



(12)发明专利

(10)授权公告号 CN 105653061 B

(45)授权公告日 2020.03.31

(21)申请号 201511020886.7

(51)Int.Cl.

(22)申请日 2015.12.29

G06F 3/023(2006.01)

(65)同一申请的已公布的文献号

(56)对比文件

申请公布号 CN 105653061 A

CN 1556458 A,2004.12.22,

(43)申请公布日 2016.06.08

CN 102866782 A,2013.01.09,

(73)专利权人 北京京东尚科信息技术有限公司

CN 102867040 A,2013.01.09,

地址 100080 北京市海淀区杏石口路65号

CN 103514236 A,2014.01.15,

西杉创意园西区11C楼东段1-4层西段

CN 104252484 A,2014.12.31,

1-4层

CN 104504142 A,2015.04.08,

专利权人 北京京东世纪贸易有限公司

审查员 张博

(72)发明人 何兴洋

(74)专利代理机构 北京成创同维知识产权代理

有限公司 11449

代理人 蔡纯 张靖琳

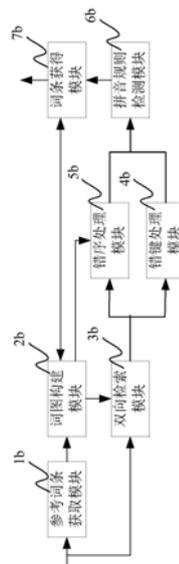
权利要求书3页 说明书8页 附图8页

(54)发明名称

针对拼音输入法的词条检索及错词检测的方法和系统

(57)摘要

本发明公开了一种针对拼音输入法的词条检索及错词检测的方法和系统,所述检索方法包括检测当前输入的拼音串前或后是否有参考词条;如果有前或后参考词条,根据前或后参考词条及当前输入的拼音串,在词库中进行匹配,获得至少一个后或前置词条;如果当前输入的拼音串的前后没有参考词条,根据当前的拼音串在词库中进行匹配获得相应的词条或词条列表。所述错词检测的方法包括根据用户当前输入的拼音串的参考词条得到至少一个前置或后置词条;构建词图;分别进行正逆向对比,错键处理或错序处理;判断外理后的拼音串是否符合拼音规则,如果符合向用户返回所述词图所对应的词条。本发明提高了输入法中的词条推荐准确率及用户输入纠错的能力。



1. 一种针对拼音输入法的词条检索方法,其中,包括以下步骤:

检测当前输入的拼音串前和/或后是否有已生成的参考词条,所述拼音串包括多个字符;

如果有前和/或后参考词条,根据前和/或后参考词条及当前输入的拼音串,在词库中进行匹配,获得至少一个后和/或前置词条,从而获得与当前输入的拼音串相对应的词条;

如果当前输入的拼音串的前后没有参考词条,根据当前的拼音串在词库中进行匹配获得相应的词条或词条列表,

如果当前输入的拼音串的前后都没有参考词条,根据当前的拼音串在词库中进行匹配时包括以下步骤:

检测当前的拼音串是否为多个子词条的拼音,如果是,根据拼音规则将当前词条分离为多个子拼音串;

根据当前子拼音串,按照从前到后的顺利,逐个进行匹配,

在根据拼音规则将当前词条分离为多个子拼音串后,还包括判断所述多个子拼音串是否符合拼音规则的步骤,如果有不符合拼音规则的子拼音串,包括以步骤:

获得与符合拼音规则的子拼音串相对应的子词条;

将与不符合拼音规则的子拼音串相邻的子词条作为参考子词条;

根据所述参考子词条及不符合拼音规则的子拼音串,在词库中进行匹配得到与不符合拼音规则的子拼音串相对应的至少一个子词条。

2. 如权利要求1所述的针对拼音输入法的词条检索方法,其中,如果当前输入的拼音串同时有前参考词条和后参考词条;在词库中进行匹配时,具体包括以下步骤:

根据前参考词条及当前输入的拼音串,在词库中匹配得到第一组词条;

根据后参考词条及当前输入的拼音串,在词库中匹配得到第二组词条;

将所述第一组词条和第二组词条的合集作为与当前输入的拼音串对应的候选词条集。

3. 一种针对拼音输入法的词条检索系统,其中,包括:

用户界面,用于提供与用户交互的接口;

检测模块,用于检测通过用户界面输入的当前拼音串前和/或后是否有已生成的参考词条,所述拼音串包括多个字符;

词条匹配模块,用于根据输入信息,在词库获得相应的词条或词条列表;

控制模块,与所述检测模块连接,根据检测模块的检测结果,向所述词条匹配模块发送相应的输入信息,从而使词条匹配模块获得与所述当前拼音串相对应的词条;和

词库,用于存储词条,

所述控制模块包括:

拼音串分离单元,用于将当前输入的拼音串按拼音规则分离成多个子拼音串,将所述多个子拼音串送给所述的词条匹配模块作为输入信息;

甄别单元,用于判断所述多个子拼音串是否符合拼音规则,并将据此得到的甄别信息判断结果作为输入信息的一部分送给所述的词条匹配模块;

所述词条匹配模块包括:

第一匹配单元,用于根据所述输入信息中的符合拼音规则的子拼音串在词库中获得相应的至少一个子词条;

参考子词条确定单元,用于根据不符合拼音规则的子拼音串在拼音串的位置确定与其相应的参考子词条;

第二匹配单元,用于根据参考子词条匹配及不符合拼音规则的子拼音串,在词库中得到与不符合拼音规则的子拼音串相对应的至少一个子词条。

4.如权利要求3所述的针对拼音输入法的词条检索系统,其中,所述词条匹配模块的输入信息包括当前输入的拼音串,或当前输入的拼音串和前或后参考词条的组合信息。

5.如权利要求3所述的针对拼音输入法的词条检索系统,其中,所述词条匹配模块包括:

输入信息解析单元,用于分析输入信息,得到两组匹配指令;

两组匹配单元,分别根据相应的匹配指令得到对应的词条组;

词条选择单元,在所述两组词条组中选出合集的词条,并将其作为与当前输入的拼音串对应的候选词条。

6.一种针对拼音输入法的错词检测方法,其中,包括以下步骤:

步骤1,针对用户当前输入的拼音串,得到至少一个前置或后置词条;

步骤2,根据所述至少一个前置和/或后置词条构建以不重复的拼音字母为节点、以词条的拼音顺序为节点相邻顺序的词图;

步骤3,分别以当前拼音串的正向拼音顺序与所述词图逐个字母正向对比和以当前拼音串的逆向拼音顺序与所述词图逐个字母逆向对比,当拼音串的字母与词图中相应的字母不同时,终止对比;

步骤4,判断终止处的词图字母与拼音串字母是否位于键盘的相邻位置,如果相邻,将所述拼音串字母替换为所述终止处的词图字母;

步骤5,判断替换后的拼音串是否符合拼音规则,如果符合,向用户返回所述词图所对应的词条,

其中,所述得到至少一个前置或后置词条,包括以下步骤:

检测当前输入的拼音串前和/或后是否有已生成的参考词条,所述拼音串包括多个字符;

如果有前和/或后参考词条,根据前和/或后参考词条及当前输入的拼音串,在词库中进行匹配,获得至少一个后和/或前置词条。

7.如权利要求6所述的针对拼音输入法的错词检测方法,其中,步骤3中终止对比之后,判断正向对比时终止处的词图字母与逆向对比终止处的词图字母是否位于词图相邻位置,如果是,则交换拼音串中两个终止字母的顺序,判断交换顺序后的拼音串是否符合拼音规则,如果符合,则向用户返回所述词图中的词条。

8.一种针对拼音输入法的错词检测系统,其中,包括:

