

[19] 中华人民共和国国家知识产权局



[12] 发明专利说明书

专利号 ZL 03800789.4

[51] Int. Cl.

G06F 17/28 (2006.01)

G06F 17/21 (2006.01)

G06F 3/02 (2006.01)

G06F 15/00 (2006.01)

G09G 5/00 (2006.01)

[45] 授权公告日 2008 年 10 月 22 日

[11] 授权公告号 CN 100428224C

[22] 申请日 2003.7.17 [21] 申请号 03800789.4

[30] 优先权

[32] 2002.7.18 [33] US [31] 60/397,253

[32] 2003.7.16 [33] US [31] 10/621,864

[86] 国际申请 PCT/US2003/022525 2003.7.17

[87] 国际公布 WO2004/010326 英 2004.1.29

[85] 进入国家阶段日期 2004.2.5

[73] 专利权人 美国在线服务公司

地址 美国弗吉尼亚州

[72] 发明人 伊斯安·布哦福特 大卫·彰·柯

[56] 参考文献

US6286064B1 2001.9.4

US6307549B1 2001.10.23

US6002390A 1999.12.14

CN1195409A 1998.10.7

US6223059B1 2001.4.24

审查员 王燕_1

[74] 专利代理机构 北京科龙寰宇知识产权代理有限公司

代理人 孙皓晨

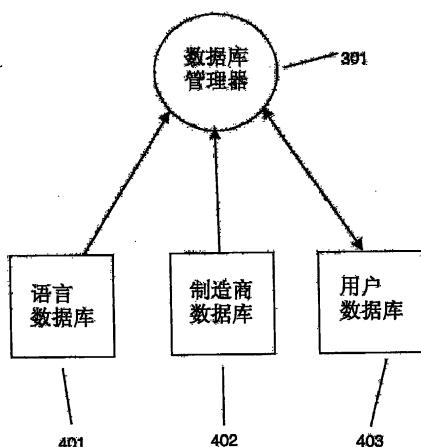
权利要求书 7 页 说明书 11 页 附图 5 页

[54] 发明名称

动态数据库重排序的设备和方法

[57] 摘要

本发明的动态数据库重排序系统提供了一种语言数据库，该语言数据库包括根据语言模型排序的单词，而所述语言模型则决定单词被呈现给用户的顺序。在一个用户通过按压按键在通信装置的小键盘上输入键击的同时，本发明预测用户正在试图输入的单词、字母、数字或者截短字。本发明通过跟踪用户用词选择，根据用户系统使用情况对语言模型排序重排序。一旦一个单词已经被选作接续按键选择的结果(接续单词)，则将一个频度值赋于所选单词，并且通过按键序列的语言数据库的语言模型，对该单词进行第一排序。当重复接续相同单词时，该接续单词的频度值将变得大于第一显示单词的频度。接续单词和第一排序单词的按键序列的随后用户输入，将导致接续单词显示在由语言模型确定的第一排序的单词之前。



1、一种重排序从数据库检索的并显示给用户的条目的方法，包括以下步骤：

- 接受用户从键盘的输入；
- 提供一个语言数据库；
- 其中所述语言数据库包含使用语言模型排序的单词；
- 从所述语言数据库动态检索单词，所述语言数据库包括通过用户键盘输入而形成的字母；
- 向用户显示所述已检索单词的目录；
- 其中使用语言数据库排序和动态重排序频度计数对所述目录中的单词排序；
- 其中对于所述目录中的单词，动态重排序计数优先于语言数据库排序；
- 将动态重排序频度计数分配给用户从所述目录中选择的单词。

2、根据权利要求 1 所述的方法，还包括以下步骤：

- 提供一个重排序数据库；
 - 其中在已提供所述重排序数据库之后，所述分配步骤将选择的单词和这些单词的相关联重排序频度插入到所述重排序数据库中。
- 3、根据权利要求 2 所述的方法，还包括以下步骤：
- 提供一个用户数据库；
 - 其中在已提供所述用户数据库之后，用户特别输入到系统中的单词被存储到所述用户数据库中；和
 - 其中在已提供所述用户数据库之后，所述检索步骤还从所述用户数据库中检索包括通过用户键盘输入所形成的字母的单词。

4、根据权利要求 2 所述的方法，其中，如果用户第一次选择一个非第一排序单词，则所述分配步骤将来自所述目录的第一排序单词和该非第一排序单词插入到所述重排序数据库中，其中如果第一排序单词未存在于所述重排序数据库中，则插入该第一排序单词。

5、根据权利要求 4 所述的方法，其中如果用户选择非第一排序单词一个预定次数，则所述目录中的第一排序单词失去其位置，然后该非第一排序单词被分配一个高于第一排序单词的频度值。

6、根据权利要求 4 所述的方法，其中输入到所述重排序数据库中的所有非第一排序单词最初被分配相等的重排序频度。

7、根据权利要求 1 所述的方法，其中每次用户选择一个单词时，就增加该单词的重排序频度。

8、根据权利要求 1 所述的方法，其中，如果用户选择所述目录中的一个单词，并且该单词位于第二排序位置之下，则所述分配步骤将单词的重排序频度指定到一个把该单词置于所述目录中第二排序位置的值。

9、根据权利要求 2 所述的方法，还包括以下步骤：

在已提供所述重排序数据库之后，对所述重排序数据库中的重排序频度周期性地执行老化处理；和

其中，在已提供所述重排序数据库之后，所述老化步骤利用一个预定因素减少所述重排序数据库中的重排序频度。

10、根据权利要求 3 所述的方法，还包括以下步骤：

在已提供所述重排序数据库之后，周期性地检查所述重排序数据库的自由空间；

其中，在已提供所述重排序数据库之后，如果所述重排序数据库的自由空间低于一个预定阈值，则从所述重排序数据库中移除具有低于预定阈值的重排序频度的单词。

11、根据权利要求 10 所述的处理，其中所述检查步骤从所述重排序数据库中移除用户定义的单词，该用户定义单词具有低于预定阈值的重排序频度，并且位于具有低于该预定阈值的重排序频度的其它单词之后。

12、根据权利要求 1 所述的方法，还包括以下步骤：

在使用语言数据库排序和动态重排序频度计数对所述目录中的单词排序之后且在将动态重排序频度计数分配给用户从所述目录中选择的单词之后，如果两个单词具有相同的重排序频度，则通过首先排序这两个单词中最新选择的单词来解决所述目录中的重排序频度碰撞。

