[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
G06F 17/30 (2006.01)

[12] 发明专利申请公布说明书

[21] 申请号 200710099474.6

[43] 公开日 2007 年 10 月 10 日

[11] 公开号 CN 101051323A

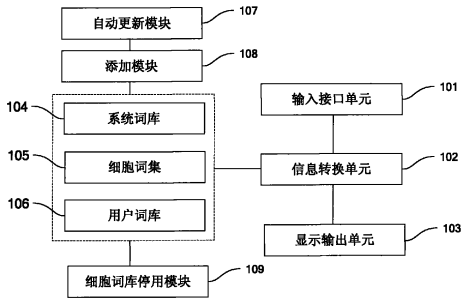
[22] 申请日 2007.5.22
[21] 申请号 200710099474.6
[71] 申请人 北京搜狗科技发展有限公司
地址 100084 北京市海淀区中关村东路 1 号院搜狐网络大厦 9 层 01 房间
[72] 发明人 马占凯 杨 磊
[74] 专利代理机构 北京集佳知识产权代理有限公司
代理人 逯长明
权利要求书 3 页 说明书 18 页 附图 2 页

[54] 发明名称

一种字符输入的方法、输入法系统及词库更新的方法

[57] 摘要

本发明提供了一种输入法系统，包括输入接口单元、信息转换单元和显示输出单元，还包括：系统词库，用于记录基础字词及其相关信息；细胞词集，用于记录扩展字词及其相关信息；所述细胞词集由从服务器端所存储的多个细胞词库中获取的至少一个细胞词库得到；每个细胞词库中的字词至少具有一个共同属性。 本发明在现有的输入法的词库容量级别上实现了动态的细胞式词库，用户通过手动或者由电脑自动添加小词库，通过每个人的个性化的选择或定制，通过自动更新和系统词库的联合使用，就能够覆盖一个用户几乎所有的词汇。 从而能够大幅提升输入法的首选词准确率，在理论上将个人的词库覆盖面扩大到最大，从而使打字的准确率有一个较大的提升。



1、一种输入法系统，包括输入接口单元、信息转换单元和显示输出单元，其特征在于，还包括:

系统词库，用于记录基础字词及其相关信息;

细胞词集，用于记录扩展字词及其相关信息;所述细胞词集由从服务器端所存储的多个细胞词库中获取的至少一个细胞词库得到;每个细胞词库中的字词至少具有一个共同属性。

2、如权利要求 1 所述的输入法系统，其特征在于，还包括:

自动更新模块，用于依据已有细胞词库列表，从服务器端获取所需的更新数据。

3、如权利要求 1 所述的输入法系统，其特征在于，所述细胞词集中存储的相关信息类型少于或等于所述系统词库中存储的相关信息类型;所述多个细胞词库中至少存在一个细胞词库由人工手动生成。

4、如权利要求 1 所述的输入法系统，其特征在于，还包括用户词库。

5、如权利要求 1 所述的输入法系统，其特征在于，还包括:

添加模块，用于将获取的细胞词库词条信息添加至所述细胞词集中;所述细胞词集为一个独立的词库或者为多个词库并列存在的词库集合。

6、如权利要求 5 所述的输入法系统，其特征在于，

所述添加方式为:完成更新下载一个细胞词库，则添加该细胞词库词条信息至所述细胞词集中;

或者，所述添加方式为:完成所有待更新细胞词库的下载后，才添加至所述细胞词集中。

7、如权利要求 6 所述的输入法系统，其特征在于，

所述添加过程在一独立的缓存词库中进行。

8、如权利要求 1 所述的输入法系统，其特征在于，还包括:

细胞词库停用模块，用于接收用户指令，从细胞词集中去除属于用户所选细胞词库的词条记录。

9、如权利要求 8 所述的输入法系统，其特征在于，

所述去除过程为:接收用户指令，将用户所选的细胞词库从列表中删除，并重新添加列表中的细胞词库，得到新的细胞词集;

或者，所述去除过程为：接收用户指令，从所述细胞词集中删除属于用户所选细胞词库的词条记录，所述细胞词集中记载有各词条所属的细胞词库；

或者，所述去除过程为：接收用户指令，在所述细胞词集中，向属于用户所选细胞词库的词条记录添加删除标记，所述细胞词集中记载有各词条所属的细胞词库。

10、一种字符输入的方法，其特征在于，包括：

加载系统词库和细胞词集；所述细胞词集用于记录扩展字词及其相关信息；所述细胞词集由从服务器端所存储的多个细胞词库中获取的至少一个细胞词库得到；每个细胞词库中的字词至少具有一个共同属性；

接收用户的输入信息；

依据所接收的输入信息，在所述系统词库和细胞词集中进行检索，得到相应的候选项；

接收用户的选择信息，将指定的候选项上屏输出。

11、如权利要求 10 所述的方法，其特征在于，

所述加载为：将细胞词集与系统词库合并为一个词库，置于缓存中；

或者，所述加载为：将细胞词集与系统词库作为两个或多个独立词库置于缓存中，并依据预置规则设定词库优先级；所述优先级用于候选项的显示排序。

12、如权利要求 10 所述的方法，其特征在于，所述细胞词集中记载有各词条所属的细胞词库以及相应的细胞词库优先级；所述优先级用于候选项的显示排序。

13、如权利要求 12 所述的方法，其特征在于，还包括：

在加载过程中，依据输入法的使用环境动态调整细胞词库优先级。

14、一种词库更新的方法，其特征在于，所更新的词库涉及用于记录扩展字词及其相关信息的细胞词集，所述细胞词集由从服务器端所存储的多个细胞词库中选取的至少一个细胞词库得到；每个细胞词库中的字词至少具有一个共同属性；

所述方法包括：

接受触发，比较已有细胞词库列表和服务器端细胞词库列表，得到所需更新的词库列表；

下载所需更新的细胞词库词条信息，并添加至细胞词集中。

15、如权利要求 14 所述的方法，其特征在于，还包括:

手动或者自动升级服务器端所存储的细胞词库，并更改相应的版本信息。

16、如权利要求 14 所述的方法，其特征在于:

所述添加方式为：完成下载一个细胞词库，则添加该细胞词库词条信息至所述细胞词集中;

或者，所述添加方式为：完成所有待更新细胞词库的下载后，才添加至所述细胞词集中。

17、如权利要求 16 所述的方法，其特征在于，所述添加过程在一独立的缓存词库中进行。

18、一种词库发布系统，其特征在于，包括:

细胞词库生成单元，包括：接口模块，用于接收输入信息；生成模块，用于依据所接收的信息生成细胞词库；标识模块，用于为每个细胞词库指定标识和版本信息；其中，每个细胞词库中的字词至少具有一个共同属性;

