



## (12) 发明专利申请

(10) 申请公布号 CN 119540587 A

(43) 申请公布日 2025. 02. 28

(21) 申请号 202411686579.1

(22) 申请日 2024.11.22

(71) 申请人 北京百度网讯科技有限公司

地址 100085 北京市海淀区上地十街10号  
百度大厦2层

(72) 发明人 邱奇

(74) 专利代理机构 中科专利商标代理有限责任  
公司 11021

专利代理师 王文思

(51) Int. Cl.

G06V 10/75 (2022.01)

G06N 5/04 (2023.01)

G06V 20/62 (2022.01)

权利要求书2页 说明书9页 附图4页

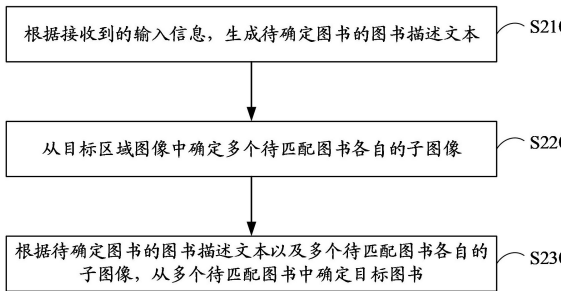
## (54) 发明名称

基于大模型技术的书籍检索方法、装置、设  
备和介质

## (57) 摘要

本公开提供了一种基于大模型技术的书籍检索方法,涉及人工智能技术领域,尤其涉及自然语言处理、计算机视觉、大模型、生成模型技术领域。具体实施方案为:根据接收到的输入信息,生成待确定图书的图书描述文本;从目标区域图像中确定多个待匹配图书各自的子图像;根据待确定图书的图书描述文本以及多个待匹配图书各自的子图像,从多个待匹配图书中确定目标图书。本公开还提供了一种基于大模型技术的书籍检索装置、电子设备和存储介质。

200



1. 一种基于大模型技术的书籍检索方法,包括:  
根据接收到的输入信息,生成待确定图书的图书描述文本;  
从目标区域图像中确定多个待匹配图书各自的子图像;以及  
根据所述待确定图书的图书描述文本以及所述多个待匹配图书各自的子图像,从所述多个待匹配图书中确定目标图书。
2. 根据权利要求1所述的方法,其中,所述输入信息包括输入语音、输入文本和输入图像中的至少之一;所述根据接收到的输入信息,生成待确定图书的图书描述文本包括:  
从所述输入语音和所述输入文本中的至少之一中提取所述待确定图书的第一特征描述文本;  
从所述输入图像中提取所述待确定图书的第二特征描述文本;以及  
根据所述第一特征描述文本和所述第二特征描述文本中的至少之一,确定所述图书描述文本。
3. 根据权利要求2所述的方法,其中,所述从所述输入语音和所述输入文本中的至少之一中提取所述待确定图书的第一特征描述文本包括:  
将所述输入语音转换为转换文本;以及  
从所述转换文本和所述输入文本中的至少之一中提取包含所述待确定图书的名称、作者、版本信息和出版信息中的至少之一的描述文本,作为所述第一特征描述文本。
4. 根据权利要求2所述的方法,其中,所述从所述输入图像中提取所述待确定图书的第二特征描述文本包括:  
从所述输入图像中提取形状、颜色和文字中的至少之一;以及  
生成所述形状、颜色和文字中的至少之一的描述文本,作为所述第二特征描述文本。
5. 根据权利要求1所述的方法,其中,所述从目标区域图像中确定多个待匹配图书各自的子图像包括:  
从所述目标区域图像中识别所述多个待匹配图书;以及  
从所述目标区域图像中截取所述多个待匹配图书各自的子图像。
6. 根据权利要求1所述的方法,还包括:  
确定所述目标区域图像中的多个待匹配图书各自的位置信息;  
在所述目标区域图像中标记所述目标图书的位置,得到标记图像;以及  
输出所述标记图像。
7. 一种基于大模型技术的书籍检索装置,包括:  
描述文本确定模块,用于根据接收到的输入信息,生成待确定图书的图书描述文本;  
图像确定模块,用于从目标区域图像中确定多个待匹配图书各自的子图像;以及  
目标图书确定模块,用于根据所述待确定图书的图书描述文本以及所述多个待匹配图书各自的子图像,从所述多个待匹配图书中确定目标图书。
8. 根据权利要求7所述的装置,其中,所述输入信息包括输入语音、输入文本和输入图像中的至少之一;所述图书描述文本确定模块包括:  
第一特征描述文本确定单元,用于从所述输入语音和所述输入文本中的至少之一中提取所述待确定图书的第一特征描述文本;  
第二特征描述文本确定单元,用于从所述输入图像中提取所述待确定图书的第二特征

描述文本;以及

图书描述文本确定单元,用于根据所述第一特征描述文本和所述第二特征描述文本中的至少之一,确定所述图书描述文本。

9. 根据权利要求8所述的装置,其中,所述第一特征描述文本确定单元包括:

转换子单元,用于将所述输入语音转换为转换文本;以及

第一提取子单元,用于从所述转换文本和所述输入文本中的至少之一中提取包含所述待确定图书的名称、作者、版本信息和出版信息中的至少之一的描述文本,作为所述第一特征描述文本。

10. 根据权利要求8所述的装置,其中,所述第二特征描述文本确定单元包括:

第二提取子单元,用于从所述输入图像中提取形状、颜色和文字中的至少之一;以及

生成子单元,用于生成所述形状、颜色和文字中的至少之一的描述文本,作为所述第二特征描述文本。

11. 根据权利要求6所述的装置,其中,所述子图像确定模块包括:

识别单元,用于从所述目标区域图像中识别所述多个待匹配图书;以及

截取单元,用于从所述目标区域图像中截取所述多个待匹配图书各自的子图像。

12. 根据权利要求6所述的装置,还包括:

位置信息确定模块,用于确定所述目标区域图像中的多个待匹配图书各自的位置信息;

