



(12) 发明专利

(10) 授权公告号 CN 101727271 B

(45) 授权公告日 2012. 11. 14

(21) 申请号 200810224815. 2

(22) 申请日 2008. 10. 22

(73) 专利权人 北京搜狗科技发展有限公司

地址 100084 北京市海淀区中关村东路 1 号  
院搜狐网络大厦 9 层 01 房间

(72) 发明人 张扬 王坚

(74) 专利代理机构 北京集佳知识产权代理有限公司 11227

代理人 逯长明

(51) Int. Cl.

G06F 3/048(2006. 01)

G06F 17/27(2006. 01)

(56) 对比文件

CN 100405371 C, 2008. 07. 23, 全文.

JP 特开平 8-115194 A, 1996. 05. 07, 全文.

CN 101114298 A, 2008. 01. 30, 全文.

JP 特开平 8-329068 A, 1996. 12. 13, 全文.

CN 101241514 A, 2008. 08. 13, 全文.

CN 1920827 A, 2007. 02. 28, 全文.

CN 1407468 A, 2003. 04. 02, 全文.

CN 101206673 A, 2008. 06. 25, 全文.

审查员 王锴

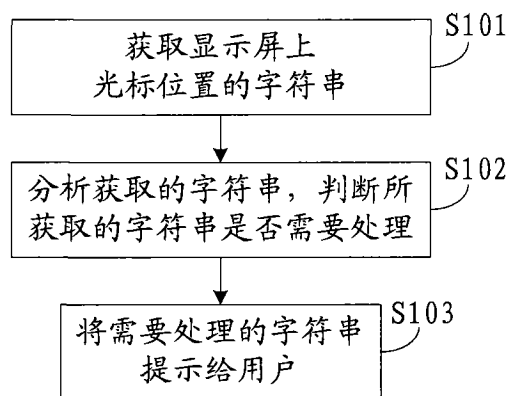
权利要求书 3 页 说明书 12 页 附图 4 页

(54) 发明名称

一种提供纠错提示的方法、装置及输入法系统

(57) 摘要

本发明公开了一种提供纠错提示的方法,该方法包括:获取显示屏上光标位置的字符串;分析获取的字符串,判断所获取的字符串是否需要处理;将需要处理的字符串提示给用户。本发明还公开了一种提供纠错提示的装置及输入法系统,通过本发明,不仅可以对用户正在使用输入法编辑的文字进行纠错提示,还可以将已经展现在显示屏上的需要处理的字符串提示给用户,并向用户提供补全或纠错候选,以此来帮助用户快速、准确地定位到需要处理的字符串并完成文本的输入及修改,使得输入过程更加流畅。由于输入法自带纠错提示功能,因此减少用户对 Word 等编辑环境的依赖,能够提升用户在不同场景下输入文字的质量。



1. 一种提供纠错提示的方法,所述方法应用于字符信息输入领域,其特征在于,包括:  
获取显示屏上光标位置的字符串;  
分析获取的字符串,判断所获取的字符串是否需要处理;  
将需要处理的字符串提示给用户;  
所述判断获取的字符串是否需要处理包括:  
将获取的字符串与预置词库中的字符串进行比较,判断预置词库中是否存在与该获取的字符串完全匹配的字符串,如果不存在,则该获取的字符串需要处理;或,  
获取显示屏上光标位置的字符串的上下文;分析所述上下文,判断该获取的字符串是否存在用词错误,如果存在,则该字符串需要处理。
2. 根据权利要求 1 所述的方法,其特征在于,还包括:  
根据预置词库,为所述需要处理的字符串提供候选字符串供用户选择。
3. 根据权利要求 2 所述的方法,其特征在于,所述根据预置词库,为所述需要处理的字符串提供候选字符串供用户选择包括:  
根据预置词库及光标位置的字符串前缀或后缀对所述需要处理的字符串进行补全,生成补全候选,并将所述补全候选提示给用户。
4. 根据权利要求 3 所述的方法,其特征在于,所述预置词库中的字符串按照预置顺序形成前缀树或后缀树结构,使用前缀树或后缀树结构生成所述补全候选。
5. 根据权利要求 2 所述的方法,其特征在于,所述根据预置词库,为所述需要处理的字符串提供候选字符串供用户选择包括:  
根据预置词库对所述需要处理的字符串进行编辑操作,生成纠错候选,并将所述纠错候选提示给用户;所述编辑操作包括插入、删除、替换或交换字符顺序。
6. 根据权利要求 5 所述的方法,其特征在于,所述预置词库中的字符串按照预置顺序形成前缀树结构,基于预置容错机制在前缀树中进行所述纠错候选的查找。
7. 根据权利要求 1 所述的方法,其特征在于,还包括:  
获取用户对所述提示的响应,在用户处理完一需要处理的字符串后,自动定位到下一需要处理的字符串。
8. 一种提供纠错提示的装置,所述装置应用于字符信息输入领域,其特征在于,包括:  
取词单元,用于获取显示屏光标位置的字符串;  
判断单元,用于分析获取的字符串,判断字符串是否需要处理;  
提示单元,用于将需要处理的字符串提示给用户;  
所述判断单元包括:  
比较子单元,用于将获取的字符串与预置词库中的字符串进行比较,判断预置词库中是否存在与该获取的字符串完全匹配的字符串;第一确定子单元,用于将所述与预置词库中的字符串均不完全匹配的字符串确定为所述需要处理的字符串;或,  
上下文获取子单元,用于获取显示屏上光标位置的字符串的上下文;分析子单元,用于分析所述展现在显示屏上的字符串的上下文,判断是否存在用词错误的字符串;第二确定子单元,用于将所述用词错误的字符串确定为所述需要处理的字符串。
9. 根据权利要求 8 所述的装置,其特征在于,还包括:  
候选提示单元,用于根据预置词库,为所述需要处理的字符串提供候选字符串供用户

选择。

10. 根据权利要求 9 所述的装置,其特征在于,所述候选提示单元包括:

补全子单元,用于根据预置词库及光标位置的字符串前缀或后缀,对所述需要处理的字符串进行补全,并生成补全候选;

第一提示子单元,用于将所述补全候选提示给用户。

11. 根据权利要求 10 所述的装置,其特征在于,所述预置词库中的字符串按照预置顺序形成前缀树或后缀树结构,所述补全子单元使用前缀树或后缀树结构生成所述补全候选。

12. 根据权利要求 9 所述的装置,其特征在于,所述候选提示单元包括:

纠错子单元,用于根据预置词库对所述需要处理的字符串进行编辑操作,生成纠错候选;所述编辑操作包括插入、删除、替换或交换字符顺序;

第二提示子单元,用于将所述纠错候选提示给用户。

13. 根据权利要求 12 所述的装置,其特征在于,所述预置词库中的字符串按照预置顺序形成前缀树结构,所述纠错子单元基于预置容错机制在前缀树中进行所述纠错候选的查找。

14. 根据权利要求 8 所述的装置,其特征在于,还包括:

用户响应获取单元,用于获取用户对所述提示的响应,在用户处理完一需要处理的字符串后,自动定位到下一需要处理的字符串。

15. 一种提供纠错提示的输入法系统,其特征在于,包括:

取词模块,用于获取显示屏上光标位置的字符串;

判断模块,用于分析获取的字符串,判断字符串是否需要处理;

提示模块,用于将需要处理的字符串提示给用户;

所述判断模块包括:

比较子模块,用于将获取的字符串与预置词库中的字符串进行比较,判断预置词库中是否存在与该获取的字符串完全匹配的字符串;第一确定子模块,用于将所述与预置词库中的字符串均不完全匹配的字符串确定为所述需要处理的字符串;或,

