

计算机操作指导丛书

中文之星使用方法与技巧



卢 山 编 著
寇国华 审

人民邮电出版社
PEOPLE'S POSTS &
TELECOMMUNICATIONS
PUBLISHING HOUSE

计算机操作指导丛书编委会

主 任：牛田佳

副主任：李树岭 寇国华

委 员：孙中臣 程 颂 戴 浩

段来盛 黄宪东 罗 为

舒智勇 唐 宏 杨宝珍

丛 书 前 言

随着近年来计算机应用的普及，越来越多的读者加入到学习计算机的行列中来，其中的很多读者，过去较少接触这一领域；还有很多读者虽然已经成为计算机的用户，但是对一些新的应用软件了解不多……许多读者在学习的过程中，常常感受到，按照一般的教材学习计算机技术费力费时，不能很好地与实际操作相结合，希望能有一套书使学习计算机技术更加直接、快捷，书籍的内容能够更加通俗易懂，一目了然。为了满足读者的这些需求，我们组织编写了这套“计算机操作指导丛书”。

这套丛书的突出特点是以指导用户学会操作使用计算机及相关软件为目的，每一本书的编排紧密结合这一主题，以简单、完整的操作步骤，清楚的操作界面图，丛替了冗长、难懂的文字叙述，对于初学者来说，通过参照书本内容边学习边操作，在实践中逐渐理解，必将使学习的过程变得轻松快捷。同时，当用户在实际操作中遇到问题时，这套书将更方便用户查找问题所在，及寻求正确答案。

对于众多的计算机用户来说，不断涌现的各种应用软件既让人兴奋不已，又令人望而却步。这套书所介绍的软件，都是经过认真筛选、较有影响的软件，在书中还配合讲解插入了大量操作界面图，对于读者了解、选择软件也会大有帮助。

我们期待这套书能给广大计算机用户带来更多、更切实的帮助，同时也欢迎广大读者对这套书提出宝贵的意见。

前 言

中文之星无疑是最出色的汉字 Windows 系统之一,许多 Windows 用户偏好使用中文之星汉字平台,一些中文技术系统(如 ORACLE)曾建议采用中文之星以确保可靠运行,许多应用开发人员总是将开发平台建立在中文之星之上。

一般而论,中文之星的优点是:

- (1) 速度快;
- (2) 汉字精美;
- (3) 功能全面;
- (4) 兼容性能好;
- (5) 可靠性高。

本书面向计算机初级或中级用户,高级用户也能够从中找到一些可借鉴的观点与方法。全书内容可分三大部分,即:第一章至第三章为汉字基础部分,涉及汉字技术的基本方法和汉字功能实现的矛盾与对策。第四章至第六章为基本操作部分,以中文之星为本,涉及汉字操作基本方法与要点;也包括了中文之星安装与配置方面的内容。第七章至第二十一章为辅助应用部分,针对中文之星的各个辅助应用,具体介绍各项功能的用途和操作技巧。

编写本书的目的之一是为了方便中文之星用户掌握操作使用方法,另一个目的是为了使用户能够知道自己使用的汉字系统究竟与西文系统有什么不同。事实上,本书中有关汉字技术方面的内容对于任何汉字 Windows 系统而言是通用的。作者的意图是使读者通过中文之星能够全面、深入地了解汉字 Windows 系统的基本概念,这对于理解、开发、使用、管理汉字系统是有利的。

当然,本书的中心内容仍然是介绍中文之星的使用方法,它涉及到几乎一切在屏幕上能够看到的中文之星的内容,即使对于一些实用价值不高的辅助应用,本书也用专门章节进行介绍。因为,中文之星的每一个辅助应用都代表着某一方面的汉字应用的需求,反映出开发者为满足需求而采取的技术措施。用户对此有所了解,有助于从整体上加深对汉字系统的理解。

中文之星的目标之一是支持西文软件处理中文,但是这样做总是需要了解和掌握某些特殊的汉字操作方法。为便于读者能够结合实际掌握要领,本书结合中文之星的操作使用,提供了较为全面的有关汉字输入编码方面的基础知识,包括汉字代码变换,拼音、五笔字型和国标/区位码规则;提供了使用西文字处理软件处理中文的实践经验,包括使用西文 Word 处理中文的技术措施;也提供了较深层次的图形/图像应用中使用汉字字形的知识。虽然这些内容没有独立的章节,但是它们贯穿于全篇内容的许多方面。

