



# (12) 发明专利

(10) 授权公告号 CN 102591847 B

(45) 授权公告日 2015. 11. 25

(21) 申请号 201110391188. 3

第 29 行至第 34 栏第 23 行及说明书附图 22-26.

(22) 申请日 2011. 11. 25

审查员 刘洋

(30) 优先权数据

2010-262876 2010. 11. 25 JP

(73) 专利权人 佳能株式会社

地址 日本东京都大田区下丸子 3-30-2

(72) 发明人 河原聪史

(74) 专利代理机构 北京怡丰知识产权代理有限公司 11293

代理人 迟军

(51) Int. Cl.

G06F 17/21(2006. 01)

(56) 对比文件

CN 101520728 A, 2009. 09. 02, 全文.

US 6336124 B1, 2002. 01. 01, 说明书第 2 栏第 12-45 行及第 19 栏第 6-16 行.

US 6369909 B1, 2002. 04. 09, 说明书第 31 栏

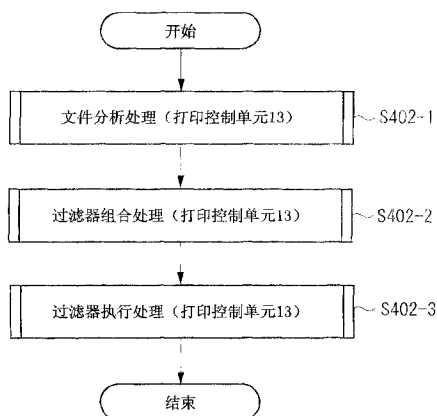
权利要求书1页 说明书10页 附图11页

(54) 发明名称

文档处理装置及文档处理方法

(57) 摘要

本发明提供文档处理装置及文档处理方法。该文档处理方法包括以下步骤：分析具有第一格式的输入文档文件；以及当经由所述分析输入文档文件的步骤分析出具有第一格式的输入文档文件包含有具有第二格式的附加文档文件时，设置以下两种不同的过滤器：用于将具有第二格式的附加文档文件转换成具有第一格式的文档文件的过滤器，以及用于将具有由第二格式转换来的第一格式的文档文件与所述输入文档文件中的具有第一格式的文档文件合并的过滤器。



1. 一种文档处理装置,该文档处理装置包括:

被构造成确定输入文档文件的格式是否是 PDF 格式的单元;

被构造成只有在确定所述输入文档文件的格式是 PDF 格式的时候、确定所述输入文档文件是否附加有另一文档文件的单元;

被构造成在确定所述输入文档文件附加有另一文档文件的情况下、确定所述另一文档文件的格式是否是 PDF 格式的单元;

被构造成从所述输入文档文件提取所述另一文档文件的单元;

被构造成生成描述至所述输入文档文件和所提取的另一文档文件的文件路径的列表文件的单元;

设置单元,其被构造成,在确定所述另一文档文件的所述格式不是 PDF 格式的情况下,设置以下两种不同的过滤器:用于将所述另一文档文件的所述格式转换成 PDF 格式的过滤器,以及用于将转换得到的所转换的文档文件与所述输入文档文件合并的过滤器;以及

被构造成将所生成的列表文件作为流传输至所设置的过滤器的单元。

2. 一种文档处理方法,所述方法包括以下步骤:

确定输入文档文件的格式是否是 PDF 格式;

只有在确定所述输入文档文件的格式是 PDF 格式的时候,确定所述输入文档文件是否附加有另一文档文件;

在确定所述输入文档文件附加有另一文档文件的情况下,确定所述另一文档文件的格式是否是 PDF 格式;

从所述输入文档文件提取所述另一文档文件;

生成描述至所述输入文档文件和所提取的另一文档文件的文件路径的列表文件;

在确定所述另一文档文件的所述格式不是 PDF 格式的情况下,设置以下两种不同的过滤器:用于将所述另一文档文件的所述格式转换成 PDF 格式的过滤器,以及用于将转换得到的所转换的文档文件与所述输入文档文件合并的过滤器;以及

将所生成的列表文件作为流传输至所设置的过滤器。

## 文档处理装置及文档处理方法

### 技术领域

[0001] 本发明涉及通过使用过滤器来处理文档的文档处理装置、文档处理方法及其程序。

### 背景技术

[0002] 用于连接多个可拆卸的模块的传统方法包括过滤器管线系统 (filter pipeline system), 在该过滤器管线系统中, 将模块作为过滤器处理并利用管线连接。存在用于在过滤器之间传输数据的各种方法, 例如, 用于将数据作为流 (streaming) 而依次传输的方法、以及将数据作为电子文件集中传输的方法。这些传输数据的方法的共同之处在于将来自某一过滤器的输出数据作为对后续过滤器的输入数据进行处理。

[0003] Microsoft XPS 过滤器管线是传统过滤器管线系统的示例。Microsoft XPS 过滤器管线可以由例如 XPS 布局过滤器和 XPS 打印过滤器构成。XPS 布局过滤器转换作为输入数据的 XPS (XML 纸规范) 文件的布局, 并输出具有转换后的布局的 XPS 文件。XPS 打印过滤器将布局后的 XPS 文件转换成 PDL (页面描述语言) 文件, 该 PDL 文件用作打印装置可识别的控制命令。

[0004] 当至该系统的输入数据是 XPS 文件时, 如果例如将 PDF (可移植文档格式) 文件作为输入数据处理, 则还可以通过建立由 PDF 布局过滤器和 PDF 打印过滤器构成的过滤器管线系统来实现该系统的目的。PDF 布局过滤器转换作为输入数据的 PDF 文件的布局, 并输出具有转换后的布局的 PDF 文件。PDF 打印过滤器将布局后的 PDF 文件转换成 PDL 文件。以下, 将上述系统称为 PDF 打印管线系统。

[0005] 过滤器管线系统具有如下原理: 来自某一过滤器的输出数据成为对后续过滤器的输入数据。基于该原理, 传统技术根据输入文件动态地连接所需要的过滤器。具体地说, 在上述示例中, 当输入数据是 XPS 文件时, 使用 XPS 布局过滤器和 XPS 打印过滤器。当输入数据是 PDF 文件时, 使用 PDF 布局过滤器和 PDF 打印过滤器。

[0006] 具体地, 以下将说明用于动态连接过滤器的传统技术。

[0007] 例如, 日本特开 2002-351709 号公报论述了如下一种技术: 解决输入数据的问题, 使得仅在数据转换时确定输出数据格式。具体地说, 当来自某一过滤器的输出数据的文件格式 (格式类型) 不同于假定的文件格式时, 将用于将该文件格式转换成假定的文件格式的过滤器动态地连接到管线的末端。该技术动态地连接与输入数据相对应的数据转换过滤器, 以获得期望的输出文件。

[0008] 日本特开 2005-258782 号公报论述了一种用于动态地确定用于处理结构化文档文件 (输入数据) 的所需要的过滤器的技术。具体地说, 该技术执行以下处理。首先检查要经受处理的文档数据的文件格式是否与后续过滤器所需的输入文件格式相一致。随后, 当文件格式彼此不一致时, 将文档数据的文件格式转换成后续过滤器所需的文件格式, 然后将文档文件传输给后续过滤器以执行处理。当各过滤器需要多个不同的数据格式时, 可以动态地配置过滤器, 以进一步提高处理效率。

[0009] 作为典型的结构化文档文件, PDF 文件和 XPS 文件是公知的。这些文件的数据格式在 URL(统一资源定位符)上公开开放。上述结构化文档文件被构造成,使得能够在其中嵌入(或者附加)其他的文档文档。因此,可以将具有不同的文件格式的多个文档文件嵌入一文档文件中。以下将这种结构化文档文件称为组合结构化文档文件。当结构化文档文件是 PDF 文件时,将其称为组合 PDF 文档文件。

[0010] 在上述 Microsoft XPS 管线系统和 PDF 打印管线系统中,存在如下情况:需要将组合结构化文档文件与其中包含的附加文档文件一起打印。各过滤器仅适用于有限的输入和输出文件。例如,在上述 PDF 打印管线系统中,PDF 布局过滤器将 PDF 文件作为输入和输出数据进行处理。然而,尽管 PDF 布局过滤器可以将 PDF 文件作为输入数据处理,但是 PDF 布局过滤器不能必要地处理在该 PDF 文件中包含的附加文档文件的格式。例如,在其中包含有附加 MS Office 文件的组合 PDF 文件的情况下,PDF 布局过滤器必须能够处理该 MS Office 文件。具体地说,有必要从组合 PDF 文件中提取 MS Office 文件,将该 MS Office 文件转换成 PDF 文件,并将转换后的 PDF 文件与组合 PDF 文件(即提取源)合并。由于过滤器的内部处理变得复杂,因此通过使用诸如 PDF 布局过滤器的单个过滤器来实现该功能是不现实的。

[0011] 另一方面,能够准备单个过滤器来将 MS Office 文件转换成 PDF 文件,并将多个 PDF 文件合并成一个 PDF 文件。因此,通过准备拼合上述过滤器的组合过滤器,能够使得组合 PDF 文件可被处理。然而,尽管在本示例中将 MS Office 文件作为附加文档文件进行了说明,但是对于可附加于组合 PDF 文件的文件格式没有限制。因此,由于组合的巨大数量,准备拼接多个上述过滤器的组合过滤器是不现实的。

