



(12) 发明专利申请

(10) 申请公布号 CN 111915424 A

(43) 申请公布日 2020. 11. 10

(21) 申请号 202010754827.7

(22) 申请日 2020.07.30

(71) 申请人 平安证券股份有限公司

地址 518000 广东省深圳市福田区福田街
道益田路5023号平安金融中心B座第
22-25层

(72) 发明人 彭晶

(74) 专利代理机构 广州三环专利商标代理有限
公司 44202

代理人 熊永强

(51) Int. Cl.

G06Q 40/02 (2012.01)

G06Q 40/06 (2012.01)

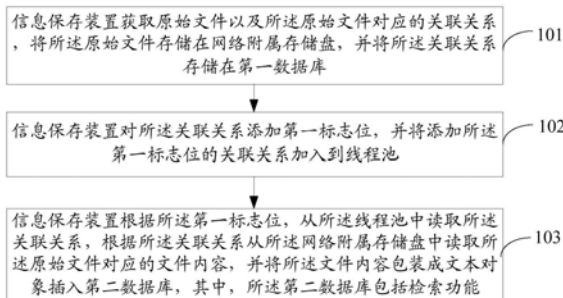
权利要求书2页 说明书10页 附图3页

(54) 发明名称

信息保存方法及相关产品

(57) 摘要

本申请涉及云存储技术领域,具体公开了一种信息保存方法及相关产品,该方法包括:获取原始文件以及所述原始文件对应的关联关系,将所述原始文件存储在网络附属存储盘,并将所述关联关系存储在第一数据库;对所述关联关系添加第一标志位,并将添加所述第一标志位的关联关系加入到线程池,所述第一标志位用于标识所述关联关系在所述线程池中处于待读取状态;根据所述第一标志位,从所述线程池中读取所述关联关系,根据所述关联关系从所述网络附属存储盘中读取所述原始文件对应的文件内容,并将所述文件内容包装成文本对象插入第二数据库,其中,所述第二数据库包括检索功能。本申请实施例在保存文件内容方面,有利于节省人力物力。



1. 一种信息保存方法,其特征在于,包括:

获取原始文件以及所述原始文件对应的关联关系,将所述原始文件存储在网络附属存储盘,并将所述关联关系存储在第一数据库;

对所述关联关系添加第一标志位,并将添加所述第一标志位的关联关系加入到线程池,所述第一标志位用于标识所述关联关系在所述线程池中处于待读取状态;

根据所述第一标志位,从所述线程池中读取所述关联关系,根据所述关联关系从所述网络附属存储盘中读取所述原始文件对应的文件内容,并将所述文件内容包装成文本对象插入第二数据库,其中,所述第二数据库包括检索功能。

2. 根据权利要求1所述的方法,其特征在于,在所述原始文件中编码有水印数据的情况下,在所述将所述原始文件存储在网络附属存储盘之前,所述方法还包括:

对所述原始文件进行解码,得到所述水印数据;

对所述水印数据进行真伪验证,在确定所述水印数据为真的情况下,执行将所述原始文件存储在网络附属存储盘的步骤。

3. 根据权利要求1或2所述的方法,其特征在于,在所述原始文件中添加有时间戳,且所述时间戳用于记录所述原始文件的保存时间的情况下,在所述将所述原始文件存储在网络附属存储盘之前,所述方法还包括:

解析所述时间戳,得到所述原始文件的保存时间;

将所述保存时间与所述关联关系中记录的所述原始文件的上传时间进行比对,在所述上传时间与所述保存时间的差值小于阈值的情况下,确定所述原始文件未被篡改,执行将所述原始文件存储在网络附属存储盘的步骤。

4. 根据权利要求1-3中任一项所述的方法,其特征在于,所述根据所述关联关系从所述网络附属存储盘中读取所述原始文件的文件内容,包括:

根据所述关联关系确定所述原始文件在所述网络附属存储盘的存储位置;

在所述原始文件为电子版文件的情况下,通过文档读写工具从所述存储位置读取所述原始文件中的文件内容;

在所述原始文件为扫描版文件的情况下,通过光学字符识别工具从所述存储位置读取所述原始文件中的文件内容。

5. 根据权利要求1-3中任一项所述的方法,其特征在于,所述根据所述关联关系从所述网络附属存储盘中读取所述原始文件的文件内容,包括:

将所述关联关系拆分为N个子关联关系,其中,N个所述子关联关系中任意两个所述子关联关系不存在关联性;

创建与所述N个子关联关系对应的N个线程,并行执行所述N个线程,从所述网络附属存储盘中分别读取与每个所述子关联关系对应的子文件内容;

将所述N个子关联关系对应的子文件内容进行合并,得到所述文件内容。

6. 根据权利要求1-5中任一项所述的方法,其特征在于,所述方法还包括:

修改所述关联关系的第一标志位,得到第二标志位,所述第二标志位用于标识所述关联关系在所述线程池中处于已读取状态。

7. 根据权利要求1-6中任一项所述的方法,其特征在于,所述方法还包括:

获取在全文检索引擎工具包中预先设定的存放位置;

从所述第二数据库中读取所述原始文件的文件内容,并将读取到的文件内容存储在所述预先设定的存放位置,得到目标文件,所述目标文件支持检索。

8. 一种信息保存装置,其特征在于,包括:

获取单元,用于获取原始文件以及所述原始文件对应的关联关系;

处理单元,用于将所述原始文件存储在网络附属存储盘,并将所述关联关系存储在所述第一数据库;

所述处理单元,还用于对所述关联关系添加第一标志位,并将添加所述第一标志位的关联关系加入到线程池,所述第一标志位用于标识所述关联关系在所述线程池中处于待读取状态;

所述处理单元,还用于根据所述第一标志位,从所述线程池中读取所述关联关系,根据所述关联关系从所述网络附属存储盘中读取所述原始文件对应的文件内容,并将所述文件内容包装成文本对象插入第二数据库,其中,所述第二数据库包括检索功能。