阅读本书需对 Windows 有所了解,能够掌握基本的操作技术,熟悉最常用的技术词汇。

本书出自中文之星用户之手,书中内容主要来源于使用中文之星的实践,如有不当,请批评指正。

参加本书编写工作的还有王平、胡建华、范庆年、殷韶、周京明、刘加明、舒志勇、章华等同志。

目 录

第一章 汉字 Windows 系统.....	1
1.1 PC系统的本质是西文的.....	1
1.2 汉字字符功能.....	2
1.3 汉字字符集.....	5
1.4 用西文系统处理汉字.....	7
第二章 汉字系统操作.....	9
2.1 从西文Windows环境中加载汉字系统.....	9
2.2 汉字输入.....	9
2.2.1 中西文输入状态识别与切换.....	10
2.2.2 选择汉字输入方法.....	10
2.3 汉字文本编辑和文件转换.....	16
2.3.1 一个汉字占用两个 ASCII 字符空间.....	16
2.3.2 高级字处理中汉字之间加空格.....	16
2.3.3 汉字文本格式转换.....	17
2.4 字库、字体和字形大小.....	17
2.5 造字.....	19
2.6 用于图形图像应用的汉字功能.....	19
2.7 将操作界面上的西文文字变成汉字.....	20
2.8 安装及配置汉字系统.....	20
2.9 小结.....	20
第三章 中文之星功能综述.....	22
第四章 中文之星的操作界面、状态及切换.....	24
4.1 启动.....	25
4.2 功能键.....	26
4.3 操作界面和状态显示.....	27
4.3.1 链形菜单管理器.....	27
4.3.2 汉字输入法提示行.....	29
4.4 汉字输入方法及状态选择.....	31
4.5 设置输入方法.....	32
4.6 字体、字大小属性及选择.....	34
第五章 汉字输入方法.....	39
5.1 新拼音.....	39

5.1.1	新全拼输入.....	40
5.1.2	新全拼词组定义	40
5.1.3	新双拼操作.....	41
5.1.4	小结.....	44
5.2	国标/区位.....	44
5.3	五笔字型.....	45
5.3.1	偏旁部首(字根)及对应按键.....	45
5.3.2	编码规则.....	46
5.3.3	简码.....	46
5.3.4	词组.....	47
5.3.5	通配键.....	47
5.4	其他输入方法.....	47
5.5	词组定义.....	47
5.6	动态键盘.....	48
第六章	系统安装及设置.....	50
6.1	系统安装.....	50
6.1.1	安装步骤.....	50
6.1.2	安装说明.....	54
6.2	安装后系统增加的内容.....	55
6.2.1	增加的内容.....	55
6.2.2	中文之星文件及类型	56
6.3	定制.....	57
6.4	设置输入方法.....	61
6.5	综合设置.....	62
6.6	修改CHINESE.INI	62
6.7	安装及设置要点.....	68
第七章	新笔字(NPS).....	69
7.1	NPS操作命令	69
7.1.1	NPS 菜单命令	69
7.1.2	NPS 工具条	71
7.1.3	NPS 键盘命令	71
7.1.4	NPS 鼠标操作	72
7.2	NPS打印属性命令	72
7.2.1	字型菜单.....	73
7.2.2	版面菜单.....	81
7.2.3	工具菜单.....	86
7.3	文件操作.....	87
7.4	打印和模拟打印.....	88

7.5 WDT格式处理.....	89
第八章 轻松表格(Easy Table)	93
8.1 操作命令.....	95
8.1.1 菜单命令.....	95
8.1.2 鼠标和键盘命令.....	96
8.1.3 操作状态显示.....	97
8.1.4 工具条、字形条和文字输入框.....	98
8.2 表处理操作.....	99
8.2.1 构造基本框架.....	100
8.2.2 数据编辑.....	100
8.2.3 调整结构.....	100
8.3 模板使用.....	101
8.4 编辑结构.....	102
8.5 设置字符属性.....	103
8.6 打印和模拟打印.....	103
8.7 OLE功能.....	104
第九章 文本格式转换器(Chinese Text Converter)	106
9.1 应用示例.....	106
9.2 说明.....	108
第十章 天工补字(Font Maker)	109
10.1 矢量图形处理与天工补字矢量字形处理	109
10.2 造字基本步骤.....	111
10.3 操作界面和命令.....	112
10.3.1 菜单命令.....	112
10.3.2 工具条命令.....	114
10.3.3 键盘/鼠标命令	114
10.4 字形编辑.....	114
10.4.1 绘画.....	115
10.4.2 点线编辑.....	117
10.4.3 特殊变换.....	118
10.5 存盘、模拟打印.....	120
10.5.1 定义代码和存新字	120
10.5.2 模拟打印.....	121
10.6 系统设置.....	121
10.6.1 用户定义笔划	121
10.6.2 系统设置.....	122
10.6.3 显示设置.....	122

