



## (12) 发明专利

(10) 授权公告号 CN 1781073 B

(45) 授权公告日 2011.06.01

(21) 申请号 200480011807. X

(22) 申请日 2004.03.11

(30) 优先权数据

10/428,709 2003.05.02 US

(85) PCT申请进入国家阶段日

2005.11.01

(86) PCT申请的申请数据

PCT/EP2004/004012 2004.03.11

(87) PCT申请的公布数据

W02004/097620 EN 2004.11.11

(73) 专利权人 普驰信息技术有限公司

地址 美国科罗拉多州

(72) 发明人 斯科特·D·马斯迪 琼·L·米切尔

(74) 专利代理机构 北京润平知识产权代理有限公司

公司 11283

代理人 周建秋 王凤桐

(51) Int. Cl.

G06F 3/12 (2006.01)

(56) 对比文件

US 5491325 A, 1996.02.13, 说明书第4栏66行至第6栏65行, 图1, 图3, 图4.

US 6181814 B1, 2001.01.30, 全文.

CN 1050102 A, 1991.03.20, 全文.

审查员 王燕

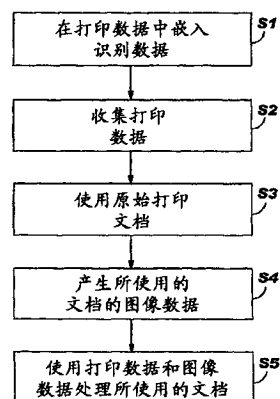
权利要求书 2 页 说明书 10 页 附图 2 页

(54) 发明名称

一种文档处理的方法和系统

(57) 摘要

联结打印文档的前端处理和后端处理的系统,方法和程序产品。尤其是,本发明需要收集可用于打印文档(即,通过持久地保存用于打印文档的数据)的打印数据,对使用后的文档进行成像,并且共同使用打印数据和图像数据处理文档。



1. 一种处理文档的方法,该方法包括步骤:  
收集可用于打印该文档的打印数据;  
在使用该文档后对该文档进行成像以产生图像数据;以及  
通过以下子步骤来处理该文档:  
检索该打印数据;  
比较该打印数据和该图像数据;  
如果在该打印数据和该图像数据之间存在匹配,则认证该文档;以及  
如果在该打印数据和该图像数据之间不存在匹配,则将该文档标记为被修改。
2. 如权利要求 1 所述的方法,其中打印数据包含图像数据、像素数据和语言数据中的至少一个。
3. 如权利要求 1 所述的方法,还包括在收集步骤之前在打印数据中嵌入文档标识数据的步骤。
4. 如权利要求 1 所述的方法,其中收集步骤包含存储打印数据和涉及文档的辅助数据。
5. 如权利要求 1 所述的方法,其中成像步骤包含存储图像数据和涉及文档的辅助数据。
6. 如权利要求 1 所述的方法,其中处理步骤包含从至少一个数据存储器中检索打印数据和图像数据中的至少一个。
7. 如权利要求 1 所述的方法,其中处理步骤包含基于文档的图像数据识别文档的打印数据。
8. 如权利要求 1 所述的方法,其中处理步骤包含比较使用后的文档的图像数据和打印数据。
9. 一种用于文档处理的系统,该系统包括:  
前端文档产生系统,包含:  
打印系统,用于基于文档产生数据产生可用于打印文档的打印数据;及  
数据收集器,用于持久地保存由打印系统产生的打印数据;及  
后端文档处理系统,用于在使用文档后处理该文档,该后端文档处理系统包含:  
访问打印数据的后端访问模块;  
成像器,用于对文档进行成像;及  
处理应用程序,用于检索该打印数据,比较该打印数据和该图像数据,如果在该打印数据和该图像数据之间存在匹配,则认证该文档,以及如果在该打印数据和该图像数据之间不存在匹配,则将该文档标记为被修改。
10. 如权利要求 9 所述的系统,其中前端文档产生系统还包含使得打印数据对后端文档处理系统可访问的前端访问模块。
11. 如权利要求 9 所述的系统,其中处理应用程序基于文档的图像数据识别文档的打印数据。
12. 如权利要求 9 所述的系统,其中处理应用程序比较使用后的文档和打印数据。
13. 如权利要求 9 所述的系统,还包括图像处理器。
14. 一种用于文档处理的系统,该系统包括:

后端文档处理系统,用于在使用文档后处理文档,包含:

用于访问由基于文档产生数据产生打印数据的前端文档产生系统产生的打印数据的第一访问模块;

成像器,用于在已经使用文档之后对文档进行成像;及

处理应用程序,用于比较该打印数据和该图像数据,如果在该打印数据和该图像数据之间存在匹配,则认证该文档,以及如果在该打印数据和该图像数据之间不存在匹配,则将该文档标记为被修改。

15. 如权利要求 14 所述的系统,其中前端文档产生系统包含:

打印系统,用于基于文档产生数据产生可用于打印文档的打印数据;及

数据收集器,用于持久地保存由打印系统产生的打印数据。

16. 如权利要求 14 所述的系统,其中前端文档产生系统还包含使得打印数据对后端文档处理系统可访问的前端访问模块。

17. 如权利要求 14 所述的系统,还包括图像处理器。

## 一种文档处理的方法和系统

### 技术领域

[0001] 本发明涉及文档处理,更具体的是涉及联结前端和后端文档处理。

### 背景技术

[0002] 无论电子通信如何演变,作为通信介质的正式文档的需求仍存在于许多行业。文档的内容和布局根据行业有所不同。例如,文档可能包含:信件,支票,定单,发票,收据,填写表格(例如,保险申请和完成的测试),证券等等。然而,文档的处理已取得进展,使得许多文档除了物理打印存在之外还具有数字生命。在需要大量文档的行业中,文档处理管理变得非常重要。文档处理管理通常可以被分为三个阶段:文档的前端产生,文档的使用,和所使用的文档的后端处理。每个阶段的内容根据行业可能有所不同。

