



# 大学生计算与信息化素养

汉字在计算机中的表示  
北京林业大学信息学院

# 汉字在计算机中的表示

## 本次课程 所讲内容

汉字编码

输入码

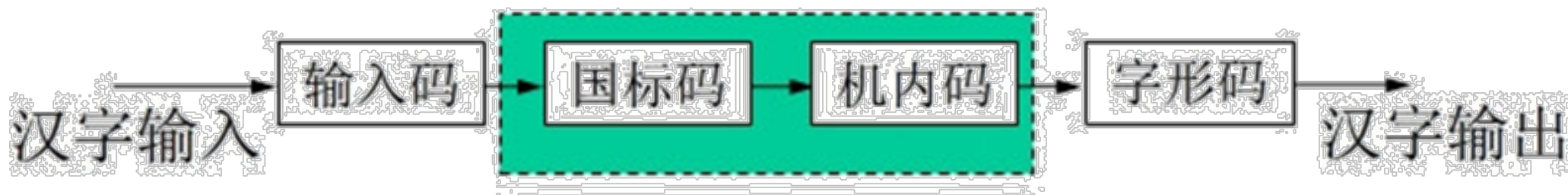
机内码、国标码、区位码

字形码

# 汉字编码



- 汉字字数多，字形复杂，计算机存储和处理比较复杂。计算机处理汉字信息的前提条件是对**每个汉字**进行编码，这些编码统称为**汉字编码**。





- **输入码：**是指在键盘上利用数字、符号或拼音字母输入汉字的代码。

## 音码类

全拼、双拼、微软拼音、智能ABC、搜狗拼音、紫光拼音、搜狗输入法等

## 形码类

五笔字型法、郑码输入法、万能五笔码、搜狗五笔码、陈桥五笔码

# 机内码、国标码、区位码



## 机内码

汉字在计算机中的存储形式（二进制形式）。

## 国标码

**国标码**又称为**汉字交换码**。汉字区位码先转换为国标码，再转换为机内码。

## 区位码

将汉字和汉字符号组成一个**94×94**的方阵，每一行称为一个“区”；每一列称为一个“位”，区、位的序号均从**01**到**94**。  
计算机中采用**2**个字节来储存区位码，第一个字节标记区码，第二个字节标记位码。

区码	位码
----	----

如汉字“啊”的区位码是1601，区码为16，位码为01，区码、位码都为十进制数

区 \ 位	01	02	.....	93	94
01					
02					
.....					
16	啊	阿			
17		雹		饼	
.....					
94					



# 机内码、国标码、区位码三者之间的关系



国标码 = 区位码 + 2020H

机内码 = 国标码 + 8080H

机内码 = 区位码 + A0A0H

# 机内码、国标码、区位码三者之间的关系



汉字机内码 = 汉字区位码 + A0A0H

“啊”的区位码为“1601D”，区码、位码分别转换成16进制为：  
“1001H”

$$1001H + A0A0H = B0A1H$$

“啊”机内码为“B0A1H”



