

增加中文 Windows 输入法初探

高 骏 张红军

(河南省政法管理干部学院, 郑州, 450002)

摘 要 以五笔字型为例, 介绍为中文增加中文 Windows 增加自定义汉字输入法, 着重介绍通过 UC DOS 5.0 的 IMDMNG(汉字输入法管理器)产生中文 Windows 的码表输入法。

关键词 汉字输入方法; 中文 Windows; UC DOS; 编码

中图分类号 TP 31

目前, Windows 作为最流行的操作系统, 提供了许多功能强大、使用简便的优秀应用软件。如: Word, Excel, Power Point 等程序设计办公自动化系统; Visual C, Visual Basic, Foxpro For Windows, Visual Foxpro 等程序设计语言; Auto CAD, Corel 等图形设计软件。但是国内真正使用中文 Windows 来代替 DOS 进行日常工作的用户并不太多, 究其原因, 除了 DOS 系统对配制要求低、性能稳定这一原因外, 中文 Windows 的汉字输入法太少, 没有给广大用户提供熟悉的汉字输入法也是一个很重要的原因。

其实, 中文 Windows 有一个被称为码表输入法的汉字输入法生成系统, 它可以让用户在中文 Windows 下建立并使用自定义的汉字输入法。下面我们就以中文 Windows 3.X 五笔字型汉字输入法为例来说明如何实现这一功能, 其它输入法可参照五笔字型来实现。本例已经在 386, 486, PENTIUM 等多种机型上通过。

1 中文 Windows 3.X 的码表输入法及其码表源文件结构

码表输入法实际上是一种汉字输入法自动生成系统。对于任何一种汉字输入法, 其汉字的输入过程就是从汉字编码到汉字/词(或符号)的映射过程。用户可以根据自己的需要和想法建立自己的汉字/编码关系。中文 Windows 的码表输入法为用户提供了将自己建立的汉字/编码映射装入工作平台的接口。

要使用码表输入法, 必须首先建立码表源文件。码表源文件是以 .TXT 为后缀的文本文件, 具体格式如下:

[Description]	[TEXT]
Name = 输入法名称	编码 1
MaxCodes = 最大码长	编码 2
UsedCodes = 码元
WildChar = 查询键	编码 n
Sort = 排序标志	

收稿日期: 1997-11-12; 修改稿返回日期: 1998-808-14

第一作者 女 1962 年 12 月生 大学本科 助教

由上述可知,中文 Windows 的码表源文件共分为两部分:编码规则定义部分(Description)和正文部分(TEXT),其各项的功能如下:

输入法名称:用户自定义码表输入法的名称。

最大码长:用户设置编码的最大长度。

码元:构成编码的基本元素。

查询键:在录入时若是对编码的某一个键把握不准,可通过查询键迅速找到所需要的字或词。

排序标志:当排序标志置为 1 时,编译程序在编译时对码表中的编码进行排序,若置为 0,则不排序。若码表的编码字串是按码表源文件中定义的码元顺序排序,则可以将排序标志置为 0,这样会大大减少编译时间,如果没有排序,则需要将排序标志置为 1。码表法对于已排好序的码表中编码条目总数没有限制,而对于未排序的码表,建议编码条目总数在 7000 条以下。

编码:中文字串+编码字串。中文字串和编码字串之间没有任何字符,其顺序按编码字串排序。

2 利用 UC DOS 5.0 的 IMDMNG(汉字输入法编码管理器)生成编码文件

目前国内流行的中文 DOS 操作系统 UC DOS 5.0 提供了一个功能很强的万能汉字输入法管理器 IMDMNG·EXE。用户通过它可以方便地维护 UC DOS 输入法的码表,并可以将 UC DOS 本身提供的汉字输入法变为编码文本文件,将这编码文本经适当变换即可成为中文 Windows 的自定义汉字输入法。下面介绍用 IMDMNG 将编码字典反编译成编码文本文件的功能。如需了解更多的内容,可参阅 UC DOS 5.0 的 README 文件。

将 UC DOS 的编码字典反编译为文本形式的编码文本文件,可用如下命令:

```
IMDMNG<ImdFile><DicFile>
```

其中:<ImdFile> 为已经编译好的编码字典文件。

<DicFile> 为编码字典文本文件

以五笔字型汉字输入法为例,假设 UC DOS 系统安装在 C 盘 UC DOS 目录下,万能输入法编码字典文件在 UC DOS \ DRV 目录下,应在 DOS 下执行命令:

```
C: \ UC DOS \ IMDMNG C: \ UC DOS \ DRV \ WB·IMD \ WB·TXT
```

以上参数可以根据实际情况改动。如果要用其他输入法,改动相应的编码字典文件即可。

执行完这一命令以后,将会产生一个 WB·TXT 文件,这就是 UC DOS 的五笔字型编码文本文件。UC DOS 的编码字典文本文件共分为两部分,编码规则定义部分和正文部分,其格式如下:

名称=【五笔】	是否执行模糊搜索=N
功能键号=5	是否使用自定义词组=Y
码元表=abcdefghijklmnopqrstuvwx	自定义词组编码方案=1
万能键=z	a 工
最大码长=4	aa 式
是否自动选择输入=Y	...

aat 堯	aadn 葺 慝
aaaa 工 恭恭敬敬	aadw 萁
aaan 工艺	...
...	