10.7 用户字形文件.....	123
第十一章 单字节字库(Single Byte Font)和单字节输入(Single Byte Input).....	124
11.1 操作.....	125
11.2 注意事项.....	127
第十二章 逻辑字形编辑器(Logical Font Editor)	128
12.1 定义逻辑字形.....	129
12.2 说明.....	130
第十三章 艺术汉字图形生成器 (Font Art).....	131
13.1 操作界面.....	131
13.1.1 菜单命令.....	131
13.1.2 按钮命令.....	131
13.1.3 列表框、数据框、显示区	132
13.2 操作及艺术效果.....	133
第十四章 汉字图形(Chinese-->Picture).....	138
第十五章 动态翻译器(Translator).....	140
15.1 启动及界面.....	140
15.2 手动翻译.....	141
15.3 自动翻译.....	142
15.4 查分字典和查主字典.....	142
第十六章 动态键盘编辑器(DynKbrd Editor)	145
16.1 操作界面.....	145
16.2 命令按钮.....	146
第十七章 词库管理器(Word Manager).....	147
第十八章 通用码表编译器(MB Compiler)	149
第十九章 英汉字典(Dictionary).....	151
第二十章 中文 DOS 窗口(Win-DOS).....	152
第二十一章 中文之星辅助应用总结.....	153
附录 A: ASCII 控制字符集.....	154
附录 B: ASCII 标准字符集.....	154

附录 C: ASCII 扩展字符集..... 155

附录 D: 大 ASCII 字符区位码表.....155

附录 E: 五笔字型简码表157

第一章 汉字 Windows 系统

本章从实现汉字功能的背景环境开始,讨论汉字系统基本功能的组成。之后,涉及汉字系统的基础:汉字字符代码。最后,讨论汉字功能在 PC 机的多种应用中的作用。本章所讨论的内容并非局限于中文之星。

本章提出和讨论以下观点,这些观点是认识、使用 and 评价汉字系统的基本点:

- (1) 汉字系统是西文加中文的系统。
- (2) 汉字系统是西文字符功能基础上的汉字字符功能的扩充。
- (3) 汉字系统的主要目标是用西文系统来处理汉字。

1.1 PC 系统的本质是西文的

对于使用过西文 Windows 同时也使用过汉字 Windows 的用户,都会觉得汉字 Windows 中的中文功能远不如西文 Windows 中的西文功能那样简便、直观和规范。使用汉字系统总是比较麻烦、比较受限制和比较难掌握。对照 Windows 的西文功能,理想的汉字系统自然应该是:凡是西文所具有的功能,中文也应该具备。因此,汉字系统应具备下列中文功能:

能够输入、显示、打印汉字,系统操作界面是汉字化的(显示画面上的文字采用汉字)。

能够在文字数据处理方面处理汉字,遵循中文行文习惯,如使用中文标点符号和按中文版式排版。

更进一步,在任何数据或数据处理过程中,只要其中能够用西文,就也可以用中文。包括在图形/图像,动画,视像,网络传输,多媒体、程序和系统管理等各个方面使用中文。比如,在图画上“写”汉字,在动画或视像中加汉字标题或字幕文字,用网络传输汉字字符数据,或用中文下达控制命令,等等。

此外,要求中文操作使用的简便程度、美观程度、处理效率等诸多方面应与西文相当。

对照以上要求,任何汉字系统的用户都会说:“那可差得太远了!”的确,目前的汉字系统是差得很远,不仅如此,就是在将来,汉字系统恐怕也很难达到上面的要求。原因是什么呢?因为:

● PC 机是“英文机”,其基础部件是以英文为基础设计的,如键盘。更英文化的是系统软件, BIOS 程序、DOS 和 Windows 以及产生系统软件的程序设计语言,它们都是以英文作为描述语言的母语,由此导致了英文化的控制命令和程序语句。