9. 一种电子设备,其特征在于,包括处理器、存储器、通信接口以及一个或多个程序,其中,所述一个或多个程序被存储在所述存储器中,并且被配置由所述处理器执行,所述程序包括用于执行权利要求1-7中任一项方法中的步骤的指令。

10. 一种计算机可读存储介质,其特征在于,所述计算机可读存储介质存储有计算机程序,所述计算机程序被处理器执行以实现如权利要求1-7中任一项所述的方法。

信息保存方法及相关产品

技术领域

[0001] 本申请涉及信息处理技术领域,具体涉及一种信息保存方法及相关产品。

背景技术

[0002] 投资银行(Investment Banks)是与商业银行相对应的一类金融机构。主要从事证券发行、承销、交易、企业重组、兼并与收购、投资分析、风险投资、项目融资等业务的非银行金融机构,是资本市场上的主要金融中介。然而,投资银行目前参与的投资项目很多文件底稿都是采用纸质底稿。但是,一个投资项目的信息需要大量的纸张才可以完成记录,而且对该投资项目的信息进行备份需要更多的纸张,因此纸质底稿的存储占用太多空间;而且,国家规定要将每个投资项目的纸质底稿保存十年以上时间,然而,纸质底稿在保存过程中容易受外界影响而损坏,保存难度很大;另外,纸质底稿只能根据文件名找到当初归档底稿的具体位置,难以进行内容节选。

[0003] 因此,采用纸质底稿保存投资银行的项目信息的方式,需要投入很多的人力和物力来制作和保存纸质底稿,而且无法从纸质底稿中进行内容节选,存在严重的检索难的问题。

发明内容

[0004] 本申请实施例提供了一种信息保存方法及相关产品。通过将文件内容包装成文本对象,生成可检索的电子文件,无需使用纸质底稿保存投资银行的项目信息,节约人力物力。

[0005] 第一方面,本申请实施例提供一种信息保存方法,包括:

[0006] 获取原始文件以及所述原始文件对应的关联关系,将所述原始文件存储在网络附属存储盘,并将所述关联关系存储在第一数据库;

[0007] 对所述关联关系添加第一标志位,并将添加所述第一标志位的关联关系加入到线程池,所述第一标志位用于标识所述关联关系在所述线程池中处于待读取状态;

[0008] 根据所述第一标志位,从所述线程池中读取所述关联关系,根据所述关联关系从所述网络附属存储盘中读取所述原始文件对应的文件内容,并将所述文件内容包装成文本对象插入第二数据库,其中,所述第二数据库包括检索功能。

[0009] 第二方面,本申请实施例提供一种信息保存装置,包括:

[0010] 获取单元,用于获取原始文件以及所述原始文件对应的关联关系;

[0011] 处理单元,用于将所述原始文件存储在网络附属存储盘,并将所述关联关系存储在第一数据库;

[0012] 所述处理单元,还用于对所述关联关系添加第一标志位,并将添加所述第一标志位的关联关系加入到线程池,所述第一标志位用于标识所述关联关系在所述线程池中处于待读取状态;

[0013] 所述处理单元,还用于根据所述第一标志位,从所述线程池中读取所述关联关系,

根据所述关联关系从所述网络附属存储盘中读取所述原始文件对应的文件内容,并将所述文件内容包装成文本对象插入第二数据库,其中,所述第二数据库包括检索功能。

[0014] 第三方面,本申请实施例提供一种电子设备,包括处理器、存储器、通信接口以及一个或多个程序,其中,所述一个或多个程序被存储在所述存储器中,并且被配置由所述处理器执行,所述程序包括用于执行如第一方面所述的方法中的步骤的指令。

[0015] 第四方面,本申请实施例提供一种计算机可读存储介质,所述计算机可读存储介质存储有计算机程序,所述计算机程序使得计算机执行如第一方面所述的方法。

[0016] 第五方面,本申请实施例提供一种计算机程序产品,所述计算机程序产品包括存储了计算机程序的非瞬时性计算机可读存储介质,所述计算机可操作来使计算机执行如第一方面所述的方法。

[0017] 实施本申请实施例,具有如下有益效果:

[0018] 可以看出,在本申请实施例中,从线程池中读取原始文件的关联关系,并根据该关联关系从网络附属存储盘中读取该原始文件的文件内容;并将该文件内容包装成文本对象插入到具有检索功能的第二数据库。因此,在该第二数据库中以电子版的方式保存了该原始文件的文件内容,无需纸质化底稿保存该原始文件,节省了人力物力;而且,在该第二数据库中可通过检索的方式来获取该原始文件的节选内容,实现对该原始文件的文件内容进行检索。

附图说明

[0019] 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单地介绍,显而易见地,下面描述中的附图是本申请的一些实施例,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。

[0020] 图1为本申请实施例提供了一种信息保存方法的流程示意图;

[0021] 图2为本申请实施例提供的另一种信息保存方法的流程示意图;

[0022] 图3为本申请实施例提供的另一种信息保存方法的流程示意图;

[0023] 图4为本申请实施例提供了一种信息保存装置的结构示意图;

[0024] 图5为本申请实施例提供了一种信息保存装置的功能单元组成框图。

具体实施方式

[0025] 下面将结合本申请实施例中的附图,对本申请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作出创造性劳动前提下所获得的所有其他实施例,都属于本申请保护的范围。

[0026] 本申请的说明书和权利要求书及所述附图中的术语“第一”、“第二”、“第三”和“第四”等是用于区别不同对象,而不是用于描述特定顺序。此外,术语“包括”和“具有”以及它们任何变形,意图在于覆盖不排他的包含。例如包含了一系列步骤或单元的过程、方法、系统、产品或设备没有限定于已列出的步骤或单元,而是可选地还包括没有列出的步骤或单元,或可选地还包括对于这些过程、方法、产品或设备固有的其它步骤或单元。