[0012] 通过使用传统技术不能够解决这一问题。日本特开 2002-351709 号公报中论述的技术在输出文件格式与指定的文件格式不同的情况下通过动态增加模块来递归调用所需的模块。该技术仅在输出文档的数据格式与假定的数据格式不同的情况下是有效的。因此,对于文档文件(例如组合结构化文档文件)包括具有不同文件格式的附加文档文件的情况,该技术是不适用的。日本特开 2005-258782 号公报中论述的技术将整个文档文件转换成后续过滤器可处理的文件格式。因此,对于输入文档文件(例如组合结构化文档文件)包括具有不同文件格式的附加文档文件的情况,该技术是不适用的。

[0013] 此外,日本特开 2002-351709 号公报和 2005-258782 号公报中论述的技术不具有用于指定要应用于组合结构化文档文件中的特定附加文档文件的处理的手段。因此,这些技术不能够配置对组合结构化文档文件所需的过滤器进行动态连接的组合过滤器。

## 发明内容

[0014] 根据本发明的一方面,提供了一种文档处理装置,该文档处理装置包括:分析单元,其被构造成分析具有第一格式的输入文档文件;以及设置单元,其被构造成,当所述分析单元分析出具有第一格式的所述输入文档文件包含有具有第二格式的附加文档文件时,设置以下两种不同的过滤器:用于将具有第二格式的附加文档文件转换成具有第一格式的文档文件的过滤器,以及用于将具有由第二格式转换来的第一格式的文档文件与所述输入文档文件中的具有第一格式的文档文件合并的过滤器。

[0015] 通过下面参照附图对示例性实施例的详细说明,本发明的其他特征和方面将变得

清楚。

### 附图说明

[0016] 包括在说明书中并构成说明书的一部分的附图，例示本发明的示例性实施例、特征和方面，并与文字说明一起，用于解释本发明的原理。

[0017] 图 1 是例示信息处理系统的示例性结构的框图。

[0018] 图 2 概念性地例示了从发出打印命令到进行打印的处理。

[0019] 图 3 例示了根据示例性实施例的打印系统的模块的结构。

[0020] 图 4 是例示根据本示例性实施例的整体处理的流程图。

[0021] 图 5 例示了在本示例性实施例中使用的示例性过滤器设置文件。

[0022] 图 6 例示了在示例性实施例中使用的示例性过滤器输入 / 输出定义文件。

[0023] 图 7 是例示过滤器中的示例性处理系统的框图。

[0024] 图 8 示意性地例示了在本示例性实施例中使用的示例性列表文件。

[0025] 图 9 概念性地例示了用于在过滤器之间传输数据的方法。

[0026] 图 10 是例示在打印控制单元内执行的文件转换处理的概要的流程图，其属于图 4 中例示的流程图。

[0027] 图 11 是例示在打印控制单元内执行的文件转换处理的详情的流程图，其属于图 4 中例示的流程图。

### 具体实施方式

[0028] 以下，将参照附图详细说明本发明的各种示例性实施例、特征和方面。

[0029] 以下将说明第一示例性实施例。在本示例性实施例中，将根据本发明的格式转换方法应用于过滤器管线系统中的客户端个人计算机 (PC)。以下将基于如下打印系统来说明本示例性实施例，在该打印系统中将组合结构化文档文件转换成要经受打印的 PDL 文件，并且由打印装置处理该 PDL 文件然后将该 PDL 文件打印在纸上。然而，本发明不限于打印系统，而是可以适用于对组合文档文件的文件格式进行转换的任何系统。类似地，用于转换格式的方法不仅适用于客户端 PC，而且适用于打印装置。用于转换格式的方法还适用于提供 Web 服务的服务器，以及适用于提供云服务的情况。

[0030] 图 1 是例示应用了根据本发明的文档处理装置的信息处理系统的示例性结构的框图。

[0031] 参照图 1，中央处理单元 1 从设置在与信息处理系统连接的介质读取装置 6 中的存储介质（例如软盘 (FD)、光盘只读存储器 (CD-ROM) 和集成电路 (IC) 存储卡）载入上述程序和相关数据。然后，中央处理单元 1 通过使用从辅助存储装置 3 向主存储单元 2 载入的系统程序 and 应用程序，处理从输入装置 4 输入的信息，并向输出装置 5 或打印装置 7 输出所述信息。在本示例性实施例中，输出装置 5 是与打印装置 7 不同的显示装置。输入装置 4 包括键盘、指示设备等。辅助存储装置 3 可以包括硬盘、磁光盘或者二者的组合。本发明在不考虑上述装置是否经由网络连接的情况下也是可行的。

[0032] 图 2 概念性地例示了如下处理：将打印处理控制程序和相关数据从介质读取装置 6 载入中央处理单元 1、从输入装置 4 输入打印命令、并将数据发送给打印装置 7 然后打印

在纸上。应用 10 在操作系统 (OS) 9 的控制下工作。

[0033] 图 3 例示了本发明的最显著特征。参照图 3, 在根据本发明的打印系统 (以下说明) 中, OS 9 控制客户端 PC 和应用 10。OS 9 经由 USB 接口与打印装置 7 (以下说明) 连接。OS 9 上的应用软件 11 控制打印装置 7。在本示例性实施例中, 将应用软件 11 称为打印应用。

[0034] 用户经由用户接口单元 12, 输入诸如对根据本发明的打印装置 7 的设置的各各种打印设置, 并指示打印装置 7 开始打印。

[0035] 打印控制单元 13 从打印应用 11 接收输入文档文件、打印设置文件以及过滤器设置文件, 生成打印装置 7 可处理的 PDF 文件和打印指令文件, 并将这些文件发送给打印装置 7。打印控制单元 13 包括根据本发明的打印系统中的主机装置侧的重要处理单元。

[0036] 过滤器控制单元 14 控制各种过滤器的处理顺序和输入输出数据。

[0037] 格式转换过滤器组 15 是一组具有文件格式转换功能的格式转换过滤器。格式转换过滤器组 15 包括 Office-to-PDF 转换过滤器 16、XPS-to-PDF 转换过滤器 17 以及 PDF-to-PDL 转换过滤器 18。Office-to-PDF 转换过滤器 16 将 MS Office 文件 (Word、Excel) 转换成 PDF 文件。XPS-to-PDF 转换过滤器 17 将 XPS 文件转换成 PDF 文件。PDF-to-PDL 转换过滤器 18 由 PDF 文件和打印设置文件, 生成打印装置 7 可打印的 PDL 文件和打印指令文件。尽管在以上内容中, 作为示例仅提及了三种不同的转换过滤器, 但是可以将具有文件格式转换功能 (将文件格式转换成另一文件格式的功能) 的任何转换过滤器添加到格式转换过滤器组 15 中。

[0038] 编辑过滤器组 19 是一组具有编辑输入文档文件的功能的编辑过滤器。编辑过滤器组 19 例如包括 PDF 布局过滤器 20 和 PDF 合并过滤器 21。PDF 布局过滤器 20 将输入的 PDF 文件转换成适合于 N-up (在一张片材上布局 N 页)、书本装订、海报打印等的 PDF 文件。PDF 合并过滤器 21 将多个输入的 PDF 文件合并成一个 PDF 文件。

[0039] Office-to-PDF 转换过滤器 16、XPS-to-PDF 转换过滤器 17、PDF-to-PDL 转换过滤器 18、PDF 布局过滤器 20 和 PDF 合并过滤器 21 具有各自的输入和输出文件规范。打印控制单元 13 用作连接这些过滤器的过滤器管线系统。

[0040] 作为 OS9 的功能, 数据收发单元 22 经由 Centronics 接口向打印装置 7 发送 / 从打印装置 7 接收数据。打印控制单元 13 经由数据收发单元 22, 向打印装置 7 发送所生成的 PDL 数据和打印指令文档。

[0041] 打印装置 7 根据从所连接的主机装置接收到的 PDL 文件和打印指令文档, 执行打印处理。

[0042] 以下将参照图 4 中的流程图, 说明根据本示例性实施例的整体处理流程。

[0043] 打印应用 11 对经由用户接口单元 12 来自输入装置 4 的输入文档文件、打印设置文件和过滤器设置文件进行识别。输入文档文件要经受打印。打印设置文件描述诸如 N-up 打印的打印方法。过滤器设置文件指定打印控制单元 13 进行处理所需的过滤器。在步骤 S401 中, 打印应用 11 向打印控制单元 13 发送这些文件和打印开始指令, 以开始打印。

[0044] 在步骤 S402 中, 打印控制单元 13 向过滤器控制单元 14 发送输入文档文件和打印设置文件。过滤器控制单元 14 载入基于过滤器设置文件的处理所需的过滤器, 基于过滤器设置文件执行处理, 并向数据收发单元 22 发送 PDL 数据 (输出文件) 和打印指令文档。以

