



(12) 发明专利申请

(10) 申请公布号 CN 116542221 A

(43) 申请公布日 2023. 08. 04

(21) 申请号 202310369133.5

(22) 申请日 2023.04.03

(71) 申请人 康键信息技术(深圳)有限公司

地址 518000 广东省深圳市前海深港合作区前湾一路1号A栋201室(入驻深圳市前海商务秘书有限公司)

(72) 发明人 顾柏进

(74) 专利代理机构 深圳市沃德知识产权代理事

务所(普通合伙) 44347

专利代理师 范腊梅

(51) Int. Cl.

G06F 40/106 (2020.01)

G06F 40/151 (2020.01)

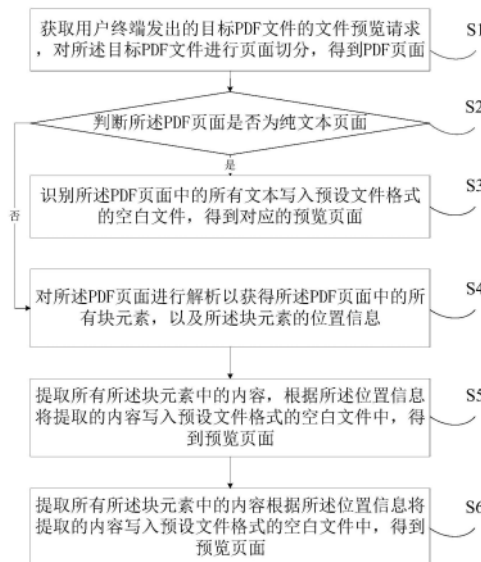
权利要求书2页 说明书9页 附图2页

(54) 发明名称

PDF文件解析预览方法、装置、设备及存储介质

(57) 摘要

本发明涉及开发辅助领域技术,揭露一种用于医疗诊断报告预览的PDF文件解析预览方法,包括:对目标PDF文件进行页面切分,得到PDF页面;当PDF页面为纯文本页面,识别PDF页面中的所有文本写入预设文件格式的空白文件,得到对应的预览页面;当PDF页面不是纯文本页面,对PDF页面进行解析以获得PDF页面中的所有块元素,以及块元素的位置信息;提取所有块元素中的内容根据位置信息将提取的内容写入预设文件格式的空白文件中,得到预览页面;将预览页面发送至用户终端,以实现目标PDF文件的预览。本发明还涉及一种区块链技术,所述预览页面可以存储在区块链节点中。本发明还提出一种PDF文件解析预览装置、电子设备以及介质。本发明可以提高PDF文件解析预览的效率。



1. 一种PDF文件解析预览方法,其特征在于,所述方法包括:
 - 获取用户终端发出的目标PDF文件的文件预览请求,对所述目标PDF文件进行页面切分,得到PDF页面;
 - 判断所述PDF页面是否为纯文本页面;
 - 当所述PDF页面为纯文本页面,识别所述PDF页面中的所有文本写入预设文件格式的空白文件,得到对应的预览页面;
 - 当所述PDF页面不是纯文本页面,对所述PDF页面进行解析以获得所述PDF页面中的所有块元素,以及所述块元素的位置信息;
 - 提取所有所述块元素中的内容,根据所述位置信息将提取的内容写入预设文件格式的空白文件中,得到预览页面;
 - 将所有所述预览页面发送至所述用户终端,以通过所述用户终端查看预览页面实现对所述目标PDF文件的预览。
2. 如权利要求1所述的PDF文件解析预览方法,其特征在于,所述判断所述PDF页面是否为纯文本页面,包括:
 - 将所述PDF页面转化为灰度图像;
 - 利用预构建的页面分类模型中的特征提取网络对所述灰度图像进行特征提取,得到特征提取矩阵;
 - 利用所述页面分类的模型中的特征映射层将所述特征提取矩阵映射为不同预设识别类别的特征值;
 - 将每个所述特征值进行归一化,得到每个识别类别的识别概率;
 - 根据所述识别概率判断所述PDF页面是否为纯文本页面。
3. 如权利要求2所述的PDF文件解析预览方法,其特征在于,所述根据所述识别概率判断所述PDF页面是否为纯文本页面,包括:
 - 判断是否存在大于预设识别阈值的识别概率;
 - 若存在大于所述识别阈值的识别概率,则判断最大识别概率对应的识别类别是否为纯文本;
 - 当最大识别概率对应的识别类别为纯文本,则所述PDF页面为纯文本页面;
 - 当最大识别概率对应的识别类别不为纯文本,则所述PDF页面不为纯文本页面;
 - 若不存在大于所述识别阈值的识别概率,则所述PDF页面不为纯文本页面。
4. 如权利要求1中所述的PDF文件解析预览方法,其特征在于,所述识别所述PDF页面中的所有文本写入预设文件格式的空白文件,得到对应的预览页面,包括:
 - 利用OCR技术识别所述PDF页面中的所有文本,得到页面文本;
 - 将所述页面文本中的所有字符转化为预设的字体后写入所述空白文件,得到所述PDF页面对应的预览页面。
5. 如权利要求1所述的PDF文件解析预览方法,其特征在于,所述将所有所述预览页面发送至所述用户终端,包括:
 - 将所有所述预览页面按照对应的PDF页面在所述目标PDF文件中的先后顺序进行组合,得到预览文件;
 - 将所述预览文件发送至所述用户终端。

