



# [12] 发明专利申请公开说明书

[21] 申请号 200410000116.1

[43] 公开日 2004年12月22日

[11] 公开号 CN 1556458A

[22] 申请日 2004.1.5

[21] 申请号 200410000116.1

[71] 申请人 郑方

地址 100084 北京市海淀区清华园清华大学  
西14-4-202

[72] 发明人 罗迈克 郑方

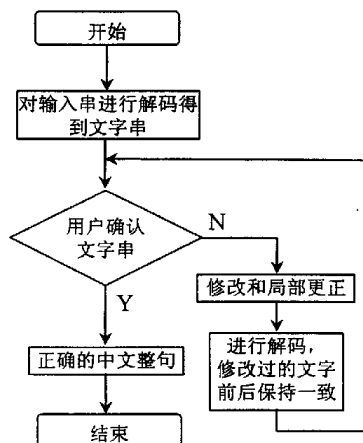
[74] 专利代理机构 北京清亦华知识产权代理事务所  
代理人 罗文群

权利要求书2页 说明书8页 附图4页

[54] 发明名称 一种中文整句输入法

[57] 摘要

本发明涉及一种中文整句输入法，属于计算机汉字输入方法技术领域。首先利用中文语言模型对用户输入的拼音串或笔画串进行解码得到文字串；由用户对上述解码文字串进行确认，并对错误进行修改和局部更正；对经过修改和更正的整句拼音串或笔画串进行解码，解码后的文字串中用户修改过的文字前后保持一致；重复上述步骤直至得到用户确认的中文整句。本发明方法具有以下优点：提供了模糊音快速重选择和对输入错误的修改方式，为输入系统的切分错误提供自动重切分修正方式；为系统的字选择错误提供选择候选字更正方式。本发明方法利用用户每次修改中所蕴涵的信息，自动进行重解码，从而快速修正句子中其他可能的错误，提高了改错效率和准确性。



- 1、一种中文整句输入法，其特征在于，该方法包括以下步骤：
  - (1) 利用中文语言模型对用户输入的拼音串或笔画串进行解码得到文字串；
  - (2) 由用户对上述解码文字串进行确认，若确认该解码结果，则为正确的中文整句，若不确认该解码结果，则用户对文字串中的任意一个错误进行修改，然后根据用户的修改作局部更正；
  - (3) 对经过上述修改和更正的整句拼音串或笔画串进行解码，解码后的文字串中用户修改过的文字前后保持一致；
  - (4) 重复步骤(2)和(3)，直至得到用户确认的中文整句。
- 2、如权利要求1所述的方法，其特征在于步骤(2)中的文字串中的错误为用户错误或系统错误，其中的用户错误为模糊音错误或输入错误，其中的系统错误为切分错误或字选择错误。
- 3、如权利要求2所述的方法，其特征在于用户对模糊音错误进行修改及局部更正的方法包括如下步骤：
  - (1) 用户选择错误文字，显示与该文字对应的拼音；
  - (2) 用户从上述错误文字的拼音相对应的模糊音菜单中选择正确的拼音。
- 4、如权利要求2所述的方法，其特征在于用户对输入错误进行修改及局部更正的方法包括如下步骤：
  - (1) 用户选择错误文字，显示与该文字对应的拼音或笔画；
  - (2) 对拼音或笔画进行修改，使其成为与正确文字相对应的拼音或笔画。
- 5、如权利要求2所述的方法，其特征在于用户对切分错误进行修改及局部更正的方法包括如下步骤：
  - (1) 用户选择两个或两个以上相邻的错误文字；
  - (2) 在用户选定的错误文字的连续拼音串中移动切分符位置，使错误文字的连续拼音串成为正确文字的连续拼音串。
- 6、如权利要求2所述的方法，其特征在于用户对字选择错误进行修改及局部更正的方法包括如下步骤：
  - (1) 用户选择错误文字，显示与该文字具有相同拼音或笔画的其它所有文字；
  - (2) 用户从上述所有文字中选择正确的文字。
- 7、如权利要求1所述的方法，其特征在于步骤(3)中，对经过修改和更正的整句拼音串或笔画串进行解码的方法包括如下步骤：
  - (1) 根据语言模型，生成一个词法树，设定一条空的搜索路径，并将其存储在路径数组中，将词法树的指针指向词法树树根；
  - (2) 按照从左往右的顺序，按步骤(3)和(4)搜索更正后的整句拼音串或笔画串；
  - (3) 从路径数组中取出一条搜索路径，若用户根据拼音或笔画选定了文字，则将取出的路径置换为一条新路径，若用户未选定文字，则根据词法树将取出的路径置换