下将详细说明打印控制单元 13 中的处理。

[0045] 在步骤 S403 中,在接收到来自打印控制单元 13 的指令后,数据收发单元 22 向打印装置 7 发送 PDL 文件(输出文件)和打印指令文档。在步骤 S404 中,打印装置 7 打印所接收的 PDL 文件。

[0046] 在详细说明打印控制单元 13 中的处理(步骤 S402)之前,在以下内容中说明一些要素。

[0047] 图 5 例示了示例性过滤器设置文件。过滤器设置文件 501 描述了如下过滤器设置,根据该过滤器设置,Office-to-PDF 转换过滤器 16 处理 MS Office 文件,PDF 布局过滤器 20 处理 PDF 文件,PDF-to-PDL 转换过滤器 18 处理 PDF 文件。在接收到来自打印控制单元 13 的指令之后,过滤器控制单元 14 按过滤器设置文件 501 中描述的顺序,执行相应的过滤器处理,以获得最终的输出数据。过滤器设置文件 502 描述了如下的过滤器设置,根据该过滤器设置,仅 PDF-to-PDL 转换过滤器 18 将 PDF 文件转换成 PDL 文件。

[0048] 以下将描述过滤器输入/输出定义文件。由于这些文件运行在过滤器管线系统上,所以 Office-to-PDF 转换过滤器 16、XPS-to-PDF 转换过滤器 17、PDF-to-PDL 转换过滤器 18、PDF 布局过滤器 20 和 PDF 合并过滤器 21 分别需要输入和输出数据规范。因此,有必要针对各过滤器准备用于指定输入和输出数据的过滤器输入/输出定义文件。通过参照该过滤器输入/输出定义文件,过滤器控制单元 14 能够使诸如过滤器设置文件 501 和 502 的过滤器设置文件有效。此外,通过基于该信息更新诸如过滤器设置文件 501 和 502 的过滤器设置文件,能够动态地连接过滤器(以下将具体说明)。图 6 例示了示例性过滤器输入/输出定义文件。过滤器输入/输出定义文件 601 指定了 Office-to-PDF 转换过滤器的过滤器名称、格式转换的过滤器类型、MS Office 文件的输入文件格式以及 PDF 文件的输出文件格式。过滤器输入/输出定义文件 602、603 和 604 分别指定了 PDF 布局过滤器 20、PDF-to-PDL 转换过滤器 18 以及 PDF 合并过滤器 21。

[0049] 以下将说明 Office-to-PDF 转换过滤器 16、XPS-to-PDF 转换过滤器 17、PDF-to-PDL 转换过滤器 18、PDF 布局过滤器 20 和 PDF 合并过滤器 21 中的处理,以及运行在打印控制单元 13 中的过滤器控制单元 14 是如何控制这些过滤器的。

[0050] 以下将参照图 7 说明 Office-to-PDF 转换过滤器 16、XPS-to-PDF 转换过滤器 17、PDF-to-PDL 转换过滤器 18、PDF 布局过滤器 20 和 PDF 合并过滤器 21 中的处理。图 7 是例示根据本示例性实施例的各个过滤器中的处理系统的框图。输入处理单元 701 接收来自在前过滤器的输出数据,作为输入数据。过滤器处理单元 702 执行相应的过滤器处理,例如,Office-to-PDF 转换过滤器 16 将 MS Office 文件转换成 PDF 文件,而 PDF 布局过滤器 20 编辑 PDF 文件的布局。输出数据生成单元 703 生成反映过滤器处理单元 702 的处理的结果的输出数据。当连续生成数据时,在这种情况下,输出数据生成单元 703 将该数据作为流依次发送。当输出存储器或硬盘中的实体文件时,列表文件生成单元 704 可能生成列出至实体文件的链接信息的列表文件,然后将该列表文件作为流进行发送。

[0051] 图 8 示意性例示了示例性列表文件。参照图 8,列表文件 801 描述了四个链接的文件(Doc 1 至 Doc 4)以及它们的文件格式(FormatType)和文件路径(FilePath)。当列表文件是至某一过滤器的输入数据时,过滤器处理该列表文件中的文件格式(FormatType)与过滤器输入/输出定义文件中指定的输入文件格式(Input/FormatType)相一致的所有

的文档文件。

[0052] 以下将说明用于在过滤器之间传输数据的方法。图 9 概念性地例示了用于在过滤器之间传输数据的方法。参照图 9, MS Office 文件是至过滤器控制单元 14 的输入数据; Office-to-PDF 转换过滤器 16、PDF 布局过滤器 20 和 PDF-to-PDL 转换过滤器 18 处理该数据; PDF-to-PDL 转换过滤器 18 输出 PDL 文件作为最终数据。

[0053] 当过滤器控制单元 14 将 MS Office 文件(输入数据)作为流传输至第一过滤器(即 Office-to-PDF 转换过滤器 16)时, Office-to-PDF 转换过滤器 16 将 MS Office 文件转换成 PDF 文件,生成指定至所生成的 PDF 文件的文件路径的列表文件,并将该列表文件作为流,传输至过滤器控制单元 14。例如,假定 PDF 布局过滤器 20 将实体 PDF 文件存储在辅助存储装置 3(硬盘等)中,则将该 PDF 文件作为流再次发送是低效率的。与此不同的是,通过向后续过滤器仅传输针对所存储的 PDF 文件的列表文件,实现了高效的数据传输。过滤器控制单元 14 确定要作为流发送的数据是否是列表文件,并且各过滤器根据来自过滤器控制单元 14 的指令来执行处理。

[0054] 过滤器控制单元 14 将来自在前过滤器的输出数据作为输入数据,馈送给后续过滤器。因此,将列表文件作为输入数据,以流的形式传输给后续的 PDF 布局过滤器 20。PDF 布局过滤器 20 根据从过滤器控制单元 14 单独接收到的打印设置,布置在接收到的列表文件中指定的 PDF 文件的布局,然后输出布置后的 PDF 文件。PDF 布局过滤器 20 生成与 Office-to-PDF 转换过滤器 16 类似的列表文件,然后将该列表文件作为流传送给过滤器控制单元 14。过滤器控制单元 14 将列表文件,作为输入数据传输给后续 PDF-to-PDL 转换过滤器 18。PDF-to-PDL 转换过滤器 18 将 PDF 文件转换成 PDL 文件,然后将该 PDL 文件作为流传输给过滤器控制单元 14。过滤器控制单元 14 输出 PDL 文件,作为格式转换过滤器组 15 的输出数据。

[0055] 将参照图 10 和图 11 中的流程图,详细说明由打印控制单元 13 进行的处理(步骤 S402)。图 10 是例示步骤 S402 中的处理的概览的流程图。图 11 是例示步骤 S402 中的处理的详情的流程图。

[0056] 以下将说明步骤 S402 中的处理的概览。在步骤 S402-1 中,打印控制单元 13 执行文件分析处理。具体地说,打印控制单元 13 指示过滤器控制单元 14 进行如下操作:分析要经受打印的输入文档;获取文件格式,并且在输入文档是结构化文档文件的情况下,获取关于输入文档是否包含任何附加文档文件的信息;以及,确定输入文档是否能够由指定的过滤器处理。

[0057] 在步骤 S402-2 中,打印控制单元 13 执行过滤器组合处理。具体地说,打印控制单元 13 确定通过添加或组合过滤器是否能够将输入文档变成可处理的;并且,如果能够将输入文档变成可处理的,则动态连接这些过滤器。

[0058] 在步骤 S402-3 中,打印控制单元 13 执行过滤器执行处理。具体地说,打印控制单元 13 实际地载入由步骤 S402-2 中的过滤器组合处理而动态组合的过滤器。

[0059] 上述步骤 S402-1、S402-2 和 S402-3 中的处理分别对应于图 11 中的步骤 S402-1-1 至 S402-1-2、S402-2-1 至 S402-2-4,以及 S402-3-1 至 S402-3-2。以下将参照图 11 说明步骤 S402 中的处理的详情。

[0060] 打印控制单元 13 指示过滤器控制单元 14 分析要经受打印的输入文档文件,以



基于接收到的过滤器设置文件来确定该输入文档是否是可处理的。在步骤 S402-1-1 和 S402-1-2 中,过滤器控制单元 14 根据该指令执行文件分析处理。所述“分析输入文档文件”包括:基于要经受打印的输入文档的文件名扩展名确定文件格式;确定输入文档是否是结构化文档文件,并且如果输入文档是结构化文档文件,则确定输入文档是否包含任何附加文件;以及,在输入文档包含有任何附加文件的情况下,获取关于附加文档文件的文件格式的信息。以下基于示例性文件更详细地说明该处理。首先,以下基于仅描述 PDF-to-PDL 转换过滤器 18 的过滤器设置文件 502(图 5)来说明该处理。当输入数据是基于针对 PDF-to-PDL 转换过滤器 18 的过滤器输入/输出定义文件 603 的 PDF 文件时,确定输入数据是可处理的(情况 1)。当输入数据是 PDF-to-PDL 转换过滤器 18 不支持的 MS Office 文件时,确定输入数据是不可处理的(情况 2)。当输入数据是组合 PDF 文件时,由于 PDF-to-PDL 转换过滤器 18 不能够处理该 PDF 文件中的附加文档文件,因此确定输入文档是不可处理的(情况 3)。

