C程序实现汉字内码与GB码 PDF转换可能丢失图片或格式，建议阅读原文 https://www.100test.com/kao_ti2020/261/2021_2022_C_E7_A8_8B_E5_BA_8F_E5_AE_9E_c97_261049.htm // HZEncode.cpp : Defines the entry point for the console application. // /* 参考文献： 汉字的编码和表示 1)汉字交换码(国标码) 汉字交换码(国标码)主要用于汉字信息交换。 国标码：以国家标准局1980年颁布的《信息交换用汉字编码字符集"基本集》(代号为GB2312 80)规定的汉字交换码作为国家标准汉字编码。 GB2312 80中共有7445个字符符号： 汉字符号6763个 一级汉字3755个(按汉语拼音字母顺序排列) 二级汉字3008个(按部首笔划顺序排列) 非汉字符号682个 GB2312 80规定，所有的国标码汉字及符号组成一个94 94的方阵。在此方阵中，每一行称为一个"区"，每一列称为一个"位"。这个方阵实际上组成一个有94个区(编号由01到94)，每个区有94个位(编号由01到94)的汉字字符集。 一个汉字所在的区号和位号的组合就构成了该汉字的"区位码"。其中，高两位为区号，低两位为位号。这样区位码可以唯一地确定某一汉字或字符.反之，任何一个汉字或符号都对应一个唯一的区位码，没有重码。 区位码分布情况如下： 区号 内 容 1区 键盘上没有的各种符号 2区 各种序号 3区 键盘上的各种符号(按中文方式给出) 4 -5区 日文字母 6区 希腊字母 7区 俄文字母 8区 标识拼音声调的母音及拼音字母名称 9区 制表符号 10- 15区 未用 16-55区 一级汉字(按拼音字母顺序排列) 56- 87区 二级汉字(按部首笔划顺序排列) 88- 94区 自定义汉字 由上可以看出，所有汉字与符号的94个区，可以分为四个组： ①1 -15区：为图形符号区。其中1 9区为标准符号区.10 15区

为自定义符号区。 ②16 -55区：为一级汉字区，包含3755个汉字。这些区中的汉字按汉语拼音顺序排序，同音字按笔画顺序列出。 ③56 -87区：为二级汉字区，包含3008个汉字。这些区中的汉字是按部首笔划顺序排序的。 ④88 -94区：为自定义汉字区。 国标码规定，每个汉字(包括非汉字的一些符号)由2字节代码表示。每个字节的最高位为0，只使用低7位，而低7位的编码中又有34个适用于控制用的，这样每个字节只有27 - 34 = 94个编码用于汉字。2个字节就有94 94=8836个汉字编码。在表示一个汉字的2个字节中，高字节对应编码表中的行号，称为区号.低字节对应编码表中的列号，称为位号。 汉字国标码的范围用二进制表示是： 00100001 00100001 01111110 01111110 (1 32)10 (1 32)10 (94 32)10 (94 32)10 7 位ASCII码是128个字符组成的字符集。其中编码值0 31(00000000 00011111)不对应任何印刷字符，通常称为控制符，用于计算机通信中的通信控制或对计算机设备的功能控制。编码值32(00100000)是空格字符SP。编码值127(1111111)是删除字符DEL。 汉字国标码的起始二进制位置选择00100001即(33)10是为了跳过ASCII码的32个控制字符和空格字符。所以，汉字国标码的高位和低位分别比对应的区位码大(32)10或(00100000)2或(20)H，即：国标码高位 = 区码 20H (H表示十六进制) 国标码低位 = 位码 20H 2) 汉字机内码(内码)(汉字存储码) 汉字机内码(内码)(汉字存储码)的作用是统一了各种不同的汉字输入码在计算机内部的表示。 为了将汉字的各种输入码在计算机内部统一起来，就有了专用于计算机内部存储汉字使用的汉字机内码，用以将输入时使用的多种汉字输入码统一转换成汉字机内码进行存储，以方便机内的汉字处理 汉字机内码是在计算机内

部存储、处理的代码。计算机既要处理汉字，又要处理英文。因此计算机必须能区别汉字字符和英文字符。英文字符的的机内码是最高为为0的8位ASCII码。为了不与7位ASCII码发生冲突，把国标码每个字节的最高位由0改为1，其余位不变的编码作为汉字字符的机内码。 汉字机内码的范围用二进制表示是： 10100001 10100001 11111110 11111110 机内码的高位和低位比对应的国标码的高位和低位大(128)10或(10000000)2或(80)H 即： 机内码高位 = 国标码高位 80H 机内码低位 = 国标码低位 80H 又因为： 国标码高位 = 区码 20H 国标码低位 = 位码 20H 所以： 机内码高位 = 区码 A0H 机内码低位 = 位码 A0H 也就是说，机内码高位和机内码低位分别比对应的区码和位码大(160)10或(10100000)2或 (A0)H 例如：汉字"啊"的区位码为"1601"，其中区码为(16)10或(10)H，位码为(01)10或(01)H。 则： 机内码高位 = 10H A0H = B0H 机内码低位 = 01H A0H = A1H 所以： 机内码= B0A1H以下是引用片段： 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