第34卷

**Computer Engineering** 

September 2008

开发研究与设计技术。

文章编号: 1000-3428(2008)18-0260-03

文献标识码: A

中图分类号: TP391.1

# 基于 Web 页面的藏文在线输入技术

# 于洪志,何向真

(西北民族大学中国民族语言文字教育部重点实验室, 兰州 730030)

摘 要:Web 页面藏文在线输入技术,能够在浏览器中脱离本机输入法而进行藏文输入,实现藏文网络在线文字交互,为网络系统提供了 跨平台的藏文输入解决方案。阐述了基于 Web 页面藏文在线输入技术的工作原理及基本设计思想,介绍藏文在线输入法的组成、内码外码 设计原则和输入法流程,对藏文输入法进行系统分析,给出实现模型,论述浏览器内嵌藏文字体信息技术,达到在线、即时的藏文输入。 采用内嵌法和外挂法,实现藏文在线输入技术与主流网页编辑器的整合。

**关键词:** Web 页面; 藏文; 在线输入; 字体嵌入

# **Tibetan Online Input Technology Based on Web Page**

YU Hong-zhi, HE Xiang-zhen

(Ministry of Education Key Laboratory for Chinese Ethnic Minority Language, Northwest University for Nationalities, Lanzhou 730030)

[Abstract] By using the technologies of Tibetan online input on the Web, Tibetan characters can be input to a Web browser directly without using local input methods. It accomplishes the online character interaction of Tibetan language, and provides a cross-platform solution of Tibetan input method for a network system. This paper explains the basic design philosophies and the principles of online input technologies of Tibetan language which is based on the Web pages. It establishes the components of the online input method of Tibetan language, the principles to design internal and external codes and the flows of the input method. It applies a system analysis and provides a real model of Tibetan input method, and addresses the information technologies of Tibetan typeface which are embedded inside browsers to achieve online and real-time inputs of Tibetan language. It also implements the combination between online input technologies of Tibetan and main webpage editors by using embedded and external methods.

**Key words** Web page; Tibetan; online input; font embedded

Web 页面藏文在线输入技术是基于网页浏览器的藏文输 入平台。普通藏文用户在全球联网计算机上,不安装任何藏 文输入系统,能即时在线输入藏文。该输入技术解决了未安 装藏文输入系统的计算机输入藏文的问题,实现藏文网络在 线文字交互。本文以现代藏文拼写输入法为例,论述在线输 入现代藏文所有字符和常用梵音藏文技术。

# Web 页面藏文在线输入法的组成

Web 页面藏文在线输入设计利用 Web 页面编程技术,藏 文的输入输出都是在 Web 浏览器上来实现,它不归 Windows 的输入法管理器管理。Web 页面藏文在线输入技术和 Windows 本机藏文输入法原理的相同之处在于:都具有键盘 事件,编码字典等;差别在于:系统的键盘事件不通过输入 法管理器,而直接对应用户编码字典,输入法再翻译键盘事 件为对应藏文, 然后将键盘事件传给 Web 浏览器的文字编辑 框中,这样就实现了Web页面的藏文输入。

键盘事件监控用户按键,来预测用户可能需要的藏文字 符,并提供一张备选藏文字符列表以供用户从中选出所需藏 文字符,备选藏文字符列表就是用户编码字典。在默认情况 下,程序为用户提供一个可以键入并选择备选藏文字符的 IME窗口。通过使用IME变换功能提供的定制接口, Web输入 法提供类似Windows输入法中的输入法管理器(Input Method Manager, IMM)、函数和消息创建自己的输入法编辑(IME)窗 □<sup>[1]</sup>,

# 1.1 状态、编码和备选窗口

用户接口为程序提供状态、编码和备选窗口。状态窗口

指示程序被打开,并向用户提供设置转换模式的方法,要让 Web 藏文输入法打开,必须检测 Windows 本机输入法是否为 英文状态输入法,如果为非英文输入法,在程序输入框内启 用 Windows 本机输入法输入文本,而非 Web 在线输入法。在 用户输入文本时,编码窗口启动,依赖转换模式显示输入文 本或转换后的文本。备选窗口同编码窗口也一同出现,包含 待选的备选字符或编码窗口中的字符,用户可以在备选列表 中选择所需字符,然后返回至编码窗口,此时发送藏文字符 至文本输入窗口。用户可以在线录入所需藏文文本,直至录 入藏文字符串结束且窗口关闭。图1所示为系统与输入法的 关系。