通信单元，用于接受触发，传输相应的细胞词库词条信息至客户端。

19、如权利要求 18 所述的词库发布系统，其特征在于，所述细胞词库生成单元还包括:

修改更新模块，用于修改更新细胞词库已存信息，并通知所述标识模块针对该细胞词库生成新的版本信息。

20、如权利要求 18 所述的词库发布系统，其特征在于，还包括:

识别模块，用于比较服务器端的细胞词库列表和客户端的细胞词库列表，所得到的比较结果用于传输所需的更新数据至客户端。

21、如权利要求 18 所述的词库发布系统，其特征在于，

依据所接收的信息得到的细胞词库中存储有多个词条信息;

或者，依据所接收的信息得到的细胞词库中存储有索引信息，所述索引信息对应其他细胞词库。

22、如权利要求 18 所述的词库发布系统，其特征在于，还包括:

合并模块，用于将多个细胞词库词条信息合并为一个下载词库，并通知通信单元将该下载词库传输至客户端。

一种字符输入的方法、输入法系统及词库更新的方法

技术领域

本发明涉及字符信息的输入领域，特别是涉及一种字符输入的方法、输入法系统以及一种词库更新的方法和一种词库发布系统。

背景技术

随着计算机技术以及互联网技术的普及与发展，不同专业领域、不同兴趣以及使用习惯的用户对于输入法系统的智能性要求越来越高。

在评价输入法智能性时，首选词的准确率是一个非常重要的评价标准，同时，候选项的排序也非常的重要，而记载有词条信息和词频信息的输入法词库是影响二者重要因素之一。因为用户所需的目标词在词库中存在，以及其相应的词频信息非常符合用户的使用习惯，则针对该用户的首选词准确率及候选项排序就会比较符合需求。

但是，目前输入法的词库一般只能够覆盖人们使用的词汇的一部分，通常主要包括一些人们普遍的常用词汇，还有一部分词汇输入法词库是不可能全部包括进来的。因为现有的输入法词库都是标准的，针对是所有用户，如果把所有用户用的词汇都加入进来，那么输入法的词库容量将在数百万的量级。词库过大，同音字过多，候选项增加，不需要使用这些词的用户会受到干扰，并且，这样一个超大的词库势必大幅占用 CPU、内存等计算设备资源，对个人电脑来说是不能接受的。

例如，每个人在使用输入法时除了输入许多常用词汇之外（例如“现在”、“时间”、“多少”等），还会输入一小部分人用的词汇，例如：一些游戏名词“艾泽拉斯”“德鲁伊”，最新的电影“云水谣”等等。这些词汇对非常小的群体来说会经常输入，例如：魔兽世界玩家，化学专业的工程师，生物学的教师等等。但是这些词汇在总体用户中的使用比例特别低，现有模式下的输入法词库是不可能把这些词汇全部包括进去，这样就会导致现有技术下，用户输入上述这些小群体的常用词汇时的首选词准确率非常低，严重影响用户的使用体验以及其思想的表达。

总之，需要本领域技术人员迫切解决的一个技术问题就是：如何改进输

入法词库，使得其既可以满足现有计算设备的资源分配，又可以大大提高各个用户的输入效率。

发明内容

本发明所要解决的技术问题是提供一种新型的输入法词库模式以及整套的输入解决方案，能够满足现有计算设备的资源分配，不会占用更多计算资源，并且可以显著提高各个用户的输入效率。

为了解决上述问题，本发明公开了一种输入法系统，包括输入接口单元、信息转换单元和显示输出单元，还包括：

系统词库，用于记录基础字词及其相关信息；

细胞词集，用于记录扩展字词及其相关信息；所述细胞词集由从服务器端所存储的多个细胞词库中获取的至少一个细胞词库得到；每个细胞词库中的字词至少具有一个共同属性。

优选的，所述的输入法系统还可以包括：自动更新模块，用于依据已有细胞词库列表，从服务器端获取所需的更新数据。

优选的，所述细胞词集中存储的相关信息类型少于或等于所述系统词库中存储的相关信息类型；所述多个细胞词库中至少存在一个细胞词库由人工手动生成。

进一步，所述的输入法系统还可以包括：用户词库。

进一步，所述的输入法系统还可以包括：添加模块，用于将获取的细胞词库词条信息添加至所述细胞词集中；所述细胞词集为一个独立的词库或者为多个词库并列存在的词库集合。优选的，所述添加过程在一独立的缓存词库中进行。

进一步，所述的输入法系统还可以包括：细胞词库停用模块，用于接收用户指令，从细胞词集中去除属于用户所选细胞词库的词条记录。

根据本发明的实施例，还公开了一种字符输入的方法，包括：

加载系统词库和细胞词集；所述细胞词集用于记录扩展字词及其相关信息；所述细胞词集由从服务器端所存储的多个细胞词库中获取的至少一个细胞词库得到；每个细胞词库中的字词至少具有一个共同属性；

接收用户的输入信息；

依据所接收的输入信息，在所述系统词库和细胞词集中进行检索，得到相应的候选项；

接收用户的选择信息，将指定的候选项上屏输出。

其中，所述加载为：将细胞词集与系统词库合并为一个词库，置于缓存中；或者，所述加载为：将细胞词集与系统词库作为两个或多个独立词库置于缓存中，并依据预置规则设定词库优先级；所述优先级用于候选项的显示排序。

优选的，所述细胞词集中记载有各词条所属的细胞词库以及相应的细胞词库优先级；所述优先级用于候选项的显示排序。

进一步，所述的方法还可以包括：在加载过程中，依据输入法的使用环境动态调整细胞词库优先级。

依据本发明的另一实施例，还公开了一种词库更新的方法，所更新的词库涉及用于记录扩展字词及其相关信息的细胞词集，所述细胞词集由从服务器端所存储的多个细胞词库中选取的至少一个细胞词库得到；每个细胞词库中的字词至少具有一个共同属性；所述方法包括：

接受触发，比较已有细胞词库列表和服务器端细胞词库列表，得到所需更新的词库列表；

下载所需更新的细胞词库词条信息，并添加至细胞词集中。

进一步，所述的方法还可以包括：手动或者自动升级服务器端所存储的细胞词库，并更改相应的版本信息。优选的，所述添加过程在一独立的缓存词库中进行。

依据本发明的另一实施例，还公开了一种词库发布系统，包括：

细胞词库生成单元，包括：接口模块，用于接收输入信息；生成模块，用于依据所接收的信息生成细胞词库；标识模块，用于为每个细胞词库指定标识和版本信息；其中，每个细胞词库中的字词至少具有一个共同属性；

