

[12] 发明专利说明书

[21] ZL 专利号 96114525.0

[45]授权公告日 2002 年 12 月 4 日

[11]授权公告号 CN 1095567C

[22]申请日 1996.11.8 [21]申请号 96114525.0

[73]专利权人 国际商业机器公司

地址 美国纽约

[72]发明人 简志敏 田 忠

[56]参考文献

US4864438 1989. 9. 5 G11B15/68

US5214768 1993. 5.25 G06F12/00

审查员 刘春霞

[74]专利代理机构 中国国际贸易促进委员会专利商标事
务所

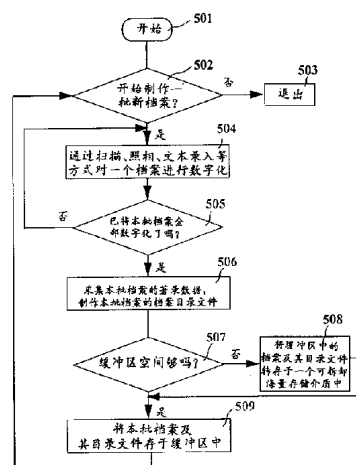
代理人 王以平

权利要求书 2 页 说明书 8 页 附图 5 页

[54]发明名称 在可拆卸海量存储介质上制作档案的方法及档案服务器

[57]摘要

公开一种在可拆卸海量存储介质上制作档案的方法,其特征在于包括以下步骤:将档案数字化为数字化档案;提取档案的著录信息;将所述数字化档案和所述著录信息存储到所述可拆卸海量存储介质上。还公开一种数字化档案服务器,用于将按照权利要求 1 的在可拆卸海量存储介质上制作档案的方法制造的档案登录到数字图书馆的数据库中,其特征在于包括:用于读取可拆卸海量存储介质上所存储的著录信息的装置;用于将所述著录信息登录到数字图书馆的数据库表中的装置。



1. 在可拆卸海量存储介质上制作档案的方法，其特征在于包括以下步骤：

将一批新档案中的一个档案数字化；

判断是否已将本批新档案中的所有档案数字化，并重复执行数字化处理，直到本批新档案中的全部档案均已数字化；

在本批新档案中的所有档案均已数字化后，采集表示本批新档案中档案详细信息的著录数据；

创建本批新档案中档案的档案目录文件；

判断一个缓冲区中是否能够容纳本批新档案中的档案及其目录文件；

如果缓冲区能够容纳本批新档案中的档案及其目录文件，则将本批新档案中的档案及其目录文件存储到该缓冲区中；以及

如果缓冲区不能容纳本批新档案中的档案及其目录文件，则将本批新档案中的档案及其目录文件转存于一个可拆卸海量存储介质上；

2. 根据权利要求1的方法，其特征在于采集著录数据进一步包括将本批新档案中各个档案的著录数据存储在不同的档案目录文件中。

3. 数字化档案服务器，用于将可拆卸存储介质上所存储的档案登录到数字图书馆的数据库中，其特征在于包括：

一个用于存储数据的缓冲区；

一个用于在可拆卸海量存储介质上读数据和写数据的海量存储介质库驱动程序；

一个用于在可拆卸海量存储介质上创建数字化档案的档案服务器，该档案服务器将一批新档案中的一个档案数字化，判断是否已将本批新档案中的所有档案数字化，并重复执行数字化处理，直到本批新档案中的全部档案均已数字化，在本批新档案中的所有档案均已数字化后，采集表示本批新档案中档案详细信息的著录数据，创建本批新档案中档案的档案目录文件，判断一个缓冲区中是否能够容纳本

批新档案中的档案及其目录文件，如果缓冲区能够容纳本批新档案中的档案及其目录文件，则将本批新档案中的档案及其目录文件存储到该缓冲区中，以及如果缓冲区不能容纳本批新档案中的档案及其目录文件，则将本批新档案中的档案及其目录文件传送到可拆卸海量存储介质上。