标记图像生成模块,用于在所述目标区域图像中标记所述目标图书的位置,得到标记图像;以及

输出模块,用于输出所述标记图像。

13. 一种电子设备,包括:

至少一个处理器;以及

与所述至少一个处理器通信连接的存储器;其中,

所述存储器存储有可被所述至少一个处理器执行的指令,所述指令被所述至少一个处理器执行,以使所述至少一个处理器能够执行权利要求1至6中任一项所述的方法。

14. 一种存储有计算机指令的非瞬时计算机可读存储介质,其中,所述计算机指令用于使所述计算机执行根据权利要求1至6中任一项所述的方法。

15. 一种计算机程序产品,包括计算机程序,所述计算机程序存储于可读存储介质和电子设备其中至少之一上,所述计算机程序在被处理器执行时实现根据权利要求1至6中任一项所述的方法。

## 基于大模型技术的书籍检索方法、装置、设备和介质

### 技术领域

[0001] 本公开涉及人工智能技术领域,尤其涉及自然语言处理、计算机视觉、大模型、生成模型技术领域。更具体地,本公开提供了一种基于大模型技术的书籍检索方法、装置、电子设备、存储介质以及计算机程序产品。

### 背景技术

[0002] 随着人工智能技术的不断发展,大模型的应用越来越广泛。例如,在面对从大量相似物品中确定人们实际想要查找的对象这一技术问题时,目前基于传统模型的解决方案往往难以准确理解用户的实际需求,从而难以帮助用户查找到目标对象。而大模型的出现,更加方便地解决这个问题。

### 发明内容

[0003] 本公开提供了一种基于大模型技术的书籍检索方法、装置、电子设备、存储介质以及计算机程序产品。

[0004] 根据第一方面,提供了一种基于大模型技术的书籍检索方法,该方法包括:根据接收到的输入信息,生成待确定图书的图书描述文本;从目标区域图像中确定多个待匹配图书各自的子图像;根据待确定图书的图书描述文本以及多个待匹配图书各自的子图像,从多个待匹配图书中确定目标图书。

[0005] 根据第二方面,提供了一种基于大模型技术的书籍检索装置,该装置包括:描述文本确定模块,用于根据接收到的输入信息,生成待确定图书的图书描述文本;图像确定模块,用于从目标区域图像中确定多个待匹配图书各自的子图像;目标图书确定模块,用于根据待确定图书的图书描述文本以及多个待匹配图书各自的子图像,从多个待匹配图书中确定目标图书。

[0006] 根据第三方面,提供了一种电子设备,包括:至少一个处理器;以及与至少一个处理器通信连接的存储器;其中,存储器存储有可被至少一个处理器执行的指令,指令被至少一个处理器执行,以使至少一个处理器能够执行根据本公开提供的方法。

[0007] 根据第四方面,提供了一种存储有计算机指令的非瞬时计算机可读存储介质,该计算机指令用于使计算机执行根据本公开提供的方法。

[0008] 根据第五方面,提供了一种计算机程序产品,包括计算机程序,所述计算机程序存储于可读存储介质和电子设备其中至少之一上,所述计算机程序在被处理器执行时实现根据本公开提供的方法。

[0009] 应当理解,本部分所描述的内容并非旨在标识本公开的实施例的关键或重要特征,也不用于限制本公开的范围。本公开的其它特征将通过以下的说明书而变得容易理解。

### 附图说明

[0010] 附图用于更好地理解本方案,不构成对本公开的限定。其中:

[0011] 图1是根据本公开的一个实施例的可以应用基于大模型技术的书籍检索方法和装置的示例性系统架构示意图；

[0012] 图2是根据本公开的一个实施例的基于大模型技术的书籍检索方法的流程图；

[0013] 图3是根据本公开的另一个实施例的基于大模型技术的书籍检索方法的流程图；

[0014] 图4是根据本公开的一个实施例的无服务器架构的服务调用示意图；

[0015] 图5是根据本公开的一个实施例的基于大模型技术的书籍检索装置的框图；以及

[0016] 图6是根据本公开的一个实施例的基于大模型技术的书籍检索方法的电子设备的框图。

### 具体实施方式

[0017] 以下结合附图对本公开的示范性实施例做出说明,其中包括本公开实施例的各种细节以助于理解,应当将它们认为仅仅是示范性的。因此,本领域普通技术人员应当认识到,可以对这里描述的实施例做出各种改变和修改,而不会背离本公开的范围和精神。同样,为了清楚和简明,以下的描述中省略了对公知功能和结构的描述。

[0018] 在从放置大量图书的书架中确定人们想要查找的图书这一场景下,传统的方案大多是基于人工规则。即使引入人工智能模型也是传统的分类模型、检测模型等,传统模型的视觉能力和语言能力都十分薄弱,不具备智能性,因此难以满足人们的需求。

[0019] 本公开的技术方案中,所涉及的用户个人信息的收集、存储、使用、加工、传输、提供和公开等处理,均符合相关法律法规的规定,且不违背公序良俗。

[0020] 在本公开的技术方案中,在获取或采集用户个人信息之前,均获取了用户的授权或同意。

[0021] 本公开的实施例提供了一种基于大模型技术的书籍检索方法,通过大模型可以直接和用户通过自然语言对话,并理解用户实际需求,并且大模型具有图像理解能力,能够准确定位出用户想要查找的图书的位置。

[0022] 用户查找图书的需求具有明显的波峰和波谷。例如,在周末和节假日,该需求可能增多,而在工作日该需求可能较少。因此,上述查找图书的需求的波谷时间可能更长,部署专用的服务器可能会造成资源浪费。

[0023] 有鉴于此,本公开的实施例采用Serverless(无服务器)的系统架构实现用户的这种需求。Serverless即无服务器运算,提供一个微型的架构,终端客户不需要部署、配置或管理服务器服务,代码运行所需要的服务器皆由云端平台来提供。