为一条或多条新路径，同时修改新路径中相应的词法树指针和搜索信息；

(4) 对上述每条置换得到的新路径进行判断，若已到达词法树的树叶，则根据上述语言模型，按下面公式计算路径中所有文字出现的累计对数概率：

累计对数概率 = 原累计对数概率 + 当前对数概率

$$\text{对数概率} = \ln P(w_n | w_{n-2}, w_{n-1})$$

然后重新把词法树指针指向词法树树根，

上式中， $w_n$  表示该路径所指的词法树树叶中的词，对应解码出来的文字中当前的词， $w_{n-1}$  和  $w_{n-2}$  分别表示解码出来的文字中当前词前面的两个词， $P(w_n | w_{n-2}, w_{n-1})$  表示词串  $(w_{n-2}, w_{n-1}, w_n)$  的出现概率；

(5) 重复步骤 (3) 和 (4)，直到路径数组中的所有路径全部置换完毕；

(6) 对所有置换后的新路径按计算得到的累计对数概率从高到低排序，根据路径数组容量，在路径数组中保留累计对数概率高的路径；

(7) 重复步骤 (2) ~ (6)，直到整句拼音串或笔画串处理完。

## 一种中文整句输入法

**技术领域** 本发明涉及一种中文整句输入法,属于计算机(包括台式计算机、笔记本电脑、掌上电脑、个人数字助理机等)汉字输入方法技术领域,特别涉及用于无线通讯设备(如手提电话、智能手机等)汉字输入方法。

**背景技术** 中文整句输入法(简称整句输入法)是按整句的方式输入汉字的方法,它允许用户在输入时不必对每个字进行选择就可以输入一整句话。整句输入法通常使用中文语言模型(简称语言模型)来智能地预测句子里的每一个可能要输入的汉字。也就是说,语言模型将把用户的输入串解码成最可能的汉字串。这个过程也称为“解码过程”或“搜索过程”。用户输入串的表现形式,可以是字的发音方式(如拼音),也可以是字的书写方式(如笔画)。

在中文整句输入法中,通常可能会发生两种错误:1、用户在输入过程中的输入错误(如不正确的插入、删除、替换,或错误的顺序),简称用户错误;2、整句输入法对输入串解码产生的错误(如进行了错误的切分或选择了不是用户要输入的汉字),简称系统错误或解码错误。

1、用户错误:用户错误可以分为以下两类。下面分别叙述。

(1) 模糊音错误:这样的错误多发生在口音或方言比较重的人中,特别是那些来自中国南方的人。“模糊音”错误有:z↔zh、c↔ch、s↔sh、f↔h、l↔n、hu↔w、an↔ang、en↔eng、in↔ing、zhi↔ji、chi↔qi、shi↔xi、y↔r等等。

现有输入法对模糊音错误的改正方法一般是用选项。如紫光输入法提供了一个模糊音输入选项,当用户设置了该选项时,输入法将在用户输入拼音串时进行模糊匹配。如用户可以用“zong”来输入“中”,用“zhong”来输入“总”。模糊音匹配候选字与正确的匹配候选字列在一起。

对于计算和存储资源受限的个人数字助理(PDA)等设备来说,这种处理方式会导致用户输入时,可候选的汉字过多,降低输入的精度和速度。

(2) 输入错误:用户输入错误特指在用户输入过程中把字母/笔画输入错了。包括:

输入了多余的字母/笔画。例如:“he”输为“hei”,“一丿、”输为“一丿、、”;  
遗漏了一个字母/笔画。例如:“hei”输为“he”,“一丿、、”输为“一丿、”;  
输入错了一个字母/笔画。例如:“nai”输为“bai”,“一丨丿、”输为“一一丿、”。

当然,在整句输入过程中上述错误可能不止一种(如将“bin”输为“ping”,“一一丨一”输为“一一丨丨”)。

现有输入法,如紫光输入法,允许用户在任何位置修改输入的拼音串。它将用户所一次输入的拼音串连续的显示在输入框上,并允许用户用方向键将光标移动到想要的位置上。在光标位置上,用户可以更正输入错误,如:添加遗漏的字母或删除不正确

的或多余的字母。

但对于一些显示分辨率较小的设备，如个人数字助理(PDA)来说，要显示用户一次输入的整个拼音串是不现实的。另外，对于触摸屏设备来说，用笔来点选位置要比用方向键方便的多。

