

一级MS--Office考试要点2 PDF转换可能丢失图片或格式，建议阅读原文

https://www.100test.com/kao_ti2020/135/2021_2022__E4_B8_80_E7_BA_A7MS--_c98_135039.htm 【考点二】数据单位、字符编码、汉字编码

1.计算机中的数据单位计算机中用到的信息单位主要有位、字节、字等。位(Bit)是计算机存储设备中的最小的信息容量单位，用0或1二进制数位来表示。如二进制数10011101是由8个位组成的，位常用b表示。字节(Byte)是计算机的最小存储单位元，常用B表示。微型机中由8个二进制位组成一个字节。如8个二进制数“10011101”构成一个字节。一个字节可存放一个半角英文字符的编码(如ASCII码)。两个字节可存放一个汉字编码。一个字节表示的无符号整数，可以从最小的00000000至最大的11111111，共28个。习惯上，210(1024)个字节称为1K字节，记为1KB。随着存储容量的增大，还有下列计量单位，它们之间的关系如下

：8b=1B210B=1024B=1KB220B=1024×1024B=1MB230B=1024×1024B×1024B=1GB字(Word)是计算机信息交换、加工、存储的基本单元。通常将组成一个字的位数叫该字的字长，用来表示数据或信息的长度。如一台计算机的字长为32位，则表示该机的一个字由4个字节组成。不同级别的计算机的字长是不同的。为什么不取整数1000，而取一个如此难记的数字1024来表示1KB字节呢?细心的读者一定不难发现

：210B=1024。正是由于计算机中采用的是二进制数，用1000来表示1千字节反而不方便了。表示行字节的KB、兆字节的MB以及千兆字节的GB可以简写成K、M、G，即：1K=1KB，1M=1MB，1G=1GB。本书在以后的章节中将采用KB、MB

或GB表示。

2. 字符编码

英文字符编码在计算机中不仅是数字，所有的数据都是用二进制数来表示的。长期以来，存在各种字符编码，难于统一，为此美国国家标准局提出了一套编码方案，它叫做ASCII码(American Standard Code for Information Interchange,美国信息交换标准代码)。它收录了128个基本字符，其中包括了数字0~9，英文大小写字母，一些运算符号如、-、*、/和一些常用符号如\$、%、#等。每一个字符用一个八位二进制数来表示，如二进制的01000001表示英文大写字母A；二进制的00110001表示数字字符1等等。为了便于记忆，常将这些字符编码以十进制形式表示。请注意在ASCII编码中所列的前32个编码所表示的字符都是计算机信息传递、加工过程中使用的一些控制字符，在屏幕上是看不出来的，打印机上也打印不出来。

汉字编码

汉字是方块的，而且结构千变万化，要将其输入计算机且表示出来，确实是一个难题。经过我国科研工作者几代的努力，这个问题已被解决。人们习惯采用一种点阵方案来表示汉字，1981年，我国制定了“中华人民共和国国家标准信息交换汉字编码”，代号为GB2312-80，这种编码称为国标码，是所有汉字编码都必须遵循的一个共同标准。GB2312-80以94个可以显示ASCII码作为基本集，共收录了汉字和图形符7445个，每个汉字用两个字节表示。汉字分为两级，一级汉字3755个，按汉字拼音字母排列；二级汉字3008个，按部首排列；非汉字字符682个。GB2312-80规定，所有的国标汉字与符号组成一个94×94的矩阵。在此矩阵中，每行称为一个“区”，每列称为一个“位”，因此，这个矩阵实际上组成了一个有94个区(区号从01至94)，每个区内有94个位(位号从01至94)的汉字

字符集。一个汉字所在的区号与位号简单地组合在一起就构成了该汉字一种外码“区位码”，它用高低两个字节来表示，高字节表示汉字所在的区号，低字节表示汉字所在的位号。汉字的区位码是唯一的。国标码与区位码之间存在如下换算关系：国标码高字节=区码 20H 国标码低字节=位码

20H GB2312-80编码的安排情况如下：(1)1~9区非汉字字符682个；(2)10~15区，空位564个；(3)16~55区，一级汉字，也称为常用字，按汉字的拼音排列；(4)56~87区，二级汉字，也称为次常用字，按汉字的部首排列；(5)88~94区，空位。近年来，为便于和夹芸国际间信息交流，国家制字新的汉字编码标准GB-13000，国际上称为ISO/IEC10646，这种汉字编码用3个字节表示一个汉字，汉字编码容量大大增加，最大的特点是包括了中、日、韩等许多国家的文字。汉字编码分为内码和外码。内码是计算机系统存储、处理汉字信息时所用的代码。汉字的输入码要转换成内码才能在计算机内存储和处理，一个内码占两个字节。汉字国标码的高低字节的取值范围在33~126之间，每个字节最高位都是0，正好和ASCII码相冲突，故不能作为机内码使用。国标码经过变换之后才能作为机内码使用，机内码与国标码之间的变换关系如下：内码高字节=国标码高字节+80H 内码低字节=国标码低字节+80H 外码是指输入码及打印码、显示码等，用于人与计算机进行交互(汉字输入/输出)时所用的代码。就输入码来说，国内外有几百种编码方案，常用的有区位码、拼音码、五笔字形码、自然码等。汉字是一种像形文字，每一个汉字可以看成是一个特定的图形，这种图形一般用点阵信息来描述。所有汉字的点阵信息按国标码规定的先后顺序组合

在一起，就形成了汉字的字库。 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