[0027] 在本文中提及“实施例”意味着,结合实施例描述的特定特征、结果或特性可以包含在本申请的至少一个实施例中。在说明书中的各个位置出现该短语并不一定均是指相同的实施例,也不是与其它实施例互斥的独立的或备选的实施例。本领域技术人员显式地和隐式地理解的是,本文所描述的实施例可以与其它实施例相结合。

[0028] 本申请中的信息保存装置可以包括智能手机(如Android手机、iOS手机、Windows Phone手机等)、平板电脑、掌上电脑、笔记本电脑、移动互联网设备MID(Mobile Internet Devices,简称:MID)或穿戴式设备等。上述信息保存装置仅是举例,而非穷举,包括但不限于上述信息保存装置。在实际应用中,上述信息保存装置还可以包括:智能车载终端、计算机设备等等。

[0029] 参阅图1,图1为本申请实施例提供的一种信息保存方法的流程示意图。该方法应用于信息保存装置。该方法包括但不限于以下步骤:

[0030] 101:信息保存装置获取原始文件以及所述原始文件对应的关联关系,将所述原始文件存储在网络附属存储盘,并将所述关联关系存储在第一数据库。

[0031] 其中,该原始文件即是投资项目的纸质稿件的电子版文件;该关联关系包括该原始文件的文件名称、在该网络附属存储(Network Attached Storage,NAS)盘中的存储路径、该原始文件对应的项目信息、人员信息、文件大小、上传时间,等等;其中,该项目信息包括该原始文件所对应的投资项目中的成员信息,该成员信息包括该投资项目中的人员负责,组内职务分明,如产品经理、项目主管、质控人员、内核人员、客户经理,等等。

[0032] 其中,该第一数据库可以为关系型数据库,例如,Oracle数据库;也可以为非关系型数据库,例如,Hbase数据库。该关联关系可以Excel表格的形式存储到该第一数据库中。

[0033] 示例性的,该原始文件可以通过生成该原始文件的用户设备中获取。

[0034] 示例性的,该关联关系从专门的项目组管理系统同步而来。具体地,原始文件的所有目录是根据证监会要求统一设置的,因此生成原始文件的过程中,该项目组管理系统就建立了与目录文件、项目与原始文件之间的对应关系,即上述的关联关系。故可从项目组管理系统中获取该原始文件的关联关系。

[0035] 102:信息保存装置对所述关联关系添加第一标志位,并将添加所述第一标志位的关联关系加入到线程池。

[0036] 其中,对该关联关系添加第一标志位实质上就是对该关联关系加行锁,对该关联关系加过行锁后只会被当前的线程读取,不会再被其他线程获取。例如,可使用Oracle数据库的for update关键词对该关联关系加行锁。其中,使用for update关键词加行锁为现有技术,不再叙述。具体地,该第一标志位用于标识所述关联关系在该线程池中处于待读取状态。

[0037] 需要说明,步骤101和步骤102是异步操作,即获取原始文件与将关联关系加入线程池可以在不同的时刻执行。

[0038] 103:信息保存装置根据所述第一标志位,从所述线程池中读取所述关联关系,根据所述关联关系从所述网络附属存储盘中读取所述原始文件对应的文件内容,并将所述文件内容包装成文本对象插入第二数据库,其中,所述第二数据库包括检索功能。

[0039] 信息保存装置读取该线程池中各个任务的标志为,在读取到该标志位为第一标志位的情况下,即该任务处于待读取状态(即待消费状态)下,则执行该任务,即从该线程池中

解析该关联关系,并根据该关联关系从NAS盘中读取该原始文件的文件内容,并将该文件内容包装成文本对象插入到第二数据库。

[0040] 可以看出,在本申请实施例中,从线程池中读取原始文件的关联关系,并根据该关联关系从网络附属存储盘中读取该原始文件的文件内容;并将该文件内容包装成文本对象插入到具有检索功能的第二数据库。因此,在该第二数据库中以电子版的方式保存了该原始文件的文件内容,无需纸质化底稿保存该原始文件,节省了人力物力;而且,在该第二数据库中可通过检索的方式来获取该原始文件的节选内容,实现对该原始文件的文件内容进行检索。

[0041] 示例性的,从该关联关系中获取该原始文件在该NAS盘中的存储位置,从该存储位置获取该原始文件,并从该原始文件中提取该原始文件的文件内容。

[0042] 示例性的,在该原始文件为电子版文件的情况下,通过文档读取工具从所述存储位置读取所述原始文件中的文件内容,其中,该文档读取工具可以是Java的POI技术,其中,该POI技术的核心是对微软的office文档的VBA函数操作的封装,能从office文档本身的正文行、段落、标题等维度获取office文档的文件内容,其通过POI技术读取文件内容为现有技术,不再具体叙述;在该原始文件为扫描版文件的情况下,则可通过光学字符识别(Optical Character Recognition,OCR)工具从该存储位置读取该原始文件中的文件内容,即通过OCR工具,采用光学的方式将该原始文件中的字符转换为黑白点阵的图像文件,然后将图像文件中的文字转换成文本格式,得到该文件内容。

[0043] 在一些可能的实施方式中,在根据该关联关系从该网络附属存储盘中读取该原始文件的文件内容之前,还可以将该关联关系拆分成N个子关联关系,且该N个子关联关系中任意两个子关联关系不存在关联性,即各个子关联关系在进行文件内容读取的过程中不会相互干扰,例如,子关联关系1为项目负责人,子关联关系2为项目内容,这两个子关联关系在进行文件内容读取的过程中可以并行读取,不会存在相互干扰;因此,创建与该N个子关联关系对应的N个线程,并行执行该N个线程,从该NAS盘中分别读取与每个子关联关系对应的子文件内容,其中,读取与该每个子关联关系对应的子文件内容就是根据获取该子关联关系对应的子文件内容在该NAS盘中的存储位置,然后,通过上述的POI技术或者OCR工具从该存储位置读取该子文件内容;最后,将该N个子关联关系对应的子文件内容进行合并,得到该文件内容。通过将该关联关系进行拆分,可并行的执行文件读取过程,提高了文件内容读取的效率。

