



(12) 发明专利申请

(10) 申请公布号 CN 103984420 A

(43) 申请公布日 2014. 08. 13

(21) 申请号 201410142863. 2

(22) 申请日 2014. 04. 10

(71) 申请人 中央民族大学

地址 100081 北京市海淀区中关村南大街
27 号

(72) 发明人 程卫军 洛桑旦增

(74) 专利代理机构 北京君尚知识产权代理事务
所(普通合伙) 11200

代理人 余长江

(51) Int. Cl.

G06F 3/023(2006. 01)

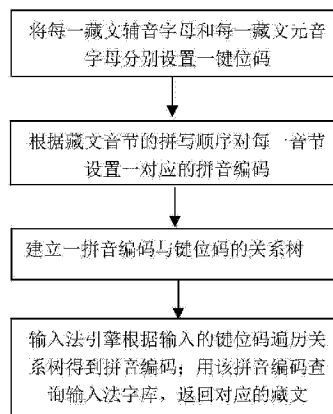
权利要求书1页 说明书5页 附图2页

(54) 发明名称

一种基于拼音的藏文智能输入法

(57) 摘要

本发明公开了一种基于拼音的藏文智能输入法。本方法为:1) 将每一藏文辅音字母和每一藏文元音字母分别设置一键位码;2) 根据藏文音节的拼写顺序对每一音节设置一对应的拼音编码,并将其保存到一输入法字库中;3) 建立一拼音编码与键位码的关系树;4) 基于所述输入法字库建立一输入法引擎,所述输入法引擎根据输入的键位码遍历所述关系树,得到对应的拼音编码;然后根据该拼音编码查询所述输入法字库,返回对应的藏文。与现有技术相比,本发明具有重码低,易于实现,易于建立和扩充词库等特点,且符合藏文书写的自然思维方式,让藏文的输入更方便、快捷、灵活。



1. 一种基于拼音的藏文智能输入法,其步骤为:
 - 1) 将每一藏文辅音字母和每一藏文元音字母分别设置一键位码;
 - 2) 根据藏文音节的拼写顺序对每一音节设置一对应的拼音编码,并将其保存到一输入法字库中;
 - 3) 建立一拼音编码与键位码的关系树;
 - 4) 基于所述输入法字库建立一输入法引擎,所述输入法引擎根据输入的键位码遍历所述关系树,得到对应的拼音编码;然后根据该拼音编码查询所述输入法字库,返回对应的藏文。
2. 如权利要求 1 所述的方法,其特征在于所述对每一音节设置一对应的拼音的方法为:对于单字符藏文音节,设置其拼音为单字符藏文音节本身;对于无上下叠加的多字符藏文音节,设置其拼音为多字符藏文音节本身;对于有上下叠加的多字符藏文音节,设置其拼音为单字符藏文音节本身。
3. 如权利要求 2 所述的方法,其特征在于同一拼音编码对应一个或多个音节。
4. 如权利要求 1 或 2 或 3 所述的方法,其特征在于所述输入法引擎根据拼音编码查找匹配的拼音,将所有与这个拼音匹配或以这个拼音为开头的藏文显示到输入法的候选字区域中,并按词频顺序排序。
5. 如权利要求 1 所述的方法,其特征在于在移动设备上采用全键盘模式或者九宫格模式作为藏文辅音字母和元音字母的输入界面。
6. 如权利要求 1 所述的方法,其特征在于在 PC 上采用喜马拉雅输入法的键位方式作为藏文辅音字母和元音字母的输入界面。

一种基于拼音的藏文智能输入法

技术领域

[0001] 本发明涉及一种输入法,尤其涉及一种基于拼音的藏文智能输入法。

背景技术

[0002] 藏文自创制以来,无论作为民族文化遗产的主要载体,还是现在作为藏区传播科技知识的主要工具,甚至作为信息化社会中一个民族的主要标示符,其独特的人类文化价值和在广大藏区发挥的巨大作用是不可估量的。

[0003] 藏文跨入信息时代数十年来,在计算机藏文信息处理的各方面得到了长足的发展,也取得了不少的成就,从藏文的打字录入到藏文排版处理、藏文电子邮件的收发、藏文网站建设、藏文应用软件的开发、藏文课件的制作等等。