[0024] 本公开的实施例的Serverless系统架构指用于实现本公开的实施例的基于大模型技术的书籍检索方法的至少部分服务不设置专用的服务器,该至少部分服务可以由云端平台来实现。在波峰阶段,可以开通云端平台的多个服务器来实现该至少部分服务。在波谷阶段,可以开通云端平台的少量的服务器来实现该至少部分服务。

[0025] 图1是根据本公开一个实施例的可以应用基于大模型技术的书籍检索方法和装置的示例性系统架构示意图。需要注意的是,图1所示仅为可以应用本公开实施例的系统架构的示例,以帮助本领域技术人员理解本公开的技术内容,但并不意味着本公开实施例不可以用于其他设备、系统、环境或场景。

[0026] 如图1所示,根据该实施例的系统架构100可以包括终端设备101、机器人102、网络

103、服务器104、105、106。网络103用以在终端设备101、机器人102与服务器104、105、106之间提供通信链路的介质。网络103可以包括各种连接类型,例如有线和/或无线通信链路等等。

[0027] 例如,终端设备101可以是手机、平板、笔记本等具有显示屏的电子设备,用户可以在显示屏所展示的页面中输入语音、文本或者上传图像。此外,终端设备101还可以设置有摄像头,用户利用摄像头可以采集书架图像。终端设备101可以调用大模型基于用户的输入信息以及采集的书架图像进行处理,从书架图像中确定用户实际想到查找的目标图书。

[0028] 又例如,机器人102具有显示屏,用户可以在显示屏展示的页面中输入语音、文本或者上传图像。此外,机器人102可以设置有摄像头,机器人102可以利用摄像头采集书架图像。然后调用大模型基于用户的输入信息以及采集的书架图像,从书架图像中确定用户实际要查找的目标图书。

[0029] 本公开的实施例通过调用大模型基于用户的输入信息以及采集的书架图像,从书架图像中确定用户实际要查找的目标图书,能够提高图书识别准确性,满足用户查找图书的需求,提高用户体验。

[0030] 图2是根据本公开的一个实施例的基于大模型技术的书籍检索方法的流程图。

[0031] 如图2所示,该基于大模型技术的书籍检索方法200包括操作S210~操作S230。

[0032] 在操作S210,根据接收到的输入信息,生成待确定图书的图书描述文本。

[0033] 用户在使用终端查找图书时,用户可以通过终端输入语音、文本或者上传图像。用户在使用机器人查找图书时,可以通过机器人的显示屏输入语音和文本,以及通过机器人的蓝牙或连接线上传图像。

[0034] 针对输入语音和输入文本,输入语音和输入文本中可以包含书名,或者是对图书的一段描述。因此,可以从输入语音和输入文本中提取待确定图书的描述信息,如图书的名称、作者、类型等,类型例如包括教科书、散文、诗集、小说等。针对上传的图像,可以从图像中提取出图书的名称、作者、颜色、出版信息等,可以将从图像中提取的信息生成描述文本。

[0035] 可以将从输入语音、输入文本中提取的描述信息以及从输入图像中提取的描述信息相结合,得到待确定图书的图书描述文本。

[0036] 在操作S220,从目标区域图像中确定多个待匹配图书各自的子图像。

[0037] 目标区域例如可以是书架。

[0038] 在用户使用终端查找图书的场景下,用户可以携带终端在书架前控制终端开启摄像头采集书架图像。终端可以调用相应的图像处理服务,将采集的书架图像传入该图像处理服务,使图像处理服务从书架图像中截取出各个图书的子图像。

[0039] 在用户使用机器人查找图书的场景下,机器人在确定用户有查找图书的需求的情况下,例如接收到用户的输入信息之后,可以移动到书架前,开启摄像头采集书架图像。机器人还可以调用图像处理服务,将采集的书架图像传入图像处理服务,使图像处理服务从书架图像中截取出每个图书的子图像。

[0040] 书架图像中的每本书均可以作为待匹配图书,图像处理服务可以对书架图像进行目标检测,检测出每本书的位置,根据位置可以从书架图像中截取出每本书的子图像,作为待匹配图书的子图像。

[0041] 在操作S230,根据待确定图书的图书描述文本以及多个待匹配图书各自的子图

像,从多个待匹配图书中确定目标图书。

[0042] 例如,可以调用多模态大模型,将图书描述文本和各个待匹配图书的子图像输入大模型,大模型通过对图书描述文本和各个子图像进行处理,从多个子图像中确定与图书描述文本相匹配的子图像,该子图像中的待匹配图书即为目标图书。

[0043] 大模型可以将目标图书的子图像返回给终端或机器人,终端或机器人可以根据该子图像在书架图像中的位置,在书架图像中标记出该目标图书的位置,并展示给用户,从而使用户快速确认要查找的图书。此外,机器人还可以根据目标图书的位置信息发出激光等射线指出该目标图像。

[0044] 本公开的实施例通过根据用户的输入信息,生成待确定图书的图书描述文本,从书架图像中截取多个待确定图书各自的子图像,基于图书描述文本和待确定图书的子图像,从待确定图书中确定目标图书,能够准确理解用户需求,提高图书识别准确率,满足用户图书检索需求,提升用户体验。

[0045] 图3是根据本公开的另一个实施例的基于大模型技术的书籍检索方法的流程图。

[0046] 如图3所示,本实施例包括操作S311~操作S314、操作S321~操作S323、以及操作S331~操作S332。操作S311~操作S314是针对用户输入信息进行处理步骤。操作S321~操作S323是针对书架图像进行处理步骤。操作S331~操作S332是调用大模型服务进行处理步骤。

[0047] 在操作S311,接收输入信息,输入信息包括输入语音、输入文本和输入图像中的至少之一。

[0048] 在操作S312,将输入语音转换为转换文本,从转换文本和输入文本中的至少之一中提取第一特征描述文本。

[0049] 根据本公开的实施例,将输入语音转换为转换文本;以及从转换文本和输入文本中的至少之一中提取包含待确定图书的名称、作者、版本信息和出版信息中的至少之一的描述文本,作为第一特征描述文本。