[0061] 在过滤器控制单元 14 确定输入文档是可处理的情况下(步骤 S402-1-2 中的“是”),例如情况 1,处理进行到步骤 S402-3-1。在步骤 S402-3-1 中,过滤器控制单元 14 依次载入过滤器设置文件 502 中描述的过滤器。在情况 1 中,过滤器控制单元 14 仅载入 PDF-to-PDL 转换过滤器 18。以下将说明情况 2 和 3 中的处理。

[0062] 在步骤 S402-3-2 中,过滤器控制单元 14 控制如上文中所述的载入的过滤器,以获得最终的输出文件。

[0063] 另一方面,在过滤器控制单元 14 确定输入文档是不可处理的情况下(步骤 S402-1-2 中的“否”),处理进行到步骤 S402-2-1。在步骤 S402-2-1 中,过滤器控制单元 14 确定通过添加过滤器设置文件 502 中未描述的新过滤器是否能够将输入文档变成可处理的。在步骤 S402-2-1 中,过滤器控制单元 14 通过使用过滤器输入/输出定义文件 601 至 604、实际输入的文档文件(情况 2 中的 MS Office 文件或者情况 3 中的组合 PDF 文件)以及过滤器定义文件 502,来进行确定。

[0064] 以下将说明情况 2 中的处理。由于过滤器设置文件 502 将 PDF-to-PDL 转换过滤器 18 描述为第一过滤器,因此,基于过滤器输入/输出文件 603 可以得出输入数据必须是 PDF 文件。具体地说,当输入数据是 MS Office 文件时,通过向过滤器管线的顶部添加 Office-to-PDF 转换过滤器 16(用于将 MS Office 文件转换成 PDF 文件),能够将输入文档变成可处理的。因此,过滤器控制单元 14 确定通过添加相关的过滤器能够将输入文档变成可处理的。

[0065] 在过滤器控制单元 14 确定输入文档是可处理的情况下(步骤 S402-2-1 中的“是”),处理进行到步骤 S402-2-2。在步骤 S402-2-2 中,过滤器控制单元 14 更新过滤器设置文件 502。在情况 2 中,过滤器控制单元 14 更新过滤器设置文件 502,使得将 Office-to-PDF 转换过滤器 16(用于将 MS Office 文件转换成 PDF 文件)添加到过滤器管线的顶部。具体地说,过滤器控制单元 14 将过滤器设置文件 502 更新为过滤器设置文件 503。

[0066] 另一方面,在过滤器控制单元 14 确定输入文档是不可处理的情况下(步骤 S402-2-1 中的“否”),处理进行到步骤 S402-2-3。在步骤 S402-2-3 中,过滤器控制单元 14 确定通过组合多个过滤器,是否能够将输入文档变成可处理的。过滤器控制单元 14 基于输入数据分析处理(步骤 S402-1-1)的结果进行所述确定。具体地说,过滤器控制单元 14 确

定要经受打印的输入文档的文件格式。过滤器控制单元 14 基于输入文档的文件格式确定输入文档是否是结构化文档文件,并且如果输入文档是结构化文档文件,则确定输入文档是否包括任何附加文档文件。如果输入文档包括任何附加文档文件,则过滤器控制单元 14 使用关于附加文档文件的文件格式的信息。在情况 3 中,以 MS-Office 文件、XPS 文件以及 MS-Office 文件的顺序向组合 PDF 文件附加两个 MS-Office 文件和一个 XPS 文件。

[0067] 在此情况下,作为输入数据分析处理(步骤 S402-1-1)的结果,获取以下信息:“输入文档是包括以下三个附加文档文件的结构化文档:MS-Office 文件、XPS 文件以及 MS-Office 文件”。

[0068] 类似于情况 2,在情况 3 中过滤器设置文件 502 将 PDF-to-PDL 转换过滤器 18 描述为第一过滤器。PDF-to-PDL 转换过滤器 18 不能够处理在组合 PDF 文件中包含的附加文档文件。然而,在过滤器控制单元 14 执行以下处理后,PDF-to-PDL 转换过滤器 18 能够处理在组合 PDF 文件中包含的附加文档文件。过滤器控制单元 14 提取附加的 XPS 文件和两个 MSOffice 文件,作为与组合 PDF 文件独立的文件,将组合 PDF 文件转换成简单的 PDF 文件,并将这些文件存储在辅助存储装置 3 中。然后,过滤器控制单元 14 生成描述至这些文件的文件路径的列表文件 802。在列表文件 802 的顶部描述的 PDF 文件是指从组合 PDF 文件转换来的简单 PDF 文件。以下的 MS-Office 文件、XPS 文件以及 MS-Office 文件是指从组合 PDF 文件中提取出的文件。然后,如果管线是通过组合 Office-to-PDF 转换过滤器 16、XPS-to-PDF 转换过滤器 17 以及 PDF 合并过滤器 21 而构成的,则能够将生成的列表文件 802 变成可处理的。因此,在此情况下,过滤器控制单元 14 确定,通过组合过滤器能够将输入文档变成可处理的。

[0069] 在过滤器控制单元 14 确定通过组合多个过滤器能够将输入文档变成可处理的情况下(步骤 S402-2-3 中的“是”),处理进行到步骤 S402-2-4。在步骤 S402-2-4 中,过滤器控制单元 14 生成上述列表文件。在情况 3 中,过滤器控制单元 14 将原始的 PDF 文件分割成 PDF 文件和提取的 MS-Office 文件和 XPS 文件。由于因过滤器管线系统的特性而难以将这些文件同时发送给过滤器,因此过滤器控制单元 14 生成列表文件。具体地说,在情况 3 中,过滤器控制单元 14 生成列表文件 802。

[0070] 如上所述,在列表文件 802 的情况下,以 MS-Office 文件、XPS 文件以及 MS-Office 文件的顺序将 XPS 文件和两个 MS-Office 文件附加到组合 PDF 文件中。

[0071] 在过滤器控制单元 14 确定通过组合多个过滤器能够将输入文档变成可处理的情况下(步骤 S402-2-3 中的“是”),与确定输入文档是可处理的情况(步骤 S402-2-1 中的“是”)类似,处理进行到步骤 S402-2-2。在步骤 S402-2-2 中,过滤器控制单元 14 更新过滤器设置文件 502 以反映步骤 S402-2-3 的结果。在情况 3 中,为了处理生成的列表文件 802,过滤器控制单元 14 将 Office-to-PDF 转换过滤器 16(用于将 MS Office 文件转换成 PDF 文件)添加到过滤器管线的顶部,并将过滤器设置文件 502 更新成过滤器设置文件 504,以使得 XPS-to-PDF 转换过滤器 17(用于将 XPS 文件转换成 PDF 文件)和 PDF 合并过滤器 21(用于组合 PDF 文件)进行处理。

[0072] 在过滤器控制单元 14 确定输入文档是不可处理的情况下(步骤 S402-2-3 中的“否”),过滤器控制单元 14 将输入文档识别为不可处理的数据,并终止处理。例如,在输入数据是 XPS 文件并且不存在用于将 XPS 文件转换成 PDF 文件的过滤器时,过滤器控制单元

14 将输入文档识别为不可处理的数据并终止处理。此外,在输入数据是组合 PDF 文件并且不存在用于将组合 PDF 文件中的附加数据转换成 PDF 文件的过滤器时,过滤器控制单元 14 将输入文档识别为不可处理的数据并终止处理。

[0073] 即使在处理进行到步骤 S402-2-1 和 S402-2-3 的情况下,处理也执行步骤 S402-3-1 和 S402-3-2。由于已经在上面说明了步骤 S402-3-1 中的处理,因此省略重复说明。尽管在以上内容中已经说明了步骤 S402-3-2 中的处理,但是以下将说明在步骤 S402-2-4 中生成的列表文件。如上文中所述,当列表文件是至各过滤器的输入数据时,过滤器处理列表文件中的、文件格式与在过滤器输入 / 输出定义文件中指定的输入文件格式相一致的所有文档文件。以下将说明情况 3。在情况 3 中,使用过滤器设置文件 504 和列表文件 802 (输入数据)。过滤器设置文件 504 描述了由 Office-to-PDF 转换过滤器 16 (第一过滤器) 处理列表文件 802 中描述的文件中的 MS Office 文件。具体地说,描述了“execute = OfficeFile”。因此,在列表文件 802 中描述的文件中,对具有 Office 文件格式 (FormatType = Office File) 的文件进行处理。在此情况下,由 Office-to-PDF 转换过滤器 16 将“yyy.doc”和“aaa.doc”转换成 PDF 文件,并且将列表文件 802 更新成列表文件 803。

[0074] 然后,过滤器设置文件 504 描述了由 XPS-to-PDF 转换过滤器 17 处理列表文件 803 中描述的文件中的 XPS 文件。具体地说,描述了“execute = XPS file”。因此,在列表文件 502 中描述的文件中,对具有 XPS 文件格式 (FormatType = XPS File) 的文件进行处理。在此情况下,由 XPS-to-PDF 转换过滤器 17 将“zzz.xps”转换成 PDF 文件,并将列表文件 803 更新成列表文件 801。