通信单元，用于接受触发，传输相应的细胞词库词条信息至客户端。

进一步，所述细胞词库生成单元还可以包括：修改更新模块，用于修改更新细胞词库已存信息，并通知所述标识模块针对该细胞词库生成新的版本信息。

进一步，所述的词库发布系统还可以包括：识别模块，用于比较服务器

端的细胞词库列表和客户端的细胞词库列表，所得到的比较结果用于传输所需的更新数据至客户端。

优选的，依据所接收的信息得到的细胞词库中存储有多个词条信息；或者，依据所接收的信息得到的细胞词库中存储有索引信息，所述索引信息对应其他细胞词库。

进一步，所述的词库发布系统还可以包括：合并模块，用于将多个细胞词库词条信息合并为一个下载词库，并通知通信单元将该下载词库传输至客户端。

与现有技术相比，本发明具有以下优点：

本发明将现有技术中面向所有用户的标准输入法词库改进为由系统词库和细胞词集两部分构成，其中，系统词库仍面向所有用户，以通用词汇为主，而细胞词集部分则通过服务器端提供多个细胞词库，由用户选择最合适自己的，然后合并得到。因此，可以保证最后该用户使用的输入法词库仍然在现有的词库容量级别上，而又通过每个人的个性化的选择和使用，使得其基本能够覆盖一个用户几乎所有的词汇，并具有相对更准确的词频信息，从而可以大大提高首选词准确率，也可以实现更符合用户使用习惯的候选项排序。

本发明在现有的输入法的词库容量级别上实现了动态的细胞式词库，用户通过手动或者由电脑自动添加小词库，通过每个人的个性化的选择或定制，通过自动更新，和系统词库的联合使用，就能够覆盖一个用户几乎所有的词汇。这样就使用户可以输入几乎所有的词汇或句子，能够大幅提升输入法的首选词准确率。在理论上将个人的词库覆盖面扩大到最大，从而使打字的准确率有一个较大的提升。

本发明通过多个细胞词库的使用，并可以通过自动升级的方式来更新细胞词库，能够使个人的词库与时代同步。个人无需动手就能够保持词汇的新鲜度，从而在互联网日新月异的发展情况下，提高打字的首选词准确率，从而较明显的提高打字速度，降低生词的出现，降低翻页次数。

并且，本发明还提供了一个词库发布系统，用于帮助各用户手动生成自己所属群体的细胞词库，以及更新、修改该细胞词库；在客户端又增加

了自动更新功能，从而可以得到分类准确的细胞词库以及实现细胞词库的自动更新，使用户与世界保持一致，永不落伍。

附图说明

图 1 是一种输入法系统的实施例的结构框图；

图 2 是一种用于完成字符输入的方法实施例的步骤流程图；

图 3 是一种词库发布系统实施例的结构框图；

图 4 是一种词库自动更新的方法实施例的步骤流程图。

具体实施方式

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。

本发明可以应用于各种输入方式的输入法平台，包括键盘符号、手写信息以及语音输入等等。即所述输入信息可以包括编码字符串，也可以包括手写输入信息以及语音输入的信息，因为这些输入方式也都需要用到词库进行候选项排序。由于这些输入方式中的信息转换都属于公知技术，在此就不详述了。下面仅仅以编码字符串输入为例进行详细说明。

参照图 1，示出了本发明一种输入法系统的实施例，具体可以包括：

输入接口单元 101，用于接收用户输入的输入信息；

信息转换单元 102，用于根据用户输入的输入信息，例如，接收键盘字符，进行编码转换，得到相应的候选项；

显示输出单元 103，用于显示候选项，并接收用户选择，上屏输出。

系统词库 104，用于记录基础字词及其相关信息；

细胞词集 105，表示细胞词库的集合，用于记录扩展字词及其相关信息；所述细胞词集由从服务器端所存储的多个细胞词库中选取的至少一个细胞词库得到；每个细胞词库中的字词至少具有一个共同属性。

在字符输入的过程中，采用预置策略，检索系统词库和细胞词集，即可完成符合该用户个性化需求的输入过程。

所述细胞词库，具体含义为某一特定群体、某一个人或一部分人使用的具有某一共性的词库（即每个细胞词库中的字词至少具有一个共同属性），例如：

最新电影词库、最新歌名词库、魔兽世界词库、生物学词库、清华大学所有人名词库、某某公司全体人名词库、海淀区地名词库等。获得细胞词库的方式可以为：通过一个管理机构或者服务器群来自动分类、解析获得细胞词库；也可以为：提供一服务器平台，由用户自发的手动生成自己所述的群体的细胞词库，以更好的满足个性化群体的需求。即优选的，本实施例中的所述多个细胞词库中至少存在一个细胞词库由用户手动生成。

在现有技术中，输入法平台可以运行在多种计算设备上，例如，个人电脑、个人数字助理、移动终端设备等等，本发明也可以适用在上述各种计算设备中，对其运行环境并不需要加以限制。

下面简单介绍一下汉字、韩文、日文等需要编码转换的字符输入的过程，以中文输入为例：

在中文里，作为基本语言单位的汉字并不与键盘上的按键存在对应关系。因此需要输入法进行输入转换。首先需要通过汉字编码将汉字转换成能够直接输入的字母、数字等。通常是用的编码就是拼音（包括简拼、双拼、模糊音等各种形式）。用户将汉字的编码字符串通过键盘输入计算机（某些情况下也可能使用鼠标，比如软键盘）。用户的键盘输入通过操作系统交给输入法，输入法进行解码。由于不同的汉字序列（词、句）可能具有相同的编码，因此输入法通常提供一个候选列表供用户从中选择。例如，对于拼音输入法可能包含以下步骤：

a、拼音解析：切分输入字符串得到拼音，比如 zhuanli➔[zhuan][li]。当然，有时候这种切分不是唯一的，比如 fangan➔[fang][an]或者[fan][gan]（分别对应“方案”“反感”）。优选的，输入法可以支持简拼，允许用户以以下形式输入：zl, zhl, zhuanl, zhli, ….考虑到某些用户发音不标准，也可以支持模糊音：zuanli。另外还可以采用双拼等形式。

b、汉字解码。根据切分得到的拼音序列到词库中查找对应的字词，或者通过一定的算法生成对应的句子。

c、用户选择所需要的内容，上屏（可能还有造词、造句的过程）。

由于不同的汉字序列可能对应相同的编码，对于特定的编码字符串，输入

法需要猜测用户真实的意图。而要做到这一点，需要词库的支持。

对于本发明而言，词库可以包含各种语言信息，例如:

（1）词条

虽然也可以在字的基础上构建输入法，但由于词才是汉语中的最小表义单位，因此现代输入法大量使用了词条信息。例如用户分别输入“zhuan”这个拼音的时候，很难确定他究竟想输入“转专赚砖……”中的哪一个字。同样，用户输入“li”的时候，也很难确定他想输入的是“里李力利……”中的哪一个字。但是，如果用户连续输入“zhuanli”这两个音节，基本上可以断定用户想输入的就是“专利”这个词。这可以大大提高输入法首选的准确度。

（2）词频

同音字大量存在，同音词也仍然是存在的。遇到这种情况，只能把所有选项列出来供用户选择。但候选位置对输入法的易用性有很大影响。一般而言，把较常用的词放到靠前的位置会对用户更有利，即词频是候选排序的重要依据。

另外，现有的很多输入法中都集成了自动构造句子的功能。此时，词频信息也是句子构造的重要依据。

上面两种语言信息是输入法词库中不可或缺的，而本发明的输入法词库还可以包括其他一些对提高输入法准确度有利的信息，例如:

语言连接关系。输入法在构造句子的过程中，除了需要考虑词频，还需要考虑词和词之间的连接关系。例如“的”常出现在形容词、名词、代词等后面，而“地”则常出现在副词后面。在这种情况下，如果用户输入了“de”，是不能只看“的”“地”哪个词频更高的。

在词库中存放了输入法所需的语言信息。用户就可以完成字符输入了。但是，不同用户所需的语言信息并不相同。比如:

（1）词条不同。几乎每个行业都有自己特殊的词汇，这些词在其他领域是很少用到的，在构造输入法词库的时候可以不必考虑。例如计算机词汇“缓存”等等。

（2）词条重要程度不同。不同的用户可能需要用到相同的词，但其重要

性却随用户的不同而不同。比如同音词“研究”和“烟酒”，前者在学术领域使用较多，而后者则在日常生活中使用较多。但两者都是可能用到的，因此当用户输入拼音“yanjiu”时，都会出现在用户的候选列表中。由于重要性不同，候选位置的相对大小会影响用户的直观感受。

对于词条相对于用户的重要程度，可以通过各种方式单独使用或者组合应用，在词库中加以体现，例如：

词频信息。词频信息通常用一个数字表示，用来表示这个词的使用频繁程度；一般使用越频繁的词词频越高。

词序信息。词序信息通常也是一个数字，但只用于表示该词条重要程度的相对含义。

或者，位置信息。为了方便，也可以省略这个数据，而用词条在词库中的相对位置来表达词条的重要程度。例如，可以认为排在词库前面的词比排在后面的词更重要，从而将前者放在候选列表的前面。

由于输入法词库不可能针对每一个用户生成一个专用的词库，因此，本发明提出，将输入法词库划分为系统词库和细胞词集两部分。系统词库用于记载常用词汇，以满足大多数人在大多数情况下的输入需求，而对于某个用户的个性化需求，则通过细胞词集进行记载。为了提高细胞词集与每个用户的贴合度，通过手动或者自动的方式生成大量的细胞词库，然后由各个用户自行选择自己所需的细胞词库，得到细胞词集，这样的细胞词集与每一个用户的贴合度都是非常好的，因为个性化的部分是其自行选择的。

对于用户选择了一个细胞词库的情况，则该细胞词库可以直接构成细胞词集。

对于用户选择了多个细胞词库时，则细胞词集可以具有多种表现形式。例如：（1）在客户端，将所述的多个细胞词库合并成为一个词库，即细胞词集以一个独立词库的形式存在；该词库中可以存储各词条的来源（即所属细胞词库）信息，也可以不存储。（2）在客户端，将所述的多个细胞词库并列存储，即细胞词集以多个独立词库并存的形式存在，依次扫描该多个细胞词库即可。（3）在客户端，将所述的多个细胞词库中的一部分词库合并（例如，某些属性比较相近的词库），即细胞词集以多个独立词库并存的形式存在，但是其中某些独

立词库是由多个细胞词库合并得到的。

对于细胞词集而言，由于某些语言信息比较复杂，例如，语言连接关系等等，一是难以获得，二是难以存储，所以优选的，对于细胞词集而言（实际上包括各个细胞词库），其中存储的语言信息的类型要少于系统词库中所存储的语言信息的类型。当然，细胞词集中所存储的语言信息的类型也有可能多于系统词库中所存储的语言信息的类型，例如，对于词序信息或者位置信息，一般存储在细胞词库中，而系统词库中一般没有。

进一步，本实施例的输入法系统中还可以包括用户词库 106，用于记录该用户的输入习惯，以更好的满足该用户的个性化需求。

在服务器提供的平台上，存在大量的细胞词库，并且也会有大量的用户为了完善这些细胞词库，对其进行修改和更新，因此，如何将最新最好的细胞词库提供给选择该细胞词库的输入法用户使用，也是本发明需要解决的技术问题之一。

优选的，本实施例还可以包括：自动更新模块 107，用于接受触发，依据已有细胞词库列表，从服务器端下载所需的更新数据。例如，该用户的输入法系统中存储有正在应用的细胞词库的信息列表，然后与服务器端的信息进行比较，如果需要更新，则根据预置的更新策略，完成下载更新。所述的更新数据可以为整个细胞词库，例如，得知该细胞词库需要更新，则直接下载该细胞词库的所有词条信息；所述的更新数据也可以为一细胞词库中的部分词条信息，例如，得知该细胞词库需要更新，则通过词条比对，仅仅下载发生变化的词条信息。当然，服务器端还可以将多个细胞词库中发生变化的词条信息合并成为一个新词库作为更新数据。

如果用户选择了多个细胞词库，则服务器端可以将这多个细胞词库合并成为一个词库，然后发送至客户端作为细胞词集，即细胞词库的数据添加任务由服务器端完成。

如果用户选择了多个细胞词库，则对于细胞词库的数据添加由输入法系统自行完成的情况下，本实施例还可以包括：添加模块 108，用于将下载的细胞词库词条信息添加至所述细胞词集中。该添加模块 108 可以采用各种可行的添加策略，例如，所述添加方式为：完成更新下载一个细胞词库，则添加该细胞

词库至所述细胞词集中；或者，所述添加方式为：完成所有待更新细胞词库的下载后，才添加至所述细胞词集中。

该添加模块 108 可以用于细胞词集第一次形成的时候，或者其词库更新的时候。该添加模块 108 可以用于下载整个细胞词库的情况，也可以用于下载一细胞词库中的部分词条信息的情况。