3 将 UC DOS 下产生的编码文本文件转化成为中文 Windows 的码表源文件

中文 Windows 的码表源文件与通过上述方法所产生的编码字典文本文件的规则定义和正文两部分都有所不同。下面说明如何建立中文 Windows 的码表源文件:

3.1 建立中文 Windows 码表源文件的编码规则定义部分

用户可以按照中文 Windows 的码表源文件格式, 建立自己的码表源文件编码规则定义部分。我们以五笔字型输入法为例来说明:

[Description]	UsedCodes=abcdefghijklmnopqrst uv wxy
Name=五笔	WildChar=z
MaxCodes=4	Sort=0

3.2 建立中文 Windows 码表源文件的正文部分

用户可以通过一定的方法将生成的编码字典文本文件的正文部分, 转变为中文 Windows 的码表源文件的正文部分。我们首先看一下这两者的不同:

(1) 生成的编码字典文本文件正文以编码开始, 而中文 Windows 码表源文件的正文以汉字开始。

(2) 生成的编码字典文本文件正文部分的编码并没有按照定义的码元顺序严格排序, 中文 Windows 码表源文件的正文部分要严格按照定义的码元顺序排序, 否则会出现码表文件生成器生成的编码不完全的情况。

(3) 生成的编码字典文本文件的编码长度最长为 12 位, 不足部分用空格代替。中文 Windows 码表源文件中汉字后紧跟编码。

(4) 生成的编码字典文本文件中一行编码可以对应多个汉字或词组。中文 Windows 码表源文件中一行编码只能有一个汉字或词组。

只要将 UC DOS 产生的编码文本文件进行适当变换, 使之符合中文 Windows 码表源文件格式, 就可以在中文 Windows 下进行编译、安装、使用。进行转换的方法很多, 如可用 C 语言编写一个程序来完成转换, 但这种方法对于大多数用户来说有点难度。下面介绍一种利用数据库系统(如 FOXBASE)来完成这一变换的方法。

首先使用字处理程序删去 UC DOS 产生的编码文本文件规则定义部分并存盘。然后进入 FOXBASE 系统, 建立一个数据库 WB.DBF, 使 WB.DBF 只有两个字段, 一个名为 WBM(字符型, 长度为 12), 一个名为 WBZ(字符型, 长度为 60) 并建立一个命令文件 UC-TOWIN.PRG。命令文件的内容如下:

select 1	append from \wb.txt type sdf
use wb	select 2
copy stru to wbb	use wbb

```

sele wb
go top
do while .not. (eof())
lenn=len(trim(wbz))
lens=at(" ",wbz)
if lens<lenn
mrec=recno()
mwbz=wbz
mwbm=wbm
replace wbz with subst(wbz,1,lens)
append blank
replace wbm with wb->wbm
replace wbz with wb->wbz
sele wb
dele
skip
else
skip
endif
enddo
select wb
pack
select wbb
go top
do while .not. (eof())
lenn=len(trim(wbz))
lens=at(" ",wbz)
if lens<lenn
mrec=recno()
mwbz=wbz
mwbm=wbm
replace wbz with subst(wbz,1,lens)
append blank
replace wbm with subst(mwbz,lens+1,lenn-
lens)
replace wbm with mwbm
go mrec+1
else
skip
endif
enddo
use
select wb
append from wbb
index on wbm to wb
copy to \winwb field wbz, wbm delimited
with blank

```

执行 UCTOWIN.PRG, 即可产生一个名为 WINWB.TXT 的文本文件, 然后再将该文件中汉字与编码之间的空格去掉, 并加上符合中文 Windows 码表格式的编码规则定义部分, 即可产生如下所示的中文 Windows 的码表源文件。

[Description]	工 a
Name=五笔	式 aa
MaxCodes=4	工 aaaa
UsedCodes=abcdefghijklmnpqrstuv wxy	恭恭敬敬 aaaa
WildChar=z	工艺 aaan
Sort=0	...
[text]	

4 在中文 Windows 中安装新的输入法

要在中文 Windows 中安装新的输入法, 应首先将码表源文件编译成为码表文件, 在主组群中找到通用码表转换程序图标并双击它 (如果在主组群中没有找到通用码表转换程序图标, 可以在程序管理器中运行 CONVERMB.EXE), 在出现的“通用码表编译程序”对话框

的“码表源文件”输入框中输入“\ WINWB.TXT”。然后单击“转换”按钮,通用码表转换程序就会自动地使用码表源文件 WINWB.TXT 生成名为 WINWB.MB 的码表文件,此码表文件位于 SYSTEM 目录下。

产生码表文件后,就可以安装输入法了。这时可执行以下步骤:

(1)打开主群组图标,双击控制面板之后再双击输入方法;在出现的“输入方法”对话框的“输入方法”列表框中找到并选择“通用码表输入法”,并单击“选用”按钮,此时在所选用输入方法框中会出现通用码表输入法,单击“设置”按钮。

(2)在随后出现的“通用码表输入法”对话框中单击“安装”按钮。

(3)在出现的“安装码表”对话框中选择 WINWB.MB 并单击“确定”按钮。新的输入法就安装完毕了。

如果在第一步的“输入方法”对话框的“输入方法”列表中没有发现通用码表输入法,可单击“安装”按钮;在随后出现的“打开”对话框的“文件名”输入框中输入 WINMB.IME,单击“确定”按钮即可完成通用码表输入法的安装。安装完之后就可在输入方法列表框中找到并选择通用码表输入法了。

使用类似方法可以为中文 Windows 3.X 增加其它汉字输入方法,还能建立符合中文 Windows 95 格式的码表源文件,从而为 Windows 95 增加自定义输入法。

Primary Discussion on Adding User-define Chinese Characters Input Methods for Simplified Chinese Windows

Gao Jun Zhang Hongjun

(Henan Law Administrative Cadre Institute)

Abstract This paper introduced adding user-define chinese characters input methods for Simplified Chinese Windows by an example of WuBiZiXing, paying special attention to the way of generate MaBiao input method for Simplified Chinese through UC DOS 's IMDMNG (Input Method Dictionary Manager).

Keywords Chinese-character input methods; simplified Chinese Windows; UC DOS; code