[0075] 按此方式,对列表文件中描述的文件之中的、具有在过滤器设置文件中指定的文件格式的文件进行处理,完全实现了高效的处理。在上述示例中,列表文件以非连续方式包括 Office 文件,即,PDF 文件、Office 文件、XPS 文件和 Office 文件。然而,Office-to-PDF 转换过滤器 16 仅需要被载入并执行一次。

[0076] 然后,至随后 PDF 合并过滤器 21 的输入数据是列表文件 801。过滤器设置文件 504 描述了由 PDF 合并过滤器 21 处理在列表文件 801 中描述的文件中的 PDF 文件。具体地说,描述了“execute = PDF File”。因此,在列表文件 801 中描述的文件中,对具有 PDF 文件格式 (FormatType = PDF File) 的文件进行处理。在此情况下,由 PDF 合并过滤器 21 将“xxx.pdf”、“yyy.pdf”、“zzz.pdf”以及“aaa.pdf”合并成一个 PDF 文件,并且将列表文件 801 更新成列表文件 804。最后,由 PDF-to-PDL 转换过滤器 18 处理在列表文件 804 中描述的 PDF 文件,并获得最终的 PDL 文件作为输出结果。然后,过滤器控制单元 14 将 PDL 文件,视为格式转换过滤器组 15 的输出数据。

[0077] 按此方式生成组合 PDF 文件 (组合结构化文档文件) 的列表文件,使得能够指定要对在组合 PDF 文件中包含的特定附加文档文件 (例如,MS Office 文件) 施加的处理。这一方法使得即使在输入数据是组合 PDF 文件 (组合结构化文档文件) 时也能够动态地构造过滤器管线。此外,由于基于针对各过滤器指定输入和输出文件格式的定义文件生成并更新处理模块信息,因此该方法容易地实现了将新的过滤器连接到过滤器管线上。

[0078] 在本发明的上述示例性实施例中,生成结构化文档文件的列表文件使得能够指定要对其中包含的特定附加文档文件施加的处理。本示例性实施例使得即使在输入数据是组合结构化文档文件时也能够动态地构造过滤器管线。此外,由于基于针对各过滤器指定输

入和输出文件格式的定义文件生成并更新处理模块信息,因此本示例性实施例容易地实现了将新的过滤器连接到过滤器管线上。

[0079] 其他实施例

[0080] 还可以由读出并执行记录在存储设备上的程序来执行上述实施例的功能的系统或装置的计算机(或诸如 CPU 或 MPU 等的设备),来实现本发明的各方面;并且可以利用由通过例如读出并执行记录在存储设备上的程序来执行上述实施例的功能的系统或装置的计算机来执行各步骤的方法,来实现本发明的各方面。为此,例如经由网络或从充当存储设备的各种类型的记录介质(例如,计算机可读介质)将程序提供给计算机。

[0081] 虽然参照示例性实施例对本发明进行了说明,但是应当理解,本发明不局限于所公开的示例性实施例。说明书中公开的各个特征以及(适当时)权利要求书和附图可以单独或以任何适当的组合方式提供。