2、系统错误：系统错误有系统切分错误和系统解码错误。下面分别叙述。

(1) 切分错误：系统切分错误与切分算法的选择密切相关。例如，最大匹配算法在没有用户干预下是不可能从“henan”解码出“河南”的，而最小匹配算法在没有用户干预下是不可能从“pingan”解码出“平安”的。总的来说，整句输入法产生的切分错误可以分成以下几类：

一个字被切分成两个字。例如：“先”(xian)被切分成“西安”(xi'an)；

两个字被合并成一个字。例如，“西安”(xi'an)被合并成“先”(xian)；

两个字被错分成另两个字。例如，“河南”(he'nan)被错分成“很安”(hen'an)。

对于切分错误，我们以PC机上的紫光输入法为例，它以“'”表示拼音分隔符，允许用户手动地编辑输入串来修改错误的切分。但用户须先删除不正确的拼音分隔符，然后在正确的位置插入拼音分隔符。根据上面所说的，这种方法同样不适用于PDA这类小屏幕设备。

微软整句输入法使用了另外一种方法来对切分错误和输入错误进行处理。它允许用户直接将光标移动到句子中错误的汉字上并对该字进行编辑，而不是对拼音串进行编辑。在光标位置上，用户需先删除错误的汉字，然后重新输入正确的拼音。这种方法的问题在于对一个小的输入错误（如漏敲了一个字母“g”）的修改时间跟一个较大输入错误的修改（如将“bin”改为“ping”）所花的时间差不多。

(2) 字选择错误：由于语言模型的精度一般都不能达到100%的正确，在加上中文的多音字和同音字问题比较普遍，因此利用语言模型的中文整句输入法在解码时有时给出的句子，不是所有的汉字都时用户希望输入的，我们称这种解码错误为字选择错误。如用户输入了“我要买机器”的拼音，结果可能是“我要卖及其”。

在汉字选择错误这种情况下，现有的整句输入法通常会提供了如下的解决方案：1) 首先用户选择句子里错误的字；2) 整句输入法将显示其他可能的汉字候选；3) 用户从中选择需要输入的字，整句输入法对所选择的字进行更新。当存在多个字词错误时，用户需要重复上述步骤直到所有的汉字都正确。一般地讲，句子里错误的字越多，用户手动修改所花的时间越长。

还有的整句输入法提供了根据用户所做的“一步手动修改”，自动更正其他错误字的功能。例如，在微软拼音输入法中，如果用户将“及其”中的“及”字改为“机”时，整句输入法将自动的将第二个字“其”改为“器”。但这种方式每次只涉及到一个词内部的相关字的自动修改，而不能影响或自动修改相关词以外的字。

