



北京理工大学
BEIJING INSTITUTE OF TECHNOLOGY

第二讲

计算机信息数字化基础



□ 汉字信息的编码与交换

■ 字符信息编码与标准交换

- 西文字符的编码与标准交换

- 汉字信息的编码与交换

- 通用字符编码集UCS

汉字的特点

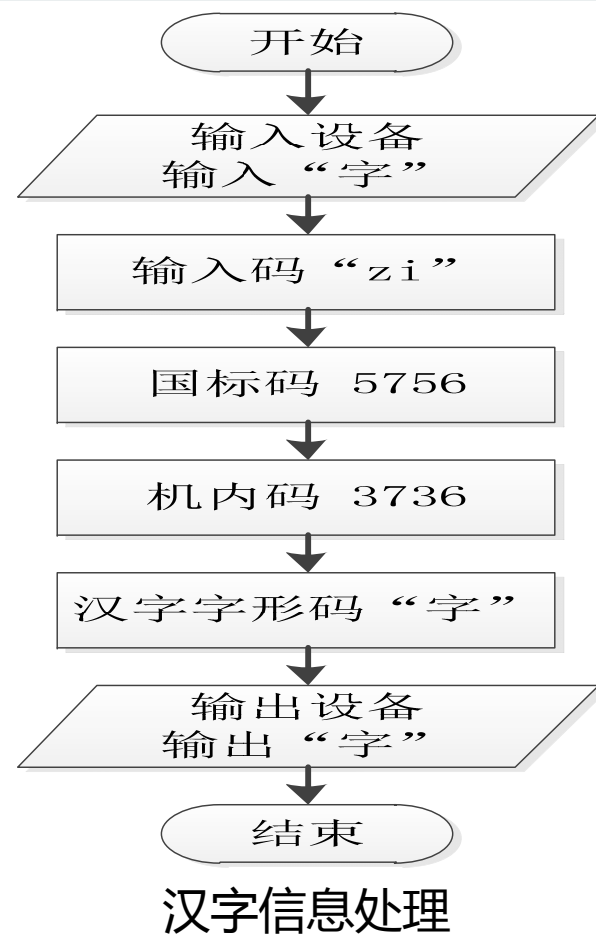
- 字数多。共6万左右，需要的编码多。
- 字形复杂。20画以上需要字模点阵多。
- 同音字多。需要输入方法灵活。

计算机对于汉字的处理实际上就是对各种汉字代码进行转换。

□ 汉字信息的编码与交换

汉字信息处理过程

涉及四种编码格式的转换过程



□ 汉字信息的编码与交换

汉字编码

- 汉字输入码
- 汉字国标码
- 汉字机内码
- 汉字字形码

输入码

- zhong 中 种 终 全拼
- zh 中 这 找 真 智能拼音

□ 汉字信息的编码与交换

汉字编码

- 汉字输入码
- 汉字国标码
- 汉字机内码
- 汉字字形码

国标码：也称**汉字信息交换码**，用区编码和位编码对汉字进行编码。

国标码分为以下几类：

- GB 2312-80
- GB 12345-90
- GBK编码
- GB 18030-2000
- BIG5 编码

□ 汉字信息的编码与交换

编码 21H 7EH		
位码 01D 94D		
区码 01D	数字、图形符号、希腊 、俄日等字符 (682个)	$15 \times 94 = 1410$
15D 16D	一级汉字 (3755个) , 按汉语拼音排列	$40 \times 94 = 3760$
55D 56D	二级汉字 (3008个) 按部首排列	$32 \times 94 = 3008$
87D 88D	空	$7 \times 94 = 658$
94D		
啊 - 区位码 : 1601D 国标码 : 3021H		