13、根据权利要求 1 所述的方法，还包括以下步骤：

在提供所述语言数据库之后，在使用语言数据库排序和动态重排序频度计数对所述目录中的单词排序之后且在将动态重排序频度计数分配给用户从所述目录中选择的单词之后，如果两个单词具有相同的重排序频度，则通过首先排序所述语言数据库中具有较高排序的单词来解决所述目录中的重排序频度碰撞。

14、根据权利要求 2 所述的方法，还包括以下步骤：

在提供所述语言数据库之后，在使用语言数据库排序和动态重排序频度计数对所述目录中的单词排序之后且在将动态重排序频度计数分配给用户从所述目录中选择的单词之后，如果在用户定义单词和来自语言数据库的单词具有一个碰撞的情况下，如果两个单词具有相同的重排序频度，则通过首先排序用户定义单词来解决所述目录中的重排序频度碰撞。

15、根据权利要求 1 所述的方法，其中用户选择的不具备与其它单词碰撞可能的单词不分配重排序频度计数。

16、一种重排序从数据库检索的并显示给用户的条目的设备，包括：

一个接受源自键盘的用户输入的模块；

一个语言数据库；

其中所述语言数据库包含使用语言模型排序的单词；

一个从所述语言数据库动态检索单词的模块，所述语言数据库包括通过用户键盘输入而形成的字母；

一个向用户显示所述已检索单词的目录的模块；

其中使用语言数据库排序和动态重排序频度计数对所述目录中的单词排序；

其中对于所述目录中的单词，动态重排序计数优先于语言数据库排序；

一个把动态重排序频度计数分配给由用户从所述目录中选择的单词的模块。

17、根据权利要求 16 所述的设备，还包括：

一个重排序数据库；

其中所述分配模块将选择的单词和与它们相关联的重排序频度插入到所述重排序数据库中。

18、根据权利要求 17 所述的设备，还包括：

一个用户数据库；

其中用户特别输入到系统中的单词被存储到所述用户数据库中；和

其中所述检索模块还从所述用户数据库中检索包括通过用户键盘输入所形成的字母的单词。

19、根据权利要求 17 所述的设备，其中，如果用户第一次选择一个非第一排序单词，则所述分配模块将来自所述目录的第一排序单词和该非第一排序单词插入到所述重排序数据库中，其中如果第一排序单词未存在于所述重排序数据库中，则插入该第一排序单词。

20、根据权利要求 19 所述的设备，其中如果用户选择非第一排序单词一个预定次数，则所述目录中的第一排序单词失去其位置，然后该非第一排序单词被分配一个高于第一排序单词的频度值。

21、根据权利要求 19 所述的设备，其中输入到所述重排序数据库中的所有非第一排序单词最初被分配相等的重排序频度。

22、根据权利要求 16 所述的设备，其中每次用户选择一个单词时，就增加该单词的重排序频度。

23、根据权利要求 16 所述的设备，其中，如果用户选择所述目录中的一个单词，并且该单词位于第二排序位置之下，则所述分配模块将单词的重排序频度指定到一个把该单词置于所述目录中第二排序位置的值。

24、根据权利要求 17 所述的设备，还包括：

一个对所述重排序数据库中的重排序频度周期性地执行老化处理的模块；和

其中，所述老化模块利用一个预定因素减少所述重排序数据库中的重排序频度。

25、根据权利要求 17 所述的设备，还包括：

一个周期性地检查所述重排序数据库的自由空间的模块；

其中，如果所述重排序数据库的自由空间低于一个预定阈值，则从所述重排序数据库中移除具有低于预定阈值的重排序频度的单词。

26、根据权利要求 25 所述的设备，其中所述检查模块从所述重排序数据库

中移除用户定义的单词，该用户定义单词具有低于预定阈值的重排序频度，并且位于具有低于该预定阈值的重排序频度的其它单词之后。

27、根据权利要求 16 所述的设备，还包括：

一个解决所述目录中的重排序频度碰撞的装置；

其中，如果两个单词具有相同的重排序频度，则所述解决模块通过首先排序这两个单词中最新选择单词来解决一个碰撞。

28、根据权利要求 16 所述的设备，还包括：

一个解决所述目录中的重排序频度碰撞的装置；

其中，如果两个单词具有相同的重排序频度，则所述解决模块通过首先排序所述语言数据库中具有较高排序的单词来解决一个碰撞。

29、根据权利要求 16 所述的设备，还包括：

一个解决所述目录中的重排序频度碰撞的装置；

其中，如果在用户定义单词和来自语言数据库的单词具有一个碰撞的情况下，如果两个单词具有相同的重排序频度，则所述解决模块通过首先排序用户定义单词来解决一个碰撞。

30、根据权利要求 16 所述的设备，其中用户选择的不具备与其它单词碰撞可能的单词不分配重排序频度计数。

31、一种重排序从数据库检索的并显示给用户的条目的方法，包括以下步骤：

接受来自键盘的用户输入；

提供一个语言数据库；

其中所述语言数据库包含使用语言模型和用于每个单词的排序频度来排序的单词；

从所述语言数据库动态检索单词，所述语言数据库包括通过用户键盘输入而形成的字母；

向用户显示所述已检索单词的目录；

其中使用语言数据库排序和动态重排序频度计数对所述目录中的单词排序；

其中对于所述目录中的单词，动态重排序计数优先于语言数据库排序；

将动态重排序频度计数分配给用户从所述目录中选择的单词，并且将用于一个单词的所述已分配的动态重排序频度计数初始化到从用于该单词的所述语言数据库中获得的排序频度值。

32、根据权利要求 31 所述的方法，还包括以下步骤：

提供一个重排序数据库；

其中在已提供所述重排序数据库之后，所述分配步骤将选择的单词和与其相关联的重排序频度插入到所述重排序数据库中。

33、根据权利要求 32 所述的方法，还包括以下步骤：

提供一个用户数据库；

其中在已提供所述用户数据库之后，用户特别输入到系统中的单词被存

储到所述用户数据库中；和

其中在已提供所述用户数据库之后，所述检索步骤还从所述用户数据库中检索包括通过用户键盘输入所形成的字母的单词。

34、根据权利要求 33 所述的方法，其中分配步骤将用户特别输入的所有单词最初分配相等的重排序频度。

35、根据权利要求 32 所述的方法，其中如果用户第一次选择一个非第一排序单词，则所述分配步骤将来自所述目录的第一排序单词和该非第一排序单词插入到所述重排序数据库中，其中如果第一排序单词未存在于所述重排序数据库中，则插入该第一排序单词。