[0003] 在文档的前端产生期间,文档产生数据作为各种文字(例如,ASCII),图形和图像存在,其通常从多个数据库中提取。数据可以用各种方式组织。在某些情况下,可以使用不可公开访问的专用格式和系统。如果文档被打印,许多打印机接受例如PostScript的文字格式,并且以操作时不存储数据的方式产生打印数据。可选地,某些打印机产生打印数据并且临时在一或多个缓冲区中存储它。然而,这个数据从不在前端产生阶段之外使用。在其它情况下,某些系统使用打印机后照相机或通过形成打印文档的另一个图像来记录打印之后的打印文档的质量检查系统。然而,这个数据从不在前端产生阶段之外使用。

[0004] 所打印的文档的归档需求可能例如因行业而有所不同。其中文档处理和归档具有重要作用的一个示例性行业是银行和金融行业。在这个行业中,例如客户借贷表或支票图像的重要数据通常被归档,使得产生哪些文档的记录存在。具有某种形式的归档文档通常可用于客户支持操作,使得客户支持代表可以检查哪些文档被发送到客户,从客户接收或返回到客户(例如,已注销支票)。这些文档的归档可包含保存文本数据,或打印就绪页,或组合(例如,某些具有选定文本数据的打印就绪页在例如IBM的ContentManager,On Demand的数据库中是常见的)。相反,每个页的像素数据,即在该页上使用的实际像素图像甚至不能临时保存。

[0005] 为了利于在后端处理打印文档期间处理和归档存储器,即在其想要的使用之后,许多组织通过扫描对所接收的所使用文档进行成像。例如,在保险行业中,某些公司扫描所有接收的信件。接着,把信函,申请表格,报告等等处理为图像以便处理。通常通过光学字符识别(OCR)程序将打印在这些文档上的信息转换为文本数据,以便能够进行文本搜寻和数据挖掘,并且辅助索引。当不使用OCR时,可以实现工作强度高和费时的人工数据键入。无论如何,显著时间和人力时常被消耗在作为所使用文档的后端处理的一部分的建立索引,对帐,出错检查,和欺诈检测。

[0006] 文档处理管理的传统方案的一个问题是前端产生数据不被用于后端处理数据。即使在前端文档产生数据与后端处理存在于相同组织中时也是如此。然而,更通常地,因为前端和后端处理不存在于相同组织,所以问题存在。例如,在银行和金融行业中,支票可以由大量机构发出并且被同样大量和独立的机构兑现。对于支票的结算,银行机构通常通宵快

速邮递支票图像的 CD-ROM 给其较大商业客户。某些机构人工地比较支票和其文本数据。在这种情况下,除非兑现银行正好已签写支票,否则非常不可能访问前端处理数据以检测错误。当前不存在这样的服务,其打印支票,并且借助 (leverage) 原始数据以通过比较每个兑现支票和打印的支票来确保支票兑现的准确性。作为另一个例子,接收并且扫描所使用的文档的保险公司时常具有由外部第三方产生的文档,使得用于打印文档的文档信息是不可访问的。过去,当这些操作发生在不同公司时,没有在接收方将文档的前端产生与后端扫描版本联系起来的方式。

[0007] 其中前端产生和后端处理的分离产生问题的另一个示例性行业是测试行业。在这个行业中,测试小册子通常被分节打印和汇编,使得一个组中的每个测试具有唯一排序的问题。使用之后,测试小册子被再次分成节,扫描并且分别发送到记录员。这个处理耗时并且冗长。另外,如果计分有问题,则纸测试小册子在仓库中被归档不同量的时间。在仓库中寻找特定的所使用测试小册子也是耗时并且工作强度高的。当前,没有关联测试的前端产生和后端评分和存档处理的方式。

[0008] 考虑到上述问题,本领域存在联结文档的前端和后端处理的需要。

## 发明内容

[0009] 本发明提供联结文档的前端处理和后端处理的系统,方法和程序产品。尤其是,本发明需要收集可用于打印文档(即,通过持久地保存用于打印文档的数据)的打印数据,对使用后的文档进行成像,并且共同使用打印数据和图像数据处理文档。

[0010] 本发明的第一方面涉及一种处理文档的方法,该方法包括步骤:收集可用于打印该文档的打印数据;在使用该文档后对该文档进行成像以产生图像数据;以及通过以下子步骤来处理该文档:检索该打印数据;比较该打印数据和该图像数据;如果在该打印数据和该图像数据之间存在匹配,则认证该文档;以及如果在该打印数据和该图像数据之间不存在匹配,则将该文档标记为被修改。

[0011] 本发明的第二方面涉及一种用于文档处理的系统,该系统包括:前端文档产生系统,包含:打印系统,用于基于文档产生数据产生可用于打印文档的打印数据;及数据收集器,用于持久地保存由打印系统产生的打印数据;及后端文档处理系统,用于在使用文档后处理该文档,该后端文档处理系统包含:访问打印数据的后端访问模块;成像器,用于对文档进行成像;及处理应用程序,用于检索该打印数据,比较该打印数据和该图像数据,如果在该打印数据和该图像数据之间存在匹配,则认证该文档,以及如果在该打印数据和该图像数据之间不存在匹配,则将该文档标记为被修改。

