(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 110083839 B
(45)授权公告日 2023.08.22

(21)申请号 201910359179.2

(22)申请日 2019.04.29

(65)同一申请的已公布的文献号
申请公布号 CN 110083839 A

(43)申请公布日 2019.08.02

(73)专利权人 珠海豹好玩科技有限公司
地址 519000 广东省珠海市横琴新区宝华路6号105室-53967(集中办公区)

(72)发明人 胡建 周振华

(74)专利代理机构 广州三环专利商标代理有限公司 44202
专利代理师 郝传鑫 熊永强

(51)Int.Cl.
G06F 40/279(2020.01)
G06F 16/16(2019.01)

(56)对比文件
CN 109189666 A,2019.01.11
CN 108829882 A,2018.11.16
CN 108304368 A,2018.07.20
CN 109388675 A,2019.02.26
CN 108279885 A,2018.07.13
CN 105607938 A,2016.05.25
CN 106293727 A,2017.01.04
CN 107251030 A,2017.10.13
CN 104462716 A,2015.03.25
US 2018322509 A1,2018.11.08
US 2014236971 A1,2014.08.21
CN 102982011 A,2013.03.20
丁俊等.大数据时代下的动态可配置数据采集系统的研究与设计.《计算机应用与软件》.2018,(第03期),

审查员 胡熠寒

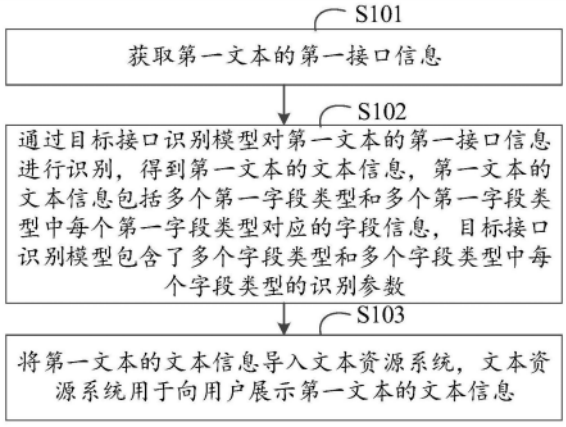
权利要求书2页 说明书11页 附图4页

(54)发明名称

文本导入方法、装置及设备

(57)摘要

本发明提供文本导入方法、装置及设备,其中,方法包括:获取第一文本的第一接口信息;通过目标接口识别模型对所述第一接口信息进行识别,得到所述第一文本的文本信息,所述文本信息包括多个第一字段类型和所述多个第一字段类型中每个第一字段类型对应的字段信息,所述目标接口识别模型包含了多个字段类型和所述多个字段类型中每个字段类型的识别参数;将所述第一文本的文本信息导入文本资源系统,所述文本资源系统用于向用户展示所述第一文本的文本信息。该技术方案可以提高文本导入文本资源系统的效率。



CN 110083839 B

1.一种文本导入方法,其特征在于,包括:

获取第一文本的第一接口信息;

通过目标接口识别模型对所述第一接口信息进行识别,得到所述第一文本的文本信息,所述文本信息包括多个第一字段类型和所述多个第一字段类型中每个第一字段类型对应的字段信息,所述目标接口识别模型包含了多个字段类型和所述多个字段类型中每个字段类型的识别参数;

将所述第一文本的文本信息导入文本资源系统,所述文本资源系统用于向用户展示所述第一文本的文本信息;

其中,所述目标接口识别模型为基本接口识别模型,所述基本接口识别模型可用于对不同提供方的文本的接口信息进行识别;

所述通过目标接口识别模型对所述第一接口信息进行识别,得到所述第一文本的文本信息,包括:

获取所述第一文本的第一提供方的标识;

若所述第一提供方的标识未包含于已识别的提供方标识集合中,则通过基本接口识别模型对所述第一接口信息进行识别,得到所述第一文本的文本信息;

其中,所述已识别的提供方标识集合中的任意一个提供方的标识用于指示已对所述提供方的文本的接口信息执行识别。

2.根据权利要求1所述的方法,其特征在于,所述方法还包括:

根据所述多个第一字段类型和所述多个第一字段类型的识别参数,生成第一接口识别模型;

建立所述第一提供方与所述第一接口识别模型的对应关系。

3.根据权利要求1所述的方法,其特征在于,所述通过目标接口识别模型对所述第一接口信息进行识别,得到所述第一文本的文本信息之前,还包括:

获取所述第一文本的第一提供方的标识;

若所述第一提供方的标识包含于已识别的提供方标识集合中,获取与所述第一提供方具有对应关系的第一接口识别模型;

将所述第一接口识别模型确定为所述目标接口识别模型。

4.根据权利要求3所述的方法,其特征在于,所述目标接口识别模型中的多个字段类型包括多个必选字段类型;

所述方法还包括:

在通过所述目标接口识别模型对所述第一接口信息未识别到至少一个必选字段类型对应的字段信息的情况下,输出识别异常信息,所述识别异常信息包括未识别到的至少一个必选字段类型。

5.一种文本导入装置,其特征在于,包括:

接口信息获取模块,用于获取第一文本的第一接口信息;

接口信息识别模块,用于通过目标接口识别模型对所述第一接口信息进行识别,得到所述第一文本的文本信息,所述文本信息包括多个第一字段类型和所述多个第一字段类型中每个第一字段类型对应的字段信息,所述目标接口识别模型包含了多个字段类型和所述多个字段类型中每个字段类型的识别参数;

文本信息导入模块，用于将所述第一文本的文本信息导入文本资源系统，所述文本资源系统用于向用户展示所述第一文本的文本信息，其中，所述目标接口识别模型为基本接口识别模型，所述基本接口识别模型可用于对不同提供方的文本的接口信息进行识别；

所述接口信息识别模块，具体用于获取所述第一文本的第一提供方的标识；

所述接口信息识别模块，具体用于若所述第一提供方的标识未包含于已识别的提供方标识集合中，则通过基本接口识别模型对所述第一接口信息进行识别，得到所述第一文本的文本信息；

