Win32 平台下女书拼音输入法的设计与实现

王鹏　　孙茂松

智能技术与系统国家重点实验室

清华大学计算机科学与技术系，北京　100084

E-mail: {GundamNew, sunmaosong}@gmail.com

摘要： 女书是中国湖南省江永县流传的世界上唯一的女性专用文字，是人类文明的一朵奇葩，目前正在申请世界非物质文化遗产。女书研究的发展对女书数字化提出了迫切需要。女书输入法是女书数字化建设的基础性关键技术之一，能够大大加快女书资料的录入速度。本文介绍了 Win32 平台下的 IME 机制、该机制的主要接口函数，以及利用该机制实现女书输入法的原理。此原理可能推广到其他语言文字输入法。

关键字： 女书，拼音输入法，Win32，IME

"Female Script" Pinyin Input Method Based on Win32 System

Wang Peng　　Sun Maosong

State Key Laboratory on Intelligent Technology and Systems

Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China

E-mail: {GundamNew, sunmaosong}@gmail.com

Abstract: The "Female Script" in Jiangyong County, Hunan Province, is the only characters used solely by women, discovered in the world to date. "Female Script" Pinyin Input Method is an important and basal technology of "Female Script" digital engineering. In this paper, we introduce the Win32 IME and its main interface functions. Then we introduce how to implement "Female Script" Pinyin Input Method with Win32 IME. This method may be extended to input methods of other languages.

Keywords: Female Script, Pinyin input method, Win32, IME.

一、前言

女书是流传在中国湖南省江永县潇水流域的农家女专用文字，主要用于记事、自述。关于女书的起源目前还尚未定论，但应已有至少数百年的历史[1][2]。女书字为斜体，呈“多”字型，属汉字流传中的变异形态，是汉字楷体的变体（图 1）。女书是记录当地土话的音符字音节表音文字，一个字符可对应多个同音、近音词，女书文字与汉字之间有多对多映射关系，如表 1。由于使用环境等原因，不同人使用的女书存在个体用字差异，但在交际上有一定的共识度。据考察，女书基本字约有 300 多个，可以完整记录当地土话。由于流传范围很小，女书目前面临着濒危的状况，遇到了传承人群匮乏、原始资料流失等问题。2004 年 9 月 20 日最后一位使用女书的老人阳焕宜逝世，标志着女书原生态历史的结束。作为世界上唯一一种女性专用文字，女书是目前湖南省唯一申请世界非物质文化遗产项目。

近年来，清华大学赵丽明教授领导的女书研究团队和女书研究会在大量调研、整理工作的基础上，完成了《中国女书合集》、《女书用字比较》等重要学术著述[1][2][3]，在女书研究方面取得了丰硕的成果。女书的进一步传承和研究对女书的数字化建设工作提出了现实需求。赵丽明教授的团队已经整理了女书字符集，建立了字体图并制作了字体，同时也在申请 Unicode 国际标准编码。ISO/UCS 女书国际编码提案已经进入投票阶段。她的团队正着手进行女书电子词典和女书数据库的建设。

女书文字的数字化是女书研究的重要课题，经过数字化，使用现有的软件工具能够给女书研

究提供很大的便利，同时对女书文字进行规范化。在之前的女书研究中，女书的整理、录入主要通过手工完成，效率较低。女书输入法因此便成为女书数字化建设的重要一环。利用女书输入法，可以大大加快女书文字的录入速度，有利于建立女书电子词典、不断完备女书资料库等。应赵教授的要求，我们设计并初步实现了一个女书拼音输入法，其截图如图 2 所示。女书拼音输入法的主要功能是借助输入的汉字拼音，得到女书文字编码，比如输入 nv，能够通过用户选择输出“𛄲”字。由于输入方式与汉字拼音输入法有相似性，原则上可以参考汉字拼音输入法来实现。在输入法使用时，文字通过对应的汉语拼音来输入，一个女书字可对应多个汉语拼音。

