

§. 汉字编码与汉字点阵字库

【汉字编码的基本概念:】

- 汉字采用多字节编码,常用的有 GB2312 和 GB18030 两种,其中 GB2312 是 2 字节编码,GB18030 是 2-4 字节编码,下面所说的汉字编码, **均特指** GB2312
- GB2312 也称为汉字的区位码,它将汉字及全角字符、全角英文、希腊字母、日文等各种字符分为 94 个区,每个区分为 94 个位,用区位对汉字进行编码,区位码一般用 10 进制表示
例:“啊”的区位码是 1601,表示该汉字是第 16 区的第 01 位
- 汉字编码也称为汉字的机内编码,即汉字在内存中存储,机内码一般用 16 进制表示,同一个汉字的机内码与区位码不同,但是可以进行换算
提示:可用 UltraEdit 等文本编辑软件的 16 进制方式查看汉字的机内码
- GB2312 的详细信息可参考: <http://www.qqxiuzi.cn/zh/hanzi-gb2312-bianma.php>
- 汉字编码是机内表示,显示时需要从汉字库中取出该汉字对应的图形信息,才能在屏幕上正确显示,汉字库常用的有点阵字库及矢量字库两种,下面所说的汉字库, **均特指** 点阵字库
- 点阵字库有 16*16/24*24 等多种不同的尺寸,下面 **均特指** 16*16 的点阵字库
- 汉字的点阵字库,以 16*16 点阵为例,一个汉字的显示为 16 行 x16 列,行列的每个位置用 1bit 表示 (0: 不显示/1: 显示); 按先行后列的顺序依次存储该汉字的点阵信息,因此一个汉字的点阵大小为 $16*16 = 256\text{bit} = 32\text{Byte}$,即一个汉字对应的点阵信息为 32 字节;所有汉字的点阵信息依次排列存储,就形成了汉字库

【提供的附件说明:】

- demo.exe: 用于理解汉字编码、汉字点阵、点阵信息的显示等概念及之间的关系
- HZK16 : GB2312 对应的汉字的简体写法的 16x16 点阵信息
- HZK16F: GB2312 对应的汉字的繁体写法的 16x16 点阵信息
- 提示: cmd 窗口屏幕的纵横比例不是 1:1 (即 16 行 x16 列画出来并不是正方形,变形较厉害),需要解决,使显示效果每个字尽量保持方形