[0012] 本发明的第三方面涉及一种用于文档处理的系统,该系统包括:后端文档处理系统,用于在使用文档后处理文档,包含:用于访问由基于文档产生数据产生打印数据的前端文档产生系统产生的打印数据的第一访问模块;成像器,用于在已经使用文档之后对文档进行成像;及处理应用程序,用于比较该打印数据和该图像数据,如果在该打印数据和该图像数据之间存在匹配,则认证该文档,以及如果在该打印数据和该图像数据之间不存在匹配,则将该文档标记为被修改。

[0013] 本发明的第四方面涉及一种计算机程序产品,包括在其中体现有计算机可读程序代码的计算机可用介质,该程序产品包括:被设置成用来基于文档产生数据产生可用于打

印文档的打印数据的程序代码；及被设置成用来持久地保存由被设置成用来产生打印数据的程序代码产生的打印数据的程序代码。

[0014] 本发明的第五方面涉及一种计算机程序产品，包括在其中体现有计算机可读程序代码的计算机可用介质，该程序产品包括：被设置成用来收使用后的文档的成像产生图像数据的程序代码；被设置成用来访问由基于文档产生数据产生可用于打印文档的打印数据的前端文档产生系统产生的打印数据的程序代码；及被设置成用来使用打印数据和图像数据处理文档的程序代码。

[0015] 通过下面结合本发明优选实施例的更具体描述可以全面地理解本发明的上述及其它特征。

## 附图说明

[0016] 参照下面附图详细描述本发明的实施例，其中类似标记表示类似单元，并且其中：

[0017] 图 1 示出了图解根据本发明的文档处理管理环境的模块图。

[0018] 图 2 示出了根据本发明的文档处理管理的流程图。

[0019] 只为了清楚，下面描述包含下列标题：I. 文档处理管理环境概述；II. 前端文档产生系统；III. 后端文档处理系统；IV. 操作；V. 示例性使用 - 处理应用程序；及 VI. 结论。

[0020] I. 文档处理管理环境概述

[0021] 参照附图，图 1 是图解本发明的文档处理管理环境 10 的模块图。文档处理管理环境 10 一般包含三个阶段：文档前端产生阶段 12，文档使用阶段 14，及所使用文档的后端处理阶段 16。

[0022] 参照使用阶段 14，将针对银行和金融行业中使用的例如支票的文档来说明本发明。然而，如下所述，本发明适用于各种行业并且不应该限于任何具体行业。使用阶段 14 可以包含文档的简单传送，或可以包含填写空白空间，标记文档，加入签名和 / 或印记，和 / 或改变文档。

[0023] 继续参考图 1，文档在环境 10 中的一般移动包含由前端文档产生系统 20 在前端产生阶段 12 产生原始文档 22，用户在使用阶段 14 使用文档 22，及接着由后端文档处理系统 21 在后端处理阶段 16 进行处理。下面会更详细地描述每个系统 20, 21。能够明白，每个阶段的内容根据行业有所不同。作为概述，本发明以电子方式捕捉和保存可用于打印文档的打印数据 28，即，在阶段 22，并且接着使得该数据在后端处理阶段 16 可访问。在这个方式中，后端处理能够使用用于实际产生原始文档的打印数据，和当所使用的文档被扫描时得到的图像数据。另外，可以使后端图像数据可用于前端产生系统 20。

[0024] II. 前端文档产生系统

[0025] 如上所述，在阶段 12，提供前端文档产生系统 20（此后“产生系统 20”）以产生原始文档 22。应认识到，虽然已经说明了一个产生系统 20，然而若干产生系统 20 可以存在于环境 10 内。例如，在银行和金融行业中，存在诸如个人，公司，例如 PayChex® 的工薪管理公司的若干支票签写实体。文档 22 的产生基于文档产生数据 24。文档产生数据 24 可以作为组合形成文档 22 的各种文本数据（例如，ASCII），图形和基本图像而存在。对支票而言，文档产生数据 24 可以包含例如：文本数据，例如支票金额，日期，支票发行者，收款

人,付款人等等;图形数据,例如验证水印,支票发行者标志等等;和基本图像,例如支票背景或其它图像。如本领域已知的,文档产生数据 24 可以从多个数据库(未示出)中提取。

[0026] 产生系统 20 包含打印系统 26,打印数据收集器 30,前端访问模块 34,处理应用程序 37 和其它系统部件 38。打印系统 26 包含任何现在已知或以后开发的用于产生打印数据 28 的机构,其可以被用于打印文档 22。打印系统 26 可以包含例如激光打印机,击打式打印机,行式打印机,点阵打印机,打字机,喷墨打印机,其它打印机类型,或其组合。在操作中,打印系统 26 得到文档产生数据 24,并且以若干现在已知或以后开发的方法中的任何方法来组织它以形成所期望的文档 22。这样,打印系统 26 产生打印数据 28。“打印数据”是打印系统 26 产生的任何可用于物理产生的打印文档 22 的数据,包含通常只临时或非持久地存储的数据,例如打印缓冲区中的像素数据。即,打印数据 28 包含打印就绪数据,其包含图像数据,像素数据(aka 像素)和语言数据中的至少一个,其由打印系统 26 从文档产生数据 24(原始数据)转换而成。于是,打印数据 28 可以包含打印内容的图像版本和编码文本版本。打印系统 26 可以包含实际打印文档 22 的机构,或可以包含产生打印数据 28 但不实际打印文档的处理器。因此,打印数据 28 不必被用于实际打印文档。

