

邓莲堂. GrADS 绘图汉字显示的技术研究与实现[J]. 气象, 2010, 36(11): 115-119.

GrADS 绘图汉字显示的技术研究与实现^{*}

邓莲堂

国家气象中心, 北京 100081

提 要: 气象常用绘图软件 GrADS 中不能用汉字设置标题和题注, 给国内的业务和科研工作带来不便。根据 GrADS 软件的特点, 在其已有功能的基础上, 对 GrADS 软件功能进行扩充, 发展了一些关键技术, 实现了 GrADS 汉字显示功能。其中, 发展的关键技术包括: 自定义汉字库创建、汉字库索引创建、中英文混合显示、快速查找自定义汉字库、解码自定义汉字库和汉字绘制。实现的汉字显示技术具有使用方便、速度快、可移植、可中英文混合显示等特点。

关键词: GrADS, 自定义汉字库, 汉字库索引, 中英文混合显示, 汉字绘制

Research and Implementation of Chinese Character Displaying in GrADS System

DENG Liantang

National Meteorological Center, Beijing 100081

Abstract: Generally, the GrADS software could not set the title or notation in Chinese. It makes some inconvenience for domestic scientists. Based on the GrADS existing functions, we develop a set of key technologies, and implement the displaying Chinese character function in GrADS. These key technologies are as follows: Set up a user-defined Chinese character base which is suitable for the GrADS system, create the index of Chinese character base, display Chinese characters along with English characters, find the exact character in the use-defined Chinese character base rapidly, decode the dot-array information of a Chinese character in the user-defined Chinese character base, and draw the Chinese character on the GrADS graphic area. The function of displaying Chinese character in GrADS is very convenient, efficient, portable, and could display Chinese along with English.

Key words: GrADS, user-defined Chinese character base, index of Chinese character database, displaying Chinese along with English, drawing Chinese character

引 言

GrADS(Grid Analysis and Display System)是当今气象界和海洋界广泛使用的一种数据处理和显示软件系统^[1-2]。该软件系统通过其集成环境,可以对气象和海洋数据进行读取、加工、图形显示和打印输出。GrADS 运行时,用户可以在 GrADS 提供的

交互式环境下输入单个绘图命令进行绘图操作。另外,GrADS 同时提供用户一个解释性的描述语言绘图环境,用户也可以使用 GrADS 描述语言进行简单编程,完成后台批处理操作。该软件具有操作简单、功能强大、显示快速、出图类型多样化、图形美观等特点。该软件兼具多种地图投影坐标系统,已经成为国内外气象、海洋数据显示的标准平台之一。

目前,GrADS 作为一种数据处理和显示软件系

* 中国气象局数值预报(GRAPES)专项和国家自然科学基金项目(40875064)共同资助
2009 年 9 月 14 日收稿; 2010 年 3 月 7 日收修定稿
作者:邓莲堂,主要从事中尺度数值预报研究. Email:denglt@cma.gov.cn

统在气象领域得到了广泛使用^[3-11]。但是,由于汉字处理技术本身的特殊性,GrADS 图形软件还不能处理汉字信息,没有进行汉字标注显示的功能。因此,如果需要在绘出了的图形上添加中文标注,则往往需要借助其他绘图软件(如 Photoshop、Windows 画图等)做一下额外的处理,添加上中文标注信息。因此,它的图形添加中文信息的步骤比较繁琐,图形处理效率很低,给国内科学工作者带来极大的不便,也不便于在实际业务系统中实现自动批处理操作。本文针对在 GrADS 软件中存在的这一无法显示汉字标注的问题进行了研究,使用户在不需要使用任何额外软件和编译工具的情况下,方便地使用 GrADS 软件进行汉字标识。