6. 如权利要求1至5中任意一项中所述的PDF文件解析预览方法,其特征在于,所述将所有所述预览页面发送至所述用户终端,包括:

将所有所述预览页面按照对应的PDF页面在所述目标PDF文件中的先后顺序进行组合,得到预览文件;

将所述预览文件存储在预设存储区域,并获取所述预览文件的存储地址;将所述存储地址发送至所述用户终端。

7. 一种PDF文件解析预览装置,其特征在于,包括:

分类判断模块,用于获取用户终端发出的目标PDF文件的文件预览请求,对所述目标PDF文件进行页面切分,得到PDF页面;判断所述PDF页面是否为纯文本页面;

页面解析模块,用于当所述PDF页面为纯文本页面,识别所述PDF页面中的所有文本写入预设文件格式的空白文件,得到对应的预览页面;当所述PDF页面不是纯文本页面,对所述PDF页面进行解析以获得所述PDF页面中的所有块元素,以及所述块元素的位置信息;提取所有所述块元素中的内容,根据所述位置信息将提取的内容写入预设文件格式的空白文件中,得到预览页面;

文件发送模块,用于将所有所述预览页面发送至所述用户终端,以通过所述用户终端查看预览页面实现对所述目标PDF文件的预览。

8. 如权利要求7所述的PDF文件解析预览装置,其特征在于,所述将所有所述预览页面发送至所述用户终端,包括:

将所有所述预览页面按照对应的PDF页面在所述目标PDF文件中的先后顺序进行组合,得到预览文件;

将所述预览文件存储在预设存储区域,并获取所述预览文件的存储地址;将所述存储地址发送至所述用户终端。

9. 一种电子设备,其特征在于,所述电子设备包括:

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计算机程序,所述计算机程序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如权利要求1至6中任一项所述的PDF文件解析预览方法。

10. 一种计算机可读存储介质,存储有计算机程序,其特征在于,所述计算机程序被处理器执行时实现如权利要求1至6中任一项所述的PDF文件解析预览方法。

PDF文件解析预览方法、装置、设备及存储介质

技术领域

[0001] 本发明涉及开发辅助技术及数字医疗领域,尤其涉及一种PDF文件解析预览方法、装置、电子设备及存储介质。

背景技术

[0002] PDF文件是一种常用的电子文件格式,当在终端上需要对服务器中的PDF文件进行预览时(如用户对PDF格式医疗诊断报告进行在线预览),需要利用数据分析技术对PDF文件解析,但由于PDF文件查看需要专用的软件,为了可以方便对PDF文件进行预览,需要将PDF文件解析转化为通用的文件格式进行预览。

[0003] 但是,当前的PDF文件解析预览方法,需要对PDF文件的每一页进行解析以解析PDF页面中的内容转化为预设的通用文件格式进行预览,导致PDF文件解析预览的效率较低。

发明内容

[0004] 本发明提供一种PDF文件解析预览方法、装置、电子设备及存储介质,其主要目的在于提高了PDF文件解析预览的效率。

[0005] 获取用户终端发出的目标PDF文件的文件预览请求,对所述目标PDF文件进行页面切分,得到PDF页面;

[0006] 判断所述PDF页面是否为纯文本页面;

[0007] 当所述PDF页面为纯文本页面,识别所述PDF页面中的所有文本写入预设文件格式的空白文件,得到对应的预览页面;

[0008] 当所述PDF页面不是纯文本页面,对所述PDF页面进行解析以获得所述PDF页面中的所有块元素,以及所述块元素的位置信息;

[0009] 提取所有所述块元素中的内容,根据所述位置信息将提取的内容写入预设文件格式的空白文件中,得到预览页面;

[0010] 将所有所述预览页面发送至所述用户终端,以通过所述用户终端查看预览页面实现对所述目标PDF文件的预览。

[0011] 可选地,所述判断所述PDF页面是否为纯文本页面,包括:

[0012] 将所述PDF页面转化为灰度图像;

[0013] 利用预构建的页面分类模型中的特征提取网络对所述灰度图像进行特征提取,得到特征提取矩阵;

[0014] 利用所述页面分类的模型中的特征映射层将所述特征提取矩阵映射为不同预设识别类别的特征值;

[0015] 将每个所述特征值进行归一化,得到每个识别类别的识别概率;

[0016] 根据所述识别概率判断所述PDF页面是否为纯文本页面。

[0017] 可选地,所述根据所述识别概率判断所述PDF页面是否为纯文本页面,包括:

[0018] 判断是否存在大于预设识别阈值的识别概率;

- [0019] 若存在大于所述识别阈值的识别概率,则判断最大识别概率对应的识别类别是否为纯文本;
- [0020] 当最大识别概率对应的识别类别为纯文本,则所述PDF页面为纯文本页面;
- [0021] 当最大识别概率对应的识别类别不为纯文本,则所述PDF页面不为纯文本页面;
- [0022] 若不存在大于所述识别阈值的识别概率,则所述PDF页面不为纯文本页面。
- [0023] 可选地,所述识别所述PDF页面中的所有文本写入预设文件格式的空白文件,得到对应的预览页面,包括:
- [0024] 利用OCR技术识别所述PDF页面中的所有文本,得到页面文本;
- [0025] 将所述页面文本中的所有字符转化为预设的字体后写入所述空白文件,得到所述PDF页面对应的预览页面。
- [0026] 可选地,所述将所有所述预览页面发送至所述用户终端,包括:
- [0027] 将所有所述预览页面按照对应的PDF页面在所述目标PDF文件中的先后顺序进行组合,得到预览文件;
- [0028] 将所述预览文件发送至所述用户终端。
- [0029] 可选地,所述将所有所述预览页面发送至所述用户终端,包括:
- [0030] 将所有所述预览页面按照对应的PDF页面在所述目标PDF文件中的先后顺序进行组合,得到预览文件;
- [0031] 将所述预览文件存储在预设存储区域,并获取所述预览文件的存储地址;将所述存储地址发送至所述用户终端。
- [0032] 为了解决上述问题,本发明还提供一种PDF文件解析预览装置,所述装置包括:
- [0033] 分类判断模块,用于获取用户终端发出的目标PDF文件的文件预览请求,对所述目标PDF文件进行页面切分,得到PDF页面;判断所述PDF页面是否为纯文本页面;
- [0034] 页面解析模块,用于当所述PDF页面为纯文本页面,识别所述PDF页面中的所有文本写入预设文件格式的空白文件,得到对应的预览页面;当所述PDF页面不是纯文本页面,对所述PDF页面进行解析以获得所述PDF页面中的所有块元素,以及所述块元素的位置信息;提取所有所述块元素中的内容,根据所述位置信息将提取的内容写入预设文件格式的空白文件中,得到预览页面;
- [0035] 文件发送模块,用于将所有所述预览页面发送至所述用户终端,以通过所述用户终端查看预览页面实现对所述目标PDF文件的预览。
- [0036] 可选地,所述将所有所述预览页面发送至所述用户终端,包括:
- [0037] 将所有所述预览页面按照对应的PDF页面在所述目标PDF文件中的先后顺序进行组合,得到预览文件;
- [0038] 将所述预览文件存储在预设存储区域,并获取所述预览文件的存储地址;将所述存储地址发送至所述用户终端。
- [0039] 为了解决上述问题,本发明还提供一种电子设备,所述电子设备包括:
- [0040] 存储器,存储至少一个计算机程序;及
- [0041] 处理器,执行所述存储器中存储的计算机程序以实现上述所述的PDF文件解析预览方法。
- [0042] 为了解决上述问题,本发明还提供一种计算机可读存储介质,所述计算机可读存

储介质中存储有至少一个计算机程序,所述至少一个计算机程序被电子设备中的处理器执行以实现上述所述的PDF文件解析预览方法。

[0043] 本发明实施例当所述PDF页面为纯文本页面,识别所述PDF页面中的所有文本写入预设文件格式的空白文件,得到对应的预览页面;当PDF页面为纯文本页面时不需要对页面进行解析,可以通过直接识别页面中的文本实现预览页面的创建,简化了预览页面的创建流程,进而提高了预览页面的创建速度,从而提高了PDF文件解析预览的效率;因此本发明实施例提出的PDF文件解析预览方法、装置、电子设备及可读存储介质降低了PDF文件解析预览的效率。

附图说明

[0044] 图1为本发明一实施例提供的PDF文件解析预览方法的流程示意图;

[0045] 图2为本发明一实施例提供的PDF文件解析预览装置的模块示意图;

[0046] 图3为本发明一实施例提供的实现PDF文件解析预览方法的电子设备的内部结构示意图;

[0047] 本发明目的的实现、功能特点及优点将结合实施例,参照附图做进一步说明。

具体实施方式

[0048] 应当理解,此处所描述的具体实施例仅仅用以解释本发明,并不用于限定本发明。

[0049] 本发明实施例提供一种PDF文件解析预览方法。所述PDF文件解析预览方法的执行主体包括但不限于服务端、终端等能够被配置为执行本申请实施例提供的该方法的电子设备中的至少一种。换言之,所述PDF文件解析预览方法可以由安装在终端设备或服务端设备的软件或硬件来执行,所述软件可以是区块链平台。所述服务端包括但不限于:单台服务器、服务器集群、云端服务器或云端服务器集群等,服务器可以是独立的服务器,也可以是提供云服务、云数据库、云计算、云函数、云存储、网络服务、云通信、中间件服务、域名服务、安全服务、内容分发网络(Content Delivery Network,CDN)、以及大数据和人工智能平台等基础云计算服务的云服务器。