[0027] 打印数据收集器 30(此后“收集器 30”)被构造成捕捉和持久地(即,长期)保存打印数据 28。打印数据收集器 30 也可以保存辅助数据 29 和打印数据 28。在产生系统 20 的语境中,“辅助数据”可以包含对后端处理有用的任何行业特定数据。在银行和金融行业中,辅助数据 29 可以包含例如帐号,支票金额,索引信息,插入标识数据,原始打印数据流,文字信息,索引信息,击打式打印机字体的扫描版本,击打式打印机字体的理想版本,原始文档的 JBIG-2(联合二值图像专家组)压缩版本,原始文档的电子版本,有关向哪里发送后端版本以归档存储的指示,或特定行业所需的任何其他数据。打印数据 28(和辅助数据 29)可以被保存在例如前端存储器 32 中。存储器 32 可以包括任何现在已知或以后开发的数据存储系统,和/或传输介质,其包含磁介质,光学介质,随机访问存储器(RAM),只读存储器(ROM),数据对象等等。此外,存储器 32 可以驻留在包括一或多个类型的数据存储器的单个物理位置上,或分布于多个物理系统上。打印数据 28 可以通过为便于索引而期望的任何方式在存储器 32 中排序。例如,针对支票,可以按日期,支票号等等排序。如下所述,存储器 32 中的打印数据 28 也可以通过关键行(keyline)数据的唯一组合来进行初级关键字索引(primary-keyindexed),使得每个支票的图像数据 42 能够唯一地和其对应打印数据 28 配对,以便由处理应用程序 37,56 使用。

[0028] 收集器 30 也可以包含被构造成插入标识数据(未示出)到打印数据 28 的标识数据插入器 36。标识数据可以包含例如文档用户标识,特定于行业的文档标识符,产生系统标识符,后端图像数据的建议索引,水印,条码,运行报头(running header),运行注脚(running footer),用于认证的保密符号等等。针对产生系统标识符,应当认识到,虽然已经说明了一个产生系统 20,然而若干产生系统可以存在于环境 10 内。例如,在银行和金融行业中,可以存在例如 PayChex® 的若干支票签写厂商。每个产生系统可以插入以后能够用于识别源产生系统 20 的标识符到其相应打印数据 28 中。

[0029] 前端访问模块 34 包含被构造成使得打印数据 28 可被后端文档处理系统 21 访问的任何硬件和/或软件。前端访问模块 34 可以包含例如用于产生系统 20 和/或前端存储器 32 的安全登录软件包,使得数据被安全地发送的加密密钥提供器,在前端和后端处理将

在其它地方处理的情况下发送第三方目的信息,等等。

[0030] 其它系统部件 38 可以提供现在已知或以后开发的用于文档产生系统 20 或辅助服务的任何其他功能。一个图解性的其它系统部件 38 可以是打印数据 28 的数字压缩软件包。

[0031] 应当认识到,产生系统 20 可以采取除简单产生文档的系统以外的各种形式。例如,产生系统 20 可以采取打印收据和存储关于收据的打印数据 28 的销售点终端的形式。

[0032] III. 后端文档处理系统

[0033] 后端文档处理系统 21(此后“处理系统 21”)被设计为在使用阶段 14 之后接收使用的文档 40,并且对使用的文档 40 或根据使用的文档 40 执行各种处理任务中的任何任务。应当认识到,虽然已经说明了一个处理系统 21,然而若干处理系统 21 可以存在于环境 10 内。例如,在银行和金融行业中,存在诸如个人,公司,例如 PayChex<sup>®</sup> 的工薪管理公司的若干支票签写实体。在这个行业中,每个接收银行可以表示一个处理系统 21。

[0034] 每个处理系统 21 可以包含成像器 50,其包含图像处理器 52,后端访问模块 54,至少一个(并且通常为各种)处理应用程序 56,及任何现在已知或以后开发的处理所使用文档的其它系统部件 58。成像器 50 被构造成接收所使用文档 40 并且通过扫描所使用文档 40 来产生其图像。图像,其部分或图像的经处理部分在这里被称作“图像数据”42。图像数据 42 可以被存储在后端存储器 44 中。辅助数据 43 也可以和图像数据 42 一起保存。在处理系统 21 的语境中,辅助数据 43 可以包含对后续处理有用的任何行业特定数据。辅助数据 43 可以包含例如关于支票金额,收款人,付款人,发出银行,支票上的基底图像,支票兑现日期,个人兑现支票的图片,兑现银行的文字信息,或特定行业所需的任何其他数据。如同前端存储器 32 那样,后端存储器 44 可以包括任何现在已知或以后开发的数据存储系统和/或传输介质,其包含磁介质,光学介质,随机访问存储器(RAM),只读存储器(ROM),数据对象等等,并且可以驻留在包括一或多个类型的数据存储器的单个物理位置上,或被分布在多个物理系统中。应当认识到,如果在无需成像器 50 的情况下可得到图像数据 42,则成像器 50 可以被省略。

[0035] 图像处理器 52 能够执行任何图像数据相关的图像处理,例如光学字符识别(OCR),旋转校正,图像失真校正,纠错,图像尺寸定标,转换到二进制图像的阈值确定,转换二进制图像到较小灰度图像,色空间转换,磁性墨字符识别(MICR)等等。这个信息可以接着被存储成图像数据 42。后端访问模块 54 被构造成允许处理系统 21 访问由产生系统 20 产生的打印数据 28。访问模块 54 包含被构造成访问前端存储器 32 中的打印数据 28 的任何硬件和/或软件。后端访问模块 54 可以包含例如用于产生系统 20 和/或前端存储器 32 的安全登录软件包,使得数据被安全地发送的加密密钥提供器,在前端和后端处理将在其它地方处理的情况下发送第三方目的信息,等等。

[0036] 如下面更详细描述,处理应用程序 37,56 提供各种特定于行业的文档处理或分析工具中的任何工具。然而,根据本发明,通过处理应用程序 56,处理系统 21 能够使用打印数据 28 和图像数据 42 进行处理。如下面进一步讨论,这个功能的优点参照操作的描述和示例性使用会变得更明显。对于前端存储器 32 和后端存储器 44,应当认识到,存储器 32 能够位于产生系统 20 位置,并且由处理系统 21 的后端访问模块 54 根据需要访问。类似地,存储器 44 能够位于处理系统 21 位置,并且由产生系统 20 的前端访问模块 34 根据需



