

基于 UNICODE 编码的蒙文编辑器

宋建斌¹，敖其尔²，马丽³

(内蒙古大学 计算机学院，内蒙古 呼和浩特 010021)

摘要：基于 UNICODE 编码的蒙文编辑器作为蒙文信息处理平台 (MIPP) 课题的一个子课题，它是蒙文信息处理工作的重点、难点之一。本文首先分析了提出该课题的背景知识和国内外研究现状、发展动态，介绍了基于 UNICODE 标准的新的蒙古文编码方案以及使用特点，在此基础上重点讲解了基于 UNICODE 编码的蒙文编辑器的设计、实现细节，给出了处理基于 UNICODE 的蒙古文文字的一些技术和理论问题。

关键字：蒙古文；UNICODE；蒙文国际标准码；MFC；蒙文编辑器

中图分类号：H085.6

文献标识码：A

一、前言

蒙古语言文字的信息处理工作在少数民族语言文字信息处理领域中起步较早、发展领先，我国计算机信息处理少数民族文字工作首先就是从蒙古文开始的。当前市面上的排版软件很多，市场占有率高的有微软的 WORD、国产优秀软件 WPS 等。目前版本的 WORD 支持竖排，但是，只支持从上到下从右向左的书写方向，蒙文的书写方向是从上到下从左向右，所以不能够支持蒙文的排版。最近金山公司推出 Windows2000 环境下的带有蒙文界面的 WPS 系统，能够支持 UNICODE 编码的蒙文排版。但是，考虑到软件版权，WPS 的技术细节并没有公布，成功的经验很难借鉴。最近北大方正推出了 Windows 环境下的 900 系统，把原来 DOS 下的排版系统升级到 Windows 环境。这些系统以文字处理（排版）为目标，并不具备通用平台的功能。

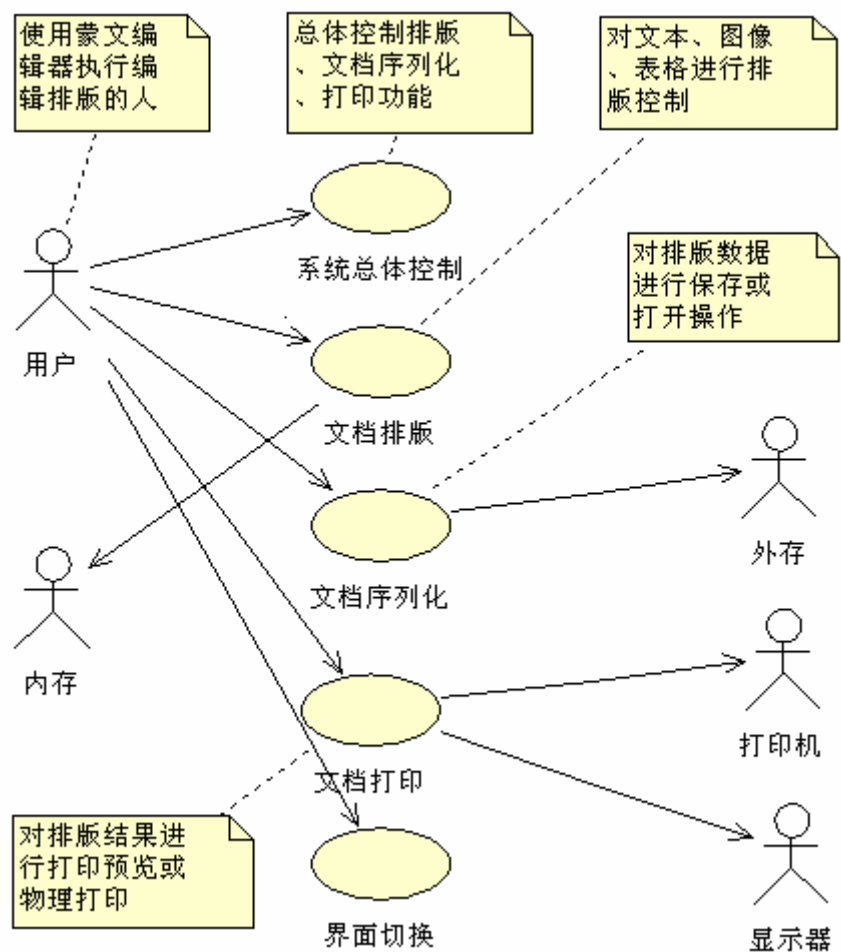


图 1：系统总体用例图

现在迫切需要一种以蒙文新标准码为编码体系的蒙文信息处理（软件开发）平台（MIPP），使蒙古文字与其它国际上的各种文字能够在相同的平台环境下工作。

MIPP 课题的基本内容包括四个方面：

（一）文字信息处理模块：（1）蒙文 True Type 字库；（2）蒙文输入法及外挂方式实现蒙文的输入；（3）蒙文编辑器。蒙文与其它文种最大的不同点在于书写方式上。蒙文书写是从上到下、由左向右的方向竖写。因此，我们需要开发一个面向蒙文的文本编辑软件；（4）搜索和排序；（5）对蒙文、满文、托忒蒙文、锡伯文等四个文种，提供名义字符串转换成显现字符串，显现字符串转换成名义字符串等编码转换功能。

（二）支持数据库的功能：（1）本平台将支持蒙文数据库的开发，客户端开发环境可以是 BCB、DELPHI、PB、VB、VC 等；（2）本系统支持 Windows 环境下的应用程序的蒙文处理，例如，可以在 WPS，WORD，PHOTOSHOP 等软件中处理蒙文。

（三）网络功能：（1）本平台支持蒙古语族文字——旧蒙文、托忒蒙文、满文和锡伯文的收发电子邮件功能；（2）支持制作蒙文网页和多媒体课件的功能；（3）支持网络数据库功能。

