



石家莊鐵道大學
SHIJIAZHUANG TIEDAO UNIVERSITY

网络精品课程

大学计算机应用基础

第1章 计算机基础知识

计算机信息编码

主讲：王书海

目录



网络精品课程

- 字符编码
- 汉字编码
- 数的编码



- 字符是计算机中使用最多的信息形式之一，也是人与计算机通信的重要媒介。
- 将字符变为指定的二进制符号称为**编码**。
- 在计算机内部，**要为每个字符指定一个确定的编码**，作为识别与使用这些字符的依据。
 - 一个编码就是一串二进制位“0”和“1”的组合，这样二进制数串的位数就决定了符号集的规模。
 - 例如，对一个由128个符号构成的符号集进行编码，就需要用7位二进制数；
 - 256个符号的字符集，需要8位二进制数，等等。



- ASCII码
 - 美国标准信息交换码 (American Standard Code for Information Interchange)
 - ASCII包括32个通用控制字符、10个十进制数码、52个英文大小写字母和34个专用符号，共128个元素，故需要用7位二进制数进行编码，以区分每个字符。
 - 通常使用一个字节(即8个二进制位)表示一个ASCII码字符，规定其最高位总是0



- BCD码

- BCD (Binary - Coded Decimal) 码又称“**二 — 十进制编码**”，专门解决用二进制数表示十进制数的问题。
- BCD码编码方法很多，最常用的是**8421** 编码，其方法是用**四位**二进制数表示一位十进制数，自左至右每一位对应的位权是8、4、2、1。
- 例：十进制数5803的8421编码：
 - 0101 1000 0000 0011



汉字编码

- 在我国推广应用计算机，必须使其具有汉字信息处理能力。
- 汉字信息的**输入**、**输出**及其**处理**远比西文困难得多，原因是汉字的编码和处理实在太复杂了。



国标码和汉字内码

- 汉字也是一种字符，但它远比西文字符量多且复杂，常用的汉字就有3000~5000个，显然无法用一个字节的编码来区分。
- 汉字通常用两个字节进行编码。
- 1981年我国公布的《通用汉字字符集(基本集)及其交换码标准》GB2312-80，共收集了7445个图形字符，其中汉字字符6763个，并分为两级，即常用的一级汉字3755个(按汉语拼音排序)和次常用汉字3008个(按偏旁部首排序)，其他图形符号682个。



国标码和汉字内码

- GB2312-80编码简称**国标码**，它规定每个图形字符由两个7位二进制编码表示，即每个编码需要占用两个字节，每个字节内占用7位信息，最高位补0。
 - 例如汉字“啊”的**国标码**为3021H，即00110000 00100001。
- **汉字内码**是汉字在计算机内部存储、处理和传输用的信息代码，要求它与ASCII码兼容但又不能相同，以便实现汉字和西文的并存兼容。通常将国标码两个字节的最高位置“1”作为汉字的内码。
 - 以汉字“啊”为例，其**内码**为B0A1H，即10110000 10100001。



汉字输入码

- 汉字输入码又称为**外码**，是指从键盘输入汉字时采用的编码。主要有：
 - **数字编码**：用一串数字代表一个汉字，最常用的是国标区位码。
 - **拼音码**：以汉语读音为基础的输入方法，如**搜狗输入法**。
 - **形码**：根据汉字形状确定的编码，常用的**五笔字型**码和表形码，
 - **音形码**：根据汉字的读音和字形进行编码，例如双拼码、五十字元等