参考词条获取模块,用于针对用户当前输入的拼音串,得到至少一个前置或后置词条;

双向检索模块,分别以当前拼音串的正向拼音顺序与词图逐个字母正向对比和以当前拼音串的逆向拼音顺序与所述词图逐个字母逆向对比,当拼音串的字母与词图中相应的字母不同时,获得拼音串终止字母和词图终止字母;

错键处理模块,用于比较拼音串终止字母和词图终止字母在键盘分布图中的位置,如果二者相邻,则将拼音串中的拼音串终止字母替换为词图终止字母;

错序处理模块,用于比较正向对比时终止处的词图字母与逆向对比终止处的词图字母在词图中的位置,如果二者相邻,则交换拼音串中两个终止字母的顺序;

拼音规则检测模块,用于检测经过所述错键处理模块或错序处理模块处理后的拼音串是否符合拼音规则;和

词条获得模块,在所述词图中获得与所述拼音串相对应的词条,

其中,所述参考词条获取模块包括以下步骤:

检测当前输入的拼音串前和/或后是否有已生成的参考词条,所述拼音串包括多个字符;

如果有前和/或后参考词条,根据前和/或后参考词条及当前输入的拼音串,在词库中进行匹配,获得至少一个后和/或前置词条;

词图构建模块,用于根据参考词条获取模块获取的前置或后置词条构建所述词图。

9. 一种针对拼音输入法的词条检索装置,包括:存储器;以及耦接至所述存储器的处理器,所述处理器被配置为基于存储在所述存储器中的指令,执行权利要求1-2任一项所述的词条检索方法。

10. 一种计算机可读存储介质,所述计算机可读存储介质存储有计算机指令,所述指令被处理器执行时实现如权利要求1-2任一项所述的词条检索方法。

11. 一种针对拼音输入法的错词检测装置,包括:存储器;以及耦接至所述存储器的处理器,所述处理器被配置为基于存储在所述存储器中的指令,执行权利要求6-7任一项所述的错词检测方法。

12. 一种计算机可读存储介质,所述计算机可读存储介质存储有计算机指令,所述指令被处理器执行时实现如权利要求6-7任一项所述的错词检测方法。

针对拼音输入法的词条检索及错词检测的方法和系统

技术领域

[0001] 本发明涉及键盘输入的数字处理技术,具体地说,涉及一种针对拼音输入法的词条检索及错词检测的方法和系统。

背景技术

[0002] 汉语是互联网世界中用量占比很大的一种语言,以搜狗输入法为代表的一系列拼音输入法迅速占领了汉语输入法的市场。这些输入法对用户输入的错词有一定的甄别能力,但是面对一些情况的表现却不是理想。

[0003] 以搜狗输入法为例,当用户输入的拼音出现错误时(拼音中字母写错或字母错序),比如用户希望拼写苹果(pingguo)但是拼写成了pungguo,搜狗输入法的做法是取出拼音字符串中的声母p和g,然后根据用户的热词词库去找以p和g开头的热词,挑选出热度排名靠前的字反馈给用户。这里所述的热词词库是指用户每次输入的词条会被输入法记录下来,由此而建立的词库。当用户再次输入相同拼音时,优先从热词词库中挑选用户的历史词条进行推荐。但是上述的纠错方式一定的问题:

[0004] 1.当汉字拼音首字母出现错误时,根据现有机制并不能有效的反馈用户需要的词条。例如,用户希望拼写“中国”(zhongguo),但是按键盘上的z键时不小心按成了旁边的x键,那么输入的字符串变为xhongguo。根据现有技术,首个提供的词条是“小红果”而非用户希望的“中国”。

[0005] 2.现有技术不能够联系上下文获得用户真正希望得到的词条,而是仅凭用户输入的拼音,利用热词词库找出词条。例如,在“图书”后边打ynigxiang,会出现常用词“影响”而不是“影像”;之后在“深远”后边打ynigxiang,会出现“影像”,因为此时的热词已变成“影像”。

[0006] 3.现有算法优先使用热词词库为用户提供待选词条,若用户输入的拼音首字母组合并不存在于用户的热词词库内,则无法进行词条的推荐(该情况通常出现于较长的词条)。例如,用户未输入过“健康一身轻”,用户实际输入jiankangyishneqing,即使声母没有拼错,现有技术也无法检测出用户原本希望的词语,结果为“健康医生呢请”。

发明内容

[0007] 为了解决现有技术中的问题,本发明提供了一种针对拼音输入法的词条检索及错词检测的方法和系统,提高输入法中的词条推荐准确率及对用户输入纠错的能力。

[0008] 为解决上述技术问题,本发明提供了一种针对拼音输入法的词条检索方法,具体包括以下步骤:

[0009] 检测当前输入的拼音串前或后是否有参考词条;

[0010] 如果有前或后参考词条,根据前或后参考词条及当前输入的拼音串,在词库中进行匹配,获得至少一个后或前置词条;

[0011] 如果当前输入的拼音串的前后没有参考词条,根据当前的拼音串在词库中进行匹

配获得相应的词条或词条列表。

[0012] 优选地,在所述的针对拼音输入法的词条检索方法中,如果当前输入的拼音串同时有前参考词条和后参考词条;在词库中进行匹配时,具体包括以下步骤:

[0013] 根据前参考词条及当前输入的拼音串,在词库中匹配得到第一组词条;

[0014] 根据后参考词条及当前输入的拼音串,在词库中匹配得到第二组词条;

[0015] 将所述第一组词条和第二组词条的合集作为与当前输入的拼音串对应的候选词条集。

[0016] 优选地,在所述的针对拼音输入法的词条检索方法中,如果当前输入的拼音串的前后没有参考词条,根据当前的拼音串在词库中进行匹配时包括以下步骤:

[0017] 检测当前的拼音串是否为多个子词条的拼音,如果是,根据拼音规则将当前词条分离为多个子拼音串;

[0018] 根据当前子拼音串,按照从前到后的顺利,逐个进行匹配。

[0019] 优选地,在所述的针对拼音输入法的词条检索方法中,在根据拼音规则将当前词条分离为多个子拼音串后,还包括判断所述多个子拼音串是否符合拼音规则的步骤,如果有不符合拼音规则的子拼音串,包括以步骤:

[0020] 获得与符合拼音规则的子拼音串相对应的子词条;

[0021] 将与不符合拼音规则的子拼音串相邻的子词条作为参考子词条;

[0022] 根据所述参考子词条及不符合拼音规则的子拼音串,在词库中进行匹配得到与不符合拼音规则的子拼音串相对应的至少一个子词条。

[0023] 为解决上述技术问题,本发明还提供了一种针对拼音输入法的词条检索系统,其中,包括:

[0024] 用户界面,用于提供与用户交互的接口;

[0025] 检测模块,用于检测通过用户界面输入的当前拼音串前或后是否有参考词条;

[0026] 词条匹配模块,用于根据输入信息,在词库获得相应的词条或词条列表;

[0027] 控制模块,与所述检测模块连接,根据检测模块的检测结果,向所述词条匹配模块发送相应的输入信息;和

[0028] 词库,用于存储词条。

[0029] 优选地,在所述的针对拼音输入法的词条检索系统中,所述词条匹配模块的输入信息包括当前输入的拼音串,或当前输入的拼音串和前或后参考词条的组合信息。

[0030] 优选地,在所述的针对拼音输入法的词条检索系统中,所述词条匹配模块包括:

[0031] 输入信息解析单元,用于分析输入信息,得到两组匹配指令;

[0032] 两组匹配单元,分别根据相应的匹配指令得到对应的词条组;