发明内容 本发明的目的是提出一种中文整句输入法，以解决现有中文整句输入法对错误修改比较繁琐的缺点，使中文输入更加快捷、方便。

本发明提出的中文整句输入法，包括以下步骤：

(1) 利用中文语言模型对用户输入的拼音串或笔画串进行解码得到文字串；

(2) 由用户对上述解码文字串进行确认，若确认该解码结果，则为正确的中文整

句，若不确认该解码结果，则用户对文字串中的任意一个错误进行修改，然后根据用户的修改作局部更正；

(3) 对经过上述修改和更正的整句拼音串或笔画串进行解码，解码后的文字串中用户修改过的文字前后保持一致；

(4) 重复步骤(2)和(3)，直至得到用户确认的中文整句。

上述方法中的文字串中的错误为用户错误或系统错误，其中的用户错误为模糊音错误或输入错误，其中的系统错误为切分错误或字选择错误。

对用户对上述文字串中的错误进行修改和局部更正的方法共有四种。

第一种是用户对模糊音错误进行修改及局部更正，包括如下步骤：

(1) 用户选择错误文字，显示与该文字对应的拼音；

(2) 用户从上述错误文字的拼音相对应的模糊音菜单中选择正确的拼音。

第二种是用户对输入错误进行修改及局部更正，包括如下步骤：

(1) 用户选择错误文字，显示与该文字对应的拼音或笔画；

(2) 对拼音或笔画进行修改，使其成为与正确文字相对应的拼音或笔画。

第三种是用户对切分错误进行修改及局部更正，包括如下步骤：

(1) 用户选择两个或两个以上相邻的错误文字；

(2) 在用户选定的错误文字的连续拼音串中移动切分符位置，使错误文字的连续拼音串成为正确文字的连续拼音串。

第四种是用户对字选择错误进行修改及局部更正，包括如下步骤：

(1) 用户选择错误文字，显示与该文字具有相同拼音或笔画的其它所有文字；

(2) 用户从上述所有文字中选择正确的文字。

上述方法中，对经过修改和更正的整句拼音串或笔画串进行解码的方法包括如下步骤：

(1) 根据语言模型，生成一个词法树，设定一条空的搜索路径，并将其存储在路径数组中，将词法树的指针指向词法树树根；

(2) 按照从左往右的顺序，按步骤(3)和(4)搜索更正后的整句拼音串或笔画串；

(3) 从路径数组中取出一条搜索路径，若用户根据拼音或笔画选定了文字，则将取出的路径替换为一条新路径，若用户未选定文字，则根据词法树将取出的路径替换为一条或多条新路径，同时修改新路径中相应的词法树指针和搜索信息；

(4) 对上述每条置换得到的新路径进行判断，若已到达词法树的树叶，则根据上述语言模型，按下面公式计算路径中所有文字出现的累计对数概率：

$$\text{累计对数概率} = \text{原累计对数概率} + \text{当前对数概率}$$

$$\text{对数概率} = \ln P(w_n | w_{n-2}, w_{n-1})$$

然后重新把词法树指针指向词法树树根。

上式中， $w_n$ 表示该路径所指的词法树树叶中的词，对应解码出来的文字中当前的

词， $w_{n-1}$ 和 $w_{n-2}$ 分别表示解码出来的文字中当前词前面的两个词， $P(w_n|w_{n-2}, w_{n-1})$ 表示词串 $(w_{n-2}, w_{n-1}, w_n)$ 的出现概率，由已有的方法估计；

(5) 重复步骤(3)和(4)，直到路径数组中的所有路径全部置换完毕；

(6) 对所有置换后的新路径按计算得到的累计对数概率从高到低排序，根据路径数组容量，在路径数组中保留累计对数概率高的路径；

(7) 重复步骤(2)~(6)，直到整句拼音串或笔画串处理完。

本发明提出的中文整句输入法，其优点是：

- 1、本发明为那些由于口音或方言习惯而不能准确拼出拼音的用户提供模糊音快速自动重选择的方法；
- 2、本发明为用户修改输入过程中的输入错误提供了方便的修改方式；
- 3、本发明为系统解码所导致的切分错误提供了方便的自动重切分修正方式；
- 4、本发明为系统解码所导致的字选择错误提供了快速的选择候选字的更正方式；
- 5、本发明充分利用用户每一次修改中所蕴涵的信息，自动进行重解码，从而可以快速修正句子中其他可能的错误，提高了改错效率和准确度。

附图说明

图1是本发明方法的流程框图。

图2是本发明方法中输入界面图。

图3、图4和图5是本发明方法中修改和更正的三个不同的实施例。

图6是本发明方法中对经过修改和更正的整句拼音串或笔画串进行解码的流程图。

图7是对修改和更正的整句拼音串或笔画串进行解码时所用词法树示例。

具体实施方式

本发明提出的中文整句输入法，其流程如图1所示，首先利用中文语言模型对用户输入的拼音串或笔画串进行解码得到文字串；由用户对上述解码文字串进行确认，若确认该解码结果，则为正确的中文整句，若不确认该解码结果，则用户对文字串中的任意一个错误进行修改，然后根据用户的修改作局部更正；对经过修改和更正的整句拼音串或笔画串进行解码，解码后的文字串中用户修改过的文字前后保持一致；重复上述过程，直至得到用户确认的中文整句。

上述方法中的文字串中的错误为用户错误或系统错误，其中的用户错误为模糊音错误或输入错误，其中的系统错误为切分错误或字选择错误。

对用户对上述文字串中的错误进行修改和局部更正的方法共有四种。

第一种是用户对模糊音错误进行修改及局部更正，包括如下步骤：

- (1) 用户选择错误文字，显示与该文字对应的拼音；
- (2) 用户从上述错误文字的拼音相对应的模糊音菜单中选择正确的拼音。

第二种是用户对输入错误进行修改及局部更正，包括如下步骤：

- (1) 用户选择错误文字，显示与该文字对应的拼音或笔画；
- (2) 对拼音或笔画进行修改，使其成为与正确文字相对应的拼音或笔画。

第三种是用户对切分错误进行修改及局部更正，包括如下步骤：

- (1) 用户选择两个或两个以上相邻的错误文字；
- (2) 在用户选定的错误文字的连续拼音串中移动切分符位置，使错误文字的连续拼音串成为正确文字的连续拼音串。