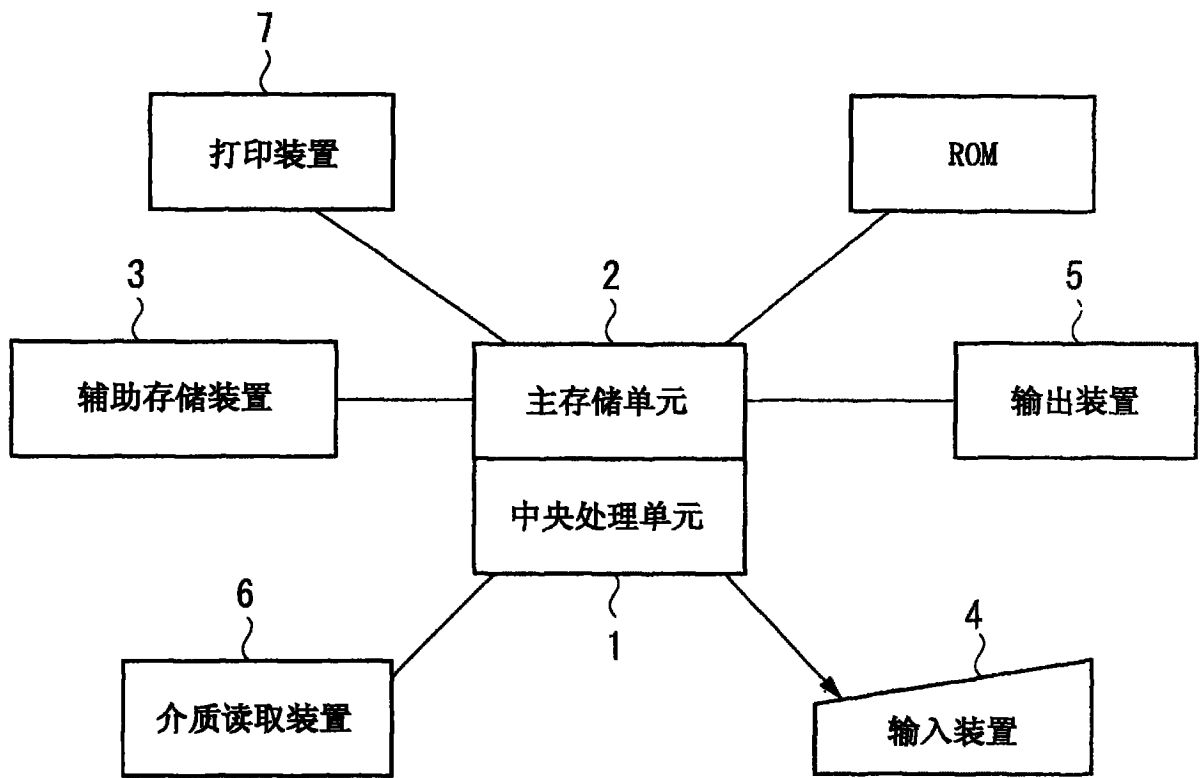


图 1

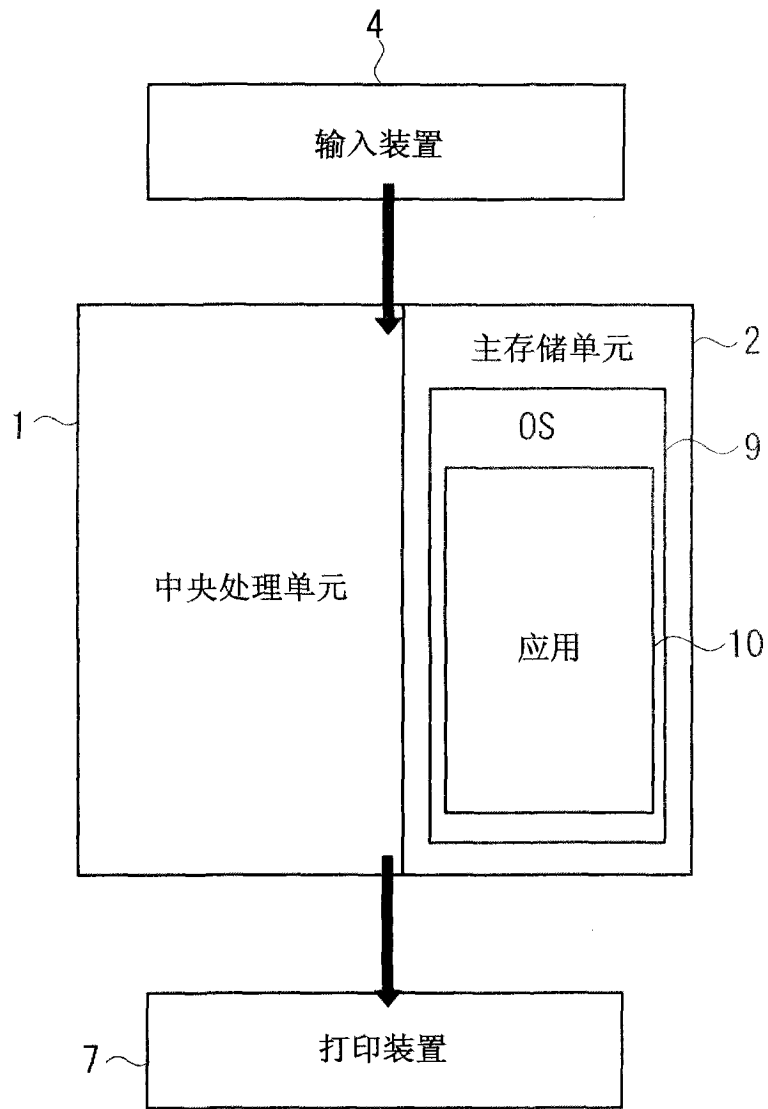


图 2

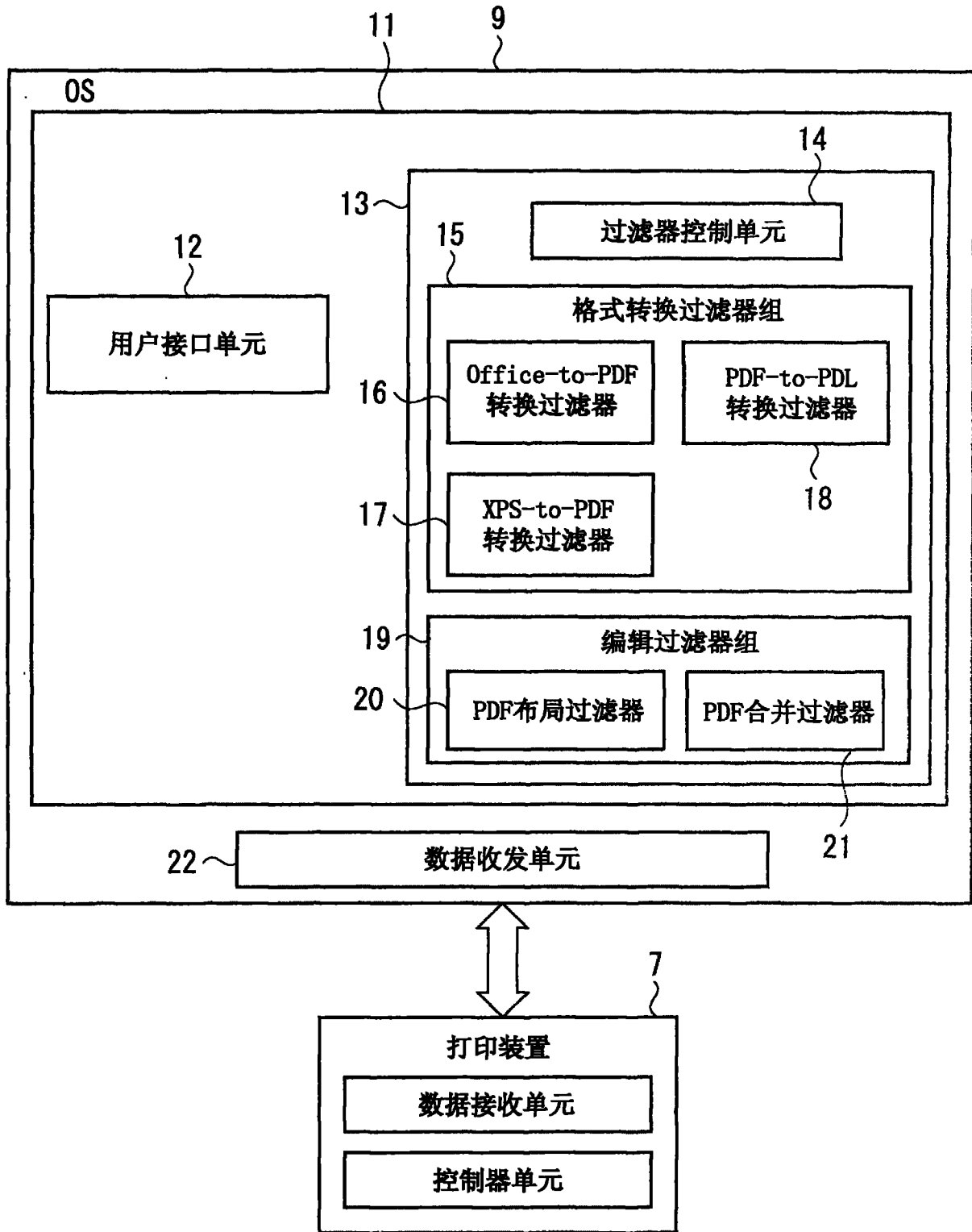


图 3

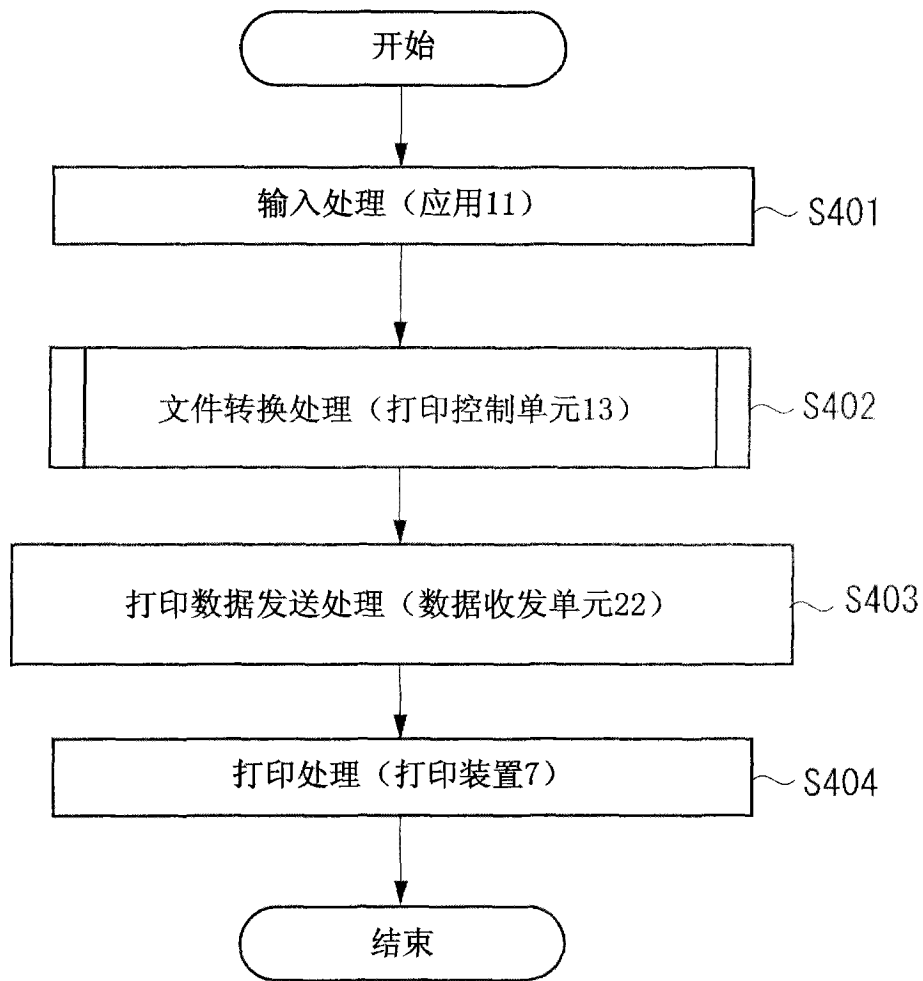


图 4



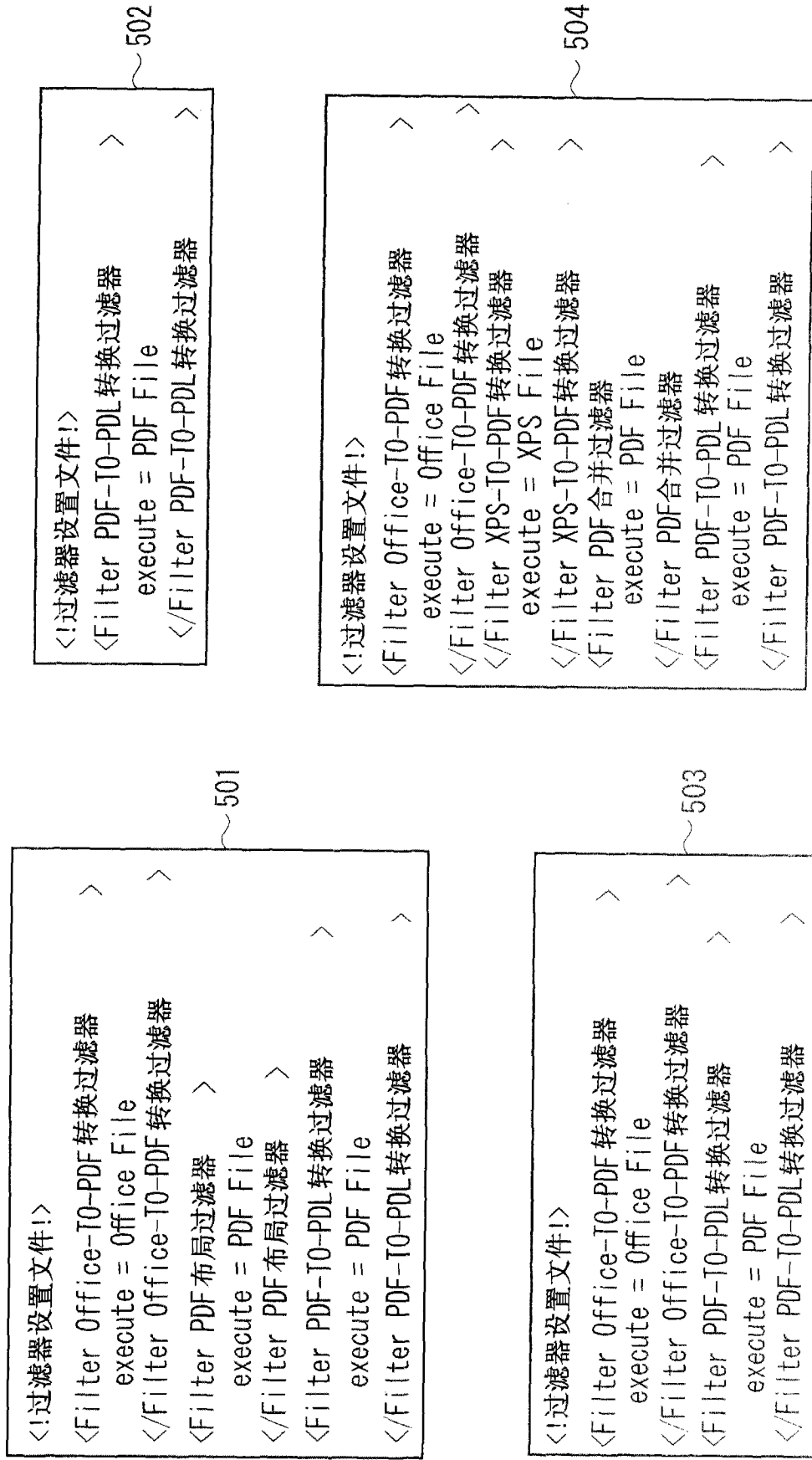


图 5

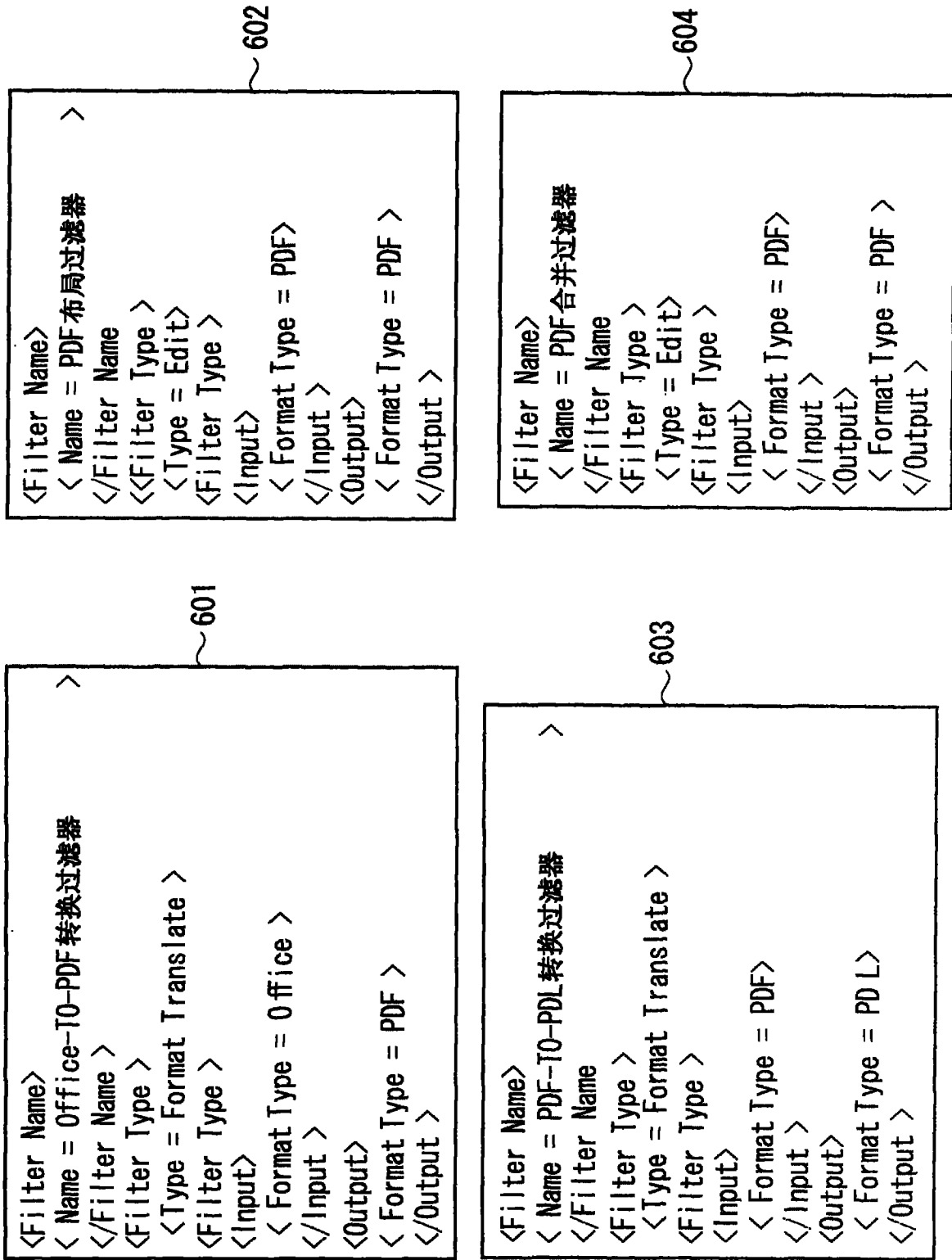


图 6

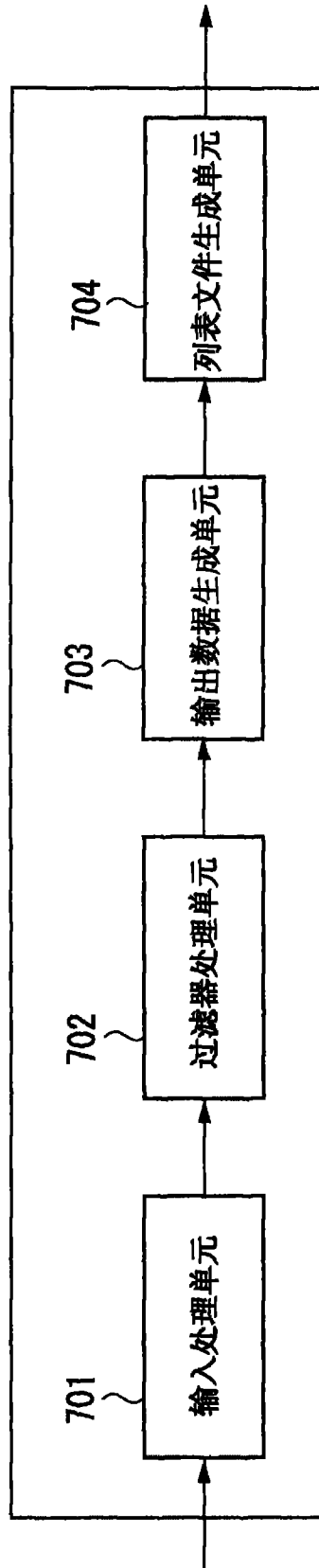


图 7

803

```

<! 列表文件 !>
<Doc 1 >
  <FormatType = PDF File>
  <FilePath = xxx.pdf>
</Doc 1>
<Doc 2 >
  <FormatType = PDF File>
  <FilePath = yyy.pdf>
</Doc 2>
<Doc 3 >
  <FormatType = XPS File>
  <FilePath = zzz.xps>
</Doc 3>
<Doc 4 >
  <FormatType = PDF File>
  <FilePath = aaa.pdf>
</Doc 4>

```

804

```

<! 列表文件 !>
<Doc 1 >
  <FormatType = PDF File>
  <FilePath = zzz.pdf>
</Doc 1>

```

802

```

<! 列表文件 !>
<Doc 1 >
  <FormatType = PDF File>
  <FilePath = xxx.pdf>
</Doc 1>
<Doc 2 >
  <FormatType = Office File>
  <FilePath = yyy.doc>
</Doc 2>
<Doc 3 >
  <FormatType = XPS File>
  <FilePath = zzz.xps>
</Doc 3>
<Doc 4 >
  <FormatType = Office File>
  <FilePath = aaa.doc>
</Doc 4>

```

801

```

<! 列表文件 !>
<Doc 1 >
  <FormatType = PDF File>
  <FilePath = xxx.pdf>
</Doc 1>
<Doc 2 >
  <FormatType = PDF File>
  <FilePath = yyy.pdf>
</Doc 2>
<Doc 3 >
  <FormatType = PDF File>
  <FilePath = zzz.pdf>
</Doc 3>
<Doc 4 >
  <FormatType = PDF File>
  <FilePath = aaa.pdf>
</Doc 4>