36、根据权利要求 35 所述的方法，其中如果用户选择非第一排序单词一个预定次数，则所述目录中的第一排序单词失去其位置，然后该非第一排序单词被分配一个高于第一排序单词的频度值。

37、根据权利要求 31 所述的方法，其中每次用户选择一个单词时，就增加该单词的重排序频度。

38、根据权利要求 31 所述的方法，其中，如果用户选择所述目录中的一个单词，并且该单词位于第二排序位置之下，则所述分配步骤将单词的重排序频度指定到一个把该单词置于所述目录中第二排序位置的值。

39、根据权利要求 32 所述的方法，还包括以下步骤：

在已提供所述重排序数据库之后，对所述重排序数据库中的重排序频度周期性地执行老化处理；和

其中，在已提供所述重排序数据库之后，所述老化步骤利用一个预定因素减少所述重排序数据库中的重排序频度。

40、根据权利要求 32 所述的方法，还包括以下步骤：

在已提供所述重排序数据库之后，周期性地检查所述重排序数据库的自由空间；

其中，在已提供所述重排序数据库之后，如果所述重排序数据库的自由空间低于一个预定阈值，则从所述重排序数据库中移除具有低于预定阈值的重排序频度的单词。

41、根据权利要求 40 所述的处理，其中所述检查步骤从所述重排序数据库中移除用户定义的单词，该用户定义单词具有低于预定阈值的重排序频度，并且位于具有低于该预定阈值的重排序频度的其它单词之后。

42、根据权利要求 31 所述的方法，还包括以下步骤：

在使用语言数据库排序和动态重排序频度计数对所述目录中的单词排序之后，且在将动态重排序频度计数分配给用户从所述目录中选择的单词并将用于一个单词的所述已分配的动态重排序频度计数初始化到从用于该单词的所述语言数据库中获得的排序频度值之后，如果两个单词具有相同的重排序频度，则通过首先排序这两个单词中最新选择的单词来解决所述目录中的重排序频度碰撞。

43、根据权利要求 31 所述的方法，还包括以下步骤：

在提供语言数据库之后，在使用语言数据库排序和动态重排序频度计数对所述目录中的单词排序之后，且在将动态重排序频度计数分配给用户从所述目录中选择的单词并将用于一个单词的所述已分配的动态重排序频度计数初始化到从用于该单词的所述语言数据库中获得的排序频度值之后，如果两个单词具有相同的重排序频度，则通过首先排序所述语言数据库中具有较高排序的单词来解决所述目录中的重排序频度碰撞。

44、根据权利要求 43 所述的方法，还包括以下步骤：

在提供语言数据库之后，在使用语言数据库排序和动态重排序频度计数对所述目录中的单词排序之后，且在将动态重排序频度计数分配给用户从所述目录中选择的单词并将用于一个单词的所述已分配的动态重排序频度计数初始化到从用于该单词的所述语言数据库中获得的排序频度值之后，如果在用户定义单词和来自语言数据库的单词具有一个碰撞的情况下，如果两个单词具有相同的重排序频度，则通过首先排序用户定义单词来解决所述目录中的重排序频度碰撞。

45、一种重排序从数据库检索的并显示给用户的条目的设备，包括以下：

一个接受来自键盘的用户输入的模块；

一个语言数据库；

其中所述语言数据库包含使用语言模型和每个单词的排序频度排序的单词；

一个从所述语言数据库动态检索单词的模块，所述语言数据库包括通过用户键盘输入而形成的字母；

一个向用户显示所述已检索单词的目录的模块；

其中使用语言数据库排序和动态重排序频度计数对所述目录中的单词排序；

其中对于所述目录中的单词，动态重排序计数优先于语言数据库排序；

一个把动态重排序频度计数分配给由用户从所述目录中选择的单词的模块，该模块还将用于一个单词的所述已分配的动态重排序频度计数初始化到，从该单词的所述语言数据库中获得的排序频度值。

46、根据权利要求 45 所述的设备，还包括：

一个重排序数据库；

其中所述分配模块将选择的单词和与其相关联的重排序频度插入到所述重排序数据库中。

47、根据权利要求 46 所述的设备，还包括：

一个用户数据库；

其中用户特别输入到系统中的单词被存储到所述用户数据库中；和

其中所述检索模块还从所述用户数据库中检索包括通过用户键盘输入所形成的字母的单词。

48、根据权利要求 47 所述的设备，其中所述分配模块将用户特别输入的所有单词最初分配相等的重排序频度。

49、根据权利要求 46 所述的设备，其中如果用户第一次选择一个非第一排序单词，则所述分配模块将来自所述目录的第一排序单词和该非第一排序单词插入到所述重排序数据库中，其中如果第一排序单词未存在于所述重排序数据库中，则插入该第一排序单词。

50、根据权利要求 49 所述的设备，其中如果用户选择非第一排序单词一个预定次数，则所述目录中的第一排序单词失去其位置，然后该非第一排序单词被分配一个高于第一排序单词的频度值。

51、根据权利要求 45 所述的设备，其中每次用户选择一个单词，就增加该单词的重排序频度。

52、根据权利要求 45 所述的设备，其中，如果用户选择所述目录中的一个单词，并且该单词位于第二排序位置之下，则所述分配模块将单词的重排序频度指定到一个把该单词置于所述目录中第二排序位置的值。

53、根据权利要求 46 所述的设备，还包括：

一个对所述重排序数据库中的重排序频度周期性地执行老化处理的模块；和

其中，所述老化模块利用一个预定因素减少所述重排序数据库中的重排序频度。

54、根据权利要求 46 所述的设备，还包括：

一个周期性地检查所述重排序数据库的自由空间的模块；

其中，如果所述重排序数据库的自由空间低于一个预定阈值，则从所述重排序数据库中移除具有低于预定阈值的重排序频度的单词。

55、根据权利要求 54 所述的设备，其中所述检查模块从所述重排序数据库中移除用户定义的单词，该用户定义单词具有低于预定阈值的重排序频度，并且位于具有低于该预定阈值的重排序频度的其它单词之后。