其中，所述已识别的提供方标识集合中的任意一个提供方的标识用于指示已对所述提供方的文本的接口信息执行识别。

6.根据权利要求5所述的装置，其特征在于，所述装置还包括：

第一模型生成模块，用于根据所述多个第一字段类型和所述多个第一字段类型的识别参数，生成第一接口识别模型；

所述第一模型生成模块，还用于建立所述第一提供方与所述第一接口识别模型的对应关系。

7.根据权利要求5所述的装置，其特征在于，所述装置还包括：

目标模型确定模块，用于获取所述第一文本的第一提供方的标识；

所述目标模型确定模块，还用于若所述第一提供方的标识包含于已识别的提供方标识集合中，获取与所述第一提供方具有对应关系的第一接口识别模型；

所述目标模型确定模块，还用于将所述第一接口识别模型确定为所述目标接口识别模型。

8.根据权利要求7所述的装置，其特征在于，所述目标接口识别模型中的多个字段类型包括多个必选字段类型；

所述装置还包括：

异常信息输出模块，用于在通过所述目标接口识别模型对所述第一接口信息未识别到至少一个必选字段类型对应的字段信息的情况下，输出识别异常信息，所述识别异常信息包括未识别到的至少一个必选字段类型。

9.一种文本导入设备，包括处理器、存储器以及输入输出接口，所述处理器、存储器和输入输出接口相互连接，其中，所述输入输出接口用于输入或输出数据，所述存储器用于存储程序代码，所述处理器用于调用所述程序代码，执行如权利要求1-4任一项所述的方法。

10.一种计算机存储介质，其特征在于，所述计算机存储介质存储有计算机程序，所述计算机程序包括程序指令，所述程序指令当被处理器执行时使所述处理器执行如权利要求1-4任一项所述的方法。

文本导入方法、装置及设备

技术领域

[0001] 本发明涉及计算机技术领域，尤其涉及文本导入方法、装置及设备。

背景技术

[0002] 目前用户查看文本信息时，一般是在文本资源系统中查看相应的文本信息，文本信息开发商需要预先将多个文本信息提供方的文本信息导入文本资源系统。

[0003] 目前的文本信息开发商将多个文本信息提供方的文本信息导入文本资源系统时，由于每个文本提供方的接口包含的各个字段类型以及各个字段类型的参数存在差异，文本信息开发商需要针对每个文本信息提供方的接口做针对性的适配，需要开发相应的适配代码。由于文本信息提供方的数量较多，且每次识别每个文本信息提供方接口都需要开发相应的适配代码，操作复杂，降低了文本导入文本资源系统的效率。

发明内容

[0004] 本发明实施例提供文本导入方法、装置及设备，解决文本识别效率低以及文本导入文本资源系统效率低的问题。

[0005] 第一方面，提供文本导入方法，包括：

[0006] 获取第一文本的第一接口信息；

[0007] 通过目标接口识别模型对所述第一接口信息进行识别，得到所述第一文本的文本信息，所述文本信息包括多个第一字段类型和所述多个第一字段类型中每个第一字段类型对应的字段信息，所述目标接口识别模型包含了多个字段类型和所述多个字段类型中每个字段类型的识别参数；

[0008] 将所述第一文本的文本信息导入文本资源系统，所述文本资源系统用于向用户展示所述第一文本的文本信息。

[0009] 结合第一方面，在一种可能的实现方式中，所述目标接口识别模型为基本接口识别模型，所述基本接口识别模型可用于对不同提供方的文本的接口信息进行识别；所述通过目标接口识别模型对所述第一接口信息进行识别，得到所述第一文本的文本信息，包括：获取所述第一文本的第一提供方的标识；若所述第一提供方的标识未包含于已识别的提供方标识集合中，则通过基本接口识别模型对所述第一接口信息进行识别，得到所述第一文本的文本信息；其中，所述已识别的提供方标识集合中的任意一个提供方的标识用于指示已对所述提供方的文本的接口信息执行识别。

[0010] 结合第一方面，在一种可能的实现方式中，所述方法还包括：根据所述多个第一字段类型和所述多个第一字段类型的识别参数，生成第一接口识别模型；建立所述第一提供方与所述第一接口识别模型的对应关系。

[0011] 结合第一方面，在一种可能的实现方式中，所述通过目标接口识别模型对所述第一接口信息进行识别，得到所述第一文本的文本信息之前，还包括：获取所述第一文本的第一提供方的标识；若所述第一提供方的标识包含于已识别的提供方标识集合中，获取与所

述第一提供方具有对应关系的第一接口识别模型;将所述第一接口识别模型确定为所述目标接口识别模型。

[0012] 结合第一方面,在一种可能的实现方式中,所述目标接口识别模型中的多个字段类型包括多个必选字段类型;所述方法还包括:在通过所述目标接口识别模型对所述第一接口信息未识别到至少一个必选字段类型对应的字段信息的情况下,输出识别异常信息,所述识别异常信息包括未识别到的至少一个必选字段类型。

[0013] 第二方面,提供文本导入装置,包括:

[0014] 接口信息获取模块,用于获取第一文本的第一接口信息;

[0015] 接口信息识别模块,用于通过目标接口识别模型对所述第一接口信息进行识别,得到所述第一文本的文本信息,所述文本信息包括多个第一字段类型和所述多个第一字段类型中每个第一字段类型对应的字段信息,所述目标接口识别模型包含了多个字段类型和所述多个字段类型中每个字段类型的识别参数;

[0016] 文本信息导入模块,用于将所述第一文本的文本信息导入文本资源系统,所述文本资源系统用于向用户展示所述第一文本的文本信息。

[0017] 结合第二方面,在一种可能的实现方式中,所述目标接口识别模型为基本接口识别模型,所述基本接口识别模型可用于对不同提供方的文本的接口信息进行识别;所述接口信息识别模块,具体用于获取所述第一文本的第一提供方的标识;所述接口信息识别模块,具体用于若所述第一提供方的标识未包含于已识别的提供方标识集合中,则通过基本接口识别模型对所述第一接口信息进行识别,得到所述第一文本的文本信息;其中,所述已识别的提供方标识集合中的任意一个提供方的标识用于指示已对所述提供方的文本的接口信息执行识别。

[0018] 结合第二方面,在一种可能的实现方式中,所述装置还包括:第一模型生成模块,用于根据所述多个第一字段类型和所述多个第一字段类型的识别参数,生成第一接口识别模型;所述第一模型生成模块,还用于建立所述第一提供方与所述第一接口识别模型的对应关系。

[0019] 结合第二方面,在一种可能的实现方式中,所述装置还包括:目标模型确定模块,用于获取所述第一文本的第一提供方的标识;所述目标模型确定模块,还用于若所述第一提供方的标识包含于已识别的提供方标识集合中,获取与所述第一提供方具有对应关系的第一接口识别模型;所述目标模型确定模块,还用于将所述第一接口识别模型确定为所述目标接口识别模型。