[0004] 藏文是拼音文字,具有横向书写性结构同时还具有纵向书写性结构,它的词组或句子是由一个个音节(或称为字)组成的,一个音节对应一个音,一个音节又有若干个藏文字母组成,粗看起来它和英文很相似,如 ;但是对于一个藏文音节来说,它又要从一个基字开始,由上加字、下加字、前加字、后加字叠加组合而成,因而又有了平面字的特点。藏文的音节的结构是以一个字母为核心,核心的字母叫“基字”,其余的字母在此基础上前后附加和上下叠写,组合成一个完整的字表结构,且每个字母的称谓都根据加在基字的部位而得名。如图 1 所示。

[0005] 藏文 30 个辅音字母均可作基字,但是,可做前、后、上、下加字的字母在文法中都是有规定的,而且数量有限。

[0006] 藏文读音也是基字辅音为中心的,一个音节只有一个元音(元音 a 可省略),因此一个音节对应一个音,在藏文拼读时,从最左边的辅音字母开始,其顺序是 1) 前加字,2) 上加字,3) 基字,4) 下加字,5) 元音,6) 后加字,7) 再后加字。

[0007] 藏文的书写是以音节为单位,从左到右横写,音节之间用小点隔开,如  音节的书写顺序与拼读的顺序一致,大部分输入法也是按照此顺序来输入藏文编码的,但是具体输入理论较为复杂,而且有很大的弊端,因为当书写上加字或下加字时某些特殊的字母需变形,所以藏文在国际编码 Unicode 中一共定义了 211 个字符,其中包含普通字符、叠加字符、数字符号、天文历算符号等。再利用 Opentype 字库标记特征,将这些普通字符与叠加字符组合而成,此功能是字体库本身具备的而非输入法,输入法将根据用户的输入形成字符编码,字体库根据此编码通过字库标记特征从而显示藏文音节。

[0008] 目前,藏文的录入速度与汉文等其它部分语言的录入速度相比仍具有明显差距,尤其在移动终端上,主要原因缺乏高效的、智能的输入法。在现有输入法当中,仅有少数具有词组输入功能,微软 Himalaya 输入法等国外藏文输入法都不支持词组或智能输入法,而国内的具有词组输入的班智达输入法采用了基字加后加字的词组编码方案,但具有不自然,难记难用且重码多,用户可随意输入字符任意组合,违反藏文的文法等缺点。因此急需开发一种易用、自然、通用性强且重码低的智能输入方案,以提高藏文的录入速度。

发明内容

[0009] 为了克服现有技术中存在的技术问题,本发明的目的在于提供一种基于拼音搜索的藏文输入法,根据藏文的文字结构、发音特征及拼写方法,本发明将一些字母作为拼音字符来标识特定音节,不考虑音节的叠加过程表示,实现拼音输入,因此本发明提出的拼音输入法就是以它为基础的。具体为利用好藏文的拼读规则,将藏文音节与对应关系的拼音存放到字库中,通过输入法形成拼音编码,再根据输入法引擎返回目标字集。

[0010] 因此本发明具有重码低,易于实现,易于建立和扩充词库等特点,且符合藏文书写的自然思维方式,容易理解及使用。

[0011] 本发明目的通过以下技术方案得以实现:

[0012] 一种基于拼音的藏文智能输入法,其步骤为:

[0013] 1) 将每一藏文辅音字母和每一藏文元音字母分别设置一键位码;

[0014] 2) 根据藏文音节的拼写顺序对每一音节设置一对应的拼音编码,并将其保存到一个输入法字库中;

[0015] 3) 建立一拼音编码与键位码的关系树;

[0016] 4) 基于所述输入法字库建立一输入法引擎,所述输入法引擎根据输入的键位码遍历所述关系树,得到对应的拼音编码;然后根据该拼音编码查询所述输入法字库,返回对应的藏文。

[0017] 进一步的,所述对每一音节设置一对应的拼音的方法为:对于单字符藏文音节,设置其拼音为单字符藏文音节本身;对于无上下叠加的多字符藏文音节,设置其拼音为多字符藏文音节本身;对于有上下叠加的多字符藏文音节,设置其拼音为单字符藏文音节本身。