上下文获取子模块,用于获取显示屏上光标位置的字符串的上下文;分析子模块,用于分析所述展现在显示屏上的字符串的上下文,判断是否存在用词错误的字符串;第二确定子模块,用于将所述用词错误的字符串确定为所述需要处理的字符串。

16. 根据权利要求 15 所述的输入法系统,其特征在于,还包括:

候选提示模块,用于根据预置词库,为所述需要处理的字符串提供候选字符串供用户选择。

17. 根据权利要求 16 所述的输入法系统,其特征在于,所述候选提示模块包括:

补全子模块,用于根据预置词库及光标位置的字符串前缀或后缀,对所述需要处理的字符串进行补全,并生成补全候选;

第一提示子模块,用于将所述补全候选提示给用户。

18. 根据权利要求 17 所述的输入法系统,其特征在于,所述预置词库中的字符串按照预置顺序形成前缀树或后缀树结构,所述补全子模块使用前缀树或后缀树结构生成所述补全候选。

19. 根据权利要求 16 所述的输入法系统,其特征在于,所述候选提示模块包括:  
纠错子模块,用于根据预置词库对所述需要处理的字符串进行编辑操作,生成纠错候选;所述编辑操作包括插入、删除、替换或交换字符顺序;  
第二提示子模块,用于将所述纠错候选提示给用户。
20. 根据权利要求 19 所述的输入法系统,其特征在于,所述预置词库中的字符串按照预置顺序形成前缀树结构,所述纠错子模块基于预置容错机制在前缀树中进行所述纠错候选的查找。
21. 根据权利要求 15 所述的输入法系统,其特征在于,还包括:  
用户响应获取模块,用于获取用户对所述提示的响应,在用户处理完一需要处理的字符串后,自动定位到下一需要处理的字符串。
22. 根据权利要求 21 所述的输入法系统,其特征在于,还包括:  
用户词库管理模块,用于当用户响应获取模块确定用户对所述需要处理的字符串进行处理时,判断用户词库中是否存在之前已判断出需要处理的字符串,如果存在,将用户词库中存在的需要处理的字符串替换为处理后的字符串;所述用户词库用于记录用户的输入历史,在用户进行字符串输入时,为用户提供字符串候选;  
所述处理后的字符串由所述用户响应获取模块获取。
23. 根据权利要求 21 所述的输入法系统,其特征在于,还包括:  
记录模块,用于当用户响应获取模块确定用户对所述需要处理的字符串进行处理时,记录用户处理前及处理后的字符串的对应关系;  
所述判断模块优先根据所述对应关系进行判断。
24. 根据权利要求 21 所述的输入法系统,其特征在于,还包括:  
新词发现模块,用于当所述用户响应获取模块确定用户忽略同一需要处理的字符串的次数超过预置阈值时,将该需要处理的字符串记录为用户新词;  
所述判断模块将再次出现的所述用户新词确定为不需要处理。

## 一种提供纠错提示的方法、装置及输入法系统

### 技术领域

[0001] 本发明涉及字符信息输入领域，特别是涉及一种提供纠错提示的方法、装置及输入法系统。

### 背景技术

[0002] 输入法用户在进行非母语字符的输入时常常会遇到一些困难，例如，非英文用户（如中文、日文、韩文等）在进行英文词条的输入时经常会出现输入错误等。这主要是因为英语的拼读、单词构成方式、语法等方面与用户通常使用的母语相比存在巨大差异，用户往往会有英文单词“记不全”、“拼错词”、“不会用”或“用错单词”的情况。因此，用户通常会希望在使用输入法开始输入英文单词的同时，输入法能够提供候选的单词，并能够在用户输入错误时能够自动纠错。这样用户一旦在输入法提供的候选词中发现需要的单词，便可以选中该单词完成输入，节省输入时间的同时，可以避免记不全、拼错词等造成的错误。

[0003] 现有的输入法通过记录用户当前正在输入的字符，可以实现针对用户当前输入的动态补全及纠错提示。其中，动态补全是指，根据用户输入的字母给出候选词，并随着用户输入字母的增加，动态调整候选词。例如，用户输入字母“b”，则输入法提供的候选词为输入法词库中以“b”为首字母的单词；接下来，如果用户输入字母“a”，则输入法提供的候选词为前两个字母为“ba”的单词，以此类推；用户可以通过选中输入法提供的某个候选词完成输入；纠错提示是指，如果输入法词库中不存在与用户当前输入相匹配的单词，则会根据拼写及读音给出修改提示。例如，用户当前输入了“bagy”，输入法可以根据其拼写及读音等，向用户做出修改为“baggy”、“baby”等纠错提示。

[0004] 但是该现有输入法只能记录用户当前输入的字符，因此其动态补全及纠错功能都是在输入法当前正在处理的字符基础上进行的，即只能针对用户当前正在输入栏输入的文字进行处理，而不能处理已经展现在显示屏上的文本。比如，如果用户打开的某文档中有存在错误的单词，或者用户完成一段文本的输入进行检查时，该现有输入法就无法向用户做出提示，用户必须根据自己的判断来修改。

### 发明内容

[0005] 有鉴于此，本发明的目的在于提供一种提供纠错提示的方法、装置及输入法系统，以解决现有输入法系统无法对已经展现在显示屏上的需要处理的文本进行提示的问题。

[0006] 为实现上述目的，本发明提供了如下方案：

[0007] 一种提供纠错提示的方法，所述方法应用于字符信息输入领域，包括：

[0008] 获取显示屏上光标位置的字符串；

[0009] 分析获取的字符串，判断所获取的字符串是否需要处理；

[0010] 将需要处理的字符串提示给用户；

[0011] 所述判断获取的字符串是否需要处理包括：

[0012] 将获取的字符串与预置词库中的字符串进行比较，判断预置词库中是否存在与该

获取的字符串完全匹配的字符串,如果不存在,则该获取的字符串需要处理;或,

[0013] 获取显示屏上光标位置的字符串的上下文;分析所述上下文,判断该获取的字符串是否存在用词错误,如果存在,则该字符串需要处理。

[0014] 优选的,还包括:

[0015] 根据预置词库,为所述需要处理的字符串提供候选字符串供用户选择。

[0016] 优选的,所述根据预置词库,为所述需要处理的字符串提供候选字符串供用户选择包括:

[0017] 根据预置词库及光标位置的字符串前缀或后缀对所述需要处理的字符串进行补全,生成补全候选,并将所述补全候选提示给用户。

[0018] 优选的,所述预置词库中的字符串按照预置顺序形成前缀树或后缀树结构,使用前缀树或后缀树结构生成所述补全候选。

[0019] 优选的,所述根据预置词库,为所述需要处理的字符串提供候选字符串供用户选择包括:

[0020] 根据预置词库对所述需要处理的字符串进行编辑操作,生成纠错候选,并将所述纠错候选提示给用户;所述编辑操作包括插入、删除、替换或交换字符顺序。

[0021] 优选的,所述预置词库中的字符串按照预置顺序形成前缀树结构,基于预置容错机制在前缀树中进行所述纠错候选的查找。

[0022] 优选的,还包括:

[0023] 获取用户对所述提示的响应,在用户处理完一需要处理的字符串后,自动定位到下一需要处理的字符串。

[0024] 一种提供纠错提示的装置,所述装置应用于字符信息输入领域,包括:

[0025] 取词单元,用于获取显示屏光标位置的字符串;