优选的，如果词库添加过程能够在较短时间内完成（比如不超过 1 秒），由于影响不大，则可以直接将添加过程插入用户的输入过程中。但如果在较短时间内无法完成以致可能影响用户的使用感受，则词库添加过程应当在一个独立的缓存词库中进行。这个过程中输入法原来的词库不受影响，用户可以正常使用。当缓存词库创建完毕后，直接替换输入法原来的词库即可。由于这个替换过程可以很快，因此可以做到对用户的正常使用干扰降到最低。

优选的，为了进一步提高用户对词库的管理，本实施例还可以包括：细胞词库停用模块 109，用于接收用户指令（例如，通过点选菜单项等方式），从细胞词集中去除属于用户所选细胞词库的词条记录，达到将某个或者某些细胞词库停用的目的。

其中，所述的去除过程可以为：接收用户指令，将用户所选的细胞词库从列表中删除，并重新添加列表中的细胞词库，得到新的细胞词集。由于被删除的细胞词库已经不在列表中存在，新得到的细胞词集将不包含其中的词，效果上等价于该词库已经被删除。对于在细胞词集中独立存在的细胞词库而言，直接删除或者加上删除标记即可达到停用的目的。

或者，所述去除过程也可以为：接收用户指令，从所述细胞词集中删除属于用户所选细胞词库的词条记录，所述细胞词集中记载有各词条所属的细胞词库。或者，所述去除过程也可以为：接收用户指令，在所述细胞词集中，向属于用户所选细胞词库的词条记录添加删除标记，所述细胞词集中记载有各词条所属的细胞词库。

即作为细胞词集的大词库中记载了每个词条的来源，当用户指定删除某个细胞词库时通知输入法系统(或者其主动)将来自该词库的词条从词库中移除。这种移出可以是直接将该词条从数据结构中删除并释放其对应的空间，也可以通过一个删除标记实现。具有删除标记的词条在后续使用中将被忽略（不释放

空间，但实现起来会容易些）。这种方式的好处是，当细胞词库很多时删除少量词库而引起的系统开销会比较小。

参照图 2，示出了一种用于完成字符输入的方法实施例，具体可以包括：

步骤 201、加载系统词库和细胞词集；所述细胞词集由从服务器端所存储的多个细胞词库中选取的至少一个细胞词库得到；每个细胞词库中的字词至少具有一个共同属性；

步骤 202、接收用户的输入信息；

步骤 203、依据所接收的输入信息，在所述系统词库和细胞词集中进行检索，得到相应的候选项；

步骤 204、接收用户的选择信息，将指定的候选项上屏输出。

本实施例中比较重要的一个问题是，当多个词库并存时，如何完成候选项的检出。步骤 201 中所述的加载过程可以为：将细胞词集与系统词库合并为一个词库，置于缓存中。

输入法在启动的时候，扫描输入法系统中具有的系统词库和细胞词集，将二者合并为一个词库后载入缓存中，这样用户在后续操作中可以按照系统词库的使用方式直接使用。其中，系统词库的加载和细胞词集的加载可以分开进行，例如，简单情况下，用户仅需要加载系统词库即可，在某些情况下，用户选择或者输入法系统自动启动（例如，符合预置策略的情况下）触发启动细胞词集的加载，然后将细胞词集合并至系统词库，置于缓存中，用于用户输入时的检索。

进一步，步骤 201 中所述的加载过程也可以为：将细胞词集与系统词库作为两个或多个独立词库置于缓存中，并依据预置规则设定的词库优先级；所述优先级用于候选项的显示排序。

即在加载过程中，将细胞词集放到系统词库以外指定的空间，并在检索系统词库的同时也检索细胞词集。优选的，此时需要指定系统词库和细胞词集的优先级，例如，默认细胞词集的优先级高于系统词库，则输出候选项时，将所有属于细胞词集的词都强制放在属于系统词库的词的前面。

对于细胞词集为一个大词库存在时，即缓存中存在两个独立的词库。而对于细胞词集也由多个细胞词库独立组成时，则缓存中可能存在多个独立的词

库。当然，此时需要设定各个词库的优先级；所述优先级用于候选项的显示排序。

优选的，对于细胞词集为一个大词库存在时，为了体现各个细胞词库的不同，也可以在细胞词集中记载有各词条所属的细胞词库以及相应的细胞词库优先级。

对于针对各个细胞词库设置有优先级的情况（包括各个细胞词库独立存在和合并为一个大词库存在的情况），则优选的，在加载过程中，可以依据输入法的使用环境动态调整细胞词库优先级。例如，细胞词集包括有“办公用语”和“网络用语”两个细胞词库，正常情况下它们的优先级是相同的。但当输入法系统识别当前应用程序为 Word 字处理程序时，可以给“办公用语”细胞词库加权，而当用户切换到 QQ 聊天程序时，则可以给“网络用语”细胞词库加权。

参照图 3，示出了一种适用于前述输入法系统（为了清楚说明，采用输入法客户端一词进行描述）的词库发布系统实施例，该词库发布系统可以用于输入法客户端首次从服务器端下载细胞词库得到细胞词集的过程，也可以用于对已有细胞词库进行更新的过程。

图 3 所示的词库发布系统具体可以包括：

细胞词库生成单元 301，包括用于接收输入信息的接口模块 3011，用于依据所接收的信息得到细胞词库的生成模块 3012，以及用于为每个细胞词库指定标识和版本信息的标识模块 3013；每个细胞词库中的字词至少具有一个共同属性；

通信单元 302，一般位于服务器端，用于接受触发信息，传输相应的细胞词库词条信息至客户端。

细胞词库生成单元 301 中一般位于服务器端，用于统一管理和维护细胞词库。当然，细胞词库生成单元 301 中的部分或者全部模块也可以位于客户端（可以为独立于输入法客户端的其他客户端）中，例如，接口模块 3011 和生成模块 3012 位于客户端，用户可以直接将生成的细胞词库文件发送至服务器端即可，由服务器端完成指定标识和版本信息的工作。

所述的触发信息可以为用户的选择操作等，也可以是输入法系统客户端自动发送的触发信息，还可以为服务器端的自动检测触发。例如，服务器或者客

户端检索用户 IP 地址或者当前输入环境，而自动推荐相应的细胞词库给用户；或者，客户端发送的更新消息也属于触发信息的一种。

细胞词库的生成可以采用手动、自动等方式，下面对手动生成细胞词库的过程进行简单说明：

词库生成人员需要通过接口模块 3011（例如，包括以词库编辑页面）提供以下信息：名称、类别、条数、版本、说明、词库作者、词条举例、词条（包括读音信息）等等。当点击提交按钮后，这些信息被保存到数据库中。然后立即启用词库生成程序。最简单的，词库生成程序直接将这些信息以文本的方式保存到一个文件中供用户下载。