[0044] 与上述拆分关联关系相适应的,可将该原始文件拆分为N个子片段,该N个子片段中任意两个子片段之间不存在关联性;然后,对每个子片段添加与该N个子关联关系对应的标识信息,并将N个子片段存在该NAS盘中的不同的文件夹下,该文件夹的名称可以为该标识信息;故在执行每个子关联关系对应的线程的过程中,可以与文件夹的文件名进行匹配,快速查询到每个子关联关系对应的子文件内容的存储位置,进一步提高文件内容的读取效率。

[0045] 示例性的,该第二数据库可以为MongoDB数据库,该MongoDB数据库具备检索功能,可在该MongoDB数据库中输入关键词,检索与该关键词对应的节选内容。具体地,将读取到的文件内容进行赋值转换,生成文本对象,将该文本对象插入到MongoDB数据库,从而实现将NAS盘中存储的原始文件中的文件内容录入到MongoDB数据库。其中,通过赋值转换生成

文本对象为现有技术,不再叙述。

[0046] 在一些可能的实施方式中,为了保证该原始文件的传输过程中的安全性,可在生成该原始文件的过程中,对该原始条件水印数据,该水印数据为数字水印,即将该原始文件和水印数据进行编码得到的,与传统的水印数据相比,这种水印数据不影响该原始文件的观赏性,以及安全性高,其中,对该原始文件编码水印数据为现有技术,不再叙述。因此,在该原始文件中编码有水印数据的情况下,在所述将所述原始文件存储在网络附属存储盘之前,该方法还包括:

[0047] 对该原始文件进行解码,得到该水印数据;

[0048] 对该水印数据进行真伪验证,在确定该水印数据为真的情况下,即确定该水印数据为该原始文件的管理者添加的水印数据,确定该原始文件未被篡改、修改或替换,等等,则将该原始文件存储在网络附属存储盘。

[0049] 可以看出,在本实施方式中,通过对原始文件添加水印数据,保证该原始文件在传输过程中的安全性。

[0050] 在一些可能的实施方式中,在该原始文件中添加有时间戳,且该时间戳用于记录所述原始文件的保存时间的情况下,在所述将所述原始文件存储在网络附属存储盘之前,所述方法还包括:

[0051] 解析所述时间戳,得到所述原始文件的保存时间;

[0052] 将所述保存时间与所述关联关系中记录的所述原始文件的上传时间进行比对,在所述上传时间与所述保存时间的差值小于阈值的情况下,确定所述原始文件未被篡改,执行将所述原始文件存储在网络附属存储盘的步骤。

[0053] 其中,该保存时间为最后一次保存该原始文件的时间,即在每次保存该原始文件的情况下,自动生成时间戳,该时间戳用于记录本次保存该原始文件的时间;而且,在下一次保存该原始文件的情况下,生成的时间戳会自动覆盖上次的时间戳。因此,上述时间戳用于记录该原始文件的最后保存时间。一般来说,最后一次保存后就需要将该原始文件上传至项目组管理系统。因此,通过将该保存时间与上传时间进行比对,判断出该原始文件在上传之后,有没有被改动,从而保证后续读取到的文件内容,为原始的文件内容,保证信息的安全。

[0054] 在一些可能的实施方式中,在读取该文件内容之后,该方法还包括:

[0055] 修改所述关联关系的第一标志位,得到第二标志位,所述第二标志位用于标识所述关联关系在所述线程池中处于已读取状态。通过修改该关联关系的标志位,即通过Commit命令释放该关联关系的行锁,避免重复读取该关联关系。

[0056] 在一些可能的实施方式中,该方法还包括:

[0057] 获取在全文检索引擎工具包中预先设定的存放位置;

[0058] 从所述第二数据库中读取所述原始文件的文件内容,并将读取到的文件内容存储在所述预先设定的存放位置,得到目标文件,所述目标文件支持检索。可通过检索获取从该文件内容中获取节选内容。

[0059] 其中,该全文检索引擎工具包可以为Lucene。

[0060] 在一些可能的实施方式中,所述方法还包括:

[0061] 确定该文件内容中的目标节选内容,该目标节选内容为读取该文件内容的过程

中,容易出现读取错误的内容,例如,该文件内容中的公式部分、图表部分,等等;将该目标节选内容在MongoDB数据库的权限修改为可编辑状态,并突出显示该目标节选内容,以引导对该目标节选内容进行核对,并在出错的情况下,对该目标节选内容进行编辑,进而提高读取该文件内容的正确率。

[0062] 参阅图2,图2为本申请实施例提供的另一种信息保存方法的流程示意图。该方法应用于信息保存装置。该方法包括但不限于以下步骤:

[0063] 201:信息保存装置获取原始文件以及所述原始文件对应的关联关系。

[0064] 202:信息保存装置对所述原始文件进行解码,得到所述水印数据。

[0065] 203:信息保存装置对所述水印数据进行真伪验证,在确定所述水印数据为真的情况下,将所述原始文件存储在网络附属存储盘,并将所述关联关系存储在第一数据库。

[0066] 204:信息保存装置对所述关联关系添加第一标志位,并将添加所述第一标志位的关联关系加入到线程池,所述第一标志位用于标识所述关联关系在所述线程池中处于待读取状态。

[0067] 205:信息保存装置根据所述第一标志位,从所述线程池中读取所述关联关系,根据所述关联关系从所述网络附属存储盘中读取所述原始文件对应的文件内容,并将所述文件内容包装成文本对象插入第二数据库,其中,所述第二数据库包括检索功能。