[0020] 结合第二方面,在一种可能的实现方式中,所述目标接口识别模型中的多个字段类型包括多个必选字段类型;所述装置还包括:异常信息输出模块,用于在通过所述目标接口识别模型对所述第一接口信息未识别到至少一个必选字段类型对应的字段信息的情况下,输出识别异常信息,所述识别异常信息包括未识别到的至少一个必选字段类型。

[0021] 第三方面,提供文本导入设备,包括处理器、存储器、以及输入输出接口,所述处理器、存储器和输入输出接口相互连接,其中,所述输入输出接口用于输入或输出数据,所述存储器用于存储文本导入设备执行上述方法的应用程序代码,所述处理器被配置用于执行上述第一方面的方法。

[0022] 第四方面,提供一种计算机存储介质,所述计算机存储介质存储有计算机程序,所

述计算机程序包括程序指令，所述程序指令当被处理器执行时使所述处理器执行上述第一方面的方法。

[0023] 本发明实施例中，根据目标接口识别模型对获取到的第一文本的第一接口信息进行识别，得到第一文本的文本信息，并将第一文本的文本信息导入文本资源系统。通过目标接口识别模型自动对第一文本的第一接口信息进行识别，得到第一文本的文本信息并自动将第一文本的文本信息导入到文本资源系统，省去了在每次识别第一文本的第一接口信息时手动操作开发相应的适配代码，提高了文本导入文本资源系统的效率，提升了用户体验。

附图说明

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

[0025] 图1是发明实施例提供的一种文本导入方法的流程示意图；

[0026] 图2是发明实施例提供的另一种文本导入方法的流程示意图；

[0027] 图3是本发明实施例提供的一种识别异常信息显示界面示意图；

[0028] 图4是本发明实施例提供的一种文本导入方法的示例图；

[0029] 图5是本发明实施例提供的另一种文本导入方法的示例图；

[0030] 图6是本发明实施例提供的一种文本导入装置的结构示意图；

[0031] 图7是本发明实施例提供的一种文本导入设备的结构示意图。

具体实施方式

[0032] 下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅是本发明一部分实施例，而不是全部的实施例。基于本发明中的实施例，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，都属于本发明保护的范围。

[0033] 本发明实施例的方案适用于识别文本提供方的接口信息并将识别到的文本信息导入文本资源系统的场景中，通过自动对文本提供方的接口信息进行识别，得到对应的文本信息并导入至文本资源系统，可以提高文本识别效率以及文本导入文本资源系统的效率。

[0034] 参见图1，图1是本发明实施例提供的一种文本导入方法的流程示意图，如图所示，该方法包括：

[0035] S101，获取第一文本的第一接口信息。

[0036] 在一些可能的场景中，第一文本可以为电子文档，例如电子小说，第一文本是由第一提供方提供的，同一个提供方可以提供不同的第一文本，其中，每个第一文本的接口信息都包含共同的特征。例如第一提供方A提供的N个第一文本的文本名称的字段名常包含字符都为bookname、文本分类的字段名常包含字符都为category、文本字数的字段名常包含字符都为words；第一提供方B提供的M个第一文本的文本名称的字段名常包含字符都为book_name、文本分类的字段名常包含字符都为cate、文本字数的字段名常包含字符都为count，

等。

[0037] 具体实现中，可以将各个提供方提供的文本的接口信息加入接口待识别列表，这样，接口待识别列表中可以包括由多个提供方提供的各个文本的接口信息，以使文本导入装置从待识别列表中获取第一文本的第一接口信息。

[0038] S102，通过目标接口识别模型对第一文本的第一接口信息进行识别，得到第一文本的文本信息，第一文本的文本信息包括多个第一字段类型和多个第一字段类型中每个第一字段类型对应的字段信息，目标接口识别模型包含了多个字段类型和多个字段类型中每个字段类型的识别参数。

[0039] 这里，目标接口识别模型中的字段类型可以为待识别的文本所具有的特征，例如字段类型可以为文本ID、文本名称、文本作者、文本分类、文本章节名、文本字数、文本内容等；字段类型的识别参数可以包括字段类型对应的字段信息的数据类型、字段类型对应的字段信息的长度、字段类型对应的字段信息数据范围、字段类型常包含字符等。

[0040] 其中，字段类型对应的字段信息的数据类型即所对应的字段信息的数据的具体类型。例如文本名称对应的字段信息(如文本名称为边城)的数据类型为字符型、文本字数对应的字段信息(如文本字数为2000000)的数据类型为数值型。

[0041] 字段类型对应的字段信息的长度即具体字段类型对应的字段信息的信息长度，例如文本名称为边城，则字段类型对应的字段信息为“边城”，“边城”的字段长度为4；文本字数为2000000，则字段类型对应的字段信息为“2000000”，“2000000”的字段长度为7。

[0042] 字段类型对应的字段信息数据范围即具体的文本名称数据范围，例如具体的文本名称(如边城)的数据范围可以为1～50范围内，即属于该数据范围的字段类型对应的字段信息为文本名称、文本字数的数据范围可以为1～2147483647范围内，即属于该数据范围的字段类型对应的字段信息为文本字数。

[0043] 字段类型常包含字符即哪些字符可以用于表示字段类型，例如，字段类型为文本名称，则常包含的字符可以为name，bookname，book_name等；字段类型为文本分类，则常包含字符可以为cate，category，class等；字段类型为文本字数，则常包含字符可以为words，count，cnt等。

[0044] 例如，对通过目标接口识别模型对第一文本的第一接口信息进行识别，得到第一文本的文本信息的过程进行举例说明。目标接口识别模型中包含的字段类型有文本名称、文本分类、文本字数等，文本名称的识别参数为字符型、1～50个字符、utf8范围字符、name/bookname/book_name；文本分类的识别参数为字符型、2～4个字符、utf8范围字符、cate/category/class；文本字数的识别参数为数值型、1～10个字符、1～2147483647、cate/category/class，等等，则识别后得到的第一文本的文本信息可以为：

[0045] "bookname":"咸鱼翻身记"，

[0046] "author":"东方不败"，

[0047] "category":"都市"，

[0048] "words":"1964556"，

[0049] "chapter count":"998"，

[0050] "cover":"http://xxx.com/yyyy.jpg"，

[0051] "introduction":"大学毕业的张三，在升职的关键，被自己的朋友陷害，名誉扫

地,被迫从事自己最不喜欢的工作……"

[0052] ……

[0053] 则多个第一字段类型包括:"bookname"、"author"、"category"、"words"、"chapter count"、"cover"、"introduction",多个第一字段类型对应的字段信息包括:"咸鱼翻身记"、"东方不败"、"都市"、"1964556"、"998"、"http://xxx.com/yyyy.jpg"、"大学毕业的张三,在升职的关键,被自己的朋友陷害,名誉扫地,被迫从事自己最不喜欢的工作……"