(四) 提供语料库 : (1) “ 汉蒙对照内蒙古地名电子词典 ” ; (2) “ 汉蒙对照机关企事业单位名称电子词典 ” ; (3) “ 英蒙对照电子词典 ” 等 ; (4) 其它语料库 , 如英蒙机器翻译用语料库等。

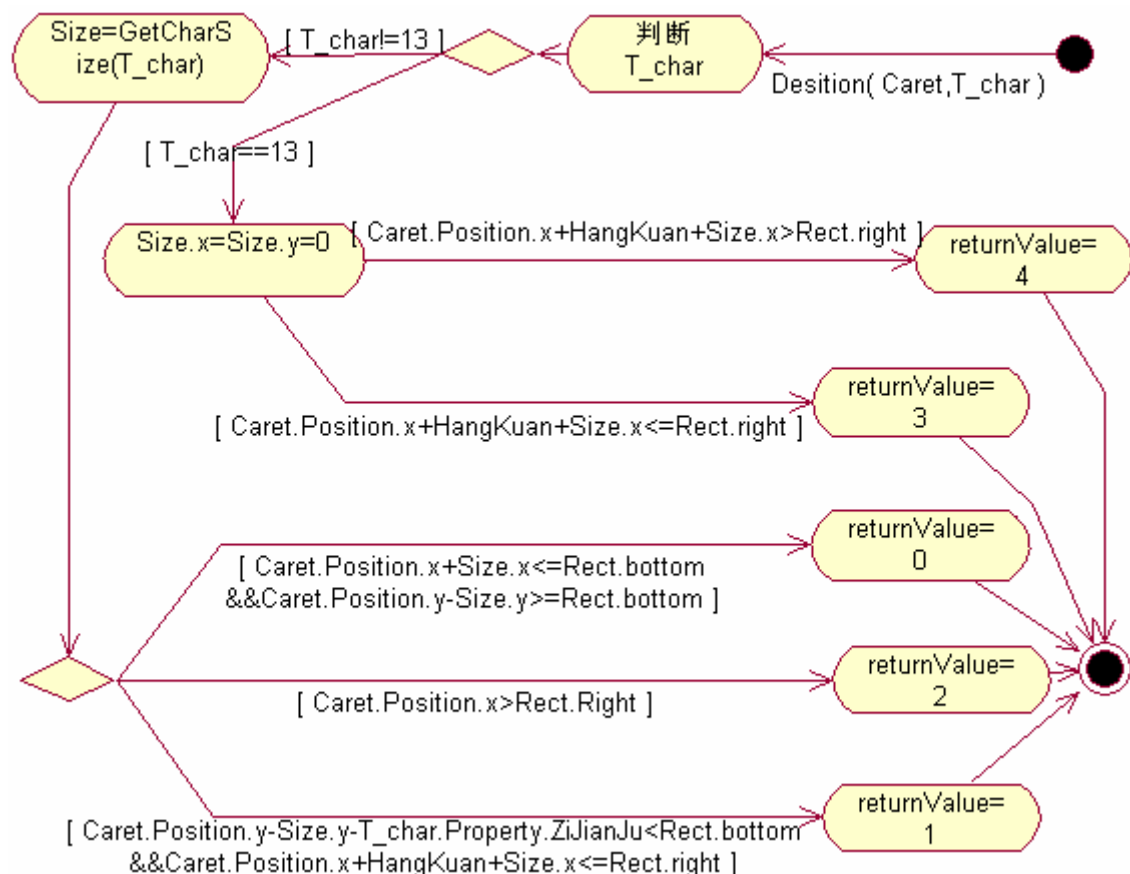


图 2 : 判断可输入字符活动图

内蒙古大学蒙文信息处理实验室早在 80 年代末 90 年代初 , 就开发出了一套处理蒙文的排版软件。这个软件在西文 DOS 操作系统下 , 成功的解决了内码冲突 , 实现了蒙文、中文和西文的混排 , 实现了图表文字的混排。在当时的技术背景下 , 取得了很好的成果。随着计算机技术的发展 , 以 Windows 为代表的图形操作系统逐步占据了主导地位。2000 年 , 出台了蒙文 UNICODE 标准 , 迫切需要研发出基于 UNICODE 编码的蒙文编辑器。

本课题的研究意义在于基于 UNICODE 编码的蒙文编辑器作为 MIPP 课题的一个子课题。它是蒙文信息处理工作的重点、难点之一 , 而且是蒙文信息处理的一项基础性研究工作。这项工作做好了 , 最直接的成果就是产生一个具有实用价值的蒙古文文字处理软件 , 形成自己独具特色的产品链 , 而且对于处理基于 UNICODE 编码的蒙古语言文字必将积累大量的技术、理论和经验 , 对后续的 MIPP 研发工作产生积极的影响。

本课题的研究方法主要是采用 Rational Rose 建模，利用它建立 MGWORD 系统的系统分析模型（主要指用例图），系统设计模型（活动图和协作图等）和系统实现模型。然后基于这些模型，利用 VC 中的 MFC 框架技术开发出合乎需求的蒙文编辑器。