[0033] 词条选择单元,在所述两组词条组中选出合集的词条,并将其作为与当前输入的拼音串对应的候选词条。

[0034] 优选地,在所述的针对拼音输入法的词条检索系统中,所述控制模块包括:

[0035] 拼音串分离单元,用于将当前输入的拼音串按拼音规则分离成多个子拼音串,将所述多个子拼音串送给所述的词条匹配模块作为输入信息。

[0036] 优选地,在所述的针对拼音输入法的词条检索系统中,所述控制模块还包括甄别单元,用于判断所述多个子拼音串是否符合拼音规则,并将所述甄别信息作为输入信息的

一部分送给所述的词条匹配模块。

[0037] 优选地,在所述的针对拼音输入法的词条检索系统中,所述词条匹配模块包括:

[0038] 第一匹配单元,用于根据所述输入信息中的符合拼音规则的子拼音串获得相应的至少一个子词条;

[0039] 参考子词条确定单元,用于根据不符合拼音规则的子拼音串在拼音串的位置确定与其相应的参考子词条;

[0040] 第二匹配单元,用于根据参考子词条匹配及不符合拼音规则的子拼音串,在词库中得到与不符合拼音规则的子拼音串相对应的至少一个子词条。

[0041] 通过以上方法和系统,本发明可以根据用户的意图推荐词条,大大提高了输入法中的词条推荐准确率。

[0042] 为解决上述技术问题,本发明还提供了一种基于前述方法和系统的针对拼音输入法的错词检测方法,其中,包括以下步骤:

[0043] 步骤1,针对用户当前输入的拼音串,根据前述的词条检索方法和系统得到至少一个前置或后置词条;

[0044] 步骤2,根据所述至少一个前置或后置词条构建以不重复的拼音字母为节点、以词条的拼音顺序为节点相邻顺序的词图;

[0045] 步骤3,分别以当前拼音串的正向拼音顺序与所述词图逐个字母正向对比和以当前拼音串的逆向拼音顺序与所述词图逐个字母逆向对比,当拼音串的字母与词图中相应的字母不同时,终止对比;

[0046] 步骤4,判断终止处的词图字母与拼音串字母是否位于键盘的相邻位置,如果相邻,将所述拼音串字母替换为所述终止处的词图字母;

[0047] 步骤5,判断替换后的拼音串是否符合拼音规则,如果符合,向用户返回所述词图所对应的词条。

[0048] 优选地,在所述的针对拼音输入法的错词检测方法中,步骤3中终止对比之后,判断正向对比时终止的词图字母与逆向对比终止时的词图字母是否为词图中的相邻字母,如果是,则交换拼音串中两个终止字母的顺序,判断交换顺序后的拼音串是否符合拼音规则,如果符合,则向用户返回所述词图中的词条。

[0049] 为解决上述技术问题,本发明还提供了一种基于前述方法和系统的针对拼音输入法的错词检测系统,其中,包括:

[0050] 参考词条获取模块,用于针对用户当前输入的拼音串,根据前述的词条检索方法和系统得到至少一个参考词条;

[0051] 词图构建模块,用于根据参考词条获取模块获取的参考词条构建词图;

[0052] 双向检索模块,分别以当前拼音串的正向拼音顺序与所述词图逐个字母正向对比和以当前拼音串的逆向拼音顺序与所述词图逐个字母逆向对比,当拼音串的字母与词图中相应的字母不同时,获得拼音串终止字母和词图终止字母;

[0053] 错键处理模块,用于比较拼音串终止字母和词图终止字母在键盘分布图中的位置,如果二者相邻,则将拼音串中的拼音串终止字母替换为词图终止字母;

[0054] 错序处理模块,用于比较正向对比时终止的词图字母与逆向对比终止时的词图字母在词图中的位置,如果二者相邻,则交换拼音串中两个终止字母的顺序;

[0055] 拼音规则检测模块,用于检测经过所述错键处理模块或错序处理模块处理后拼音串是否符合拼音规则;和

[0056] 词条获得模块,在所述词图中获得与所述拼音串相对应的词条。

[0057] 通过以上方法和系统,本发明可以根据上下文判断用户输入的真正意图,解决现有技术无法解决的首字母拼错时无法正确推荐词条的问题和用户拼写错误时根据热词词库推荐词条不准确的弊端,并且有效自动纠正用户输入时按错键或输入错序的问题,因而有效地提高了对用户输入纠错的能力,尤其是应对拼音首字母出错的情况。提升了用户体验和输入法程序的智能化程度。

附图说明

[0058] 通过以下参照附图对本发明实施例的描述,本发明的上述以及其它目的、特征和优点将更为清楚,在附图中:

[0059] 图1为本发明所述词条检索时的词条关系图;

[0060] 图2为本发明所述针对拼音输入法的词条检索方法的一个具体实施例的流程图;

[0061] 图3为本发明所述针对拼音输入法的词条检索系统的一个具体实施例的原理框图;

[0062] 图4为本发明所述对拼音输入法的词条检索系统一具体实施例中的词条匹配模块原理框图;

[0063] 图5为本发明所述对拼音输入法的词条检索系统另一具体实施例中控制模块原理框图;

[0064] 图6为本发明所述对拼音输入法的词条检索系统另一具体实施例中的词条匹配模块原理框图;

[0065] 图7为本发明所述词条检索时的另一词条关系图。

[0066] 图8为本发明所述针对拼音输入法的错词检测方法的一具体实施例的流程图;

[0067] 图9为本发明所述词图构建过程示意图;

[0068] 图10为本发明所述词图的一个具体实施例;和

[0069] 图11为本发明所述针对拼音输入法的错词检测系统的一具体实施例的原理框图。

具体实施方式

[0070] 以下基于实施例对本发明进行描述,但是本发明并不仅仅限于这些实施例。在下文对本发明的细节描述中,详尽描述了一些特定的细节部分。对本领域技术人员来说没有这些细节部分的描述也可以完全理解本发明。为了避免混淆本发明的实质,公知的方法、过程、流程、元件和电路并没有详细叙述。

[0071] 此外,本领域普通技术人员应当理解,说明书附图中提供的附图都是为了说明的目的,并非是按比例、实际布局绘制。

[0072] 本发明提供了一种可根据上下文提供搭配词条的检索方法和系统。一个词条可能会有多种搭配词条,如图1所示,一个词条可能会有多个前置词条,也可能会有多个后置词条。本发明可以帮助用户准确地找到希望的词条,具体参考如下说明。

[0073] 如图2所示,为本发明所述针对拼音输入法的词条检索方法的一个具体实施例的

流程图。在本实施例中,包括了多种情况。具体如下:

[0074] 步骤S1,用户输入拼音串。

[0075] 步骤S2,检测当前用户输入的拼音串前后是否有参考词条。例如当用户在正常拼写时,除了首次输入,当前输入的拼音串会有前置参考词条,即此时的检索类型可以认为是字-音型。或者在某个词条前插入拼写时,会有后置参考词条,即音-字型。如果是这种情况,转到步骤S3。如果此时前和后置参考词条都没有,即音-音型,则转到步骤S8。

[0076] 步骤S3,判断是否前后都有参考词条。此时判断属于前述的正常拼写或首词条插入的情况,还是属于插入的情况,即字-音-字型,即前后都有参考词条的情况。如果只有前或后置参考词条,即字-音型/音-字型,则转到步骤S7,根据当前拼音串和前/后置参考词条在词库中得到与当前拼音串相对应的至少一个词条,并返回给用户。如果前后都有参考词条,即字-音-字型,则转到步骤S4。