[0054] S103,将第一文本的文本信息导入文本资源系统,文本资源系统用于向用户展示第一文本的文本信息。

[0055] 这里,文本资源系统可以存储该文本信息。用户可以通过文本资源系统查看到第一文本的文本信息,具体可以查看到第一文本的文本ID、文本名称、文本作者、文本分类、文本章节名、文本字数、文本内容等。在一种可能的情况下,文本资源系统可以为小说资源系统,例如可以为XX小说软件对应的小说资源库,小程序对应的小说资源库。

[0056] 具体实现中,可以预先建立目标接口识别模型与文本资源系统的对应关系。其中,目标接口识别模型包含了多个字段类型,文本资源系统包含了多个业务字段类型,即建立多个字段类型与文本资源系统中多个业务字段类型的对应关系。

[0057] 这里,对应关系可以为映射关系,根据目标接口识别模型中的字段类型以及目标接口识别模型与文本资源系统的映射关系,确定与目标接口识别模型中的字段类型具有映射关系的文本资源系统中的业务字段类型,并将识别得到的第一字段类型对应的字段信息,映射给与目标接口识别模型中字段类型对应的业务字段类型所对应的字段信息,从而实现将通过目标接口识别模型识别得到的第一文本的文本信息导入文本资源系统。

[0058] 例如,目标接口识别模型中的第一字段类型为文本名称对应于文本资源系统中的业务字段类型为文本名称;目标接口识别模型中的第一字段类型为文本分类对应于文本资源系统中的业务字段类型为文本分类;目标接口识别模型中的第一字段类型为文本字数对应于文本资源系统中的业务字段类型为文本字数。通过目标接口识别模型识别得到的第一字段类型以及第一字段类型对应的字段信息分别为"bookname":"咸鱼翻身记"、"category":"都市"、"book words":"1964556",则文本资源系统中"文本名称"对应的字段信息为"咸鱼翻身记"、文本资源系统中"文本分类"对应的字段信息为"咸鱼翻身记"、文本资源系统中"文本字数"对应的字段信息为"1964556"。

[0059] 可选地,还可以建立目标接口识别模型与文本资源系统的类型转换关系,即建立目标接口识别模型中多个字段类型与文本资源系统中多个业务字段类型的类型转换关系,包括:

[0060] 例如识别后得到的第一文本的第一文本信息为"book words":"1964556",识别出来的"1964556"为字符型,与目标接口识别模型中文本字数对应的文本资源系统中业务字段为"文本字数",文本资源系统中"文本字数"对应的字段信息应该为数值型,则对应关系为字符型转数值型,则"文本字数"对应的字段信息为数值型982278而非字符型"1964556"。

[0061] 又如,识别后得到的第一文本的第一文本信息为"date":"2019/1/14",与目标接口识别模型中文本日期对应的文本资源系统中业务字段为"日期",文本资源系统中"日期"对应的字段信息应该为数值型,则对应关系为字符型转日期型,则"日期"对应的字段信息

为日期型2019年1月14日而非字符型"2019/1/14"。可选的，日期可以为文本的上市日期、文本的写作完成日期等等。

[0062] 在一种可能的情况下，识别得到的第一文本的文本信息为：

[0063] "知否知否应是绿肥红瘦"，

[0064] "author":"关心则乱"，

[0065] "言情"，

[0066] "words":"1964556"，

[0067] ……

[0068] 其中，"知否知否应是绿肥红瘦"与"言情"为不同的两个第一字段类型对应的字段信息，两者的类型都为字符型、长度都为1～50个字符内，则可以将字符长度多的字段信息确定为文本名称对应的字段信息，将字符长度少的字段信息确定为文本分类对应的字段信息，即将"知否知否应是绿肥红瘦"确定为文本名称、将"言情"确定为文本分类。可选地，文本导入人员可以对文本资源系统中的业务字段类型对应的字段信息进行人工调整，以确定文本资源系统中的业务字段类型对应的字段信息导入准确。

[0069] 本发明实施例中，根据目标接口识别模型对获取到的第一文本的第一接口信息进行识别，得到第一文本的文本信息，并将第一文本的文本信息导入文本资源系统。通过目标接口识别模型自动对第一文本的第一接口信息进行识别，得到第一文本的文本信息并自动将第一文本的文本信息导入到文本资源系统，省去了在每次识别第一文本的第一接口信息时手动操作开发相应的适配代码，提高了文本导入文本资源系统的效率，提升了用户体验。

[0070] 在一种可能的实现方式中，可以通过判断第一文本的第一接口信息是否为首次识别，从而确定识别第一文本的第一接口信息的目标接口识别模型，可以提高文本识别的效率以及准确率。参见图2，图2是发明实施例提供的另一种文本导入方法的流程示意图，如图所示，该方法包括：

[0071] S201，获取第一文本的第一接口信息。

[0072] 具体获取第一文本的第一接口信息的方法参见步骤S101中的描述，此处不再赘述。

[0073] S202，获取第一文本的第一提供方的标识。

[0074] 这里，第一文本的第一提供方的标识用于唯一地指示该第一文本的第一提供方。具体地，该第一文本的第一提供方的标识可以为第一文本的第一提供方的名称、该第一文本的第一提供方的名称缩写、该第一文本的第一提供方的图标等标识中的一种或多种。

[0075] S203，判断第一提供方的标识是否包含于已识别的提供方标识集合中。

[0076] S204，若否，则通过基本接口识别模型对第一接口信息进行识别，得到第一文本的文本信息。

[0077] 这里，已识别的提供方标识集合中的任意一个提供方的标识用于指示已对提供方的文本的接口信息执行识别，基本接口识别模型可用于对不同提供方的文本的接口信息进行识别，基本接口识别模型中包含有所有提供方的文本可能包含的特征。

[0078] 具体地，若第一提供方的标识未包含于已识别的提供方标识集合中，则表示第一提供方的文本的接口信息为首次识别，则通过基本接口识别模型对第一接口信息进行识别。在通过基本接口识别模型对第一接口信息进行第一次识别后，将该第一提供方的标识

加入已识别的提供方标识集合中。这里，基本接口识别模型可以为步骤S102中的目标接口识别模型，具体通过目标接口识别模型识别第一接口信息的方法参见步骤S102中的描述，此处不再赘述。