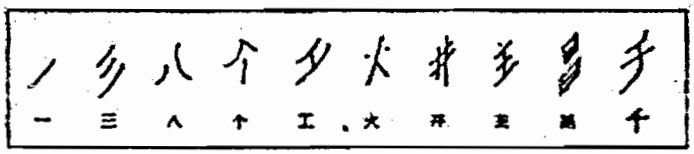

图 1 女书文字样例（上排为女书，下排为所参考的汉字）

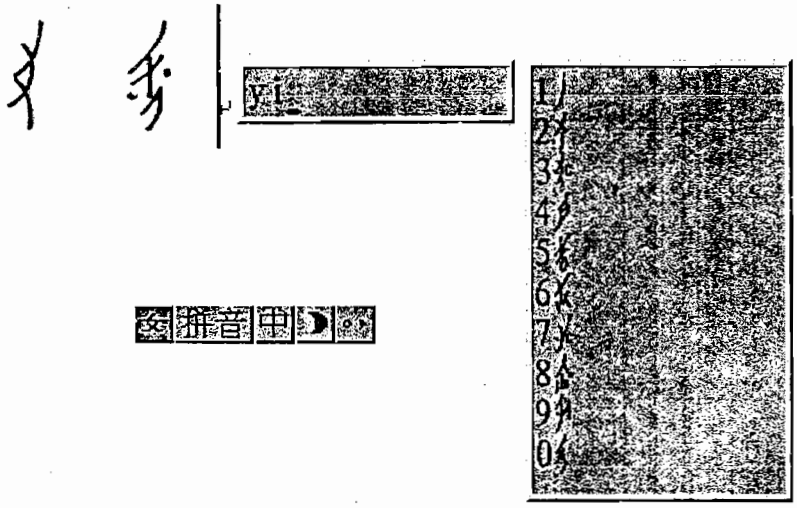

图 2 女书拼音输入法

女书字	汉语拼音	女书字	汉语拼音
𛄁	yi	𛄲	ren
𛄂	er	𛄳	ba
𛄃	ru	𛄴	shi
𛄄	ri	𛄵	piao
𛄅	liang	𛄶	pi
𛄆	qi	𛄷	yu
𛄇	cha	𛄸	pian
𛄈	cuo	𛄹	biao

表 1 女书文字与对应汉语拼音示例

本文主要描述了 Win32 平台下女书输入法的实现方法：根据整理好的女书字体和拼音码表，

实现 Win32 系统提供的接口函数。其中实现 Win32 系统的接口函数是完成输入法的核心内容，也是本文的主要工作。本文所完成的女书拼音输入法的基本机制，经过简单修改便可用于其他少数民族语言文字的拼音输入法。

二、Win32 平台下女书拼音输入法的实现

1、Win32 平台的 IME 机制介绍

Windows 系统中文输入法的基本原理是将键盘输入的 ascii 字符串通过一定规则转换成文字编码输出。针对当前的应用程序，Windows 系统的 user.exe 通过监控键盘输入，将键盘事件传输到输入法管理器（Input Method Manager，IMM）中。输入法管理器将键盘事件再传输到输入法（Input Method Editor，IME），输入法将键盘输入转换成汉字编码，再通过输入法管理器 IMM、Windows 系统的 user.exe 将汉字编码输出到应用程序。由于女书输入法的目的是完成汉语拼音到女书编码的转换，因此可以参考汉字输入法的机制。汉字输入法的关键是完成 IME。IME 程序的扩展名是.ime，实际上是一个动态链接库程序（DLL）[4][5]。根据 Windows 输入法机制的要求，在这个动态链接库程序中要实现 19 个接口函数。这些接口函数由 Windows 系统调用传递信息，包括：