[0050] 参照图1所示的本发明一实施例提供的PDF文件解析预览方法的流程示意图,在本发明实施例中,所述PDF文件解析预览方法包括以下步骤:

[0051] S1、获取用户终端发出的目标PDF文件的文件预览请求,对所述目标PDF文件进行页面切分,得到PDF页面;

[0052] 本发明实施例中预设服务端接收所述用户终端发出的目标PDF文件的文件请求,为了更好的对所述PDF文件进行预览转化,得到对所述目标PDF文件进行页面切分,得到对应的PDF文件,所述目标PDF文件可以为医疗诊断报告。

[0053] 进一步地,本发明实施例中尽管可以将PDF页面通过截图、扫描的方式直接转发为图片的格式,但是往往转化的图片格式无法统一,清晰度也无法满足预览要求,因此,还需要准确识别PDF页面中的内容,并根据识别的内容构建统一格式的。

[0054] S2、判断所述PDF页面是否为纯文本页面;

[0055] 本发明实施例中由于PDF页面中包含的元素类型不同,当所述PDF页面中只有文本是可以直接识别页面中的文本,不需要对PDF页面进行解析,提高了页面中内容的提取速

度。

[0056] 进一步地,本发明实施例中通过机器视觉,利用人工智能模型判断所述PDF页面是否为纯文本页面,当所述PDF页面为纯文本页面时可以直接识别页面中的文本,不需要对PDF页面进行解析,提高了页面中内容的提取速度。

[0057] 具体地,本发明实施例中判断所述PDF页面是否为纯文本页面,包括:

[0058] 将所述PDF页面转化为灰度图像;

[0059] 利用预构建的页面分类模型中的特征提取网络对所述灰度图像进行特征提取,得到特征提取矩阵;

[0060] 利用所述页面分类的模型中的特征映射层将所述特征提取矩阵映射为不同预设识别类别的特征值;

[0061] 将每个所述特征值进行归一化,得到每个识别类别的识别概率;

[0062] 根据所述识别概率判断所述PDF页面是否为纯文本页面。

[0063] 进一步地,本发明实施例中利用预构建的页面分类模型中的特征提取网络对所述灰度图像进行特征提取,得到特征提取矩阵,包括:

[0064] 将所述灰度图像转化为图像矩阵;

[0065] 利用所述特征提取网络对所述图像矩阵进行卷积池化,得到所述特征提取矩阵。

[0066] 具体地,本发明实施例中获取所述灰度图像中所有像素的灰度值,并将所述灰度值作为空白矩阵中的元素构建矩阵,得到所述图像矩阵。

[0067] 进一步地,本发明实施例中所述特征提取层由多层卷积层及多层池化层按照一定的连接顺序串联得到的,本发明实施例对卷积层及池化层的层结构不做限制,所述特征提取层中第一层为卷积层,其余连接顺序不做限制;所述特征映射层由多层全连接层串联构成,其中最后一层全连接层的输出节点的数量与识别类别的数量一致,每个所述识别类别与一个输出节点一一对应,输出节点的输出值即为对应识别类别的特征值。

[0068] 具体地,本发明实施例中所述识别类别包括:纯文本。

[0069] 详细地,本发明实施例中所述根据所述识别概率判断所述PDF页面是否为纯文本页面,包括:

[0070] 判断是否存在大于预设识别阈值的识别概率;

[0071] 若存在大于所述识别阈值的识别概率,则判断最大识别概率对应的识别类别是否为纯文本;

[0072] 当最大识别概率对应的识别类别为纯文本,则所述PDF页面为纯文本页面;

[0073] 当最大识别概率对应的识别类别不为纯文本,则所述PDF页面不为纯文本页面;

[0074] 若不存在大于所述识别阈值的识别概率,则所述PDF页面不为纯文本页面。

[0075] S3、当所述PDF页面为纯文本页面,识别所述PDF页面中的所有文本写入预设文件格式的空白文件,得到对应的预览页面;

[0076] 本发明实施例中当所述PDF页面为纯文本页面时,利用OCR技术识别所述PDF页面中的所有文本,得到页面文本;将所述页面文本中的所有字符转化为预设的字体后写入所述空白文件,得到所述PDF页面对应的预览页面。具体地,本发明实施例中所述预设的字体为所述服务端中存在的字体,

[0077] 进一步地,本发明实施例中所述预设文件格式为所述用户终端可以解析的文件格

式,可选地,所述文件格式可以为jpg、png等文件格式。

[0078] S4、当所述PDF页面不是纯文本页面,对所述PDF页面进行解析以获得所述PDF页面中的所有块元素,以及所述块元素的位置信息;

