

JAVA基础：GB2312汉字编码字符集 PDF转换可能丢失图片或格式，建议阅读原文

https://www.100test.com/kao_ti2020/145/2021_2022_JAVA_E5_9F_BA_E7_A1_80_c104_145425.htm GB 2312 码是中华人民共和国

国家标准汉字信息交换用编码，全称《信息交换用汉字编码字符集基本集》，标准号为 GB 2312 - 80（GB 是“国标”二字的汉语拼音缩写），由中华人民共和国国家标准总局发布，1981年5月1日实施。习惯上称国标码、GB 码，或区位码。它是一个简化字汉字的编码，通行于中国大陆地区。新加坡等地也使用这一编码。GB 2312-80 收录简化汉字及一般符号、序号、数字、拉丁字母、日文假名、希腊字母、俄文字母、汉语拼音符号、汉语注音字母，共 7445 个图形字符。其中汉字以外的图形字符 682 个，汉字 6763 个。GB 2312-80 规定，“对任意一个图形字符都采用两个字节（Byte）表示。每个字节均采用 GB 1988 - 80 及 GB 2311 - 80 中的七位编码表示。两个字节中前面的字节为第一字节，后面的字节为第二字节。”习惯上称第一字节为“高字节”，第二字节为“低字节”。GB 2312-80 将代码表分为 94 个区（Section），对应第一字节；每个区 94 个位（Position），对应第二字节。两个字节的值，分别为区号值和位号值各加 32（20H）。GB 2312-80 规定，01 ~ 09 区（原规定为 1 ~ 9 区，为表示区位码方便起见，今改称 01 ~ 09 区）为符号、数字区，16 ~ 87 区为汉字区。而 10 ~ 15 区、88 ~ 94 区是有待于“进一步标准化”的“空白位置”区域。但第 10 区推荐与第 3 区的 94 个图形字符（即 GB 1988 - 80 中的 94 个图形字符）相同，字形宽度为其宽度的一半。GB 2312-80 把收录的汉字分成两级。第一

级汉字是常用汉字，计 3755 个，置于 16 ~ 55 区，按汉语拼音字母 / 笔形顺序排列；第二级汉字是次常用汉字，计 3008 个，置于 56 ~ 87 区，按部首 / 笔画顺序排列。字音以普通话审音委员会发表的《普通话异读词三次审音总表初稿》

(1963 年出版) 为准，字形以中华人民共和国文化部、中国文字改革委员会公布的《印刷通用汉字字形表》(1964 年出版) 为准。例：汉字“啊”，第一字节为 0110000，第二字节为 0100001，即 16 区、01 位，用 16 - 01 表示。注：1. 有些外挂式的中文平台，不能正确显示 01 ~ 15 区的空白位置。2. 02 区的 0201 ~ 0210、06 区的 0664 ~ 0685 和 08 区的 0827 ~ 0832 位置，在 GB 2312 - 80 中原本规定为有待于进一步标准化的空白位置。但如果你使用的操作系统为 Windows 95 简体中文版（及更高的版本），或者使用了其他支持 GBK 码、CJK 码的汉字系统，在某些情况下会看到以上位置分别为 10 个小写罗马数字、19 个中文竖排用符号和 6 个汉语拼音符号。这些符号系 GB 5007.1、GB/T 12345-90 及 GBK 增补的符号，由于 Windows 95 简体中文版使用了 GBK 字体，所以显示出这些符号。此外第 10、11、12 区显示的汉字，也属类似的情况。3. GB 码的编码范围为 2121H ~ 777EH，与 ASCII 码有重叠，通行将 GB 码两个字节的最高位均置 1 (MSB=1)，以示区别。因此，实际应用的 GB 码是 GB 码高位置 1 后的变形码，编码范围为 0A1A1H ~ 0F7FEH。如汉字“啊”，第一字节为 10110000，第二字节为 10100001。100Test 下载频道开通，各类考试题目直接下载。详细请访问 www.100test.com