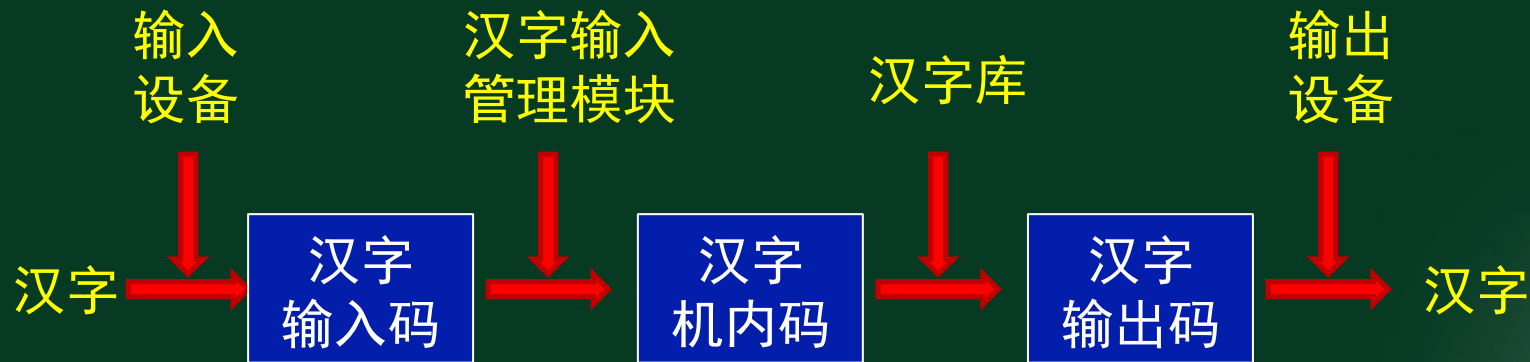


汉字字形码

- 每一个汉字的字形都必须预先存放在计算机内，例如GB2312国标汉字字符集的所有字符的形状描述信息集合在一起，称为字形信息库，简称**字库**。
- 字库通常分为**点阵字库**和**矢量字库**。
- 字库中存储了每个汉字的字形点阵代码。
- 不同的字体（如宋体、仿宋、楷体、黑体等）对应着不同的字库。
- 在输出汉字时，计算机要先到字库中去找到它的字形描述信息，然后再把字形送去输出。



各种编码之间的关系



汉字在计算机中的处理过程



数的编码

- 将数值型数据全面、完整地表示成一个机器数，应该考虑三个因素：机器数的**范围**、机器数的**符号**和机器数中**小数点的位置**。

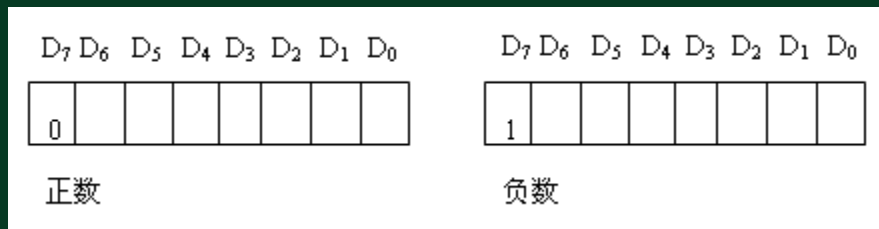


- 机器数的范围：
 - 机器数的范围由硬件 (CPU中的寄存器) 决定。
 - 当使用16位寄存器时，字长为16 位，所以一个无符号整数的最大值是 $(FFFF)_{16} = (65535)_{10}$ ，机器数的范围为 $0 \sim 65535$ 。



数的编码

- 机器数的符号：
 - 通常规定**最高位为符号位**，并用**0** 表示**正**，用**1** 表示**负**
 - 这种把符号数字化，并和数值位一起编码的方法，很好地解决了带符号数的表示方法及其计算问题。