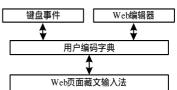


图 1 操作系统与 Web 页面藏文输入法的关系

# 1.2 IME 窗口类

" IME " 窗口类是预定义的定义标准 IME 窗口行为和外

基金项目: 国家 "863" 计划基金资助项目(2003AA115530); 国家社 会科学基金资助项目(02BYY038);教育部科研基金资助重点项目 (105172)

**作者简介:**于洪志(1947 - ),女,教授、博士生导师,主研方向:藏

文信息处理;何向真,硕士研究生

**收稿日期:**2007-10-30 E-mail: yhz@xbmu.cn 观系统全局类。程序用 IME 类创建自己的 IME 窗口。这就允许程序利用默认的 IME 窗口进程以及窗口位置布局。

#### 1.3 IME 消息

当某个键盘事件发生时,系统送出 IME 窗口消息到程序窗口程序来影响 IME 窗口。例如,当激活一个窗口,系统给文本输入窗口送出 IMEDOWN(EVENT)消息。

#### 1.4 录入字符串

录入字符串是录入窗口的当前文本。即 IME 转换到最终字符的文本。每个录入字符串由一个或多个 Clauses 组成,这里 Clauses 是 IME 能转换到最终字符的字符最小组合。用户在录入窗口输入文本时,IME 跟踪录入窗口的状态,包括属性信息、clause 信息、键入信息和光标位置。

#### 1.5 热键

热键使用户可以快速地改变 IME 输入模式或切换到英文 IME。程序为用户增加了几个很必要的热键,这些键可以帮助用户快速的来使用此输入法,但在不同的 BROWSER 中,键盘事件有相冲突的地方,此时针对不同的 BROWSER 而设置不同的热键。

# 2 藏文输入流程

# 2.1 Web 页面藏文输入法键盘布局

Web页面下藏文输入法键盘布局,以藏文书写规则为主,考虑键盘的常用习惯,在藏文键盘字母布局表中,合理设计了上加字、下加字、反字、基字及常用字符的位置<sup>[2]</sup>,如图 3 所示。

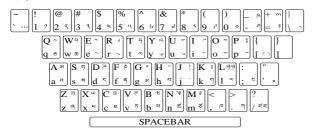


图 2 Web 页面藏文在线输入法键盘布局

# 2.2 藏文输入的基本结构

程序系统结构及藏文输入流程如图 3 所示。

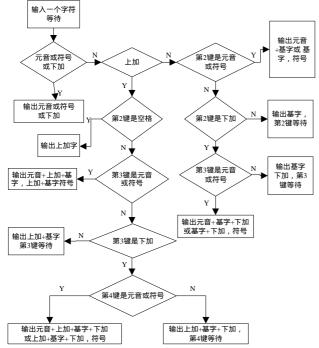


图 3 程序系统结构及藏文输入流程

藏文输入流程的程序有 2 个窗口:输入法转换接口和输入法用户接口。输入法转换接口由转换函数提供,这些函数被类似Windows系统输入法中的IMM调用。输入法用户接口由一组窗口提供,这些窗口接收消息并输出相应藏文<sup>[3]</sup>。

程序的多平台支持是由一组判断函数提供,程序运行首先判断用户浏览器类型,然后跳转至 Internet Explorer Browser 或 Netscaped Browser 相应的代码,实现多浏览器多平台的支持,如图 3 所示。

键盘事件监控用户的按键,来预料用户可能需要的字符,并提供一张备选字符列表以供用户从中选出所需字符,备选字符列表就是用户编码字典。

当用户敲击键盘时,程序作出判断,如果用户敲击的是元音或符号或下加字时,则程序直接输出字符至文本输入窗口;如果用户敲击的不是元音或符号或下加字时,程序再次进行判断,如果是上加字,即用户按下 q, w, e 按键时,并且第 2 键是空格的话,则输出上加字;如果第 2 键是元音或符号时,程序再次进行判断,依图中所示的那样,依次输出藏文或者符号。

