



(12) 发明专利申请

(10) 申请公布号 CN 112528666 A

(43) 申请公布日 2021.03.19

(21) 申请号 201910818035.9

(22) 申请日 2019.08.30

(71) 申请人 北京猎户星空科技有限公司
地址 100025 北京市朝阳区姚家园南路一
号惠通时代广场8号

(72) 发明人 李鹏 韩伟

(74) 专利代理机构 北京同达信恒知识产权代理
有限公司 11291
代理人 郭晓丽

(51) Int. Cl.

G06F 40/30 (2020.01)

G06F 40/247 (2020.01)

G06F 16/332 (2019.01)

G06F 16/33 (2019.01)

权利要求书2页 说明书14页 附图3页

(54) 发明名称

一种语义识别方法、装置及电子设备

(57) 摘要

本申请公开了一种语义识别方法、装置及电子设备,涉及语义识别技术领域,能够基于基础词库利用相似词挖掘自动抽象出一些通用词语,从而更准确的理解用户意图。本申请方法包括:对于获取到的文本信息中的待识别词语,在基础词库中查找与所述待识别词语匹配的目标基础词;若未查到所述目标基础词,在相似语义词库中查找与所述待识别词语匹配的目标相似词,其中,所述相似语义词库中的相似词为根据所述基础词库中的基础词按照设定规则进行处理得到的;若查找到所述目标相似词,根据基础词与相似词之间的映射关系,确定所述目标相似词对应的目标基础词;根据所述目标基础词确定所述文本信息的语义解析结果。



1. 一种语义识别方法,其特征在于,该方法包括:

对于获取到的文本信息中的待识别词语,在基础词库中查找与所述待识别词语匹配的目标基础词;

若未查到所述目标基础词,在相似语义词库中查找与所述待识别词语匹配的目标相似词,其中,所述相似语义词库中的相似词为根据所述基础词库中的基础词按照设定规则进行处理得到的;

若查找到所述目标相似词,根据基础词与相似词之间的映射关系,确定所述目标相似词对应的目标基础词;

根据所述目标基础词确定所述文本信息的语义解析结果。

2. 如权利要求1所述的方法,其特征在于,根据基础词与相似词之间的映射关系,确定所述目标相似词对应的目标基础词,包括:

若根据基础词与相似词之间的映射关系,确定出所述目标相似词对应的基础词为多个,控制智能设备输出提示信息,所述提示信息用于提示对多个基础词进行选择;

获取反馈的确认信息,将所述确认信息所选定的基础词确定为所述目标基础词。

3. 如权利要求1所述的方法,其特征在于,根据所述基础词库的基础词按照设定规则进行处理得到相似词,包括如下至少一种方式:

对于所述基础词库中的每个基础词,获取所述基础词的指定位置上设定字数的词片段,对于具有相同词片段的基础词,提取所述相同词片段作为相似语义词库中对应的相似词;

对所述基础词库中包括的每个基础词进行分词处理,得到所述基础词对应的分词,将所得到的分词确定为所述基础词对应的相似词;

对于所述基础词库中的基础词,从预设同义词库中选择与该基础词具有相同语义的同义词,根据所选择的同义词,确定所述基础词对应的相似词。

4. 如权利要求1~3任一所述的方法,其特征在于,所述基础词库配置有对应的生效范围,所述方法还包括:

将基于任一基础词库得到的相似语义词库的生效范围配置为所述基础词库的生效范围;或者

若所述相似语义词库是基于多个基础词库得到的,将所述相似语义词库的生效范围配置为所述多个基础词库的生效范围中最大的生效范围。

5. 如权利要求4所述的方法,其特征在于,在相似语义词库中查找与该词语匹配的目标相似词,包括:

根据智能设备的属性信息以及各相似语义词库的生效范围,确定对所述智能设备有效的目标相似语义词库;

在所述目标相似语义词库中,查找与该词语匹配的目标相似词。

6. 如权利要求3所述的方法,其特征在于,所述基础词库包括专属基础词库和通用基础词库,根据所述基础词库的基础词按照设定规则进行处理得到相似词,还包括:

根据所述专属基础词库中的专属词,按照设定规则进行处理得到专属相似词,基于得到的专属相似词形成专属相似语义词库;以及

根据所述通用基础词库中的标准词,按照设定规则进行处理得到标准相似词,基于得

到的标准相似词形成通用相似语义词库。

7. 如权利要求4所述的方法,其特征在于,所述基础词库的生效范围,包括如下至少一种:对于指定智能设备生效、对于指定企业生效、对于指定业务线生效、对于指定应用程序生效、对于指定领域生效、对于所有智能设备生效;

所述相似语义词库的生效范围,包括如下至少一种:对于指定智能设备生效、对于指定企业生效、对于指定业务线生效、对于指定应用程序生效、对于指定领域生效、对于所有智能设备生效。

8. 一种语义识别装置,其特征在于,该装置包括:

基础词查找单元,对于获取到的文本信息中的待识别词语,在基础词库中查找与所述待识别词语匹配的目标基础词;

相似词查找单元,用于若未查到所述目标基础词,在相似语义词库中查找与所述待识别词语匹配的目标相似词,其中,所述相似语义词库中的相似词为根据所述基础词库中的基础词按照设定规则进行处理得到的;

确定单元,用于若查找到所述目标相似词,根据基础词与相似词之间的映射关系,确定所述目标相似词对应的目标基础词;

解析单元,用于根据所述目标基础词确定所述文本信息的语义解析结果。

9. 一种电子设备,其特征在于,该电子设备包括处理器和存储器,其中,所述存储器存储可执行程序,所述程序执行时,所述处理器实现如权利要求1~7任一所述的方法。

10. 一种计算机存储介质,其上存储有计算机程序,其特征在于,该程序被处理器执行时实现如权利要求1~7任一所述方法的步骤。

一种语义识别方法、装置及电子设备

技术领域

[0001] 本申请涉及智能机器人技术领域,尤其涉及一种语义识别方法、装置及电子设备。

背景技术

[0002] 对于智能接待机器人而言,引领用户到达指定地点是一个非常重要而实用的功能,比如用户说“带我去会议室”,机器人便可以理解用户意图并带领用户到达指定地点。

[0003] 目前的智能服务机器人可以针对某个企业定义这个企业的所有地点,它们大多数存储有企业地点词库,当收到用户的询问后,机器人可以明白用户的意图并带领用户到达上述企业地点词库中的一些指定地点,例如在企业地点词库中定义了卫生间、电梯、会议室,机器人收到用户关于卫生间、电梯、会议室的询问后,便可以明白用户意图并带领用户到达卫生间、电梯、会议室。

[0004] 然而企业的地点词库无法覆盖用户的所有问法,机器人必须准确的收到询问中关于企业地点词库中完整的词语,才能理解用户的准确地点,例如一个企业的地点词库有“会议室A”、“会议室B”、“会议室C”和“会议室D”,用户说“带我去会议室”,机器人便无法理解用户的准确地点。