[0079] 在第一提供方的文本的接口信息为首次识别的情况下，通过基本接口识别模型对第一接口信息进行识别，得到第一文本的文本信息后，还可以根据第一文本的文本信息中的多个第一字段类型和多个第一字段类型的识别参数，生成第一接口识别模型。换句话说，第一接口识别模型包含与该第一接口识别模型所对应的文本提供方的所有文本所共有的接口信息。不同提供方对应的第一接口识别模型中的字段类型常包含字符可能不同，例如提供方A的第一接口识别模型A中字段类型为文本名称，则文本名称常包含字符可能为bookname、字段类型为文本分类，则文本分类常包含字符可能为cate、字段类型为文本字数，则文本字数常包含字符可能为words。提供方B的第一接口识别模型B中字段类型为文本名称，则文本名称常包含字符可能为name、字段类型为文本分类，则文本分类常包含字符可能为category、字段类型为文本字数，则文本字数常包含字符可能为count。

[0080] 可选的，在首次识别生成第一接口识别模型后，还可以建立第一提供方与第一接口识别模型的对应关系，具体是将第一提供方的标识与第一接口识别模型对应起来，以便于需要识别该第一提供方的其他文本的接口信息的情况下，可以根据第一提供方的标识查找与其对应的第一接口识别模型。

[0081] 这里，每个第一提供方的标识对应一个第一接口识别模型，通过确定第一提供方的标识可以确定与该标识对应的第一接口识别模型，通过第一接口识别模型识别第一提供方的文本的接口信息，可以得到该文本的文本信息。

[0082] 通过步骤S201-S204的方法可以实现通过基本接口识别模型识别每个提供方的文本的接口信息，得到每个提供方的第一文本的文本信息，进而生成每个提供方对应的第一接口识别模型。

[0083] 举例来进行说明，例如有提供方A、B、C，通过基本接口识别模型分别识别提供方A、B、C分别提供的文本的接口信息，得到提供方A对应的第一字段类型A1～A8以及第一字段类型的识别参数a1～a8、提供方B对应的第一字段类型B1～B7以及第一字段类型的识别参数b1～b7、提供方C对应的第一字段类型C1～C9以及第一字段类型的识别参数c1～c7。通过提供方A对应的第一字段类型A1～A8以及第一字段类型的识别参数a1～a8生成第一接口识别模型A、通过提供方B对应的第一字段类型B1～B7以及第一字段类型的识别参数b1～b7生成第一接口识别模型B、通过提供方C对应的第一字段类型C1～C9以及第一字段类型的识别参数c1～c9生成第一接口识别模型C；则第一提供方与第一接口识别模型的对应关系为：提供方A对应第一接口识别模型A、提供方B对应第一接口识别模型B、提供方C对应第一接口识别模型C，由此，可得到每个提供方对应的第一接口识别模型。

[0084] S205，若是，获取与第一提供方具有对应关系的第一接口识别模型。

[0085] 这里，若第一提供方的标识包含于已识别的提供方标识集合中，则表示第一提供方的文本的接口信息为非首次识别，则获取与该第一提供方具有对应关系的第一接口识别模型，如该第一提供方B对应第一接口识别模型B为步骤S204中的提供方B，则与提供方B具有对应关系的第一接口识别模型为第一接口识别模型B。

[0086] S206，将第一接口识别模型确定为目标接口识别模型。

[0087] 例如,将步骤S205中举例的第一接口识别模型B确定为目标接口识别模型。

[0088] S207,通过目标接口识别模型对第一文本的第一接口信息进行识别,得到第一文本的文本信息。

[0089] 这里,具体通过目标接口识别模型识别第一文本的第一接口信息的方法参见步骤S102中的描述,此处不再赘述。

[0090] 针对本实施例的一种可能的实现方案,在通过目标接口识别模型对第一接口信息未识别到至少一个必选字段类型对应的字段信息的情况下,输出识别异常信息,识别异常信息包括未识别到的至少一个必选字段类型。这一情况下,可选地,目标接口识别模型为基本接口识别模型时,基本接口识别模型中可以包含多个必选字段和多个非必选字段类型,目标接口识别模型为第一接口识别模型时,第一接口识别模型中也可以包含多个必选字段和多个非必选字段类型。

[0091] 这里,必选字段类型可以为文本ID、文本名称、文本作者、文本分类、文本章节名、文本字数、文本内容等用于描述文本的必要特征;识别异常信息可以为在显示界面上显示“XX字段类型识别异常”,识别异常信息也可以为通过语音提示“XX字段类型识别异常”。例如,通过目标接口识别模型对第一接口信息未识别到必选字段中文本内容对应的字段信息的情况下,输出识别异常信息“文本内容识别异常”,如图3所示,图3是本发明实施例提供的一种识别异常信息显示界面示意图。

[0092] 可选地,目标接口识别模型中的多个字段类型还可以包括多个非必选字段类型,例如文本的非必要特征,如文本阅读者信息,即文本阅读用户是否为会员,或者文本阅读者年龄等,即文本阅读用户须在某个年龄区间,例如年龄区间可以为[8,60]等。在这种情况下,通过目标接口识别模型对第一接口信息未识别到至少一个非必选字段类型对应的字段信息的情况下,可以根据具体情况确定是否输出识别异常信息。例如可以根据该非必选字段类型的重要性确定是否输出识别异常信息。

[0093] S208,将第一文本的文本信息导入文本资源系统,文本资源系统用于向用户展示第一文本的文本信息。

[0094] 这里,具体将第一文本的文本信息导入文本资源系统的方法参见步骤S103的描述,此处不再赘述。

[0095] 本发明实施例中,通过判断第一文本的第一提供方的标识是否包含于已识别的提供方标识集合中,从而确定第一提供方的文本的接口信息是否为首次识别,在第一提供方的文本的接口信息为首次识别的情况下,根据识别到的第一文本的文本信息生成与第一提供方对应的第一接口识别模型,由于该第一接口识别模型具有第一提供方的第一文本的文本信息的特征,所以使用第一接口识别模型识别第一提供方的第一文本的文本信息可以提高识别效率,使得识别结果更准确,在通过第一接口识别模型识别第一提供方的文本的接口信息时,未识别出所有必选字段对应的字段信息的情况下,输出针对未识别出的必选字段的识别异常信息,可以使文本导入人员了解当前识别异常原因从而进行相应的调整,提高文本识别效率以及文本导入文本信息库的效率。

[0096] 在一种可能的实现方式中,文本导入方法可以应用在以下系统的结构图中。其中,第一种系统结构图可以如图4所示,图4是本发明实施例提供的一种文本导入方法的示例图,该文本导入系统中包括第一提供方401、文本导入装置402,其中,文本资源系统402包含