要访问。可选地,打印数据 28 的分组可以被传送到处理系统 21,用于与后端存储器 44 中存储的图像数据 42 进行比较。类似地,图像数据 42 的分组可以被传送到产生系统 20,用于与前端存储器 32 中存储的打印数据 28 进行比较。在另一个例子中,第三方可以被实现为从任何数量的产生系统 20 和 / 或处理系统 21 收集数据,并且控制数据的可访问性。例如,文档产生能够发生在若干产生系统 20 中,例如在传真机中,在多个提供商打印机上,或在多个本地打印机上。这个前端数据能够被第三方收集到单个源中,并且从而使得处理系统 21 容易访问。文档可以包含关于在哪里寻找原始打印文档的电子版本的嵌入其中的信息。数据的采集也可以相对于具有若干处理系统 21 的环境反向进行,以便产生系统 20 容易地访问。总之,使得“原始”前端打印数据 28 可用于处理图像数据 42。

[0037] 其它系统部件 58 可以包含例如排序机构。

[0038] IV. 操作

[0039] 组合参照图 1 和 2,现在描述基于本发明的文档处理管理。步骤 S1 表示可选步骤,其中文档标识数据被标识数据插入器 36 嵌入打印数据 28 中。例如,具有照片形式的收款人标识能够被嵌入支票中,用于当收款人要兑现支票时进行确认。在另一个例子,工薪管理公司的标识可以被嵌入打印数据 28 中,使得处理系统 21 知道将使用后端访问模块 54 访问若干打印数据存储器 32 的哪个。

[0040] 在步骤 S2,数据收集器 30 收集打印数据 28。步骤 S2 也可以包含存储打印数据 28 和涉及文档的辅助数据 29。

[0041] 在步骤 S3,原始文档 22 被用于其想要的目的。对支票而言,支票将被发送到收款人并且拿到他 / 她的银行进行兑现。在步骤 S4,所使用文档 40 被成像器 50 成像以产生图像数据 42。对支票而言,接收银行会对支票进行成像以产生其图像数据 42。成像步骤可以包含存储图像数据 42 和涉及文档的辅助数据 43。成像步骤也可以包含使用图像处理器 52 执行例如光学字符识别的图像处理。

[0042] 在步骤 S5,通过使用打印数据 28 和图像数据 42 由一或多个处理应用程序 37,56 处理所使用文档 40。步骤 S5 也可以包含从至少一个数据存储器,即存储器 32 和 / 或存储器 44 中检索打印数据 28 和图像数据 42 中的至少一个。一个示例性处理应用程序 37,56 包含基于所使用文档的图像数据 42 识别所使用文档 40 的打印数据 28。另一个示例性处理应用程序 37,56 包含比较所使用文档 40 的图像数据 42 和对应打印数据 28 以检测文档的变化。将在下面描述其它处理应用程序 37,56。

[0043] V. 示例性使用 - 处理应用程序

[0044] 现在描述按行业的若干示例性使用,和说明本发明的一起使用前端打印数据 28 和后端图像数据 42 的优点的相关示例性处理应用程序 37,56。然而,应当认识到,本发明可应用于除这里讨论的特定例子,处理应用程序和行业以外。因此,本发明不应限于下述的任何例子,处理应用程序和 / 或行业。另外,应当认识到,处理应用程序 37,56 可以在处理系统 21 和 / 或产生系统 20 中实现。

[0045] A. 银行和金融行业

[0046] 继续支票例子,支票经过两个极单独的处理:支票的产生和由例如银行的金融机构结算支票的单独处理。上述发明允许闭合打印的支票和结算的支票之间的工作流循环。尤其是,能够实现若干处理应用程序 37,56 (图 1),其利用本发明的使用前端打印数据 28

和后端图像数据 42 的优点。在支票产生期间,文档产生数据 24 被转换成打印系统 26 内的打印数据 28。支票通常受到复杂的控制,包含通过唯一支票号进行跟踪,允许上游软件,例如 IBM 的 Infoprint Workflow 和 InfoprintManager 跟踪和保证所有必要支票被适当重印和没有复制。在支票结算期间,同样复杂的处理读取支票的图像数据 42,验证关键支票信息,并且适当地排序支票,包含那些需要人工干预的处理。支票图像数据 42 能够被归档,使得“所清除的”图像能够被长期保存。

[0047] 本发明通过允许访问及处理前端打印数据 28 和后端图像数据 42,闭合这 2 个处理循环。处理应用程序 37,56 可以提供下列功能和相关优点,例如:

[0048] 1. 进行比较以匹配数据:可以提供比较器,用于比较图像数据 42 和打印数据 28,以便匹配对应数据。例如,能够对提供给银行的支票图像执行 OCR 以确定例如发行银行(支票产生系统)标识和支票号。这个信息提供可用于识别适当前端存储器 32 和检查打印数据 28 以进行匹配的特定于支票的关键字。在某些情况下,支票可能不在打印数据 28 中。例如,由个人签写的个人支票在呈现于接收银行之前不会在打印数据 28 中,在这种情况下,支票和对应图像数据 B 将以传统方式处理。相反,如果对于支票的图像数据存在匹配,则对应打印数据 28(文本数据和原始支票图像)可以被其它处理应用程序 37,56 用于例如支票验证和清算处理的功能。这个处理应用程序单独允许大量支票打印,排序和存储库服务的合并。