[0026] 判断单元,用于分析获取的字符串,判断字符串是否需要处理;

[0027] 提示单元,用于将需要处理的字符串提示给用户;

[0028] 所述判断单元包括:

[0029] 比较子单元,用于将获取的字符串与预置词库中的字符串进行比较,判断预置词库中是否存在与该获取的字符串完全匹配的字符串;第一确定子单元,用于将所述与预置词库中的字符串均不完全匹配的字符串确定为所述需要处理的字符串;或,

[0030] 上下文获取子单元,用于获取显示屏上光标位置的字符串的上下文;分析子单元,用于分析所述展现在显示屏上的字符串的上下文,判断是否存在用词错误的字符串;第二确定子单元,用于将所述用词错误的字符串确定为所述需要处理的字符串。

[0031] 优选的,还包括:

[0032] 候选提示单元,用于根据预置词库,为所述需要处理的字符串提供候选字符串供用户选择。

[0033] 优选的,所述候选提示单元包括:

[0034] 补全子单元,用于根据预置词库及光标位置的字符串前缀或后缀,对所述需要处理的字符串进行补全,并生成补全候选;

[0035] 第一提示子单元,用于将所述补全候选提示给用户。

[0036] 优选的,所述预置词库中的字符串按照预置顺序形成前缀树或后缀树结构,所述

补全子单元使用前缀树或后缀树结构生成所述补全候选。

[0037] 优选的,所述候选提示单元包括:

[0038] 纠错子单元,用于根据预置词库对所述需要处理的字符串进行编辑操作,生成纠错候选;所述编辑操作包括插入、删除、替换或交换字符顺序;

[0039] 第二提示子单元,用于将所述纠错候选提示给用户。

[0040] 优选的,所述预置词库中的字符串按照预置顺序形成前缀树结构,所述纠错子单元基于预置容错机制在前缀树中进行所述纠错候选的查找。

[0041] 优选的,还包括:

[0042] 用户响应获取单元,用于获取用户对所述提示的响应,在用户处理完一需要处理的字符串后,自动定位到下一需要处理的字符串。

[0043] 一种提供纠错提示的输入法系统,包括:

[0044] 取词模块,用于获取显示屏上光标位置的字符串;

[0045] 判断模块,用于分析获取的字符串,判断字符串是否需要处理;

[0046] 提示模块,用于将需要处理的字符串提示给用户;

[0047] 所述判断模块包括:

[0048] 比较子模块,用于将获取的字符串与预置词库中的字符串进行比较,判断预置词库中是否存在与该获取的字符串完全匹配的字符串;第一确定子模块,用于将所述与预置词库中的字符串均不完全匹配的字符串确定为所述需要处理的字符串;或,

[0049] 上下文获取子模块,用于获取显示屏上光标位置的字符串的上下文;分析子模块,用于分析所述展现在显示屏上的字符串的上下文,判断是否存在用词错误的字符串;第二确定子模块,用于将所述用词错误的字符串确定为所述需要处理的字符串。

[0050] 优选的,还包括:

[0051] 候选提示模块,用于根据预置词库,为所述需要处理的字符串提供候选字符串供用户选择。

[0052] 优选的,所述候选提示模块包括:

[0053] 补全子模块,用于根据预置词库及光标位置的字符串前缀或后缀,对所述需要处理的字符串进行补全,并生成补全候选;

[0054] 第一提示子模块,用于将所述补全候选提示给用户。

[0055] 优选的,所述预置词库中的字符串按照预置顺序形成前缀树或后缀树结构,所述补全子模块使用前缀树或后缀树结构生成所述补全候选。

[0056] 优选的,所述候选提示模块包括:

[0057] 纠错子模块,用于根据预置词库对所述需要处理的字符串进行编辑操作,生成纠错候选;所述编辑操作包括插入、删除、替换或交换字符顺序;

[0058] 第二提示子模块,用于将所述纠错候选提示给用户。

[0059] 优选的,所述预置词库中的字符串按照预置顺序形成前缀树结构,所述纠错子模块基于预置容错机制在前缀树中进行所述纠错候选的查找。

[0060] 优选的,还包括:

[0061] 用户响应获取模块,用于获取用户对所述提示的响应,在用户处理完一需要处理的字符串后,自动定位到下一需要处理的字符串。

[0062] 优选的,还包括:

[0063] 用户词库管理模块,用于当用户响应获取模块确定用户对所述需要处理的字符串进行处理时,判断用户词库中是否存在之前已判断出需要处理的字符串,如果存在,将用户词库中存在的需要处理的字符串替换为处理后的字符串;所述用户词库用于记录用户的输入历史,在用户进行字符串输入时,为用户提供字符串候选;

[0064] 所述处理后的字符串由所述用户响应获取模块获取。

[0065] 优选的,还包括:

[0066] 记录模块,用于当用户响应获取模块确定用户对所述需要处理的字符串进行处理时,记录用户处理前及处理后的字符串的对应关系;

[0067] 所述判断模块优先根据所述对应关系进行判断。

[0068] 优选的,还包括:

[0069] 新词发现模块,用于当所述用户响应获取模块确定用户忽略同一需要处理的字符串的次数超过预置阈值时,将该需要处理的字符串记录为用户新词;

[0070] 所述判断模块将再次出现的所述用户新词确定为不需要处理。

[0071] 根据本发明提供的具体实施例,本发明公开了以下技术效果:

[0072] 第一,本发明通过获取显示屏上光标位置的字符串,并对获取的字符串进行分析,发现其中需要处理的字符串,并将需要处理的字符串提示给用户,可见,本发明不仅可以提示用户正在输入的文字,还可以将已经展现在显示屏上的需要处理的字符串提示给用户,以此来帮助用户快速、准确地发现需要处理的字符串并完成文本的输入及修改,使得输入过程更加流畅。

[0073] 第二,可以对需要处理的字符串进行补全或纠错,将生成的候选提示给用户,进一步加快了字符串的处理效率。由于还可以获取字符串的上下文,因此不仅能够获取到存在拼写错误的字符串,还可以获取到存在使用错误的字符串,然后可以对需要处理的字符串进行纠错或补全,以处理本身拼写正确但存在用词错误的情况。比如通过光标取词获取文字片段“please fill the form before you check out from the hotel”后,快速定位到其中的第一个 from,并给出“form”的纠错提示。另外,补全单元可以根据获取的字符串中的正确部分的前缀或后缀做出补全候选。例如,在用户进行退格操作时,由于可以获取已展现在显示屏上的单词的前缀部分,因此可以根据该前缀部分在预置词库中匹配并进行动态补全,而不需要用户从头输入这部分单词前缀。

[0074] 第三,可以使用前缀树或后缀树结构进行补全候选或纠错候选的查找匹配,查找效率高。

[0075] 第四,具有用户词学习功能,能够实现对用户反馈的能动适应。处理前后的字符串对(错误形式,正确形式)经用户确认后记录下来(如放入缓存),在之后的应用中可进行优先查找和匹配,提升效率。如果多次提示同一个“错误”拼写或“使用错误”的字符串,用户都不予以修改,学习这个词条并记录下来,下次再出现这个词条时可不予提示。从而实现了针对用户的个性化纠错提示。

[0076] 第五,通过输入法系统来提供对已经展现在显示屏上的文本纠错提示功能,用户可以基于输入法完成对已展现在显示屏上的文本的纠错,从而缓解了对文本编辑环境(如 Microsoft Word 等文本编辑软件)的依赖。