56、根据权利要求 45 所述的设备，还包括：

一个解决所述目录中的重排序频度碰撞的模块；

其中，如果两个单词具有相同的重排序频度，则所述解决模块通过首先排序这两个单词中最新选择的单词来解决一个碰撞。

57、根据权利要求 45 所述的设备，还包括：

一个解决所述目录中的重排序频度碰撞的模块；

其中，如果两个单词具有相同的重排序频度，则所述解决模块通过首先排序所述语言数据库中具有较高排序的单词来解决一个碰撞。

58、根据权利要求 43 所述的设备，还包括：

一个解决所述目录中的重排序频度碰撞的模块；

其中，如果在用户定义单词和来自语言数据库的单词具有一个碰撞的情况下，如果两个单词具有相同的重排序频度，则所述解决模块通过首先排序用户定义单词来解决一个碰撞。

动态数据库重排序的设备和方法

技术领域

本发明涉及对从数据库中提取的元素的排序。特别是，本发明涉及以用户实际选择的数据库元素的顺序对源自数据库的显示元素的排序。

背景技术

技术进步已经使制造商造出各种小型个人电子装置。一些例子是个人数字助理（PDA）、蜂窝电话、小形状系数的数据输入单元、和其它小形状系数的通信单元。

由于这些小型电子数据装置的尺寸减小，使得这些装置上的数据输入键盘的尺寸缩小。减小键盘尺寸的方案是减少键盘上按键的数量。减少按键的数量产生了许多问题。最明显的是，造成诸如蜂窝电话上的键盘超载。一个单键可以代表许多字符。当输入文本到缩减键盘的装置中时，使用户输入合理量的文本变得又慢又难。超载按键通常需要多个处理来获取正确字符。

键盘消除歧义系统（比如，本申请人所拥有的美国专利 US5818437、US5953541、US6011554 和 US6286064 公开的键盘消除歧义系统）通过处理用户键击并形成单词和将单词呈现给与按压按键关联的用户，来解决文本输入问题。完成的单词被呈现给开始于由按键按压所代表的字母的用户。呈现与按键按压关联的一列单词使用户节约了拼写整个单词的输入附加键击，同时还节省了时间。用户只需选择该列的第一个单词或者向下滚动并选择所需的单词。

呈现给用户的单词被存储在词汇数据库中。本申请人所拥有的美国专利

特征（排序特征）将极大地便于用户节约更多的时间，并且能够使用户更快和更有效地输入文本。

提供一个动态数据库重排序系统将是有利的，该系统能够根据用户实际的使用单词的情况，将与按键按压相关联的单词以一定的顺序显示给用户。提供一个不在主数据库中存储使用信息的频度的动态数据库重排序系统，也将是有利的。

发明内容

本发明提供了一种动态数据库重排序系统。本发明以基于用户实际使用单词（如果有的话）的顺序向用户显示与按键按压相关联的单词。此外，本发明不在主数据库中存储使用信息的频度，从而需要最小的存储空间。

本发明的优选实施例提供了一种包含根据语言模型排序的单词的语言数据库，该语言模型支配被呈现给用户的单词的排序。用户在通信装置的小键盘上输入键击。在用户按压按键输入的同时，本发明预测用户正在试图输入的单词、字母、数字或者截短单词（word stub）。完成的单词被动态地显示给用户，而那些单词开始于由按键按压所代表的字母。用户通常按压与数据库中一个以上单词相关联的按键序列。为了节省贮存语言数据库的空间，首先对语言数据库预排序，然后再置入产品中。

本发明提供了基于用户的系统使用情况对语言模型顺序的重新排序。如果有一个以上的单词共享相同的按键序列，则最常使用的单词作为第一选择显示于显示目录中。如果用户不需要该单词，而是要与按键序列相关联的另一个单词，则用户可以通过按压下一按键或者按压上/下翻滚按键，来翻滚单词的显示目录。一旦用户发现了所需的单词，用户就激活一个选择按键，于是系统在插入点将所需的单词输入到用户的文本消息中。

本发明的优先实施例跟踪用户的单词选择。一旦一个单词已经被选择为下一按键选择的结果（接续单词），就把一个频度值应用于该所选单词，并通过语言数据库中的语言模型将该单词列在该按键序列的关联单词的首位。

当一个单词第一次被用户接续选择（nexted），接续单词的频度值通常低于显示目录中第一排序单词的频度值。在该接续单词再次被接续选择时，相对于第一排序单词的频度值，该接续单词的频度值增加。每次选择一个单词

时，都会调整频度值。

本发明另一个实施例查找接续单词在显示目录中的位置。如果接续单词被定位在显示目录中的第二单词之后，则将该接续单词提升到第二单词位置。这使接续单词频度增加到第二单词的频度或者高于第二单词频度的频度，但是低于第一单词的频度。

当接续单词再三接续同一单词时，该接续单词的频度值将变得大于第一显示单词的频度值。随后，用户输入用于接续单词和第一排序单词的按键序列会导致，该接续单词显示在由语言模型第一排序的单词之前。

在本发明的一个实施例中，当一个单词被接续选择的次数为第一排序单词的三倍（或预定数）以上时，该单词的频度变得大于第一排序单词或者其它具有较高频度的单词。

本发明的其它方面和优点将通过下面结合附图的详细说明，变得更加清楚，其中附图通过实施例说明了本发明的原理。

附图说明

图 1 是根据本发明原理的具有简化键盘的便携通信装置的示意图；

图 2 是根据本发明原理的蜂窝电话键盘的示意图；

图 3 是根据本发明原理的一个任务观点的方框示意图；

图 4 是根据本发明原理的语言、制造商、用户数据库成分的方框示意图；

图 5 是根据本发明原理的存储于用户数据库中的频度信息的示意图。

具体实施方式

本发明被实施于动态数据库重排序系统。本发明的系统以基于用户实际使用单词（如果有的话）的顺序向用户显示与按键按压关联的单词。此外，本发明不在主数据库中存储使用信息的频度，从而需求最小的存储空间。

本发明的一个优选实施例提供了一种方法，该方法可向用户显示从语言数据库中检索得到的结果，而此结果与用户在键盘上的按键按压相匹配。将源自显示结果的由用户选择的任何单词赋予一由用户的单词使用模式确定的频度值。该频度值最好不存储在主语言数据库中，并且仅把用户实际使用的单词赋予一频度。用户使用的不会与其它单词碰撞的某些单词不需要有指定

的频度。本方法将所需的存储量减小到为通常跟踪用户的使用所需存储量的 1/7。

参见图 1，这里显示了一个具有简化键盘的个人通信装置 101。键盘消除歧义系统（比如本申请人拥有的美国专利 US5818437、US5953541、US6011554 和 US6286064 中公开的键盘消除歧义系统）解决了输入按键 102 超载和单一按键可代表几个字符的文本输入问题。在键盘 102 上用户的键击被处理并显示在 103 上。在用户按压按键的同时，系统预测用户正在试图输入的单词 106、107、字母 108、数字 109 或者截短字（未显示）。完成的单词动态地显示在 105 上，而该些单词开始于按键按压所代表的字母。