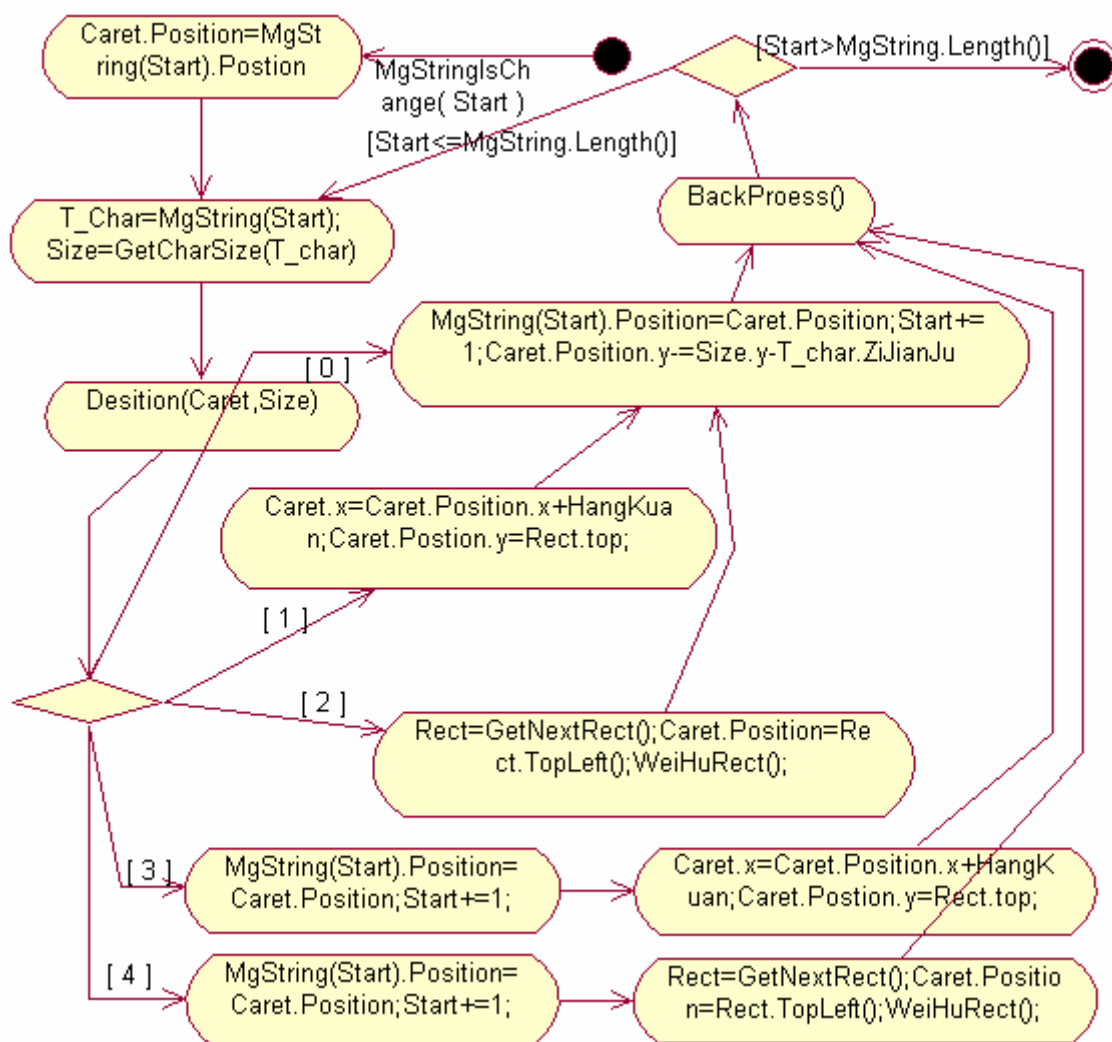


图 3：处理可输入字符活动图

二、蒙古文的编码方案

随着计算机技术的发展，蒙古文的编码方案也经历了“百家争鸣”到“百川归一”的发展历程。国内外较早的具有代表性的几种蒙古文编码有：“NEC”编码（1992）、“SUDAR”编码（1992）、“MLS”编码（1993）、“GB”编码（1987）、“ZN”编码（1991）、“FZ”编码（1990）。这些编码体系在当时的技术背景下起到了一定的作用，但随着计算机技术的飞速发展，编码方案的不统一阻碍了蒙古文信息处理的国际化发展。因此，迫切需要制订蒙古文国际标准码。

UNICODE 是目前用来解决同一平台上处理多个文种的比较好的解决方案。2000 年 2 月，蒙古文国际标准码得到国际标准化组织的正式通过，并得到 UNICODE 技术委员会的认可，为蒙古

文字字符集和编码的统一制定了标准。蒙古文国际标准码包括传统蒙古文、托忒文、锡伯文和满文以及蒙、托、满三种文字用于转写藏文和梵文的阿礼嘎礼 (Ali Gali) 字母、标点符号、数字和控制符的蒙古文字符的编码集。托忒文、锡伯文、满文均用蒙古文字符。蒙古文国际标准码只对“名义字符”编码,对强制性合体字没有编码。传统蒙古文、托忒文、锡伯文和满文的大部分字母根据他们在词里的位置(词首、词中、词尾),有不同的变体,我们称它为变形显现字符。根据国际标准 ISO/IEC 10646 有关规定,在蒙古文国际标准码中,对蒙古文的变形显现字符集没有进行编码。甚至于同一个字母在传统蒙古文、托忒文、锡伯文和满文中都有时,只占用一个编码位置。这些现象和事实的存在对于蒙古文的信息处理带来许多麻烦。但是为了与国际接轨,我们必须用蒙古文国际标准码来开发新的蒙文信息处理系统。

蒙古文编码国际标准由“名义字符”和“显现字符”两部分组成。“名义字符”,是每一个蒙古文字母的多个变体中最具代表性的一个形式,它代表该字母进行编码;“显现字符”,是蒙古文的每一个字母的多个变体字形,是显现、输出用的字符。

三、MGWORD 系统的分析

本课题的要求是编制一个适合处理基于 UNICODE 编码的蒙古文文字编辑器,这个编辑器称为 MGWORD 系统。我们对 MGWORD 系统的需求作了深入的调研和分析,得到了系统的一些功能级用例模型图。