文本导入装置4021。首先，第一提供方401提供了第一文本的第一接口信息；接着，文本导入装置4021获取第一文本的第一接口信息，通过文本导入装置4021中的目标接口识别模型对第一文本的第一接口信息进行识别，得到第一文本的文本信息；最后，文本导入装置4021根据第一文本的文本信息与文本资源系统402的对应关系将第一文本的文本信息导入到文本资源系统402中的各个业务字段类型对应的字段信息，实现文本导入文本资源系统402。

[0097] 第二种系统结构图可以如图5所示，图5是本发明实施例提供的另一种文本导入方法的示例图，该文本导入系统中包括第一提供方501、文本导入装置502以及文本资源系统503。首先，第一提供方501提供了第一文本的第一接口信息；接着，文本导入装置502获取第一文本的第一接口信息，通过文本导入装置502中的目标接口识别模型对第一文本的第一接口信息进行识别，得到第一文本的文本信息；最后，根据文本导入装置502中的目标接口识别模型与文本资源系统503的对应关系将第一文本的文本信息导入与目标接口识别模型对应的文本资源系统503中的各个业务字段类型对应的字段信息，实现文本导入文本资源系统503。

[0098] 本发明实施例涉及的文本导入装置可以是具备处理能力的设备，例如：平板电脑、手机、电子阅读器、个人计算机(Personal Computer，PC)、笔记本电脑、服务器等设备；或者可以为嵌入在文本资源系统中的文本导入模块。本发明实施例对此不做限定。

[0099] 上面介绍了发明实施例的方法，下面介绍发明实施例的装置。

[0100] 参见图6，图6是本发明实施例提供的一种文本导入装置的结构示意图，该装置包括：

[0101] 接口信息获取模块601，用于获取第一文本的第一接口信息；

[0102] 接口信息识别模块602，用于通过目标接口识别模型对所述第一接口信息进行识别，得到所述第一文本的文本信息，所述文本信息包括多个第一字段类型和所述多个第一字段类型中每个第一字段类型对应的字段信息，所述目标接口识别模型包含了多个字段类型和所述多个字段类型中每个字段类型的识别参数；

[0103] 文本信息导入模块603，用于将所述第一文本的文本信息导入文本资源系统，所述文本资源系统用于向用户展示所述第一文本的文本信息。

[0104] 在一种可能的设计中，所述目标接口识别模型为基本接口识别模型，所述基本接口识别模型可用于对不同提供方的文本的接口信息进行识别；

[0105] 所述接口信息识别模块602，具体用于获取所述第一文本的第一提供方的标识；

[0106] 所述接口信息识别模块602，具体用于若所述第一提供方的标识未包含于已识别的提供方标识集合中，则通过基本接口识别模型对所述第一接口信息进行识别，得到所述第一文本的文本信息；

[0107] 其中，所述已识别的提供方标识集合中的任意一个提供方的标识用于指示已对所述提供方的文本的接口信息执行识别。

[0108] 在一种可能的设计中，所述装置还包括：

[0109] 第一模型生成模块604，用于根据所述多个第一字段类型和所述多个第一字段类型的识别参数，生成第一接口识别模型；

[0110] 所述第一模型生成模块604，还用于建立所述第一提供方与所述第一接口识别模型的对应关系。

[0111] 在一种可能的设计中,所述装置60还包括:

[0112] 目标模型确定模块605,用于获取所述第一文本的第一提供方的标识;

[0113] 所述目标模型确定模块605,还用于若所述第一提供方的标识包含于已识别的提供方标识集合中,获取与所述第一提供方具有对应关系的第一接口识别模型;

[0114] 所述目标模型确定模块605,还用于将所述第一接口识别模型确定为所述目标接口识别模型。

[0115] 在一种可能的设计中,所述目标接口识别模型中的多个字段类型包括多个必选字段类型;

[0116] 所述装置60还包括:

[0117] 异常信息输出模块606,用于在通过所述目标接口识别模型对所述第一接口信息未识别到至少一个必选字段类型对应的字段信息的情况下,输出识别异常信息,所述识别异常信息包括未识别到的至少一个必选字段类型。

[0118] 需要说明的是,图6对应的实施例中未提及的内容可参见方法实施例的描述,这里不再赘述。

[0119] 本发明实施例中,根据目标接口识别模型对获取到的第一文本的第一接口信息进行识别,得到第一文本的文本信息,并将第一文本的文本信息导入文本资源系统。通过目标接口识别模型自动对第一文本的第一接口信息进行识别,得到第一文本的文本信息并自动将第一文本的文本信息导入到文本资源系统,省去了在每次识别第一文本的第一接口信息时手动操作开发相应的适配代码,提升了文本导入文本资源系统的效率;通过判断第一文本的第一提供方的标识是否包含于已识别的提供方标识集合中,从而确定第一提供方的文本的接口信息是否为首次识别,在第一提供方的文本的接口信息为首次识别的情况下,生成与第一提供方对应的第一接口识别模型,由于该第一接口识别模型具有第一提供方的第一文本的文本信息的特征,所以使用第一接口识别模型识别第一提供方的第一文本的文本信息可以提高识别效率,使得识别结果更准确;在通过第一接口识别模型识别第一提供方的文本的接口信息时,未识别出所有必选字段对应的字段信息的情况下,输出针对未识别出的必选字段的识别异常信息,就可以使文本导入人员了解当前识别异常原因从而进行相应的调整,提高文本识别效率以及文本导入文本资源系统的效率。

[0120] 参见图7,图7是本发明实施例提供的一种文本导入设备的结构示意图,该设备70包括处理器701、存储器702以及输入输出接口703。处理器701连接到存储器702和输入输出接口703,例如处理器701可以通过总线连接到存储器702和输入输出接口703。

[0121] 处理器701被配置为支持所述文本导入设备执行图1-图2所述的文本导入方法中相应的功能。该处理器701可以是中央处理器(central processing unit,CPU),网络处理器(network processor,NP),硬件芯片或者其任意组合。上述硬件芯片可以是专用集成电路(application specific integrated circuit,ASIC),可编程逻辑器件(programmable logic device,PLD)或其组合。上述PLD可以是复杂可编程逻辑器件(complex programmable logic device,CPLD),现场可编程逻辑门阵列(field-programmable gate array,FPGA),通用阵列逻辑(generic array logic,GAL)或其任意组合。

[0122] 存储器702存储器用于存储程序代码等。存储器702可以包括易失性存储器(volatile memory,VM),例如随机存取存储器(random access memory,RAM);存储器702也