## 附图说明

- [0077] 图 1 是本发明提供的方法的流程图；
- [0078] 图 2 是现有输入法对已经展现在显示屏上的字符串的补全示意图；
- [0079] 图 3 是本发明使用的前缀树结构词库的示意图；
- [0080] 图 4 是本发明提供的第一装置的示意图；
- [0081] 图 5 是本发明提供的第二装置的示意图；
- [0082] 图 6 是本发明提供的第三装置的示意图；
- [0083] 图 7 是本发明提供的第一输入法系统的示意图；
- [0084] 图 8 是本发明提供的第二输入法系统的示意图；
- [0085] 图 9 是本发明提供的第三输入法系统的示意图。

## 具体实施方式

[0086] 为使本发明的上述目的、特征和优点能够更加明显易懂，下面结合附图和具体实施方式对本发明作进一步详细的说明。

[0087] 参见图 1，本发明实施例提供的纠错提示的方法包括以下步骤：

[0088] S101：获取显示屏上光标位置的字符串；

[0089] 在使用现有的输入法进行字符输入时，所述显示屏上光标位置的字符串已经不受输入法控制处理，即便某些字符串需要处理（如，存在拼写错误、使用错误等），现有的输入法也无法做出任何提示，因为现有的输入法只能记录用户当前输入的字符，并根据当前记录的字符进行提示。因此，在本发明实施例提供的方法中，首先要获取显示屏上光标位置的字符串，在此基础上来进行后续的分析判断步骤。

[0090] 其中，所述获取显示屏上光标位置的字符串的具体实现可以有多种，例如，在 Windows 系统下，当光标移动到某位置时，可以强制重绘窗口来获取文本信息，根据光标位置通过函数钩子来取得当前字符串；或者记录输入历史，在需要获取光标位置展现在显示屏上的字符串时，直接将当前输入的输入历史取回，等等。

[0091] 关于字符串的边界，以英文输入为例，可以有以下两种情况：如果是纯英文输入，可以通过前后的空格或结束符来确定；如果是中英文混合输入等情况，则非英文字符、空格、结束符等都可以用于字符串边界的判定。

[0092] S102：分析获取的字符串，判断所获取的字符串是否需要处理；

[0093] 所述需要处理通常可以包括存在拼写错误或用词错误等，对于存在拼写错误的字符串，可以通过以下方式判断：将获取的字符串与预置词库中的字符串进行比较，判断预置词库中是否存在与该获取的字符串相匹配的字符串，如果不存在，说明该字符串存在拼写错误，那么该获取的字符串需要处理。

[0094] 其中，所述预置词库中保存了所有可能用到的字符串，例如，对于英文而言，该预置词库相当于英文词典，可以保存所有的英文单词。在本发明实施例中，所述预置词库不仅可以用于对显示屏上的字符串是否需要处理的判断过程，还可以用于为需要处理的字符串提供候选字符串的过程，后者会在后续的部分进行详细地介绍。

[0095] 对于存在用词错误的字符串，需要在获取光标位置展现在显示屏上的字符串时，

还获取该字符串的前后若干个字符串作为该字符串的上下文,则可以通过分析所述上下文,判断该获取的字符串是否存在用词错误,如果存在,则该字符串需要处理。例如,通过获取文字片段“please fill the form before you check out from the hotel”后,对该片段进行分析后,可以判断出其中的第一个 from 存在用词错误(虽然其不存在拼写错误),则该字符串需要处理。

[0096] S103:将需要处理的字符串提示给用户。

[0097] 将需要处理的字符串提示给用户,可以帮助用户定位到需要处理的字符串,并进行必要的修改,而无需用户亲自去查找,节省了用户的时间。

[0098] 在实际应用中,还可以对需要处理的字符串给出一些修改的候选项,这样如果候选项中存在用户认为正确的字符串,则用户可以直接选择来完成修改,进一步节省了用户的时间;另一方面,用户可能会由于忘记了正确的拼写方式或其他原因而不知如何修改,因此,从这一点来讲,提供修改的候选项能够为用户提供更大的便利。下面就对本发明实施例采用的提供修改候选项的方法进行详细地描述。

[0099] 实施例一、可以通过对光标位置的字符串前缀或后缀进行补全,生成补全候选,并提示给用户。其中所述补全的过程可以是在用户正在修改需要处理的字符串时进行的。为了更好地体现出本发明的效果,现在与现有输入法进行对比并举例进行说明。

[0100] 在用户使用现有的输入法(以英文输入法为例)时,当用户需要修改(包括字母的插入、删除、替换等)单词时,现有输入法只会以用户新输入的字母为起点给出补全内容。比如用户在搜索引擎中输入单词“bagy”,点击搜索后发现该单词少了一个字母“g”,当用户开始从字母“y”起退格,或者直接将光标移到字母“g”之前或之后时,输入法不会做出任何提示,而当用户开始插入字母“g”时,该输入法会认为用户是在进行以当前输入的字母“g”为首字母的新的输入,并开始进行以“g”为首字母的动态补全,同时将输入法词库中以“g”为首字母的单词提供给用户,参见图2。

[0101] 而在使用本发明实施例提供的方法时,如果用户需要修改单词,则会以光标位置的单词前缀和/或后缀进行补全。仍以上述例子为例,使用本发明实施例提供的方法,当用户开始从字母“y”起退格,或者直接将光标移到字母“g”之前或之后时,本发明实施例会根据光标位置的单词前缀“bag”给出补全候选,因此,其中有非常大的可能性包括“baggy”这个候选,此时,用户便可以通过直接选择该候选来完成输入了。当前,生成的补全候选可能有很多个,因此可以预设容错查找范围,例如将编辑距离限定在某阈值之内等。其中,编辑距离是指将一个字符串转变为另一个字符串所需要的编辑开销。这里涉及的字符串可以是英语字母、汉字和/或标点符号等。编辑开销如全部定义为1,那么将 zipo 转变为 zippo 的编辑距离是1,将 englsi 转变为 english 的编辑距离是2,反过来也是如此。

[0102] 实施例二、可以根据所述预置词库对所述需要处理的字符串进行编辑操作,生成纠错候选,并将所述纠错候选提示给用户,其中,所述编辑操作包括插入、删除、替换或交换字符顺序等。例如,对于错误拼写“zipo”,可以做出“zipp”、“zip”、“zippy”等纠错候选。其中,所述交换字符顺序通常是指两个字母在输入顺序上发生错误的情况,例如经常将“what”输入成“waht”,这时就需要将字符 a 和字符 h 的顺序进行交换,将“what”作为纠错候选提示给用户。与生成补全候选时类似,由于这些编辑操作可能会生成非常多的候选,因此,可以预先规定生成纠错候选的限制条件,在一定容错查找范围内进行,例如在一定编辑

距离内进行编辑操作等。

[0103] 上述实施例一及实施例二给出了两种为用户提供修改候选项的方法,包括补全及纠错,可以在用户开始输入退格时,或直接根据光标位置,或当光标位置发生变化时进行补全或纠错的提示。在实际应用中,补全候选或纠错候选可以通过在预置的词库中进行查找获得。因此可以通过一定的策略来提高查找的效率。以生成纠错候选的过程为例,对于错误的拼写“zipo”,最普遍的做法是将 zipo 与所述预置词库中所有的条目比较一遍,这显然比较浪费时间。一个改进的办法可以是,在 zipo 的每个位置上进行 26 个字母的插入、删除、替换和交换操作,查看生成的词条是否在所述预置词库中存在。例如,其中一条路径是假设 zipo 没有错误,则可以在第三个字母 p 之后插入一个字母,生成 zip<x>o 等 26 个候选,再分别将这 26 个候选与预置词库中的词条进行比较,查找可能正确的词条,