呈现与按键按压相关联的单词目录使用户省去了输入拼写整个单词的附加键击，并且还节省了时间。用户选择目录中的第一单词，或者向下翻滚目录并选择所需的单词。在该实施例中，用户可以按压空格键 112，以接受目录中的第一单词。所选单词出现在用户文本输入位置 104 中。用户还可以继续按压按键，以进一步缩减或者推敲所显示单词的选择。

呈现给用户的单词被存储在词汇数据库中。如果通过每个单词的使用频度对数据库中的单词整理，那么，相同的单词总是以相同的排序呈现在 105 上给用户。

本发明的 T9[®]语言数据库（LDB）包含根据语言模型排序的单词，该语言模型支配向用户呈现的单词的顺序。用户通常按压与数据库中一个以上单词关联的按键的一个序列。

为了节省存储 LDB 的空间，LDB 在装入产品之前被预先排序。利用一个语言模型对单词进行排序，该语言模型测量用于数据库中每个单词的通用频度值。数据库利用频度排序来汇编。编译数据库时，频度值将不与单词一起存储到数据库中，从而减少存储 LDB 的空间。

例如，图 2 显示了典型的蜂窝电话键盘 201。如果语言模型指示单词“in”具有高于“go”的频度，那么当按此排序选择传统蜂窝电话键盘上的 4 按键 202 和 6 按键 203 时，在显示目录中“in”出现在“go”之前。

本发明提供了根据用户对系统的使用情况对语言模型顺序进行重排序。

再参见图 1，如果一个以上的单词共享相同按键序列，则最常用的单词将被作为第一选择呈现在显示目录中。如果用户不需要该单词，而是要另一

个与按键序列关联的单词，那么用户可以通过按压 0 或者 NEXT 按键或者上/下翻滚按键（如果有的话），来翻滚单词的显示目录 105。一旦用户发现了所需的单词，则用户可激活一选择按键或者空格键 112，于是系统在插入点 104 将所需的单词输入到用户的文本消息 103 中。

本发明的优选实施例跟踪用户单词选择。一旦一单词被选择为 NEXT 按键选择的结果，就将一频度值应用于所选单词，并通过 LDB 的语言模型将该单词排在与该按键序列相关的单词列表的首位。应用于通过语言模型第一排序单词的频度值以及被选作 NEXT 按键(接续单词)结果的单词依赖于多个因素。一个示例因素包括接续单词相对于第一排序单词的使用通用性。

在用户接续选用一个单词的首次，在显示目录中该接续单词的频度值通常低于第一排序单词的频度值。该接续单词被再次接续选择时，相对于第一排序单词的频度值，该接续单词的频度值增加。当每次选择一个单词时，就调整频度值。

本发明的另一个实施例在显示目录中查找接续单词的位置。如果该接续单词被定位在显示目录中的第二单词之后，那么把接续单词提升到第二单词位置。这将接续单词的频度增加到第二单词的频度或高于第二单词的频度，但是低于第一单词的频度。下面将讨论碰撞问题。

在接连接续相同单词的某个点上，接续单词的频度值将变得大于第一显示单词的频度值。这样，对于接续单词和第一排序单词的按键序列的随后的用户输入将导致：在被由语言模型第一排序的单词之前显示该接续单词。

在本发明的一个实施例中，当选择一个单词的次数是第一排序单词的三倍（或预定数）以上时，该单词的频度变得大于第一排序单词或者具有较高频度的其它单词。

在本发明的另一个实施例中，将单词和该单词的排序存储到语言数据库中，并且将作为接续结果的频度值存储到一个单独数据库（一个重排序数据库）中。重排序数据库中由频度值标识的排序的优先级高于语言数据库中单词的排序。当频度值被存储在语言数据库中时，语言数据库中的频度为重排序数据库中的频度提供初始值。

仍在本发明的另一实施例中，如果语言数据库中没有存储频度值，则语言数据库中的单词的排序用来在重排序数据库中合成频度的初始值。

参见图 3，图中示出了本发明的文本处理器的一部分。键盘管理器 303 监视用户的按键按压。每个按键按压被发送到数据库管理器 301。数据库管理器 301 收集每个按键按压，并执行预测单词处理。

数据库管理器 301 访问在主机装置的存储装置 304 上的语言数据库，并且从匹配按键按压的语言数据库中提取第一 n 个单词，形成预测单词目录。n 值取决于制造商或者用户推荐的目录的长度。如果目录太长，则用于翻滚目录的键击数量将大于打字输入完整单词所需的按键按压的总量。数据库管理器 301 将单词目录传送到显示管理器 302。显示管理器 302 向用户显示单词的目录。

当每次按压翻滚按键或者选择按键时，键盘管理器 303 通知显示管理器 302。显示管理器 302 利用翻滚按键按压增亮显示目录中的适当单词。当用户按压选择按键时，显示管理器 302 将所选单词插入到用户的文本输入字段中，并通知数据库管理器 301 已选择了该单词。

如果一个单词正在被跟踪，数据库管理器 301 则调整所选用单词的频度记录。该频度通过各用户单词使用模式来设置。本发明不跟踪使用的每个单词。一些单词不会发生碰撞并且不需要对频度计数。本发明对因频度值彼此相等而出现的碰撞排序。该技术方案通常将存储频度数据所需的存储量减至典型技术方案所需的 1/7。

显示管理器 301 必须限制所跟踪单词的频度，因为如果不限制，计数可能超过装置上寄存器的存储容量。系统必须适应用户的使用变化。本发明的老化算法能完成已记录频度的计算，并扣除较老的使用频度。较老的使用不反映用户的当前习惯，因此价值低。

该部分详细说明了学习用户使用模式以及生成模仿用户使用模式的显示部分目录的过程。

参见图 4，本发明数据库实施的一个优选实施例包含：一个语言数据库 (LDB) 401，一个制造商数据库 (MDB) 402，和一个用户数据库 (UDB) 403。语言数据库 401 是一个固定的预编辑数据库，含有按照语言模型测量的单词使用频度排序的单词。