在可拆卸海量存储介质上 制作档案的方法及档案服务器

本发明涉及在数字图书馆中使用的在可拆卸海量存储介质上制作数字化档案的方法，并且涉及用于将这种可拆卸海量存储介质上的数字化档案登录到数字图书馆的数据库中的数字化档案服务器。

数字图书馆是“信息高速公路”的一个重要部分。一般来说，数字图书馆主要由网络、数据库服务器、海量存储器、用户界面等部分组成。通常，对于一份档案资料，其进入数字图书馆的过程如下：

1. 通过人工录入、光电识别、电子扫描、数字录音/录象等方式将档案资料数字化；
2. 将已数字化的档案资料存储到数字图书馆的海量存储器池中；
3. 提取有关此份档案资料的著录项目以及已数字化的档案资料在本数字图书馆的海量存储介质中的物理或逻辑地址，并将著录项目和物理或逻辑地址登录在数字图书馆对应的数据库中。

一份档案一旦进入数字图书馆，读者或用户就可以通过连接在网络上的计算机终端访问数字图书馆，以获取所需资料。

在现今的“信息高速公路”环境下，人们除了利用网络直接传输进行信息交流之外，主要是通过可拆卸的海量存储介质作为物理载体进行信息交流。在目前，CD - ROM 是最重要的媒体之一。在数字图书馆中，论文、照片、电影、音乐以及其他档案资料常常被数字化后存储于 CD - ROM 中，形成 CD - ROM 档案。然而目前，CD - ROM 档案存在如下几个问题：

1. 档案在 CD - ROM 上的著录格式不统一；
2. CD - ROM 上没有相应的档案标注信息；
3. 必须通过人工将 CD - ROM 档案登录到电子化数字图书馆的数据

库管理系统中。

以上这些问题带来的后果是：将 CD - ROM 档案登录到数字图书馆的数据库中时将浪费大量的人力、物力和时间，为一个特定的数字图书馆制作的 CD - ROM 档案只能在该系统中使用。如果要将一个数字图书馆的 CD - ROM 档案使用在另一个数字图书馆中，则需要人工地对 CD - ROM 档案进行数据库登录。这样，CD - ROM 档案的传播和交流受到了极大的限制。

本发明第一个目的在于提供一种在可拆卸海量存储介质上制作数字化档案的方法。

本发明第二个目的在于提供一种数字化档案服务器，用于将可拆卸海量存储介质上的数字化档案登录到数字图书馆的数据库中。

为实现上述第一目的，本发明提供一种在可拆卸海量存储介质上制作数字化档案的方法，包括以下步骤：将档案数字化为数字化档案；提取档案的著录信息；将所述数字化档案和所述著录信息存储到所述可拆卸海量存储介质上。

为实现上述第二目的，本发明提供一种数字化档案服务器，用于将可拆卸海量存储介质上的数字化档案登录到数字图书馆的数据库中，该数字化档案服务器包括：

用于读取可拆卸海量存储介质上所存储的著录信息的装置；

用于将所述著录信息登录到数字图书馆的数据库表中的装置。

利用本发明的方法，可以统一数字化档案在可拆卸海量存储介质上的著录格式，便于各数字图书馆之间交流可拆卸海量存储介质上的数字化档案。在数字图书馆中利用本发明的档案服务器，可以将可拆卸海量存储介质上的数字化档案自动地登录到数字图书馆的数据库管理系统中，节省了大量的人力、物力和时间。

通过以下详细描述，并结合附图，本发明的其他优点、特征将会更加明显，其中

图 1 示出了本发明档案服务器在常规数字图书馆中的一种应用；

图 2 是在本发明方法和档案服务器中使用的光盘档案目录文件的一种

文件格式;

图3是图2的文件格式中所含著录条目的一种格式;

图4是图3的条目格式中所含档案项目的一种格式;

图5是本发明的在可拆卸海量存储介质上制作数字化档案的方法的一种流程图;

图6是本发明的数字化档案服务器的一种工作流程图。

下面结合附图对本发明的一种实施方式进行详细说明。

本发明中所采用的可拆卸海量存储介质可以是CD-ROM、PD、CD-R和DVD等。对于这些不同的存储介质,应用本发明的方法和档案服务器时原理是一样的。下面,虽然以CD-ROM为例进行说明,但是并不构成对本发明的限制。