可以包括非易失性存储器(non-volatile memory,NVM),例如只读存储器(read-only memory,ROM),快闪存储器(flash memory),硬盘(hard disk drive,HDD)或固态硬盘(solid-state drive,SSD);存储器702还可以包括上述种类的存储器的组合。

[0123] 所述输入输出接口703用于输入或输出数据。

[0124] 处理器701可以调用所述程序代码以执行以下操作:

[0125] 获取第一文本的第一接口信息;

[0126] 通过目标接口识别模型对所述第一接口信息进行识别,得到所述第一文本的文本信息,所述文本信息包括多个第一字段类型和所述多个第一字段类型中每个第一字段类型对应的字段信息,所述目标接口识别模型包含了多个字段类型和所述多个字段类型中每个字段类型的识别参数;

[0127] 将所述第一文本的文本信息导入文本资源系统,所述文本资源系统用于向用户展示所述第一文本的文本信息。

[0128] 需要说明的是,各个操作的实现还可以对应参照上述方法实施例的相应描述;所述处理器701还可以与输入输出接口703配合执行上述方法实施例中的其他操作。

[0129] 本发明实施例还提供一种计算机存储介质,所述计算机存储介质存储有计算机程序,所述计算机程序包括程序指令,所述程序指令当被计算机执行时使所述计算机执行如前述实施例所述的方法,所述计算机可以为上述提到的文本导入设备的一部分。例如为上述的处理器701。

[0130] 本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程,是可以通过计算机程序来指令相关的硬件来完成,所述的程序可存储于计算机可读取存储介质中,该程序在执行时,可包括如上述各方法的实施例的流程。其中,所述的存储介质可为磁碟、光盘、ROM或RAM等。

[0131] 以上所揭露的仅为本发明较佳实施例而已,当然不能以此来限定本发明之权利范围,因此依本发明权利要求所作的等同变化,仍属本发明所涵盖的范围。

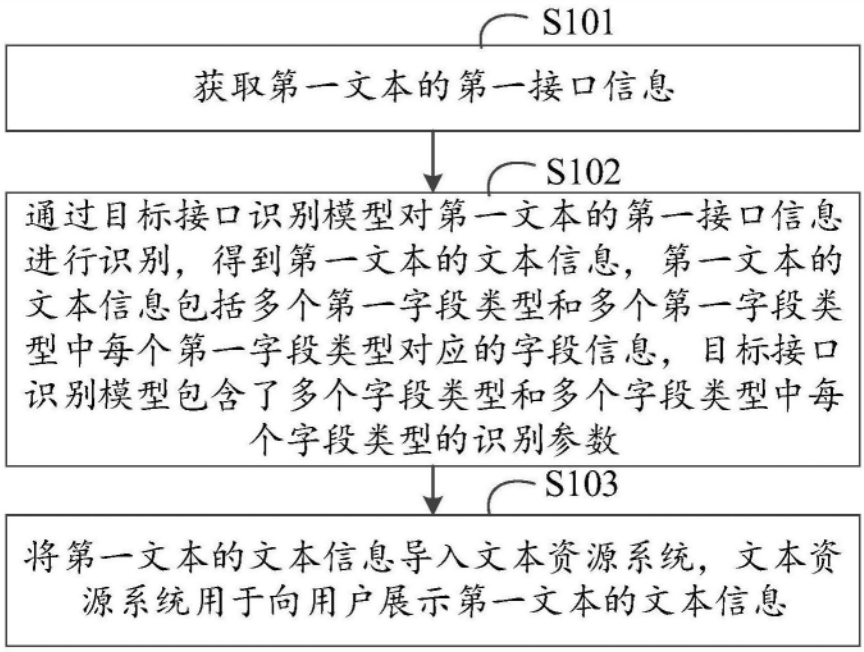

S101
获取第一文本的第一接口信息
S102
通过目标接口识别模型对第一文本的第一接口信息进行识别，得到第一文本的文本信息，第一文本的文本信息包括多个第一字段类型和多个第一字段类型中每个第一字段类型对应的字段信息，目标接口识别模型包含了多个字段类型和多个字段类型中每个字段类型的识别参数
S103
将第一文本的文本信息导入文本资源系统，文本资源系统用于向用户展示第一文本的文本信息