[0049] 2. 验证和自动欺诈/伪造检测:可以提供比较器以比较打印数据 28 和图像数据 42,以允许处理应用程序 37,56 进行各种欺诈,伪造和/或变更检测处理。对欺诈或变更而言,例如,如果成像支票的金额不匹配于打印支票的金额,则支票能够被转移以进行调查而不是清算。不匹配可能指示欺诈,伪造或 OCR/图像处理器 52 的误读。另外,如果打印数据 28 是二进制的,则它能够在处理系统 21 处被转换为具有多级的正确尺寸的图像。例如,如果支票被打印在具有预先打印的背景的纸票面上,则只有每英寸 600 像素的二进制图像可以被打印在纸张上,包含个人化和格式化的信息,例如支票号,付款人的姓名和地址,发行银行,收款人姓名和地址,作为数值的支票金额,填写的支票金额,及签名行。图像数据 42 可以是较低分辨率灰度图像,例如每英寸 100 像素。原始二进制每英寸 600 像素图像的灰度逼近能够通过累加每个 6x6 像素正方形中的黑色像素的数量,并且调整动态范围以匹配所扫描的图像来产生。当在所扫描的图像数据 42 上以不同颜色叠加时,任何显著的差异均表现为明显色彩差异。因此,支票的变更变得易于辨别。特别是对于支票号变更,由于打印数据 28 只包含实际打印的那些号码和每张支票上的确切数据,所以罪犯很难产生看上去合法但是欺诈性的支票。

[0050] 对复制支票而言,处理应用程序 37,56 会能够检测呈现的复制支票,在第一次清算时产生支票的打印数据 28 和图像数据 42,和匹配支票号的任何其他呈现支票。这将允许签名,及支票图像上的任何其它标记的比较。对伪造而言,邮寄打印支票和存放支票之间的时间可用于寻找和存储具有前端打印数据 28(即作为辅助数据 29)的期望接收方的签名。接着,能够在数月内比较图像数据 42 以检测签名的显著变化。

[0051] 总体上,自动欺诈/伪造/变更检测消除了对文档进行人工干预的必要。此外,因为通过与打印数据 28 比较来暴露错误,能够只进行一次 OCR 尝试而不是例如 5 次 OCR 尝试,所以处理系统 21 能够变得更加快速。另外,如果对处理应用程序 37,56 的访问足够快,则

虽然兑现者仍然在银行,也可以使银行 / 接收方知道问题。总之,伴随有图像数据 42 的打印数据 28 可以作为犯罪的证据被传递给警察。

[0052] 对于上述 1. 和 2.,应当认识到,在不同位置完成任务可能是有利的。例如,当匹配没有被检测到时,处理系统 21 可以返回图像数据 42 到产生系统 20(或前端存储器 32 的第三方拥有者)以查询错误匹配的合法原因。

[0053] 3. 确认支票清算:当通过 OCR 扫描(图像数据 42)时,支票跟踪号和支票金额现在可以通过参照打印数据 28 验证原始支票来确认。

[0054] 4. 允许打印和清算支票的长期存储:支票现在可以与其原始打印数据 28 匹配,使得打印和清算的支票的图像能够在长期归档中可得到。

[0055] 5. 提供例如状态查询的附加服务:因为能够通知产生系统 20 处理系统 21 已经接收一张支票,所以能够提供附加服务。在一个例子中,支票发行者能够发出状态查询以确切发现其哪个支票已经清算,访问 确认的支票金额及打印和清算的支票图像。这种查询能够被用于客户跟踪,管理可接收帐户和债务,及用于确认清算支票的金额。在另一个例子中,现在可以为银行客户提供有关现金流,盘点等等的最新信息。其它附加服务也是可能的。

[0056] 6. 自动支票重印:当支票未出示给银行(通常从发行日开始 60 天内)时,可能需要重印。因为产生系统 20 存储原始支票的打印数据 28,所以该支票的重印可在以后的时间进行。在这种情况下,打印系统 26 或标识数据插入器 36 可以在支票的某个位置处包含重印标记,例如"reprint#\_1"以作为灰度图像。这将允许重印支票与原始支票区分开。注意,重印能够在 60 天之后(或任何时间段后)自动进行,使得不浪费时间重新键入客户帐户信息,支票金额等等。注意,处理应用程序 37,56 能够自动注册原始支票图像为"无效(void)",使得它在超出有效日期后出现原始支票的情况下不能清算。处理应用程序 37,56 能够接着在后端存储器 44 中将支票的整个"族系(family)"组合在一起,例如,的打印原始支票,重印版本及任何出示的支票。

[0057] 7. 允许数据存储尺寸缩减及索引简化:在后端图像数据 42 匹配"完好(perfect)"前端打印数据 28 的情况下,较少的关于支票的信息可能需要存储。例如,前端打印数据 28 可以被认为是"原始"支票,其中图像数据 42 通常将被认为是原始的。在这种情况下,由于所有其它相关信息是冗余及已知的,所以例如支票的接收日期与时间的不太新的信息可能有必要。此外,由于支票的打印数据 28 可能以特定的有用顺序(例如,按支票号,按公司,按发票号或按日期)产生,所以支票的索引可更易于自动化。

[0058] 8. 提供数据质量优点:打印数据 28 可以标准图像格式产生,因此能够被用于例如查看,而无需了解产生技术,打印机,公司,用于产生它的程序或存储器。另外,由于打印数据 28 是无噪声的,所以它的压缩优于扫描产生的图像数据 42,并且可以具有出众的图像质量。另外,文本数据能够完好地从图像数据 42 中提取,因为 OCR 结果能够与打印数据 28 中已可得到的文字数据比较,因而允许测试,调整及离线改进 OCR 处理。

