



(12) 发明专利申请

(10) 申请公布号 CN 104345899 A

(43) 申请公布日 2015. 02. 11

(21) 申请号 201310344128. 5

(22) 申请日 2013. 08. 08

(71) 申请人 阿里巴巴集团控股有限公司

地址 英属开曼群岛大开曼资本大厦一座四
层 847 号邮箱

(72) 发明人 杨扬

(74) 专利代理机构 北京国昊天诚知识产权代理
有限公司 11315

代理人 许志勇

(51) Int. Cl.

G06F 3/023(2006. 01)

G06F 17/30(2006. 01)

G06F 17/27(2006. 01)

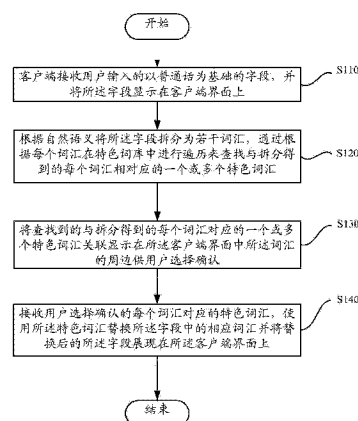
权利要求书3页 说明书11页 附图7页

(54) 发明名称

用于输入法的字段转换方法及客户端

(57) 摘要

本申请公开了用于输入法的字段转换方法及客户端,其中字段转换方法包括:接收输入的以普通话为基础的字段并将字段显示在客户端界面上;根据自然语义将字段拆分为若干词汇,通过根据每个词汇在特色词库中进行遍历来查找与拆分得到的每个词汇相对应的特色词汇,特色词库以字符串方式对应保存以普通话为基础的词汇及相应的特色词汇;将查找到的与拆分得到的每个词汇对应的特色词汇关联显示在客户端界面中词汇的周边供选择确认;接收选择确认的每个词汇对应的特色词汇,使用特色词汇替换字段中的相应词汇并将替换后的字段展现在客户端界面上。本申请能够以普通话词汇输入,输出特色词汇,方便任何不熟悉特色词汇或非普通话输入法的人书写特色词汇。



1. 一种用于输入法的字段转换方法,其特征在于,包括:

接收输入的以普通话为基础的字段,并将所述字段显示在客户端界面上;

根据自然语义将所述字段拆分为若干词汇,通过根据每个词汇在特色词库中进行遍历来查找与拆分得到的每个词汇相对应的一个或多个特色词汇,所述特色词库以字符串方式对应保存以普通话为基础的词汇及相应的一个或多个特色词汇;

将查找到的与拆分得到的每个词汇对应的一个或多个特色词汇关联显示在所述客户端界面中所述词汇的周边供选择确认;

接收选择确认的每个词汇对应的特色词汇,使用所述特色词汇替换所述字段中的相应词汇并将替换后的所述字段展现在所述客户端界面上。

2. 如权利要求1所述的用于输入法的字段转换方法,其特征在于,

将查找到的与拆分得到的每个词汇对应的一个或多个特色词汇关联显示在所述客户端界面中所述词汇的周边供选择确认的步骤,进一步包括:

通过根据每个词汇在特色词库中进行遍历来查找与拆分得到的每个词汇相对应的多个特色词汇时,同时还读取所述多个特色词汇中每个特色词汇的使用热度;

将查找到的与拆分得到的每个词汇对应的一个或多个特色词汇关联显示在所述客户端界面中所述词汇的周边供选择确认的步骤,进一步包括:

将查找到的与拆分得到的每个词汇对应的多个特色词汇按照对应的使用热度由高到低的排列顺序关联显示在所述客户端界面中所述词汇的周边供选择确认。

3. 如权利要求2所述的用于输入法的字段转换方法,其特征在于,通过根据每个词汇在特色词库中进行遍历来查找与拆分得到的每个词汇相对应的多个特色词汇时,同时还读取所述多个特色词汇中每个特色词汇的使用热度的步骤前,还包括:

在距当前时刻的一总时长区域内,划分以当前时刻为起点的多个时长范围;

按所述时长范围的时长从短到长的顺序为每个时长范围赋予一权重值,所述时长范围的时长最短的权重值最高;

在每个时长范围内,分别统计每个特色词汇被选择确认的次数以及所有特色词汇被选择确认的次数;

根据每个特色词汇被选择确认的次数以及所有特色词汇被选择确认的次数,计算在每个时长范围内每个特色词汇被选择确认的次数所占的比重,作为每个特色词汇被选择确认的概率;

针对每个特色词汇,在各个时长范围被选择确定的概率及相应时长范围的权重,统计出每个特色词汇在总时长区域内被选择确认的概率作为每个特色词汇的使用热度并进行记录以供读取。

4. 如权利要求1所述的用于输入法的字段转换方法,其特征在于,客户端接收输入的以普通话为基础的字段的步骤前,还包括:

获取所述客户端的位置信息;

根据获取的所述位置信息查找相应的特色词库,比较所述位置信息相应的特色词库与上一次使用的特色词库;

如果不同,将所述位置信息相应的特色词库、上一次使用的特色词库及用户偏好设置的特色词库提供给用户供选择;如果相同,将所述位置信息相应的特色词库及用户偏好设

置的特色词库提供给用户供选择；

切换到用户选择的特色词库。

5. 如权利要求 1 所述的用于输入法的字段转换方法,其特征在于,

接收输入的以普通话为基础的字段的步骤,进一步包括:

自动记录输入以普通话为基础的字段时的输入的拼音;

根据每个词汇在特色词库中进行遍历的步骤,进一步包括:

在查找到对应的一个或多个特色词汇后,还根据所述输入的拼音与所述特色词库中保存的拼音通过模糊比对进行验证确认,如果所述输入的拼音与所述特色词库中保存的拼音相同或相近,则验证确认通过;其中,所述特色词库还保存有与以普通话为基础的词汇、相应的一个或多个特色词汇相对应的以普通话为基础的词汇的拼音。

6. 如权利要求 1 所述的用于输入法的字段转换方法,其特征在于,

一个特色词库中所保存的特色词汇,为与相应以普通话为基础的词汇的字音相同、字面不同的同一类特色词汇;

所述特色词汇包括以下至少一种:方言词汇、网络用语。

7. 一种用于输入法字段转换的客户端,其特征在于,包括:

本地接收模块,用于接收输入的以普通话为基础的字段;

词汇拆分模块,用于根据自然语义将所述字段拆分为若干词汇;

上报查询模块,用于分别将每个词汇上报到云端服务器从而通过根据每个词汇在所述云端服务器的特色词库中进行遍历来查找与拆分得到的每个词汇相对应的一个或多个特色词汇,所述特色词库以字符串方式对应保存以普通话为基础的词汇及相应的一个或多个特色词汇;

下发接收模块,用于接收所述云端服务器下发的查找到的与拆分得到的每个词汇对应的一个或多个特色词汇;

交互展现模块,用于将所述字段显示在客户端界面上,还用于将所述一个或多个特色词汇关联显示在所述客户端界面中所述词汇的周边供选择确认,接收选择确认的每个词汇对应的特色词汇,使用所述特色词汇替换所述字段中的相应词汇并将替换后的所述字段展现在所述客户端界面上。

8. 如权利要求 7 所述的客户端,其特征在于,

所述下发接收模块,还用于在接收所述云端服务器下发的查找到的与拆分得到的每个词汇对应的多个特色词汇时,接收所述多个特色词汇中每个特色词汇的使用热度;

所述交互展现模块,进一步用于将查找到的与拆分得到的每个词汇对应的多个特色词汇按照相应的使用热度由高到低的排列顺序关联显示在所述客户端界面中所述词汇的周边供选择确认。

9. 如权利要求 8 所述的客户端,其特征在于,

所述下发接收模块,还用于接收所述云端服务器下发的在所述特色词库中查找到的每个特色词汇对应的使用热度,其中,所述云端服务器在距当前时刻的一总时长区域内,划分以当前时刻为起点的多个时长范围;按所述时长范围的时长从短到长的顺序为每个时长范围赋予一权重值,所述时长范围的时长最短的权重值最高;在每个时长范围内,分别统计每个特色词汇被选择确认的次数以及所有特色词汇被选择确认的次数;根据每个特色词汇被

选择确认的次数以及所有特色词汇被选择确认的次数,计算在每个时长范围内每个特色词汇被选择确认的次数所占的比重,即为每个特色词汇被选择确认的概率;针对每个特色词汇,在各个时长范围被选择确定的概率及相应时长范围的权重,统计出每个特色词汇在总时长区域内被选择确认的概率作为每个特色词汇的使用热度并记录保存在所述特色词库中。

10. 如权利要求 7 所述的客户端,其特征在于,还包括:定位模块和词库配置模块;

所述定位模块,用于获取所述客户端的位置信息;

所述上报查询模块,用于将所述定位模块获取的位置信息上报到所述云端服务器从而查找相应的特色词库,还用于将所述交互展现模块发来的选择的特色词库的标识上报给所述云端服务器从而使所述云端服务器切换到所述选择的特色词库;

所述下发接收模块,用于接收所述云端服务器下发的所述位置信息相应的特色词库的标识;