[0050] 例如,可以将由输入语音转换得到的转换文本与输入文本进行组合,得到组合文本。接下来可以调用大语言模型对组合文本进行总结和提炼,从组合文本中提取出用户要查找的图书的名称、作者、类型、版本信息、出版信息等。类型例如包括教科书、散文、诗集、小说等。从组合文本中提取出这些特征的描述文本作为第一特征描述文本。

[0051] 在操作S313,从输入图像中提取待确定图书的第二特征描述文本。

[0052] 根据本公开的实施例,从输入图像中提取形状、颜色和文字中的至少之一;以及生成形状、颜色和文字中的至少之一的描述文本,作为第二特征描述文本。

[0053] 例如,可以调用多模态大模型从输入图像中提取图书的形状、颜色、文字等,将形状、颜色用文本描述出来,再将形状、颜色的描述文本与图像中的文字组合,得到第二特征描述文本。

[0054] 在操作S314,根据第一特征描述文本和第二特征描述文本中的至少之一,确定图书描述文本。

[0055] 例如,可以将第一特征描述文本和第二特征描述文本进行整合,可以合并相同的特征,得到图书描述文本。

[0056] 在操作S321,采集书架图像。

[0057] 在操作S322,从书架图像中识别多个待匹配图书以及每个待匹配图书的位置。

[0058] 在操作S323,从书架图像中截取多个待匹配图书各自的子图像。

[0059] 例如,用户携带终端在书架前采集书架图像或者机器人移动到书架前采集书架图像。终端或者机器人能够调用图像处理服务,将采集的书架图像传入图像处理服务,图像处理服务能够从书架图像中识别出每个图书的位置,并从书架图像中截取出每个图书的子图像。

[0060] 在操作S331,根据图书描述文本以及多个待匹配图书各自的子图像,从多个待匹配图书中确定目标图书。

[0061] 例如,可以调用多模态大模型,将图书描述文本和多个待匹配图书各自的子图像输入多模态大模型,多模态大模型通过对图书描述文本和各个子图像进行处理,从多个子图像中确定与图书描述文本相匹配的子图像,该子图像中的图书即为目标图书。

[0062] 在一个示例中,颜色作为图书的重要特征,在用户指定要查找的图书的颜色时,可以先过滤掉不是指定颜色图书的子图像,再利用大模型进行匹配,可以提升速率。

[0063] 在操作S332,在书架图像中标记目标图书的位置,得到标记图像,输出标记图像。

[0064] 由于每一本书的子图像不能表征该图书在书架的位置,因此,为了让用户找到目标图书,可以根据该目标图书在书架图像中的位置,在书架图像中标记出该图书,得到标记图像。终端或者机器人可以展示上述标记图像,使用户快速确认要查找的图书。

[0065] 在一个示例中,用户可以反馈目标图书是否是实际要查找的图书,将该反馈信息输入给大模型,使大模型进行微调,提高大模型识别图书的准确性,使得大模型查找图书的准确率越来越高。

[0066] 本公开的实施例通过调用大模型从输入语音和输入文本中提取待确定图书的第一特征描述文本以及从输入图像中提取待确定图书的第二特征描述文本,根据第一特征描述文本和第二特征描述文本生成待确定图书的图书描述文本,从书架图像中截取出多个待匹配图书各自的子图像,再调用大模型基于图书描述文本和待匹配图书的子图像,从待匹配图书中确定目标图书,能够提高图书识别准确率。

[0067] 图4是根据本公开的一个实施例的无服务器架构的服务调用示意图。

[0068] 本公开的实施例可以采用Serverless无服务器架构实现基于大模型技术的书籍检索方法。如图4所示,Serverless无服务器架构提供接入服务40,用来调用各个其他服务来处理用户的请求,其他服务可以包括语音转换服务41、文字特征提取服务42、图像特征提取服务43、检索服务44以及图像处理服务45等。

[0069] 例如,用户通过终端或者机器人的显示页面输入语音、文本或者上传图像,终端可以通过发起http请求将用户输入的语音、文本或图像传入接入服务40。针对输入语音,接入服务40可以调用语音转换服务41来将输入语音转换为转换文本。针对转换文本和用户输入的文本,接入服务40可以调用文字特征提取服务42进行处理,文字特征提取服务42具体可以调用大模型服务431针对转换文本和输入文本进行特征提取,得到第一特征描述文本。例如,大模型服务431可以包括大语言模型,大模型服务431可以调用大语言模型针对转换文本和输入文本进行总结,得到第一特征描述文本。针对输入图像,接入服务40可以调用图像特征提取服务43进行处理,图像特征提取服务43具体可以调用大模型服务431对输入图像进行特征提取,得到第二特征描述文本。例如,大模型服务431可以包括多模态大模型,大模



型服务431可以调用多模态大模型对输入图像进行特征提取,并生成所提取图像特征的描述文本,即第二特征描述文本。

[0070] 例如,终端或者机器人还可以通过发起http请求将摄像头采集的书架图像传入接入服务40。接入服务40调用图像处理服务45对书架图像进行处理。图像处理服务45具体可以调用目标检测服务451来进行处理,例如,目标检测服务451可以利用目标检测模型从书架图像中识别出待匹配图书,并从书架图像中截取出每个待匹配图书的子图像。目标检测模型例如可以是YOLO模型。图像处理服务45还可以调用缓存服务452,将目标检测服务451截取出的各个待匹配图书的子图像进行缓存。

[0071] 检索服务44可以从缓存服务452中获取各个待匹配图书的子图像,然后调用大模型服务431根据图书描述文本和各个待匹配图书的子图像,从多个待匹配图书中确定目标图书。

[0072] 检索服44还可以将目标图书的子图像以及位置信息返回给接入服务40,接入服务40可以在书架图像中重新标记出目标图书的位置,得到标记图像,并将标记图像返回给终端。终端向用户展示标记图像,可以使用户直观地观察到要查找的图书的位置。