如图 1 所示, MGWORD 系统运行过程中的角色包括用户、内存、外存、打印机和显示器等。用户是系统的驱动力。内存、外存、打印机和显示器在系统的运行过程中与系统中的不同用例发生关系。这些实体共同组成了 MGWORD 系统运行的软硬件环境。

MGWORD 系统中有一个系统总体控制用例,负责整个系统的控制协调。包括三个主要的子控制用例,即界面控制、排版控制和打印控制。排版控制又包括键盘控制和鼠标控制,因为在文档排版过程中,相应的功能一般都支持鼠标和键盘两种硬设备的输入。

文档排版用例又包括子用例有文本排版用例、图像排版用例、表格排版用例和文本框排版用例等。文本排版用例包括剪贴板操作、行距、字间距操作、颜色、字体操作。剪贴板操作包括常用的复制、剪切和粘贴(目前还没有实现拖放操作);行距、字间距操作可以对文本的行间距进行调整。同样也支持字间距的设置;通过颜色设置可以改变背景色,可以设置选中字符的字体、大小、样式、颜色等,这项功能由字体操作用例来完成。图像排版包括剪贴板操作、缩放控制用例。缩放控制用例具体捕获缩放操作点,可以对图像由上、下、左、右四个方向进行非等比例缩放,从左上、左下、右上、右下四个方向进行等比例缩放,拖动左上角一定范围的小矩形可以整体移动图像。表格排版同样包括剪贴板操作,对于单元格中文本的字体设置,行间距和字间距设置要共用字体操作用例和行距、字间距操作用例来完成。表格操作中重要的操作要算边框控制了,边框竖线的左右单元格向左向右移动都有不同的逻辑要求,边框横线的

上下单元格向上向下移动也有不同的逻辑要求，还有选中单元格的拆分和合并，都有特定的拆分前提和合并条件，这些功能是边框控制用例完成的。文本框操作，实际上是表格单元格的一个变种，很多操作与表格相似，文本框支持横竖两种，两种显示也在缩放控制用例中实现。

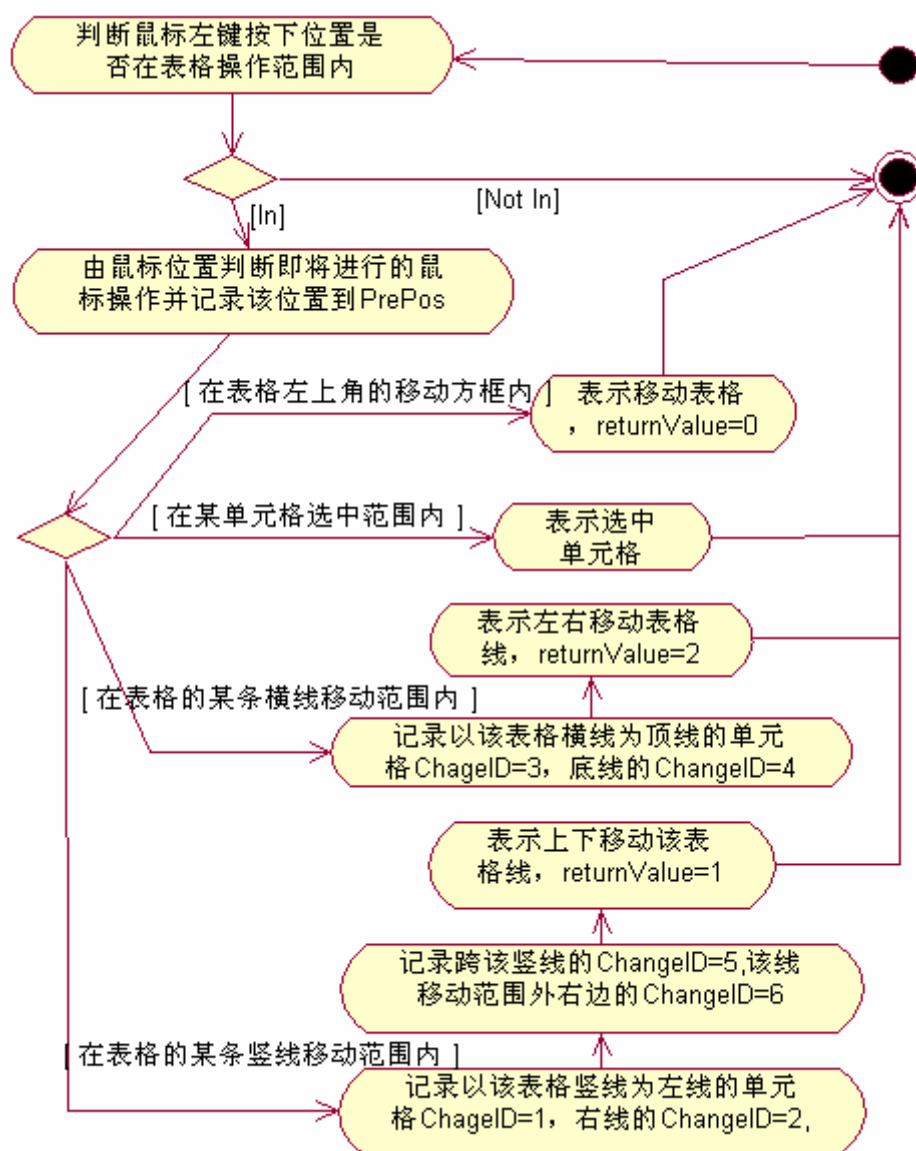


图 4：左键按下的表格操作活动图

MGWORD 系统中，文档的序列化具体包括文档的保存和文档的打开。文档保存用例负责按照一定格式保存文本、表格、图像等对象。文档打开对应于文档的保存。按照一定的格式，反向从文件中读取文本、表格、图像等对象。文件类型是 Mongolia Document，扩展名为 mdoc。文件头部分是文档总体信息，如版本号（这样可以控制多版本的兼容问题）、建立日期、作者等属性，接着是文档的数据部分，按照特定格式对文本、图像、表格等依次存放。