[0079] 本发明实施例中对所述PDF页面进行解析以获得所述PDF页面中的所有块元素,其中,所述块元素的类型包括文字、表格或者图形。

[0080] S5、提取所有所述块元素中的内容,根据所述位置信息将提取的内容写入预设文件格式的空白文件中,得到预览页面;

[0081] 本发明实施例中根据所述块元素中的位置信息在所述空白文件中构建空白区域,得到该块元素的内容插入区域;将每个所述块元素的提取内容填入每个所述该块元素的内容插入区域,得到所述预览页面。

[0082] 进一步地,本发明实施例中将每个所述块元素的提取内容填入每个所述该块元素的内容插入区域,得到所述预览页面时,会识别每个块元素提取内容中的字符的字体类型,若识别的字体类型在所述服务端中不存在,则将提取内容中的字符转化为预设的字体存入所述空白区域。

[0083] 本发明另一实施例中所述预览页面可以存储在区块链节点中,利用区块链节点高吞吐的特性,提高数据的取用效率。

[0084] S6、将所有所述预览页面发送至所述用户终端,以通过所述用户终端查看预览页面实现对所述目标PDF文件的预览。

[0085] 本发明实施例中将所有所述预览页面发送至所述用户终端,包括:

[0086] 将所有所述预览页面按照对应的PDF页面在所述目标PDF文件中的先后顺序进行组合,得到预览文件;

[0087] 将所述预览文件发送至所述用户终端。

[0088] 本发明实施例中将所述预览文件发送至所述用户终端,以响应所述文件预览请求,使得用户可以通过查看所述预览文件的形式实现对所述目标PDF文件的预览,从而实现在不支持PDF文件的终端上实现PDF文件的解析预览。

[0089] 进一步地,本发明实施例中还可以将所述预览文件间接发送至所述用户终端,具体地,本发明实施例中所述将所述预览文件发送至所述用户终端步骤还可以替换为将所述预览文件存储在预设存储区域,并获取所述预览文件的存储地址;将所述存储地址发送至所述用户终端,以便所述用户终端根据所述存储地址浏览所述预览文件,从而实现对所述目标PDF文件的预览。可选地,本发明实施例中所述预设存储区域为所述服务端的数据存储区域。

[0090] 如图2所示,是本发明PDF文件解析预览装置的功能模块图。

[0091] 本发明所述PDF文件解析预览装置100可以安装于电子设备中。根据实现的功能,所述PDF文件解析预览装置可以包括分类判断模块101、页面解析模块102、文件发送模块103,本发所述模块也可以称之为单元,是指一种能够被电子设备处理器所执行,并且能够完成固定功能的一系列计算机程序段,其存储在电子设备的存储器中。

[0092] 在本实施例中,关于各模块/单元的功能如下:

[0093] 所述分类判断模块101用于获取用户终端发出的目标PDF文件的文件预览请求,对所述目标PDF文件进行页面切分,得到PDF页面;判断所述PDF页面是否为纯文本页面;

[0094] 所述页面解析模块102用于当所述PDF页面为纯文本页面,识别所述PDF页面中的所有文本写入预设文件格式的空白文件,得到对应的预览页面;当所述PDF页面不是纯文本页面,对所述PDF页面进行解析以获得所述PDF页面中的所有块元素,以及所述块元素的位置信息;提取所有所述块元素中的内容,根据所述位置信息将提取的内容写入预设文件格式的空白文件中,得到预览页面;

[0095] 所述文件发送模块103用于将所有所述预览页面发送至所述用户终端,以通过所述用户终端查看预览页面实现对所述目标PDF文件的预览。

[0096] 详细地,本发明实施例中所述PDF文件解析预览装置100中所述的各模块在使用时采用与上述图1中所述的PDF文件解析预览方法一样的技术手段,并能够产生相同的技术效果,这里不再赘述。

[0097] 如图3所示,是本发明实现PDF文件解析预览方法的电子设备的结构示意图。

[0098] 所述电子设备可以包括处理器10、存储器11、通信总线12和通信接口13,还可以包括存储在所述存储器11中并可在所述处理器10上运行的计算机程序,如PDF文件解析预览程序。

[0099] 其中,所述存储器11至少包括一种类型的可读存储介质,所述可读存储介质包括闪存、移动硬盘、多媒体卡、卡型存储器(例如:SD或DX存储器等)、磁性存储器、磁盘、光盘等。所述存储器11在一些实施例中可以是电子设备的内部存储单元,例如该电子设备的移动硬盘。所述存储器11在另一些实施例中也可以是电子设备的外部存储设备,例如电子设备上配备的插接式移动硬盘、智能存储卡(Smart Media Card,SMC)、安全数字(Secure Digital,SD)卡、闪存卡(Flash Card)等。进一步地,所述存储器11还可以既包括电子设备的内部存储单元也包括外部存储设备。所述存储器11不仅可以用于存储安装于电子设备的应用软件及各类数据,例如PDF文件解析预览程序的代码等,还可以用于暂时地存储已经输出或者将要输出的数据。