例如，一个细胞词库为一个文件，其中包含的数据可能有：

词库序号	00015214
链接网址	http://abc.com/dict/00015214
名称	魔兽世界
类别	游戏
条数	188
版本	0008
日期	2006.12.6
说明	我做的细胞哦。
词库作者	张三 李四
词条举例	艾泽拉斯……
词条，读音，词频数据	具体数据

为了提高细胞词库添加的效率，还可以对细胞词库的格式进行必要处理。例如对其内部的词条进行排序，当然，这些工作都可以在生成模块 3012 中完成，然后将词条排序后的数据文件作为细胞词库文件提供给用户下载。

出于版权信息保护等目的，还可以对细胞词库进行加密处理。对应的，需要在安装细胞词库时对其进行解密。即优选的，服务器端还可以包括一加密模块，输入法客户端还可以包括一解密模块。

为了便于更新，标识模块 3013 同时会为每一个细胞词库指定一个唯一 ID

和一个版本号。

图 3 所示实施例中的细胞词库可以具有多种表现形式，例如：一般情况下，细胞词库中直接存储多个词条信息；或者，细胞词库中也可以仅仅存储索引信息，所述索引信息对应其他细胞词库。存储索引信息的细胞词库一般可以应用于：服务器端存储有多个依据所接收的信息得到的细胞词库，然后根据这些细胞词库的某个共性，生成一个新的细胞词库（即间接利用所接收的信息），为了实现简便，则可以仅仅在该新细胞词库中存储索引信息即可，用户需要该词库时，再由服务器端合并各相应词库后进行传输。

进一步，为了满足细胞词库的快速更新，则本实施例中词库发布系统的细胞词库生成单元 301 还可以包括：修改更新模块 3014，用于修改更新细胞词库已存信息，并通知所述标识模块针对该细胞词库生成新的版本信息。所述修改可以为人工完成，也可以为依据一定的预置策略对细胞词库进行调整而完成，例如：其他用户向某个细胞词库中添加新的词条；或者，依据预置策略，将两个细胞词库中的词条合并为一个细胞词库；或者，依据互联网词频统计结果，将某个细胞词库中互联网词频不符合预置条件的词条进行删除或者进行排序调整。

图 3 所示的实施例至少可以通过以下两种方式完成细胞词库的数据添加。

一是先将细胞词库下载至本地，然后通过双击打开这个文件，完成数据的添加。细胞词库是带有某一特定后缀名的文件，例如.scd 后缀。当输入法系统在安装的时候，会通过注册表将.scd 后缀与一个特定的应用程序关联。当用户双击后缀为.scd 文件的时候，操作系统会根据这个关联规则启动对应的应用程序模块（例如，图 1 所示实施例中的添加模块），完成细胞词库数据的添加。

二是通过点击页面上的链接，直接在线完成细胞词库数据的添加。用户点击页面上的细胞词库链接后，有两种方式：保存和执行。如果用户保存了细胞词库文件，同前一种方式。如果用户选择了执行，系统会将细胞词库文件保存在系统的临时文件夹中，然后运行它。其内部实现机制和第一种方式也是相同的，区别在于文件被下载到了系统临时文件夹，因此不需要用户指定下载位置。同时，系统会在必要时对临时目录进行清理，因此虽然细胞词库已经被下载到

临时目录中，但实际对用户而言是不可见的。

优选的，将所下载的细胞词库添加至细胞词集的过程，还可以包含一个转换步骤，例如对词库中原来无序的词条进行排序以便提高添加的效率。如果存在这个转换步骤，将使用转换后的词库文件；否则直接使用原词库文件。当然，如果服务器端在词库生成过程中已经完成了转换排序的工作，则客户端在数据添加时就不需要重复了。

在数据添加过程中，输入法系统（即输入法客户端）需要维护一个当前所应用的细胞词库的列表。所述细胞词库列表可以采用各种可行的形式，例如，将所有活动的细胞词库拷贝到一个指定的目录中，或者保存一个文件名的列表（这个列表可以放在本地磁盘文件中，也可以存放在注册表中，或者存放在远程，例如网络上）。

对于将细胞词库的数据添加至细胞词集的过程，可以在下载完成之后立即操作（例如，通知输入法客户端开始添加操作）；也可以等待输入法主动发现更新（例如用户下次启动输入法）的时候，再开始添加操作：扫描细胞词库列表，依次读入并将每个细胞词库添加到细胞词集中。

以细胞词集的表现形式为一个独立存在的大词库为例进行说明，具体的添加过程可以有两种方式：增量、批量。

批量方式是一次性将所有细胞词库中的词合并成一个大的临时词库，然后一次性加入细胞词集。这种方式实现起来会比较简单，但用户必须等待所有词库都合并完成后才能使用新加入的细胞词库。增量方式为：当读入若干个词条就将其加入细胞词集，如果合并时间很长的话，用户可以边合并边使用，但这对系统设计的要求较高。

对于增量合并方式，在合并过程中就可以使用，因此当合并完成后不需要通知输入法系统。但对于批量合并方式，需要在合并完成后通知输入法系统新的词库已经可以使用了。一种替代的做法是，直接访问输入法的存储空间并对数据进行更新，这样虽然输入法没有得到通知，但数据已经被更新，因此实际已经可以使用新的数据了。

优选的，在数据添加的过程中，还可以包括优化步骤，用于对词库中重复的词进行优化，例如，将重复的词条合并。当然，为了准确记录该词，可以在

其来源属性中记录其所述的多个细胞词库的标识等信息。进一步，还可以记录该词所述的多个细胞词库的不同的优先级，用于对于不同的输入环境，采用不同的细胞词库的优先级进行候选项排序。

为了帮助输入法客户端更好的完成更新任务，则本实施例中的词库发布系统可以将更新的识别工作设置在服务器端完成。即优选的，本实施例中的词库发布系统还可以包括：识别模块 303，用于比较服务器端保存的细胞词库列表和客户端发送的细胞词库列表，所得到的比较结果用于传输所需的更新数据至客户端。例如，可以将发生变化的细胞词库形成列表发送给客户端，由客户端确定和发起下载请求；或者，也可以直接由服务器将发生变化的细胞词库推送给客户端，完成更新。所述的更新数据可以为整个细胞词库，例如，识别得知该细胞词库需要更新，则传输该细胞词库的所有词条信息；所述的更新数据也可以为一细胞词库中的部分词条信息，例如，识别得知该细胞词库需要更新，则进一步通过词条比对，仅仅传输发生变化的词条信息即可。