所述词库配置模块,用于比较所述位置信息相应的特色词库的标识与上一次使用的特色词库的标识,如果不同,将所述位置信息相应的特色词库的标识、上一次使用的特色词库的标识及用户偏好设置的特色词库的标识发送给所述交互展现模块;如果相同,将所述位置信息相应的特色词库的标识及用户偏好设置的特色词库的标识发送给所述交互展现模块;

所述交互展现模块,用于将所述词库配置模块发来的特色词库的标识提供给用户供选择,并将用户选择的特色词库的标识发送给所述上报查询模块。

11. 如权利要求 7 所述的客户端,其特征在于,

所述本地接收模块,还用于记录在输入以普通话为基础的字段时输入的拼音并通知所述上报查询模块;

所述上报查询模块,还用于将所述本地接收模块发来的所述输入的拼音上报到云端服务器,在所述云端服务器通过以普通话为基础的词汇及特色词汇的对应关系查找到相应的特色词汇后,还将所述输入的拼音与在所述特色词库中对应保存的拼音进行模糊比对,当比对结果相同或相近时向所述下发接收模块下发所查找到的特色词汇,其中,所述特色词库还保存有与以普通话为基础的词汇、相应的一个或多个特色词汇相对应的以普通话为基础的词汇的拼音。

12. 如权利要求 7 所述的客户端,其特征在于,

一个特色词库中所保存的特色词汇,为与相应以普通话为基础的词汇的字音相同、字面不同的同一类特色词汇;

所述特色词汇包括以下至少一种:方言词汇、网络用语。

用于输入法的字段转换方法及客户端

技术领域

[0001] 本申请涉及计算机领域,具体地说,涉及一种用于输入法的字段转换方法及客户端。

背景技术

[0002] 目前,人们通过电子设备进行沟通时,依然以文本输入为主流,除了输入普通话文本以外,针对用户希望输入方言的需求,出现了方言输入法,主要解决的是不同方言咬字发音的差异所导致的拼音差异。比如上海吴语输入法,就是利用了这种方言的独特发音新创方言拼音来输入,用户可以使用上海吴语输入法“linfhakqin”,这个输入在普通话为基础的输入法中是没有含义的,但是在上海吴语输入法中,根据上海吴语的独特发音,它的意思为“拎勿清”,意思为“搞不清楚重点”。

[0003] 但是这种方言输入法只能为熟悉此种方言及这种方言输入法的用户使用,无法由普通话输入法而得到方言词汇,并且这种输入法只能为用户提供方言输出,不能提供其他特色词汇如网络语言的输出。

[0004] 与此同时,现在更多的用户更加追求输入法的趣味性,希望能够输入带有方言味道的普通话,用户希望在普通话输入法中输入基于普通话的文字,并希望在表达基本意思的前提下,以“带有方言的字音”来输出词汇,这些输出的文字字面意思虽然不求准确,但是字音完全可以使用户能准确理解语言含义,同时还带有亲切感。例如:以普通话输入“生命是多么的辉煌,生命是如此的精彩”,而输出的是:“绳命是刹么的回晃,绳命是入刺的井猜”,输出的文字本身的字形和在普通话中的字意已经变化,但是其读音和输入的普通话接近,使读者能够通过字音读出其含义,饶有趣味。但目前还有没输入法可以完成这个功能。

发明内容

[0005] 有鉴于此,本申请所要解决的技术问题是提供了一种用于输入法的字段转换方法及客户端,以解决用户希望以普通话词汇输入,以特色词汇输出(包括方言、网络语言、同音异形、近音异形等等特色词汇输出)的需求。

[0006] 为了解决上述技术问题,本申请公开了一种用于输入法的字段转换方法,包括:接收输入的以普通话为基础的字段,并将所述字段显示在客户端界面上;根据自然语义将所述字段拆分为若干词汇,通过根据每个词汇在特色词库中进行遍历来查找与拆分得到的每个词汇相对应的一个或多个特色词汇,所述特色词库以字符串方式对应保存以普通话为基础的词汇及相应的一个或多个特色词汇;将查找到的与拆分得到的每个词汇对应的一个或多个特色词汇关联显示在所述客户端界面中所述词汇的周边供选择确认;接收选择确认的每个词汇对应的特色词汇,使用所述特色词汇替换所述字段中的相应词汇并将替换后的所述字段展现在所述客户端界面上。

[0007] 进一步地,将查找到的与拆分得到的每个词汇对应的一个或多个特色词汇关联显示在所述客户端界面中所述词汇的周边供选择确认的步骤,包括:通过根据每个词汇在特

色词库中进行遍历来查找与拆分得到的每个词汇相对应的多个特色词汇时,同时还读取所述多个特色词汇中每个特色词汇的使用热度;将查找到的与拆分得到的每个词汇对应的一个或多个特色词汇关联显示在所述客户端界面中所述词汇的周边供选择确认的步骤,包括:将查找到的与拆分得到的每个词汇对应的多个特色词汇按照对应的使用热度由高到低的排列顺序关联显示在所述客户端界面中所述词汇的周边供选择确认。

[0008] 进一步地,通过根据每个词汇在特色词库中进行遍历来查找与拆分得到的每个词汇相对应的多个特色词汇时,同时还读取所述多个特色词汇中每个特色词汇的使用热度的步骤前,还包括:在距当前时刻的一总时长区域内,划分以当前时刻为起点的多个时长范围;按所述时长范围的时长从短到长的顺序为每个时长范围赋予一权重值,所述时长范围的时长最短的权重值最高;在每个时长范围内,分别统计每个特色词汇被选择确认的次数以及所有特色词汇被选择确认的次数;根据每个特色词汇被选择确认的次数以及所有特色词汇被选择确认的次数,计算在每个时长范围内每个特色词汇被选择确认的次数所占的比重,作为每个特色词汇被选择确认的概率;针对每个特色词汇,在各个时长范围被选择确定的概率及相应时长范围的权重,统计出每个特色词汇在总时长区域内被选择确认的概率作为每个特色词汇的使用热度并进行记录以供读取。

[0009] 进一步地,客户端接收输入的以普通话为基础的字段的步骤前,还包括:获取所述客户端的位置信息;根据获取的所述位置信息查找相应的特色词库,比较所述位置信息相应的特色词库与上一次使用的特色词库;如果不同,将所述位置信息相应的特色词库、上一次使用的特色词库及用户偏好设置的特色词库提供给用户供选择;如果相同,将所述位置信息相应的特色词库及用户偏好设置的特色词库提供给用户供选择;切换到用户选择的特色词库。

[0010] 进一步地,接收输入的以普通话为基础的字段的步骤,包括:自动记录输入以普通话为基础的字段时的输入的拼音;根据每个词汇在特色词库中进行遍历的步骤,包括:在查找到对应的一个或多个特色词汇后,还根据所述输入的拼音与所述特色词库中保存的拼音通过模糊比对进行验证确认,如果所述输入的拼音与所述特色词库中保存的拼音相同或相近,则验证确认通过;其中,所述特色词库还保存有与以普通话为基础的词汇、相应的一个或多个特色词汇相对应的以普通话为基础的词汇的拼音。

[0011] 进一步地,一个特色词库中所保存的特色词汇,为与相应以普通话为基础的词汇的字音相同、字面不同的同一类特色词汇;所述特色词汇包括以下至少一种:方言词汇、网络用语。

[0012] 为了解决上述技术问题,本申请公开了一种用于输入法字段转换的客户端,包括:本地接收模块,用于接收输入的以普通话为基础的字段;词汇拆分模块,用于根据自然语义将所述字段拆分为若干词汇;上报查询模块,用于分别将每个词汇上报到云端服务器从而通过根据每个词汇在所述云端服务器的特色词库中进行遍历来查找与拆分得到的每个词汇相对应的一个或多个特色词汇,所述特色词库以字符串方式对应保存以普通话为基础的词汇及相应的一个或多个特色词汇;下发接收模块,用于接收所述云端服务器下发的查找到的与拆分得到的每个词汇对应的一个或多个特色词汇;交互展现模块,用于将所述字段显示在客户端界面上,还用于将所述一个或多个特色词汇关联显示在所述客户端界面中所述词汇的周边供选择确认,接收选择确认的每个词汇对应的特色词汇,使用所述特色词汇

替换所述字段中的相应词汇并将替换后的所述字段展现在所述客户端界面上。

[0013] 进一步地,所述下发接收模块,还用于在接收所述云端服务器下发的查找到的与拆分得到的每个词汇对应的多个特色词汇时,接收所述多个特色词汇中每个特色词汇的使用热度;所述交互展现模块,进一步用于将查找到的与拆分得到的每个词汇对应的多个特色词汇按照相应的使用热度由高到低的排列顺序关联显示在所述客户端界面中所述词汇的周边供选择确认。

[0014] 进一步地,所述下发接收模块,还用于接收所述云端服务器下发的在所述特色词库中查找到的每个特色词汇对应的使用热度,其中,所述云端服务器在距当前时刻的一总时长区域内,划分以当前时刻为起点的多个时长范围;按所述时长范围的时长从短到长的顺序为每个时长范围赋予一权重值,所述时长范围的时长最短的权重值最高;在每个时长范围内,分别统计每个特色词汇被选择确认的次数以及所有特色词汇被选择确认的次数;根据每个特色词汇被选择确认的次数以及所有特色词汇被选择确认的次数,计算在每个时长范围内每个特色词汇被选择确认的次数所占的比重,即为每个特色词汇被选择确认的概率;针对每个特色词汇,在各个时长范围被选择确定的概率及相应时长范围的权重,统计出每个特色词汇在总时长区域内被选择确认的概率作为每个特色词汇的使用热度并记录保存在所述特色词库中。