[0104] 在本发明的优选实施例中,采用了更加有效、更加节省计算量的方法,即首先将所述预置词库中的词条按照预置顺序排列,形成前缀树的结构,然后在前缀树结构下进行查找。以英文单词词库为例,则该词库中的单词可以按照字母顺序形成前缀树结构。参见图 3 所示的一个以字母 z 开头的前缀树结构,<S> 表示开始,<E> 表示单词结束,因仅为示意,所以只显示了 z 字母开头的 zinc、zorro、zip、zippo 等几个单词。仍以上述假设 zipo 没有错误为例,在前缀树结构的词库中可以发现以 zip 为前缀的单词只存在 <E> 和 p 两条路径,这样就不用再进行除 p 以外的 25 个字母的查找匹配过程了,从而大大减少了计算量。图 3 中红色箭头显示的是另外一个例子,即在 zi- 处进行插入一个字母的插入操作,针对 zi- 下的 n、p 进行跨字母匹配:首先在 zin- 后查看是否存在 zinpo,结果查找失败;接下来在 zip- 后查看是否存在 zippo,查找成功,这时就可以把 zippo 选取出来作为一个 zipo 的纠错候选。同样,删除、替换及交换操作也可以通过类似的办法,通过前缀树结构来实现快速高效地查找。当然,利用前缀树结构进行查找匹配也适用于生成补全候选的过程,而且,如果是针对当前字符串后缀的前向补全,则可以使用后缀树结构,也就是将词库中的词条逐字母倒过来存入后缀树中。生成补全候选时,可以将当前字符串后缀从后向前逐字母与词库中的词条进行比对。

[0105] 当然,其他语种的词库也可以使用前缀树、后缀树结构,例如,对于汉语词库,可以形成两万多汉字的汉语“字母表”。因此本发明同样适用于汉字等其他文字的补全和纠错,比如汉语成语的补全等。

[0106] 另外,本发明实施例中的补全不一定是针对用户当前的输入片段或用户正在修改的字符串,也可以是为单词提供其复合词形式,例如,可以为单词“English”提供“Englishman”、“Englishwomen”、“non-English”等。基于这种原因,补全操作的容错查找范围通常要比纠错操作的查找范围要大一些。例如,补全时,所述需要处理的字符串与生成的候选字符串之间的编辑距离要大于纠错操作中限定的阈值。

[0107] 对于纠错候选,虽然不能直接利用前缀树生成,但是可以采用基于预置的容错策略在前缀树中进行纠错候选的查找,从而缩小了编辑操作的查找范围,节省了处理时间。

[0108] 与本发明实施例提供的纠错提示的方法相对应,本发明实施例还提供了一种提供纠错提示的装置,参见图 4,该装置包括:

[0109] 取词单元 U401,用于获取显示屏上光标位置的字符串;

[0110] 判断单元 U402,用于分析获取的字符串,判断字符串是否需要处理;

[0111] 提示单元 U403,用于将需要处理的字符串提示给用户。

[0112] 取词单元 U401 可以获取显示屏上光标位置的字符串,然后判断单元 U402 便可以判断该获取的字符串是否需要进行处理,如果需要处理,则将需要处理的字符串提示给用户。这样用户便可以根据提示,快速知悉哪个字符串是需要处理的,然后开始对其进行必要的处理。因此,提高了处理字符串的效率。

[0113] 其中,参见图 5,判断单元 U502 可以包括以下子单元:

[0114] 比较子单元 U5021,用于将获取的字符串与预置词库中的字符串进行比较,判断是否存在与预置词库中的字符串均不匹配的字符串;

[0115] 第一确定子单元 U5022,用于将所述与预置词库中的字符串均不匹配的字符串确定为所述需要处理的字符串。

[0116] 通过上述两个子单元,可以发现存在拼写错误的字符串,在本发明的优选实施例中,也可以发现存在用词错误的字符串,因此,该装置的判断单元 U502 也可以包括以下子单元:

[0117] 上下文获取子单元 U5023,用于获取显示屏上光标位置的字符串的上下文;

[0118] 分析子单元 U5024,用于分析所述展现在显示屏上的字符串的上下文,判断是否存在用词错误的字符串;

[0119] 第二确定子单元 U5025,用于将所述用词错误的字符串确定为所述需要处理的字符串。

[0120] 判断单元 U502 可以同时包括比较子单元 U5021 到第二确定子单元 U5025 这五个子单元,这样判断单元 U502 既可以判断出拼写错误,又可以判断出用词错误。

[0121] 其中,图 5 中的取词单元 U501、提示单元 U503 与图 4 中的取词单元 U401、提示单元 U403 相同。

[0122] 在实际应用中,除了将需要处理的字符串提示给用户外,还可以在用户开始修改字符串时或尚未修改时,向用户做出修改候选提示,这样,如果修改提示中包含用户需要的正确的字符串,则用户便可以通过直接选择来完成输入,从而进一步提高了处理字符串的效率。因此,参见图 6,该装置还可以包括:

[0123] 候选提示单元 U604,用于根据预置词库,为所述需要处理的字符串提供候选字符串供用户选择。

[0124] 其中,候选提示单元 U604 可以提示的候选可以包括补全候选及纠错候选,提示补全候选时,候选提示单元 U604 可以包括以下子单元:

[0125] 补全子单元 U6041,用于在用户开始修改字符串时,根据预置的词库及光标位置的字符串前缀或后缀(字符串中正确的部分),对所述需要处理的字符串进行补全,并生成补全候选;

[0126] 第一提示子单元 U6042,用于将生成的补全候选提示给用户。

[0127] 提示纠错候选时,候选提示单元 U604 可以包括以下子单元:

[0128] 纠错子单元 U6043,用于在判断出存在需要处理的字符串后,在用户开始修改前,便可以根据预置的词库,对所述需要处理的字符串进行编辑操作,生成纠错候选;所述编辑操作可以包括插入、删除、替换或交换字符顺序等;

[0129] 第二提示子单元 U6044,用于将生成的纠错候选提示给用户,以供用户选择。

[0130] 其中,候选提示单元 U604 可以同时包括用于提示补全候选的子单元和用于提示纠错候选的子单元,这样候选提示单元 U604 既可以向用户做出补全提示,又能够做出纠错提示。

[0131] 由于补全候选及纠错候选的生成过程都要用到预置的词库,因此在本发明的优选实施例中,该预置的词库中的字符串采用按照预置顺序(如英文单词可以按照字母顺序)实现的前缀树结构,这样生成补全或纠错候选时,便可以使用该前缀树结构进行,从而能够实现快速、高效的查找。其中,补全子单元 U6041 可以直接使用前缀树或后缀树结构生成所述补全候选;纠错子单元 U6043 可以基于预置容错机制在前缀树中进行所述纠错候选的查找。

[0132] 在本发明的优选实施例中,还可以包括用户响应获取单元 U605,用于获取用户对所述提示的响应,在用户处理完一待处理的字符串后,可以自动定位到下一需要处理的字符串。

[0133] 其中,图 6 中的取词单元 U601、判断单元 U602、提示单元 U603 与图 4 中的取词单元 U401、判断单元 U402、提示单元 U403 相同。

[0134] 本发明旨在提供一种能够对已经展现在显示屏上的字符串进行纠错提示的技术方案,而如果将该方案实现在输入法系统中,则将解决现有输入法系统无法对已经展现在显示屏上的字符串进行纠错的问题。因此,本发明实施例还提供了一种提供纠错提示的输入法系统,参见图 7,该系统包括:

[0135] 取词模块 U701,用于获取显示屏上光标位置的字符串;

[0136] 判断模块 U702,用于分析获取的字符串,判断字符串是否需要处理;

[0137] 提示模块 U703,用于将需要处理的字符串提示给用户。

[0138] 本发明实施例提供的输入法系统可以具有输入模式及纠错模式,在输入模式下,与通常的输入法系统基本相似;在用户输入完一段文字后,可以切换到纠错模式,则该输入法系统将会通过以上各模块进行查找及提示。

[0139] 其中,在本发明的优选实施例中,除了将需要处理的字符串提示给用户外,还可以在用户开始修改字符串时或尚未修改时,向用户做出修改候选提示,这样,如果修改提示中包含用户需要的正确的字符串,则用户便可以通过直接选择来完成输入,从而进一步提高了处理字符串的效率。因此,参见图 8,该输入法系统还可以包括:

[0140] 候选提示模块 U804,用于根据预置词库,为所述需要处理的字符串提供候选字符串供用户选择。

[0141] 其中,候选提示模块 U804 可以提示的候选可以包括补全候选及纠错候选,提示补全候选时,候选提示模块 U804 可以包括以下子模块:

[0142] 补全子模块 U8041,用于在用户开始修改字符串时,根据预置的词库及光标位置的字符串前缀或后缀,对所述需要处理的字符串进行补全,并生成补全候选;

[0143] 第一提示子模块 U8042 将生成的补全候选提示给用户。而在用户选定补全选项后,输入法会自动将待补全前缀的后续部分补全,形成一个完整的单词,节省用户的编辑时间。这里的方式可以是:提供单词的完整形式;或者,只提供缺失的若干前缀或后缀。如对“English”提供的候选项只有“non-”、“-man”、“-woman”等。很显然前者对用户而言更加友好些。这两种情况下当用户选中某选项时只补全缺失的部分。

[0144] 提示纠错候选时,候选提示模块 U804 可以包括以下子模块

[0145] 纠错子模块 U8043,用于在判断出存在需要处理的字符串后,在用户开始修改前,便可以根据预置的词库,对所述需要处理的字符串进行编辑操作,生成纠错候选;所述编辑操作可以包括插入、删除、替换或交换字符顺序等;

[0146] 第二提示子模块 U8044 也可以将生成的纠错候选提示给用户,以供用户选择。

[0147] 由于补全候选及纠错候选的生成过程都要用到预置词库,因此为提高查找效率,在本发明的优选实施例中,该预置词库中的字符串采用按照预置顺序实现的前缀树结构,这样生成补全或纠错候选时,便可以使用该前缀树结构进行,从而能够实现快速、高效的查找。其中,补全子模块 U8041 可以直接使用前缀树或后缀树结构生成所述补全候选;纠错子模块 U8042 可以基于预置容错机制在前缀树中进行所述纠错候选的查找。

[0148] 在本发明的优选实施例中,还可以包括用户响应获取模块 U805,用于获取用户对所述提示的响应,在用户处理完一待处理的字符串后,可以自动定位到下一需要处理的字符串。

[0149] 本发明的优选实施例中,可以实现针对特定用户的个性化纠错。由于输入法系统中通常会存在用户词库,用于记录用户的历史输入,在用户使用输入法进行字符串的输入时,可以根据该用户词库中记录的词条,为用户提供字符串候选。也就是说,只要是用户输入并确定后的字符串都会记录在该用户词库中,即使输入的某些字符串存在拼写错误等,但只要被用户确认,就会记录在用户词库中。显然,根据用户词库为用户提供字符串候选时,会有较大的几率将存在错误的字符串提示给用户,以致伤害到用户的体验。因此,该输入法系统还可以包括用户词库管理模块 U806,用于当用户在用户响应模块 U805 中响应了补全模块或纠错模块给出的提示,或者采用手动修改等其他方式对需要处理的字符串进行了修改时,如果发现该用户的用户词库中存在这个处理前的字符串,则可以将该处理前的字符串从用户词库中删除,并替换为用户处理后的字符串,同时更新用户词库中的记录信息。这样,避免了因为用户的错误输入对用户词库带来的噪音

[0150] 需要说明的是,本发明所述的预置词库与此处的用户词库不同,所述预置词库是用于在纠错模式下,判断字符串是否需要处理以及提供补全或纠错候选的;而用户词库是用于在输入模式下,为用户提供输入字符串候选的。

[0151] 另外,该输入法系统还可以包括记录模块 U807,对于从用户响应模块 U805 获得的处理前字符串及处理后字符串,该记录模块 U807 可以将其对应关系记录下来(如,放入缓存中),判断模块 U802 可以优先根据所述对应关系进行判断,为查找和匹配工作节省时间。

[0152] 再者,有些字符串可能是用户自造的,或者是一些缩略语等,而预置词库中可能不包含该字符串,因此判断模块 U802 会判断出该字符串需要处理,并由提示模块 U803 提示给用户。但是此时用户并不需要修改,盲目的提示会影响到用户的体验。因此在本发明的优选实施例中,还可以包括新词发现模块 U808,通过用户响应获取模块 U805 来判断用户对同一需要处理的字符串不予理睬的次数是否超过了某预置的阈值,如果是,则可以学习这个字符串作为用户新词,下次再出现该字符串时,判断模块将 U802 所述用户新词确定为不需要处理,提示模块 U803 将不予提示,以此来实现用户词学习功能,更能满足用户的个性化需求。

[0153] 其中,图 8 中的取词模块 U801 与图 7 中的取词模块 U701 相同。

[0154] 另外,由于该提供纠错提示的功能是由输入法系统实现的,而输入法系统在任何的编辑环境下都可以使用,因此在任何输入环境下都能够对已经展现在显示屏上的文字进行纠错提示。虽然一些文字编辑软件(如Microsoft Word等)也具有对已经展现在显示屏上的文字进行纠错提示的功能,但是并不是所有的文字都是在这类文字编辑软件输入的,例如,用户正在使用某聊天工具聊天,则用户在输入完一段文字之后,如果想要检查这段文字是否存在拼写或用词错误,则只能将这段文字拷贝到Word等编辑环境中检查,修改完毕后再拷贝回原来的聊天工具编辑框中。可见,纠错的功能主要依赖于Word等编辑环境。而如果实现了本发明实施例提供的输入法系统,则可以直接基于输入法对已经展现在显示屏上的文字进行纠错,从而大大减少了用户对Word等编辑环境的依赖。这样能够提升用户在不同场景(聊天、BBS发帖等非正式文字编辑场合)下输入文字的质量。

[0155] 与提供纠错提示的装置类似,参见图9,该提供纠错提示的输入法系统的判断模块U902也可以包括以下子模块:

[0156] 比较子模块U9021,用于将获取的字符串与预置词库中的字符串进行比较,判断判断预置词库中是否存在与该获取的字符串完全匹配的字符串;

[0157] 第一确定子模块U9022,用于将所述与预置词库中的字符串均不完全匹配的字符串确定为所述需要处理的字符串。

[0158] 通过上述两个子模块,可以发现存在拼写错误的字符串,在本发明的优选实施例中,也可以发现存在用词错误的字符串,因此,该输入法系统的判断模块U902也可以包括以下子模块:

[0159] 上下文获取子模块U9023,用于获取显示屏上光标位置的字符串的上下文;

[0160] 分析子模块U9024,用于分析所述展现在显示屏上的字符串的上下文,判断是否存在用词错误的字符串;