制造商数据库 402 是 OEM 提供的定制数据库。制造商数据库 402 是任选的，并且是在除语言数据库 401 中备有的数据库外，OEM 需要一个附加的定

制数据库时提供。

用户数据库 403 包含用户定义的多个单词以及一个重排序数据库(RDB)，该重排序数据库是跟踪用户选择的单词以及这些单词频度的用户数据库 403 的一个区域。图 5 显示了某些信息 501 的示例性说明，这些信息包含在跟踪用户单词使用的用户数据库 403 中。该信息列出了每个单词 502 的所有字母、单词 503 的使用频度、单词 504 的输入按键序列、包含在语言数据库 401 中的单词在 LDB505 中的对象编号。

下面各项在文本中按以下方式使用：

UDB 重排序单词—一个经由单词选择添加到 RDB 上的 LDB 单词。

有效单词—一个具有匹配当前 LDB 的 LDB 字段的 UDB 重排序单词。当下面讨论老化和垃圾收集算法时，有效单词的概念是重要的。

最近删除截止频度—这是利用垃圾收集算法从 RDB 中删除的最近 UDB 重排序单词的频度。

重排序单词—重排序单词是其频度大于或等于最近删除截止频度的单词。这些单词将排在选择显示目录的首位，它们包括 UDB 添加单词和 UDB 重排序单词。

UDB 添加单词—UDB 添加单词和 UDB 单词具有相同含义；它们涉及已经被用户添加到 UDB 上的单词，该单词或者是由用户直接添加，或者是由系统检测到的由用户构成的新单词。

重排序数据库 (RDB) —存储 UDB 重排序单词的 UDB 部分。RDB 驻留在与 UDB 相同的存储器内或者存储空间中。

第一单词和非第一单词—第一单词和非第一单词都是 LDB 单词。相对于 LDB，第一单词是第一显示的 LDB 单词，而非第一单词则不第一显示。

本发明的数据库管理器 301 通过保持跟踪多个 UDB 添加单词和每个 UDB 重排序单词的频度或者使用计数，学习用户的使用模式。根据单词的使用模式更新和设置每个单词的使用计数(频度)。使用计数被用来为垃圾收集、老化算法确定显示选择目录中的单词相对位置。

保持对 LDB 单词使用模式的跟踪的第一步是将该 LDB 单词添加到 RDB 中。当一个单词被接受时，踢开该添加处理。下面描述接受一个单词的按键事件。一旦一个 LDB 单词被添加到 RDB 上，就将其称作 UDB 重排序单词。表

1 描述了把单词添加到 RDB 上的规则。在表 1 中，列涉及选择目录中第一对象的类型，行涉及当前选择的对象类型。表中的编号涉及添加规则，如下表所示：

表 1-添加规则

	UDB 添加单词	LDB 单词	MDB 单词
UDB 添加单词	规则_0	规则_1	规则_0
LDB 单词	规则_3	规则_2	规则_3
MDB 单词	规则_0	规则_0	规则_0

基于当前选择对象类型和显示选择目录中的第一项对象类型，该表显示了以下规则：

- 规则_0 目录中的第一对象是 UDB 添加单词或是 MDB 单词以及所选对象也同样是 UDB 添加单词或是 MDB 单词的情况下。在这种情况下，显示选择目录中所选对象和第一对象都不会添加到 RDB 中。
- 规则_1 在选择目录中的第一对象是 LDB 第一单词以及选择对象是 UDB 添加单词的情况下，本发明将 LDB 第一单词（显示选择目录中的第一对象）添加到 RDB 上。
- 规则_2 该规则应用于显示选择目录中的第一对象是 LDB 第一单词以及所选对象是 LDB 非第一单词的时候。此时这两个单词都被添加到 RDB 中。
- 规则_3 当选择对象是 LDB 单词（LDB 第一或者 LDB 非第一单词）以及显示选择目录中的第一对象或者是 UDB 添加单词或者是 MDB 单词时，将所选的 LDB 单词添加到 RDB 中。

当接受有效单词时，基于添加规则，该单词被添加到或者不被添加到 RDB 上。如果该单词已经在 RDB/UDB 中，则提升该单词的频度。下面描述接受单词的事件。

- 当通过切换到一个新语言、注册或者不注册 MDB、或者终止数据库管理器来改变建立单词目录的上下文时。

- 当按压右箭头按键时。
- 当按压空格键时，当空格键或者箭头按键被按压后，随后再按空格键将不起任何作用，因为该单词已经被接受。
- 当用户输入标点或者与已知单词不匹配的显式字符时。

如上所述，当接受 UDB 添加单词时，该单词被添加到 UDB 上。初始频度将被设置为比第一单词高 3 个使用计数的值（可以被设置到与第一单词无关的任何值）。当使用该单词时，提升 UDB 添加单词的频度。

本发明通过对用户使用模式的渐进学习促使单词的排序，来使用“不侵犯”学习原则，其中非第一单词的单一使用将不会迫使第一单词到达显示选择目录中的第一位置。“不侵犯”原则可以通过仔细更新和设置频度来实现。

- 当把第一单词和非第一单词添加到 RDB 上时，设置第一单词的频度，以便对于第一单词失去显示选择目录中第一放置位置来说，将得到非第一单词的三次（或一个预定数）不对消的使用。用户按压空格键选择非第一单词的第一次被认为是一次使用。例如，如果正好添加了这两个单词，那么非第一单词的二次以上的使用将使该单词移动到显示选择目录的第一位置。
- 所有非第一单词开始于相同的初始频度。它们的频度增加或者减少将取决于每隔多长时间使用它们。频度在老化处理时减少。
- 如果从显示选择目录中选择一个第一单词，并且该第一单词已经在 RDB 中，则提升该单词的频度。
- 如果试图将第一单词添加到 RDB（当选择非第一单词时），并且第一单词已经在 RDB 中，则不增加其频度。假定该第一单词已经处于正确位置。
- UDB 添加单词的初始频度是高于第一单词的两个使用计数。可以为不同实施方式调整 δ 值。

在某些时间间隔，将老化所有的 UDB 添加和 UDB 重排序单词。老化是指通过某些因素减少这些单词的频度。每隔多长时间执行老化取决于所使用的时间间隔的单位，时间间隔的单位由数据库管理员维持。由于数据库没有时间概念，本发明通过维护一个更新计数试探地计算时间间隔。每当按压空格键选择一个单词时，该更新计数就递增 1。当更新计数达到某一值时，就踢开（中止）老化处理。这一过程按以下方式执行：