[0059] 9. 提供过程性质量保证:假定例如福利和社会安全支票的文档以例如按月的定期计划进行打印。为了跟踪过程符合性,前端存储器 32 能够标记未及时打印的支票。

[0060] 10. 改进会计:支票的打印数据 28 能够被传递到会计部门,使得例如帐号,金额,付款人等等的支票信息能够在支票清算之前被输入到必要的会计和征税系统。当支票清算时,能够从处理系统 21 发送清算通知,其中只有清算相关数据,例如清算日期。通过这种方

式,会计部门不必等待支票清算以便更新会计记录,这使得记录保持更加快速,并且可以增加现金流。

[0061] B. 标准化测试行业

[0062] 对于标准化测试行业,处理应用程序 37,56 可以提供下列功能及相关优点,例如:

[0063] 1. 特殊格式化:测试小册子通常被分节打印和装配,使得一个组中的每个测试具有唯一排序的问题。在其中填写测试的使用之后,测试小册子被再次分成节,用识别盖板扫描,并且单独发送到计分员。一旦被计分,每个测试小册子的结果必须针对每个测试接受者重新编辑。使用打印数据 28 和图像数据 42 的处理允许以支持这个处理的逐个小册子的方式进行信息的特殊格式化。例如,按需打印技术可以允许测试小册子被打印成具有个人测试接受者标识和测试页/问题顺序标识。标识可以包含例如条码。在这种情况下,因为打印数据 28 能够提供用于图像数据 42 的索引信息,所以重新编辑被简化。因此,由于每个测试小册子包含允许打印数据 28 的检索的信息,所以测试小册子不再需要用某些独立识别盖板成批扫描。

[0064] 2. 测试小册子的归档:在对计分有疑问的情况下,将纸测试小册子在仓库中归档不同量的时间。在仓库中寻找特定使用的测试小册子是耗时和工作强度高的。图像数据 42 和打印数据 28 的使用通过存储原始测试和扫描的已使用测试的电子版本来减轻这个负担。因此,代替寻找纸张,能够通过例如自动搜寻来检索电子版本。

[0065] 3. 自动计分优点:如果使用多个选择框,能够根据知道正确答案的打印数据 28 在每个问题只标记一个答案的情况下对返回的测试小册子进行自动计分。扫描图像能够与原始打印数据比较。对每个提问,用户只填写(希望如此)一个选择框。填写的选择框的位置能够由处理应用程序 56 确定,并且与和打印数据 28 一起接收的辅助数据 29 中提供的正确答案比较。如果不止一个选择框已经涂黑,则最黑的一个(假定其它涂黑为擦除)能够与正确答案比较。针对指定问题填写多个选择框的错误答案可能需要在被计分之前转移到人工进行判断。自动计分将允许对测试接受者进行更快反馈。如果不止一个答案被标记,则测试小册子的打印数据 28 和图像数据 42 以及正确答案的图像能够被呈现给人工以进行比较。如果正确答案在正确答案图像中被呈现为更大的彩色选择框,则正确答案表现为彩色轮廓的黑选择框。相反,错过的正确答案表现为实彩色选择框,并且错误答案表现为无彩色边缘的黑选择框。可选地,错误答案能够被指定不同彩色选择框。接着,如果不止一个答案是正确的,则接受具有正确颜色的任何标记。基于这种处理,来自产生系统 20 的个别化测试能够被存为打印的,并且在自动评分期间在运行时产生答案。在这种情况下,不存在要保护的答案纸。由于每个测试是个别化的,并且不产生和打印答案纸,所以通过偷窃答案纸进行的欺骗被阻碍。在前端信息不可用于后端计分处理的情况下,必须产生答案纸。由于联结了前端和后端文档处理,所以保持附加到前端数据的答案是可行的。作为增加的安全措施,前端数据不能被公开,直到测试日之后。正确和错误答案的颜色编码是和打印数据 28 一起存储并且在后端处理中使用的辅助数据 29 的例子。例如,船上的见习船员接受测试。通常,答案纸必须传真给考官,以便评出结果。对于这个方案,考官能够独自检查自动计分的测试,或和见习船员一起,并且不需要附具答案的单片纸。由于每个测试可以不同,所以这是有用的。

[0066] C. 保险行业

[0067] 对于保险行业,处理应用程序 37,56 可以提供下列功能及相关优点,例如:

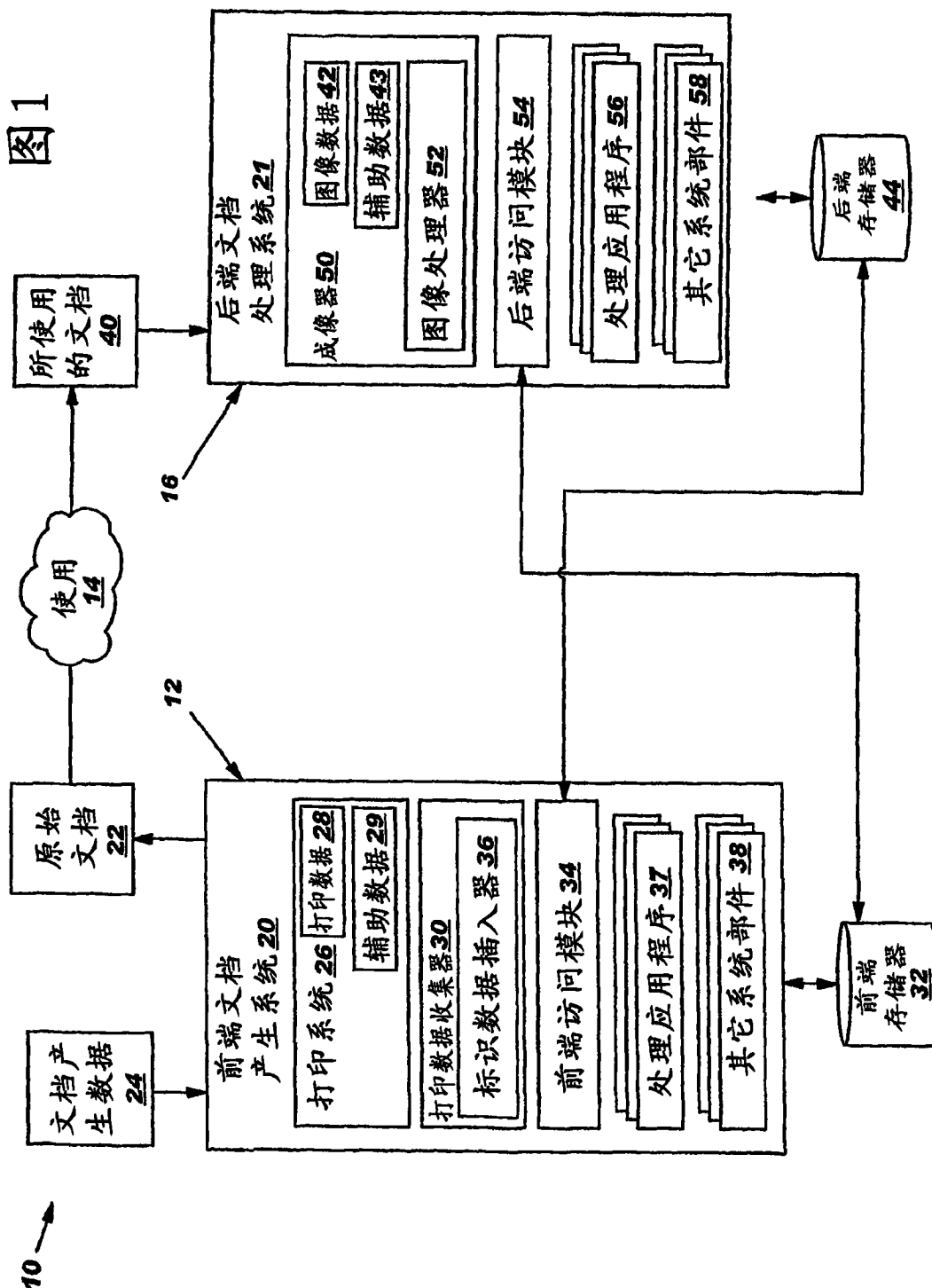
[0068] 1. OCR 精度检查:由第三方产生的信件可以包含曾经扫描和处理的标识符以允许访问前端打印数据 28。在这种情况下,图像数据 42 和打印数据 28 的比较可以允许检查从图像数据 42 得到的 OCR 数据。

[0069] 2. 欺诈检测:由于不具有前端存储器 32 中的对应数据,所以任何欺诈信件可以容易被检测。

[0070] VI. 结论

[0071] 在前面的讨论中,可以理解,所讨论的方法步骤可以通过例如计算机系统的中央处理单元、执行在存储器中存储的程序产品的指令的处理器来进行。可以理解,这里描述的各种设备,模块,机构和系统可以在硬件,软件或硬件和软件的组合中实现,并且可以划分为除了如图所示的之外。它们可以通过任何类型的计算机系统或适于执行这里描述的方法的其它设备来实现。硬件和软件的典型组合可以是具有计算机程序的通用计算机系统,当加载和执行该计算机程序时,其控制计算机系统,使得它执行这里描述的方法。可选地,能够使用包含用于执行本发明的一或多个功能任务的专用硬件的专用计算机。本发明也可以体现为计算机程序产品,其包括允许实现这里描述的方法和功能的所有特征,并且(当在计算机系统中加载时)其能够执行这些方法和功能。在当前上下文中,计算机程序,软件程序,程序,程序产品或软件是指以任何语言,代码或符号对一组指令的任何表述,其旨在使得系统具有信息处理能力以直接或在下面处理之后执行特定功能:(a) 转换到另一个语言,代码或符号;和/或(b)以不同材料形式再现。

[0072] 虽然结合上面概述的特定实施例描述了本发明,然而显然许多候选,修改和变化对本领域的技术人员是显而易见。因此,如上面提出的本发明的实施例是说明性的,而不是限制性的。可以在不偏离如下面权利要求定义的本发明的实质和范围的前提下进行各种变化。



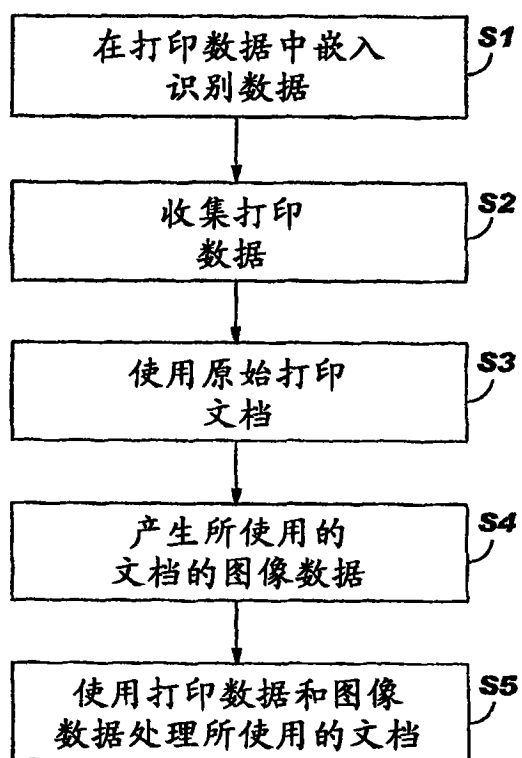


图 2