文档的打印是 MGWORD 系统的一个重要功能需求，是 MGWORD 系统中比较难的一块儿。打印包括打印预览和物理打印。打印预览功能是让用户在物理打印之前预览自己的排版效果，为了

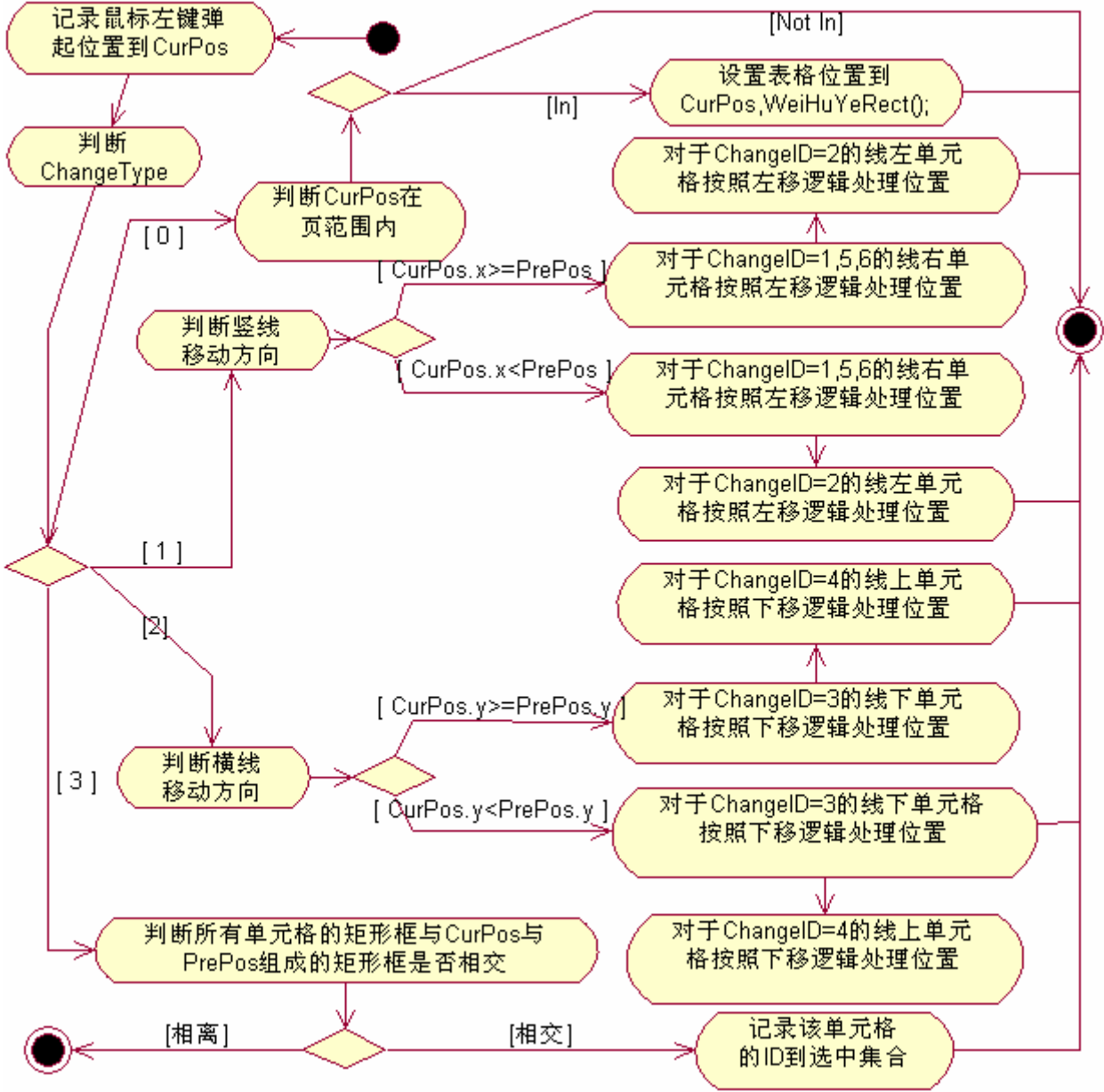


图 5：左键弹起的表格操作活动图

达到打印预览的目的，系统必须进行物理尺寸和逻辑尺寸的恰当转换。物理打印就是最终将排版文档输出到硬设备，如打印机。为了实现这项功能，必须对物理设备的相关参数进行访问和设置，尽管复杂，VC 中还是提供了方法。为了实现打印功能，对于文档中的各种对象都要做出不同的处理。分别由不同的用例完成。

四、MGWORD 系统的设计

MGWORD 系统是一个复杂的系统。在该系统中，处理的对象有文本、表格、图形图像和文本框等对象。对于每个对象都要考虑编辑、存取、打印等。涉及到的技术也很多，如剪贴板技术、避免屏闪技术、橡皮筋技术、打印技术、存取技术等等。同样使用了建模工具 Rational Rose，具体用 UML 中的活动图、协作图等表示。在设计过程中，光标控制设计和表格控制设计是 MGWORD 系统中较重要的两大块。每一个需求都要考虑到，因此，设计是复杂的，产生的图也就特别多。