[0015] 进一步地,还包括:定位模块和词库配置模块;所述定位模块,用于获取所述客户端的位置信息;所述上报查询模块,用于将所述定位模块获取的位置信息上报到所述云端服务器从而查找相应的特色词库,还用于将所述交互展现模块发来的选择的特色词库的标识上报给所述云端服务器从而使所述云端服务器切换到所述选择的特色词库;所述下发接收模块,用于接收所述云端服务器下发的所述位置信息相应的特色词库的标识;所述词库配置模块,用于比较所述位置信息相应的特色词库的标识与上一次使用的特色词库的标识,如果不同,将所述位置信息相应的特色词库的标识、上一次使用的特色词库的标识及用户偏好设置的特色词库的标识发送给所述交互展现模块;如果相同,将所述位置信息相应的特色词库的标识及用户偏好设置的特色词库的标识发送给所述交互展现模块;所述交互展现模块,用于将所述词库配置模块发来的特色词库的标识提供给用户供选择,并将用户选择的特色词库的标识发送给所述上报查询模块。

[0016] 进一步地,所述本地接收模块,还用于记录在输入以普通话为基础的字段时输入的拼音并通知所述上报查询模块;所述上报查询模块,还用于将所述本地接收模块发来的所述输入的拼音上报到云端服务器,在所述云端服务器通过以普通话为基础的词汇及特色词汇的对应关系查找到相应的特色词汇后,还将所述输入的拼音与在所述特色词库中对应保存的拼音进行模糊比对,当比对结果相同或相近时向所述下发接收模块下发所查找到的特色词汇,其中,所述特色词库还保存有与以普通话为基础的词汇、相应的一个或多个特色词汇相对应的以普通话为基础的词汇的拼音。

[0017] 进一步地,一个特色词库中所保存的特色词汇,为与相应以普通话为基础的词汇的字音相同、字面不同的同一类特色词汇;所述特色词汇包括以下至少一种:方言词汇、网络用语。

[0018] 本申请与现有的方案相比,本申请所获得的技术效果有,能够以普通话词汇输入,输出特色词汇,方便任何不熟悉特色词汇或非普通话输入法的人书写特色词汇;能够依据

不同特色词汇的被选择概率,向用户提供有顺序的方便快捷的特色词汇选择方式,加快用户输入速度;能够依据选定的时间段,有侧重地提供所述时间段常用的特色词汇,实现输出特色词汇带有时间特征;能够依据选定的地理位置,有侧重地提供所述地理位置对应的常用的特色词汇,实现输出词汇带有地理特征,能够通过用户本人对应的特色词汇库方便地维持自己的语言风格;能够依据云端用户全体的使用概率选择流行的特色词汇,提升了用户语言的流行性、娱乐性等等效果。

[0019] 当然,实施本申请的任一产品必不一定需要同时达到以上所述的所有技术效果。

附图说明

[0020] 此处所说明的附图用来提供对本申请的进一步理解,构成本申请的一部分,本申请的示意性实施例及其说明用于解释本申请,并不构成对本申请的不当限定。在附图中:

[0021] 图 1 是本申请实施例的一种用于输入法的字段转换方法的流程框图。

[0022] 图 2a、2b 是本申请实施例图 1 中步骤 S130 的具体方法流程框图。

[0023] 图 3 是本申请实施例的一种使用地理位置信息介入用户对特色词汇库的选择的用于输入法的字段转换方法的流程框图。

[0024] 图 4 是本申请实施例的一种应用在本地和云端的用于输入法的字段转换方法流程框图。

[0025] 图 5 是本申请实施例的一种用于输入法的字段转换的客户端 500 的结构框图。

[0026] 图 6 是本申请实施例的一种包括定位模块 560 及词库配置模块 570 的用于输入法的字段转换的客户端 600 的结构框图。

具体实施方式

[0027] 为使本申请的目的、技术方案和优点更加清楚,下面将结合本申请具体实施例及相应的附图对本申请技术方案进行清楚、完整地描述。显然,所描述的实施例仅是本申请一部分实施例,而不是全部的实施例。基于本申请中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都属于本申请保护的范围。

[0028] 参考图 1,图 1 是本申请实施例的一种用于输入法的字段转换方法的流程框图,其包括步骤 S110~步骤 S140,其中:

[0029] 在步骤 S110 处:客户端接收用户输入的以普通话为基础的字段,并将所述字段显示在客户端界面上。

[0030] 具体而言,所述客户端可以位于普通用户使用的个人电脑、平板电脑、手机、PDA 等等设备上,这种普通用户的设备可以称为本地,所述客户端可以为本地的一个程序、应用、插件等等,其可以有形地出现在用户界面,也可以无形地附着在其它程序应用上。当用户需要输入一段文字时,一般会将光标落在一个可以输入文本的位置,某种中文输入法被启动,所述中文输入法可以为笔画、拼音、手写等等方式的输入法,用户凭借所述中文输入法可输入一段以普通话为基础的字段。所述普通话为基础的含义为,其主体结构是普通话,一般而言,主体结构中的词语在普通话词典中对应的语义即为用户想表达的意思,除主体结构外,字段中可以夹杂英文、专有名词、表情、符号等等内容。所述客户端接收所述字段,并且用户输入的所述字段显示在客户端的界面上,这时用户可以看见自己输入的字段,所述客户端

界面包括客户端附着的其它程序应用的界面。

[0031] 在步骤 S120 处:根据自然语义将所述字段拆分为若干词汇,通过根据每个词汇在特色词库中进行遍历来查找与拆分得到的每个词汇相对应的一个或多个特色词汇,所述特色词库以字符串方式对应保存以普通话为基础的词汇及相应的一个或多个特色词汇。

[0032] 具体而言,一般一个字段包含若干词汇,每一个词汇包含其独立的意思,一个字段的的意思即由所有词汇的独立的意思结合语法组合而成。在这个步骤中,本方法利用分词技术将所述字段拆分为所述若干词汇,尽可能使得拆分后的词汇的自然语义和其在字段中的自然语义相同或相近。所述分词技术为现有技术,其包括字符串匹配的分词方法、基于理解的分词方法、基于统计的分词方法等等,并且在不断地发展中,本方法并不限制分词所采用的技术方案,并可以采用现有的或未来出现的分词技术。例如,用户输入“生命是多么的辉煌”这一字段,在本步骤中,可用分词技术将其分为“生命”、“是”、“多么”、“的”、“辉煌”这些词汇。这种分词的效果是,尽可能地将较长的字段分为较短的具有独立意思的词汇,这样更易于在特色词库中查找其到对应的特色词汇,并且,还可以分别向用户提供多个被拆分的词汇的特色词汇选择,丰富了用户对特色词汇的可选择性。

[0033] 将字段拆分为若干词汇后,将每个词汇在特色词库中进行遍历查找。所述特色词库可以包括若干词条,每一词条中有一个以普通话为基础的词汇,并有与其对应的一个或多个特色词汇。所述特色词汇可以为所述普通话词汇对应的一种方言,例如,“你”、“喜欢”、“吃”、“什么”在对应的上海话词汇分别为:“侬”、“欢喜”、“切”、“萨”;所述特色词汇还可以是网络语言,例如“蟑螂”对应的网络语言词汇可以是“小强”,“同学”对应的网络语言词汇可以是“童鞋”。

[0034] 优选的,所述特色词库中所保存的特色词汇,可以为与以普通话为基础的词汇的字音相同或相近、字面不同的同一类特色词汇。例如:以普通话为基础的词汇分别为“生命”、“多么”、“辉煌”对应的特色词汇可以分别是“绳命”、“剁么”、“回晃”,这三个特色词汇与以普通话为基础的词汇发音相近,字面不同,当字段“生命是多么的辉煌”被特色词汇取代为“绳命是剁么的回晃”时,通过其读音,读者可以猜到其本来的意思“生命是多么的辉煌”,并由于读音不完全是本来的读音,而产生了一种趣味。这种字音相同或相近、字面不同的同一类特色词汇也可以包括:方言词汇、网络用语等。例如“什么”对应的网络用语是“神马”。

[0035] 特别地,对于与以普通话为基础的词汇的字音相同或相近、字面不同的同一类特色词汇,他们之间的字音相同或相近,所以应该对应相似或相近的拼音。因此所述特色词库中除了以字符串方式对应保存以普通话为基础的词汇及相应的一个或多个特色词汇之外,还将以普通话为基础的词汇的拼音与以普通话为基础的词汇、相应的一个或多个特色词汇对应保存,这个拼音可以用于之后在步骤 S120 中查找到特色词汇后对于查找的准确性进行验证。例如,如果用户是通过输入拼音来输入以普通话为基础的字段时,在步骤 S110 中客户端还自动记录所述拼音,在步骤 S120 中通过以普通话为基础的词汇及特色词汇的对应关系查找到相应的特色词汇后,还获取对应保存的拼音,并与通过步骤 S110 自动记录的拼音进行模糊比对,如果两者相同或相近,则验证通过。