[0100] 所述处理器10在一些实施例中可以由集成电路组成,例如可以由单个封装的集成电路所组成,也可以是由多个相同功能或不同功能封装的集成电路所组成,包括一个或者多个中央处理器(Central Processing Unit,CPU)、微处理器、数字处理芯片、图形处理器及各种控制芯片的组合等。所述处理器10是所述电子设备的控制核心(Control Unit),利用各种接口和线路连接整个电子设备的各个部件,通过运行或执行存储在所述存储器11内的程序或者模块(例如PDF文件解析预览程序等),以及调用存储在所述存储器11内的数据,以执行电子设备的各种功能和处理数据。

[0101] 所述通信总线12可以是外设部件互连标准(Peripheral Component Interconnect,简称PCI)总线或扩展工业标准结构(Extended Industry Standard Architecture,简称EISA)总线等。该总线可以分为地址总线、数据总线、控制总线等。所述通信总线12总线被设置为实现所述存储器11以及至少一个处理器10等之间的连接通信。为便于表示,图中仅用一条粗线表示,但并不表示仅有一根总线或一种类型的总线。

[0102] 图3仅示出了具有部件的电子设备,本领域技术人员可以理解的是,图3示出的结构并不构成对所述电子设备的限定,可以包括比图示更少或者更多的部件,或者组合某些部件,或者不同的部件布置。

[0103] 例如,尽管未示出,所述电子设备还可以包括给各个部件供电的电源(比如电池),

优选地,电源可以通过电源管理装置与所述至少一个处理器10逻辑相连,从而通过电源管理装置实现充电管理、放电管理、以及功耗管理等功能。电源还可以包括一个或一个以上的直流或交流电源、再充电装置、电源故障分类电路、电源转换器或者逆变器、电源状态指示器等任意组件。所述电子设备还可以包括多种传感器、蓝牙模块、Wi-Fi模块等,在此不再赘述。

[0104] 可选地,所述通信接口13可以包括有线接口和/或无线接口(如WI-FI接口、蓝牙接口等),通常用于在该电子设备与其他电子设备之间建立通信连接。

[0105] 可选地,所述通信接口13还可以包括用户接口,用户接口可以是显示器(Display)、输入单元(比如键盘(Keyboard)),可选地,用户接口还可以是标准的有线接口、无线接口。可选地,在一些实施例中,显示器可以是LED显示器、液晶显示器、触控式液晶显示器以及OLED(Organic Light-Emitting Diode,有机发光二极管)触摸器等。其中,显示器也可以适当的称为显示屏或显示单元,用于显示在电子设备中处理的信息以及用于显示可视化的用户界面。

[0106] 应该了解,所述实施例仅为说明之用,在专利申请范围上并不受此结构的限制。

[0107] 所述电子设备中的所述存储器11存储的PDF文件解析预览程序是多个计算机程序的组合,在所述处理器10中运行时,可以实现:

[0108] 获取用户终端发出的目标PDF文件的文件预览请求,对所述目标PDF文件进行页面切分,得到PDF页面;

[0109] 判断所述PDF页面是否为纯文本页面;

[0110] 当所述PDF页面为纯文本页面,识别所述PDF页面中的所有文本写入预设文件格式的空白文件,得到对应的预览页面;

[0111] 当所述PDF页面不是纯文本页面,对所述PDF页面进行解析以获得所述PDF页面中的所有块元素,以及所述块元素的位置信息;

[0112] 提取所有所述块元素中的内容,根据所述位置信息将提取的内容写入预设文件格式的空白文件中,得到预览页面;

[0113] 将所有所述预览页面发送至所述用户终端,以通过所述用户终端查看预览页面实现对所述目标PDF文件的预览。

[0114] 具体地,所述处理器10对上述计算机程序的具体实现方法可参考图1对应实施例中相关步骤的描述,在此不赘述。

[0115] 进一步地,所述电子设备集成的模块/单元如果以软件功能单元的形式实现并作为独立的产品销售或使用,可以存储在一个计算机可读取存储介质中。所述计算机可读介质可以是非易失性的,也可以是易失性的。所述计算机可读介质可以包括:能够携带所述计算机程序代码的任何实体或装置、记录介质、U盘、移动硬盘、磁碟、光盘、计算机存储器、只读存储器(ROM,Read-Only Memory)。

[0116] 本发明实施例还可以提供一种计算机可读存储介质,所述可读存储介质存储有计算机程序,所述计算机程序在被电子设备的处理器所执行时,可以实现:

[0117] 获取用户终端发出的目标PDF文件的文件预览请求,对所述目标PDF文件进行页面切分,得到PDF页面;

[0118] 判断所述PDF页面是否为纯文本页面;

[0119] 当所述PDF页面为纯文本页面,识别所述PDF页面中的所有文本写入预设文件格式的空白文件,得到对应的预览页面;

[0120] 当所述PDF页面不是纯文本页面,对所述PDF页面进行解析以获得所述PDF页面中的所有块元素,以及所述块元素的位置信息;