第四种是用户对字选择错误进行修改及局部更正，包括如下步骤：

- (1) 用户选择错误文字，显示与该文字具有相同拼音或笔画的其它所有文字；
- (2) 用户从上述所有文字中选择正确的文字。

上述方法中，对经过修改和更正的整句拼音串或笔画串进行解码的方法的流程如图 6 所示，根据语言模型，生成一个词法树，设定一条空的搜索路径，并将其存储在路径数组中，将词法树的指针指向词法树树根；按照从左往右的顺序，搜索更正后的整句拼音串或笔画串；从路径数组中取出一条搜索路径，若用户根据拼音或笔画选定了文字，则将取出的路径置换为一条新路径，若用户未选定文字，则根据词法树将取出的路径置换为一条或多条新路径，同时修改新路径中相应的词法树指针和搜索信息；对每条置换得到的新路径进行判断，若已到达词法树的树叶，则根据上述语言模型，按下面公式计算路径中所有文字出现的累计对数概率：累计对数概率 = 原累计对数概率 + 当前对数概率

$$\text{对数概率} = \ln P(w_n | w_{n-2}, w_{n-1})$$

然后重新把词法树指针指向词法树树根。上式中， $w_n$  表示该路径所指的词法树树叶中的词，对应解码出来的文字中当前的词， $w_{n-1}$  和  $w_{n-2}$  分别表示解码出来的文字中当前词前面的两个词， $P(w_n | w_{n-2}, w_{n-1})$  表示词串  $(w_{n-2}, w_{n-1}, w_n)$  的出现概率，由已有的方法估计；

重复上述直到路径数组中的所有路径全部置换完毕；对所有置换后的新路径按计算得到的累计对数概率从高到低排序，根据路径数组容量，在路径数组中保留累计对数概率高的路径；重复上述过程直到整句拼音串或笔画串处理完。

下面以基于拼音的中文整句输入法说明本发明方法中对经过修改和更正的整句拼音串或笔画串进行解码的过程。词法树的示例如图 7 所示，是按音节组织的，其中  $\Phi$  表示空，不与任何拼音匹配。按图中箭头的方向，从树根走到树叶，会得到一个拼音串，该拼音串对应于相应的树叶中所保存的词，该拼音串用以与用户的输入拼音串进行匹配。所有树叶中包含的词组成了语言模型中的词表。

以用户输入“zhong guo ren min”且其中的“guo”已经被用户选定为“国”为例，说明搜索过程。

(1) 搜索开始时，只有一条空路径，其中的词法树指针指向词法树的树根，记录累计对数概率的程序变量清为 0。

(2) 按照从左往右的顺序，按步骤 (3) 和 (4) 搜索更正后的整句拼音串或笔画串。



(3) 从路径数组中取出一条搜索路径，若用户根据拼音或笔画选定了文字，则将取出的路径置换为一条新路径，若用户未选定文字，则根据词法树将取出的路径置换为一条或多条新路径，同时修改新路径中相应的词法树指针和搜索信息；比如，(a) 若当前路径中词法树指针指向树根，而当前正在匹配第一个拼音“zhong”，检查词法树树根下所指，与用户输入“zhong”相匹配的只有一个方向，沿着该方向下去，另外有两条无需匹配任何音节( $\Phi$ )即可到达树叶结点的方向，因此把该路径置换为三条新路径，其中第一条新路径的词法树指针指向词法树中“zhong”所指的那个结点，第二条新路径的词法树指针指向词法树中“中”那个树叶结点，第三条新路径的词法树指针指向词法树中“忠”那个树叶结点；(b) 若当前词法树指针指向词法树中“zhong”所指的结点，且用户输入指向用户已经选定的“国”，检查词法树中“zhong”所指结点的后续结点所对应的叶子结点中是否有与“国”匹配的，发现“中国”、“中心”、“忠心”、“中”和“忠”5个树叶中只有一个可以与国匹配，那么将该路径置换为一个新的路径，词法树指针指向词法树中“guo”所指的那个结点，而“guo”后是一个 $\Phi$ ，因此直接将指针指向“中国”树叶结点。