1 研究思路

在研究开发 GrADS 软件的汉字显示功能时,遵循模块化设计、程序易于维护、用户使用方便等的原则。尽量使用户在使用时操作简单,不用任何其他额外的软件工具,且不用进行编译和安装等操作。这就要求研究开发时根据 GrADS 软件的特点,充分利用 GrADS 提供的功能,实现 GrADS 软件在绘图区域绘制汉字的功能。根据这一要求,本研究的基本思路如下:

(1) 考虑到 GrADS 软件本身具有标注英文和符号的功能,其实现的方法是将英文字母及符号的形状信息提取出来,然后以图形的方式绘制到绘图区。根据此特点,可以将汉字看作特殊的标识符号,只要将汉字形状信息提取出来,然后以图形的方式绘制到绘图区。

(2) 将汉字的字形信息数值化后,保存在一个数据文件中,然后利用 GrADS 的数据读取功能来读取汉字的形状信息。

(3) GrADS 为用户提供的描述语言,具有简单的编程功能。可以对汉字形状信息进行简单地编程处理,并利用描述语言中的绘图语句进行绘图,从而实现汉字显示功能。

2 汉字显示的主要技术

2.1 自定义汉字库的创建

在 GrADS 绘图中实现汉字显示就是利用 GrADS 的绘图功能,在绘图区绘制出汉字的形状。

因此实现汉字显示的技术关键就是获取每个汉字的形状信息,创建自定义汉字库。本研究在创建自定义汉字库时,使用点阵汉字技术中的汉字点阵信息来表示汉字字形,采用了 16×16 点阵和 24×24 点阵共两种。将汉字的点阵信息经过编码后,以 GrADS 的二维数据片的形式写入到数据文件中。为了 GrADS 能正确读取自定义汉字库的数据信息,还必须配备相应的数据描述文件。自定义汉字库的具体创建步骤如下:

(1) 应用辅助软件,从标准字库中提取汉字的点阵信息。标准字库使用我国 1981 年公布的 GB2312-80 方案,其一级汉字为 3755 个(按拼音排序),二级汉字 3008 个(按部首排序)。由于本研究使用的辅助软件是网络资源中搜索的免费测试软件,只能提取到一级字库的 3755 个汉字。因此,自定义汉字库仅包含标准汉字库中的一级汉字库,即总共 3755 个常用汉字。

(2) 利用 FORTRAN 程序对提取信息进行处理。统计出每个汉字的有效点数、每个有效点在点阵数组的位置,并统计出 3755 个汉字的最大有效点数。

(3) 将自定义汉字库数据文件的 x 轴维数大小定义为所有汉字的最大点数加 1。 y 轴维数大小定义为所有汉字个数,这里应该为 3755。

(4) 将汉字的点阵信息存储在数据文件中。在 GrADS 软件中, x 轴表示行, y 轴表示列。一个汉字的点阵信息占用一行,即第一个汉字存在第一行,最后一个汉字存在第 3755 行。每一行的信息包括:①该汉字的有效点数;②每个有效点在点阵中的位置。其中,第一列存放该汉字的有效点数,其后各列依次存放每个有效点在点阵中的位置信息。为了节省数据存储空间,将有效点在点阵中的位置 x 和 y 值进行压缩编码,即按照 $x \times 100 + y$ 的编码规则将 x 和 y 坐标值合成一个实型数,作为该有效点的位置信息存放在一个数值单元中。

(5) 创建相应的数据描述文件。数据描述文件中,数据文件名即为自定义汉字库的文件名, x 方向格点数为最大点数加 1, y 方向格点数为 3755, z 方向格点数为 1,时次(t)数为 1,具体时间可以随意给定,变量数为 1。变量名称为了避免与通常的变量名重名,所以取名为‘cnziku01’。

从以上步骤来看,自定义汉字库的 x 方向格点数为点阵字库中所有汉字的最大有效点数加 1。汉

字的不同点阵数、不同字体的最大有效点数也就不同,自定义汉字库的 x 方向格点数的也不同。一般来说,汉字点阵数越大,最大有效点数也就越大。就字体而言,楷体字的最大有效点数最小,宋体字的有效点数次之,黑体字的最大有效点数最大(见表 1)。