图1是例示将本发明的档案服务器应用在数字图书馆中的示意图。标号1代表用户终端,标号2代表网络,标号3代表数字图书馆服务器,标号4代表数据库,标号5代表本发明的档案服务器,标号6代表光盘库驱动器,标号7代表用于存放数字化档案的海量存储器池(本例中,为光盘库)。

在光盘库7中存有大量光盘,每张光盘上的内容包括两部分,即光盘头和光盘体。光盘头由一组光盘档案目录文件组成,这组光盘档案目录文件是光盘档案的总揽和概要。档案目录文件的格式将在后文参照图2、3、4加以描述。光盘体由一组数字化档案文件组成。以上两部分的内容是利用本发明的方法在光盘上制作的。

光盘库驱动器6能够根据档案服务器5的命令,从光盘库7选择一张光盘,读取光盘头中的光盘档案目录文件,将目录文件中所含的档案著录信息及相应数字化档案的地址信息登录到数据库4中。如图所示,本发明的档案服务器5在体系结构上位于数据库4和光盘库驱动器6之间。档案服务器5的工作流程将在下文参照图6详细描述。

图2是本发明的方法和档案服务器使用的光盘档案目录文件的一种文件格式。光盘档案目录文件是在进行一次光盘著录时形成的档案条目集文件,它说明了各光盘著录条目及其在该文件中的排列顺序,该文件是光盘

档案自动进入数字图书馆并提供检索依据的工具。一般来说,为了便于分批制作数字化档案,并充分利用光盘存储空间,在光盘上形成一组档案目录文件及相应的数字化档案。

一个光盘档案目录文件分为三部分,如图2所示,为文件头201,文件体202和文件结束符203。

文件头202包括本目录文件所涉及的档案种类数2011、各类档案信息的依次说明2012和本目录文件涉及的档案个数2013。对于每一类档案,档案信息的说明包括:关于该类档案归档的建议(即其在数据库中的记录信息),该类档案的个数,该类档案中所有档案的著录条目名称。

如图2所示,在文件头201中包括以下内容:

档案目录文件所说明的档案的种类数(m);\n

关于第1类档案归档的建议;\n

第1类档案著录项目的个数(m1);\n

第1类档案第1个著录项目名;\n

...

第1类档案第m1个著录项目名;\n

关于第2类档案归档的建议;\n

第2类档案著录项目的个数(m2);\n

第2类档案第1个著录项目名;\n

...

第2类档案第m2个著录项目名;\n

关于第m类档案归档的建议;\n

第m类档案著录项目的个数(mm);\n

第m类档案第1个著录项目名;\n

...

第m类档案第mm个著录项目名;\n

本档案目录文件所说明的档案的个数(即条目的个数k)

文件体202由一组档案条目组成,条目的个数与档案个数2013相等。每一条目的格式将在后文参照图3描述。

文件结束符 203 是一个 ASCII 码 EOF。

如图所示，在文件头 201 中，每一项之后都随一个符号“\n”表示行尾，根据操作系统的不同，“\n”有不同的含义，如在基于 MS - DOS 的系统中表示回车和换行两个字符，而在基于 UNIX 的系统中表示一个回车符。后文所述的“\n”与这里的含义相同。

图 3 是图 2 中文件体 202 中档案条目的一种格式。光盘著录条目是单份文件著录的结果，它说明了著录项目及其排列顺序。如图所示，每一条目包括三部分，即本档案的卷号、本档案所涉及的所有项目、条目档案体的文件名（含路径名）。

图 4 是图 3 中所含项目的一种格式。光盘档案著录项目是根据光盘的易读难写的特性选择的能充分揭示档案内容和特性并便于检索的记录事项，它说明各著录项目的名称、内容及排列顺序。如图所示，每一项目是一个以“\n”结束的字节流，字节流代表光盘著录项目的内容。