- 当更新计数达到其最大值（本实施例中是 1000），老化处理中止。在统计学中，1000 适用于一个快速 T9® 用户，该用户可以使用 T9® 键入 20wpm（每分钟 20 个单词），50 分钟内在持续的状态下输入文本。也就是约为 50 个信息，每条信息有 20 个单词（~120 个字符）。
- 所有单词的频度被老化的一个（31/32）因数减小。本领域熟练技术人员将容易地理解任何老化因数都可以用来实现预期的衰减速率。例如，如果非第一单词的频度是 54，则它将被减小到 52 ($54 \times (31/32)$)。

对于垃圾收集，“来得容易去得快”原则用于从 RDB 和 UDB 存储空间中先删除 UDB 重排序单词，然后删除 UDB 添加单词。建议使用因数 2 来删除重排序单词。这样，在删除具有 X 频度的 UDB 添加单词之前，本发明首先移除频度小于或等于 $2^* X$ 的所有重排序单词。下面说明垃圾收集算法：

- 1、移除已经被标上删除标记的所有 UDB 添加单词。
- 2、如果步骤 1 执行后自由空间的空间量大于或者等于总 UDB 数据空间的 1/8，则进行垃圾收集。
- 3、反之，增加最近删除截止频度，并删除具有该频度的所有 UDB 重排序单词和具有一半该频度的所有 UDB 添加单词。该处理持续进行直至 UDB 空间的 1/8 是自由空间。

本领域的熟练技术人员将容易理解可以调整 1/8 的阈值，以达到预期的 UDB 自由空间量。

新显示选择目录由第一（从顶部看）重排序单词、UDB、LDB 和 MDB 单词（依赖于 MDB 栅栏（fence））组成。MDB 栅栏用来设置可出现在 MDB 单词目录之前的 LDB 单词的最大量。这确保了 OEM 将可以显示其单词。显示选择目录中的重排序单词的数量由最近删除截止频度或者非第一单词初始频度所确定的，其数量总是较少的—称之为 RDB 计数。其频度大于或等于 RDB 计数的所有 UDB 添加单词和 UDB 重排序单词，是重排序单词，并且将出现在显示选择目录中的首位。下面分层次说明它们出现在显示选择目录中的排序：

- 1、其频度高于截止频度的所有重排序单词。
- 2、UDB 终结单词。
- 3、达到 MDB 栅栏的 LDB 单词。

4、MDB 终结单词。

5、UDB 词干。

6、MDB 词干。

当具有相同频度的重排序单词在竞争显示选择目录的前 5 个位置时，会涉及到连接中断器。下面的规则适用于中断连接：

1、如果两个 UDB 添加单词被连接，则添加到 UDB 上的最新单词获胜。

2、如果 UDB 添加和 UDB 重排序单词被连接，则 UDB 添加单词获胜。

3、如果两个 UDB 重排序单词被连接，则具有较小 LDB 对象编号的单词获胜。

每个 UDB 重排序单词与它的 LDB 对象编号一起被存储为按键序列。具有了它们来自哪个 LDB 的信息和消息，就可以重构单词。该技术使用了更少的存储器来存储 RDB 单词。一个字符只使用半个字节，而不是每字符使用一个或者两个字节。

一个 UDB 重排序单词将平均消耗 8 个字节-两个频度字节，一个长度字节，一个 LDB 对象编号字节，一个语言 ID 字节，加上用于六个字符单词的三个字节（平均单词长）。RDB 空间的 4K 字节能够容纳约 512 个 UDB 重排序单词，3K 字节将能够容纳约 384 个单词，2K 字节将能够容纳约 256 个单词，1K 字节容纳约 128 个单词。

用户可以接通和断开 RDB。下面描述其操作动作：

如果还未接通 RDB，则接通 RDB 功能部件。

断开 RDB 将具有以下两个效果：

- 显示选择目录回到它的原始排序。
- 在垃圾收集之前，RDB 中的任何 UDB 排序单词都将保留。

本领域熟练技术人员将容易理解，尽管在上文中已经使用了术语“单词”，但本发明同样适用于其它语言输入单位，比如音节（字节）或者词组（短语）。例如，在中文中可以输入单音节，在日本语中可以输入整个词组。

尽管已经参照优选实施例说明了本发明，但是本领域熟练技术人员将容易理解，在不背离本发明精神和范围的条件下，也可以用其它应用取代这里所述的应用。因此，本发明仅仅由其权利要求限定。