发明内容

[0005] 本申请提供一种语义识别方法、装置及电子设备,能够基于基础词库利用相似词挖掘自动抽象出一些通用词语,使得智能设备尽可能的识别用户的所有问法,从而更准确的理解用户意图。

[0006] 第一方面,本申请提供一种语义识别方法,该方法包括:

[0007] 对于获取到的文本信息中的待识别词语,在基础词库中查找与所述待识别词语匹配的目标基础词;

[0008] 若未查到所述目标基础词,在相似语义词库中查找与所述待识别词语匹配的目标相似词,其中,所述相似语义词库中的相似词为根据所述基础词库中的基础词按照设定规则进行处理得到的;

[0009] 若查找到所述目标相似词,根据基础词与相似词之间的映射关系,确定所述目标相似词对应的目标基础词;

[0010] 根据所述目标基础词确定所述文本信息的语义解析结果。

[0011] 第二方面,本申请提供一种语义识别装置,该装置包括:

[0012] 基础词查找单元,对于获取到的文本信息中的待识别词语,在基础词库中查找与所述待识别词语匹配的目标基础词;

[0013] 相似词查找单元,用于若未查到所述目标基础词,在相似语义词库中查找与所述待识别词语匹配的目标相似词,其中,所述相似语义词库中的相似词为根据所述基础词库中的基础词按照设定规则进行处理得到的;

[0014] 确定单元,用于若查找到所述目标相似词,根据基础词与相似词之间的映射关系,

确定所述目标相似词对应的目标基础词；

[0015] 解析单元,用于根据所述目标基础词确定所述文本信息的语义解析结果。

[0016] 第三方面,本申请提供一种电子设备,该设备包括处理器和存储器,其中,所述存储器存储可执行程序,所述程序执行时,所述处理器实现如下过程:

[0017] 对于获取到的文本信息中的待识别词语,在基础词库中查找与所述待识别词语匹配的目标基础词;

[0018] 若未查到所述目标基础词,在相似语义词库中查找与所述待识别词语匹配的目标相似词,其中,所述相似语义词库中的相似词为根据所述基础词库中的基础词按照设定规则进行处理得到的;

[0019] 若查找到所述目标相似词,根据基础词与相似词之间的映射关系,确定所述目标相似词对应的目标基础词;

[0020] 根据所述目标基础词确定所述文本信息的语义解析结果。

[0021] 第四方面,本申请提供一种计算机存储介质,其上存储有计算机程序,该程序被处理器执行时实现上述第一方面所述方法的步骤。

[0022] 本申请提供一种语义识别设备,具有以下有益效果:

[0023] 针对智能设备中由于配置的基础词库的词汇量有限,无法解析用户的所有语义,对所述基础词库进行扩词后,通过扩词后的相似语义词库查找待识别词语,将查找到的所有目标相似词根据基础词与相似词之间的映射关系,确定所述目标相似词对应的目标基础词,最终根据所述目标基础词确定所述文本信息的语义解析结果,从而更好的理解用户意图。

附图说明

[0024] 为了更清楚地说明本申请实施例中的技术方案,下面将对实施例描述中所需要使用的附图作简要介绍,显而易见地,下面描述中的附图仅仅是本申请的一些实施例,对于本领域的普通技术人员来讲,在不付出创造性劳动性的前提下,还可以根据这些附图获得其他的附图。

[0025] 图1为本申请实施例提供的一种语义识别方法流程图;

[0026] 图2为本申请实施例提供的一种语义识别方法具体流程图;

[0027] 图3为本申请实施例提供的一种语义识别装置示意图;

[0028] 图4为本申请实施例提供的一种电子设备示意图。

具体实施方式

[0029] 为了使本申请的目的、技术方案和优点更加清楚,下面将结合附图对本申请作进一步地详细描述,显然,所描述的实施例仅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它实施例,都属于本申请保护的范围。

[0030] 下面对文中出现的一些词语进行解释:

[0031] 1、本申请实施例中术语“和/或”,描述关联对象的关联关系,表示可以存在三种关系,例如,A和/或B,可以表示:单独存在A,同时存在A和B,单独存在B这三种情况。字符“/”一

般表示前后关联对象是一种“或”的关系。

[0032] 2、本申请实施例中的智能设备可以是具有一定功能的智能机器人,如具有引领功能、接待功能、导览功能、讲解功能、聊天功能、业务查询功能、播放音乐等功能的智能设备等,智能设备也可以为便携设备(例如:手机、平板、笔记本电脑等)或者个人电脑(PC, Personal Computer),还可以是智能音箱等,本发明实施例中不对智能设备的具体形式进行限定。

[0033] 本申请实施例描述的应用场景是为了更加清楚的说明本申请实施例的技术方案,并不构成对于本申请实施例提供的技术方案的限定,本领域普通技术人员可知,随着新应用场景的出现,本申请实施例提供的技术方案对于类似的技术问题,同样适用。其中,在本申请的描述中,除非另有说明,“多个”的含义是两个或两个以上。

[0034] 目前应用基础词库进行语义识别时,基础词库可存储于智能设备侧,或者存储于服务器侧,若存储于服务器侧,智能设备可以通过网络连接访问服务器,由服务器基于基础词库进行相关查询并反馈结果。

[0035] 上述基础词库可以为一个或多个,可以进一步分为专属基础词库和通用基础词库,其中专属基础词库中存储专属词,通用基础词库中存储标准词,专属词为针对特定应用场景定义的词语,如针对不同企业定义专属于该企业的专属词,具体示例为针对“咖啡厅”,针对企业A定义了专属词“意恋咖啡厅”、企业B定义了专属词“福莱斯特咖啡厅”,等等,可根据文本信息“带我去意恋咖啡厅”,在专属基础词库中查询,根据查询结果进行引领;上述标准词为适用所有应用场景的词语,具体示例为适用所有企业的词语如“卫生间”,智能设备接收到文本信息“请带我去卫生间”时,通过查询通用基础词库,根据查询结果进行引领。

[0036] 但是不管是专属词还是标准词常常都不止一种表达形式,如标准词“卫生间”,也可被表达为“洗手间”、“厕所”、“WC”、“Toilet”、“茅厕”等词语,专属词“意恋咖啡厅”,也可被表达为“意恋”,按照目前语义识别方式,仅基于基础词库对“厕所”或“意恋”进行语义识别时,基础词库中并不能查到这些词语,进而不能正确理解用户意图,导致无法进行正确引领。

[0037] 为了解决上述技术问题,本申请提出一种语义识别方法,基于基础词库和相似语义词库进行语言识别,其中,相似语义词库中存储有相似词,所存储的相似词具体为根据基础词库中的基础词按照设定规则进行处理得到的,在基于基础词库按照设定规则进行处理得到对应的相似词,进一步建立基础词和该基础词处理得到的相似词之间的映射关系。

[0038] 如图1所示,为本申请提供的语义识别方法,应用于智能设备,或者应用于服务器,主要包括以下步骤:

[0039] 步骤100,对于获取到的文本信息中的待识别词语,在基础词库中查找与所述待识别词语匹配的目标基础词。

[0040] 作为一种可选的实施方式,从文本信息中确定待识别词语的具体过程为:对文本信息进行领域识别,确定上述文本信息对应的领域;对上述文本信息进行意图识别,确定上述文本信息对应的意图类别;根据确定的领域和意图类别按照对应的方式,对上述文本信息进行槽位抽取,确定抽取出的槽位对应的槽位值为上述待识别词语。

[0041] 其中,领域(domain),是指同一类型的数据或者资源,以及围绕这些数据或资源提供的服务,比如导航、百科、天气、音乐、火车票、闲聊等。意图(intent),是通过用户的交互

输入识别出的用户所要表达的目的。通常,意图是指对于领域数据的操作,一般以动宾短语来命名,比如询问天气、查找音乐、购买火车票等。槽位(slot),是将初步用户意图转化为明确用户指令所需要补充的信息。一个槽位与一件事情的处理中所需要获取的一种信息相对应。例如,要想执行查看天气的任务,需要知道地点、时间这些必要的元素,这些必要元素就是槽位。

[0042] 对上述文本信息进行槽位抽取时,得到各槽位项对应的槽位值,可以但不局限于将所得到的各个槽位值确定为上述待识别词语,其中可以但不局限于通过如下方式对文本信息进行槽位解析:

[0043] 槽位解析方式1:通过文本规则匹配方式,对上述文本信息进行槽位解析;

[0044] 槽位解析方式2:通过自然语言理解NLU模型,对上述文本信息进行槽位解析。

[0045] 在以槽位解析的方式获取上述步骤中的待识别词语之后,在基础词库中查找与所述待识别词语匹配的目标基础词。

[0046] 该方法应用于智能设备时,智能设备直接获取文本信息,或者接收语音信息,并将语音信息通过语音识别技术转换为文本信息,其中基础词库存储于智能设备;

[0047] 该方法应用于服务器时,获取智能设备直接上传的文本信息,或者获取智能设备上传的语音信息,通过语音识别技术转换为文本信息,其中基础词库存储在服务器。

[0048] 其中,上述智能设备上传的文本信息,可以是智能设备直接获取的文本信息,也可以是智能设备对获取的语音信息进行处理得到的文本信息。

[0049] 智能设备可以通过语音采集装置如麦克风采集语音信息,通过输入显示模块如触摸屏接收文本信息。

[0050] 步骤101,若未查到目标基础词,在相似语义词库中查找与待识别词语匹配的目标相似词;

[0051] 该方法应用于智能设备时,相似语义词库存储于智能终端,该方法应用于服务器时,相似语义词库存储在服务器。

[0052] 在相似语义词库中查找与待识别词语匹配的目标相似词,可以通过比较待识别词语与相似语义词库中的各词语的相似度来判断是否匹配,将相似语义词库中与待识别词语的相似度高于设定阈值的词语,确定为与待识别词语匹配的目标相似词。

[0053] 步骤102、若查找到所述目标相似词,根据基础词与相似词之间的映射关系,确定所述目标相似词对应的目标基础词。

[0054] 步骤103、根据所述目标基础词确定所述文本信息的语义解析结果。

[0055] 本申请实施例在进行语义识别时,在基础词库中未查到目标基础词时,进一步在相似语义词库进行查询,并映射到对应的目标基础词,因此能够针对同一语义的多种相似询问进行查询,从而正确解析语义,提高了智能设备语义识别的准确度。

[0056] 作为一种可选的实施方式,若根据基础词与相似词之间的映射关系,确定出所述目标相似词对应的基础词为多个,控制智能设备输出提示信息,提示信息用于提示对多个基础词进行选择,可以通过智能设备显示模块显示的文本提示信息,也可以通过智能设备语音模块输出的语音提示信息。

[0057] 进一步的,获取反馈的确认信息,将确认信息所选定的基础词确定为所述目标基础词,相应地,可以通过智能设备显示模块接收的文本确认信息,也可以通过智能设备

的语音模块采集的语音确认信息。

[0058] 进一步的,本实施例根据语义解析结果,确定文本信息对应的响应信息,并控制智能设备输出响应信息,其中,响应信息可以是多媒体信息(如图片、视频、音频等),也可以是控制指令(如控制智能设备移动、控制智能设备跳舞等等),也可以是语音播报等等。

[0059] 实施例1

[0060] 结合上述实施例提供的语义识别方法,本实施例给出基础词库和相似语义词库各种可能的实施方式,及具体的待识别词语查询方式。

[0061] 本实施例中基础词库包括存储有专属词的专属基础词库和存储有标准词的通用基础词库,具体如前描述,这里不再重述。

[0062] 专属基础词库的数量为至少一个,通用基础词库的数量为至少一个,可以预先为各专属基础词库或通用基础词库分别配置对应的生效范围,为方便描述,下面将各专属基础词库或通用基础词库称为各基础词库,所配置的生效范围可以但不局限于包括:

[0063] 1) 设备级生效

[0064] 设备级生效是指对指定设备生效,可以根据场景需要或实际需要,配置对各基础词库生效的指定设备,如配置基础词库1的生效范围为智能设备A和智能设备B,基础词库2的生效范围为智能设备C,则对于智能设备B而言,对其生效的基础词库为基础词库1。

[0065] 设备级生效还可以指对指定设备类型的设备生效,可以根据场景需要或实际需要,配置上述任一基础词库生效范围为对于指定设备类型的设备生效,如智能设备A和智能设备B的设备类型都为设备类型a,智能设备C的设备类型都为设备类型b,当配置基础词库1的生效范围为设备类型a时,则基础词库1对于智能设备A和智能设备B均生效;

[0066] 其中,上述指定设备类型可以但不局限于为指定功能的设备等,如具有引领功能的智能设备等。

[0067] 2) 企业级生效

[0068] 企业级生效是指对指定企业生效,可以根据场景需要或实际需要,配置上述任一基础词库生效范围为对指定企业的设备生效,如配置基础词库1的生效范围为企业A,基础词库2的生效范围为企业B,则对于企业A下的智能设备而言,对其生效的基础词库为基础词库1,而对于企业B下的智能设备而言,对其生效的基础词库为基础词库2。

[0069] 3) 业务线级生效

[0070] 业务线级生效是指对指定业务线生效,可以根据场景需要或实际需要,配置上述任一基础词库生效范围为对指定业务线的设备生效,如配置基础词库1的生效范围为业务线A,基础词库2的生效范围为业务线B,则对业务线A下的智能设备而言,对其生效的基础词库为基础词库1。

[0071] 4) 应用程序级生效

[0072] 应用程序级生效是指对指定应用程序生效,可以根据场景需要或实际需要,配置上述任一基础词库生效范围为对指定应用程序的设备生效,如配置基础词库1的生效范围为应用程序A,基础词库2的生效范围为应用程序B,则对于执行应用程序A的智能设备而言,对其生效的基础词库为基础词库1。