[0036] 在步骤 S130 处:将查找到的与拆分得到的每个词汇对应的一个或多个特色词汇关联显示在所述客户端界面中所述词汇的周边供用户选择确认。

[0037] 具体而言,在经过步骤 120 的遍历之后,可以查找到若干词汇对应的一个或多个

特色词汇,然后将所述一个或多个特色词汇关联显示在所述客户端界面中,所述关联显示,即一种可以使用户感知到显示的特色词汇和某一普通话为基础的词汇具有对应关系的显示方式,例如,可以直接显示在以普通话为基础的词汇的上方,或者仅当鼠标靠近某以普通话为基础的词汇时,将特色词汇显示在其上方或旁侧,并可以让用户选择。还可以的关联显示方式是,直接默认地替换以普通话为基础的词汇而显示一个特色词汇,以向用户直接提供默认选择,加快用户选择速度,当用户希望更改特色词汇时,可以再提供其它特色词汇的选择。

[0038] 当然的,有些以普通话为基础的词汇在特色词汇库中并没有对应的特色词汇,对这些词汇就可以不显示特色词汇或提示用户没有特色词汇。

[0039] 在步骤 S140 处:接收用户选择确认的每个词汇对应的特色词汇,使用所述特色词汇替换所述字段中的相应词汇并将替换后的所述字段展现在所述客户端界面上。

[0040] 具体而言,当用户选择了某以普通话为基础的词汇对应的特色词汇后,则隐去所述以普通话为基础的词汇,而显示用户已经选择的特色词汇。当然,还可以提供用户针对此特色词汇的选择,让其回选其在步骤 110 中输入的以普通话为基础的词汇,或者选择其它的与以所述普通话为基础的词汇对应的特色词汇。

[0041] 优选的,在图 1 所述步骤 130 处拆分得到的每个词汇对应多个特色词汇时,还可以在图 1 所述步骤 140 处以特色词汇的使用热度来决定特色词汇的显示方式。如图 2a 所示,图 2a 是本申请实施例图 1 中步骤 S131 是步骤 S130 的一种针对每个词汇对应多个特色词汇时的具体处理步骤,步骤 S141 是步骤 S140 的一种针对每个词汇对应多个特色词汇时显示的处理步骤:

[0042] 在步骤 S131 处,通过根据每个词汇在特色词库中进行遍历来查找与拆分得到的每个词汇相对应的多个特色词汇时,同时还读取所述多个特色词汇中每个特色词汇的使用热度。

[0043] 具体而言,以特色词汇的使用热度来决定特色词汇的显示方式,特色词汇的使用热度可以通过被选择确认的概率的方式来客观衡量。一种方案可以是,特色词汇库始终在本地时,则可以依据此本地的特色词汇库中的特色词汇被选择的历史记录来确定特色词汇的被选择确认的概率;另一种方案可以是,当特色词汇库在本地、网络服务器或在云端时,特色词汇库与用户身份对应,例如,一个用户登录其本地或网络侧、云端的账号后可以与其账号对应的特色词汇库关联,则可以依据此用户对特色词汇选择的历史记录来确定特色词汇的被这个用户选择确认的概率;还有一种方案可以是,当特色词汇库在网络服务器或在云端,并为接入同一网络服务器或者云端的多个用户共同使用的特色词汇库,则可以以多个用户整体对特色词汇选择的历史记录来确定特色词汇的被选择确认的概率。此外,本方法不限制采用其它概率确定方式。

[0044] 在步骤 S141 处,将查找到的与拆分得到的每个词汇对应的多个特色词汇按照对应的使用热度由高到低的排列顺序关联显示在所述客户端界面中所述词汇的周边供用户选择确认。

[0045] 具体而言,如前所述,特色词汇的使用热度通过被选择确认的概率的方式来客观衡量。这种有排列顺序的关联显示可以让用户可最方便地选择高概率的特色词汇。例如可以将概率最高的特色词汇显示在离以普通话为基础的词汇最近处,较低概率的显示在较远

处;还可以将较高概率的特色词汇以饱和度较高的颜色显示,较低概率的特色词汇以较灰暗的颜色显示,等等。此技术方案的效果是,可以方便用户选择使用被选择概率最高的特色词汇,如果用户有自己的语言风格,则通过与用户本人对应的特色词汇库方便地维持自己的语言风格;而当用户希望使用社会性的流行用语时,使用在云端被用户共同使用特色词汇库,即使用户不知道当前流行什么特色词汇,依然可以依据云端用户全体的使用概率选择流行的特色词汇,提升了用户语言的流行性、娱乐性。

[0046] 优选的,步骤 S130 中,在图 2a 显示的步骤 131 之前,云端服务器可以通过如下的方式设计概率,并通过这一概率来反应特色词汇的使用热度,具体而言,如图 2b 所示:

[0047] 在步骤 S1301 处,在距当前时刻的一总时长区域内,划分以当前时刻为起点的多个时长范围;

[0048] 在步骤 S1302 处,按所述时长范围的时长从短到长的顺序为每个时长范围赋予一权重值,所述时长范围的时长最短的权重值最高;

[0049] 在步骤 S1303 处,在每个时长范围内,分别统计每个特色词汇被选择确认的次数以及所有特色词汇被选择确认的次数;

[0050] 在步骤 S1304 处,根据每个特色词汇被选择确认的次数以及所有特色词汇被选择确认的次数,计算在每个时长范围内每个特色词汇被选择确认的次数所占的比重,即为每个特色词汇被选择确认的概率;

[0051] 在步骤 S1305 处,针对每个特色词汇,在各个时长范围被选择确定的概率及相应时长范围的权重,统计出每个特色词汇在总时长区域内被选择确认的概率并进行记录以供读取下发;

[0052] 之后执行步骤 S131 ~ S132。

[0053] 这里需要注意,随着时间的推移,总时长区域和各个时长范围的长度保持不变,各个时长范围的权重也保持不变,但是各个时长范围内的被选择确认的次数以及所有特色词汇被选择确认的次数会随着当前时刻的推移而推移到其他时长范围内,这样每个特色词汇在总时长区域内被选择确认的概率会发生变化,因此随着当前时刻的推移,每个特色词汇在总时长区域内被选择确认的概率会实际进行计算更新。

[0054] 本实施例中,使用总时长区域内特色词汇被选择确认的概率来表示总时长区域内这些特色词汇的热度。对总时长区域内特色词汇的概率来说最重要的是离当前时刻较近的一个时长范围内特色词汇的热度,因为如果特色词汇的热度只是在离当前时刻较远的一个时长范围保持很高的,而在离当前时刻较近的时刻这些特色词汇的热度变得很低,说明这些特色词汇的热度和被选择概率呈现了急剧的下降趋势,但是因为在离当前时刻较远的一个时长范围内特色词汇的热度保持很高,导致总时长区域内的概率还是很高,这显然与现实的概率趋势不符,因此对总时长区域进行细度的划分和并赋予权重,这样可以相对客观地、预见性地反应出总时长区域内特色词汇的被选择确认的概率。

[0055] 下面以一个应用实例来解释上述实施例,但并不对上述实施例做出限定。

[0056] 例如,以“周”为单位,从当前时刻至前 1 周的时长范围(以下称为第一时长范围)内,统计每个特色词汇被选择确认的次数在所有特色词汇被选择确认的次数中所占的概率,并为这个时长范围赋予权重 0.6;

[0057] 从当前时刻至前 2 周的时长范围内(以下称为第二时长范围,这个第二时长范围

包括了第一时长范围),统计每个特色词汇被选择确认的次数在所有特色词汇被选择确认的次数中所占的概率,并为这个第二时长范围赋予权重 0.3。

[0058] 从当前时刻至前 1 个月的时长范围内(以下称为取样时长范围,这个取样时长范围就是概率统计所考虑的总时长区域,这个取样时长范围包括了第一、第二时长范围),统计每个特色词汇被选择确认的次数在所有特色词汇被选择确认的次数中所占的概率,并为这个第三时长范围赋予权重 0.1。

[0059] 分别根据第一时长范围内任一特色词汇(例如“小盆友”这个词汇)被选择确定的概率 0.8 和第一时长范围的权重 0.6,第二时长范围内所述特色词汇“小盆友”被选择确定的概率 0.5 和第二时长范围的权重 0.3,以及取样时长范围内所述特色词汇“小盆友”被选择确定的概率 0.37 和取样时长范围的权重 0.1,加权作和从而得到所述特色词汇“小盆友”在总时长区域内的被选择确定的概率为 $0.8 \times 0.6 + 0.5 \times 0.3 + 0.37 \times 0.1 = 0.667$ 。通过这样的方式获得在总时长区域内每个特色词汇的概率。

[0060] 当用户输入以普通话为基础的词汇后,根据所述词汇查找到对应的多个特色词汇,同时还获取每个特色词汇的概率,根据上述概率的由高到低的顺序对所述多个特色词汇排序。

[0061] 本方法的另一实施例可以使用地理位置信息介入用户对特色词汇的选择。如图 3,图 3 是本申请实施例的一种使用地理位置信息介入用户对特色词汇库的选择的用于输入法的字段转换方法的流程框图。