图 5 是本发明的在可拆卸海量存储介质上制作数字化档案的方法的一种实施流程图。该实施从步骤 501 开始。在步骤 502 提示是否开始制作一批新档案。如果回答为“是”，则进入步骤 504，否则由步骤 503 退出。在步骤 504，通过扫描、照相、文字录入等方式对一个档案进行数字化，然后进入步骤 505。在步骤 505，判断是否已将本批档案全部数字化。如果在步骤 505 判定已经将本批全部档案数字化，则进入步骤 506，否则返回步骤 504。在步骤 506，采集本批档案的著录数据，制作本批档案的档案目录文件（档案目录文件的一种格式如上文参照图 2、3 和 4 所述）。进入步骤 507 后，判断缓冲区是否够用。如果判定缓冲区够用，则进入步骤 509，否则进入步骤 508。在步骤 509，将本批档案及相应目录文件存入缓冲区，然后返回步骤 502，进行下一批档案的制作过程。在步骤 508，将缓冲区中的档案及相应目录文件转存于一个可拆卸海量存储介质中，清空缓冲区，然后进入步骤 509。

在上述实施中，缓冲区是在可擦写存储介质上（如硬盘）的一个大小与相应可拆卸海量存储介质一致的存储空间。并且假定每批档案及其目录文件所占空间不会超过一个可拆卸海量存储介质的存储容量。

在同一个可拆卸海量存储介质（如光盘）上，可以多次利用上述实施，对多个档案分几次进行著录。在此介质上形成一组档案目录文件。这样，有助于存储介质读写系统充分利用此介质的容量。

在本实施方式中，第一次著录时形成的档案目录文件的文件名为 DOCS.001。以后可能进行的第二次或第三次著录时所形成的档案目标文件的文件名分别为 DOCS.002, DOCS.003，等等。所有这些档案目录文件构成一组档案目录文件。

为了进一步说明本发明的上述实施方法和上述档案目录文件的格式，下面举一个例子。

现有两部纪录短片。其一是《周总理访问朝鲜》，中国新闻纪录制片厂摄制，1971年5月。其二是《小麦种植》，中国农业电影制片厂摄制，1980年12月。现欲将这两部电影数字化，并放入某数字图书馆，以提供在线服务。

现在，根据本发明的上述实施方法制作数字化档案。

首先，用数字摄影机将这两部电影数字化为文件“zhou.mpg”和“xiao.mpg”，并存于“缓冲区”中（假设这两个文件所占存储空间的大小没有超过“缓冲区”大小，并假设“缓冲区”中仍有空间可以存放下面将要形成的档案目录文件）。然后，采集著录数据，制作档案目录文件。假定《周总理访问朝鲜》应该登录在“新闻影片”数据库中，其著录项目根据档案著录标准选为“片名”，“摄制日期”，“制片厂”。假定《小麦种植》应该登录在“科技影片”数据库中，其著录项目应为“片名”，“摄制日期”，“制片厂”。注意，著录项目是根据档案著录标准选定的，并与相应数据库表中的域名一致。这样，档案目录文件的内容如下（注：档案目录文件是一个文本文件，括号中的文字为对文件数据的说明，并非文件体）：

（文件开始）

2\n	（本档案目录文件说明两类档案）
新闻影片\n	（第1类档案为新闻影片）
3\n	（第1类档案有3个著录项目）

片名\n (第1类档案的第1个著录项目名为“片名”)
 摄制日期\n (第1类档案的第2个著录项目名为“摄制日期”)
 制片厂\n (第1类档案的第3个著录项目名为“制片厂”)
 科技影片\n (第2类档案为科技影片)
 3\n (第2类档案有3个著录项目)
 片名\n (第2类档案的第1个著录项目名为“片名”)
 摄制日期\n (第2类档案的第2个著录项名为“摄制日期”)
 制片厂\n (第2类档案的第3个著录项名为“制片厂”)
 2\n (本档案目录文件说明两个档案)
 (第1个条目)
 周总理访问朝鲜\n (第1个条目的“片名”)
 1971.5\n (第1个条目的“摄制日期”)
 中国新闻纪录制片厂\n (第1个条目的“制片厂”)
 zhou.mpg (第1个条目档案体的文件名)
 (第2个条目)
 小麦种植\n (第2个条目的“片名”)
 1980.12\n (第2个条目的“摄制日期”)
 中国农业电影制片厂\n (第2个条目的“制片厂”)
 xiao.mpg (第2个条目档案体的文件名)
 (文件结束)
 EOF (本档案目录文件的结束符)