图1

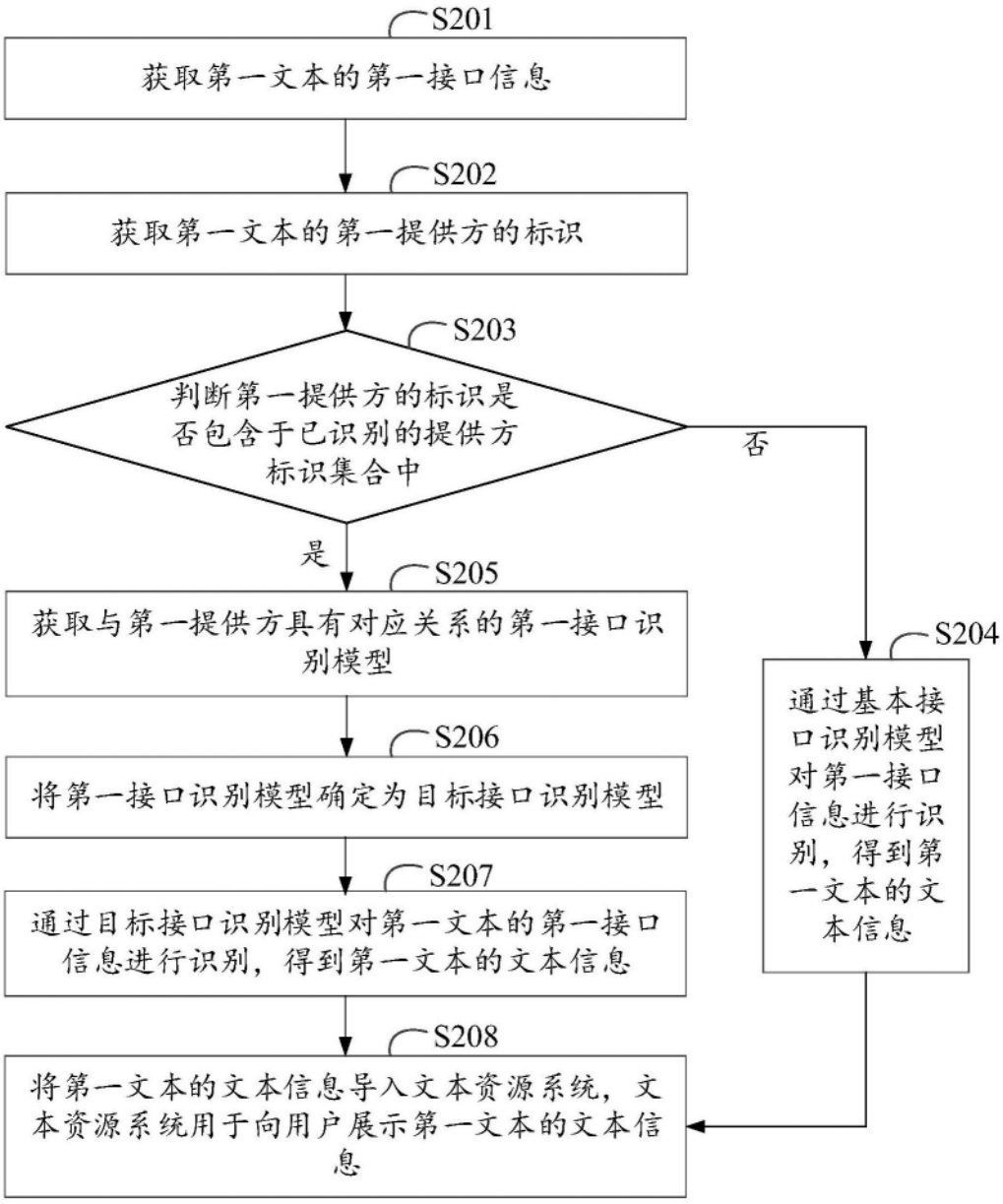

S201
获取第一文本的第一接口信息
S202
获取第一文本的第一提供方的标识
S203
判断第一提供方的标识是否包含于已识别的提供方标识集合中
否
是
S205
获取与第一提供方具有对应关系的第一接口识别模型
S206
将第一接口识别模型确定为目标接口识别模型
S207
通过目标接口识别模型对第一文本的第一接口信息进行识别，得到第一文本的文本信息
S204
通过基本接口识别模型对第一接口信息进行识别，得到第一文本的文本信息
S208
将第一文本的文本信息导入文本资源系统，文本资源系统用于向用户展示第一文本的文本信息

图2

图3

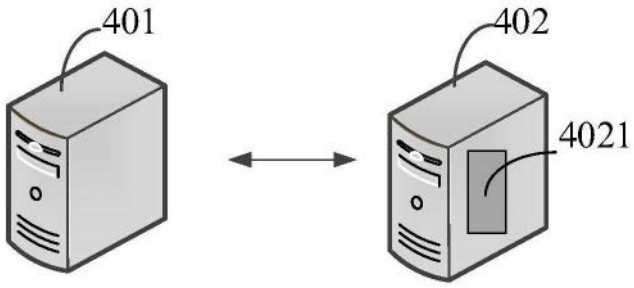


图4

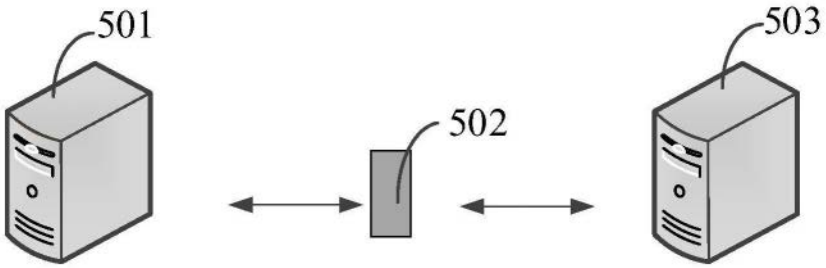


图5

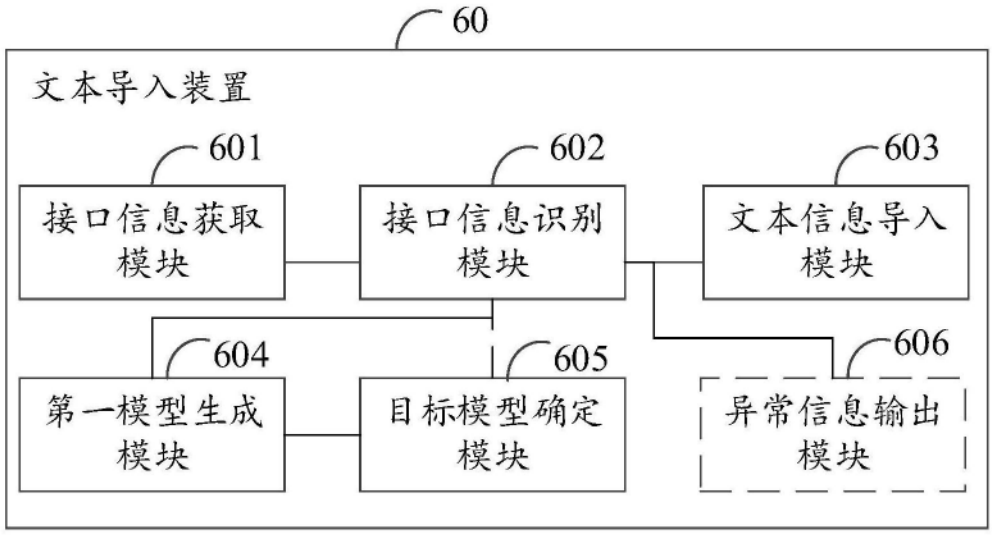

60
文本导入装置
601
接口信息获取模块
602
接口信息识别模块
603
文本信息导入模块
604
第一模型生成模块
605
目标模型确定模块
606
异常信息输出模块

图6

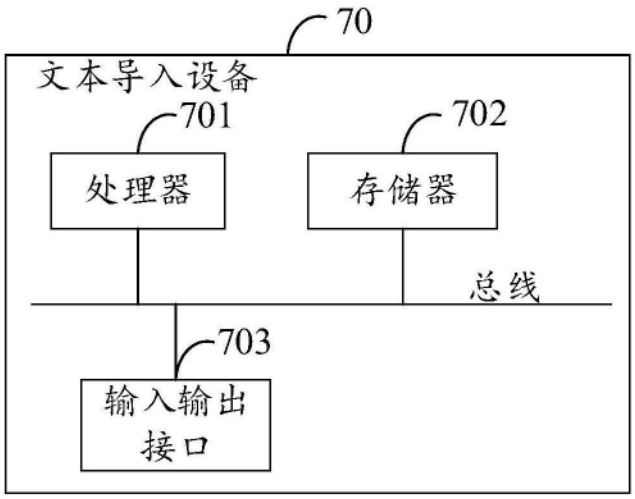

70
文本导入设备
701
处理器
702
存储器
总线
703
输入输出接口

图7