[0077] 步骤S4,根据前置参考词条和当前拼音串在词库中得到包括至少一个词条的第一组词条。

[0078] 步骤S5,根据后置参考词条和当前拼音串在词库中得到包括至少一个词条的第二组词条。

[0079] 步骤S6,将两组词条合集集中的词条作为与当前拼音串相对应的候选词条集,并返回给用户。

[0080] 步骤S8,此时既没前置参考词条,也没有后置参考词条,只有拼音串的情况,即音-音型,判断此时的拼音串是单个词条的拼音串还是多个词条的拼音串,如果此时是单个词条的拼音串,则在步骤S13,根据拼音串得到词条列表并返回给用户,如果是多个词条的拼音串,则转到步骤S9。

[0081] 步骤S9,按照拼音规则将当前拼音串分离成多个子拼音串。

[0082] 步骤S10,检测每一个子拼音串是否符合拼音规则,如果符合,则在步骤S13,根据拼音串得到词条或词条列表,并返回给用户,如果不符合,步骤S11。

[0083] 步骤S11,匹配符合拼音规则的子拼音串,将得到的子词条作为参考子词条;

[0084] 步骤S12,根据参考子词条和不符合拼音规则的子拼音串在词库中得到与不符合拼音规则的子拼音串对应的词条或词条列表,并返回给用户。

[0085] 图3本发明所述针对拼音输入法的词条检索系统的一个具体实施例的原理框图。具体包括用户界面1a、检测模块2a、控制模块3a、词条匹配模块4a和词库5a。其中,用户界面1a用于提供与用户交互的接口。用户可以通过该界面输入拼音串,并获得系统返回的词条或词条列表。检测模块2a用于检测当前输入的拼音串前或后是否有参考词条。词条匹配模块4a用于根据输入信息,在词库获得相应的词条或词条列表。控制模块3a与所述检测模块2a连接,根据检测模块2a的检测结果,向所述词条匹配模块4a发送相应的输入信息。其中,输入信息可以为包括拼音串的信息,或者为包括拼音串和参考词条的组合信息。词库5a用于存储词条。本发明中的词库是一种词条搭配统计的训练集,其可以根据一个词条提供出其前后所搭配的词条,这是根据一些样本集训练而得。这个词库在网络条件允许的情况下可以放在云端;若网络条件不允许,则可以根据用户的输入习惯建立词库。

[0086] 对应于图2中的拼音串前后都有参考词条的情况,即字-音-字型时,所述检索系统中的词条匹配模块具体如图4所示。具体包括输入信息解析单元41a、两组匹配单元:第一匹

配单元42a和第二匹配单元43a和词条选择单元44a。其中,输入信息解析单元41a用于分析输入信息,得到两组匹配指令,一组包括前参考词条和拼音串,另一组包括后参考词条和拼音串,并将这两组匹配指令分别发送给第一匹配单元42a和第二匹配单元43a。第一匹配单元42a依据得到匹配指令a到词库中匹配得到若干词条a1、词条a2.....词条am。第二匹配单元43a根据匹配指令b到词库中匹配得到若干词条b1、词条b2.....词条bn。词条选择单元44a找到所述两组词条组的合集,并将合集集中的词条a2 (b2) 作为与当前输入的拼音串对应的候选词条提供给用户,具体如图7所示。

[0087] 对应于当拼音串为多个拼音串的情况,即音-音型时,本发明所述检索系统中的所述控制模块一具体实施例如图5所示。包括拼音串分离单元31a,用于将当前输入的拼音串按拼音规则分离成多个子拼音串,将所述多个子拼音串送给所述的词条匹配模块4a作为输入信息。更好地,还可以包括甄别单元32a,用于判断所述多个子拼音串是否符合拼音规则,并将所述甄别信息作为输入信息的一部分送给所述的词条匹配模块4a。因此,输入给所述词条匹配模块的输入信息则至少包括两部分,一部分为已分离的子拼音串和该拼音串是否符合拼音规则的说明信息。与此相对应的词条匹配模块4a具体如图6所示。当所述词条匹配模块4a接收到所述输入指令时,可以通过说明信息确定当前的子拼音串是否符合拼音规则,并控制第一匹配单元411a根据符合拼音规则的子拼音串在词库中获得相应的子词条;由参考子词条确定单元412a根据不符合拼音规则的子拼音串在拼音串的位置确定与其相应的参考子词条。由第二匹配单元413a根据参考子词条匹配和不符合拼音规则的子拼音串在词库中得到与不符合拼音规则的子拼音串相对应的至少一个子词条。例如当一个不符合拼音规则的子拼音串位于整个拼音串的末尾时,由其前面的已得到的子词条作为前参考词条,根据前参考词条得到一个可供挑选的词条列表。如果一个不符合拼音规则的子拼音串位于整个拼音串的中间,则以其前后相邻的已得到的子词条作为参考词条,通过前述拼音串前后都有参考词条的情况(即字-音-字型)得到一个词条。

[0088] 依据前述的检索方法和系统,本发明提供了一种错词检测的方法和系统,如图8所示,为本发明所述针对拼音输入法的错词检测方法的一具体实施例的流程图。

[0089] 步骤1,根据用户当前输入的拼音串的参考词条,应用前述方法和系统得到至少一个前置或后置词条。例如,用户当前输入的前置参考词条为“我爱”,则根据前述方法和系统,词库推荐的置后词条为“祖国”(zuguo)、“中国”(zhongguo)、“你”(ni)、“苹果”(pingguo)。

[0090] 步骤2,根据所述至少一个参考词条构建以不重复的拼音字母为节点、以词条的拼音顺序为节点相邻顺序的词图。词图的建立过程如图9所示。首先针对词条“祖国”(zuguo),以不重复的拼音字母为节点,以拼音顺序为连接顺序,依次连接,得到第一个词条的词图C1。而后,在此基础上,增加第二个词条“中国”(zhongguo),得到第二个词条的词图C2。然后,在此基础上,增加第三个词条“你”(ni),得到第三个词条的词图C3。最后,在前面的基础上,增加最后一个词条“苹果”(pingguo),得到第四个词条的词图C4。去掉表示连接方向的箭头,最终得到如图10所示的词图。

[0091] 步骤3,正逆向双向检索。即分别以当前拼音串的正向拼音顺序与所述词图逐个字母正向对比和以当前拼音串的逆向拼音顺序与所述词图逐个字母逆向对比,当拼音串的字母与词图中相应的字母不同时,出现了错误,则终止对比。例如,用户想输入的为zhongguo,

但是不小心输成了zhingguo。进行正逆向双向检索时,分别找到当前拼装串的首尾字母z和o,从正向开始,按照z-h-i-n-g-g-u-o的顺序分别与词图逐个字母正向对比,当对比到拼音串字母为i时,词图中对应位置的字母是o,二者不同,出现了错误,则到此终止正向检索,错误位置为字母o。同理进行逆向检索,按照o-u-g-g-n-i-h的顺序进行对比。当对比到拼音串中的字母为h时,词图中对应位置为字母p,二者不同,则到此终止逆向检索。因此,正向对比终止时,词图字母为o,拼音串字母为i,逆向对比终止时,词图字母为p,拼音串字母为h。

[0092] 步骤4,判断正向对比时终止的词图字母与逆向对比终止时的词图字母是否为词图中的相邻字母,即判断词图字母o与词图字母p是否在词图中相邻,参照词图,二者不相邻,则执行下一步。

[0093] 步骤5,判断终止处的词图字母与拼音串字母是否位于键盘的相邻位置。由于本实施例中,有两处终止处,因此需要单独分别判断。对于正向对比时,判断词图字母o与拼音串字母i是否在键盘中处于相邻的位置,很明显二者相邻,则执行下一步。如果二者不相邻,则在步骤10以现有任何一种技术检索得到用户想要的词条。