□ 汉字信息的编码与交换

汉字编码

- 汉字输入码
- 汉字国标码
- 汉字机内码
- 汉字字形码

机内码：为了在计算机内部对汉字进行存储、处理的汉字代码。由2字节组成。内码用于计算机内部处理、存储和传输汉字。由国标码演化而来。

□ 汉字信息的编码与交换

汉字编码

- 汉字输入码
- 汉字国标码
- 汉字机内码
- 汉字字形码

机内码编码

编码原则：码长短、有序且连续、与交换码对应中西文兼容

编码方式：多种，常用的是：将交换码的两个7位变为两个字节，两个字节的最高位为1

汉字“啊”

● **交换码** 3021H - - 00110000

00100001

● **内 码** B0A1H - - 10110000

10100001

□ 汉字信息的编码与交换

汉字编码

- 汉字输入码
- 汉字国标码
- 汉字机内码
- 汉字字形码

字形码：按汉字字形得到的编码，**又称字模**。有 16×16 点阵、 24×24 点阵等。**字模保存在计算机字库中**。当需要输出某个汉字时，找出该汉字的字模，输出设备按照字模进行输出。

汉字字模	精密型	96×96 (1152B)
		简易型 16×16 (32B)
	通用型	普通型 24×24 (72B)
		提高型 32×32 (128B)

□ 汉字信息的编码与交换

汉字编码

- 汉字输入码
- 汉字国标码
- 汉字机内码
- 汉字字形码

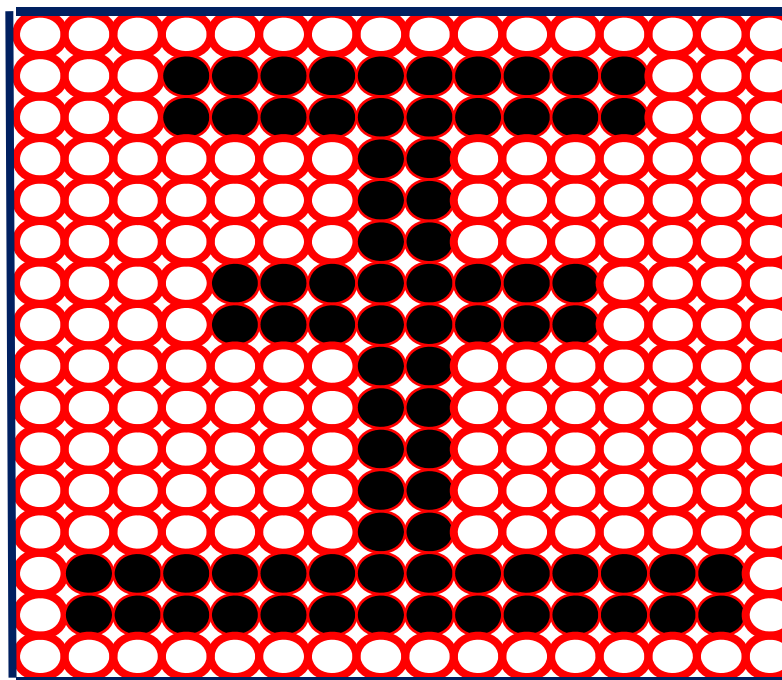
汉字点阵类型	点阵	占用字节数
简易型	16×16	32
普及型	24×24	72
提高型	32×32	128
精密型	96×96	1152

点阵数越高字型质量越好，但占用存储空间越大。
精密型汉字字形通常采用信息压缩存储技术。

□ 汉字信息的编码与交换

字形码示例

十六位编码



0	0	0	0
1	F	F	8
1	F	F	8
0	1	8	0
0	1	8	0
0	1	8	0
0	F	F	0
0	F	F	0
0	1	8	0
0	1	8	0
0	1	8	0
0	1	8	0
7	F	F	E
7	F	F	E
0	0	0	0

字模=2B×16行=32B

□ 汉字信息的编码与交换

汉字字形码与汉字库

汉字的字模保存在计算机里，称为**字库**。

三种字库

汉字库：字模的集合

软字库：将字模信息存放在外存上

硬字库：将字模信息固化在芯片上

当需要输出某个汉字时，将该汉字的字模找出，输出设备按照字模进行输出。不同的字体和效果有不同的字库，如宋体、黑体、楷体等。