[0068] 需要说明的是,图2所示的方法的各个步骤的具体实现过程可参见上述图1所述方法的具体实现过程,在此不再叙述。

[0069] 可以看出,在本申请实施例中,从线程池中读取原始文件的关联关系,并根据该关联关系从网络附属存储盘中读取该原始文件的文件内容;并将该文件内容包装成文本对象插入到具有检索功能的第二数据库。因此,在该第二数据库中以电子版的方式保存了该原始文件的文件内容,无需纸质化底稿保存该原始文件,节省了人力物力;而且,在该第二数据库中可通过检索的方式来获取该原始文件的节选内容,实现对该原始文件的文件内容进行检索;并且,对该原始文件编码有水印数据,从而可提高该原始文件传输过程中的安全性,防止信息泄露。

[0070] 参阅图3,图3为本申请实施例提供的另一种信息保存方法的流程示意图。该方法应用于信息保存装置。该方法包括但不限于以下步骤:

[0071] 301:信息保存装置获取原始文件以及所述原始文件对应的关联关系。

[0072] 302:信息保存装置对所述原始文件进行解码,得到所述水印数据。

[0073] 303:信息保存装置对所述水印数据进行真伪验证,在确定所述水印数据为真的情况下,将所述原始文件存储在网络附属存储盘,并将所述关联关系存储在第一数据库。

[0074] 304:信息保存装置对所述关联关系添加第一标志位,并将添加所述第一标志位的关联关系加入到线程池,所述第一标志位用于标识所述关联关系在所述线程池中处于待读取状态。

[0075] 305:信息保存装置根据所述第一标志位,从所述线程池中读取所述关联关系,根据所述关联关系从所述网络附属存储盘中读取所述原始文件对应的文件内容,并将所述文件内容包装成文本对象插入第二数据库,其中,所述第二数据库包括检索功能。

[0076] 306:信息保存装置修改所述关联关系的第一标志位,得到第二标志位,所述第二标志位用于标识所述关联关系在所述线程池中处于已读取状态。

[0077] 307:信息保存装置获取在全文检索引擎工具包中预先设定的存放位置。

[0078] 308:信息保存装置从所述第二数据库中读取所述原始文件的文件内容,并将读取到的文件内容存储在所述预先设定的存放位置,得到目标文件,所述目标文件支持检索。

[0079] 需要说明的是,图3所示的方法的各个步骤的具体实现过程可参见上述图1所述方法的具体实现过程,在此不再叙述。

[0080] 可以看出,在本申请实施例中,可以看出,在本申请实施例中,从线程池中读取原始文件的关联关系,并根据该关联关系从网络附属存储盘中读取该原始文件的文件内容;并将该文件内容包装成文本对象插入到具有检索功能的第二数据库。因此,在该第二数据库中以电子版的方式保存了该原始文件的文件内容,无需纸质化底稿保存该原始文件,节省了人力物力;而且,在该第二数据库中可通过检索的方式来获取该原始文件的节选内容,实现对该原始文件的文件内容进行检索;并且,对该原始文件编码有水印数据,从而可提高该原始文件传输过程中的安全性,防止信息泄露。

[0081] 参阅图4,图4为本申请实施例提供的一种信息保存装置的结构示意图。如图4所示,信息保存装置400包括处理器、存储器、通信接口以及一个或多个程序,其中,上述一个或多个程序被存储在上述存储器中,并且被配置由上述处理器执行,上述程序包括用于执行以下步骤的指令:

[0082] 获取原始文件以及所述原始文件对应的关联关系,将所述原始文件存储在网络附属存储盘,并将所述关联关系存储在所述第一数据库;

[0083] 对所述关联关系添加第一标志位,并将添加所述第一标志位的关联关系加入到线程池,所述第一标志位用于标识所述关联关系在所述线程池中处于待读取状态;

[0084] 根据所述第一标志位,从所述线程池中读取所述关联关系,根据所述关联关系从所述网络附属存储盘中读取所述原始文件对应的文件内容,并将所述文件内容包装成文本对象插入第二数据库,其中,所述第二数据库包括检索功能。

[0085] 在一些可能的实施方式中,在所述原始文件中编码有水印数据的情况下,在所述将所述原始文件存储在网络附属存储盘之前,上述程序还用于执行以下步骤的指令:

[0086] 对所述原始文件进行解码,得到所述水印数据;

[0087] 对所述水印数据进行真伪验证,在确定所述水印数据为真的情况下,执行将所述原始文件存储在网络附属存储盘的步骤。

[0088] 在一些可能的实施方式中,在所述原始文件中添加有时间戳,且所述时间戳用于记录所述原始文件的保存时间的情况下,在所述将所述原始文件存储在网络附属存储盘之前,上述程序还用于执行以下步骤的指令:

[0089] 将所述保存时间与所述关联关系中记录的所述原始文件的上传时间进行比对,在所述上传时间与所述保存时间的差值小于阈值的情况下,确定所述原始文件未被篡改,执行将所述原始文件存储在网络附属存储盘的步骤。

[0090] 在一些可能的实施方式中,在根据所述关联关系从所述网络附属存储盘中读取所述原始文件的文件内容方面,上述程序具体用于执行以下步骤的指令:

[0091] 根据所述关联关系确定所述原始文件在所述网络附属存储盘的存储位置;

[0092] 在所述原始文件为电子版文件的情况下,通过文档读写工具从所述存储位置读取所述原始文件中的文件内容;

[0093] 在所述原始文件为扫描版文件的情况下,通过光学字符识别工具从所述存储位置读取所述原始文件中的文件内容。

[0094] 在一些可能的实施方式中,在根据所述关联关系从所述网络附属存储盘中读取所述原始文件的文件内容方面,上述程序具体用于执行以下步骤的指令:

[0095] 将所述关联关系拆分为N个子关联关系,其中,N个所述子关联关系中任意两个所述子关联关系不存在关联性;