[0073] 本公开的实施例采用Serverless无服务器架构,无需设置专用的服务器,通过将用户输入信息以及摄像头采集的书架图像传入接入服务,由接入服务调用各个服务来实现基于大模型技术的书籍检索方法,能够减少资源占用,节约成本。

[0074] 根据本公开的实施例,本公开还提供了一种基于大模型技术的书籍检索装置。

[0075] 图5是根据本公开的一个实施例的基于大模型技术的书籍检索装置的框图。

[0076] 如图5所示,该基于大模型技术的书籍检索装置500包括描述文本确定模块510、图像确定模块520以及目标图书确定模块530。

[0077] 描述文本确定模块510用于根据接收到的输入信息,生成待确定图书的图书描述文本。

[0078] 图像确定模块520用于从目标区域图像中确定多个待匹配图书各自的子图像。

[0079] 目标图书确定模块530用于根据待确定图书的图书描述文本以及多个待匹配图书各自的子图像,从多个待匹配图书中确定目标图书。

[0080] 根据本公开的实施例,输入信息包括输入语音、输入文本和输入图像中的至少之一。描述文本确定模块510包括第一特征描述文本确定单元、第二特征描述文本确定单元和图书描述文本确定单元。

[0081] 第一特征描述文本确定单元用于从输入语音和输入文本中的至少之一中提取待确定图书的第一特征描述文本。

[0082] 第二特征描述文本确定单元用于从输入图像中提取待确定图书的第二特征描述文本。

[0083] 图书描述文本确定单元用于根据第一特征描述文本和第二特征描述文本中的至少之一,确定图书描述文本。

[0084] 第一特征描述文本确定单元包括转换子单元和第一提取子单元。

[0085] 转换子单元用于将输入语音转换为转换文本。

[0086] 第一提取子单元用于从转换文本和输入文本中的至少之一中提取包含待确定图书的名称、作者、版本信息和出版信息中的至少之一的描述文本,作为第一特征描述文本。

[0087] 第二特征描述文本确定单元包括第二提取子单元和生成子单元。

[0088] 第二提取子单元用于从输入图像中提取形状、颜色和文字中的至少之一。

[0089] 生成子单元用于生成形状、颜色和文字中的至少之一的描述文本,作为第二特征描述文本。

[0090] 图像确定模块520包括识别单元和截取单元。

[0091] 识别单元用于从目标区域图像中识别多个待匹配图书。

[0092] 截取单元用于从目标区域图像中截取多个待匹配图书各自的子图像。

[0093] 基于大模型技术的书籍检索装置500还包括位置信息确定模块、标记图像生成模块以及输出模块。

[0094] 位置信息确定模块用于确定目标区域图像中的多个待匹配图书各自的位置信息。

[0095] 标记图像生成模块用于在目标区域图像中标记目标图书的位置,得到标记图像。

[0096] 输出模块用于输出标记图像。

[0097] 根据本公开的实施例,本公开还提供了一种电子设备、一种可读存储介质和一种计算机程序产品。

[0098] 图6示出了可以用来实施本公开的实施例的示例电子设备600的示意性框图。电子设备旨在表示各种形式的数字计算机,诸如,膝上型计算机、台式计算机、工作台、个人数字助理、服务器、刀片式服务器、大型计算机、和其它适合的计算机。电子设备还可以表示各种形式的移动装置,诸如,个人数字处理、蜂窝电话、智能电话、可穿戴设备和其它类似的计算装置。本文所示的部件、它们的连接和关系、以及它们的功能仅仅作为示例,并且不意在限制本文中描述的和/或者要求的本公开的实现。

[0099] 如图6所示,设备600包括计算单元601,其可以根据存储在只读存储器(ROM) 602中的计算机程序或者从存储单元608加载到随机访问存储器(RAM) 603中的计算机程序,来执行各种适当的动作和处理。在RAM 603中,还可存储设备600操作所需的各种程序和数据。计算单元601、ROM 602以及RAM 603通过总线604彼此相连。输入/输出(I/O)接口605也连接至总线604。

[0100] 设备600中的多个部件连接至I/O接口605,包括:输入单元606,例如键盘、鼠标等;输出单元607,例如各种类型的显示器、扬声器等;存储单元608,例如磁盘、光盘等;以及通信单元609,例如网卡、调制解调器、无线通信收发机等。通信单元609允许设备600通过诸如因特网的计算机网络和/或各种电信网络与其他设备交换信息/数据。

[0101] 计算单元601可以是各种具有处理和计算能力的通用和/或专用处理组件。计算单元601的一些示例包括但不限于中央处理单元(CPU)、图形处理单元(GPU)、各种专用的人工智能(AI)计算芯片、各种运行机器学习模型算法的计算单元、数字信号处理器(DSP)、以及任何适当的处理器、控制器、微控制器等。计算单元601执行上文所描述的各个方法和处理,例如基于大模型技术的书籍检索方法。例如,在一些实施例中,基于大模型技术的书籍检索方法可被实现为计算机软件程序,其被有形地包含于机器可读介质,例如存储单元608。在一些实施例中,计算机程序的部分或者全部可以经由ROM 602和/或通信单元609而被载入和/或安装到设备600上。当计算机程序加载到RAM 603并由计算单元601执行时,可以执行上文描述的基于大模型技术的书籍检索方法的一个或多个步骤。备选地,在其他实施例中,计算单元601可以通过其他任何适当的方式(例如,借助于固件)而被配置为执行基于大模

型技术的书籍检索方法。

[0102] 本文中以上描述的系统和技术各种实施方式可以在数字电子电路系统、集成电路系统、场可编程门阵列 (FPGA)、专用集成电路 (ASIC)、专用标准产品 (ASSP)、芯片上系统的系统 (SOC)、复杂可编程逻辑设备 (CPLD)、计算机硬件、固件、软件、和/或它们的组合中实现。这些各种实施方式可以包括:实施在一个或者多个计算机程序中,该一个或者多个计算机程序可在包括至少一个可编程处理器的可编程系统上执行和/或解释,该可编程处理器可以是专用或者通用可编程处理器,可以从存储系统、至少一个输入装置、和至少一个输出装置接收数据和指令,并且将数据和指令传输至该存储系统、该至少一个输入装置、和该至少一个输出装置。