1. BOOL ImeInquire(LPIMEINFO lpIMEInfo, LPTSTR lpszWndClass, LPCTSTR 或者 dword lpszData)

用于初始化 IME。

2. LRESULT CALLBACK CandWndProc(HWND hCandWnd,UINT uMsg,WPARAM wParam,LPARAM lParam)

Candidates Windows 注册函数。Candidates Windows 即用户用于选择候选汉字的 UI 窗口。

3. LRESULT CALLBACK CompWndProc(HWND hCompWnd,UINT uMsg,WPARAM wParam,LPARAM lParam)

Composition Windows 注册函数。Composition Windows 即显示用户输入编码的 UI 窗口。

4. LRESULT CALLBACK StatusWndProc(HWND hStatusWnd,UINT uMsg,WPARAM wParam,LPARAM lParam)

Status Windows 注册函数。Status Windows 即显示输入法状态的 UI 窗口。

5. DWORD IMEConversionList(HIMC hIMC, LPCTSTRlpSrc, LPCANDIDATELIST lpDst, DWORD dwBufLen, UINT uFlag)

用于将字符串转换成目标字符串。

6. LRESULT CALLBACK UIWndProc(HWND hUIWnd,UINT uMsg,WPARAM wParam,LPARAM lParam)

用户界面接口函数。

7. BOOL ImeConfigure(HKL hKL, HWND hWnd, DWORD dwMode, LPVOID lpData)

用于提供输入法设置的接口。

8. BOOL ImeDestroy(UINT uReserved)

用于关闭输入法。

9. LRESULT ImeEscape(HIMC hIMC, UINT uEscape, LPVOID lpData)

用于让应用程序访问输入法内部信息或功能的接口，这些功能通常无法用 IMM 函数调用实现。

10. BOOL ImeSetActiveContext(HIMC hIMC, BOOL fFlag)

用于在当前窗口激活或搁置输入法。

11. BOOL ImeProcessKey(HIMC hIMC, UINT uVirKey, DWORD lParam, CONST LPBYTE lpbKeyState)

用于处理所有程序传入的键盘敲击事件。

12. BOOL NotifyIME(HIMC hIMC, DWORD dwAction, DWORD dwIndex, DWORD dwValue)

用于修改输入法编辑器的状态。

13. BOOL ImeSelect(HIMC hIMC, BOOL fSelect)

用于打开或关闭输入法时对输入法进行初始化或恢复释放。

14. BOOL WINAPI ImeSetCompositionString(HIMC hIMC, DWORD dwIndex, LPCVOID lpComp, DWORD dwCompLen, LPCVOID lpRead, DWORD dwReadLen)

用于将编码窗口输入的编码保存，并发送编码完成的消息给系统。

15. UINT ImeToAsciiEx(UINT uVirKey, UINT uScanCode, CONST LPBYTE lpbKeyState, LPDWORD lpdwTransBuf, UINT fuState, HIMC hIMC)

用于将输入法上下文的编码转换成相应字符。

16. BOOL WINAPI ImeRegisterWord(LPCTSTR lpszReading, DWORD dwStyle, LPCTSTR lpszString)

用于向 IME 词典里添加新词。

17. BOOL WINAPI ImeUnregisterWord(LPCTSTR lpszReading, DWORD dwStyle, LPCTSTR lpszString)

用于把某个词从 IME 词典中去掉。

18. UINT WINAPI ImeGetRegisterWordStyle(UINT nItem, LPSTYLEBUF lpStyleBuf)

用于取得当前 IME 支持的词风格的列表。

19. UINT WINAPI ImeEnumRegisterWord(hKL, REGISTERWORDENUMPROC lpfnEnumProc, LPCTSTR lpszReading, DWORD dwStyle, LPCTSTR lpszString, LPVOID lpData)

用于列出给定条件的所有字符串。

这 19 个函数中，UI 窗口注册函数等 UI 相关函数构成了用户界面的主要部分，ImeProcessKey 和 ImeToAsciiEx 函数则构成了输入法编码转换部分的主要部分。

输入法的用户界面主要有 3 个：Status Windows、Composition Windows、Candidates Windows。Status Windows 主要显示输入法的状态，比如图标、中/英文、全/半角等；Composition Windows 显示用户当前的输入序列，比如图 2 中显示"yi"的窗口；Candidates Windows 显示转换后的候选序列，比如图 2 中显示 10 个女书文字的窗口。