[0096] 创建与所述N个子关联关系对应的N个线程,并行执行所述N个线程,从所述网络附属存储盘中分别读取与每个所述子关联关系对应的子文件内容;

[0097] 将所述N个子关联关系对应的子文件内容进行合并,得到所述文件内容。

[0098] 在一些可能的实施方式中,上述程序还用于执行以下步骤的指令:

[0099] 修改所述关联关系的第一标志位,得到第二标志位,所述第二标志位用于标识所述关联关系在所述线程池中处于已读取状态。

[0100] 在一些可能的实施方式中,上述程序还用于执行以下步骤的指令:

[0101] 获取在全文检索引擎工具包中预先设定的存放位置;

[0102] 从所述第二数据库中读取所述原始文件的文件内容,并将读取到的文件内容存储在所述预先设定的存放位置,得到目标文件,所述目标文件支持检索。

[0103] 参阅图5,图5本申请实施例提供的一种信息保存装置的功能单元组成框图。信息保存装置500包括:获取单元501和处理单元502,其中:

[0104] 获取单元501,用于获取原始文件以及所述原始文件对应的关联关系;

[0105] 处理单元502,用于将所述原始文件存储在网络附属存储盘,并将所述关联关系存储在所述第一数据库;

[0106] 处理单元502,还用于对所述关联关系添加第一标志位,并将添加所述第一标志位的关联关系加入到线程池,所述第一标志位用于标识所述关联关系在所述线程池中处于待读取状态;

[0107] 处理单元502,还用于根据所述第一标志位,从所述线程池中读取所述关联关系,根据所述关联关系从所述网络附属存储盘中读取所述原始文件对应的文件内容,并将所述文件内容包装成文本对象插入第二数据库,其中,所述第二数据库包括检索功能。

[0108] 在一些可能的实施方式中,在所述原始文件中编码有水印数据的情况下,在所述将所述原始文件存储在网络附属存储盘之前,处理单元502,还用于:

[0109] 对所述原始文件进行解码,得到所述水印数据;

[0110] 对所述水印数据进行真伪验证,在确定所述水印数据为真的情况下,执行将所述原始文件存储在网络附属存储盘的步骤。

[0111] 在一些可能的实施方式中,在所述原始文件中添加有时间戳,且所述时间戳用于记录所述原始文件的保存时间的情况下,在所述将所述原始文件存储在网络附属存储盘之前,处理单元502,还用于:

[0112] 将所述保存时间与所述关联关系中记录的所述原始文件的上传时间进行比对,在所述上传时间与所述保存时间的差值小于阈值的情况下,确定所述原始文件未被篡改,执行将所述原始文件存储在网络附属存储盘的步骤。

[0113] 在一些可能的实施方式中,在根据所述关联关系从所述网络附属存储盘中读取所

述原始文件的文件内容方面,处理单元502,具体用于:

[0114] 根据所述关联关系确定所述原始文件在所述网络附属存储盘的存储位置;

[0115] 在所述原始文件为电子版文件的情况下,通过文档读写工具从所述存储位置读取所述原始文件中的文件内容;

[0116] 在所述原始文件为扫描版文件的情况下,通过光学字符识别工具从所述存储位置读取所述原始文件中的文件内容。

[0117] 在一些可能的实施方式中,在根据所述关联关系从所述网络附属存储盘中读取所述原始文件的文件内容方面,处理单元502,具体用于:

[0118] 将所述关联关系拆分为N个子关联关系,其中,N个所述子关联关系中任意两个所述子关联关系不存在关联性;

[0119] 创建与所述N个子关联关系对应的N个线程,并行执行所述N个线程,从所述网络附属存储盘中分别读取与每个所述子关联关系对应的子文件内容;

[0120] 将所述N个子关联关系对应的子文件内容进行合并,得到所述文件内容。

[0121] 在一些可能的实施方式中,处理单元502,还用于:

[0122] 修改所述关联关系的第一标志位,得到第二标志位,所述第二标志位用于标识所述关联关系在所述线程池中处于已读取状态。

[0123] 在一些可能的实施方式中,处理单元502,还用于:

[0124] 获取在全文搜索引擎工具包中预先设定的存放位置;

[0125] 从所述第二数据库中读取所述原始文件的文件内容,并将读取到的文件内容存储在所述预先设定的存放位置,得到目标文件,所述目标文件支持检索。

[0126] 本申请实施例还提供一种计算机存储介质,所述计算机可读存储介质存储有计算机程序,所述计算机程序被处理器执行以实现如上述方法实施例中记载的任何一种信息保存方法的部分或全部步骤。

[0127] 本申请实施例还提供一种计算机程序产品,所述计算机程序产品包括存储了计算机程序的非瞬时性计算机可读存储介质,所述计算机程序可操作来使计算机执行如上述方法实施例中记载的任何一种信息保存方法的部分或全部步骤。

[0128] 需要说明的是,对于前述的各方法实施例,为了简单描述,故将其都表述为一系列的动作组合,但是本领域技术人员应该知悉,本申请并不受所描述的动作顺序的限制,因为依据本申请,某些步骤可以采用其他顺序或者同时进行。其次,本领域技术人员也应该知悉,说明书中所描述的实施例均属于可选实施例,所涉及的动作和模块并不一定是本申请所必须的。

[0129] 在上述实施例中,对各个实施例的描述都各有侧重,某个实施例中未详述的部分,可以参见其他实施例的相关描述。

[0130] 在本申请所提供的几个实施例中,应该理解到,所揭露的装置,可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,例如所述单元的划分,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,例如多个单元或组件可以结合或者可以集成到另一个系统,或一些特征可以忽略,或不执行。另一点,所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口,装置或单元的间接耦合或通信连接,可以是电性或其它的形式。

[0131] 所述作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元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单元来实现本实施例方案的目的。

[0132] 另外,在本申请各个实施例中的各功能单元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,也可以采用软件程序模块的形式实现。