假定这是第1批写入“缓冲区”的档案，则制作好的档案目录文件的文件名取为“DOCS.001”，并将其写入缓冲区。

假定缓冲区已没有空间存储第2批档案了，则将缓冲区中的内容写入一个相应的可拆卸海量存储介质（如一张CD）中。这时，该张CD中含有三个文件，即DOCS.001，zhou.mpg和xiao.mpg。

至此，完成了在可拆卸海量存储介质上的档案制作。

图6是本发明的数字化档案服务器的一种工作流程图。在步骤601，

启动数字化档案服务器。在步骤 602，判断在光盘库上是否有未经登录的 CD - ROM。如果，在步骤 602 中判定有未经登录的 CD - ROM，则进入步骤 603，否则由步骤 606 退出。在步骤 603，读入该 CD - ROM 中的每一个档案目录文件。然后在步骤 604，逐一分析档案目录文件，并将该文件中所记录的名类档案中的各个档案的相关数据写入到相应的数据库的表中。之后，在步骤 605 中判断是否还有其他档案目录文件未被分析。如果在步骤 605 中判定还有其他档案目录文件未被分析，则返回步骤 603，否则由步骤 606 结束登录过程。

下面，以前面的例子说明上述档案服务器的工作流程。

将按照上述方法制作的 CD 插入数字图书馆的光盘库中。

启动“数字化档案服务器”，开始新 CD 的登录。

“数字化档案服务器”查找光盘库，并发现了该未经登录的 CD。

“数字化档案服务器”打开 CD 上的档案目录文件 DOCS.001，并分析它，将两个档案的著录数据分别填入数字图书馆中的“新闻影片”数据库和“科技影片”数据库的表中。

“数字化档案服务器”关闭文件 DOCS.001，完成将档案登录到数字图书馆的过程。

虽然以上结合附图对本发明的方法和档案服务器进行了详细说明，但是应该理解到，对于本领域熟练的技术人员，在不背离本发明的实质和范围的情况下，可以做出许多修改和变更。本发明的范围仅由权利要求限定。