[0062] 在步骤 S310 处,获取所述客户端的位置信息并上报。具体而言,定位可以使用 GPS 定位技术,或者可以使用互联网定位技术,或者使用移动通信网等等定位技术等等现有及未来的自动定位技术,还可以由用户自己选择其位置。所述上报,可以上报给在本地或者在网络服务器、云端的负责特色词汇库的模块。

[0063] 在步骤 S320 处,根据获取的所述位置信息查找相应的特色词库,比较所述位置信息相应的特色词库与用户上一次使用的特色词库。具体而言,所述位置信息相应的特色词库包括可以为一种方言,例如位置信息显示客户端在四川,则其对应的特色词库可以为四川方言。此外,本发明并不排除其它位置信息与特色词库的对应方式。

[0064] 在步骤 S330 处,如果步骤 320 中的比较结果为不同,将所述位置信息相应的特色词库、用户上一次使用的特色词库及用户偏好设置的特色词库提供给用户供选择;如果比较结果为相同,将所述位置信息相应的特色词库及用户偏好设置的特色词库提供给用户供选择。其中,所述“用户偏好设置的特色词库”为用户事先设定的其偏好的特色词库。

[0065] 在步骤 S340 处,切换到用户选择的特色词库。后续步骤中,将以此用户选择的特色词库为默认使用的特色词库。

[0066] 完成步骤 340 后,继续执行图 1 中的步骤 110-步骤 140。

[0067] 此技术方案的效果是,可以依据用户的位置提供其对应特色词汇库,从而可以使用户随着其处所的变化方便地使用与处所相适应的语言,提升用户的使用感受。

[0068] 本方法的另一种实施例特别针对客户端位于本地、特色词库位于云端服务器的实现方式进行描述。如图 4,图 4 是本申请实施例的一种应用在本地和云端的用于输入法的字段转换方法流程框图,包括步骤 S410-S440;

[0069] 在步骤 S410 处,本地客户端接收用户输入的以普通话为基础的字段,并将所述字

段显示在客户端界面上,并将所述字段上报至云端服务器。所述本地客户端可以为一个程序、应用,或者在程序中的插件,例如在浏览器中嵌入的一个插件。

[0070] 在步骤 S420 处,云端服务器根据自然语义将所述字段拆分为若干词汇,通过根据每个词汇在位于云端服务器的特色词库中进行遍历来查找与拆分得到的每个词汇相对应的一个或多个特色词汇,所述特色词库以字符串方式对应保存以普通话为基础的词汇及相应的一个或多个特色词汇;所述位于云端特色词库可以为对应于一个用户使用的一个特色词库(例如,当用户登录其账户后,关联其对应的特色词库),也可以为对应于公共使用的一个特色词库。

[0071] 在步骤 S430 处,云端服务器将查找到的与拆分得到的每个词汇对应的一个或多个特色词汇下发给本地客户端;

[0072] 在步骤 S440 处,本地客户端将所述一个或多个特色词汇关联显示在所述客户端界面中所述词汇的周边供用户选择确认;

[0073] 在步骤 S450 处,本地客户端接收用户选择确认的每个词汇对应的特色词汇,使用所述特色词汇替换所述字段中的相应词汇并将替换后的所述字段展现在所述客户端界面上。

[0074] 此技术方案的技术效果是,可以缩小本地客户端程序的大小,通过将存储和计算工作置于云端,从而减轻本地的存储消耗、运算消耗,方便用户在不同的本地客户端登录云端,方便地按照自己的偏好使用特色词汇库实现自己的语言风格,或者方便地使用公共特色词汇库使用流行的语言风格。

[0075] 参考图 5,图 5 是本申请实施例的一种用于输入法的字段转换的客户端 500 的结构框图:

[0076] 本地接收模块 510,用于接收用户输入的以普通话为基础的字段,并将所述字段发送给交互展现模块 550。

[0077] 词汇拆分模块 520,用于根据自然语义将所述字段拆分为若干词汇;

[0078] 上报查询模块 530,用于分别将每个词汇上报到云端服务器从而通过根据每个词汇在所述云端服务器的特色词库中进行遍历来查找与拆分得到的每个词汇相对应的一个或多个特色词汇,所述特色词库以字符串方式对应保存以普通话为基础的词汇及相应的一个或多个特色词汇。

[0079] 优选的,一个特色词库中所保存的特色词汇,为与相应以普通话为基础的词汇的字音相同、字面不同的同一类特色词汇;优选的,所述特色词汇包括:方言词汇、网络用语等。

[0080] 特别地,对于与以普通话为基础的词汇的字音相同或相近、字面不同的同一类特色词汇,他们之间的字音相同或相近,所以应该对应相似或相近的拼音。因此所述特色词库除了以字符串方式对应保存以普通话为基础的词汇及相应的一个或多个特色词汇之外,还将以普通话为基础的词汇的拼音与以普通话为基础的词汇、相应的一个或多个特色词汇对应保存,这个拼音可以用于当上报查询模块 530 将每个词汇上报到云端服务器,通过云端服务器查找到特色词汇后,所述云端服务器对于查找的准确性进行自我验证。例如,如果用户是通过输入拼音来输入以普通话为基础的字段时,本地接收模块 510 还用于自动记录所述输入的拼音并通知上报查询模块 530;上报查询模块 530,还将所述本地接收模块 510 发

来的所述拼音上报到云端服务器,在云端服务器通过以普通话为基础的词汇及特色词汇的对应关系查找到相应的特色词汇后,还将所述输入的拼音与所述特色词汇对应保存的拼音进行模糊比对,如果两者相同或相近则验证通过,向所述下发接收模块 540 下发根据每个词汇所查找到的对应的一个或多个特色词汇。

[0081] 下发接收模块 540,用于接收所述云端服务器下发的查找到的与拆分得到的每个词汇对应的一个或多个特色词汇。

[0082] 优选的,所述下发接收模块,还用于在接收所述云端服务器下发的查找到的与拆分得到的每个词汇对应的多个特色词汇时,接收所述多个特色词汇中每个特色词汇的使用热度。特色词汇的使用热度通过被选择确认的概率的方式来客观衡量,所述选择确认的概率包括客户端所在本地对应的选择确认的概率、与当前用户使用历史相应的选择确定的概率、网络中多用户共同选择确定的概率等等。

[0083] 优选的,所述概率由以下方式获得:云端服务器在距当前时刻的一总时长区域内,划分以当前时刻为起点的多个时长范围;按所述时长范围的时长从短到长的顺序为每个时长范围赋予一权重值,所述时长范围的时长最短的权重值最高;在每个时长范围内,分别统计每个特色词汇被选择确认的次数以及所有特色词汇被选择确认的次数;根据每个特色词汇被选择确认的次数以及所有特色词汇被选择确认的次数,计算在每个时长范围内每个特色词汇被选择确认的次数所占的比重,即为每个特色词汇被选择确认的概率;针对每个特色词汇,在各个时长范围被选择确定的概率及相应时长范围的权重,统计出每个特色词汇在总时长区域内被选择确认的概率并进行记录以供下发至下发接收模块 540。

[0084] 交互展现模块 550,用于将本地接收模块 510 接收的所述字段显示在客户端界面上;还用于将下发接收模块 540 接收的所述一个或多个特色词汇关联显示在所述客户端界面中所述词汇的周边供用户选择确认,接收用户选择确认的每个词汇对应的特色词汇,使用所述特色词汇替换所述字段中的相应词汇并将替换后的所述字段展现在所述客户端界面上。

[0085] 优选的,所述交互展现模块,进一步用于将查找到的与拆分得到的每个词汇对应的多个特色词汇按照被选择确认的概率由高到低的排列顺序关联显示在所述客户端界面中所述词汇的周边供用户选择确认。

[0086] 参考图 6,图 6 是本申请实施例的一种包括定位模块 560 及词库配置模块 570 的用于输入法的字段转换的客户端 600 的结构框图,其中模块 510-550 包括用于实现与图 5 中 510-550 相同的功能,除此以外下述模块的功能还包括:

[0087] 定位模块 560,用于获取所述客户端的位置信息。具体而言,定位可以使用 GPS 定位技术,或者可以使用互联网定位技术,或者使用移动通信网等等定位技术等等现有及未来的自动定位技术,还可以由用户自己选择其位置。

[0088] 上报查询模块 530,还用于将所述定位模块 560 获取的位置信息上报到所述云端服务器从而查找相应的特色词库,还用于将所述交互展现模块 550 发来的用户选择的特色词库的标识上报给所述云端服务器从而使所述云端服务器切换到所述用户选择的特色词库。

[0089] 所述下发接收模块 540,还用于接收所述云端服务器下发的所述位置信息相应的特色词库的标识。

[0090] 所述词库配置模块 570, 用于比较所述位置信息相应的特色词库的标识与用户上一次使用的特色词库的标识, 如果不同, 将所述位置信息相应的特色词库的标识、用户上一次使用的特色词库的标识及用户偏好设置的特色词库的标识发送给所述交互展现模块; 如果相同, 将所述位置信息相应的特色词库的标识及用户偏好设置的特色词库的标识发送给所述交互展现模块;