[0133] 所述集成的单元如果以软件程序模块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器中。基于这样的理解,本申请的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的全部或部分可以以软件产品的形式体现出来,该计算机软件产品存储在一个存储器中,包括若干指令用以使得一台计算机设备(可为个人计算机、服务器或者网络设备)执行本申请各个实施例所述方法的全部或部分步骤。而前述的存储器包括:U盘、只读存储器(ROM,Read-Only Memory)、随机存取存储器(RAM,Random Access Memory)、移动硬盘、磁碟或者光盘等各种可以存储程序代码的介质。

[0134] 本领域普通技术人员可以理解上述实施例的各种方法中的全部或部分步骤是可以通程序来指令相关的硬件来完成,该程序可以存储于一计算机可读取存储器中,存储器可以包括:闪存盘、只读存储器(英文:Read-Only Memory,简称:ROM)、随机存取器(英文:Random Access Memory,简称:RAM)、磁盘或光盘等。

[0135] 以上对本申请实施例进行了详细介绍,本文中应用了具体个例对本申请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申请的方法及其核心思想;同时,对于本领域的一般技术人员,依据本申请的思想,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申请的限制。

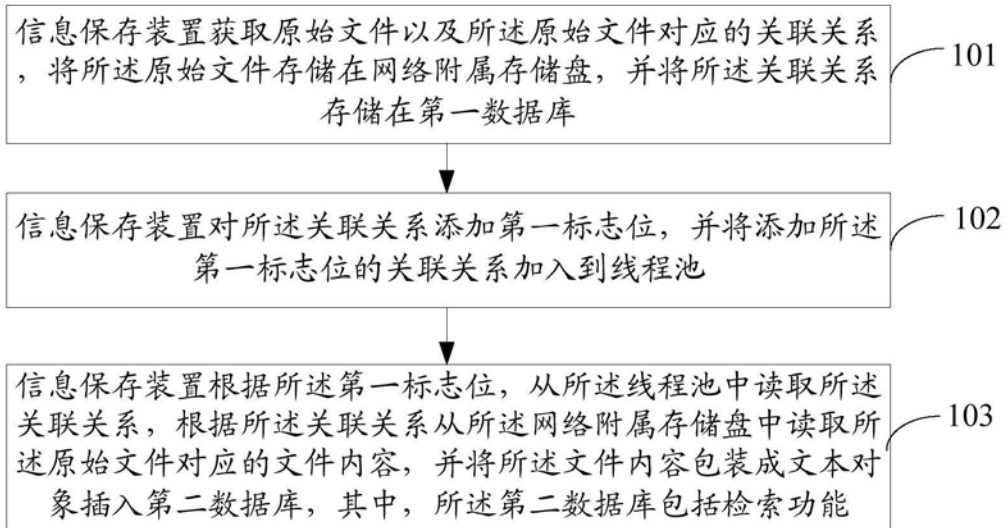


图1

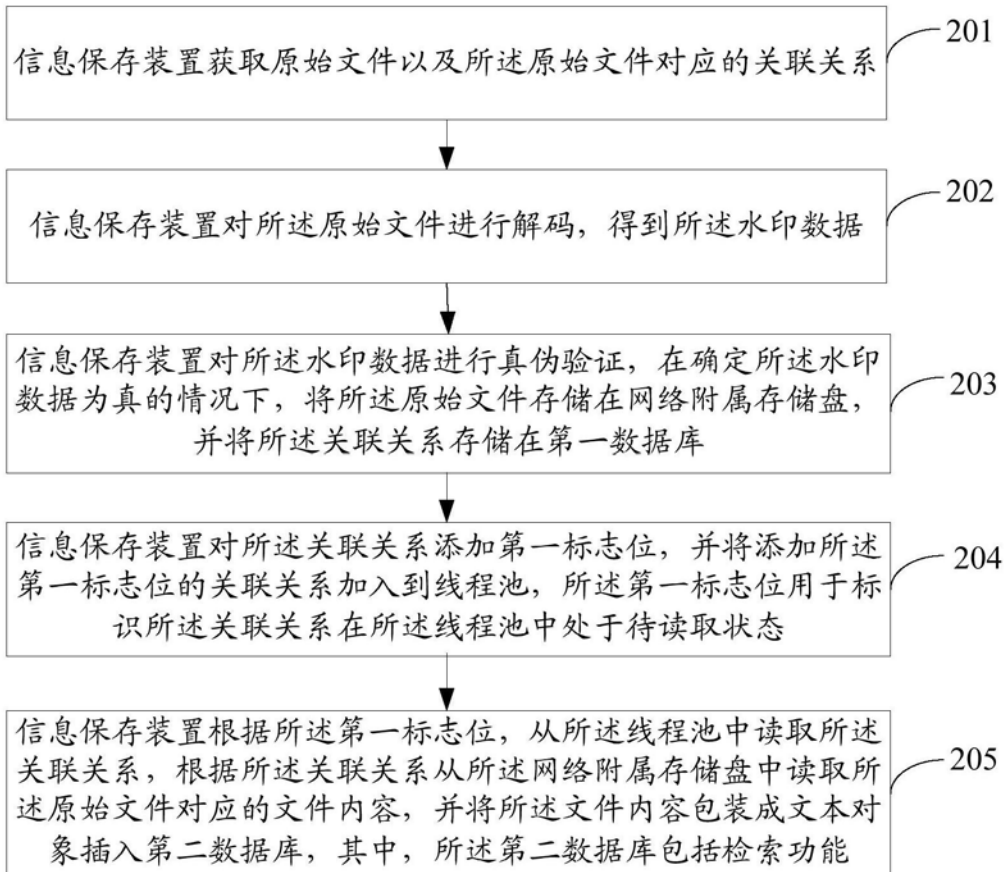


图2

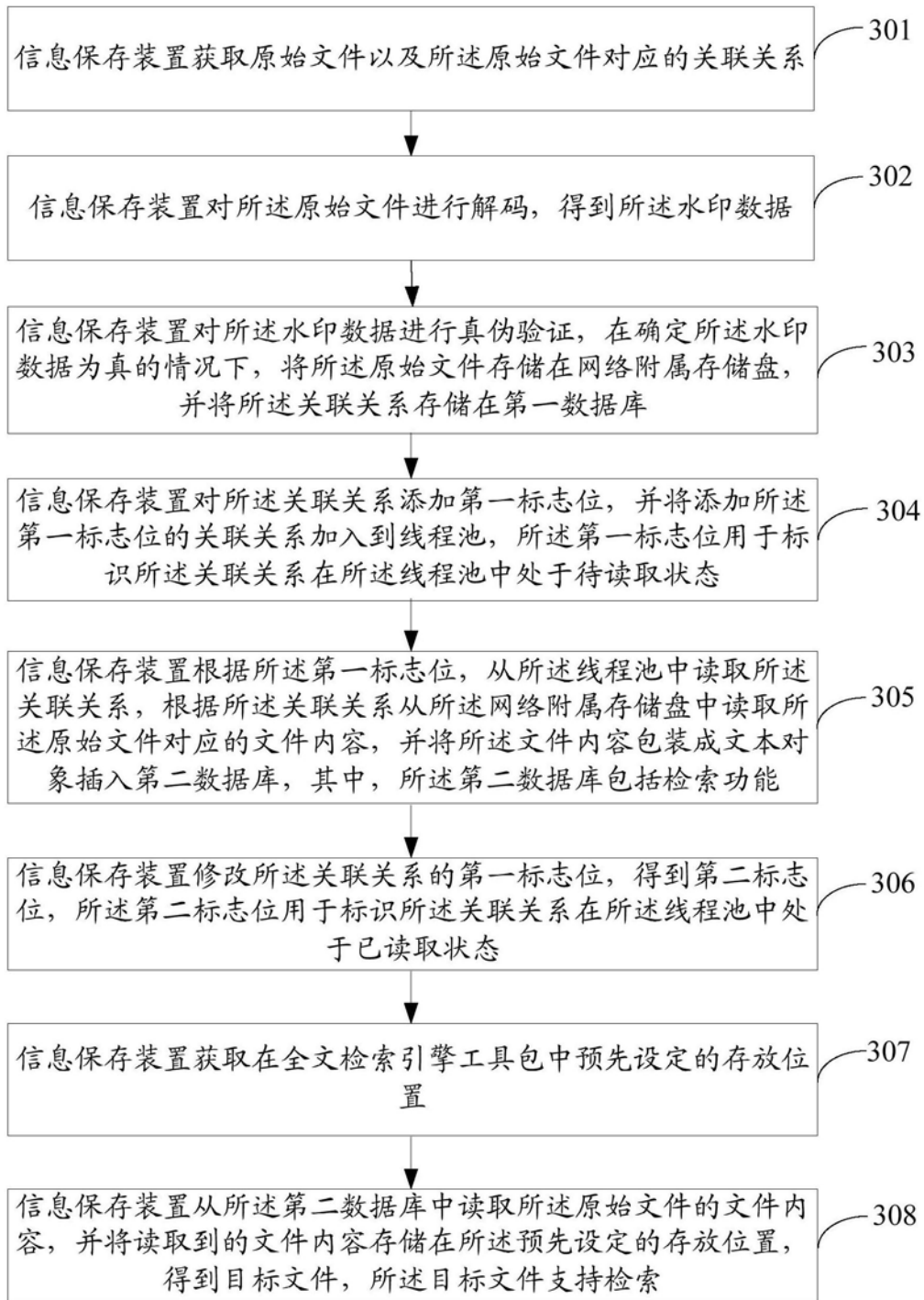


图3

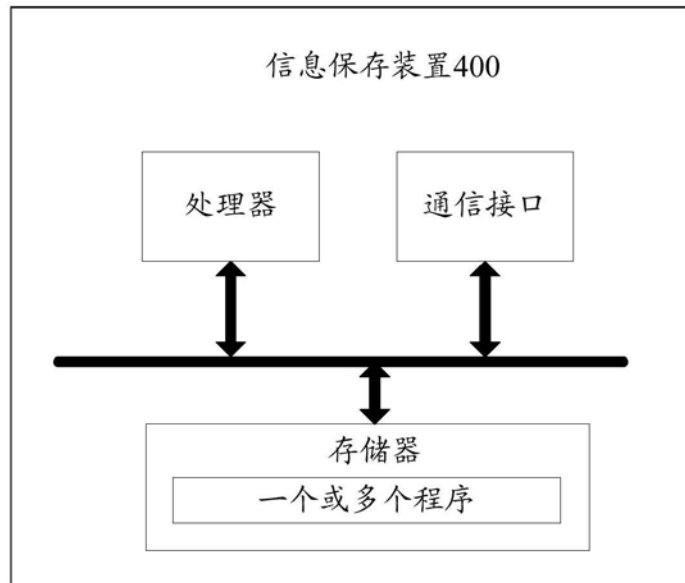


图4

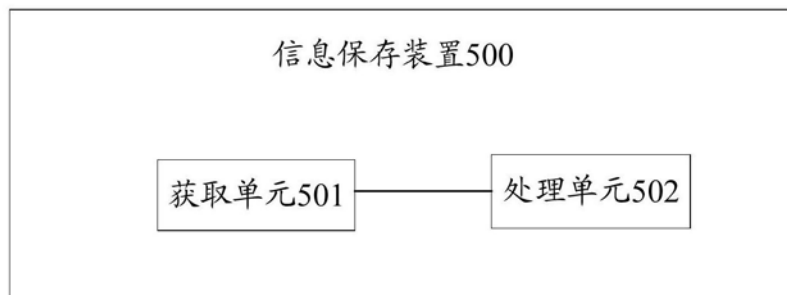


图5