[0094] 步骤6,将所述拼音串字母替换为所述终止处的词图字母。即将拼音串中的i替换为o,此时拼音串变为zhongguo。

[0095] 步骤7,判断替换后的拼音串zhongguo是否符合拼音规则。很明显符合拼音规则,则进行下一步。如果不符合拼音规则,则在步骤10用现有任何一种技术来得到用户想要的词条。

[0096] 步骤8,将相应的词图上的词条“中国”返回给用户。

[0097] 对于本实施例中的逆向对比,由于逆向对比时,词图字母p与拼音串字母h不处于键盘中的相邻位置,因此,不需进行如步骤6所述的替换。

[0098] 在输入法中,用户输入错误的一个主要原因是将正确字母按成了键盘上的相邻字母,本发明通过如上所述方法可以方便快捷地纠正这一错误,并返回给用户正确的词条。在该方法中,基于键盘按键的分布图来判断当前的检索处终止字母是否为用户按错相邻键所致。如前所述,当用户希望输入“中国”,拼音输成“zhinghuo”,本发明所述的检索系统可以检测出此拼音串中的错误字母为i,当检测键盘按键的分布图时,发现o和i为相邻字母,因而确定可能是用户错误地将o按成了i,因而可以向用户返回拼音串“zhongguo”对应的词条。

[0099] 再例如,如果用户将“中国”输入成“xhongguo”,在对比词图时发现用户拼写错误且错误字母为x时,根据键盘按键的分布图判断z和x为相邻字母,可能是用户按错位,返回用户输入为zhongguo。针对用户首字母拼错的情况,这种检索方式十分有效。

[0100] 在输入法中,用户输入错误的另一个主要原因是错序,即用户将正确的字母输错了顺序。例如,用户希望输入中国,拼音输成“zhnogguo”。参照前述方法过程,在进行双向检索过程,正向对比终止于词图字母o,逆向对比终止于词图字母n,在步骤4,判断正向对比时终止的词图字母与逆向对比终止时的词图字母是否为词图中的相邻字母。在该实施例中,字母o与字母n为词图中的相邻字母,因而,在步骤9,交换拼音串中两个终止字母的顺序,判断交换顺序后的拼音串是否符合拼音规则,如果符合,则向用户返回所述词图中的词条。当交换字母o和n后,拼音串变为“zhongguo”,且符合正确的拼音规则,则这种交换是可以接受的,因而,在步骤8,将相应的词图上的词条“中国”返回给用户。

[0101] 基于上述方法,本发明还提供了一种针对拼音输入法的错词检测系统,如图11所示,为本发明所述针对拼音输入法的错词检测系统的一具体实施例的原理框图。所述错词检测系统包括参考词条获取模块1b,词图构建模块2b、双向检索模块3b、错键处理模块4b、错序处理模块5b、拼音规则检测模块6b和词条获得模块7b。

[0102] 参考词条获取模块1b的输入为用户当前输入的拼音串及参考词条,根据用户当前输入类型,如前面所述字-音型/音-字型等得到至少一个后/前置词条。将并得到的多个参考词条发送给所述的词图构建模块2b,所述词条构建模块2b根据所述多个参考词条构建词图。双向检索模块3b根据用户当前输入的拼音串,分别以当前拼音串的正向拼音顺序与所述词图逐个字母正向对比和以当前拼音串的逆向拼音顺序与所述词图逐个字母逆向对比,当拼音串的字母与词图中相应的字母不同时,此处为错误位置,获得拼音串终止字母和词图终止字母。并将错误位置处的拼音串终止字母和词图终止字母发送给错键处理模块4b和错序处理模块5b。错键处理模块4b比较拼音串终止字母和词图终止字母在键盘分布图中的位置,如果二者相邻,则将拼音串中的拼音串终止字母替换为词图终止字母。错序处理模块5b比较正向对比时终止的词图字母与逆向对比终止时的词图字母在词图中的位置,如果二者相邻,则交换拼音串中两个终止字母的顺序。拼音规则检测模块6b,用于检测经过错键处理模块4b或错序处理模块5b处理后的拼音串是否符合拼音规则。词条获得模块7b根据得到的拼音串,在所述词图中获得与所述拼音串相对应的词条。

[0103] 通过上述方法和系统,解决了现有技术无法解决的首字母拼错时无法正确推荐词条的问题,也解决了用户拼写错误时根据热词词库推荐词条不准确的弊端,有效地自动纠正用户输入时按错键或输入错序的问题,因而有效地提高了输入法对用户输入纠错的能力,尤其是应对拼音首字母出错的情况。并且本发明提供的方法和系统可以根据用户的意图推荐词条,而非仅依赖于热词词库,提高了用户的输入速度。

[0104] 以上所述仅为本发明的优选实施例,并不用于限制本发明,对于本领域技术人员而言,本发明可以有各种改动和变化。凡在本发明的精神和原理之内所作的任何修改、等同替换、改进等,均应包含在本发明的保护范围之内。