[0091] 所述交互展现模块 550, 还用于将所述词库配置模块发来的特色词库的标识提供给用户供选择, 并将用户选择的特色词库的标识发送给所述上报查询模块 530。

[0092] 在一个典型的配置中, 计算设备包括一个或多个处理器 (CPU)、输入 / 输出接口、网络接口和内存。

[0093] 内存可能包括计算机可读介质中的非永久性存储器, 随机存取存储器 (RAM) 和 / 或非易失性内存等形式, 如只读存储器 (ROM) 或闪存 (flash RAM)。内存是计算机可读介质的示例。

[0094] 计算机可读介质包括永久性和非永久性、可移动和非可移动媒体可以由任何方法或技术来实现信息存储。信息可以是计算机可读指令、数据结构、程序的模块或其他数据。计算机的存储介质的例子包括, 但不限于相变内存 (PRAM)、静态随机存取存储器 (SRAM)、动态随机存取存储器 (DRAM)、其他类型的随机存取存储器 (RAM)、只读存储器 (ROM)、电可擦除可编程只读存储器 (EEPROM)、快闪记忆体或其他内存技术、只读光盘只读存储器 (CD-ROM)、数字多功能光盘 (DVD) 或其他光学存储、磁盒式磁带, 磁带磁磁盘存储或其他磁性存储设备或任何其他非传输介质, 可用于存储可以被计算设备访问的信息。按照本文中的界定, 计算机可读介质不包括暂存电脑可读媒体 (transitory media), 如调制的数据信号和载波。

[0095] 还需要说明的是, 术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含, 从而使得包括一系列要素的过程、方法、商品或者设备不仅包括那些要素, 而且还包括没有明确列出的其他要素, 或者是还包括为这种过程、方法、商品或者设备所固有的要素。在没有更多限制的情况下, 由语句“包括一个……”限定的要素, 并不排除在包括所述要素的过程、方法、商品或者设备中还存在另外的相同要素。

[0096] 本领域技术人员应明白, 本申请的实施例可提供为方法、系统或计算机程序产品。因此, 本申请可采用完全硬件实施例、完全软件实施例或结合软件和硬件方面的实施例的形式。而且, 本申请可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质 (包括但不限于磁盘存储器、CD-ROM、光学存储器等) 上实施的计算机程序产品的形式。

[0097] 以上所述仅为本申请的实施例而已, 并不用于限制本申请。对于本领域技术人员来说, 本申请可以有各种更改和变化。凡在本申请的精神和原理之内所作的任何修改、等同替换、改进等, 均应包含在本申请的权利要求范围之内。