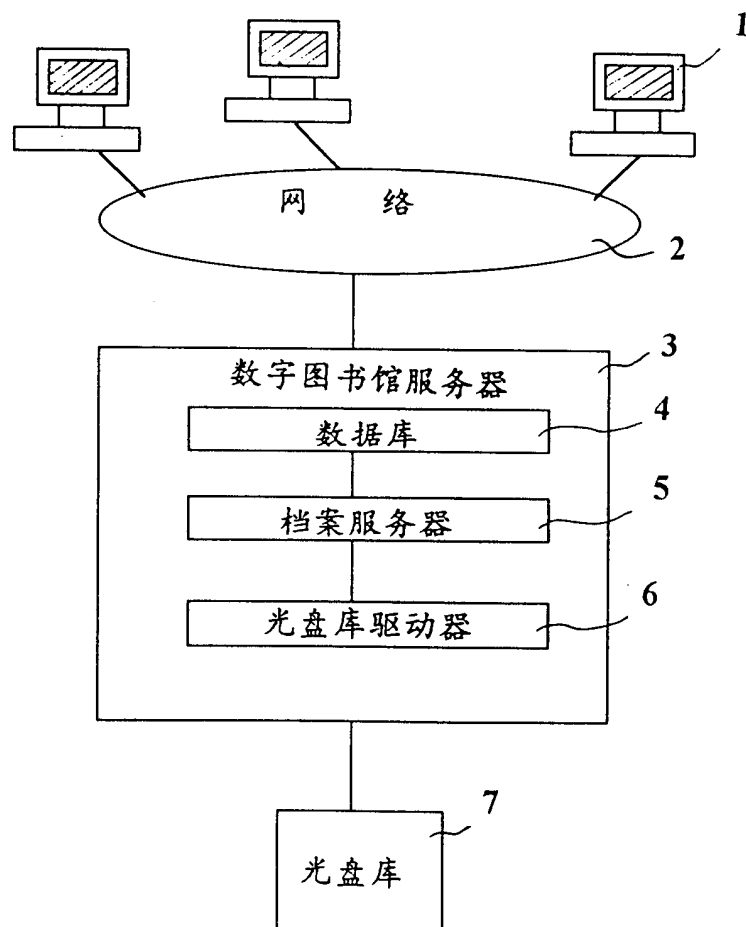


图1

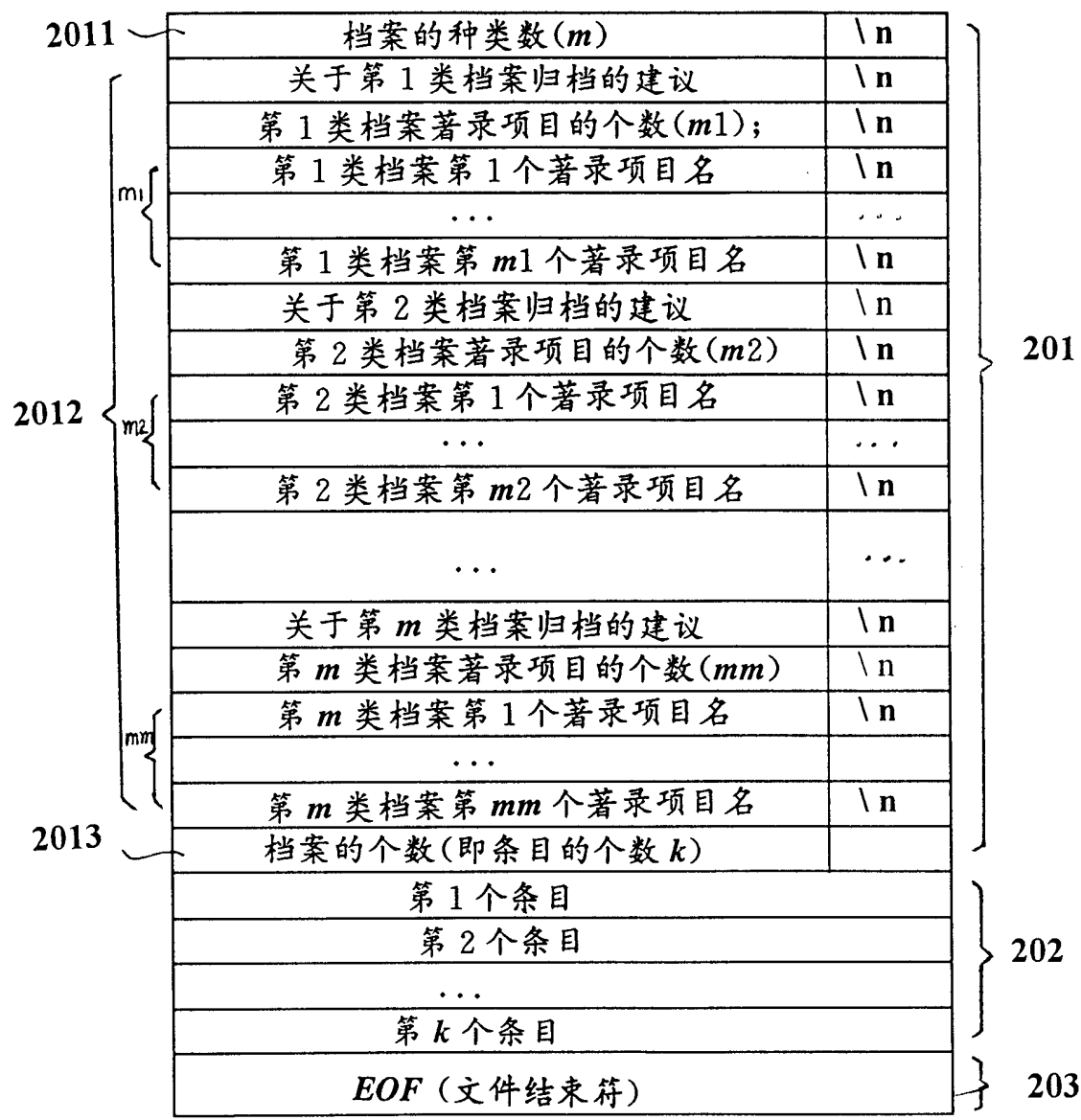


图 2

