GB2312 编码规则

GB2312 标准共收录 6763 个汉字，其中一级汉字 3755 个，二级汉字 3008 个；同时，GB2312 收录了包括拉丁字母、希腊字母、日文平假名及片假名字母、俄罗斯语西里尔字母在内的 682 个全形字符。

GB2312 的出现，基本满足了汉字的计算机处理需要，它所收录的汉字已经覆盖 99.75%的使用频率。GB2312 中对所收汉字进行了“分区”处理，每区含有 94 个汉字/符号。这种表示方式也称为区位码。

01-09 区为特殊符号。

16-55 区为一级汉字，按拼音排序。

56-87 区为二级汉字，按部首/笔画排序。

10-15 区及 88-94 区则未有编码。

举例来说，“啊”字是 GB2312 之中的第一个汉字，它的区位码就是 1601。字节结构在使用 GB2312 的程序中，通常采用 EUC 储存方法，以便兼容于 ASCII。每个汉字及符号以两个字节来表示。第一个字节称为“高位字节”，第二个字节称为“低位字节”。“高位字节”使用了 0xA1-0xF7(把 01-87 区的区号加上 0xA0)，“低位字节”使用了 0xA1-0xFE(把 01-94 加上 0xA0)。例如“啊”字在大多数程序中，会以 0xB0A1 储存。（与区位码对比：0xB0=0xA0+16,0xA1=0xA0+1）。

所以 GB2312 编码中汉字区码的十进制是从 176 到 247，位码是从 161 到 255.之所以存储了 6763 小于 72*94=6768，是因为在区码为 215，位码为 250-254 之间共五个编码没有汉字编码，所以 6768-5=6763 个。