● DOS 和 Windows 原本是为西文处理而设计的操作系统,任何汉字系统都是以这样的西文系统为基础,经过修改、补充、扩充而成为汉字系统的。

● 在汉字系统开发时,开发的原则是:首先是要保持原有的西文功能有效,其次才是要尽可能地扩充汉字功能。

在上述条件的限制下开发出的汉字系统当然是“不够理想的”。

打个形象的比喻,目前的汉字系统很像一个学习汉语的西方国家“留学生”,用汉语讲

话难免带着“洋腔调”，而用英文话说则十分流利，特别是它只能够“听”懂英文的命令。再进一步地将电脑比喻人脑，无论电脑的中文水平有多高，其“大脑”中的信息交换母语是英语。从语言文字发展的进程上看，电脑语言的词汇已经反映在人们的日常语言之中，一些汉字系统上的翻译新名词牵强附会，这些新名词的产生，不仅是因为它们多出自非语言行家的程序员之手，它更是在汉化西文系统时、在代码和显示长度有严格限制的状态下产生的（长度不能超过原有的英文词句）。总之，汉字系统不可能是很“纯粹”的。

目前，实用的汉字系统的中文功能已能够满足一般的要求。一些汉字系统功能较强，中文水平更高些，比如中文之星更多地考虑到图形/图像处理中使用汉字的问题，四通利方则考虑到 Internet 网上的中文传输。

另一方面，汉字系统的中文“不够理想”，但其他方面却保持了原来的水准。说“其他方面”而不说“西文方面”是因为文字功能只是 PC 系统多种功能中的一种。除文字功能外，还有运算功能、网络功能、图形功能等等，汉字系统的 PC 机，原有的各种功能依然十分强大，同时还在文字功能中增加了中文文字功能。

文字，是诸数据类型中最简练、最精确而且表现力最强的一种，文字功能的使用不仅只是局限在文字数据处理的范围之内，它也经常使用在其他类型数据的处理之中，甚至用来控制管理 PC 机系统。因此，在文字之中增加汉字不仅要考虑文字处理，同时还要考虑到与文字有关的方方面面。在扩充了汉字的系统上，总是存在着这样的问题：在某些情况下能够用汉字，在某些情况下能够有条件地用汉字，而在某些情况下不能够用汉字；在使用汉字时，某些情况下十分方便，某些情况下很不方便，某些情况下非常困难。

1.2 汉字字符功能

本节讨论在西文系统中扩充中文首先需要解决的最基本问题，即“汉字字符”问题。

汉字系统的设计目标是实现汉字字符功能，因此将其称之为汉字字符系统更为确切，其他的中文功能都是围绕汉字字符功能而展开的。那么，汉字字符具体是什么？它与西文字符有何不同呢？

字符，是 PC 系统中一种最基本的数据类型，对应人们日常书写使用的文字。人识字是识别字的字形，而计算机识字是识别字的代码值，如代码值“41H”对应英文字母“A”，代码值“BOA1H”对应汉字“啊”。另一方面，计算机显示文字如同人们写字，它在屏幕上“写”出的是人能够识别的“字形”而不是代码值。对计算机系统而言，它所能够使用的字符具有以下三方面的特征：

- 它有一个独特的数值作为代码。
- 它有对应的、能够显示和打印出的、人们能够识别的字形。
- 它存在于系统所能够识别的所有的字符组成的集合中；集合中字符的数量是有限的、有序排列的，其序列的位置往往对应于代码值和固定的字形。

在西文系统上，存在着一套西文字符集，字符集中的字符能够用来书写西文语句。而在西文系统上扩充汉字时，首先要在系统中扩充出汉字字符集。注意，所谓扩充，是指要在保留原先的西文字符的基础上增加汉字字符。

开发汉字系统，要做的最基本的工作包括：

1. 采用独特的汉字代码体制

西文的最小组成是字母，由字母组合组成单词，再由单词和西文标点符号组成文章。

中文的最小组成是汉字，由汉字和中文标点符号组成文章。

计算机系统中基本字符是英文，它包括字母、数字、算术符号和西文标点等。基本字符的总数远远少于汉字字符。数量上的差别决定了它们必须采用不同的代码体系：

西文字符代码：长 8 位，代码空间为 2 的 8 次方等于 256。在 256 个西文代码中，前 128 个称为标准 ASCII 字符集，后 128 个称为扩展 ASCII 字符集。标准 ASCII 集中，前 32 个代码直接用作控制命令，因此又称控制代码或控制字符；后 96 个代码对应大小写英文字母、标点、数字、算术符号等等，它们都是有“形”的，因此又称显示字符。PC 机系统能够识别标准 ASCII 代码组合而成的命令字或关键字。西文字符代码为八位二进制数，一字节存储空间正好存储一个字符代码：