- 这类编码方法，常用的有**原码**、**反码**、**补码**三种。



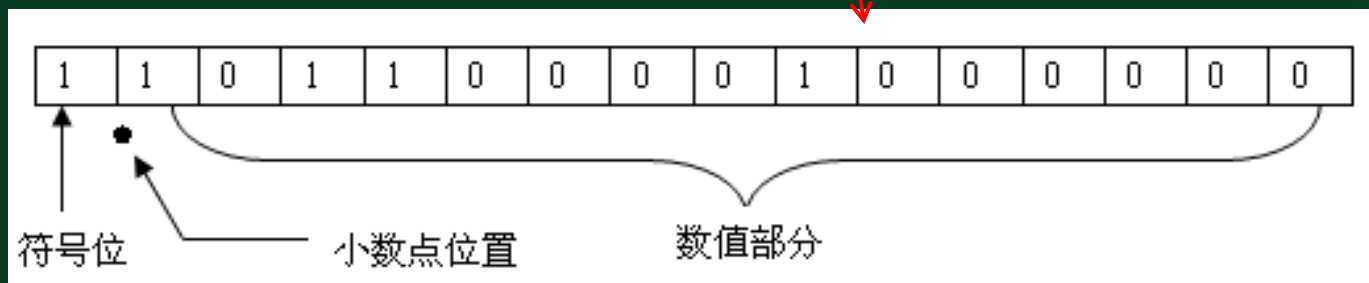
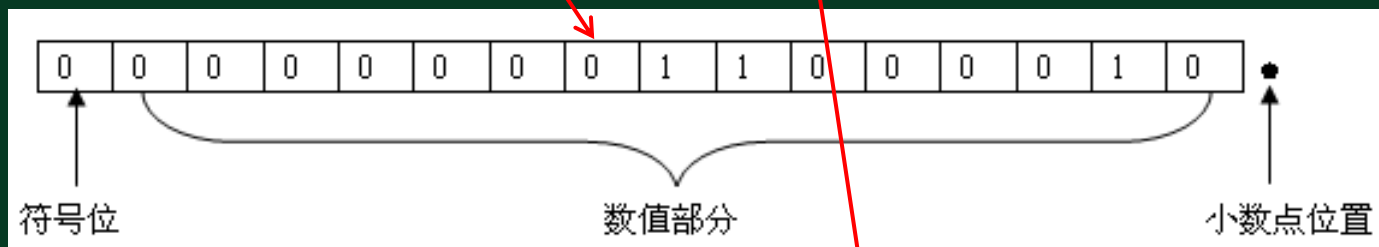
数的编码

- 在计算机内部难以表示小数点。故**小数点的位置是隐含的**。隐含的小数点位置可以是固定的，也可以是变动的，前者表示形式称为“**定点数**”，后者表示形式称为“**浮点数**”。



数的编码

- 定点数：
 - 在定点数中，小数点的位置一旦固定，就不再改变。
 - 定点数中又有**定点整数**和**定点小数**之分。



- 浮点数：

- 如果要处理的数既有整数部分，又有小数部分，则采用定点数便会遇到麻烦。为此引出浮点数，即**小数点位置不固定**。

- 一个二进制数 N 也可以表示为：

$$N = \pm S \times 2^{\pm P}$$

- N 、 P 、 S 均为二进制数。 S 称为 N 的**尾数**，即全部的有效数字(数值小于1)， S 前面的 \pm 号是尾数的符号； P 称为 N 的**阶码**(通常是整数)，即指明小数点的实际位置， P 前面的 \pm 号是阶码的符号。



数的编码

- 浮点数：



图 1-10 · 浮点数的存放形式

- 尾数的符号和阶码的符号各占一位，
- 阶码是定点整数，
- 阶码的位数决定了所表示的数的范围，
- 尾数是定点小数，
- 尾数的位数决定了数的精度。



习题讲解

- ASCII码可以表示_____种字符。
 - A. 255
 - B. 256
 - C. 127
 - D. 128



习题讲解

- BCD码是专门用二进制数表示_____的编码。
 - A. 字母符号
 - B. 数字字符
 - C. 十进制数
 - D. 十六进制数



习题讲解

- 最常用的BCD码是8421码，它用____位二进制数表示一位十进制数。
 - A. 1
 - B. 2
 - C. 4
 - D. 8



习题讲解

- 输入汉字时所采用的编码是_____。
 - A. 汉字国标码
 - B. 汉字机内码(内码)
 - C. 汉字输入码(外码)
 - D. 汉字字形码



- 汉字在计算机系统内部进行存储、加工处理和传输所采用的编码是_____。
 - A. 汉字国标码
 - B. 汉字机内码(内码)
 - C. 汉字输入码(外码)
 - D. 汉字字形码



- 汉字在屏幕上显示或在打印机上输出所采用的编码是_____。
 - A. 汉字国标码
 - B. 汉字机内码(内码)
 - C. 汉字输入码(外码)
 - D. 汉字字形码



- 数以某种表示方式存储在计算机中，称为“机器数”，_____称为“字长”。
 - A. 机器数的表示范围
 - B. 机器数的二进制位数
 - C. 机器数的最大值
 - D. 机器数的尾数的位数



习题讲解

- 定点整数的小数点约定在_____。
 - A. 符号位之后
 - B. 符号位之前
 - C. 最低位右边
 - D. 最低位前边



小结

- ASCII 字符编码
 - 7个二进制位，128个编码
- 汉字在计算机中的几种编码形式及其关系
 - 国标码、机内码、输入码、字型码
- 数的编码：定点数和浮点数