为了实现这些函数，Win32 系统提供了一些数据结构用于 IME 和 IMM 的通信。比如：

CANDIDATEINFO：用于记录 Candidates Windows 的信息；

CANDIDATELIST：用于记录 Candidates Windows 内的数据；

INPUTCONTEXT：用于记录用户输入等信息。

Win32 系统还提供了一些 IMM 管理函数，用于编写完成 IME 接口函数。比如：

BOOL WINAPI ImmGenerateMessage(HIMC)

用于将输出编码发送到应用程序中；

LPINPUTCONTEXT WINAPI ImmLockIMC(HIMC);

用于获取当前应用程序的 INPUTCONTEXT 信息，并增加当前应用程序的计数器；

BOOL　WINAPI ImmUnlockIMC(HIMC);

用于释放当前应用程序的计数器；

DWORD WINAPI ImmGetIMCLockCount(HIMC);

用于查询当前应用程序的计数器；

以及用于 INPUTCONTEXT 结构成员处理的 ImmCreateIMCC、ImmDestroyIMCC、ImmLockIMCC、ImmUnlockIMCC、ImmReSizeIMCC、ImmGetIMCCSize、ImmGetIMCCLockCount 等等。

利用 Win32 系统提供的 IMM 函数和数据结构，完成 19 个 IME 接口函数，就是完成一个 Win32 系统 IME 机制输入法的主要任务。IME 机制实现的细节在此不详细描述。

2、女书拼音输入法的实现原理

由于女书文字的特殊性，目前的主流字体中并不包括女书文字。输入法需要女书字体的协助进行候选字显示；输出的文字编码也需要女书字体来显示。女书拼音输入法使用的字体是清华大学赵丽明教授的研究团队提供的女书字体。该字体包含 457 个女书文字，除了女书基本字，还包括一些由于使用习惯不同产生的变形。

完成输入法还需要汉字拼音与女书文字编码的对应关系，也就是码表。输入法使用的是赵教授的团队提供的拼音码表，共有 1179 条对应关系。这些对应关系中包含了全部女书字体中的女书字，，平均每个女书字有 2.58 个对应关系。女书字对应的拼音数量如表 2，可见大部分女书字对应 1、2 个拼音。由于女书文字的数量和表音性质，对应的汉语拼音数量相比汉字要多，甚至还有对应 15 个汉语拼音的女书字。

单个女书字对应汉语拼音数	女书字数	单个女书字对应汉语拼音数	女书字数
1	171	7	7
2	121	8	9
3	60	9	4
4	34	10	1
5	33	13	1
6	15	15	1

表 2 女书字对应汉语拼音数量汇总

女书拼音输入法除了输入法 IME 文件本身，还提供了两个外部文件。词库文件用于记录输入法码表、词库等信息。规则文件用于记录拼音规则，即输入法将哪些拼音当成单字处理，如该文件记录了“yi”，没有记录“yc”，则用户输入“yi”时输入法将“yi”当成单字的拼音输入并查找词库；用户输入“yc”时则会当成两个字的不完全拼音处理，如图 3 所示。这样做的好处是当拼音规则与汉字不完全相同时，可以通过修改.tab 文件方便地进行扩展，同时也为将输入法修改成其他语言文字使用提供了方便。