X X X X X X X X

这里注意，标准 ASCII 代码与扩展 ASCII 代码的二进制数值的区别在于字节高位是 0 或 1。

中文字符代码：长 16 位，代码空间为 2 的 14 次方等于 16384(理论上，16 位数的代码空间为 2 的 16 次方等于 65536)。其二进制的代码形式为：

1 X X X X X X X X 1 X X X X X X X

显然，中文代码值与扩展 ASCII 代码值是相似的，单独看一个高位为 1 的字节数值，它可能是扩展 ASCII 代码，也可能是汉字代码的一部分。人对此数值的识别要参考前后文，甚至要了解字义之后才能下定论。机器的识别十分机械，西文系统只认为它是扩展字符；中文系统认为它是双字节汉字代码的一半，其前后还会有一个同样具有高位为 1 的数值，那就是另外半个汉字代码。

由于系统汉化处理要保持原有的控制功能，而系统命令及关键字必须是标准 ASCII 字符拼写的。因此，中文系统保留了单字节的标准 ASCII 代码，而它的代码值又完全能够明确地与汉字代码值区别开来，因为标准 ASCII 代码的二进制数值的高位总是 0。

0 X X X X X X X

综上所述，实际的中文系统的字符集中包括中文和西文两种字符，区别在于代码的高位是“1”或是“0”。是“0”的，代码长度 8 位，是西文代码集中的前 128 个字符。是“1”的，代码长度为 16 位，即是汉字字符。

依照代码区分方法，系统在西文代码处理程序前增加了高位识别指令开关，对高位为“0”的代码，控制沿用原先的处理；对高位为“1”的代码，控制转向扩充的汉字模块进行汉字处理。这样一来，西文系统中高位为“1”的扩展 ASCII 字符代码被汉字代码“挤掉”，这使系统字符集中减少了 128 个西文扩充字符，但增加了许许多多的汉字字符。由于扩展 ASCII 代码不能用来组成系统命令，因此取消它并不影响系统的功能而只是降低了西文字符

的“表现力”。

汉字字符集及 16 位中文代码的编码规则是标准化的，通常是按照 GB2312-80 定义的，或是由它扩展而来的。GB2312-80 定义了最常用的 6763 个汉字的次序和代码，有关说明稍后给出。

2. 扩充汉字字形

实现中文字符功能，自然要求系统能够显示和打印出人们能够直观识别的汉字字形，因此需要为系统的显示或打印提供汉字字形数据。字形数据又称字库，具体指字形数据文件。字库中的字形数据通常是按字符集中字符排列顺序分组存放的，一个字库包括某一字符子集中全部字符的字形数据，如标准 ASCII 字库中包括 128 个标准 ASCII 字符的字形数据，一个汉字库包括全部标准汉字字符集中每一个汉字的字形数据。当系统显示文字时，依照字符代码值确定字库文件名和计算出对应字形数据组在字库文件中的地址，显示程序按地址访问字形数据并用于显示字符。字形数据有缺或错误将使某些字符的显示有误，但这不影响代码数据的有效性。

部分常用汉字字形也有国家标准，如 16×16、24×24 的汉字点阵字形。

从字形上看，汉字字形与西文字母字形的特征不同，汉字字形要宽，而且宽度一致。如果不对原有的显示程序加以改造，用西文字符的显示功能来显示汉字，肯定是一片“混乱”。

3. 解决汉字字符输入问题

定义了汉字代码和扩充了汉字字形后，系统具备了识别汉字代码和在屏幕上显示汉字字形的功能。但是，这样的系统中的汉字数据如何而来呢？实用的字符处理系统必须具有字符输入功能，这也是建立汉字系统首先要解决的基本问题。输入汉字当然用键盘是最方便的，这样做的最大好处是无须在系统上增加特殊的汉字输入设备。