[0073] 5) 领域级生效

[0074] 领域级生效是指对指定领域生效,可以根据场景需要或实际需要,配置上述任一

基础词库生效范围为对指定领域的设备生效,如配置基础词库1的生效范围为领域A,基础词库2的生效范围为领域B,则对于领域A下的智能设备而言,对其生效的基础词库为基础词库1。

[0075] 6) 对所有设备生效

[0076] 可以配置上述任一基础词库生效范围为对所有设备生效,如配置基础词库1的生效范围为对所有设备生效,则基础词库1对于任一智能设备均生效。

[0077] 本实施例中根据基础词库得到相似语义词库,一种可能的方式为根据一个基础词库对应得到一个相似语义词库,另一种可能的方式为根据多个基础词库对应得到一个相似语义词库。

[0078] 基于上述方式配置的各基础词库的生效范围,本发明实施例中各相似语义词库预先配置有对应的生效范围,具体可以但不限于采用如下任一配置方式:

[0079] 配置方式1,将基于任一基础词库得到的相似语义词库的生效范围配置为所述基础词库的生效范围;

[0080] 即对于基于一个基础词库得到的一个相似语义词库,将该相似语义词库的生效范围配置为该基础词库的生效范围。

[0081] 配置方式2,若相似语义词库是基于多个基础词库得到的,将该相似语义词库的生效范围配置为该多个基础词库的生效范围中最大的生效范围。

[0082] 对应地,本发明实施例中,相似语义词库的生效范围包括但不限于如下任一或任多种:

[0083] 1) 对于指定设备生效,可以根据场景需要或实际需要,配置上述相似语义词库生效范围为对于指定设备(即上述智能设备)或者指定设备类型的设备生效;

[0084] 2) 对于指定企业生效,可以根据场景需要或实际需要,配置上述相似语义词库生效范围为对于指定企业的设备生效;

[0085] 3) 对于指定业务线生效,可以根据场景需要或实际需要,配置上述相似语义词库生效范围为对于指定业务线的设备生效;

[0086] 4) 对于指定应用程序生效,可以根据场景需要或实际需要,配置上述相似语义词库生效范围为对于指定应用程序的设备生效;

[0087] 5) 对于指定领域生效,可以根据场景需要或实际需要,配置上述相似语义词库生效范围为对于指定领域的设备生效;

[0088] 6) 对于所有设备生效,可以根据场景需要或实际需要,配置上述相似语义词库生效范围为对所有设备生效。

[0089] 本实施例中还可以根据各基础词库的生效范围设置基础词库的优先级,可以但不限于将基础词库的优先级从高到低按如下顺序配置:

[0090] 对于指定设备生效的基础词库、对于指定企业生效的基础词库、对于指定应用程序生效的基础词库、对于指定业务线生效的基础词库、对于指定领域生效的基础词库、对于所有设备生效的基础词库。

[0091] 应当理解的是,所述基础词库的生效范围越小,其优先级越高。

[0092] 本实施例中还可以根据各相似语义词库的生效范围设置相似语义词库的优先级,可以但不限于将相似语义词库的优先级从高到低按如下顺序配置:

[0093] 对于指定设备生效的相似语义词库、对于指定企业生效的相似语义词库,对于指定应用程序生效的相似语义词库、对于指定业务线生效的相似语义词库,对于指定领域生效的相似语义词库、对于所有设备生效的相似语义词库。

[0094] 应当理解的是,所述相似语义词库的生效范围越小,其优先级越高。

[0095] 本实施例在基础词库中进行查询时,可以根据智能设备的属性信息及各基础词库的生效范围,确定对智能设备有效的目标基础词库,并在有效的目标基础词库中查询目标基础词。

[0096] 进一步地,对智能设备有效的目标基础词库为多个时,可以根据该多个基础词库的优先级从高到低的顺序,依次从多个基础词库中查找与所述待识别词语匹配的目标基础词。

[0097] 具体的,先从优先级最高的基础词库中查找是否存在与待识别词语匹配的词语,若存在,则将查找到的词语确定为待识别词语匹配的目标基础词;若不存在,则继续从优先级次高的基础词库中查找是否存在与待识别词语匹配的词语,以此类推,直至在任一基础词库中查找到与待识别词语匹配的词语或者任一基础词库中均未查找到与待识别词语匹配的词语。

[0098] 本实施例在相似语义词库中进行查询时,可以根据智能设备的属性信息以及各相似语义词库的生效范围,确定对所述智能设备有效的目标相似语义词库;在所述目标相似语义词库中,查找与该词语匹配的目标相似词。

[0099] 进一步地,对智能设备有效的目标相似语义词库为多个时,可以根据该多个目标相似语义词库的优先级从高到低的顺序,依次从多个目标相似语义词库中查找与所述待识别词语匹配的目标相似词。

[0100] 智能设备的属性信息用以标识上述智能设备的身份信息以及上述智能设备的使用场景等,上述属性信息可以但不局限于包括:设备身份标识ID、设备类型,所属的企业等中的至少一种信息。

[0101] 该方法应用于服务器时,智能设备向服务器上传文本信息或者语音信息时,可以同时向服务器上传该智能设备的属性信息。

[0102] 需要说明的是,上述各个基础词库及相似语义词库可以保存在其生效的智能设备中,也可以保存在服务器中。

[0103] 实施例2

[0104] 结合上述实施例提供的语义识别方法,本实施例给出基于基础词库中的基础词按照设定规则,得到相似语义词库中相似词的可能实施方式。

[0105] 本实施例中相似语义词库的相似词可以通过对基础词库的基础词,按照如下任一方式进行处理后得到,或者将如下任多种方式组合后得到:

[0106] 处理方式1、对于所述基础词库中的每个基础词,获取所述基础词的指定位置上设定字数的词片段,对于具有相同词片段的基础词,提取所述相同词片段作为相似语义词库中对应的相似词。

[0107] 本实施例中的指定位置包括但不限于基础词的前缀、中段、后缀;所述设定字数包括但不限于1个字、2个字、3个字及以上等;其中指定位置和设定字数可以进行多种组合,可以是获取所述基础词在前缀位置上2个字的词片段,也可以是获取所述基础词在中段位置

上3个字的词片段,还可以是获取所述基础词在后缀位置上2个字的词片段。

[0108] 例如获取所述基础词在前缀位置上2个字的词片段,则对基础词“电梯A”、“电梯B”、“电梯C”,能够获取的词片段为“电梯”,则将“电梯”作为基础词的相似词。

[0109] 处理方式2、对所述基础词库中包括的每个基础词进行分词处理,得到所述基础词对应的分词,将所得到的分词确定为所述基础词对应的相似词。

[0110] 一种可能的实施方式是,将基础词库中的基础词进行分词处理后,进行去重处理,得到各个分词。进一步将所得到的各个分词与基础词库中的基础词进行去重处理,具体的:对于任一分词,若基础词库中已存在该分词,则不将该分词作为基础词对应的相似词,若基础词库中不存在该分词,则将该分词作为基础词对应的相似词。

