

区位码，汉字交换码和汉字机内码的概念是什么？

它们之间有什么关系？

区位码：

1981 年国家颁布了 GB2312 汉字标准共有 6763 个，其中一级 3755，二级 3008，还有 682 非汉字字符。并为每个字符规定了标准编码，便于在计算机内部相互转换。

作为 GB2312 标准只是定义了一张 94×94 的二维表。其中行为区号，列为位号。这样可以利用区号和位号来找到其中的汉字。这种编码就是我们所说的区位码。

比如

陈(1934) 区号：19 位号：34，为了处理与存储的方便，每个汉字的区号和位号在计算机内部分别用一个字节来表示

区位码无法于汉字进行通信，因为 ASCII 码中规定了 00H-1F 作为控制码使用，这样就发生了冲突。由于计算机不是中国人发明的所以只能听从于国际标准 ISO2022 规定区号和位号都加上 32 (20H)，这样就防止冲突。经过加上 32 以后的编码称为国际交换码

陈-区号：19+32= 51

00010011+00100000=00110011

位号：34+32=66

00100010+00100000=01000010

即 5166 16 进制 3342

由于文本中通常混合使用汉字和西文字符，汉字信息如果不予以特别标识，就会与单字节的 ASCII 码混淆。此问题的解决方法之一是将一个汉字看成是两个扩展 ASCII 码，使表示 GB2312 汉字的两个字节的最高位都为 1。

这种高位为 1 的双字节汉字编码即为 GB2312 汉字的机内码，简称为“内码”。

00110011 最高位变为 1 则从 33 变为 B3

01000010 最高位变为 1 则从 42 变为 C2

这样一来，陈的机内码应该为 B3C2

这里要说明的是不管你是采用什么样的输入法输入汉字，其汉字的机内码都是相同的。

如果要从一个汉字的机内码转换为区位码，其实就是相反的方向进行运算。