(4) 对上述每条置换得到的新路径进行判断，若已到达词法树的树叶，则根据上述语言模型，按下面公式计算路径中所有文字出现的累计对数概率：

$$\text{累计对数概率} = \text{原累计对数概率} + \text{当前对数概率}$$

$$\text{对数概率} = \ln P(w_n | w_{n-2}, w_{n-1})$$

然后重新把词法树指针指向词法树树根，上式中， $w_n$ 表示该路径所指的词法树树叶中的词，对应解码出来的文字中当前的词， $w_{n-1}$ 和 $w_{n-2}$ 分别表示解码出来的文字中当前词前面的两个词， $P(w_n | w_{n-2}, w_{n-1})$ 表示词串 $(w_{n-2}, w_{n-1}, w_n)$ 的出现概率，由已有的方法估计；比如在某条路径中，已经处理到第二个音节，词法树指针指向“国”树叶结点，而该路径保留的前续词为“中”，且累计对数概率为-2.99，词串“中、国”的概率为 $P(\text{国} | \text{中}) = 0.1$ ，那么累计对数概率 $= -2.99 + \ln 0.1 = (-2.99) + (-2.30) = -5.29$ 。

(5) 重复步骤(3)和(4)，直到路径数组中的所有路径全部置换完毕。

(6) 对所有置换后的新路径按计算得到的累计对数概率从高到低排序，根据路径数组容量，在路径数组中保留累计对数概率高的路径。比如，在搜索到拼音“ren”时，我们得到这样的一些路径及其相应的累计对数概率

路径(a): “中、国、人”，	-8.34
路径(b): “中、国、ren”，	-5.29
路径(c): “忠、国、人”，	-10.56
路径(d): “忠、国、ren”，	-8.60
路径(e): “中国、人”，	-5.10
路径(f): “中国、ren”	-3.78
路径(g): “中国人”	-4.90

而数组容量为 5，那么将把路径(f)、(g)、(e)、(b)、(d)保留在路径数组中。

(7) 重复步骤(2)~(6)，直到整句拼音串处理完，最后得到“中国人民”。

上述四种修改和局部更正的方法，在移动设备如 PDA、手机等上，需要一个用户接口，如可以显示至少 6 个字符（对应最长的拼音如“zhuang”）宽度的显示区域，如图 2。

以下结合附图介绍本发明的实施例。

图 3 所示是修改模糊音错误并进行局部更正的实施例：

用户选择错误文字，显示与该文字对应的拼音；用户从上述错误文字的文字拼音相对应的模糊音菜单中选择正确的拼音。

例如，用户错误的输入“yizangpiao”后，系统解码为“亿藏票”。然后，用户选择“藏”字，在拼音区域将显示“zang”。接着用户按住拼音区域，这时将弹出一个有“zan”、“zhan”和“zhang”三个选项的浮动菜单。用户选择“zhang”后，系统将原先的拼音串更新为“yizhangpiao”。对于具有自动拼音串重新解码功能的系统来说，系统将对更新后的拼音串重新进行解码，得到汉字串“一张票”。

图 4 所示是修改输入错误并进行局部更正的实施例：

用户选择错误文字，显示与该文字对应的拼音或笔画；对拼音或笔画进行修改，使其成为与正确文字相对应的拼音或笔画。

例如，用户错误的输入“nirushuo”后，系统解码为“你如说”。这时，用户选择“你”字，系统将在拼音区域显示“ni”。接着用户点击该拼音串使之成为可编辑状态。然后用户将“ni”改为“bi”，并点击该拼音串以结束本次修改。由于“bi”是一个有效的拼音串，因此，系统将原先的拼音串更新为“birushuo”。对于具有自动拼音串重新解码功能的系统来说，系统将对更新后的拼音串重新进行解码，得到汉字串“比如说”。

图 5 所示是修改切分错误并进行局部更正的实施例：

用户选择两个或两个以上相邻的错误文字；在错误文字的连续拼音串中移动切分符位置，使错误文字的连续拼音串成为正确文字的连续拼音串。

例如，用户输入串为“henansheng”，而系统错误的切分为“hen'an'sheng”并输出汉字串“很安生”。这时，用户用笔选择了“很安”。系统将得到左字拼音串“hen”与右字拼音串“an”，并在中间用一个拼音分隔将这两个串连接称为一个串“hen'an”。系统用子算法来判断该串是否存在其他的切分方式。子算法将拼音分隔符从当前位置左移一位，既拼音串变为“he'nan”，然后对该串用局部解码算法进行解码。这时解码正确，则解码结果返回给系统。系统将把原先的拼音串改为“he'nan'sheng”。对于具有自动拼音串重新解码功能的系统来说，对输入串重新解码得到“河南省”。