如前所述，PC 机的键盘是西文化的。按动键盘按键即可直接输入西文字符。西文输入状态下，键盘的键与可输入的字符是一一对应的。现在借用这样的西文键盘来输入汉字，显然必须要解决用几十个键去对应成千上万个汉字字符的问题。

解决问题的方法也不难想到，“一对一”不够数，就用“多对一”方法。

汉字输入，采用顺序的多次按键来输入一个汉字的基本方法，按键的序列称之为汉字输入编码。对汉字系统来说，要从输入编码到汉字代码进行换算；对操作员来说是要从汉字到输入编码进行“换算”。例如输入“中国”二字，操作员要动脑筋“算出”编码 KHK 和 LGY(五笔字型)。编码输入系统后，系统用编码换算出代码：D6D0，B9FA。

从上述分析中可以总结出一条基本规律：应用系统是人-机系统；扩充的应用系统，人、机两方面的功能都要“扩充”。具体讲，人要“扩充”汉字操作技术，其中最重要的是掌握汉字输入方法。

输入汉字对操作者提出了一个十分苛刻但又是绝对必须的要求：即操作者必须能够记住数千个汉字的键盘输入编码。许多人认为掌握计算机比较困难，主要就是指掌握输入方法太难。为了方便人们掌握汉字输入编码，人们设计出了各种各样的汉字输入编码方案，力图创造出易于掌握的或者“免于记忆”的输入编码。

汉字输入方法的本质是输入编码与汉字字符代码的映射关系，这种映射应当是统一和标准化的。但是，目前在汉字输入方法上没有专门的标准。下列的输入方法具有实用标准的意义：

国标/区位码，其原理源于 GB2312-80。

拼音码，以国家汉语拼音方案为基础。

电报码，邮电行业标准，在电报业务中使用。

五笔字型码，以偏旁部首组字方法为基础，是最流行的输入方法，为许多行业选用。

实用的汉字系统一般具有国标/区位、拼音、五笔字型这三种输入方法。

了解多种汉字系统后，更能体会到标准化的迫切性。汉字 Windows 系统的标准化是有限的。在输入编码、汉字字形、操作方式等许多方面，各系统各具特色，各有不兼容现象存在。

可以认为，实现了汉字代码、扩充了汉字字库和增加了汉字输入功能的系统是一个具备实用价值的基本汉字系统。这一基本系统能够用来输入、显示、打印、存储汉字数据。

1.3 汉字字符集

本节深入讨论字符集方面的问题。如前所述，任何系统字符来源于系统的字符集，字符集中的每一个字符具有特定的代码值。中文系统的字符集中包括高位为“0”的单字节长度的标准 ASCII 字符(子)集和高位为“1”的双字节长度的汉字字符(子)集。

汉字字符集，由 6763 个汉字组成，分一级字和二级字两部分，一级字 3755 个，由最常用的汉字组成，按汉语拼音字母顺序排列。常规文章中用到的汉字，一级字符占 95%以上。二级字 3008 个，多为名词专用字符或其他，按偏旁部首笔划顺序排列。此外，在 6763 个汉字之外，实际的汉字字符集中还包括扩充定义的近千个特殊符号，它们具有汉字字符的特点，是汉字字符集的重要组成，简称为全角字符或大 ASCII 字符。

汉字字符集“挤占”了 ASCII 扩展字符集的位置，同时为加大代码数量，将原来的单字节对应一个代码的定义改为双字节对应一个代码，这使扩展集空间在理论上扩展了 128 倍，即 $128 \times 128 = 16384$ 。这一扩展方案使用了两个 7 位的二进制数来表示一个汉字，它实际占用了两个 8 位的字节，其中包括每一字节的高位为“1”的特征位。如此一来，一个汉字代码由两个扩展 ASCII 代码组成，每一字节的值必然是 80H 以上。理论上顺序排列的双字节汉字代码值自 8080H 开始，80FFH 之后的代码是 8180H。

实际的汉字代码方案与理论分析不同。实际方案中，将可用空间分成长度为 128 的段，然后将 6763 个字符顺序地排列在各段中。为技术实现的方便，排列不是连续的，具体的排列考虑到下列情况：

(1) 分区。为方便地实现十进制的区位码方法(分区和区中位置编码值为两位十进制数)，定义每 94 个汉字为一个区，占用了长度为 128 的代码段中的后 33~126 个代码位置。