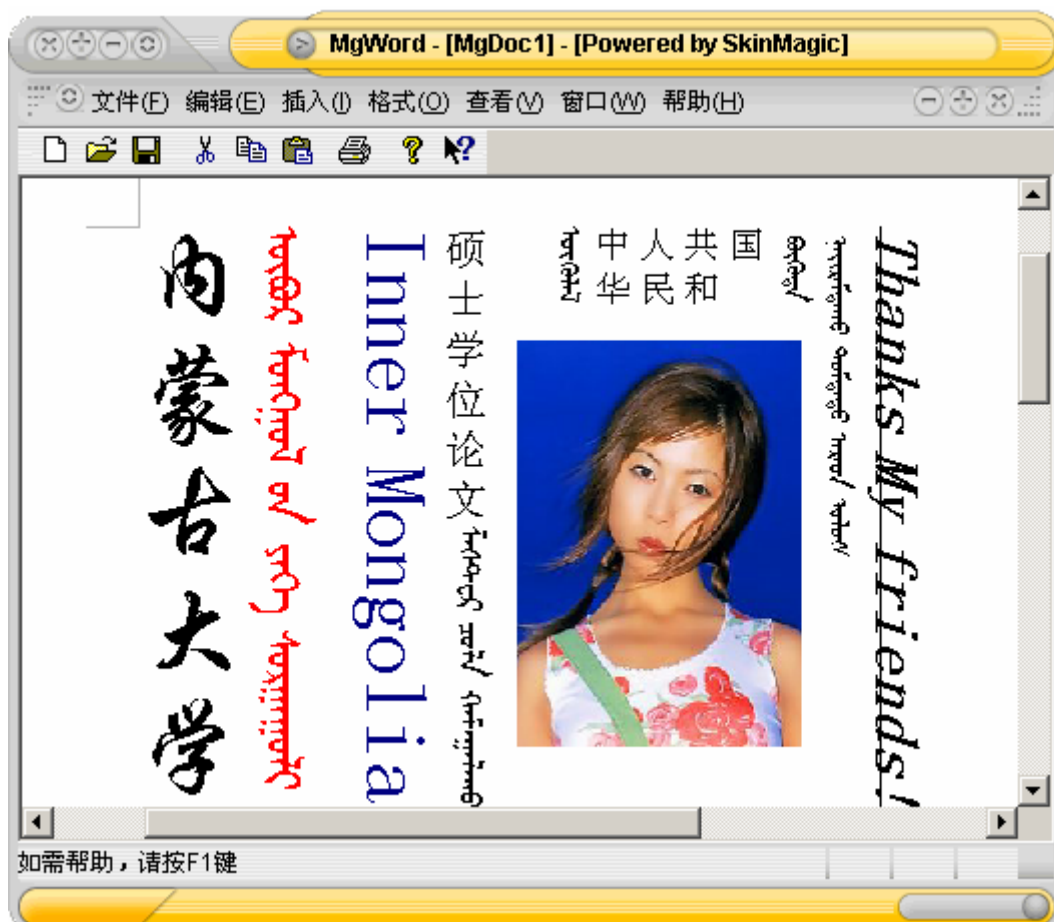


图 6：MGWORD 系统运行效果图一

（一）光标控制

蒙古文要求从上到下从左向右书写，已有的编辑控件像编辑框控件和 RichEdit 控件是不能用了。它们目前不支持蒙文的书写要求。MGWORD 系统中的排版工作必须也只能从底层开始，也就是说排版工作需要从光标控制（代表插入点）开始。光标控制是一项比较复杂的工作。MGWORD 系统要考虑字体改变、换行、换页、行间距、字间距、字符键、方向键、BackSpace 键、Delete 键、Home 键、End 键、Insert 键、PageUp 键、PageDown 键、鼠标左键单击和鼠标左键框选等等操作变化引起的光标位置的改变，这确实是一项复杂的工作。

用户按下一个键或者组合键，那么应该有一个模块负责接收此次按键，并且要根据键值引导系统作出不同的反应。按键首先分成两大类：功能键和字符键。如果是功能键，要区分为 VK_LEFT, VK_RIGHT, VK_END, VK_HOME, VK_INSERT, VK_PAGEUP, VK_PAGEDOWN, VK_SHIFT,

VK_CTRL 等。区分了这些键值，那么引导系统去做相应的处理。比如是 VK_LEFT，光标应该是左移一行，左移一行的过程中要碰到各种情况都要考虑到：已经是第一行了，要移到的行要比移前光标位置小，这些情况都要处理。另一类是字符键，图 2 和图 3 是用 Rational Rose 建立的键盘输入字符时的光标控制活动图，由图也可以窥出光标控制的复杂了。

当输入字符是回车键时，意味着回车字符在本行末显示，下一个输入字符要换行显示。但要考虑下一行是否超过当前显示区域的右边界，没超过则直接换行即可。返回值 `returnValue=3`，如果超过则要得到下一个显示区域 `Rect`，返回值 `returnValue=4`，置光标到 `Rect` 的左上角。

当输入字符不是回车键时，光标的重置面临三种情况：一种是输入字符不用换行，直接显示；另一种是换行不换显示区域；最后一种是换显示区域。这三种情况分别返回值 `returnValue=0`，`returnValue=1`，`returnValue=2`。上述各种情况分别作了不同的光标位置控制，各种情况的处理详见图 3 所示。

（二）表格控制

表格的边框控制是另一个比较复杂的控制。经过多次需求分析，参照 WPS、WORD、FrontPage 等优秀的排版软件中表格边框的控制方法，我们开发出了适合于蒙文竖排的表格。MGWORD 表格的边框控制与 WORD 中表格的边框控制正好相反，即 MGWORD 中表格中横线的控制和 WORD 中竖线的控制是相同的，而 MGWORD 中表格的竖线控制和 WORD 中横线的控制是相同的。这种区别正是由文字的书写方向决定的。