进一步提高词库发布的效率，本实施例还可以包括：合并模块 304，用于将多个细胞词库词条信息合并为一个下载词库，并通知通信单元 302 将该下载词库传输至客户端。所述合并单元可以用于各种可能的场景，例如，将用户所选的多个细胞词库合并为一个词库后进行传输；或者，将多个需要更新的细胞词库中的发生变化的词条信息进行合并，得到一个新词库，然后进行传输；或者，将细胞词库中索引信息相应的细胞词库进行合并，得到一个新词库，然后进行传输。

参照图 4，示出了一种词库更新的方法实施例，所需更新的词库涉及到在输入法系统中记录扩展字词及其相关信息的细胞词集，所述细胞词集由从服务器端所存储的多个细胞词库中选取的至少一个细胞词库得到；每个细胞词库中的字词至少具有一个共同属性；

所述方法实施例具体可以包括：

步骤 401、接受触发，比较已有细胞词库列表和服务器端细胞词库列表，得到所需更新的词库列表；所述触发可以手动触发，也可以自动触发；

步骤 402、下载所需更新的细胞词库词条信息，并添加至细胞词集中。

优选的，所述方法实施例还可以包括步骤 403：手动或者自动升级服务

器端所存储的细胞词库，并更改相应的版本信息。所述升级可以为人工完成，也可以为依据一定的预置策略对细胞词库进行调整而完成，例如：其他用户向某个细胞词库中添加新的词条；或者，依据预置策略，将两个细胞词库中的词条合并为一个细胞词库；或者，依据互联网词频统计结果，将某个细胞词库中互联网词频不符合预置条件的词条进行删除或者进行排序调整。

为了便于更新，每个细胞词库都具有一个唯一的 ID，这个唯一 ID 可以是一个自然增长的整数，也可以是一个网络地址或者其他信息（只要保证两个不同的细胞词库具有不同的 ID 就可以）。每个细胞词库还可以具有一个版本信息，这个版本信息可以是一个流水号，也可以是最后一次修改的时间。该版本信息发生了改变，则表明该词库文件需要更新。例如，采用客户端最后一次更新时间作为版本信息，如果与服务器上保存的文件更新时间相比前者有变化，那么该词库文件需要更新。

对于步骤 401 中的比较过程的实现可以采用多种实现方式，例如：

（1） 输入法客户端将现有细胞词库列表发送给服务器，可以通过 TCP/IP 协议发送，或者通过 HTTP 协议发送；由服务器判断与列表中的 ID 相应的细胞词库是否需要更新。

（2） 输入法客户端发起更新请求，服务器将所有的细胞词库的列表信息发回，由输入法客户端判断哪些已有词库需要更新。

（3） 输入法客户端将现有细胞词库列表发送给服务器，服务器将列表中的 ID 相应的细胞词库的版本信息发回，由输入法客户端判断哪些已有词库需要更新。

上述几种方式对于带宽和设备计算压力各有所不同，本领域技术人员根据实际需要选用即可。

对于由服务器完成识别过程的情况而言，服务器可以将发生变化的细胞词库形成列表发送给客户端，由客户端确定和发起下载请求（例如，从中选择部分词库进行更新）；或者，也可以直接由服务器将发生变化的细胞词库推送给客户端，完成更新。

对于步骤 402 中所下载的数据，可以为整个词库，也可以为一细胞词库中的部分词条信息，例如，发生变化的词条信息。

对于步骤402中的数据添加过程，可以采用增量模式、批量模式或者二者的结合。例如，所述添加方式为：完成更新下载一个细胞词库，则添加该细胞词库词条信息至所述细胞词集中；或者，所述添加方式为：完成所有待更新细胞词库的下载后，才添加至所述细胞词集中。

对于增量模式，可以更新一个词库就安装一个词库，其优点是已下载的词库不受未下载词库的影响，可以立即生效。但当下载词库较多时可能导致频繁的词库添加操作，加重系统负担。而批量模式则要求所有词库都下载到本地后才进行添加。由于添加操作较少，系统负荷较低。但当下载过程较长，特别是中间还可能发生下载失败的情况时，就会出现已下载的词库长期无法使用的问题。实际使用中可以将两种模式进行结合，比如每下载成功一个词库就检查距上次添加操作是否已经过了一个预定义的时间间隔。如果超过，就执行词库添加操作。

如果词库添加过程能够在较短时间内完成（比如不超过1秒），由于影响不大，可以直接插入用户的输入过程中。但如果在较短时间内无法完成以致可能影响用户的使用感受，则词库添加过程应当在一个独立的缓存词库中进行。这个过程中输入法原来的词库不受影响，用户可以正常使用。当缓存词库创建完毕后，直接替换输入法原来的词库。由于这个替换过程可以很快，因此可以做到避免对用户的正常使用构成干扰。

以上对本发明所提供的一种输入法系统、一种字符输入的方法以及一种词库更新的方法和一种词库发布系统，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。