# 3 藏文输入流程

藏文输入流程的程序有 2 个窗口:输入法转换接口和输入法用户接口。输入法转换接口由转换函数提供,这些函数被类似 Windows 系统输入法中的 IMM 调用。输入法用户接口由一组窗口提供,这些窗口接收消息并输出相应藏文。

键盘事件监控用户的按键,来预料用户可能需要的字符,并提供一张备选藏文字符列表以供用户从中选出所需藏文字符,备选藏文字符列表就是用户编码字典。

# 3.1 藏文输入字符的选取原则

Web 页面藏文输入法字符选取依据国家标准藏文编码字符集扩充集 A。

# 3.2 外码设计原则

根据藏文基字的使用频率,并参考了GB/T 17543-1998 《信息技术 藏文编码字符集(基本集)键盘字母数字区的布局》标准<sup>[4]</sup>,结合藏文的书写特点和习惯,设计了藏文键盘布局,将现代藏文和梵音藏文分布在一个键盘表中,有效地离散了重码。

#### 3.3 用户编码字典

考虑到程序的执行速度、后续维护、重码率等方面,程序采用独立的用户编码字典文件,并以 JS 文件存放。为程序的维护带来方便,并且可以自由扩充。

#### 3.4 藏文输入法系统分析

藏文属于拼音文字,每个词都是由字符上下左右拼写叠加组成。

在设计 Web 藏文输入法时采用键盘映射方式,即将每个按键按下后传送给浏览器的输入控件,程序扫描码映射为需要输出的藏语文字符编码。这主要是针对世界上基于字母的语言使用的。在键盘抽象中,需要给出一个转换表(符号表),并有状态相关的转换规则,而且系统需要确定状态的表示方法;此外还需要定义特殊按键的控制作用。输入法中将英文与藏文分别设计成一组键,用一个修饰键在组之间切换,同时其他如汉字输入法仍然使用 Windows 本机输入法的方式。从而使 Web 编辑器可以同时输入藏文、汉字与英文。设定切换藏汉英文输入的快捷方式,可以达到顺利输入所需文字的 接藏汉英文输入的快捷方式,可以达到顺利输入所需文字的目的。当系统输入法默认是汉字输入法时,此时,Web 编辑器中则为汉字输入,当系统输入法默认为英文输入状态时,

就可以在 Web 编辑器中输入藏文。英文、汉字的输入和藏文的输入可以随时切换。

#### 3.5 实现模型

藏文在线输入由于是在浏览器中实现,因此采用 Browser/Server 模型来实现。

#### 3.6 程序描述

#### 3.6.1 源代码模块组成

藏文在线输入法由 4 大模块组成:(1)TIBETINPUT,运行在用户的第一视线,主要表现为藏文输入法窗口模块。(2)TIBET\_MAIN,定义按键功能、搜索藏文用户字典等。(3)TIBET\_CODING ,藏 文 用 户 编 码 字 典 。 (4)TIBET\_LAYOUT,主要提供用户帮助,是一个藏文键盘布局表。3.6.2 初始化模块

向程序说明针对不同浏览器某些特殊按键的初始化,以 及该输入法将用到的所有按键排列。

#### 3.6.3 浏览器检测

藏文在线输入法最终是要运行在浏览器中的,如今浏览器种类众多,许多浏览器都不是基于 Web 标准的浏览器,因此,要保证在几款知名浏览器中程序能正确运行,必须给不同的浏览器设定不同的键值及属性。

在输入藏文前,采用一组判断函数来判断用户浏览器类型,类型确定后,跳转至 Internet Explorer Browser 或 Netscaped Browser 相应的代码,实现多浏览器多平台的支持。3.6.4 自定义按键

藏文输入法键盘布局采用 26 个小写字母 + 26 个大写字母 + 标点符号来定义码表。下面详细列出所定义的按键功能及方法:

- (1)字母按键。程序根据藏文外码设计原则,为避免重码和以后扩充梵文输入的需要,将大写字母和小写字母区分,统一安排藏文字符键位,所以字母按键是最为基本的码表组合键。
- (2)数字按键。将键盘上的 0~9 数字键安排为藏文的数字键,对应的藏文为: 233~450~。
- (3)符号按键。符号按键功能和字母按键功能类似,均安排了藏文编码。将这些符号按键排列在大写字母之后,小写字母之前。
- (4)快捷键设计。主要是设计一些特殊按键,利于快速编辑。包括:Backspace, Tab等。

#### 3.6.5 浏览器内嵌藏文字体信息

为了使用户能很好地享受在线藏文输入法带来的便利,必须包含字体信息,否则无法显示藏文字体,如果让用户自己安装字体的话,就体现不出在线输入法的便捷特点,只有将藏文字体信息嵌入到浏览器中,才能真正实现在线的、即时的输入法。

本输入法设计支持多平台,即支持多种浏览器。在 Internet Explorer 中,脱字体显示及藏文字体嵌入的方法是利 用微软提供的 WEFT 工具,生成包含字体信息的 EOT 文件。

前面提到藏文在线输入技术是基于 B/S 模式的,这时把相应的 EOT 文件也拷贝到指定的 Web 服务器的目录下面,在浏览器中就可以显示出藏文字体信息。

而在Linux等非Windows系统中,网页是在Mozilla或者Netscaped中浏览的,上述的方法只是针对IE浏览器,如果要实现类似IE中脱字体的效果,必须用到Bitstream WebFont Wizard工具来实现<sup>[5]</sup>。

#### 3.7 输入法用户界面

本输入法用户界面包括状态窗口、编码输入窗口、藏文 候选字窗口及输入法窗口。

- (1)输入法窗口即文本输入窗口,用来接收经用户按键转换的藏文字符,同时也接收经 Windows 或 Linux 本机输入法输入的字符。
- (2)候选字列表窗口(候选窗口),根据藏文的书写顺序,动态提示当前藏文。
- (3)编码输入窗口即字母组合窗口,其形式基于转换模式,同时显示已经输入的字母或者是已经转换了的字符。它将用户的输入操作通知给编码字典,并且处理由输入窗口为了响应用户行为而发送给它的控制消息。
- (4)状态窗口显示当前的输入法状态,例如是藏文输入还 是英文输入、以及控制自动结束符等信息。

# 4 与网页编辑器的整合方法

藏文网页中只要有输入控件的地方,均可以用 Web 页面藏文在线输入法解决输入藏文的问题,不论是<input>还是<textarea>控件,网页编辑器也是采用类似的输入框来编辑文字信息,输入法和编辑器的整合有下面2种方法。

#### 4.1 内嵌法

可以将 Web 页面藏文在线输入法嵌入当前流行的即时网页编辑器中,例如:嵌入 InnovaStudio WYSIWYG Editor 中。嵌入的方法应根据各个编辑器的具体情况而定,因为如今的编辑器输入框可以归纳为以下情况:

- (1)输入框为一个内嵌的空白网页;
- (2)输入框和 Web 页面藏文输入法的输入框相似,均为 <textarea>标签控件。

对于第(1)种情况,采取修改其 JS 代码,和输入法的 JS 文件结合;而第(2)种情况较为简单,只是将输入法的 <textarea>标签属性进行移植就可以。图 4 为采用内嵌法和网页编辑器整合后的效果。

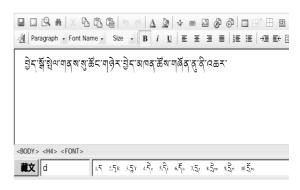


图 4 整合后的效果

# 4.2 外挂法

外挂法是不直接改变编辑器的输入框参数,在输入藏文文字信息时,只是将 Web 页面藏文输入法调用为弹出页面,当藏文输入结束时,将信息提交到需要输入的编辑器中,此方法简单实用,和编辑器完美结合,开发效率高。图 5 为采用外挂法和网页编辑器 FCKeditor 的整合效果,该编辑器提供了用户扩展插件功能,整合后,输入法就是该编辑器的一个 Plug-Ins。用户只需点击"藏文输入法"按钮弹出输入法页面,在输入法页面中编辑好藏文后,单击"提交"按钮即可将藏文提交到上级 FCKeditor 编辑器中,达到编辑藏文的目的。