

汉字编码表示与显示

一、汉字的编码

1981年5月，我国国家标准总局颁布了《信息交换用汉字编码字符集》(GB2312-80)，简称国家标准汉字编码，也叫国标码。国标码共收进标准字符7445个。其中一级汉字3755个，二级汉字3008个，共计6763个汉字。由于汉字的字符多，一个字节即8位二进制代码不足以表示所有的常用汉字。同时为了不与西文的ASCII码混淆，汉字国标码的每个汉字或符号都使用2个字节(16位二进制)代码来表示。西文字符采用一个字节表示，即ASCII码，一般只用七位来表示128个字符，而把最高位用作奇偶校验(或者不用)。

国标码介绍：在GB2312-80代码表中，纵向分为0~93,共94行；横向也是0~93,共94列。行与列分别用b7b6b5b4b3b2b1七位二进制码表示,第一字节表示行,第二字节表示列。其值从0100001到1111110(十六进制为21-7E)。这正是ASCII码的可打印字符的编码范围。国标码是将第一字节和第二字节连写而得。由于二进太长，一般用十六进制表示。

区位码介绍：在国标GB2312-80中，国标码除了用双七位二进制表示外，还可以表示成区位码的形式。即在国标代码表中，将行号称为区号，列号称为位号，分别有94个区和94个位。区号和位号用十进制表示，不足两位前面补0。这样每个汉字或符号都可用4位十进制表示。区位码因此可以用来作输入码。是汉字输入的基本编码方法之一。

机内码介绍：在计算机中双字节汉字与单字节西文字符混合使用、处理，汉字编码的各个字节若不予以特别标识，就会与单字节的ASCII码混淆不清；为此，将标识汉字的两个字节编码的最高位置为1，这种最高位为1的双字

节汉字编码就是中国大陆普遍采用的汉字机内码，简称内码，是计算机内部存储、处理汉字所使用的代码。

内码、国标码、区位码三者的关系是：

高字节内码=高字节国标码+80H=区码+20H+80H=区码+0A0H=区码+160

低字节内码=低字节国标码+80H=位码+20H+80H=位码+0A0H=位码+160

繁体汉字在一些地区和领域仍在使用，国家又制定出相应的繁体汉字字符集，国家标准代号是 GB12345-90“信息交换用汉字编码字符集——辅助集”，包含了 717 个图形符号和 6866 个繁体汉字。BIG5 是我国台湾地区计算机系统中使用的汉字编码字符集，包含了 420 个图形符号和 13070 个繁体汉字（不用简体字）。

二、汉字的字模库

汉字的输出主要是指汉字字形的输出。输出的方式主要是显示和打印两种。汉字输出时，用一个点阵来表示一个汉字。点阵的每个点位只有两种状态：有点或无点。若用二进制代码来表示即为该位取值为 1 表示有点，取值为 0 表示无点。汉字的输出原理与西文的输出原理是相同的。不同的是汉字笔划较多，要能很好地表示一个汉字，起码需要 16×16 点阵才行。如果要求字型逼真美观，点阵的点数还要增加。如用 24×24 、 32×32 、 48×48 等，因此汉字的存储空间比西文要大很多，需要用大量的存储空间来存放字模。

描述一个汉字点阵信息的二进制代码串称为汉字的“字模”。所有汉字和各种符号的点阵信息就组成汉字的“字模库”（简称字库）。字模的表示顺序为：先从左到右，再从上到下。也就是先画第一行左上方的 8 个点，再是右上

方的 8 个点，然后是第二行左边 8 个点，右边 8 个点，以此类推。

三、汉字的显示原理

1. 从键盘输入的汉字经过键盘管理模块，变换成机内码。
2. 然后经字模检索程序，查到机内码对应的点阵信息在字模库的地址。
3. 从字库中检索出该汉字点阵信息。
4. 利用显示驱动程序将这些信息送到显示卡的显示缓冲存储器中。
5. 显示器的控制器把点阵信息整屏顺次读出，并使每一个二进制位与屏幕的一个点位相对应，就可以将汉字字形在屏幕上显示出来。

四、用区位码获取汉字的点阵信息

以 16×16 的点阵汉字库文件为例。一个汉字用了 256 个点共 32 个字节表示。汉字共分 94 区，每个区有 94 位汉字。机内码用两个字节表示，第一个字节存储区号 (qh)，为了和 ASCII 码相区别，范围从十六进制的 A1H 开始（小于 80H 地为 ASCII 码字符），对应区码的第一区；第二个字节是位号 (wh)，范围也从 A1H 开始，对应某区中的第一个位码。这样，将汉字机内码减去 A0A0H 就得到该汉字的区位码。

从而可以得到汉字在字库中的具体位置：

$$\text{location} = (94 * (\text{qh} - 1) + \text{wh} - 1) * \text{一个汉字字模占用的字节数}$$

对于 16×16 的点阵汉字库，汉字在字库中的具体位置的计算公式就是： $(94 * (\text{qh} - 1) + \text{wh} - 1) * 32$ 。例如，“房”的机内码为十六进制的 B7BF，则其区位码

是 B7BFH-A0A0H=171FH，转化为十进制就是 2331，在汉字库中的位置就是 $32*[94*(23-1)+(31-1)]=67136$ 字节以后的 32 个字节为“房”的显示点阵。

本 文 来 自 CSDN 博 客 ， 转 载 请 标 明 出 处 ：

<http://blog.csdn.net/buxoman/archive/2005/02/23/299110.aspx>