表 1 自定义汉字库中不同字体信息
Table 1 Information of different type in user-defined Chinese character database

点阵大小	字体	最大有效点数	x 方向格点数
16×16	楷体	115	116
16×16	宋体	165	166
24×24	楷体	233	234
24×24	宋体	260	261
24×24	黑体	357	358

在实际的业务工作中,标准汉字库的一级汉字(共 3755 个常用汉字)基本能够满足需求。如果有标准汉字库的二级汉字和其他字符的点阵信息,那么也可以对自定义汉字库进行扩充。扩充自定义汉字库时,需按照上面介绍的自定义汉字库方法,将扩展汉字的点阵信息追加在自定义汉字库数据文件中。

2.2 创建自定义汉字库索引

用户输入汉字是与标准汉字库相联系的,而用于汉字显示的汉字信息则是与自定义汉字库相联系的。汉字显示程序需在标准汉字库与自定义汉字库之间建立一个索引,将标准汉字库与自定义汉字库之间建立一一对应的联系,从而实现根据输入汉字信息进行汉字显示的功能。

首先,利用 GrADS 的描述语言定义字符串的功能,将自定义汉字索引定义成一个长字符串,并赋值给字符串变量“zikuidx”。标准汉字库的一级汉字有 3755 个字符,如果作为一个整行一次输入,则在文本编辑器中往往会造成无法显示的尴尬,既不美观,也不利于程序的维护。因此,这里采用了 GrADS 的描述语言的字符串连接技术,将汉字分成若干行,每行 10 个汉字,对汉字库索引字符串进行累加连接。这样处理后,汉字库索引字符串整齐有序,便于查看和维护。

写入汉字库索引字符串的内容包括:①标准汉字库中的一级汉字;②每个汉字在自定义汉字库中的位置。其中,一级汉字共 3755 个,在索引字符串中的排放顺序与标准汉字库中的顺序一致。自定义

汉字库的汉字存放顺序也与标准汉字库的顺序一致,在 GrADS 打开后,可以按照 GrADS 的纵坐标(y 轴)来进行查找定位,数值从 1 到 3755。为了便于汉字的快速查找,这里将单个汉字的信息固定长度为 6 个字节,汉字本身占 2 个字节,自定义汉字库中的位置数值写成四位整数,占 4 个字节,不够的数位以 0 补充。比如汉字“啊”,其位置数值为 1,因此,写入索引长字符串中的写法是“啊 0001”。

创建汉字库索引的程序代码较长,如果直接放在主程序代码中不符合模块化编程的原则,也不便于程序的维护。因此,将汉字索引创建的功能封装成一个独立的函数子程序 zrcode()。这样在主程序中只要使用以下语句:

```
zikuidx = zrcode()
```

即可方便地获取自定义汉字库索引字符串。

2.3 中英文混合显示技术

显示汉字程序从输入参数中获取输入汉字串,接下来即可对逐个汉字单独处理显示。但是,为了方便,通常用户在输出汉字会夹杂一些英文字符。这就要求显示汉字程序具有处理中英文混合输入的功能。要实现中英文混合显示,则从中英文混合输入的汉字串中分离出汉字和英文字符是关键。

在计算机的文本文件中,汉字与英文字符的区别主要有:每个汉字占用 2 个字节长度存储空间,而英文字符只占一个字节存储空间;汉字的 2 个字节的最高位都为 1,单个字节数值都大于 160(16 进制为 A0H),而英文字符的内码则小于 128(16 进制为 80H)。因此,当某个单字节字符的内码小于 128(80H)时,可以判别该字符为英文字符;否则可以判别该单字节为汉字的一部分。