(2) 大 ASCII 字符空间。汉字字符集除 6763 个汉字外还包括了近千个双字节代码的特殊符号，其中有多种西文符号、日文片假名符号、俄文符号、中文标点符号、制表符等等。这些符号代码占用了汉字代码之前的 16 个代码分段。

(3) 留出全部空间中 最前面的 33 个段(4224 个代码)的代码位置不用。

因此，实际的汉字及代码定义如表 1-1 所示。

表 1-1

汉字字符及代码对照表

	序号	汉字	内码	传输码	区位码
大 ASCII 字符	0001	空格	A1A1	2121	0101
	0002	、	A1A2	2122	0102
.....					
一级字:	0001	啊	B0A1	3021	1601
	0002	阿	B0A2	3022	1602
.....					
	0094	剥	B0FE	307E	1694
	0095	薄	B1A1	3121	1701
.....					
二级字:	3756	亍	D8A1	5821	5601
	6763	龠	F7FE	777E	8794

表 1-1 中:

“序号”指字符在字符集中的顺序位置,该值是在字库中访问字形数据的相对地址值。

“内码”即系统内的代表汉字的代码,它与标准的汉字字符代码不同。

“传输码”是标准的汉字字符代码,汉字数据在一些网络上传输须用传输码。传输码由两个字节组成,每个字节的代码值小于 7FH,字节的高位总为“0”。它与内码的区别也仅仅在于字节的高位是“0”或是“1”。在 PC 机系统内,传输码形式的汉字数据不可识别。

“区位码”是最原始的汉字输入编码,它用十个数字键来输入汉字。从区位码值能够直接换算出:序号、内码和传输码。

由上表情况又可以看出,有效的汉字代码在代码空间中的分布是相当松散的。在理论上可用的代码空间($128 \times 128 = 16384$)中,总数 8000 多个汉字的分布是不连续的,从前到后占用了大部分扩展得到的空间,而未使用的空间支离破碎,这给进一步扩充汉字造成不利,如图 1-1 所示。但是,据专家估计,汉字总数在十万左右,常用古文汉字也大约有两万个。从此又可以看出,即使节省使用扩展代码空间也不足以满足实际要求。为此,一些汉字系统采取了另外的增大代码空间的技术措施,如采用三字节代码。在国际标准方面,已经制定了通用的双字节字符代码方案,在一些系统中双字节方案已经得以实现。双字节方案于中文系统的建立十分有利,其意义不仅在于能够扩大汉字的数量,更重要的是中西文统一在一个代码体系中,使文字数据处理走向统一。

历史上曾经有过多种内码方案,但是目前各个汉字系统均采用高位加“1”方案。高位加“1”方案实现简单,但不可避免地使西文软件中原来用到的 ASCII 扩展代码被当成汉字,造成了严重的不兼容问题:表格线字符不能正确显示,这一问题在 DOS 环境下十分突出。如图 1-2 所示。在图形操作系统 Windows 上,表格线字符的问题依然存在,但影响不大,因为 Windows 的表格线大多不使用表格线字符表示而采用图形画线方法画出。

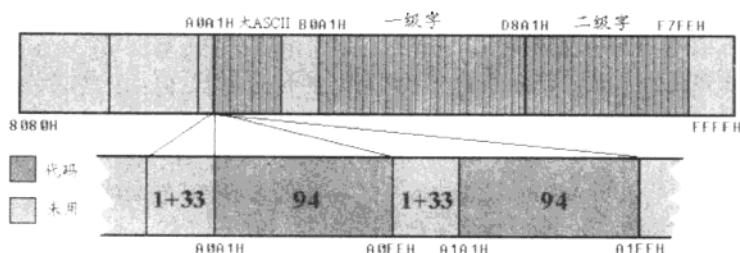


图 1-1 汉字代码在代码空间中的分布

另一方面，一些西文应用软件只能够处理单字节的字符，这使得系统的汉字功能对它们无效。中文之星为此专门设计了动态单字节汉字方法，可临时定义单字节汉字代码供它们使用。

此外，几乎所有的汉字系统都能够利用剩余的代码空间进行用户自定义字符的扩充。扩充汉字是一件颇为麻烦的事情，对扩充字必须分配一个特殊的代码值，还要为字形库“造”出对应的字形，有时要造出不同字体的多个字形。