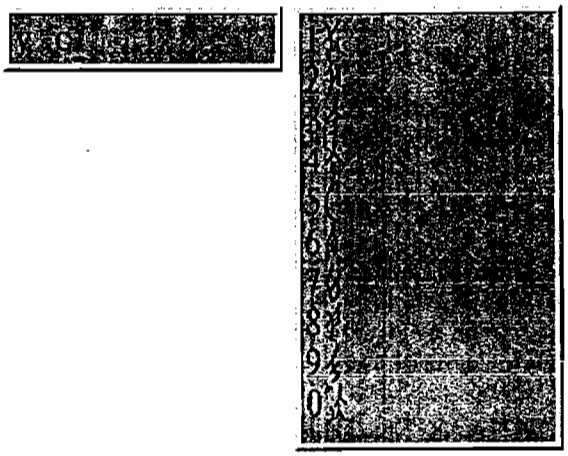

图 3 不符合拼音规则的输入序列处理样例

如前文所述，Win32 平台下汉字输入法主要要完成 19 个 IME 接口函数，用于与输入法管理器 IMM 通信。在输入法初始化阶段，输入法会初始化各个 UI 窗口，并读取.tab 文件中的拼音规则，用于对按键事件进行处理。IME 接口函数除了完成自己的初始化、显示、处理编码等工作，还要完成对 Win32 系统发出的事件、消息进行反应，这些事件包括休眠/激活输入法、输入法状态信息反馈、开始/结束编码事件、窗口管理等等。关于 UI 管理、信息反馈等细节处理，本文不做详细描述。

当用户输入拼音序列时，对于每一次按键，都会有一个按键事件传递到输入法程序。输入法用 ImeProcessKey 接口函数判断这次按键是否合法，即是否应该由输入法处理，如果不是则交由系统自身处理按键，如果是则系统会要求输入法进一步处理。

接口函数 ImeToAsciiEx 主要用来将用户输入的编码转换成输出的编码，可以说是输入法功能的核心。对用户输入的序列，计算其与拼音规则的最大匹配并进行分割。对尚未选择输出编码的第一个匹配，通过处理拼音的函数，根据码表查找对应的女书字显示在 Candidates Windows 供用户选择。用户选择后将更新 Composition Windows 中的显示或者将女书字/词的编码发送给输入法管理器 IMM。对于用户在输入过程中进行的 BackSpace、PageDown、PageUp 等按键，输入法也会进行相应的处理，更新输入序列的编码匹配和 Candidates Windows、Composition Windows 的显示。

3、女书拼音输入法的功能

女书拼音输入法的最主要功能是实现了用户输入的拼音序列到女书编码的转换。女书拼音输入法具备了普通中文拼音输入法包含的功能，如中英输入转换、半角全角转换、中英文标点转换等。用户在输入时可以输入多个字，输入法可以记录用户输入的“词”并记录在用户词库中。

三、结束语

本文主要介绍了 Win32 平台下的中文输入法实现机制，介绍了 19 个接口函数的大概功能，并描述了用 IME 机制实现女书输入法的主要原理。在更多了解女书语言文字的基础上，今后还可以给输入法添加女书文字与汉字的对照功能、智能组词功能等。

女书源于汉字，不同于使用字母的语言如英语，因此可以有更多样的输入方式。女书文字与

当地土话的发音对应，采用汉字的拼音方式输入未必就是最好的输入方式。在女书拼音输入法的基础上，还可以编写如女书部件输入法等多种输入方式，以更利于女书研究的展开。

女书拼音输入法基于汉字拼音输入法相似的原理，因此与汉字输入原理相似的其他语言文字也可以用类似的方式设计输入法。通过修改女书拼音输入法的规则文件和处理拼音的函数，可以较方便地实现其他文字的输入法。本文提供的这种利用类似机制实现其他语言输入法的思路，希望对其他少数民族语言输入法的研发能够有所帮助。

本输入法已于早前提交给了赵丽明教授的女书研究团队，相信已在女书编码的 Unicode 申请中发挥了应有的作用。

参考文献

[1] 赵丽明. 女书基本字与字源考. 2004, 女书的历史、现状与未来国际学术研讨会.

[2] 赵丽明, 等. 女书用字比较. 2006, 知识产权出版社.

[3] 赵丽明编著. 中国女书合集. 2005, 中华书局.

[4] David Iseminger 主编. Win32 开发人员参考库. 2001, 机械工业出版社.

[5] Microsoft Corp. Microsoft Win32 Multilingual IME Application Programming Interface for IME Development, Windows DDK[R].