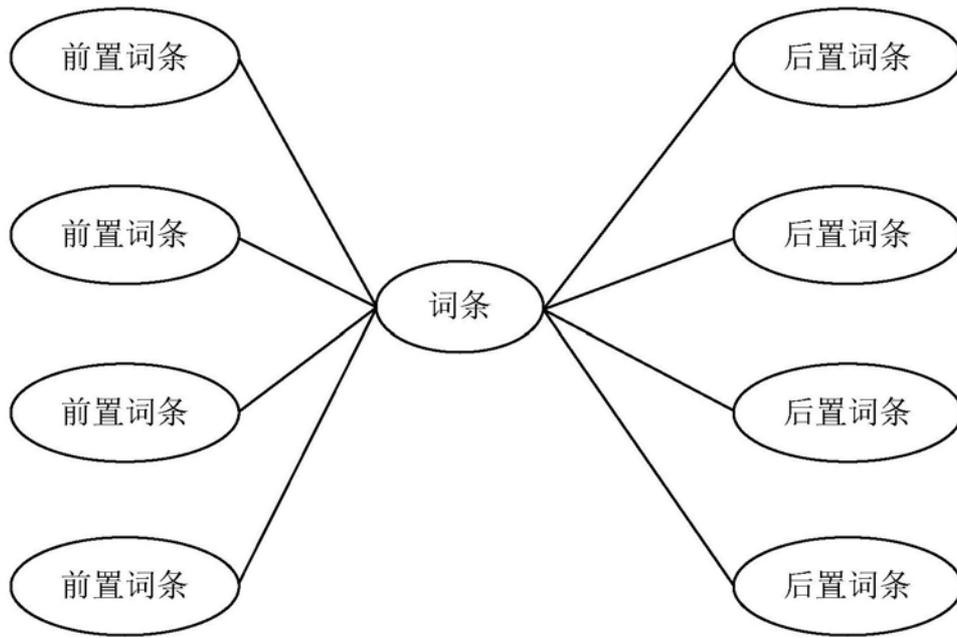


图1

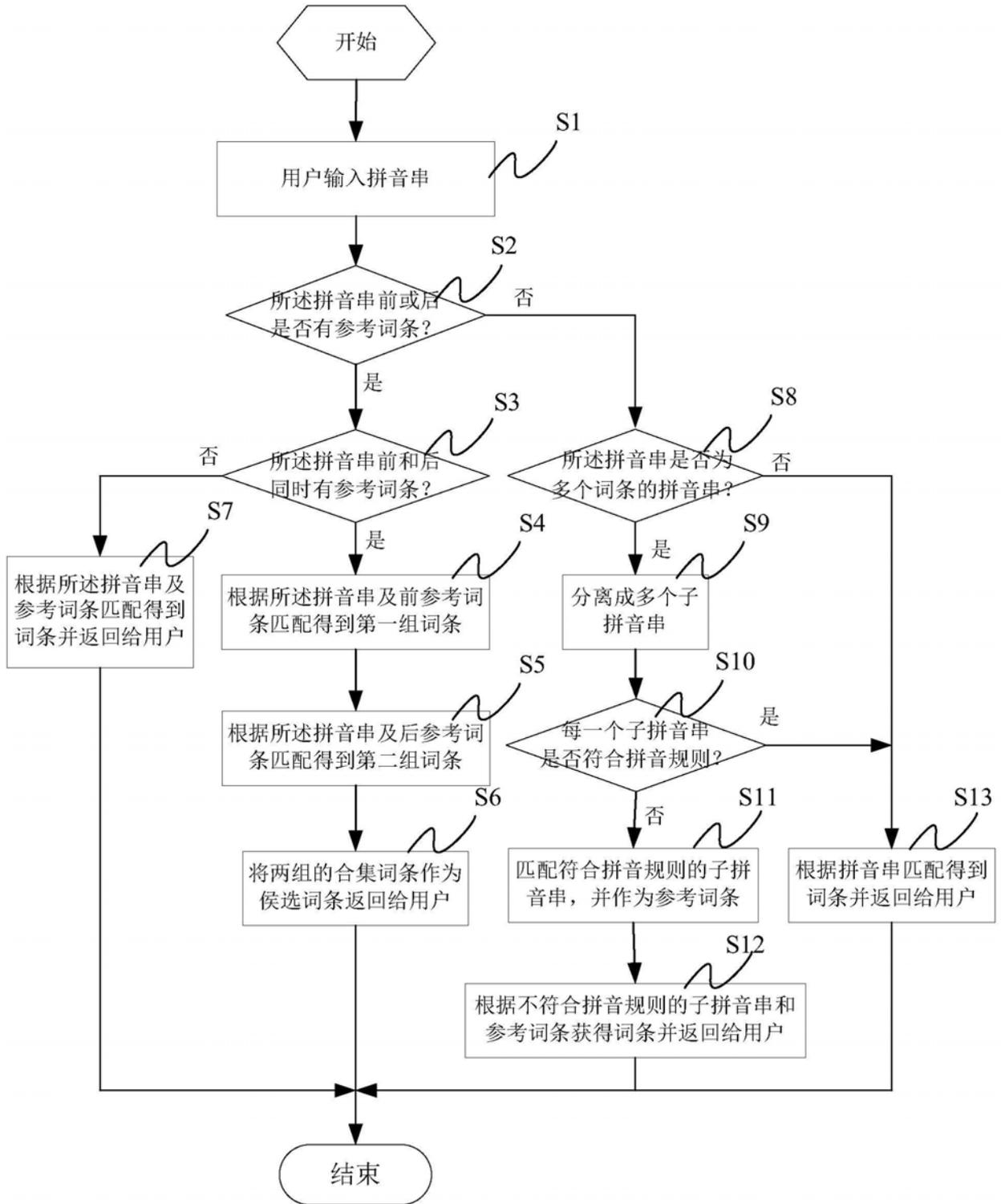


图2

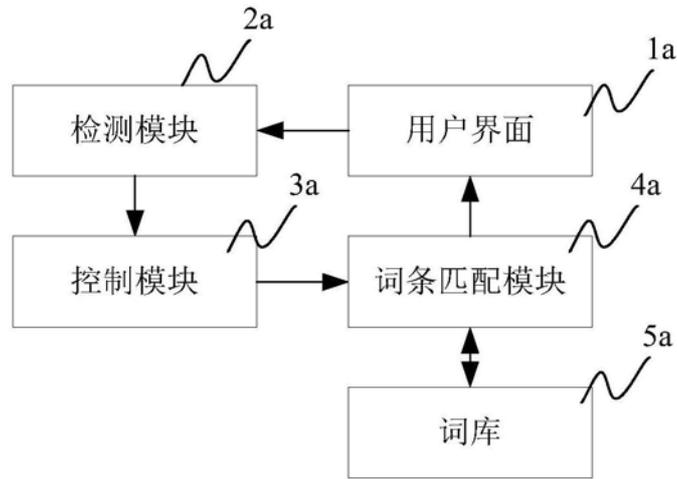