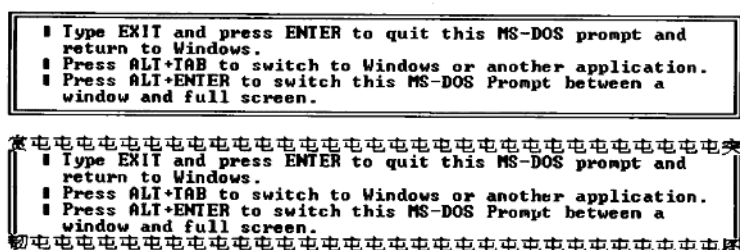


图 1-2 扩展 ASCII 字符在汉字状态下不兼容实例

1.4 用西文系统处理汉字

如上所述，基本的汉字功能包括输入、显示、打印和存储汉字代码数据，这样的系统当然能够实现汉字的字符处理。

除此之外，中文系统可能具备更强的汉字功能，例如：

- 操作界面的汉化，即将操作界面上原先的英文文字翻译成中文。一般的中文系统采用修改系统程序中的文字代码方法实现，一些系统则采用动态置换方法在运行前或运行中翻译界面上的文字。

- 使系统具备更多更强的字/表数据的功能，如方便准确地使用汉字标点符号、货币符号和时间中文格式，文字排版按照中文行文规则进行等等，这些功能往往实现于字处理应用软件中。一些中文系统直接提供专门开发的汉字软件，如微软中文系统的中文 Office 和中文之星的 NPS、Easy Table。但是，对多数系统来说，中文功能的提高是靠借用西文软件的功能实现的，它们基于这样一种前提：中文系统的功能非常之好，好到能够支持用功能强大

的西文软件来处理汉字。

在图形/图像处理中使用汉字。在作图时经常需要在图上写汉字，大部分应用软件是直接使用操作系统的字符/字形功能的，因此写汉字不成问题。但是有些应用软件为追求精美而使用自备的特殊字符/字形功能，此时实现汉字功能需要采取特殊技术措施。一些情况下是对应用软件做汉化处理，将其改造成为中文软件；也有一些系统采用外部汉字字符/字形模块补充不足。一些汉字 Windows 系统标记有特殊的版本号，表明能够提供某个或某类应用软件专用的汉字模块，如对 CAD 或三维动画应用。

网络应用中实现汉字字符传输。首先要明确，网络传输控制命令和参数必须是西文的，不得使用中文。网络传输文字数据分两种情况，第一，对于二进制数据传输，数据值无任何限制，因此数据是任意的，也无须区分中文或西文。另一种情况是流控制下的文字数据传输，系统逐字监视传输中的数据代码。此时，传输的数据必须是系统所能够识别的字符代码数值，其他数值可能会理解为是错码或控制命令。网络也是一种西文化的信息系统，它原本不能够识别汉字字符，因此直接用内码传输，网络不能将其理解为汉字代码。此时，在流控制的网络传输方式下传输汉字，必须要对汉字代码作特殊处理。

一般的方法是去掉汉字代码字节高位的“1”（使内码成为传输码），同时在代码串的两端加汉字标志“括号”，“括号”表明其中的数据是汉字的传输码而不是 ASCII 代码。

比如，用“~{”和“~}”做汉字标志“括号”，传输“中国”的过程为：

(1) 将“中国”的内码“D6D0 B9FA”，去掉高位的“1”成为“5650 397A”，对应的标准 ASCII 字符为“9z vp”；

(2) 在代码的前后加汉字标志“括号”，得到字符串：~{9zvp~}；

(3) 将 ~{9zvp~} 传出去；

(4) 接收方收到传来的字符串，识别到“~{”后，将其后至“~}”间的代码高位置“1”，还原成为“D6D0 B9FA”，即“中国”。

这一过程简单描述为：内码-->传输码，传输，传输码-->内码，过程中两次代码转换是汉字系统扩充的功能。

数据库，用西文的数据库系统来处理中文数据，需要考虑一些特殊情况，如能够使用汉字来定义库结构，用汉字作为查找关键字，按汉字排序规则使用排序功能等等。

综上所述，汉字系统的功能不仅涉及 Windows 系统，也涉及到多种多样的应用系统，汉字功能的开发已经深入到了一些流行的应用软件方面。深入开发的方式大致上有：提供汉字应用软件；给西文应用软件配套中文功能模块；对西文应用系统进行改造，使之能够用来处理汉字。所有这些工作的根本目的都是为了更好、更多、更方便地提供汉字功能。