[0103] 用于实施本公开的方法的程序代码可以采用一个或多个编程语言的任何组合来编写。这些程序代码可以提供给通用计算机、专用计算机或其他可编程数据处理装置的处理器或控制器,使得程序代码当由处理器或控制器执行时使流程图和/或框图所示的功能/操作被实施。程序代码可以完全在机器上执行、部分地在机器上执行,作为独立软件包部分地在机器上执行且部分地在远程机器上执行或完全在远程机器或服务器上执行。

[0104] 在本公开的上下文中,机器可读介质可以是有形的介质,其可以包含或存储以供指令执行系统、装置或设备使用或与指令执行系统、装置或设备结合地使用的程序。机器可读介质可以是机器可读信号介质或机器可读储存介质。机器可读介质可以包括但不限于电子的、磁性的、光学的、电磁的、红外的、或半导体系统、装置或设备,或者上述内容的任何合适组合。机器可读存储介质的更具体示例会包括基于一个或多个线的电气连接、便携式计算机盘、硬盘、随机存取存储器 (RAM)、只读存储器 (ROM)、可擦除可编程只读存储器 (EPROM 或快闪存储器)、光纤、便捷式紧凑盘只读存储器 (CD-ROM)、光学储存设备、磁储存设备、或上述内容的任何合适组合。

[0105] 为了提供与用户的交互,可以在计算机上实施此处描述的系统和技术,该计算机具有:用于向用户显示信息的显示装置 (例如,CRT (阴极射线管) 或者LCD (液晶显示器) 监视器);以及键盘和指向装置 (例如,鼠标或者轨迹球),用户可以通过该键盘和该指向装置来将输入提供给计算机。其它种类的装置还可以用于提供与用户的交互;例如,提供给用户的反馈可以是任何形式的传感反馈 (例如,视觉反馈、听觉反馈、或者触觉反馈);并且可以用任何形式 (包括声输入、语音输入或者、触觉输入) 来接收来自用户的输入。

[0106] 可以将此处描述的系统和技术实施在包括后台部件的计算系统 (例如,作为数据服务器)、或者包括中间件部件的计算系统 (例如,应用服务器)、或者包括前端部件的计算系统 (例如,具有图形用户界面或者网络浏览器的用户计算机,用户可以通过该图形用户界面或者该网络浏览器来与此处描述的系统和技术实施方式交互)、或者包括这种后台部件、中间件部件、或者前端部件的任何组合的计算系统中。可以通过任何形式或者介质的数字数据通信 (例如,通信网络) 来将系统的部件相互连接。通信网络的示例包括:局域网 (LAN)、广域网 (WAN) 和互联网。

[0107] 计算机系统可以包括客户端和服务端。客户端和服务端一般远离彼此并且通常通过通信网络进行交互。通过在相应的计算机上运行并且彼此具有客户端-服务器关系的计算机程序来产生客户端和服务端的关系。

[0108] 应该理解,可以使用上面所示的各种形式的流程,重新排序、增加或删除步骤。例

如,本公开中记载的各步骤可以并行地执行也可以顺序地执行也可以不同的次序执行,只要能够实现本公开公开的技术方案所期望的结果,本文在此不进行限制。

[0109] 上述具体实施方式,并不构成对本公开保护范围的限制。本领域技术人员应该明白的是,根据设计要求和因素,可以进行各种修改、组合、子组合和替代。任何在本公开的精神和原则之内所作的修改、等同替换和改进等,均应包含在本公开保护范围之内。