[0161] 第二确定子模块U9025,用于将所述用词错误的字符串确定为所述需要处理的字符串。

[0162] 判断模块U902可以同时包括用于发现拼写错误的子模块及用于发现用词错误的子模块,这样判断模块U902既可以判断出拼写错误,又可以判断出用词错误。

[0163] 其中图9中的取词模块U901、提示模块U903与图7中的取词模块U701、提示模块U703相同。

[0164] 为了更好地理解本发明实施例提供的技术方案,下面通过使用输入法系统进行纠错提示的具体实例进行详细的介绍。

[0165] 小明最近找了一个澳大利亚的网友,偶尔跟他聊聊天,希望借此提升一下自己的英语水平。虽然聊天不算正式,他还是希望能尽可能少犯错误,避免沟通中不必要的误解。因为对方聊天打字的速度较快,小明由于不习惯英文输入等原因无法在输入过程中发现拼写错误或用词错误。

[0166] 这时使用本方案可以很好的解决这个问题。在输入的过程中,可以使用该输入法对英文词条的补全功能,在输入完一句话复查时,小明使用预置的纠错快捷键可以很快的发现句子中的拼写错误或用词错误,例如,其中一个拼错的单词是“Englsih”。输入法帮助小明定位到这个错误后。小明使用退格依次删除了h、i、s三个字母。由于这是针对之前输入法取得的已经展现在显示屏上的文字的修改,同时输入法根据预加载词库的前缀匹配情

况判定“Engl”是合法前缀,于是给出“English”、“England”等若干候选,而不是从头开始接收用户的输入并以此作为前缀进行匹配。

[0167] 根据与原始输入“Englsih”的匹配程度,“English”排在候选项第一位。小明自己也意识到“English”才是正确的输入,并直接选择以完成修改。接下来他调用输入法中寻找错误的快捷键,迅速定位到下一个错误位置进行修改。输入法发现之前的用户词库存在“Englsih”这样被用户确认的错误拼写,于是将该词条删除,代之以“English”并将其使用频率加1。接下来输入法又定位到输入中的下一个错误。通过本方案,小明能够很畅快地输入,完成与国外网友聊天的输入内容。

[0168] 本发明所述输入法系统是指计算机用户使用键盘、手写板等输入工具向计算机输入文字、符号的编辑工具。输入法使用一定编码规则,将用户输入序列(拼音序列、五笔、手写板输入等)转化为机器可处理的文字编码。输入法是用户尤其非罗马字符文字用户进行计算机信息处理的入口程序。这里涉及的输入法系统是与平台无关的,不具体到PC、无线或其他终端设备上的任一特定操作系统或平台。

[0169] 本发明中的提供纠错提示的方法主要以英文的输入为例进行介绍,但是该方法同样适用于中文、日文、法语等其他语言文字。对于中文、日文等需要进行字符转换的文字而言,区别在于字符的表示方法与英文不同,可以另外引入分词的步骤以切分字词边界。而对于法语、德语、西班牙语等罗马字符文字,则可以直接应用本发明提供的方法。

[0170] 以上对本发明所提供的一种提供纠错提示的方法、装置及输入法系统,进行了详细介绍,本文中应用了具体个例对本发明的原理及实施方式进行了阐述,以上实施例的说明只是用于帮助理解本发明的方法及其核心思想;同时,对于本领域的一般技术人员,依据本发明的思想,在具体实施方式及应用范围上均会有改变之处。综上所述,本说明书内容不应理解为对本发明的限制。