图3

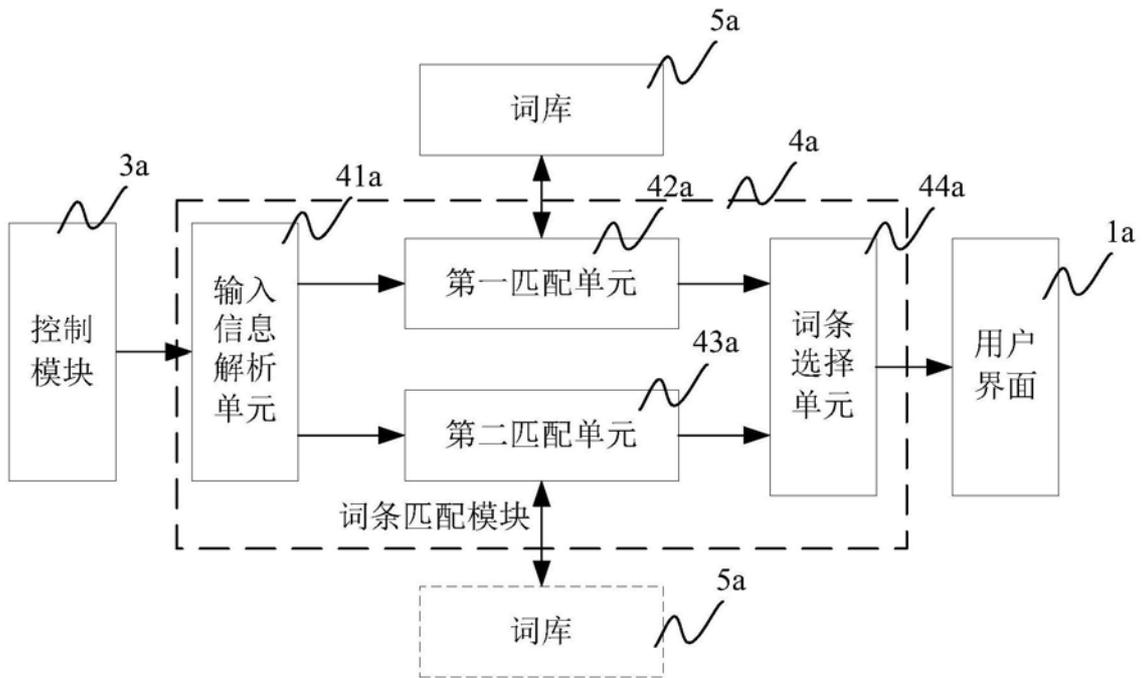
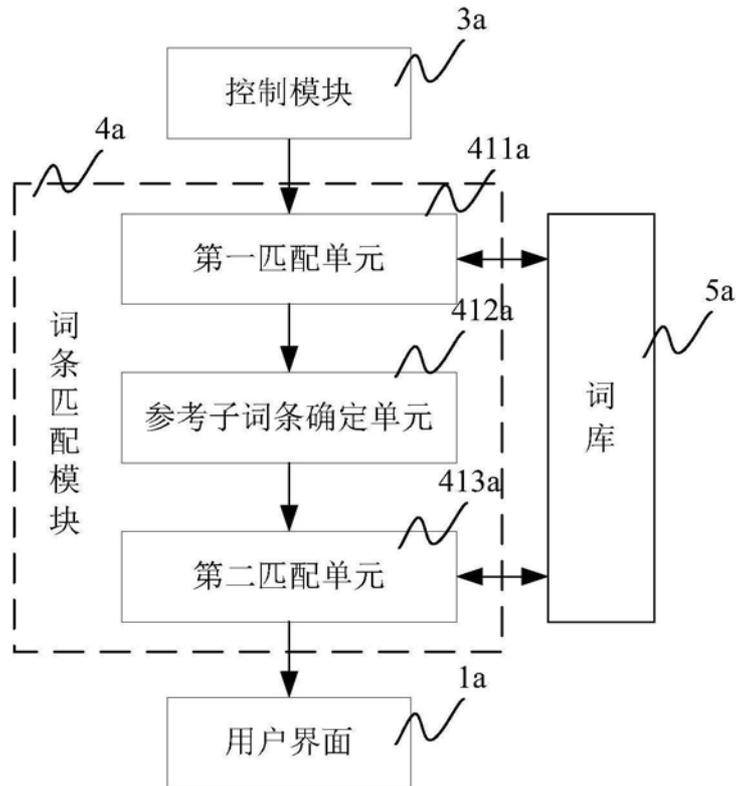
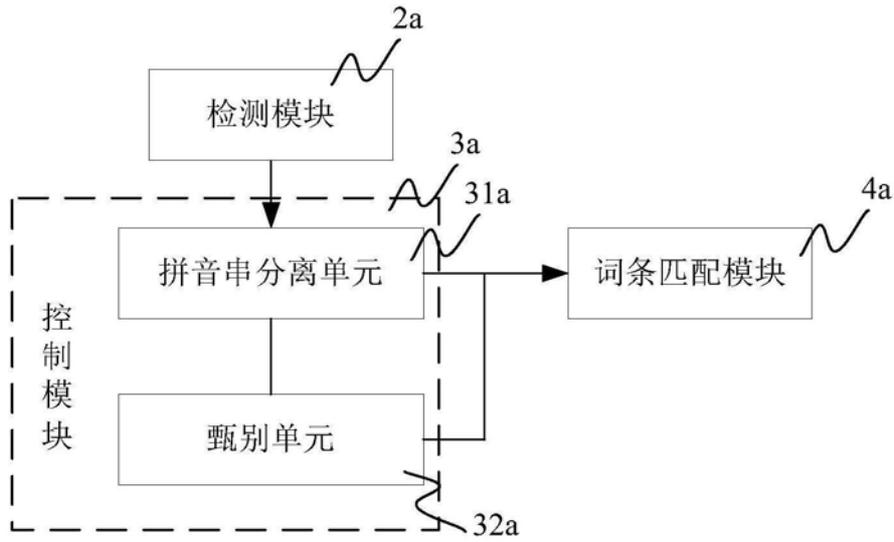