实际上,在构成输入字符串的英文字符中,内码最小的字符为‘!’(33),内码最大的字符为‘~’(126)。为了程序实现的方便,在输入字符串中判别是否为汉字,可将单个字符(以变量 c 表示)与字符‘!’和‘~’进行比较,当‘!’ $\leq c \leq$ ‘~’时是英文字符,当 $c \leq$ ‘~’或者 $c \geq$ ‘~’时这个字符为汉字字符,它和下一个字符将合成一个汉字。

由于 GrADS 系统是采用高级语言 C 开发的。因此,在 UNIX 系统中,字符的内码数据在内存中存储为无符号整数,汉字的内码都大于英文字符的内码,因此字符是否大于‘~’可以作为区分中英文字符的判据。而在 WINDOWS 系统和 LINUX 系

统中,字符的内码数据存储为有符号整数,汉字的两个字节最高位都为 1,其内码都是负整数,比英文字符的内码小,所以字符是否小于‘!’可以作为区分中英文字符的判据。利用这一特性,可将汉字显示技术在不同业务平台中移植,使其适合于不同的业务系统,如 Windows、Linux 和 UNIX 系统等。

2.4 快速查找自定义汉字库

在自定义汉字库索引中按照逐个汉字顺序查找的方式查找输入汉字,速度非常慢。因此,前面在创建自定义汉字库索引时,就考虑了快速查找的需要,在索引长字符串中使用顺序排放的方式,并且单个记录长度固定,以便于快速查找。本研究采用快速查找方法中的折半查找方法^[13]。应用折半查找方法时,首先将汉字库索引字符串视为共有 3755 个记录的有序系列,单个记录长度为 6 个字节。每个记录中前 2 个字节是汉字,用于与输入汉字进行直接比较。每个记录中后 4 个字节是汉字在自定义汉字库中的位置数据,数值范围为 1~3755。查找时将索引字符串分成两段,以第 1878 个记录为界,从 1~1877 个记录是低值区,从 1879~3755 个记录是高值区。将输入汉字与第 1878 个记录中的汉字比较,如果输入汉字与此汉字相同,则表示查找成功。在查找不成功时,如果输入汉字小于此汉字,则在索引长字符串系列的低值区重新查找;否则在高值区重新查找。重新查找时,将低值区或高值区中的字符串视为一个新的有序系列,应用前面的方法进行分段查找,如此循环,直到找到输入汉字或者查完整个有序系列为止。如果输入汉字查找成功,则可以直接获取此汉字在索引中记录的后四个字节的位置数值,得到汉字在自定义字库中的确切位置。应用快速查找技术,使查找速度提高了 5~10 倍,提高了程序的运行效率。

2.5 解码自定义汉字库信息

自定义汉字库中的信息是对汉字点阵信息经过编码处理后的信息。因此,在程序读取自定义汉字库信息时,需要对自定义字库的信息进行解压处理。首先,读出自定义汉字库中汉字的第一个记录单元的数值,即得到此汉字的总有效点数。其次,根据有效点数的数值,读取每个有效点的压缩编码。然后,对压缩编码进行解码,得到 x 坐标值和 y 坐标值。解码的具体公式为

$$x = \text{INT}(\text{code}/100)$$

$$y = \text{code} - \text{INT}(\text{code}/100) * 100$$

其中,code 是汉字的压缩编码,INT()为取整函数。

在实际的编程中,由于 GrADS 描述语言的数值和字符串可以直接等价使用。因此,在进行解码时,可以在读出复合数值后,将压缩编码的数值当成字符串来处理,即将 4 位数字字符串的后 2 位数字赋给 y ,将前 1 位或者 2 位数字赋给 x 。

2.6 绘制汉字

绘制汉字时,需要根据函数获取的字体大小、颜色、位置等输入参数,结合汉字的点阵信息中每个有效点的 x 和 y 坐标的数值,计算出每个点在绘图区域的 x 和 y 坐标值。程序显示输出汉字时,往往不是仅仅显示输出一个汉字,而是输出中英文混合的字符串。因此,在计算每个有效点在绘图区的位置时,还要考虑汉字在输入字符串中的相对位置信息。再利用 GrADS 的画点语句‘draw mark’,在绘图区的相应位置上绘制汉字的有效点,从而完成汉字的显示输出。