# 字形码

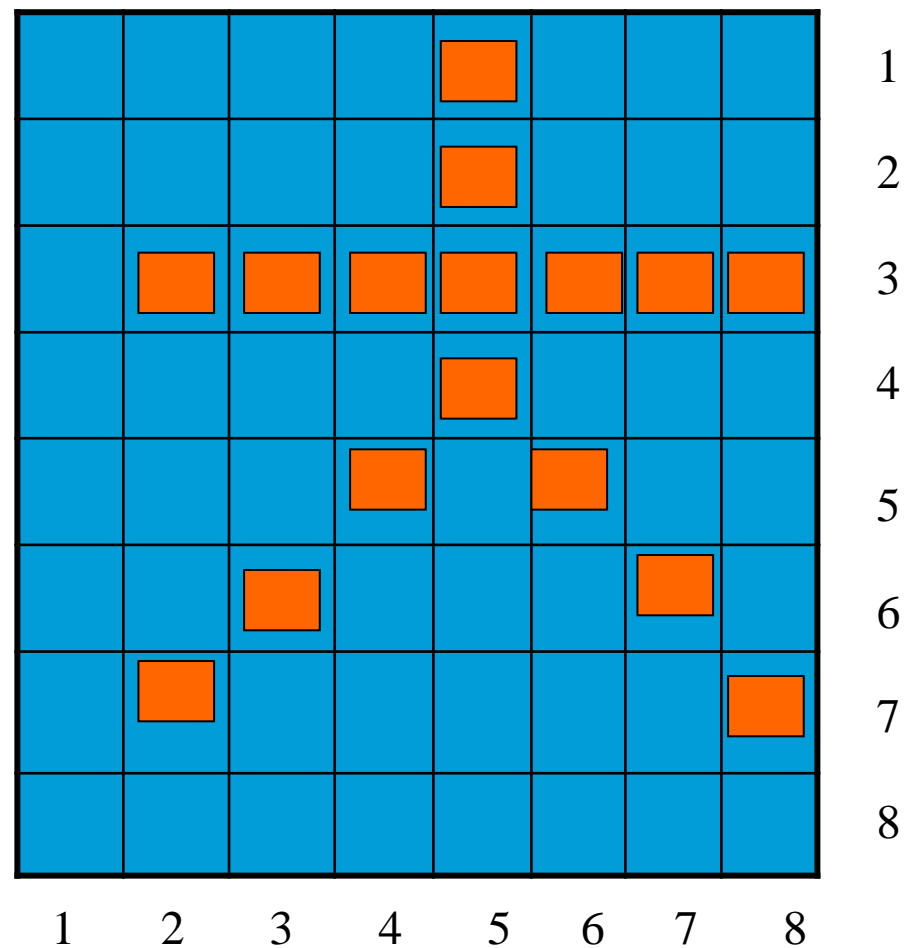


汉字字库中存储的汉字字形的数字化信息，用于汉字的显示和打印。  
字形码是汉字字形的字模数据，以点阵方式或矢量函数表示。

8×8的点阵

- 用点阵表示字形时，汉字字形码指的就是这个汉字字形点阵的代码。
- 字形码也称字模码，是由点阵表示的汉字字形代码，它是汉字的输出形式。

08	00001000
08	00001000
7F	01111111
08	00001000
14	00010100
22	00100010
41	01000001
00	00000000

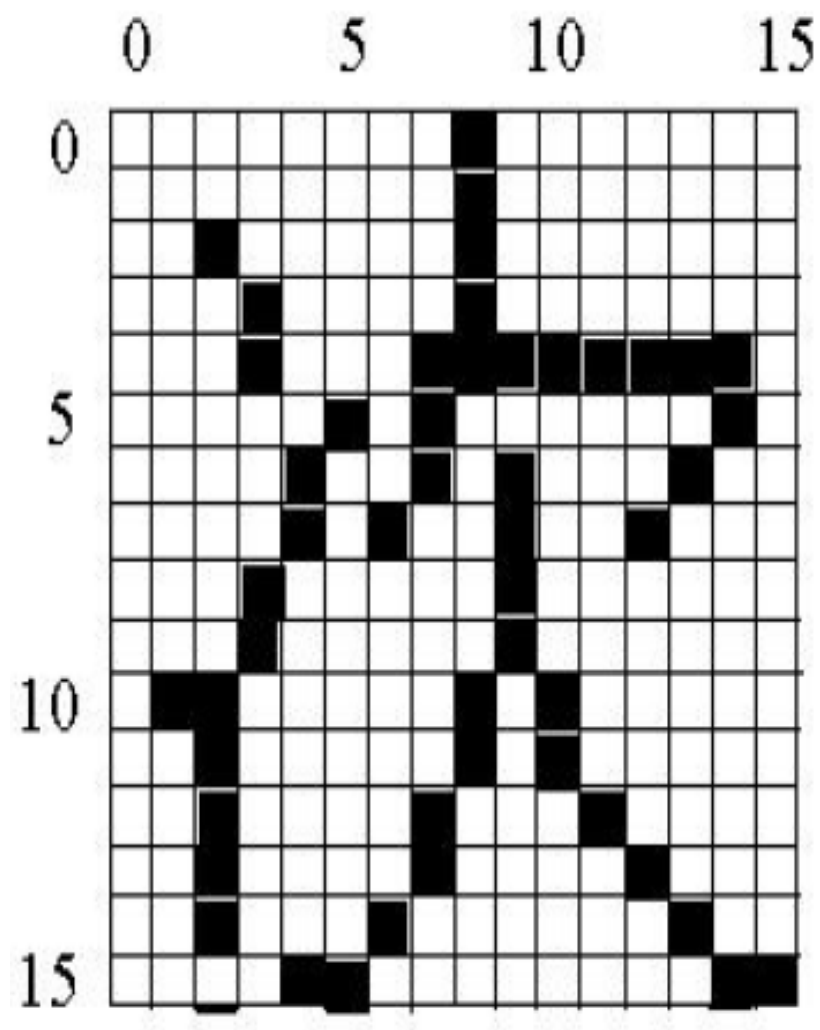




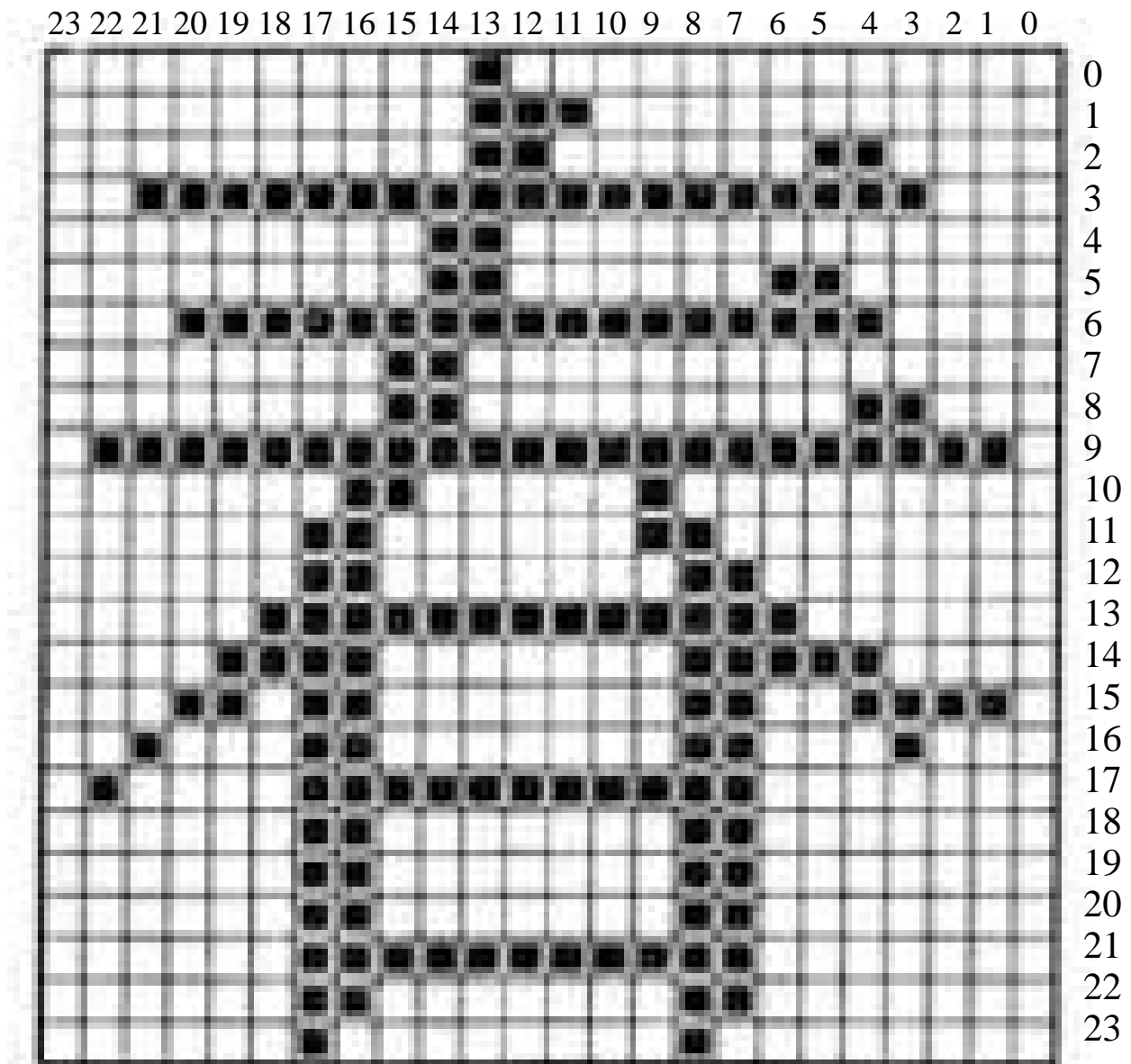
# 汉字点阵



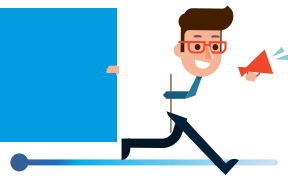
- 汉字点阵的信息量很大，所以占用的存储空间也很大：如 $8 \times 8$ 点阵每个汉字占8个字节。
- 字模点阵就构成了所谓的“字库”，用于显示和打印。
- 根据汉字形式的精密程度要求的不同，每个汉字点阵的大小也不同。
- 点阵形式有： $8 \times 8$ 、 $16 \times 16$ 、 $24 \times 24$ 、 $32 \times 32$ 、 $48 \times 48$ 、 $64 \times 64$ 、 $72 \times 72$ 、 $128 \times 128$  .....



0000,0000,1000,0000	00,80
0000,0000,1000,0000	00,80
0010,0000,1000,0000	20,80
0001,0000,1000,0000	10,80
⋮	
0010,0000,1010,0000	20,A0
0010,0001,0001,0000	21,10
0010,0010,0000,0100	22,04
0000,1100,0000,0011	0C,03



# 矢量方式



保存的是对每一个汉字的**描述**信息，比如一个笔划的起始、终止坐标，半径、弧度等等。在显示、打印这一类字库时，要经过一系列的**数学运算**才能输出结果，但是这一类字库保存的汉字理论上可以被无限地放大，笔划轮廓仍然能保持圆滑。