[0018] 进一步的,同一拼音编码对应一个或多个音节。

[0019] 进一步的,所述输入法引擎根据拼音编码查找匹配的拼音,将所有与这个拼音匹配或以这个拼音为开头的藏文显示到输入法的候选字区域中,并按词频顺序排序。

[0020] 进一步的,在移动设备上采用全键盘模式或者九宫格模式作为藏文辅音字母和元音字母的输入界面。

[0021] 进一步的,在 PC 上采用喜马拉雅输入法的键位方式作为藏文辅音字母和元音字母的输入界面。

[0022] 本发明的一种基于拼音的藏文智能输入法流程图如图 2 所示,其具体步骤如下:

[0023] 一、规定 30 个藏文辅音字母和 4 个元音字母来组成拼音字符,并且根据藏文拼写顺序组合形成每音节对应的拼音。

[0024] 如表 1 为藏文的辅音字母:

[0025] 表 1 为藏文辅音字母表

[0026]

ཀ	ཁ	ག	ང	ཅ	ཆ	ཇ	ཉ	ཏ	ཐ	ད	ན	པ	ཕ	བ
མ	ཙ	ཛ	ཱ	ེ	ཱ	ེ	ཱ	ེ	ཱ	ེ	ཱ	ེ	ཱ	ེ

[0027] 如表 2 为藏文的元音字母:

[0028] 表 2 为藏文元音字母表

[0029]

འི	འི	འི	འི
----	----	----	----

[0030] 根据藏文的拼写顺序规定每个音节对应的拼音,其具体如下:

[0031] 1. 单字符音节的拼音为它本身。如表 3:

[0032] 表 3 为单字符音节

[0033]

藏文音节	拼写顺序	规定拼音
ཀ	ཀ	ཀ
ཁ	ཁ	ཁ

[0034] 2. 多字符音节并无上下叠加(除了元音字符叠加)时拼音为它本身。如表 4:

[0035] 表 4 为无上下叠加的多字符音节

[0036]

藏文音节	拼写顺序	规定拼音
གང	ག ང	གང
བསམ	བ ས མ	བསམ
གཏོང	ག ཏ འ ང	གཏོང
སེམས	ས ེ མ ས	སེམས
སུམ	ས ུ མ	སུམ

[0037] 3. 多字符音节有叠加时拼音由该字的拼写顺序决定。如下表 5:

[0038] 表 5 为有上下叠加的多字符音节

[0039]

[0040]

藏文音节	拼写顺序	规定拼音
སྒྲོལ	ས ག འ རྒ རྒ	སྒྲོལ
སྒྲོལ	ས ག འ རྒ རྒ	སྒྲོལ
བྱུང	བ འ རྒ རྒ	བྱུང
བསྒྲོལ	བ འ རྒ རྒ འ རྒ རྒ	བསྒྲོལ

[0041] 根据上述的三条规则我们基本能够确定每个音节对应拼音,同时也发现同一个拼音对应多个音节。

[0042] 二、创建字库并字库中添加特定关系的拼音与对应的音节。

[0043] 目标音节与拼音存储在字库中,从数据结构上两者存储结构存在两种关系,一种是,一对一关系,即一串拼音字符只表示一个音节,另一种是一对多关系,一串拼音表示多个音节,由于藏文的特殊性,一个拼音最多对应三个音节。

[0044] 三、输入法引擎

[0045] 输入法引擎是实现智能输入的核心部分,它为输入法提供一个适配器,即接收用户输入的码值,将在适配器中找到该码值所对应的拼音码,再将拼音码给字库进行搜索,将搜索的结果返回给用户,从而完成输入。

[0046] 与现有输入法相比,本发明的有益效果:

[0047] 1) 通用性强

[0048] 本发明的字符编码方案基于国际标准的 Unicode 编码,在不同的设备上易于实现。

[0049] 2) 键位排版灵活

[0050] 由于本发明的拼音字符数量少,只需要 34 个键位即可,这样我们在 PC 机上完全可以采用当前比较普遍使用的喜马拉雅输入法的键位方式,而在手机、平板等终端上很巧妙地实现全键盘模式或者九宫格模式,现有手机藏文输入法还没有九宫格键盘模式。