3 结论与讨论

根据 GrADS 软件本身的功能特点,探讨了 GrADS 中汉字显示技术问题,在 GrADS 中实现了汉字显示功能。文中介绍的汉字显示技术,有以下特点:

(1) 程序执行效率高。程序开发中使用了快速查找技术,使程序的执行速度大大提高。

(2) 研究了字符串的中英文分离技术,实现了中英文混合输入的功能。

(3) 程序使用方便,它可以作为 GrADS 的一个用户自定义函数来使用,用户使用时不需要任何额外的软件支持,也不需要额外的编译安装。

(4) 研究了不同操作平台中数据存储格式的差异,解决了在不同的操作平台中移植问题。

因此,本文介绍的技术弥补了 GrADS 绘图软件不能显示汉字的不足,使 GrADS 绘图软件使用起来更为方便,可直接应用于科研工作和实际业务系统中进行自动批处理操作。目前,该技术已经广泛应用于国内气象业务部门的日常业务运行当中。如国家气象中心用于数值天气预报业务系统的图形批量处理业务;广东东莞市气象局用于地面气象要

素可视化业务;上海台风研究所用于气象要素诊断业务等。同时,该技术也被收编到中国科学院大气物理研究所的《GrADS 使用手册》中,在许多气象工作者的科研工作中得到推广应用。另外,文中详细介绍的关键方法和技术,除了 2.1 部分内容外,其他技术都不完全局限于 GrADS 软件,在其他软件的相似工作中具有一定的借鉴意义。

参考文献

- [1] Institute of Global Environment and Society. GrADS Documentation Index[EB/OL]. [2006-06-12]. <http://www.iges.org/grads/gadoc/gadocindex.html>.
- [2] 中国科学院大气物理研究所大气科学和地球流体力学数值模拟国家重点实验室(LASG). GrADS 使用手册[S]. 2005.
- [3] 郑沛楠,白志鹏,李欢. GrADS 绘图软件简介[J]. 科协论坛, 2008,12:84-85.
- [4] 张娟,冯蜀青. 浅谈 GRADS 气象绘图软件对站点资料的使用[J]. 青海气象,2006,3:54-55.
- [5] 黄锋,陈明先,关鸿志. 用 GrADS 实现东莞市地面气象要素的可视化显示[J]. 广东气象,2007,29(2):54-55.
- [6] 黄明策,姚才. GrADS 在绘制站点等值线图及热带气旋路径中的应用[J]. 广西气象,2005,26(1):45-47.
- [7] 杨兆礼,万齐林. 实例分析如何用 GrADS 实现 NetCDF 格式转换[J]. 广东气象,2008,30(5):47-48.
- [8] 高歌,张强,高波,等. 中国地面气候要素图形分析业务系统[J]. 气象,2001,27(6):47-50.
- [9] 孙景荣,曹晓钟,张秀年. 基于 VC++ 的气象要素客观预报集成系统软件开发[J]. 应用气象学报,2006,17(suppl.): 147-151.
- [10] 严军,刘健文. NCEP/NCAR40 年再分析资料图形显示系统[J]. 气象,2002,28(5):37-43.
- [11] 陈国勇,朱健,诸晓明,等. 中尺度自动气象站资料的图形化分析应用[J]. 气象,2006,32(2):111-115.
- [12] 田会方,张洪昌. 计算机汉字库开发液晶汉字显示数据的方法[J]. 武汉理工大学学报(信息与管理工程版). 2005,27(4):83-86.
- [13] 谭浩强. C 语言程序设计教程(第二版)[M]. 北京:高等教育出版社,1998:152-155.