[0121] 提取所有所述块元素中的内容,根据所述位置信息将提取的内容写入预设文件格式的空白文件中,得到预览页面;

[0122] 将所有所述预览页面发送至所述用户终端,以通过所述用户终端查看预览页面实现对所述目标PDF文件的预览。

[0123] 进一步地,所述计算机可用存储介质可主要包括存储程序区和存储数据区,其中,存储程序区可存储操作系统、至少一个功能所需的应用程序等;存储数据区可存储根据区块链节点的使用所创建的数据等。

[0124] 在本发明所提供的几个实施例中,应该理解到,所揭露的设备,装置和方法,可以通过其它的方式实现。例如,以上所描述的装置实施例仅仅是示意性的,例如,所述模块的划分,仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式。

[0125] 所述作为分离部件说明的模块可以是或者也可以不是物理上分开的,作为模块显示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部模块来实现本实施例方案的目的。

[0126] 本申请实施例可以基于人工智能技术对相关的数据进行获取和处理。其中,人工智能(Artificial Intelligence, AI)是利用数字计算机或者数字计算机控制的机器模拟、延伸和扩展人的智能,感知环境、获取知识并使用知识获得最佳结果的理论、方法、技术及应用系统。

[0127] 另外,在本发明各个实施例中的各功能模块可以集成在一个处理单元中,也可以是各个单元单独物理存在,也可以两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现,也可以采用硬件加软件功能模块的形式实现。

[0128] 对于本领域技术人员而言,显然本发明不限于上述示范性实施例的细节,而且在不背离本发明的精神或基本特征的情况下,能够以其他的具体形式实现本发明。

[0129] 因此,无论从哪一点来看,均应将实施例看作是示范性的,而且是非限制性的,本发明的范围由所附权利要求而不是上述说明限定,因此旨在将落在权利要求的等同要件的含义和范围内的所有变化涵括在本发明内。不应将权利要求中的任何附关联图标记视为限制所涉及的权利要求。

[0130] 本发明所指区块链是分布式数据存储、点对点传输、共识机制、加密算法等计算机技术的新型应用模式。区块链(Blockchain),本质上是一个去中心化的数据库,是一串使用密码学方法相关联产生的数据块,每一个数据块中包含了一批网络交易的信息,用于验证其信息的有效性(防伪)和生成下一个区块。区块链可以包括区块链底层平台、平台产品服务层以及应用服务层等。

[0131] 此外,显然“包括”一词不排除其他单元或步骤,单数不排除复数。系统权利要求中陈述的多个单元或装置也可以由一个单元或装置通过软件或者硬件来实现。第二等词语用来表示名称,而并不表示任何特定的顺序。

[0132] 最后应说明的是,以上实施例仅用以说明本发明的技术方案而非限制,尽管参照较佳实施例对本发明进行了详细说明,本领域的普通技术人员应当理解,可以对本发明的技术方案进行修改或等同替换,而不脱离本发明技术方案的精神和范围。