[0051] 3) 输入速率高重码少

[0052] 采用了基于搜索的拼音输入,且拼音与字的一对多关系量少,不仅重码少,而且能准确确定首选字,提高了输入速度。如词组 ~~ལྷོ་ཁོང་གི་མཉམ་འཇུག་པུ་སྒྲིག་པོ་~~ 普通输入法需要 11 键,而本输入法仅需 8 键,甚至更少。对于较短的字,首次输入就能准确的确定首选字。

[0053] 4) 与藏文的书写思维一致,好学好用

[0054] 本发明是基于拼音的输入法,而且藏文本身也是拼音文字,拼音的书写完全于藏文的书写思维吻合,只要有藏文的书写基础既能轻松学会本输入法。

[0055] 5) 易于建立和扩充词库

[0056] 本发明中定义了基本系统字库,它包含所有藏文音节,我们将以它为基础建立用户词库,该词库不仅需要良好的数据结构,还需要良好的扩充性、兼容性,采用本发明的编码方案,能将网络中各类藏文词汇添入词库中,使得词库的扩充方便实现。

[0057] 6) 实现词频记录,输入更快

[0058] 与当前主流的中文输入法一样,我们的输入法也具备词频的自动统计与调整,实现适应用户特点的输入方式,还可以实现用户词组的记忆功能,让藏文的输入更方便、快捷、灵活。

附图说明

[0059] 下面将结合附图,对本发明作进一步的详细描述:

[0060] 图 1 为藏文字的结构解析示意图;

[0061] 图 2 为本发明方法流程图;

[0062] 图 3 为本发明一具体实施例的输入法原理框图。

具体实施方式

[0063] 结合附图,下面对本发明进一步说明,通过以下步骤实施本基于拼音的智能输入方案:

[0064] 1. 按照已给定的拼音规则,建立输入法字库。一种高效的输入法都必然对应着其特定结构的字库,在本发明中,在编码上采用夸平台的 Unicode 编码,按表 6 所示来设计输入法字库的结构,字库由编码(ID)、音节(Vlaue)、拼音码(PinyinCode)以及频率(Frequency)四部分组成。

[0065] 表 6 为输入法字库

[0066]

ID	Vlaue	PinyinCode	Frequency
1	ཐུགས་	ཐུགས་ཐུགས་	1
2	མེས་	མེས་མེས་	3
3	དབང་	དབང་དབང་	1
...	...		

[0067] 字库的建立按以下步骤进行：

[0068] a. 字库中音节材料的收集，查阅相关的词汇表、词典等材料，由于所有拼音字符都是本发明特定的，必须由人工获取或编写程序录入拼音。

[0069] b. 将获取的拼音与音节归并到输入法字库中，可设计专门的程序模块负责此项工作。

[0070] 2. 建立基于字库的输入法引擎

[0071] 基于字库的输入法引擎主要是为用户与字库间提供一个适配器，如图 3 所示。具有匹配用户输入的键位码与对应的拼音值、输出拼音值并查询字库、返回字库查询的结果等的功能。

[0072] 输入法中定义了一颗所有拼音值与键位码的关系树，适配器将遍历这棵树得到与用户当前输入键位码所匹配的拼音值，根据所得到的拼音值搜索字库，并返回结果。当用户开始输入时，所有与这个拼音匹配或以这个拼音为开头的字显示到输入法的候选字区域中，并按词频顺序排序，此时用户可以在候选字区域选择目标字，也可以完整输入拼音来选字。根据藏文的文法输入法还可以自动添加需要的结尾符，藏文中有四种结尾符，并由相应的文法规定。

[0073] 藏文输入过程举例如下：

[0074] a. 假设键盘布局为全键盘，拟输入 བསྟན་འཛིན་ 字，其对应拼音为：བསྟན་འཛིན་ 那么当输入第一个字符 བ 时输入法将所有以 བ 为开头的拼音的字返回给用户界面，如 (བས) (བལ) (བེ) (བད) (བུ) (བཿ) 等字，当用户输入第二个字符 (འ) 时，输入法将以 (བས) 开头的拼音的字返回给用户，如 (བསྟ) (བསལ) (བསུ) (བསྟད) (བསེ) (བསཿ) 等，以此类推，进行不断的筛选，最后得到要输入的字。

[0075] b. 在本发明中拼音与音节(字)的关系主要以一对一关系存在，而且搜索的过程来看，发现并不用输完整的拼音才得到结果，而在输入过程中输入法引擎将最优的结果返回给用户，我们只要从中选择就可以了。