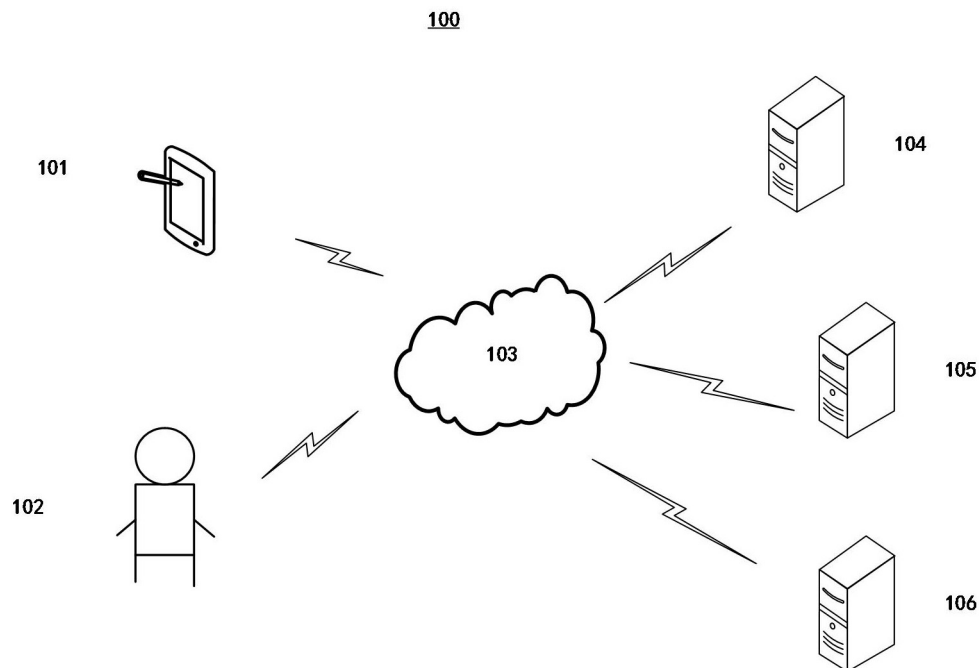


图1

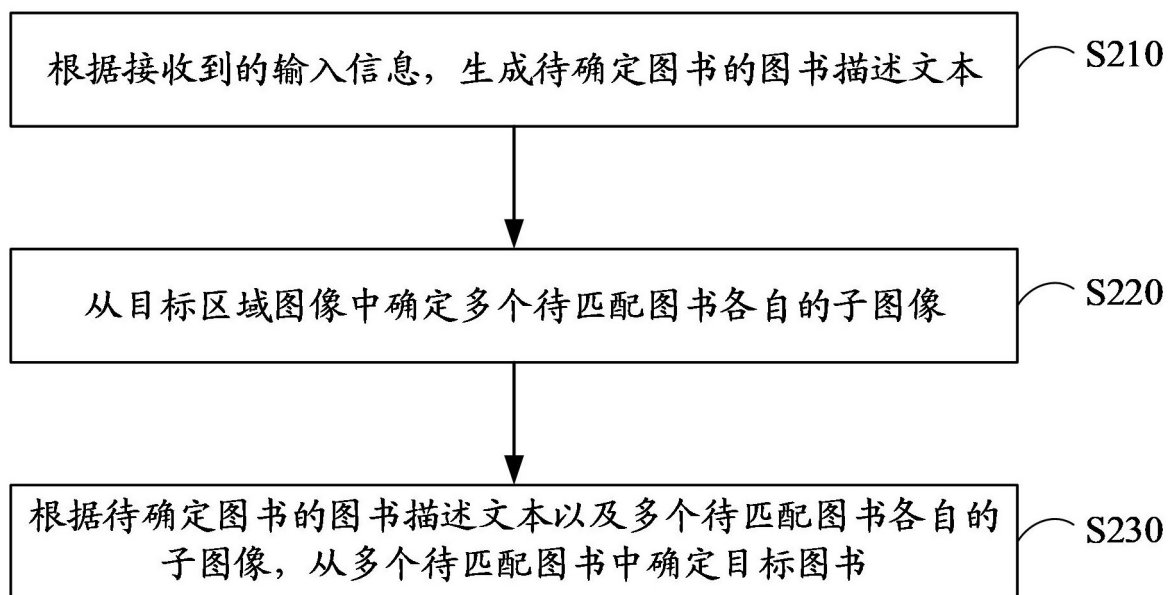
200

图2

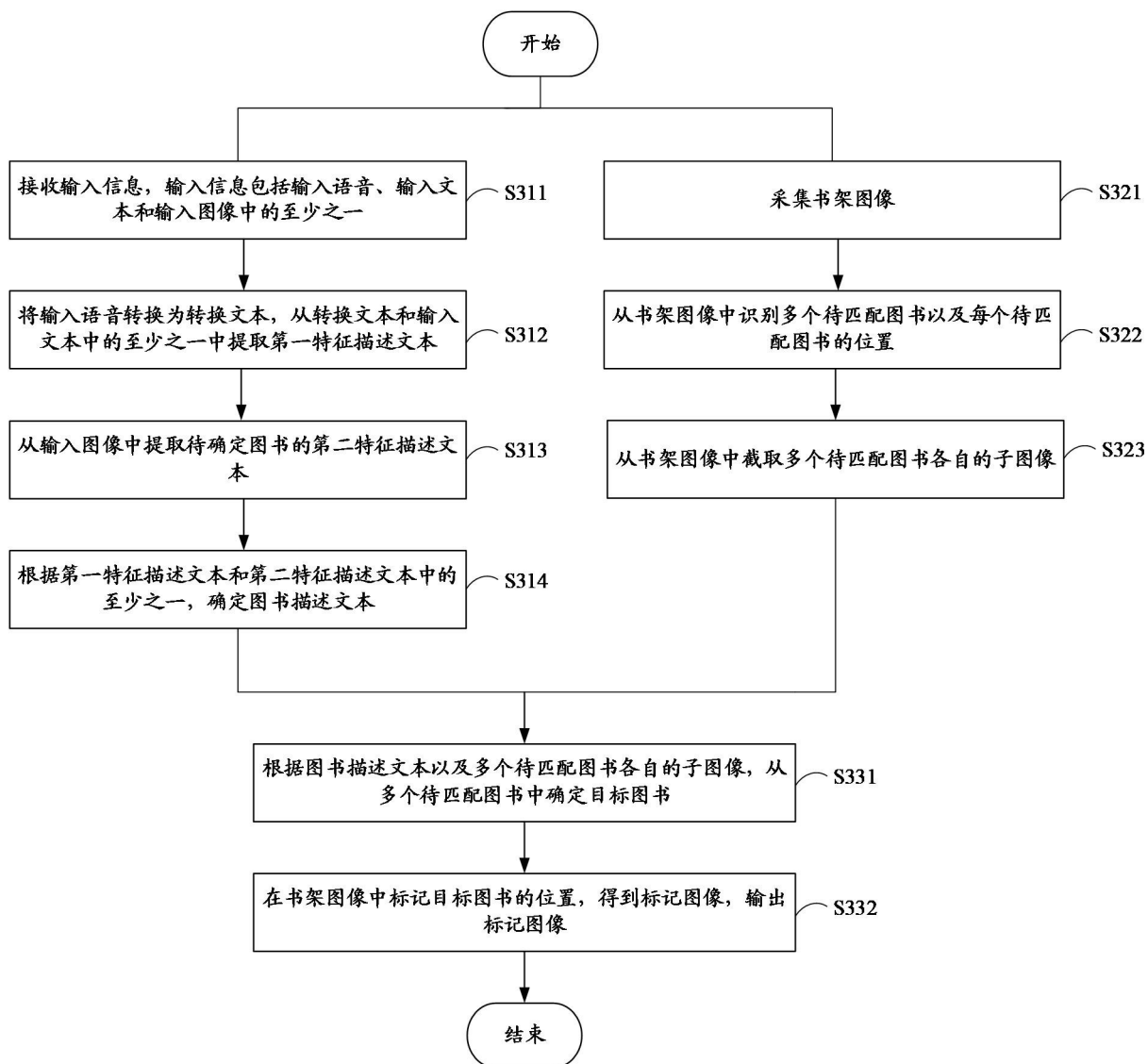


图3