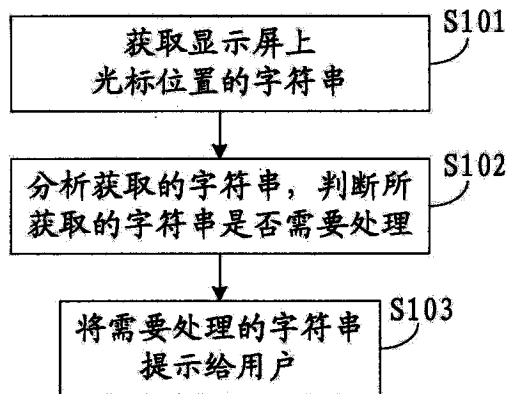


图 1



图 2

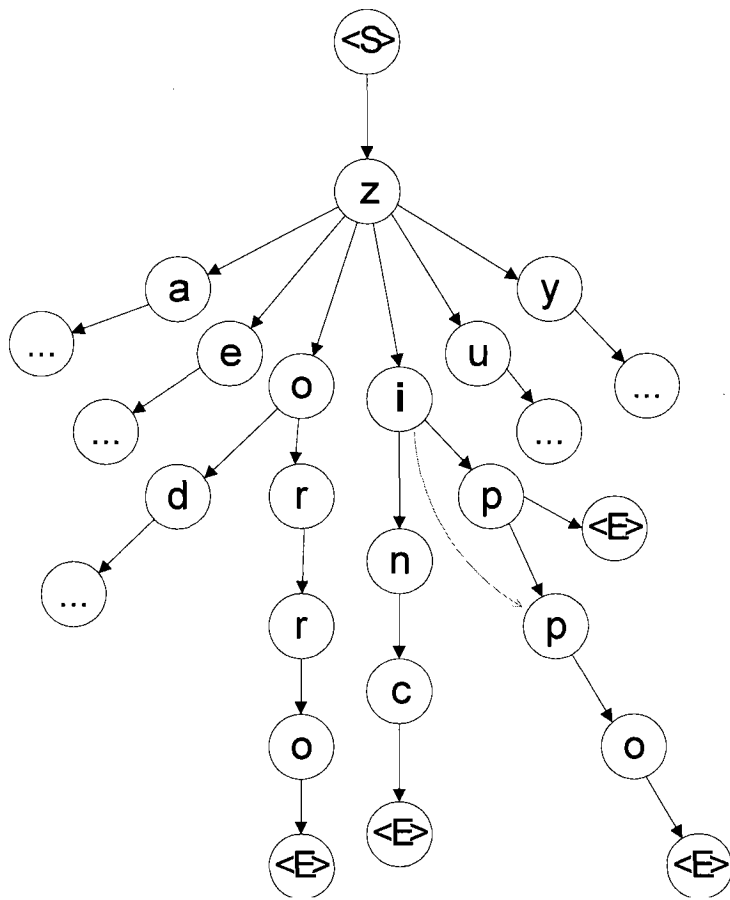


图 3

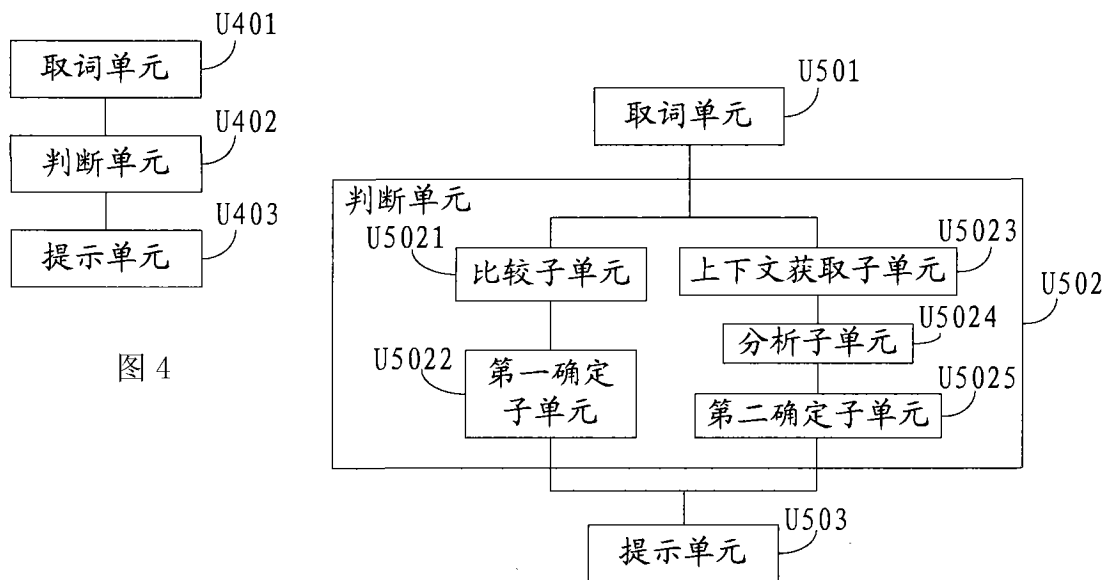


图 5

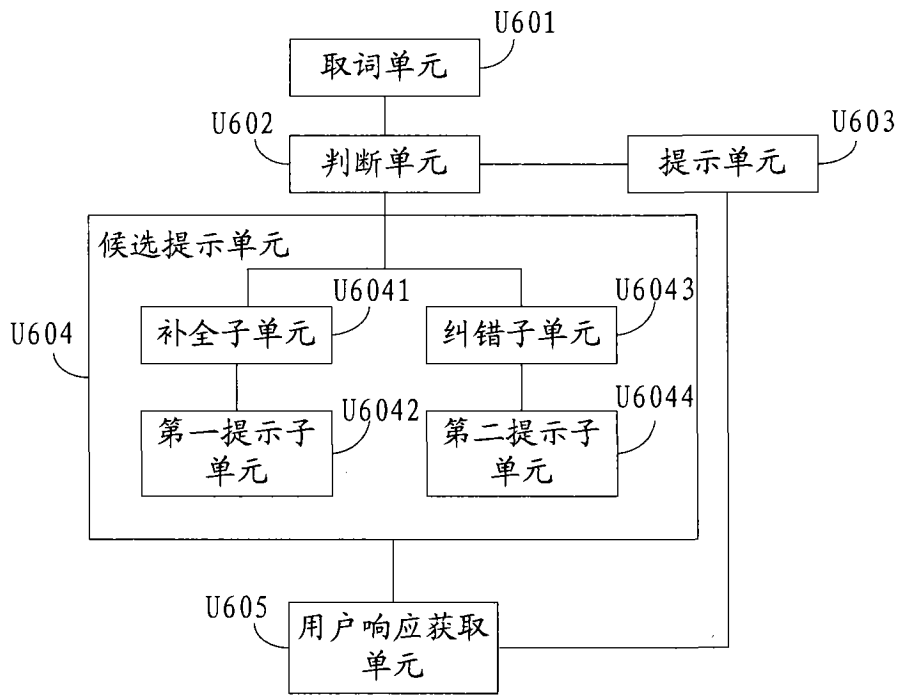


图 6

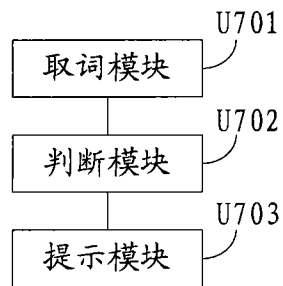


图 7

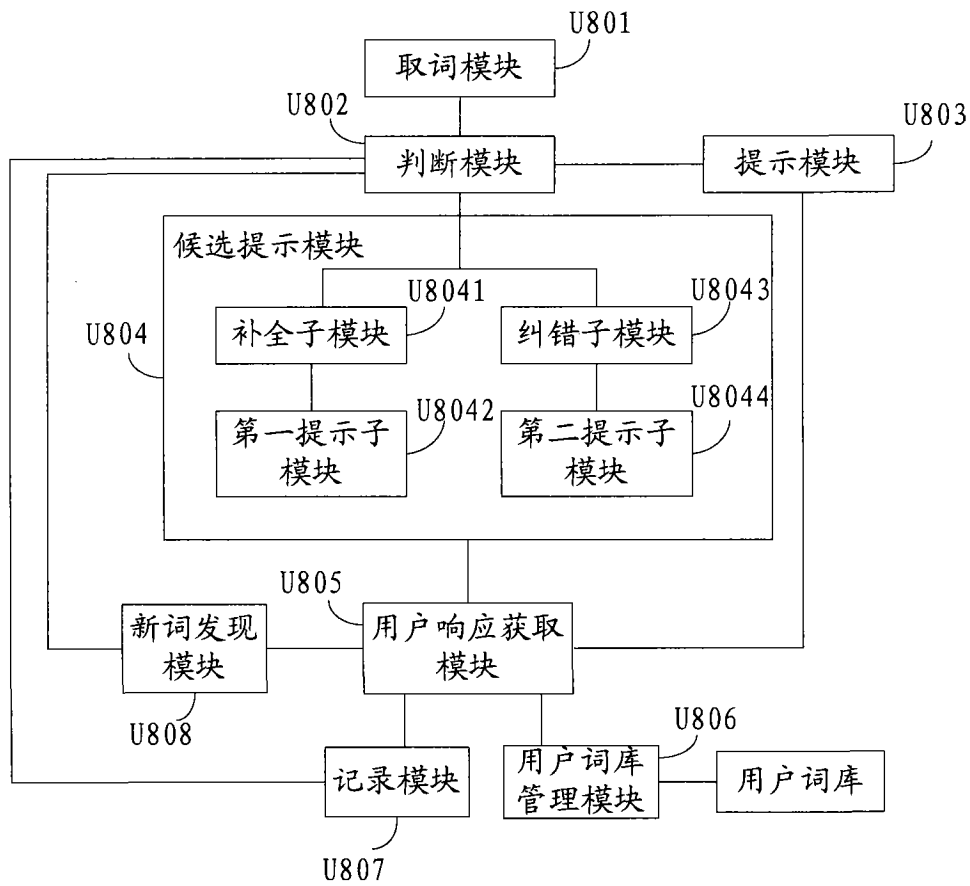


图 8

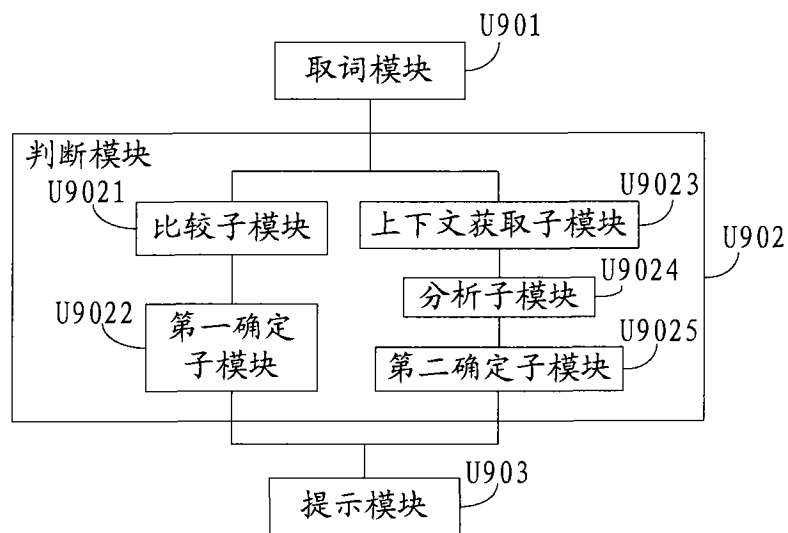


图 9