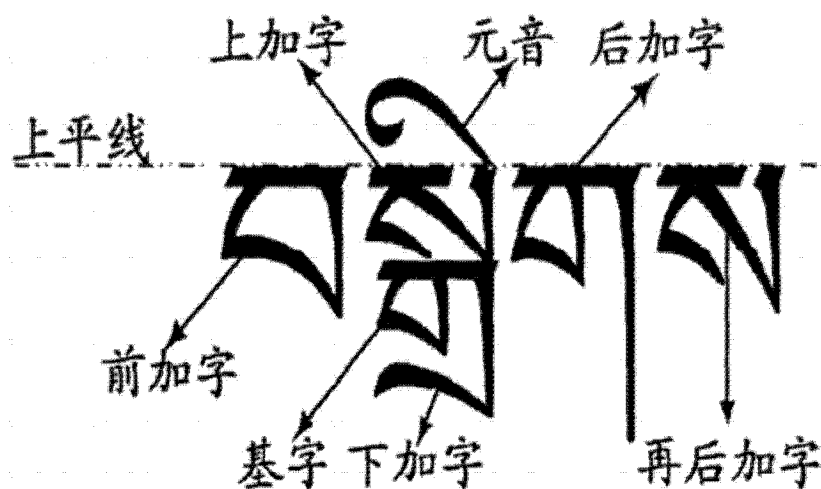


图 1

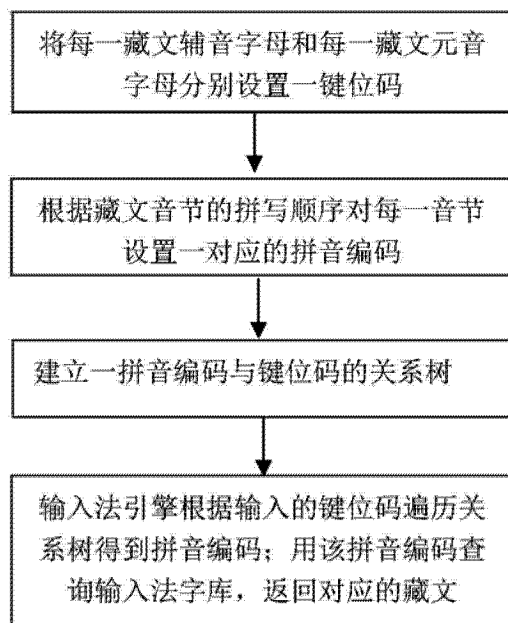


图 2

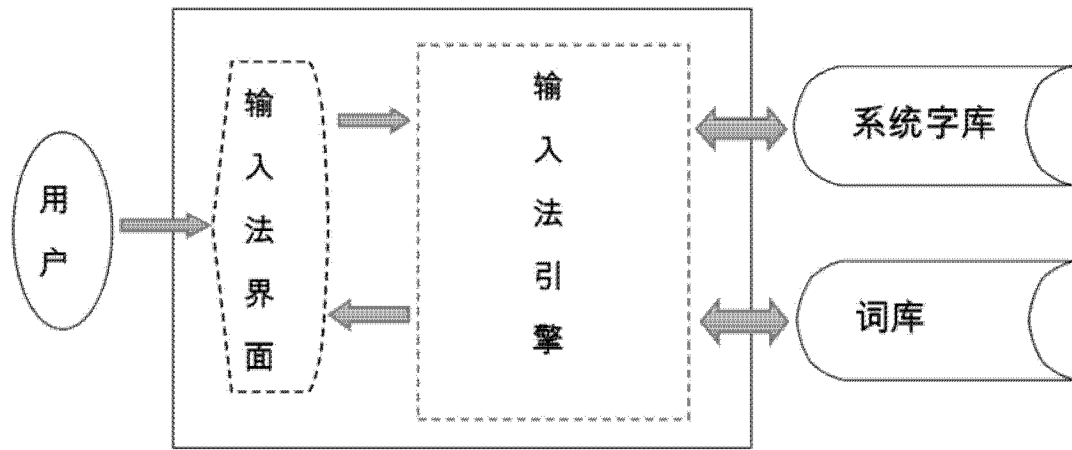


图 3