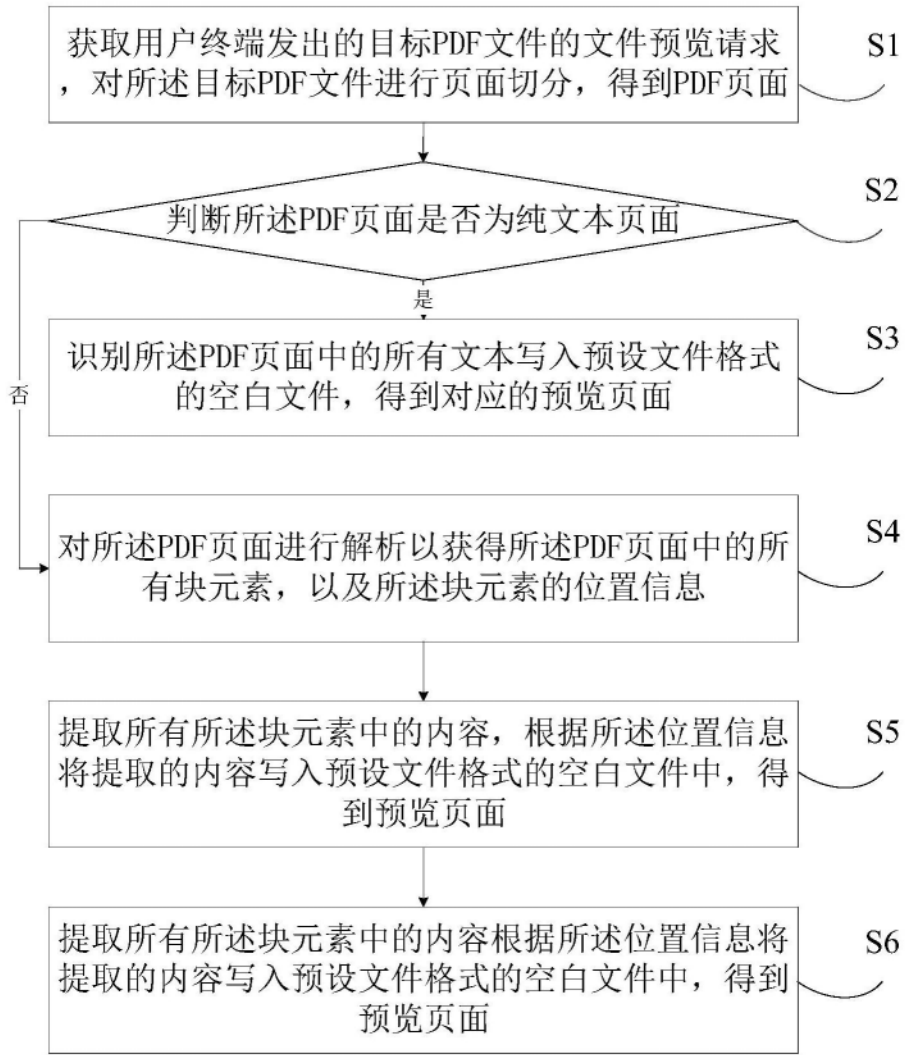


图1

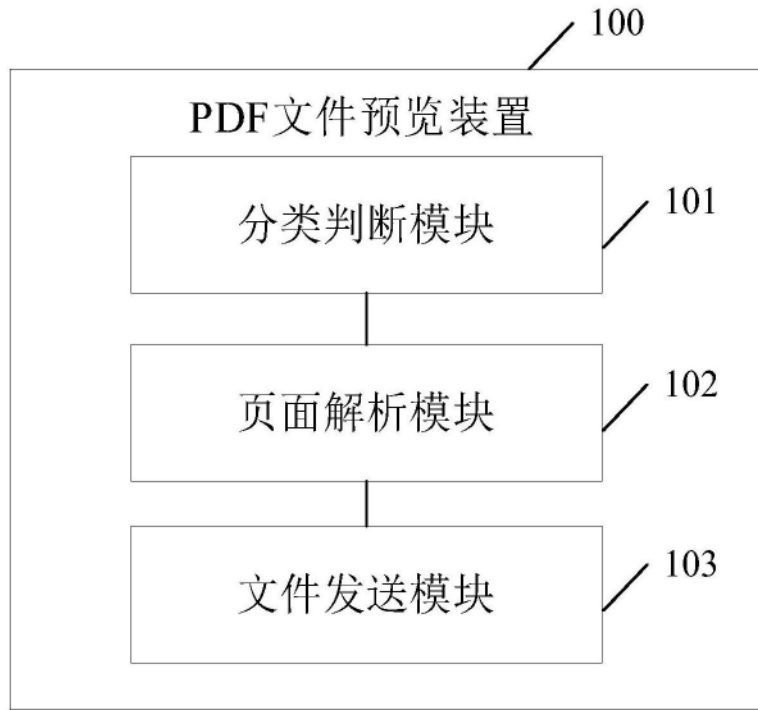


图2

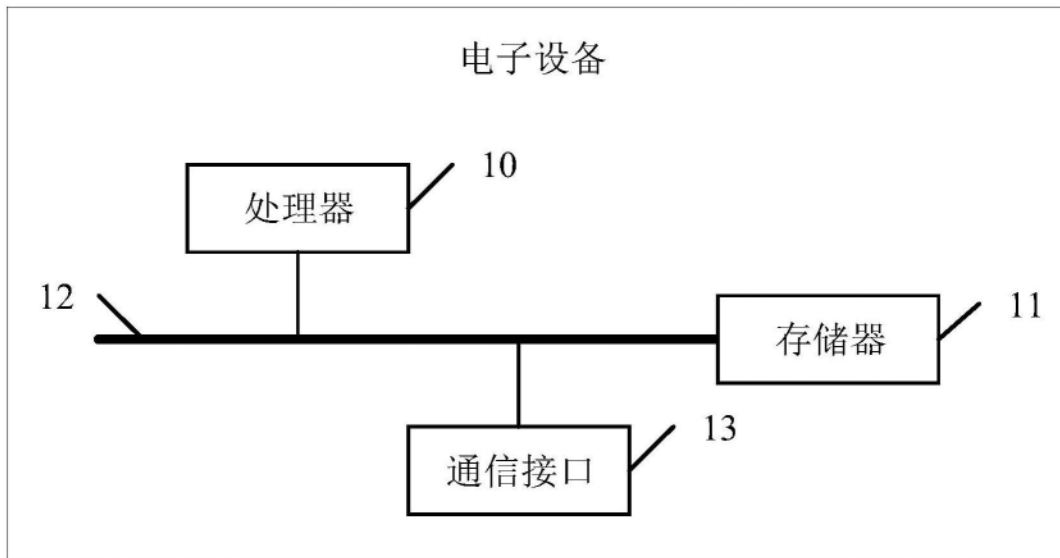


图3