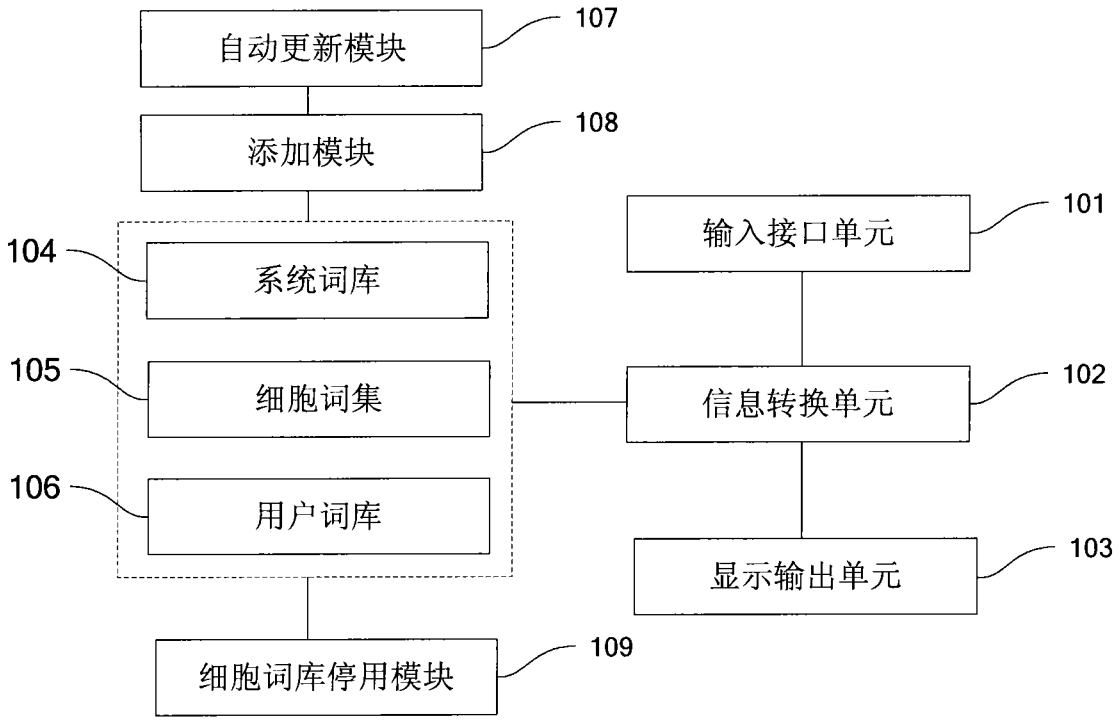

自动更新模块
107
添加模块
108
104
系统词库
105
细胞词集
106
用户词库
输入接口单元
101
信息转换单元
102
显示输出单元
103
细胞词库停用模块
109

图 1

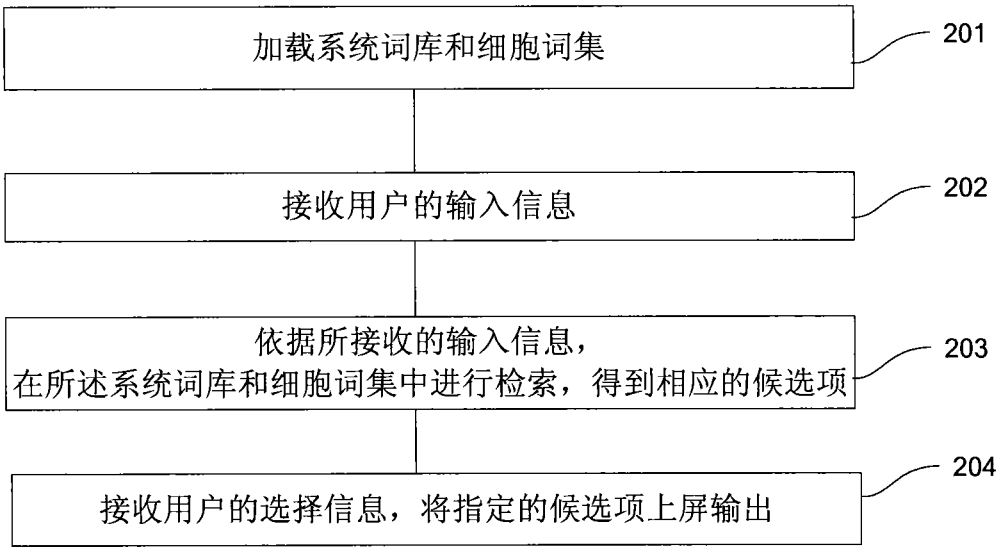

加载系统词库和细胞词集
201
接收用户的输入信息
202
依据所接收的输入信息，
在所述系统词库和细胞词集中进行检索，得到相应的候选项
203
接收用户的选择信息，将指定的候选项上屏输出
204

图 2

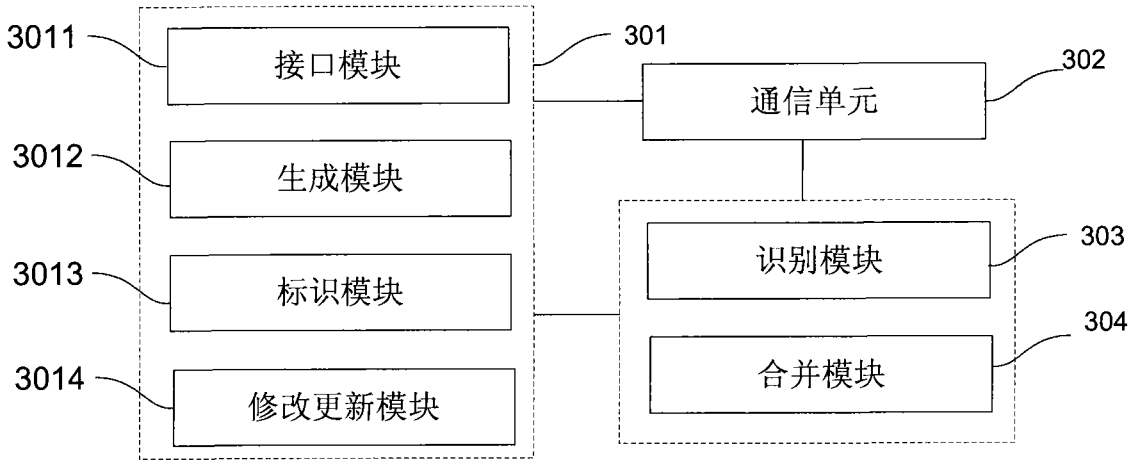

3011
接口模块
3012
生成模块
3013
标识模块
3014
修改更新模块
301
通信单元
302
识别模块
303
合并模块
304

图 3

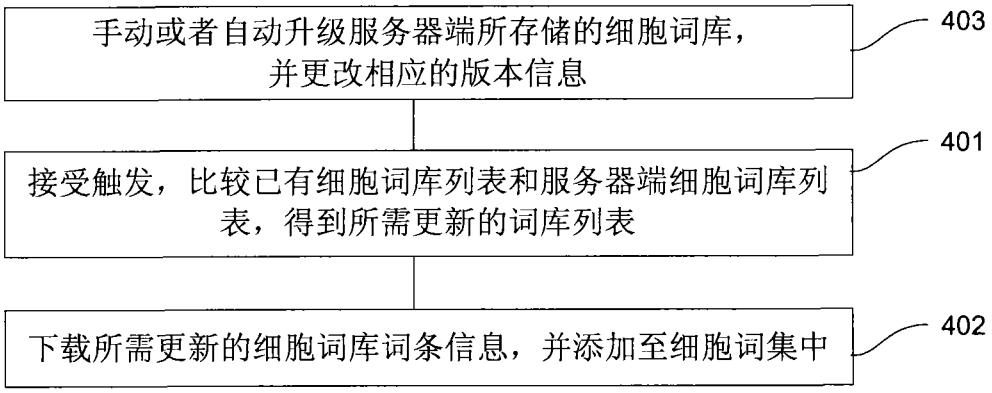

手动或者自动升级服务器端所存储的细胞词库，并更改相应的版本信息
403
接受触发，比较已有细胞词库列表和服务器端细胞词库列表，得到所需更新的词库列表
401
下载所需更新的细胞词库词条信息，并添加至细胞词集中
402

图 4