[0111] 例如,基础词库中包括“北京会议室”,其中“北京会议室”进行分词处理后,划分为2个分词“北京”和“会议室”,从基础词库中查找与“北京”和“会议室”一致的基础词,若都未查找到,则将“北京”和“会议室”作为相似词,若从基础词库中查找到“北京”,未查找到“会议室”,则将“会议室”作为相似词,若从基础词库中查找到“会议室”,未查找到“北京”,则将“北京”作为相似词。

[0112] 处理方式3、对于基础词库中的基础词,从预设同义词库中选择与该基础词具有相同语义的同义词,根据所选择的同义词,确定所述基础词对应的相似词。

[0113] 例如,基础词库中包括“咖啡馆”,根据具有同一语义的多个同义词之间的映射关系,在预设同义词库中预先建立了“咖啡馆”与对应“意恋咖啡”、“福莱斯特咖啡厅”、“欧米奇咖啡厅”的同义词之间的映射关系,找到与该“咖啡馆”具有相同语义的同义词“意恋咖啡”、“福莱斯特咖啡厅”、“欧米奇咖啡厅”,则将所述“意恋咖啡”、“福莱斯特咖啡厅”、“欧米奇咖啡厅”作为“咖啡馆”的相似词语。

[0114] 其中,本实施例可以将上述三种处理方式进行任意组合:

[0115] 例如,本实施例还可以将上述处理方式1和处理方式2结合,对所述对于所述基础词库中的部分基础词采用处理方式1进行处理,部分基础词采用处理方式2进行处理,例如:若采用处理方式1对所述基础词进行处理,则对于在指定位置上具有相同的设定字数的词片段的基础词,提取相同的词片段,若采用处理方式2对所述基础词进行处理,则对所述基础词进行分词处理;也可以对所述基础词库中的所有基础词均分别采用处理方式1和处理方式2进行处理,得到相似词,然后再对得到的所有相似词进行去重处理,去除掉重复的相似词。

[0116] 作为一种可能的实施方式,本实施例还可以将上述处理方式1和处理方式3结合,一种可能的情况是,对所述基础词库中的每个基础词采用处理方式1进行处理,并且对所述基础词库中的每个基础词采用处理方式3进行处理。

[0117] 作为另一种可能的实施方式,本实施例还可以将上述处理方式2和处理方式3结合,一种可能的情况是,对所述基础词库中的每个基础词采用处理方式2进行处理,并且对所述基础词库中的每个基础词采用处理方式3进行处理。

[0118] 实施例3

[0119] 结合上述实施例提供的语义识别方法中,基础词库包括存储有专属词的专属基础词库和存储有标准词的通用基础词库,及根据基础词库得到相似语义词库,一种可能的方式为根据一个基础词库对应得到一个相似语义词库,另一种可能的方式为根据多个基础词

库对应得到一个相似语义词库,本发明实施例给出基于不同类型的基础词库得到对应相似语义词库各种可能的实施方式。

[0120] 本实施例中根据专属基础词库中的专属词,按照设定规则进行处理得到专属相似词,基于得到的专属相似词形成专属相似语义词库,具体有两种方式:

[0121] 一种可能的方式为根据一个专属基础词库对应得到一个专属相似语义词库,另一种可能的方式为根据多个专属基础词库对应得到一个专属相似语义词库,专属相似语义词库的生效范围及查询方式参见实施例1的描述,基于专属词得到对应专属相似词参见实施例2的描述,这里不再重述。

[0122] 本实施例中根据所述通用基础词库中的标准词,按照设定规则进行处理得到标准相似词,基于得到的标准相似词形成通用相似语义词库,具体有两种方式:

[0123] 一种可能的方式为根据一个通用基础词库对应得到一个通用相似语义词库,另一种可能的方式为根据多个通用基础词库对应得到一个通用相似语义词库,通用相似语义词库的生效范围参见实施例1的描述,基于标准词得到对应标准相似词参见实施例2的描述,这里不再重述。

[0124] 这里需强调的是,由于相似语义词库的生效范围是根据基础词库的生效范围对应配置的,因此根据智能设备的属性信息确定对智能设备生效的相似语义词库,是由对智能设备生效的基础词库按预设规则得到,可以根据智能设备的属性信息及相似语义词库的生效范围,确定对智能生效的相似语义词库为目标相似语义词库。

[0125] 在目标相似语义词库中,查找与待识别词语匹配的目标相似词,具体可能存在如下几种情况:

[0126] 情况1)目标相似语义词库仅为一个。

[0127] 若目标相似语义词库为专属相似语义词库,则在该专属相似语义词库查找与待识别词语匹配的目标专属相似词;

[0128] 若目标相似语义词库为通用相似语义词库,则在该通用相似语义词库中查找与待识别词语匹配的目标标准相似词。

[0129] 情况2)目标相似语义词库既包括专属相似语义词库,又包括通用相似语义词库。

[0130] 查找与待识别词语匹配的相似词,可以但不局限于如下两种方式:

[0131] 方式1:分别在通用相似语义词库和专属相似语义词库中,查找与待识别词语匹配的目标相似词。

[0132] 在具体实施中,可以同时分别在专属相似语义词库和通用相似语义词库中进行相似词判断,以查找与所述待识别词语匹配的目标相似词,可以包括如下几种确定目标相似的方式:

[0133] 作为一种可选的实施方式,不区分专属相似语义词库和通用相似语义词库,将查找到的所有相似词作为待识别词语的目标相似词,即若仅在专属相似语义词库中查到与待识别词语匹配的目标专属相似词,将查到的所有目标专属词作为目标相似词,若仅在通用相似语义词库中查到与待识别词语匹配的目标标准相似词,将查到的所有目标标准相似词作为目标相似词,若既查到目标专属相似词也查到目标标准相似词,将查到的所有目标专属相似词和目标标准相似词作为目标相似词。

[0134] 作为另一种可选的实施方式,若仅在专属相似语义词库中查到与待识别词语匹配

的目标专属相似词,按照查到目标专属相似词所在的专属相似语义词库的优先级,选择在最高优先级的专属相似语义词库查到的目标专属相似词作为目标相似词;若仅在通用相似语义词库中查到与待识别词语匹配的目标标准相似词,按照查到目标标准相似词所在的通用相似语义词库的优先级,选择在最高优先级的通用相似语义词库查到的目标标准相似词作为目标相似词;若既在专属相似语义词库中查到目标专属相似词,也在通用相似语义词库中查到目标标准相似词,则根据专属相似语义词库中和通用相似语义词库的优先级顺序,由于专属相似语义词库的优先级高于通用相似语义词库,可优先将在专属相似语义词库中查找到的目标专属相似词作为目标相似词。若在多个专属相似语义词库中均查找到与待识别词语匹配的目标专属相似词,则选择在最高优先级的专属相似语义词库查到的目标专属相似词作为目标相似词。