总体的思路如下：左键按下时判断是否对表格操作，如果是的话，判断出是什么操作，并且记录相关信息（如左键按下点，操作类型）；在鼠标拖动过程中，也就是鼠标左键按下移动鼠标的过程中，应用了橡皮筋技术。利用这项技术可以及时反映表格位置、表格线或框选的动态情况，给用户一个直观的印象；左键弹起的时候按照操作类型在特定的逻辑要求下对表格实际操作，修改数据结构。控制逻辑如图 4 和图 5 所示。

五、系统特点

MGWORD 编辑器的设计开发是蒙古文信息处理平台（MIPP）的一个重要部分。这个系统由于使用 VC++ 开发，而且主要使用了 VC++ 中的框架技术 MFC，因此，开发出的系统具有传统 Windows 的特征，是一个典型的 Win32 应用程序。由于 MGWORD 的特殊用户需求，再加上别具匠心的设计，使得 MGWORD 系统又有了许多新的特点。

（1）特制的启动界面：通过导入 MFC 内建组件 `CSplashWnd`，给 MGWORD 系统添加了启动界面，可以简要说明系统信息，如图 7 所示。

（2）可换的系统界面：MGWORD 系统引进了一个精心设计的具有界面切换功能的静态链接库——`SkinMagicLib`，使得 MGWORD 系统界面可以随时切换，MGWORD 系统已经带有 8 种不同分格的界面，用户还可以编辑添加新的个性皮肤。

（3）多文档界面：MGWORD 启动以后，可以同时在不同的文档之间切换。用户可以在启动一个系统的情况下，同时对多份蒙古文文档进行排版、存储、打印等操作，方便了用户。

(4) MGWORD 系统基于 UNICODE 码编写，因此该系统理论上能够在统一的平台上支持所有竖写的 UNICODE 编码字符。在实际的开发过程中，本系统通过测试能够很好地支持蒙文、英文和中文的混排，解决了蒙文编辑器中蒙文和其它文种混排的问题。



图 7：MGWORD 系统启动界面

(5) MGWORD 系统支持图像（目前版本仅限于 DIB 图像）的操作，包括图像的插入、尺寸、位置的改变、删除、存储、剪贴板操作、打印，并且实现了与文本、表格、文本框等其它对象的混排。对图像的相关处理主要封装在类 CDib 中。系统运行效果如图 6 所示。

(6) MGWORD 系统支持表格处理。在考察了 WPS、WORD、FrontPage 等多种编辑软件中对于表格的处理方式、特点后，作者利用自己编制的算法，实现了适合蒙古文竖写的表格类—Ctable。该类支持表格位置的改变、尺寸的改变、单元格的拆分、单元格的合并、表格中文字的处理、表格的剪贴板操作、表格的存储、表格的打印等。实现了表格与文本、图像等对象的混排。系统运行效果如图 8 所示。

(7) MGWORD 系统实现了文档中各种对象的剪切、复制、粘贴等剪贴板操作。这些操作封装在类 CClipboard 中。

(8) MGWORD 系统实现了文档的序列化。这些操作由 CMgWordDoc 类中的 Serialise 成员函数调用，由各对象对应类中相应的序列化成员函数完成存储与打开操作。

(9) MGWORD 系统实现了打印预览和打印的功能，能够对打印做相应的设置，如页类型、页边距、页方向等都可以由用户设置。



图 7：MGWORD 系统运行效果图二

六、技术总结

在具体实现 MGWORD 系统的过程中 ,用到了很多编程技术 ,总结了很多处理 UNICODE 码文字的经验。

(一) VC 中处理 UNICODE 码文字的技术

在产生应用程序框架代码之前 ,AppWizard 允许开发人员决定是否支持 UNICODE。为了成功编译支持 UNICODE 的 MFC 程序，必须使用 MFC 的 UNICODE 版本库。默认的 VC 安装并没有安装 MFC 的 UNICODE 版本库，在定制安装 Visual C++ 时必须选中安装。为了在 Visual C++ 6.0 中激活 UNICODE 标准，可以这样做：打开工程文件后，从主菜单中选择“Project | Settings”，打开工程设置对话框 => 然后选择“C/C++”标签 => 在“Preprocessor definitions”编辑框中添加 UNICODE 或者 _UNICODE 预处理宏指令。设置之后，在程序中对于字符串的处理应该有变化。

C 程序员一般是用 char 关键字象下面这样来声明一个字符串数组：

```
char str[100];
```

象下面这样声明函数原形：

```
void strcpy( char *out, char *in );
```

为了将上面的声明改成支持双字节的 UNICODE 字符集，可以用下面的方法：

```
wchar_t str[100]; 或者 void wcscpy( wchar_t *out, wchar_t *in );
```

(二) 避免闪烁的方法

窗口背景色往往与绘图内容反差很大，在短时间内背景色与显示图形的交替出现，使得显示窗口看起来在闪。为了避免闪烁，首先当然是去掉 MFC 提供的背景绘制过程。实现的方法很多，介绍两种：

(1) 可以在窗口形成时给窗口的注册类的背景刷赋 NULL；

(2) 要简单也可以重载 OnEraseBkgnd(CDC* pDC) 直接返回 TRUE。

这样背景没有了，结果图形显示的确不闪了，但是显示变得一团乱。要想解决这个问题，这必须运用双缓存的方法。双缓存就是除了在屏幕上有图形进行显示以外，在内存中也有图形在绘制。我们可以把要显示的图形先在内存中绘制好，然后再一次性的将内存中的图形按照一个点一个点地覆盖到屏幕上去（这个过程非常快，因为是非常规整的内存拷贝）。这样在内存中绘图时，随使用什么反差大的背景色进行清除都不会闪，因为看不见。当贴到屏幕上时，因为内存中最终的图形与屏幕显示图形差别很小（如果没有运动，当然就没有差别），这样看起来就不会闪。

