 通用多八位编码字符集（UCS）第一部分：体系结构与基本多文种平面》一直未被业界采用。



4、GB18030

- GB18030最新版本为中华人民共和国国家标准GB18030-2005《信息技术 中文编码字符集》，是中华人民共和国现时最新的标准字符集。
- GB18030与GB2312完全兼容，与GBK基本兼容，支持GB13000及Unicode的全部统一汉字，共收录汉字70244个。