```

图 8

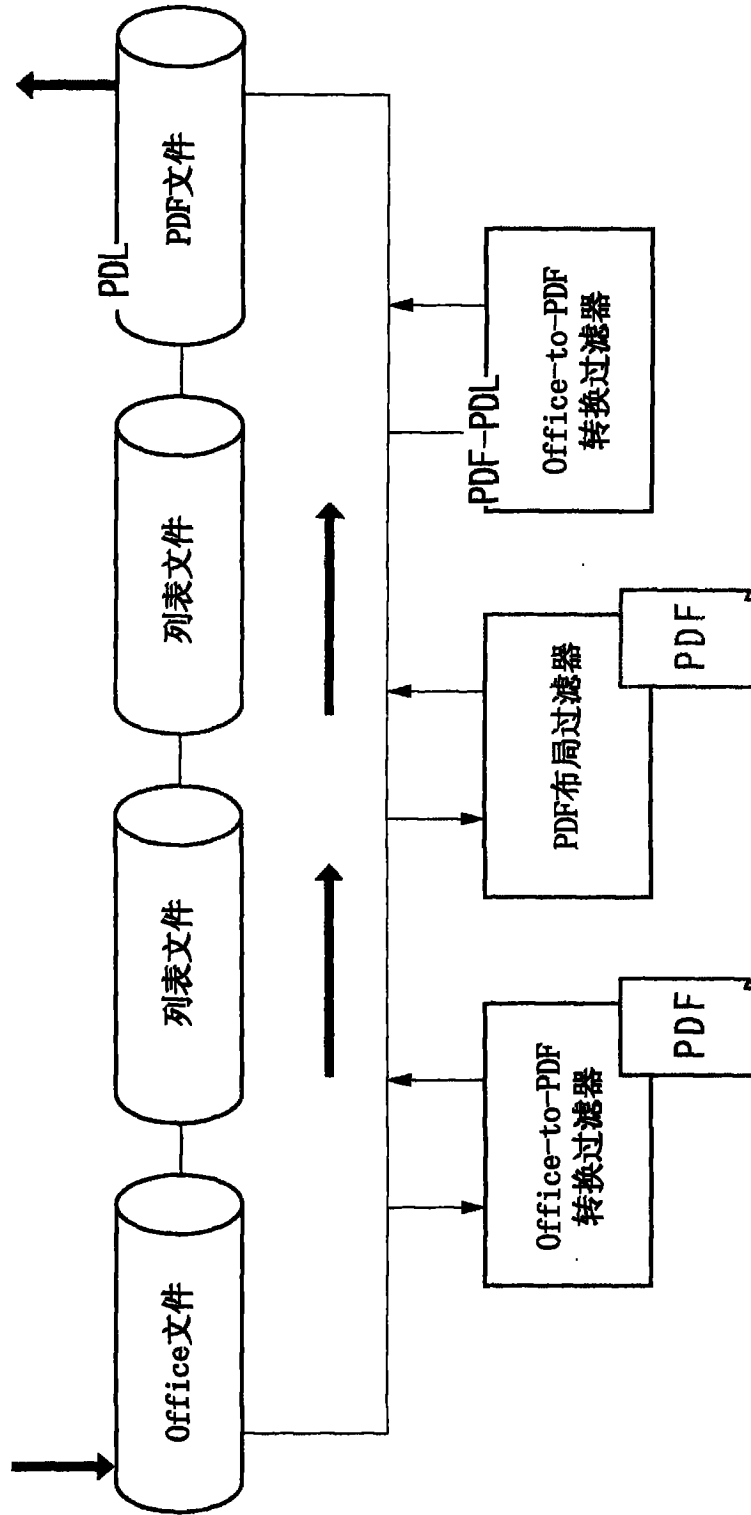


图 9

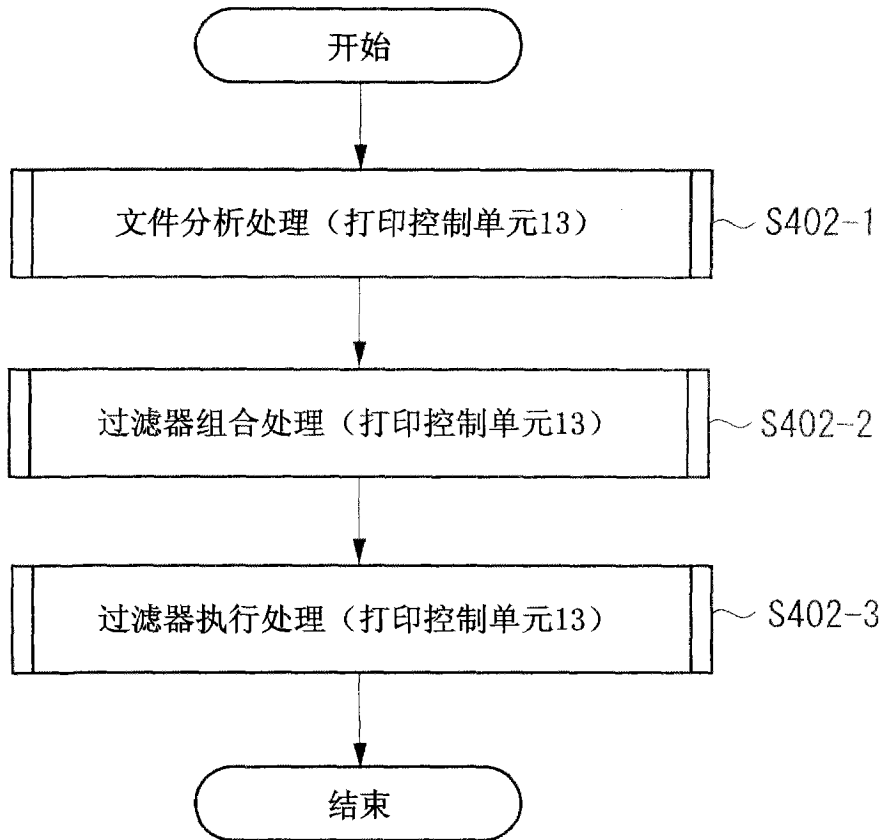


图 10

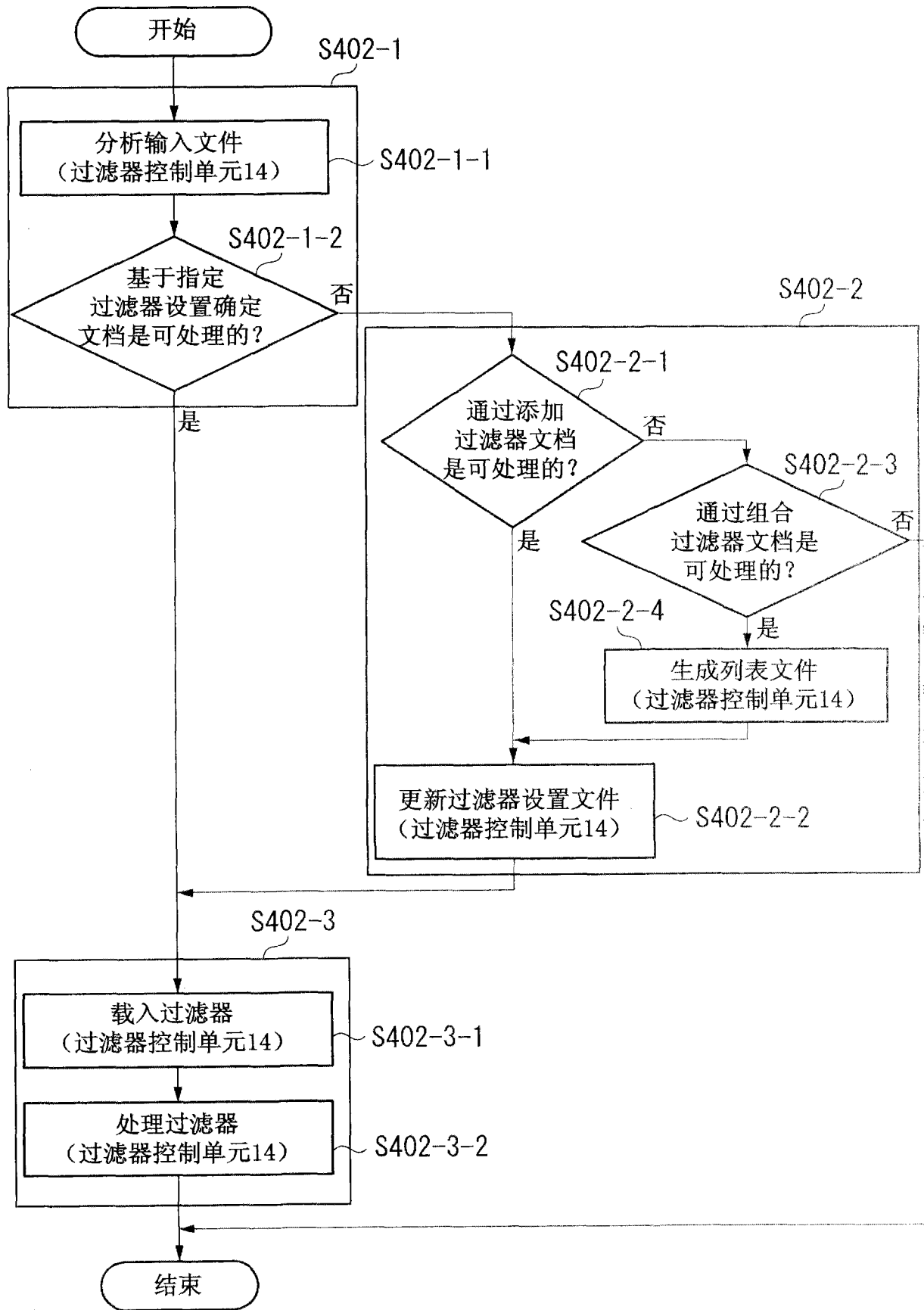


图 11