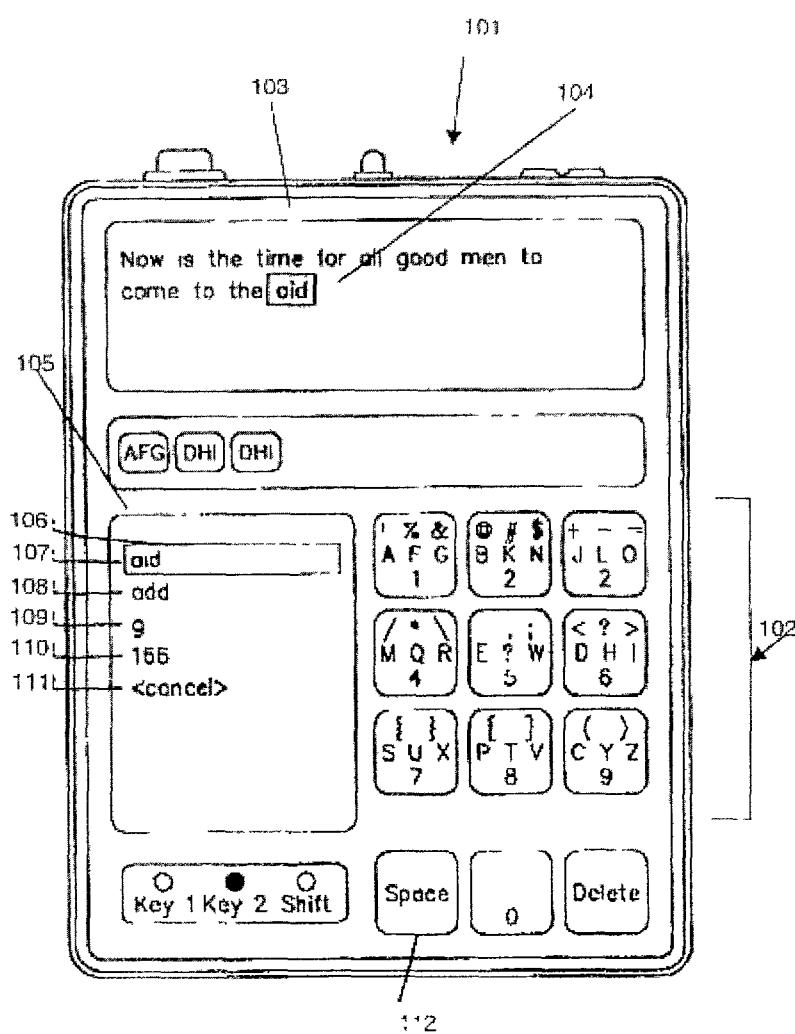


图 1

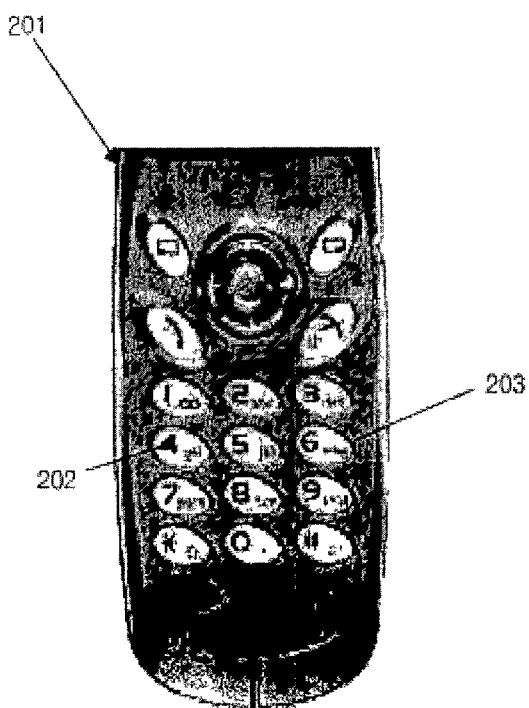


图 2

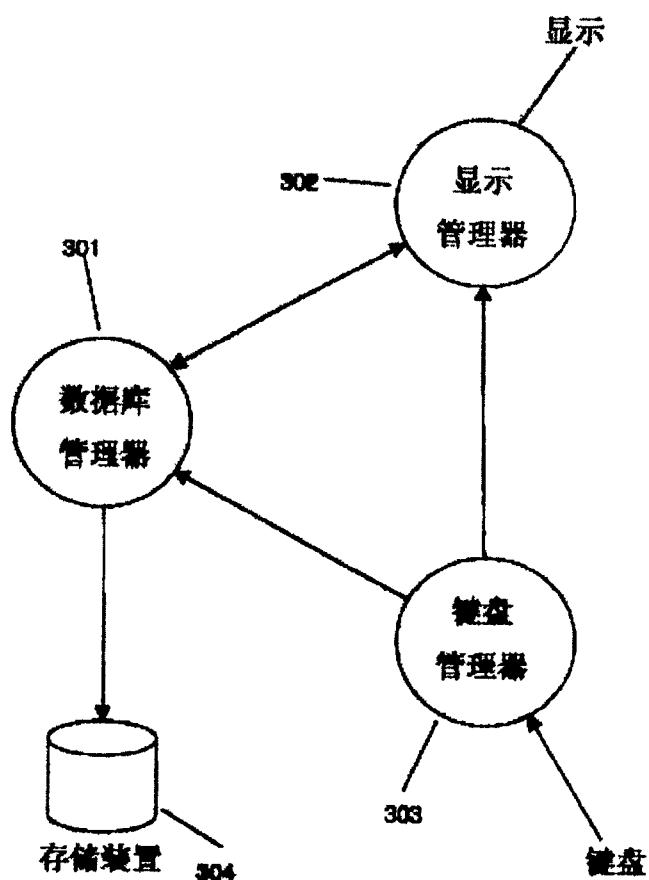


图 3

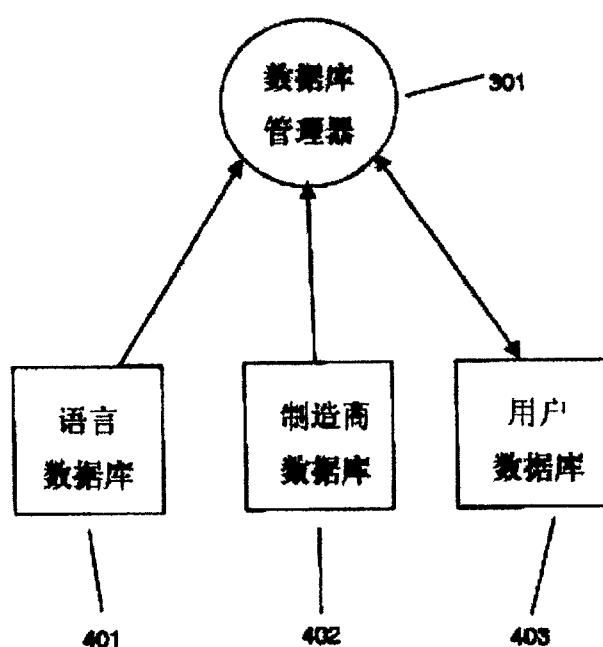


图 4

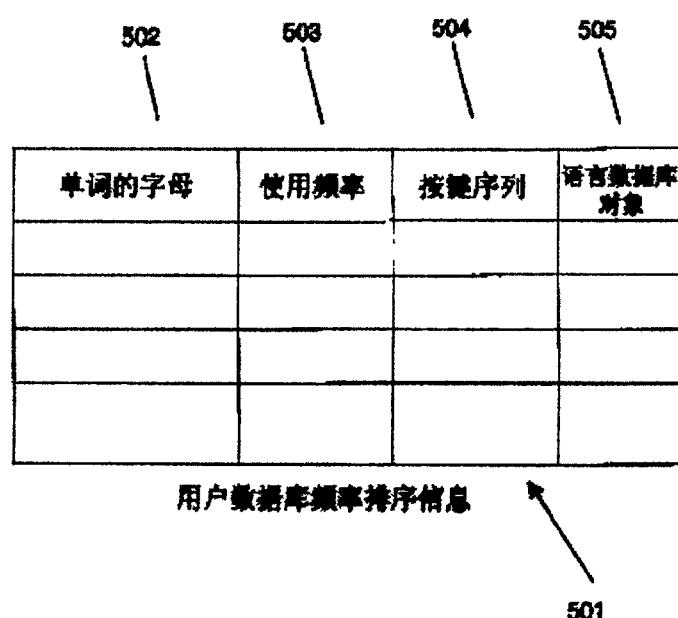


图 5