(三) 提高绘图的效率

如果真要在一次重绘过程重画大量的图形，可想而知，这个过程是非常漫长的。实际上，在 OnDraw(CDC *pDC) 中绘制的图并不是所有都显示的，这是因为 MFC 本身为了提高重绘的效率设置了裁剪区。裁剪区的作用就是只有在这个区内的绘图过程才会真正有效，在区外的是无效的，即使在区外执行了绘图函数也是不会显示。因为多数情况下窗口重绘的产生大多是因为窗口部分被遮挡或者窗口有滚动发生，改变的区域并不是整个图形而只有一小部分，这一部分需要改变的就是 pDC 中的裁剪区了。因为显示（往内存或者显存都叫显示）比绘图过程的计算要花费时间长得多。有了裁剪区后显示的就只是应该显示的部分，大大提高了显示效率。但是这个裁剪区是 MFC 设置的，它已经为我们提高了显示效率。为提高刷新效率，我们可以只刷新用户区的一小部分，其余没有发生变化的我们可以不予刷新。窗口函数可以通过调用函数 InvalidateRect 显式地使用户区内的一个矩形无效。而且只有当窗口客户区的某一部分失效时，其窗口函数才会收到 WM_PAINT 消息。设备上下文先用 pDC->GetClipBox() 得到裁剪区，然后在绘图时判断你的图形是否在这个区内，如果在就画，不在就不画。按照这样处理，绘图效率就大大提高了。

MGWORD 系统是第一个从底层全面支持 UNICODE 编码的蒙文编辑器，为 MIPP 其它课题的开展奠定了基础，积累了经验，开发 MGWORD 系统是 MIPP 课题的重要子课题。经过测试，MGWORD 系统是一个具有实际使用价值的蒙文编辑器，对于一定篇幅的文档已经具有基本的排版功能、存储和打印功能。尽管 MGWORD 系统已具备了基本的排版功能，但是由于编辑器的开发具有相当的难度，再加上蒙文信息处理组内开发基于 UNICODE 码的蒙文编辑器编程经验不多，可以肯定 MGWORD 系统在许多方面存在不足甚至是 BUG，在将来的工作中，还需要不断地修改或完善，

包括：

- (1) 进一步确保系统在各种运行环境下的健壮性；
- (2) 继续优化程序的运行效率；
- (3) 优化数据结构，文件格式和某些调用频率较高的函数比如光标定位函数；
- (4) 进一步增强程序功能，开发出适合蒙文排版要求的文件模板，向导程序等。

参考文献

- [1]敖其尔.几种蒙文信息处理系统的设计与实现[A].内蒙古大学第三次蒙古学国际学术讨论会论文[C].1998年8月,呼和浩特.
- [2]敖其尔.蒙文信息处理的现状与展望[A].99朝鲜语信息处理国际学术会议论文[C].1999年8月,延边.
- [3]敖其尔.蒙古文信息处理平台(MIPP)研究[J].呼和浩特:内蒙古大学蒙古语研究中心网站,
- [4]敖其尔, 巩正, 廷方等.一种基于 UNICODE 编码的蒙文信息处理系统[A].Advances in Computation of Oriental Languages[C].北京:清华大学出版社, 2003.
- [5]白云莉.蒙古文信息处理平台—蒙古文、锡伯文和满文的文字输入法的研究[M].内蒙古大学硕士学位论文, 2003 年 5 月.
- [6]高光来,侯宏旭,李彦等.蒙汉混排文字处理系统的研究设计[J].内蒙古大学学报(自然科学版), 1994, (6).
- [7]那顺乌日图, 确精扎布等.蒙古文国际标准编码的构成原则[J].内蒙古大学学报(人文社会科学版), 1997, (6).
- [8]张胜男, 张霞, 黄有群, 赵双强.基于 UNICODE 码的中英文字符的区分方法[J].沈阳工业大学学报, 2000, (2).
- [9] Microsoft. MSDN Library Visual Studio6.0 . 1998

The editor for the mongolian script based on unicode

SONG Jian-bin¹, Ochir², MA Li³

(College Of Computer Science Inner Mongolia University, HUHEHOT 010021)

Abstract: The Editor for the Mongolian Script based on UNICODE is one of the question for study of Mongolian Information Processing Platform (MIPP), and which is very difficult and important in Mongolian Information Processing work. This paper first analysed the background and present conditions of studying and developing trends of the Editor for the Mongolian Script, then emphatically introduced the new Mongolian coding scheme and using trait based on UNICODE. On the base, this paper primarily narrated the implementation details and posed some technics and theoretical experience using to processing the Mongolian Script based on UNICODE.

Key words: the Mongolian Script ; UNICODE ; the Universal Standard Code for the Mongolian Script ; MFC ; the Editor for the Mongolian Script

收稿日期：2004-5-15；

基金项目：教育部人文社会科学研究重大项目“蒙古文信息处理平台（MIPP）的研究”（02JAZJD850003）；国家自然科学基金项目“基于WEB的蒙文图书信息管理系统”（60163003）

作者简介：1，宋建斌（1977-），男，汉族，内蒙古呼和浩特市人。内蒙古大学计算机学院研究生，主要研究蒙文信息处理和计算语言学；

2，敖其尔（1941-），男，蒙古族，内蒙古兴安盟人。内蒙古大学计算机学院教授，内蒙古大学蒙古学研究中心兼职教授，硕士生导师，主要研究蒙文信息处理和计算语言学；

3，马丽（1978-），女，汉族，内蒙古鄂尔多斯市人。内蒙古大学计算机学院研究生，主要研究信息处理和数据库技术。