图4



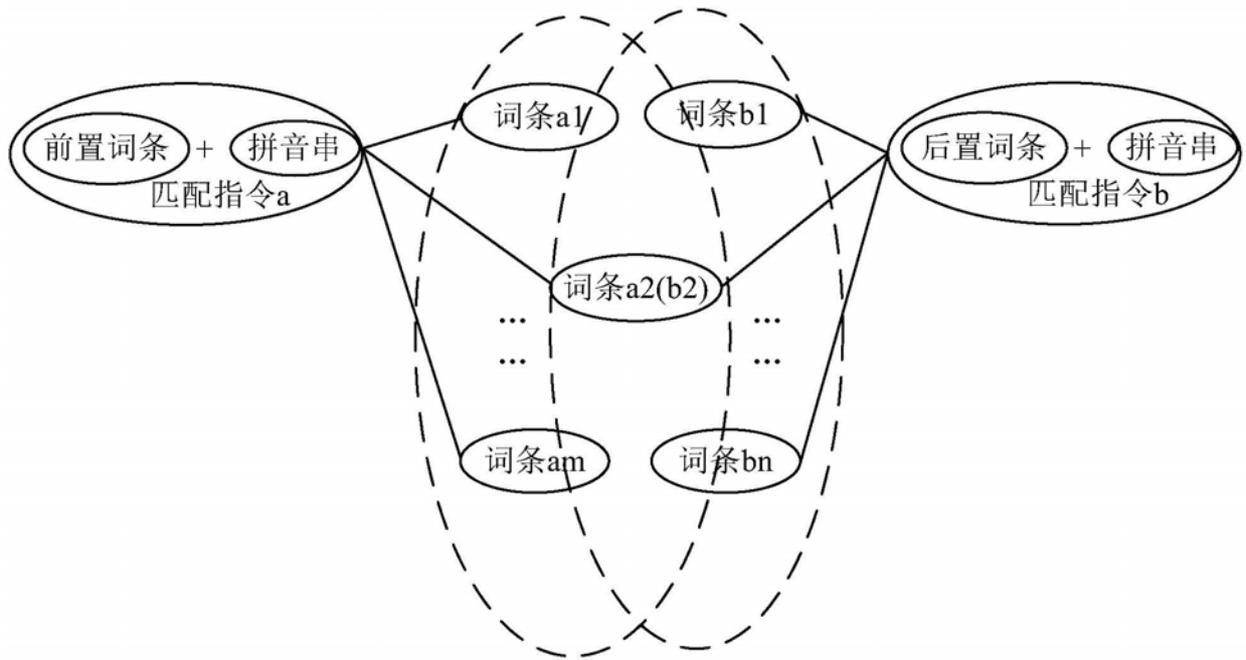


图7

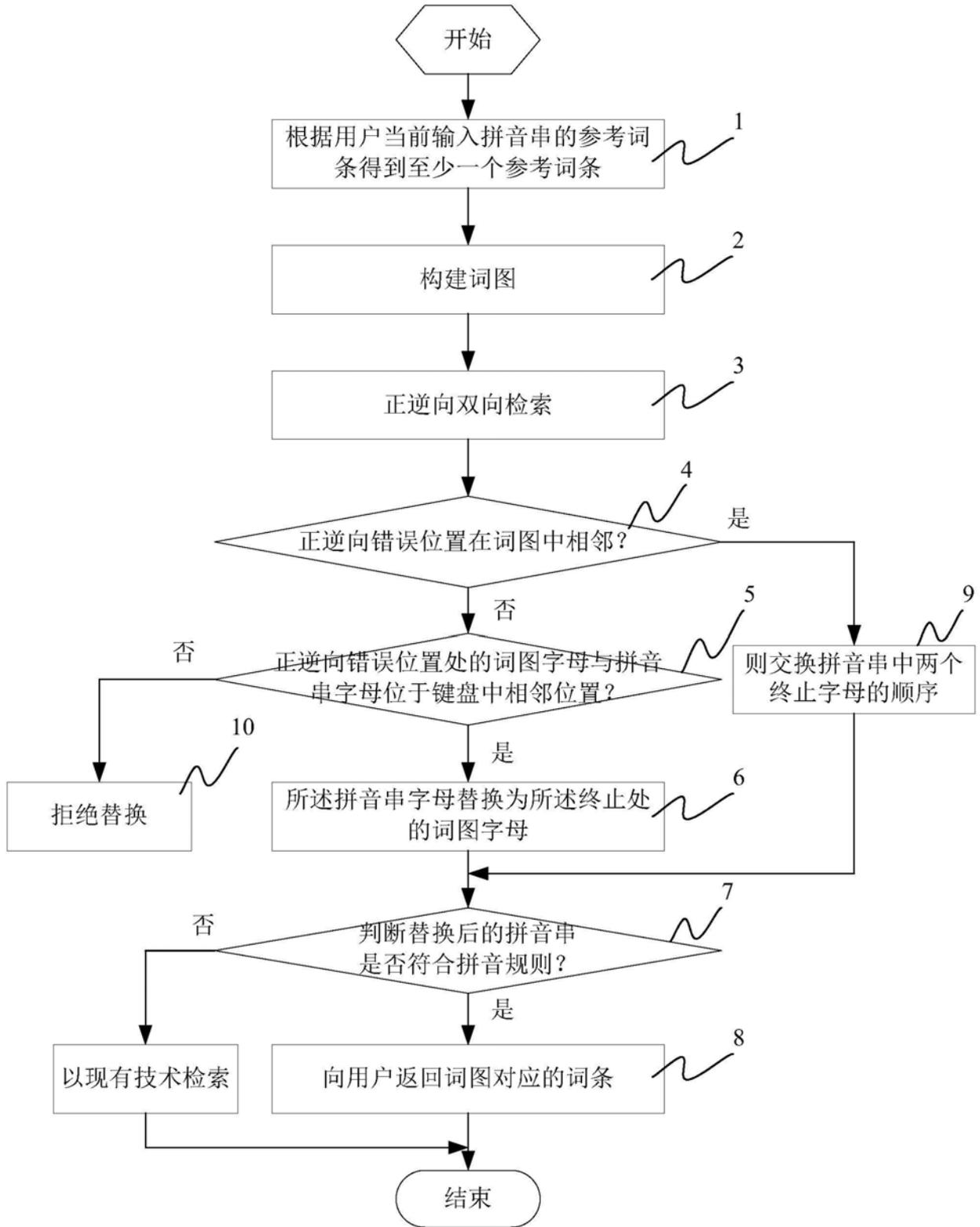


图8

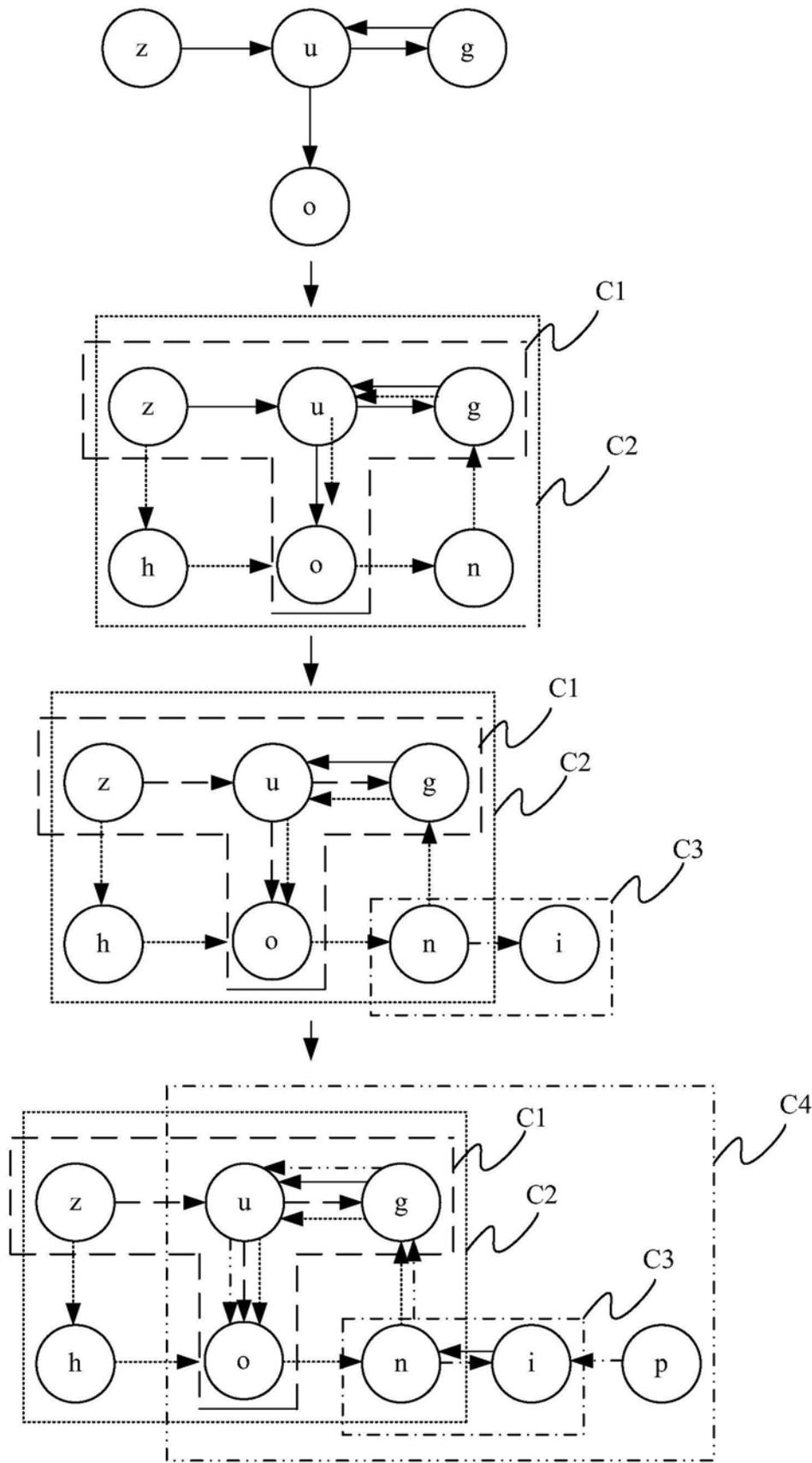


图9

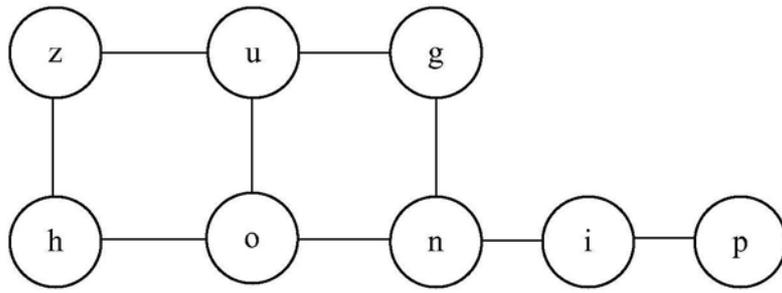


图10

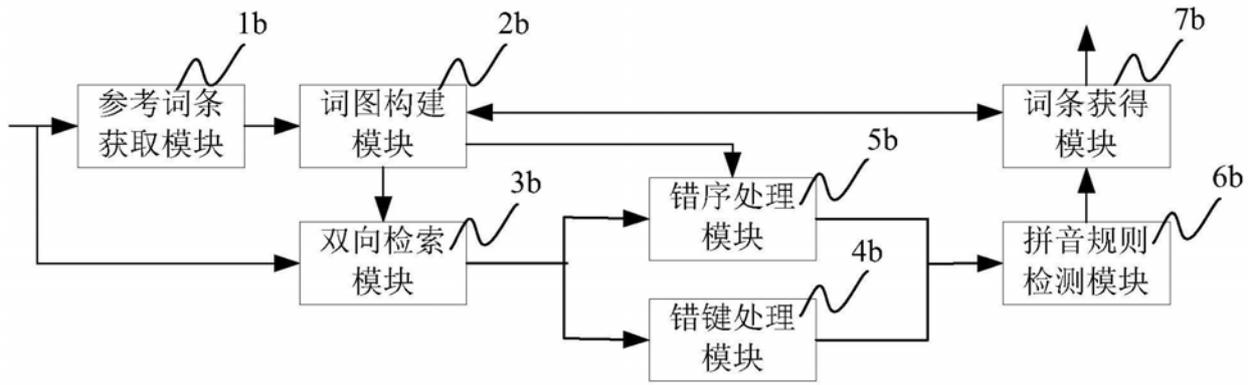


图11