[0135] 方式2:先在专属相似语义词库中,查找与待识别词语匹配的目标相似词,若未查到,再在通用相似语义词库中,查找与待识别词语匹配的目标相似词。

[0136] 所配置专属相似语义词库的优先级高于通用相似语义词库的优先级,在相似语义词库中查询时,一种可能的方式为,优先查询对智能设备生效的专属相似语义词库,若对智能设备生效的专属相似语义词库有多个,则同时查找,或者按照专属相似语义词库的优先级从高到低的顺序进行查询,若在专属相似语义词库未查到,进一步查询对智能设备生效的通用相似语义词库,若对智能设备生效的通用相似语义词库为多个,则同时查找,或者按照通用相似语义词库的优先级从高到低的顺序查询。

[0137] 进一步的,若仅在任一专属相似语义词库查找到与待识别词语匹配的目标专属相似词,则将查找到的目标专属相似词确定为目标相似词;若同时在多个专属相似语义词库均查找到目标专属相似词,则按照专属相似语义词库的优先级从高到低的顺序,选择在最高优先级的专属相似语义词库中查到的目标专属相似词作为目标相似词。

[0138] 若仅在任一通用相似语义词库查找到与待识别词语匹配的目标标准相似词,则将查找到的目标标准相似词确定为目标相似词;若同时在多个通用相似语义词库均查找到目标标准相似词,则按照通用相似语义词库的优先级从高到低的顺序,选择在最高优先级的通用相似语义词库中查到的目标标准相似词作为目标相似词。

[0139] 为了便于理解本实施例提供的语义识别方法,以下提供一个示例性的完整过程的说明,本实施例中以地点名称为例进行说明。如图2所示,具体包括:

[0140] 步骤201、通过智能设备获取文本信息;

[0141] 步骤202、通过语义理解系统中的领域分类模块对文本信息进行领域识别,确定上述文本信息对应的领域;

[0142] 步骤203、通过语义理解系统中的意图分类模块对文本信息进行意图识别,确定上述文本信息对应的意图类别;

[0143] 步骤204、通过语义理解系统中的槽位解析模块对上述文本信息进行槽位解析,得到各槽位项对应的槽位值,将所得到的槽位值确定为待识别词语;

[0144] 本示例中,仅以地点名称作为槽位对应的槽位值为例进行说明;

[0145] 步骤205、根据智能设备的属性信息以及各个基础词库的生效范围,确定对该智能设备有效的基础词库;

[0146] 其中,上述基础词库可以包括一个或多个通用基础词库和一个或多个专属基础词

库；

[0147] 步骤206、判断在有效的基础词库中是否查找到与所述待识别词语匹配的目标基础词，若未查找到执行步骤207，否则执行步骤209；

[0148] 步骤207、在相似语义词库中查找与所述待识别词语匹配的目标相似词；

[0149] 其中，所述相似语义词库中的相似词为根据所述基础词库中的基础词按照设定规则进行处理得到的；

[0150] 步骤208、根据基础词与相似词之间的映射关系，确定所述目标相似词对应的目标基础词；

[0151] 步骤209、根据所述目标基础词确定所述文本信息的语义解析结果。

[0152] 如图3所示，基于相同的发明构思本发明实施例还提供一种语义识别装置，该装置包括：

[0153] 基础词查找单元300，对于获取到的文本信息中的待识别词语，在基础词库中查找与所述待识别词语匹配的目标基础词；

[0154] 相似词查找单元301，用于若未查到所述目标基础词，在相似语义词库中查找与所述待识别词语匹配的目标相似词，其中，所述相似语义词库中的相似词为根据所述基础词库中的基础词按照设定规则进行处理得到的；

[0155] 确定单元302，用于若查找到所述目标相似词，根据基础词与相似词之间的映射关系，确定所述目标相似词对应的目标基础词；

[0156] 解析单元303，用于根据所述目标基础词确定所述文本信息的语义解析结果。

[0157] 作为一种可能的实施方式，所述确定单元302具体用于：

[0158] 若根据基础词与相似词之间的映射关系，确定出所述目标相似词对应的基础词为多个，控制智能设备输出提示信息，所述提示信息用于提示对多个基础词进行选择；

[0159] 获取反馈的确认信息，将所述确认信息所选定的基础词确定为所述目标基础词。

[0160] 作为一种可能的实施方式，所述相似词查找单元301具体用于：

[0161] 对于所述基础词库中的每个基础词，获取所述基础词的指定位置上设定字数的词片段，对于具有相同词片段的基础词，提取所述相同词片段作为相似语义词库中对应的相似词；

[0162] 对所述基础词库中包括的每个基础词进行分词处理，得到所述基础词对应的分词，将所得到的分词确定为所述基础词对应的相似词；

[0163] 对于所述基础词库中的基础词，从预设同义词库中选择与该基础词具有相同语义的同义词，根据所选择的同义词，确定所述基础词对应的相似词。

[0164] 作为一种可能的实施方式，所述基础词库配置有对应的生效范围，根据如下方式配置所述相似语义词库的生效范围：

[0165] 将基于任一基础词库得到的相似语义词库的生效范围配置为所述基础词库的生效范围；或者

[0166] 若所述相似语义词库是基于多个基础词库得到的，将所述相似语义词库的生效范围配置为所述多个基础词库的生效范围中最大的生效范围。

[0167] 作为一种可能的实施方式，所述相似词查找单元301具体用于：

[0168] 根据智能设备的属性信息以及各相似语义词库的生效范围，确定对所述智能设备

有效的目标相似语义词库；

[0169] 在所述目标相似语义词库中,查找与该词语匹配的目标相似词。

[0170] 作为一种可能的实施方式,所述相似词查找单元301具体还用于:

[0171] 根据所述专属基础词库中的专属词,按照设定规则进行处理得到专属相似词,基于得到的专属相似词形成专属相似语义词库;以及

[0172] 根据所述通用基础词库中的标准词,按照设定规则进行处理得到标准相似词,基于得到的标准相似词形成通用相似语义词库。

[0173] 作为一种可能的实施方式,所述基础词库的生效范围,包括如下至少一种:

[0174] 对于指定智能设备生效、对于指定企业生效、对于指定业务线生效、对于指定应用程序生效、对于指定领域生效、对于所有智能设备生效;

[0175] 所述相似语义词库的生效范围,包括如下至少一种:对于指定智能设备生效、对于指定企业生效、对于指定业务线生效、对于指定应用程序生效、对于指定领域生效、对于所有智能设备生效。

[0176] 作为一种可能的实施方式,通过如下方式获取文本信息中的待识别词语:

[0177] 对所述文本信息进行槽位解析,得到各槽位项对应的槽位值,将所得到的槽位值确定为待识别词语。

[0178] 如图4所示,基于相同的发明构思本发明实施例还提供一种电子设备,该设备包括处理器400和存储器401,其中,上述存储器存储可执行程序,上述程序执行时,上述处理器实现如下过程:

[0179] 对于获取到的文本信息中的待识别词语,在基础词库中查找与所述待识别词语匹配的目标基础词;

[0180] 若未查到所述目标基础词,在相似语义词库中查找与所述待识别词语匹配的目标相似词,其中,所述相似语义词库中的相似词为根据所述基础词库中的基础词按照设定规则进行处理得到的;

[0181] 若查找到所述目标相似词,根据基础词与相似词之间的映射关系,确定所述目标相似词对应的目标基础词;

[0182] 根据所述目标基础词确定所述文本信息的语义解析结果。

[0183] 作为一种可能的实施方式,所述处理器具体用于:

[0184] 若根据基础词与相似词之间的映射关系,确定出所述目标相似词对应的基础词为多个,控制智能设备输出提示信息,所述提示信息用于提示对多个基础词进行选择;

[0185] 获取反馈的确认信息,将所述确认信息所选定的基础词确定为所述目标基础词。

[0186] 作为一种可能的实施方式,所述处理器具体用于:

[0187] 对于所述基础词库中的每个基础词,获取所述基础词的指定位置上设定字数的词片段,对于具有相同词片段的基础词,提取所述相同词片段作为相似语义词库中对应的相似词;

[0188] 对所述基础词库中包括的每个基础词进行分词处理,得到所述基础词对应的分词,将所得到的分词确定为所述基础词对应的相似词;

[0189] 对于所述基础词库中的基础词,从预设同义词库中选择与该基础词具有相同语义的同义词,根据所选择的同义词,确定所述基础词对应的相似词。

[0190] 作为一种可能的实施方式,所述基础词库配置有对应的生效范围,所述处理器具体还用于:

[0191] 将基于任一基础词库得到的相似语义词库的生效范围配置为所述基础词库的生效范围;或者

[0192] 若所述相似语义词库是基于多个基础词库得到的,将所述相似语义词库的生效范围配置为所述多个基础词库的生效范围中最大的生效范围。

[0193] 作为一种可能的实施方式,所述处理器具体用于:

[0194] 根据智能设备的属性信息以及各相似语义词库的生效范围,确定对所述智能设备有效的目标相似语义词库;

[0195] 在所述目标相似语义词库中,查找与该词语匹配的目标相似词。

[0196] 作为一种可能的实施方式,所述基础词库包括专属基础词库和通用基础词库,所述处理器具体还用于:

[0197] 根据所述专属基础词库中的专属词,按照设定规则进行处理得到专属相似词,基于得到的专属相似词形成专属相似语义词库;以及

[0198] 根据所述通用基础词库中的标准词,按照设定规则进行处理得到标准相似词,基于得到的标准相似词形成通用相似语义词库。

[0199] 作为一种可选的实施方式,所述基础词库的生效范围,包括如下至少一种:

[0200] 对于指定智能设备生效、对于指定企业生效、对于指定业务线生效、对于指定应用程序生效、对于指定领域生效、对于所有智能设备生效;

[0201] 所述相似语义词库的生效范围,包括如下至少一种:对于指定智能设备生效、对于指定企业生效、对于指定业务线生效、对于指定应用程序生效、对于指定领域生效、对于所有智能设备生效。

[0202] 作为一种可选的实施方式,确定所述文本信息中每个待语义识别的词语,包括:

[0203] 对所述文本信息进行槽位解析,得到各槽位项对应的槽位值,将所得到的槽位值确定为待语义识别的词语。

[0204] 本申请实施例还提供一种计算机存储介质,其上存储有计算机程序,该程序被处理器执行时实现如下步骤:

[0205] 对于获取到的文本信息中的待识别词语,在基础词库中查找与所述待识别词语匹配的目标基础词;

[0206] 若未查到所述目标基础词,在相似语义词库中查找与所述待识别词语匹配的目标相似词,其中,所述相似语义词库中的相似词为根据所述基础词库中的基础词按照设定规则进行处理得到的;

[0207] 若查找到所述目标相似词,根据基础词与相似词之间的映射关系,确定所述目标相似词对应的目标基础词;

[0208] 根据所述目标基础词确定所述文本信息的语义解析结果。

[0209] 本领域内的技术人员应明白,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和光学存储器等)上实施的计算机程序产品的形

式。

[0210] 本申请是参照根据本申请实施例的方法、设备(系统)、和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器,使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的设备。

[0211] 这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中,使得存储在该计算机可读存储器中的指令产生包括指令设备的制品,该指令设备实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。

[0212] 这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上,使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理,从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

[0213] 显然,本领域的技术人员可以对本申请进行各种改动和变型而不脱离本申请的精神和范围。这样,倘若本申请的这些修改和变型属于本申请权利要求及其等同技术的范围之内,则本申请也意图包含这些改动和变型在内。