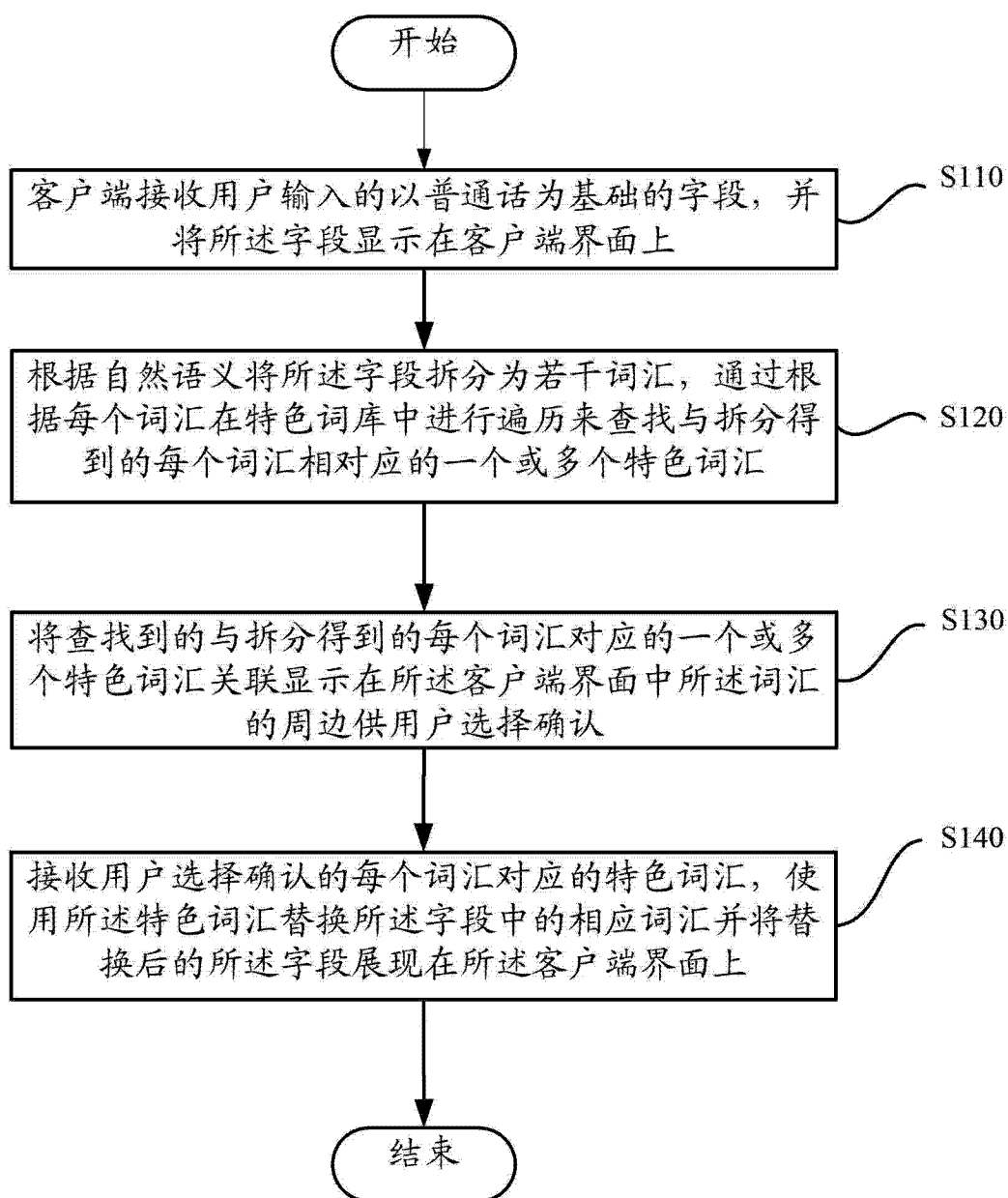


图 1

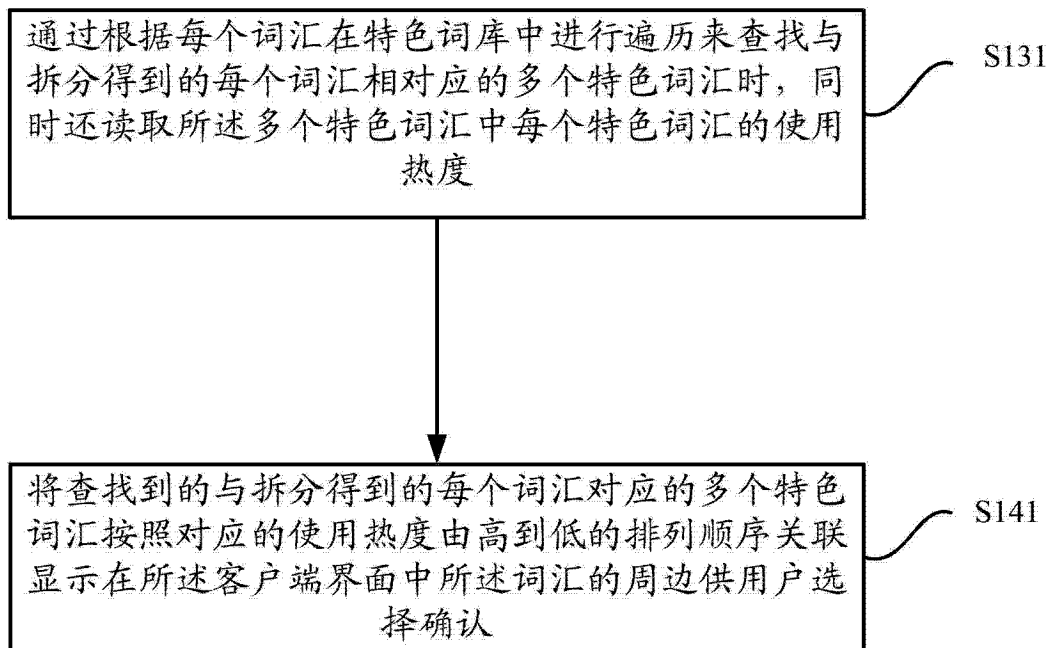


图 2a

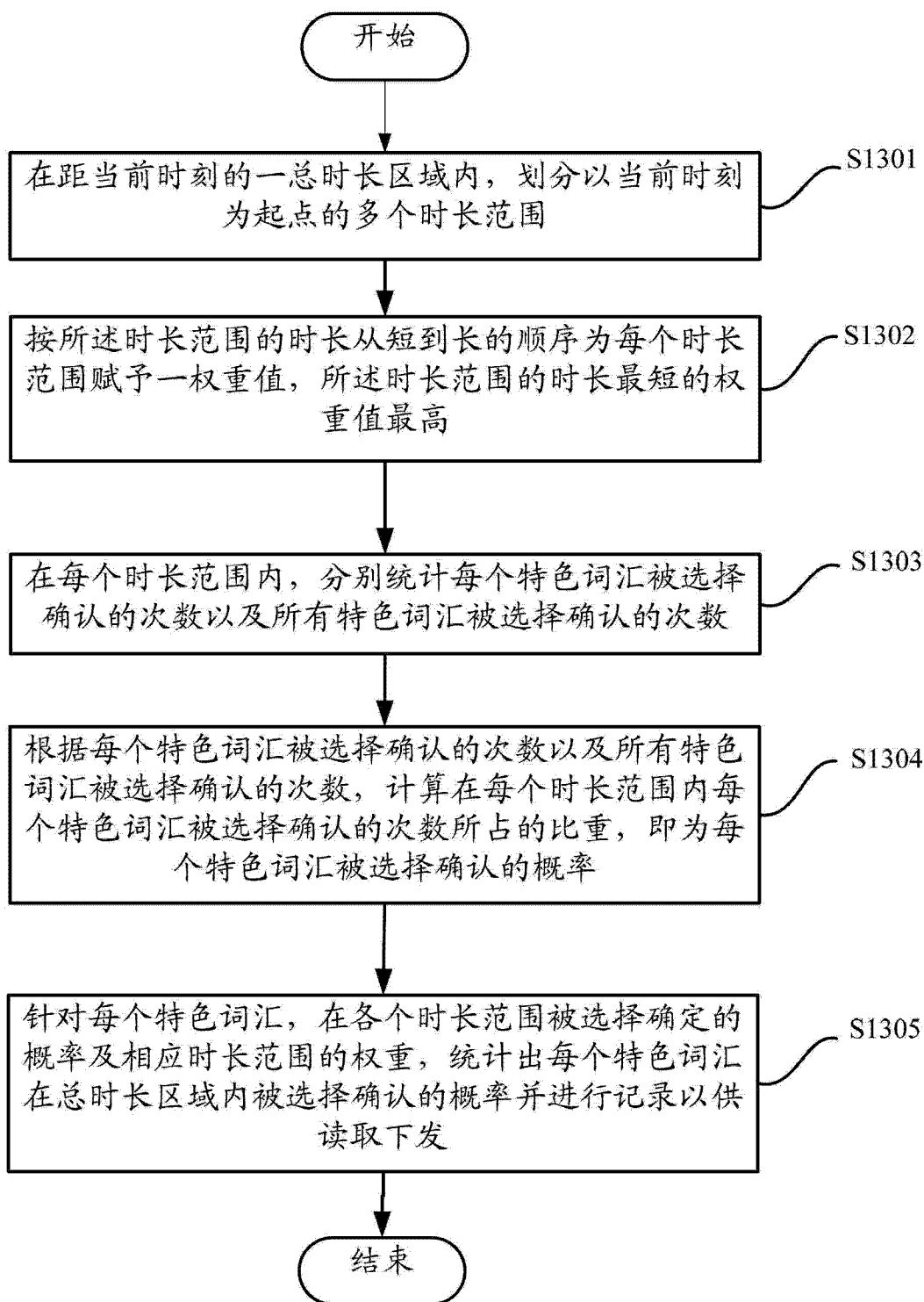


图 2b

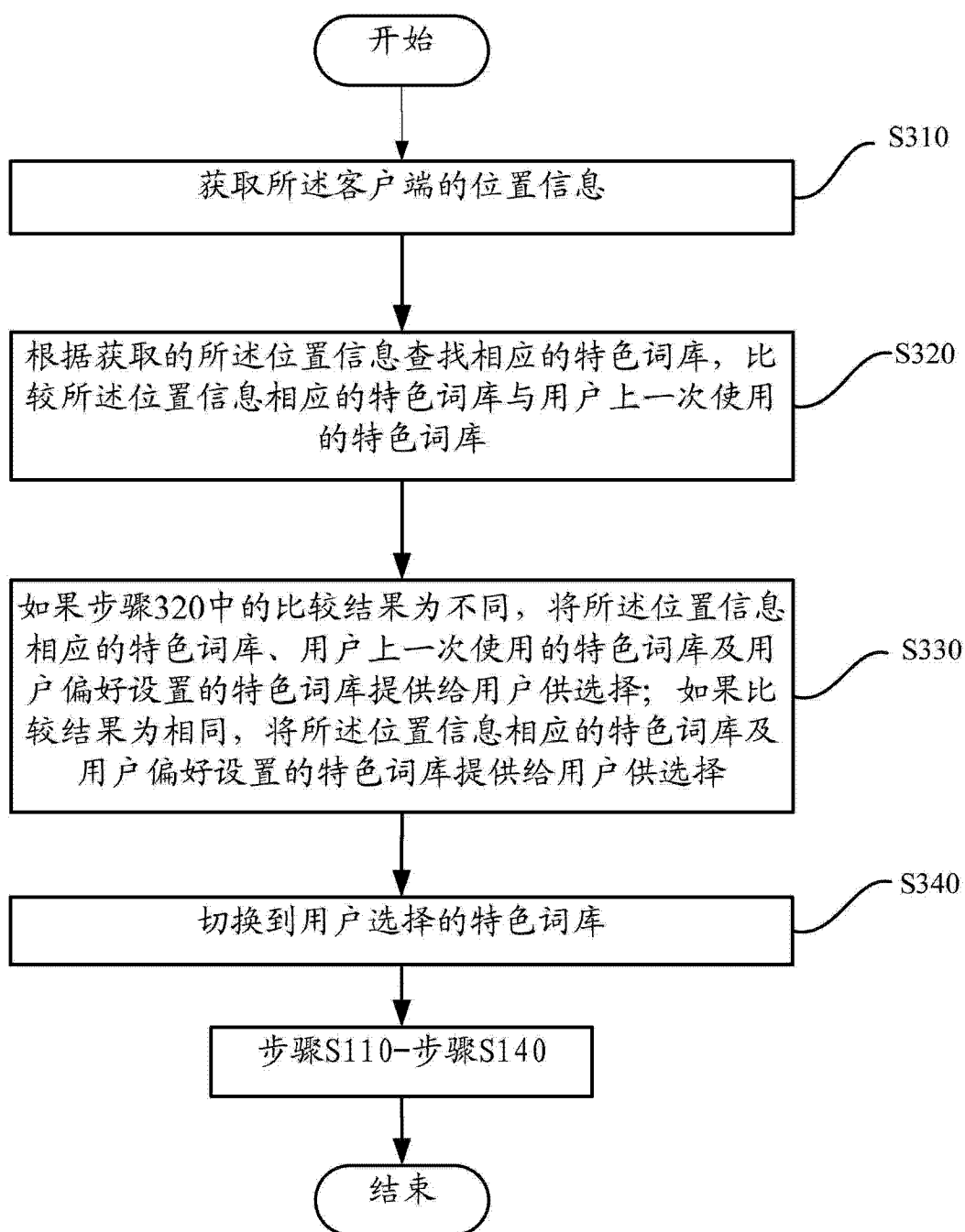