本档案类号	\n	
第1个项目		
第2个项目		
...		
第n个项目		
条目档案体的文件名 (含路径名)	\n	

图 3

项目的内容	\n
-------	----

图 4

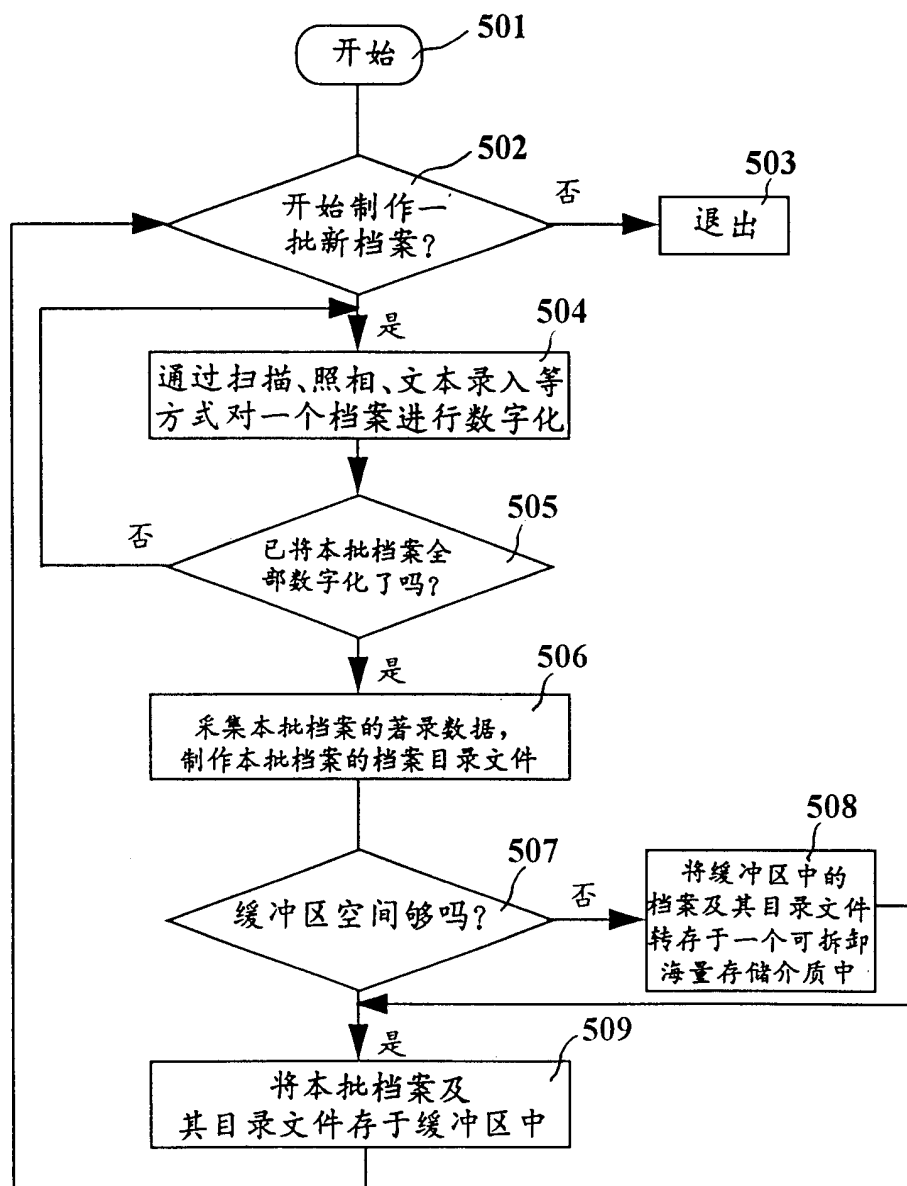


图 5

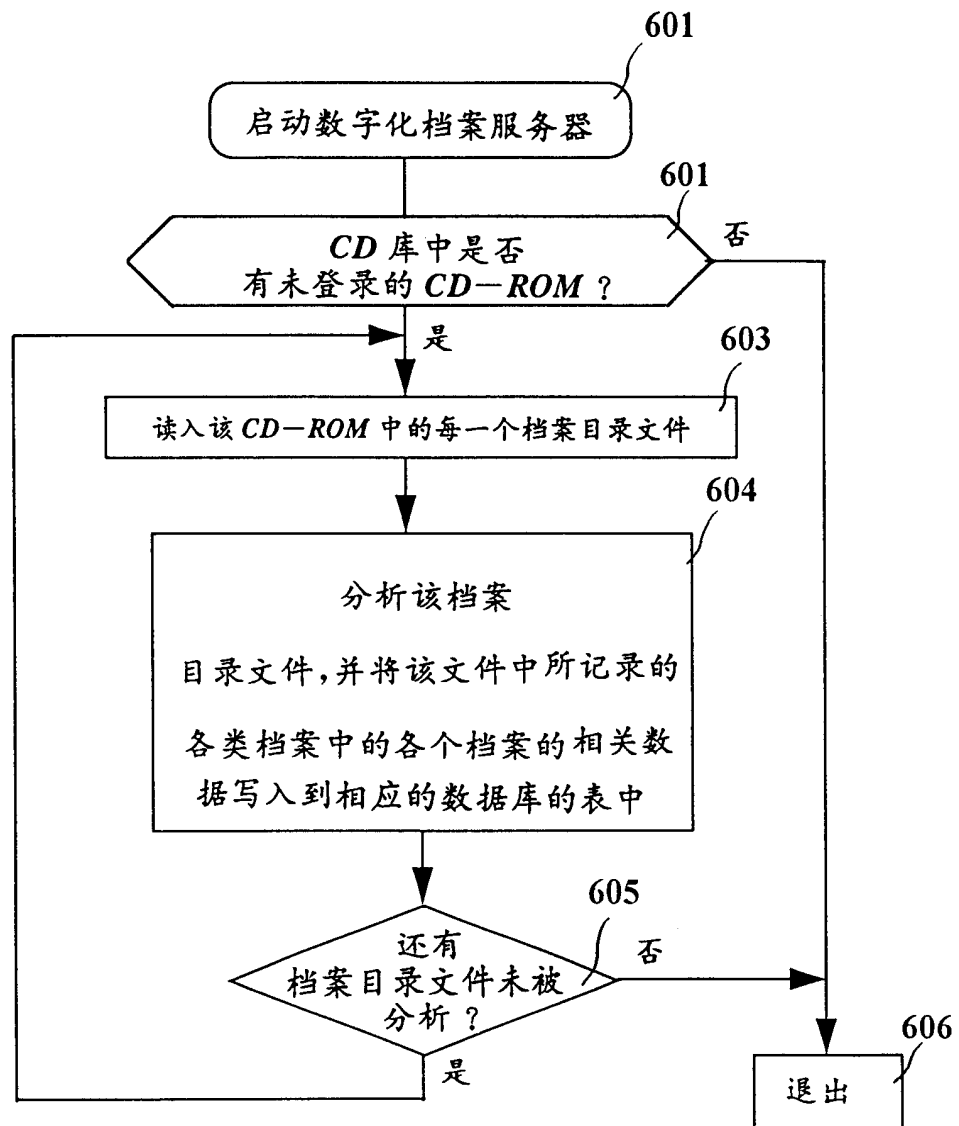


图 6