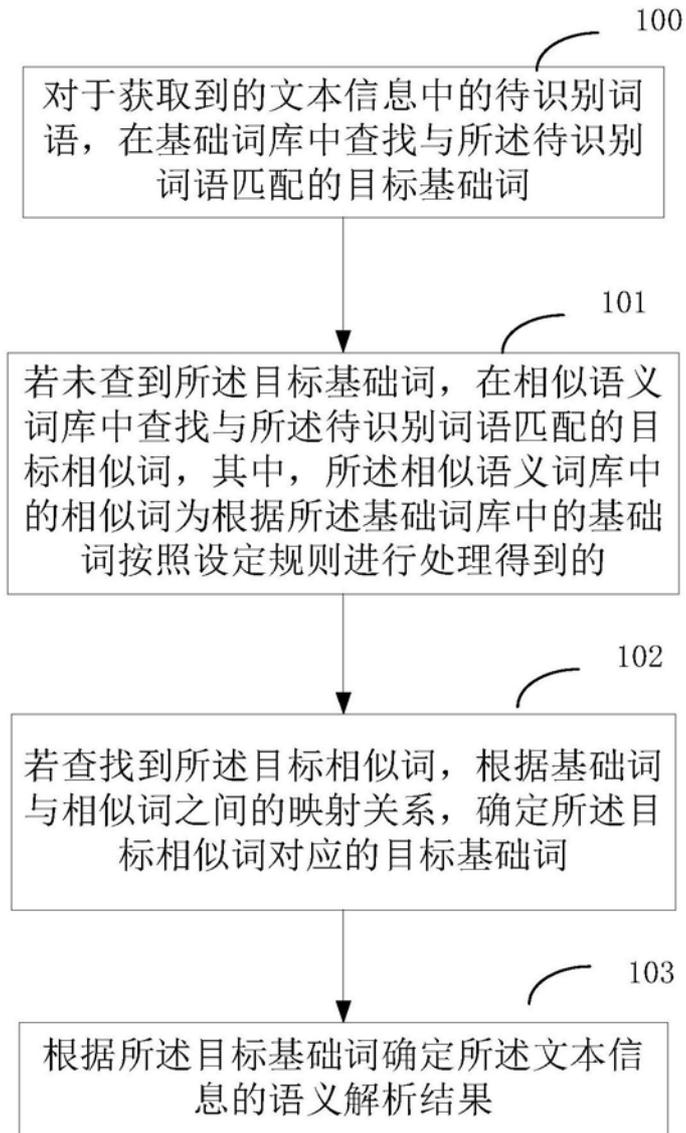


图1

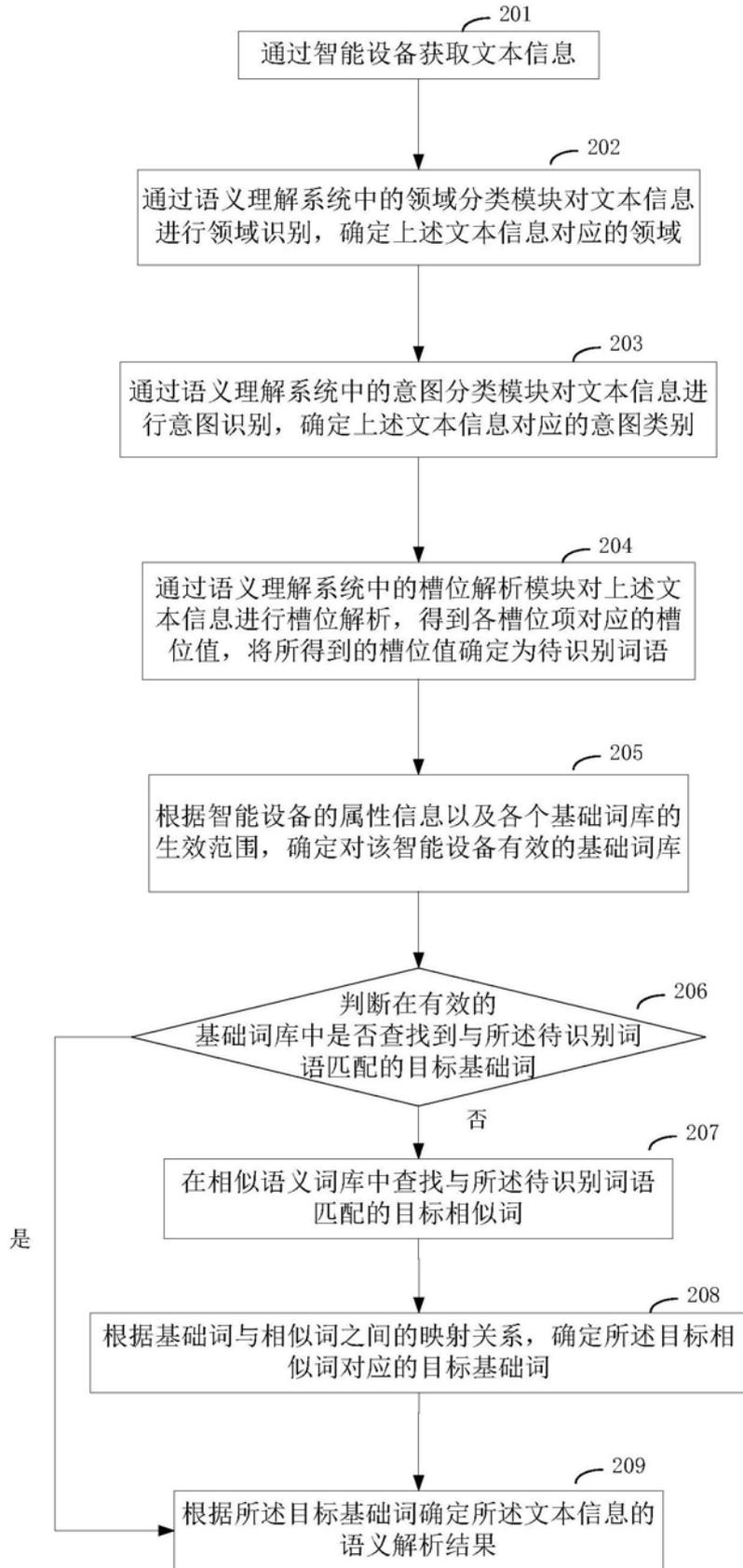


图2

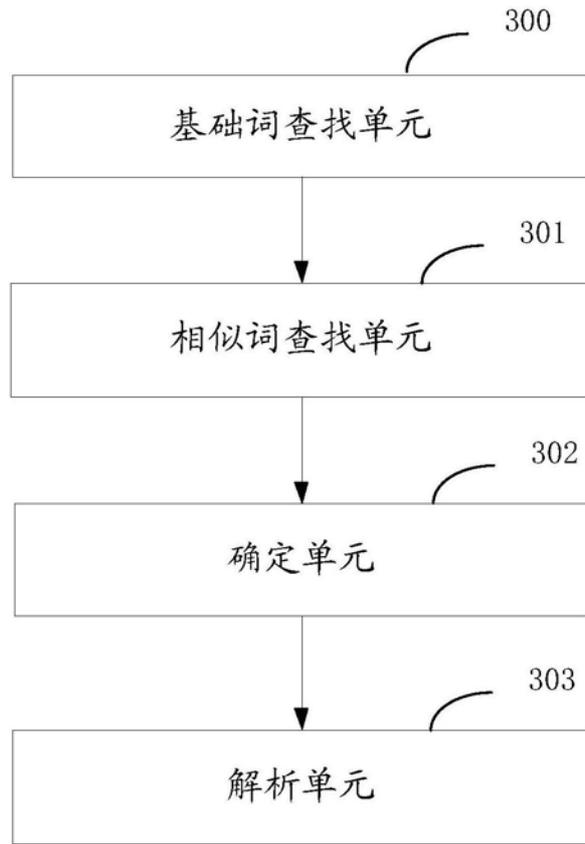


图3

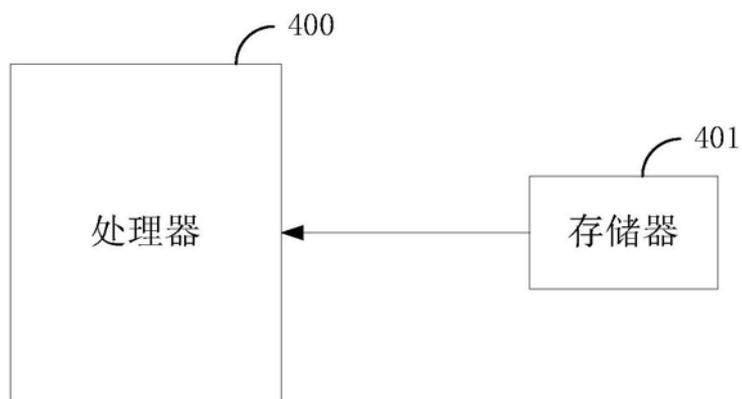


图4