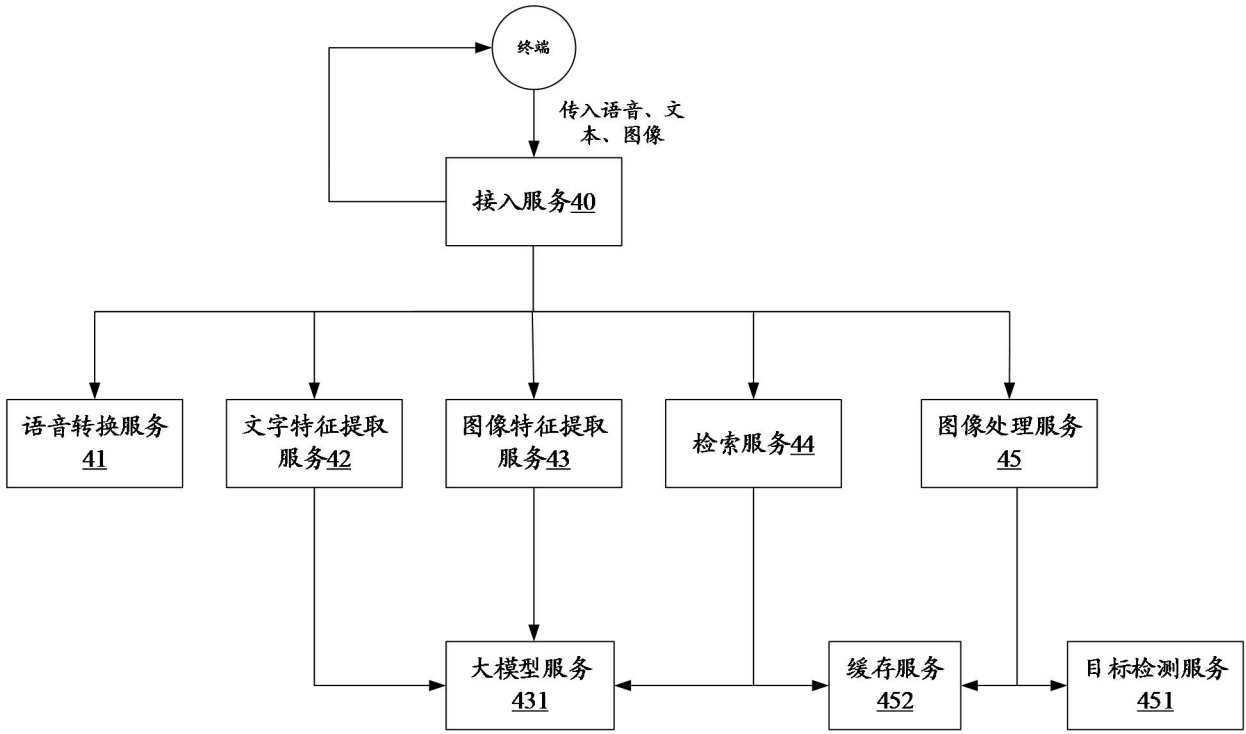


图4

500

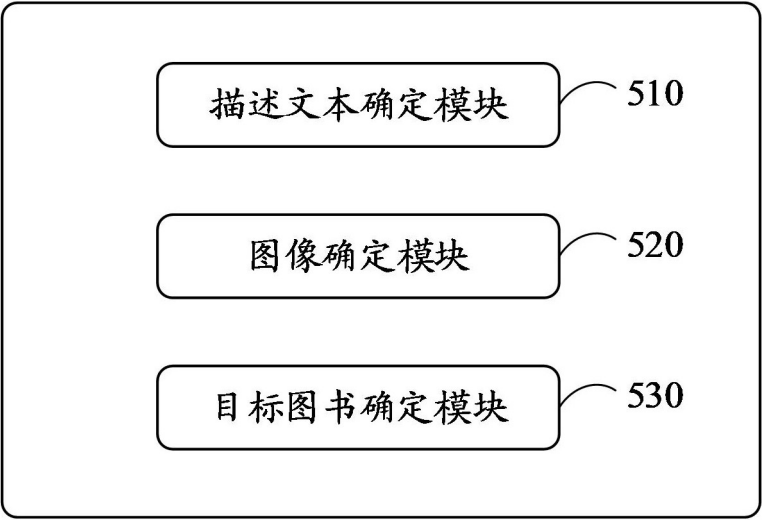


图5

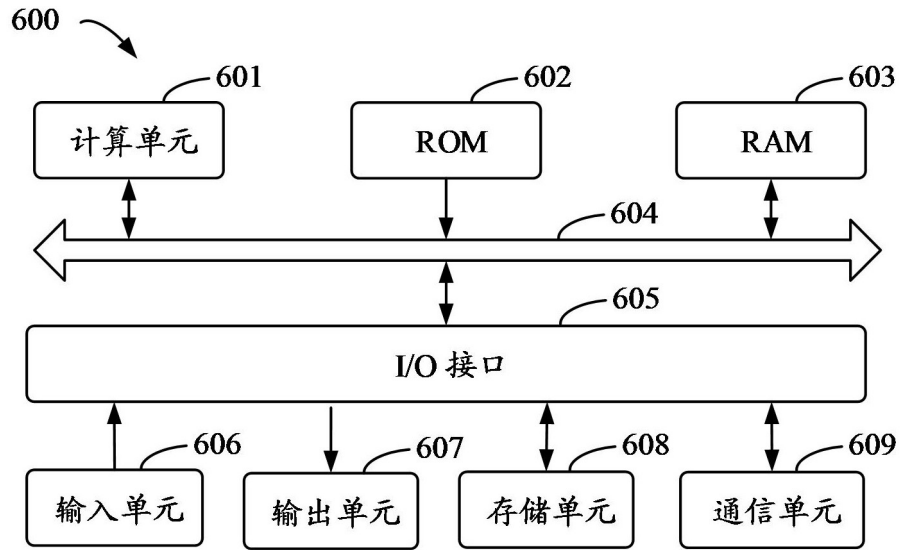


图6