修改各种错误并进行局部更正的综合示例

例如，用户输入拼音串“suijienishuijiaoshichishuijiao”后，系统错误的解码结果为“岁届泥水教师吃睡觉”。

这时，用户将句子里的第一个字“岁”选中，希望进行模糊音选择，系统提示有“sui”和“shui”，用户选择“shui”，系统将锁定修改并自动重新解码，得到结果“水解泥水教师吃睡觉”。用户选定第一个汉字，系统列出具有相同拼音“shui”的所有汉

字，用户从中锁定汉字“谁”，此时系统将锁定修改并自动重新解码，结果为“谁届泥水教师吃睡觉”。

用户把第二个字的拼音错误改正为“jiao”后，系统将锁定修改并自动重新解码为“谁叫你睡觉时吃睡觉”。

然后用户将句子里的最后一个字“觉”改为“饺”，系统将锁定修改并对整个拼音串重解码，输出结果为正确的汉字串“谁叫你睡觉时吃水饺”。

因此，虽然句子里总共有8个错误的字，但用户只需做四次修改就可以得到正确的句子。如果不存在模糊音错误和输入错误的问题，则两次修改即可修正8个错误，效率大大提高。

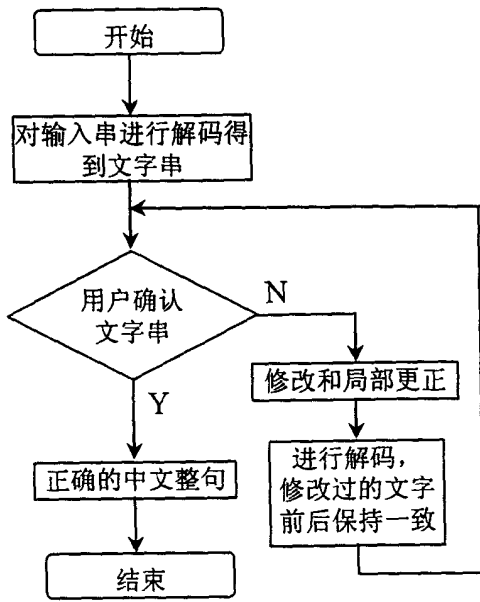


图 1

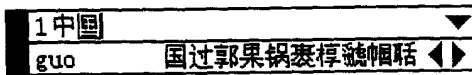


图 2

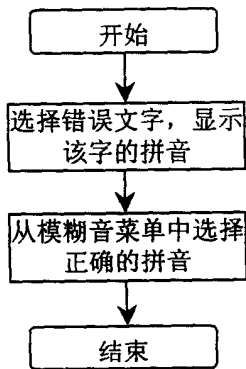


图 3

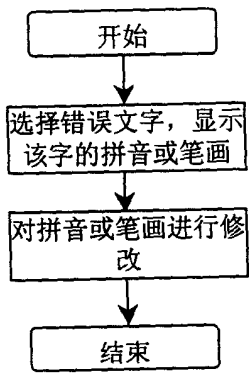


图 4

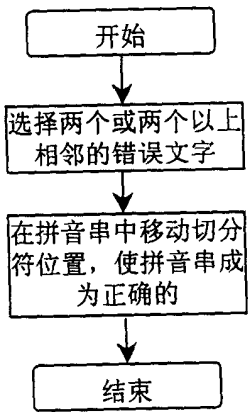


图 5

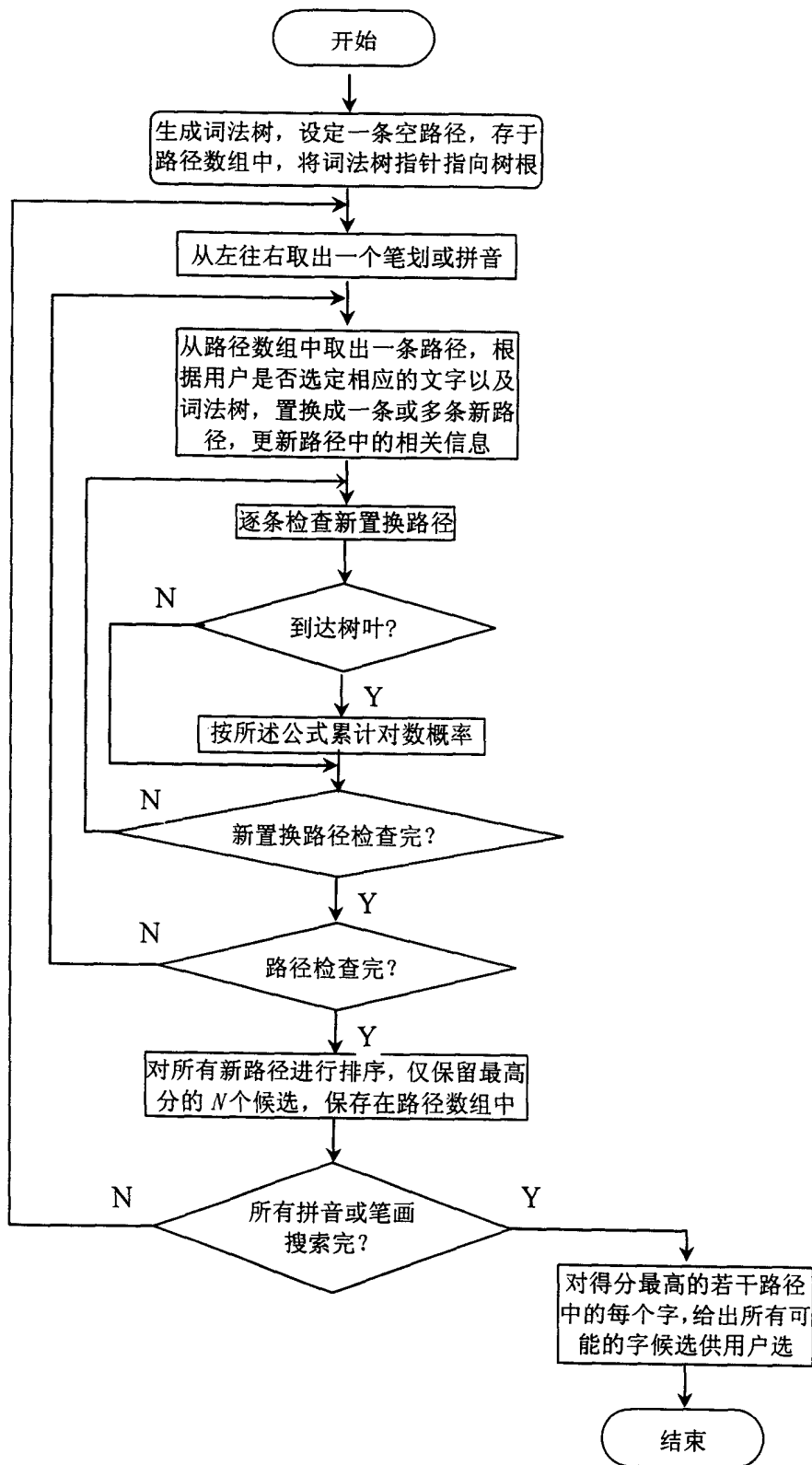


图 6

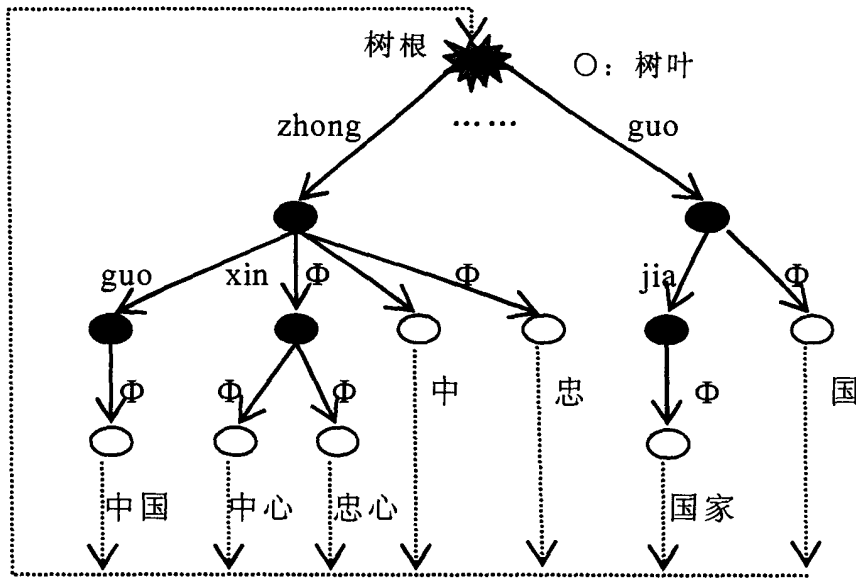


图 7