图 3

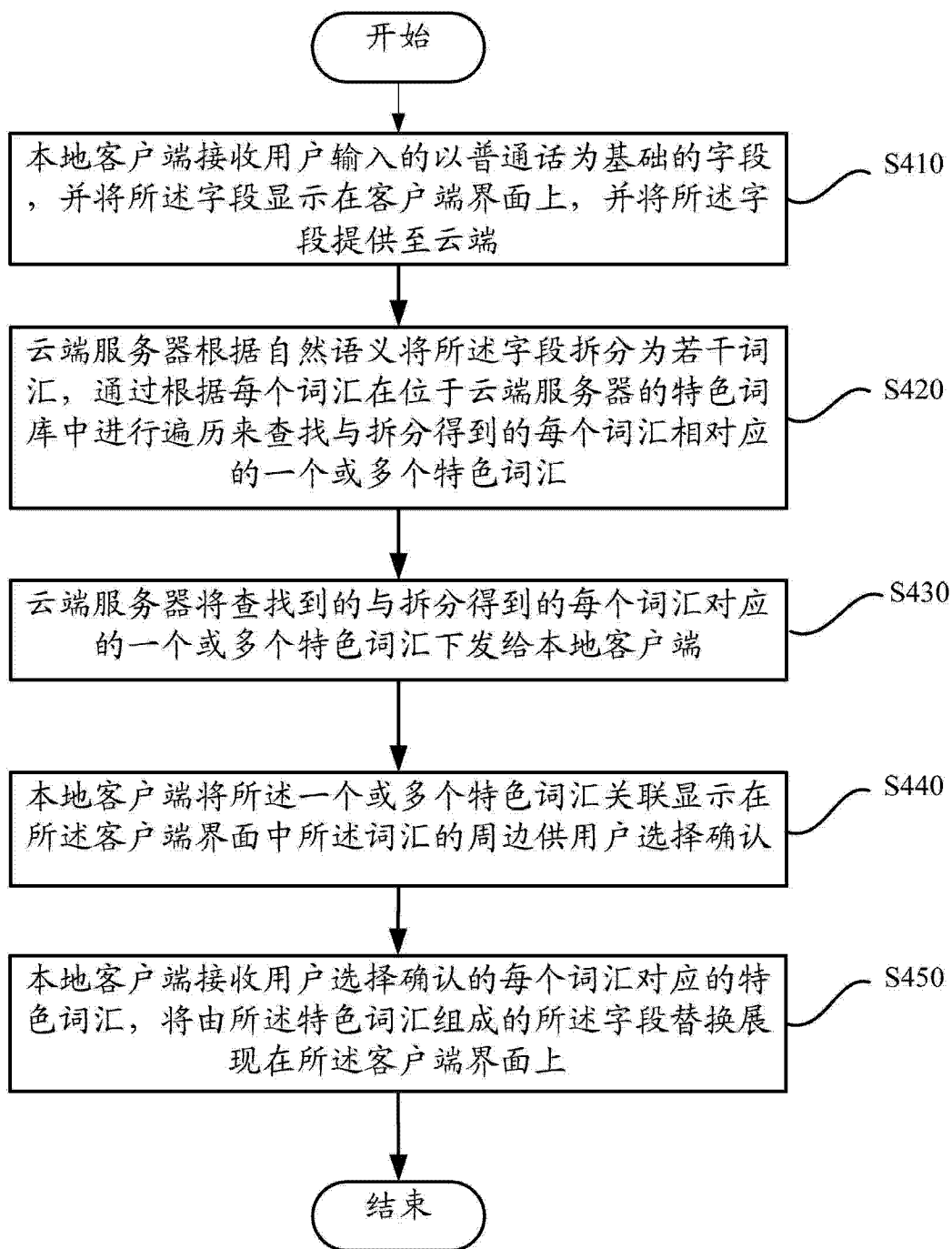


图 4

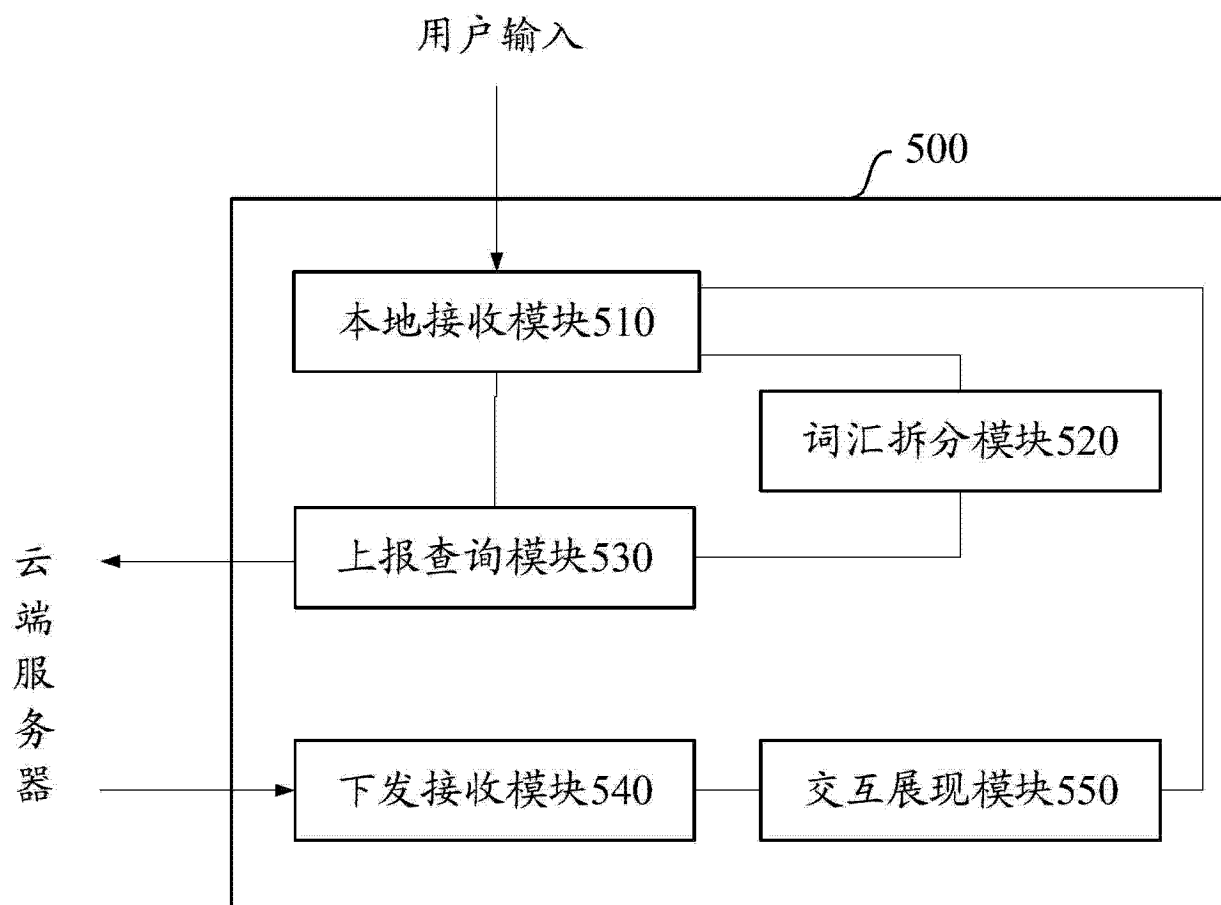


图 5

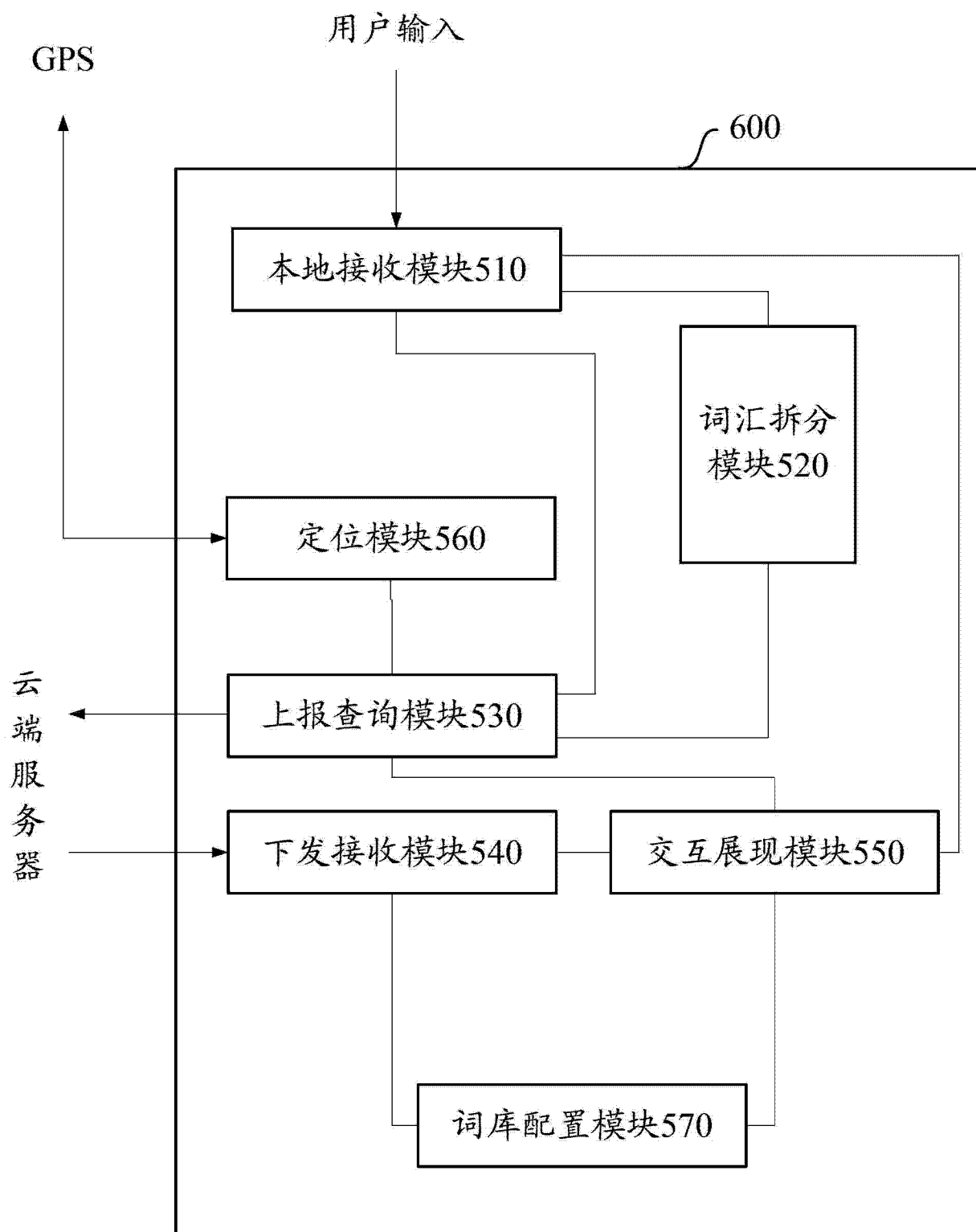


图 6