



(12) 发明专利

(10) 授权公告号 CN 1571968 B

(45) 授权公告日 2010.04.28

(21) 申请号 02820705.X

(74) 专利代理机构 中国专利代理(香港)有限公司 72001

(22) 申请日 2002.08.14

代理人 程天正 王勇

(30) 优先权数据

60/312,788 2001.08.17 US

(51) Int. Cl.

10/086,522 2002.03.04 US

G06F 17/30(2006.01)

(85) PCT申请进入国家阶段日

审查员 张妍

2004.04.19

(86) PCT申请的申请数据

PCT/US2002/025700 2002.08.14

(87) PCT申请的公布数据

W02003/017141 EN 2003.02.27

(73) 专利权人 通用商业矩阵有限责任公司

地址 美国内华达州

(72) 发明人 R·S·布洛克 G·冈农 U·科塔里

J·劳 F·查普斯

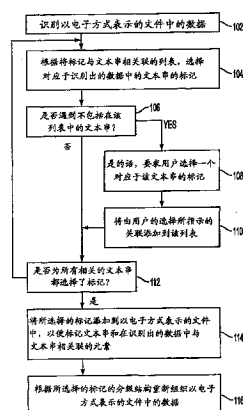
权利要求书 2 页 说明书 10 页 附图 6 页

(54) 发明名称

向数据的集合中添加与该数据相关联的元数据的方法

(57) 摘要

一种用于向数据添加标记(图1B)(例如可扩展标记语言兼容或可扩展商业报告语言兼容的标记)的方法,包括:a)识别以电子方式表示的文件中的数据(102);b)根据将标记与文本串相关联的列表来选择对应于识别出的数据中的文本串的标记(104);以及c)将所选择的标记添加到以电子方式表示的文件中,以便标记文本串以及识别出的数据中与文本串相关联的元素(114)。所述标记包括关于数据的信息并且在一个或多个分类中被定义。当列表没有将标记与文本串相关联时,可以提示用户选择一个对应于识别出的数据中的文本串的标记。然后可以将由用户的选择所指示的关联添加到将标记与文本串相关联的列表中(116)。



1. 一种计算机实现的用于将元数据添加到数据和与该数据相关联的第一元数据的集合中的方法,该方法包括以下步骤:

基于所述第一元数据和所述数据和 / 或所述第一元数据在所述集合中的一个或多个位置识别所述集合中的数据;以及

基于所识别出的数据将第二元数据添加到所述集合中。

2. 如权利要求 1 所述的方法,包括捕获作为数据流的所述数据和第一元数据的集合。

3. 如权利要求 2 所述的方法,包括:

存储捕获到的数据流;

其中对所存储的数据流执行识别步骤。

4. 如权利要求 2 所述的方法,其中所述捕获步骤和识别步骤是在不同的位置执行的。

5. 如权利要求 1 所述的方法,其中:

所述识别步骤包括识别第一货币中的第一值;以及

所述方法包括根据该第一值和一个换算系数来确定第二货币中的第二值。

6. 如权利要求 5 所述的方法,其中用户指定第二货币和换算系数。

7. 如权利要求 6 所述的方法,其中用户通过指示一个换算系数已知的日期来指定该换算系数。

8. 如权利要求 1 所述的方法,其中:

所述第一元数据根据第一标准来组织识别出的数据;以及

所述元数据根据第二标准来组织识别出的数据。

9. 如权利要求 8 所述的方法,其中:

所述第一标准是美国公认会计准则和国际公认会计准则中的一个;以及

所述第二标准是美国公认会计准则和国际公认会计准则中的另一个。

10. 如权利要求 1 所述的方法,其中所述第二元数据将识别出的数据映射到一个可扩展标记语言分类。

11. 如权利要求 1 所述的方法,其中所述第二元数据将识别出的数据映射到一个电子表格。

12. 如权利要求 1 所述的方法,其中所述第二元数据将识别出的数据映射到一个数据库。

13. 如权利要求 1 所述的方法,其中所述第二元数据将识别出的数据映射到一个平面文件。

14. 如权利要求 13 所述的方法,包括输出一个定义该平面文件的结构的数据定义。

15. 如权利要求 14 所述的方法,其中所述结构指示被映射的数据在该平面文件内的位置。

16. 如权利要求 2 所述的方法,其中所述数据流具有输出到计算机显示屏的数据格式。

17. 如权利要求 2 所述的方法,其中所述数据流具有输出到计算机数据端口的数据格式。

18. 如权利要求 2 所述的方法,其中所述数据流具有输出到数据存储设备的数据格式。

19. 如权利要求 18 所述的方法,其中所述数据存储设备是计算机中的随机存取存储器。

20. 如权利要求 18 所述的方法,其中所述数据存储设备是盘驱动器
21. 如权利要求 2 所述的方法,其中所述数据流是在实现该方法的计算机的操作系统级产生的。
22. 如权利要求 1 所述的方法,其中:
所述第二元数据包括从将标记与所述第一元数据相关联的列表中选择标记。
23. 如权利要求 1 所述的方法,其中:所述第二元数据将识别出的数据映射到文件结构、模式和分类中的至少一个中。
24. 如权利要求 22 所述的方法,包括将所述第一元数据从所述数据集合中去除。
25. 如权利要求 22 所述的方法,包括通过将所选择的标记与至少识别出的数据相组合来创建一个文件。
26. 如权利要求 22 所述的方法,包括以下步骤:当该列表没有把标记与元数据相关联时,要求用户选择与在所述第一元数据中的元数据相对应的标记。
27. 如权利要求 26 所述的方法,包括以下步骤:将由用户的选择所指示的关联添加到将标记与所述第一元数据相关联的列表中。
28. 如权利要求 22 所述的方法,其中所述列表包括多个与元数据相关联的标记。
29. 如权利要求 28 所述的方法,其中所述多个标记包括同义的标记。
30. 如权利要求 28 所述的方法,其中根据数据流中对应于所述元数据的标记来识别数据流中的元数据。
31. 如权利要求 30 所述的方法,其中所选择的标记不同于数据流中的标记。
32. 如权利要求 28 所述的方法,其中:
所述识别步骤包括识别数据流中的元数据;以及
所述映射步骤包括从多个与识别出的元数据相关联的标记中选择一个标记。
33. 如权利要求 32 所述的方法,其中所述多个中的标记是不同语言的。
34. 如权利要求 33 所述的方法,其中所选择的标记是不同语言中特定的一种。
35. 如权利要求 34 所述的方法,其中所述特定的语言是可扩展商业标记语言。
36. 如权利要求 34 所述的方法,其中所选择的标记是特定的人类的口语。
37. 如权利要求 22 所述的方法,其中所述标记符合可扩展标记语言。
38. 如权利要求 37 所述的方法,其中所述标记符合可扩展商业报告语言技术规范。
39. 如权利要求 38 所述的方法,其中在至少一个可扩展商业报告语言分类中定义所述标记。
40. 如权利要求 22 所述的方法,其中所述元数据包括至少一个文本串。
41. 如权利要求 1 所述的方法,包括将数据流从目标程序提供到转换程序的步骤,其中所述转换程序 a) 执行识别和映射步骤,以及 b) 在目标程序看来其形式为设备驱动器。
42. 如权利要求 41 所述的方法,其中所述转换程序独立于目标程序。
43. 如权利要求 41 所述的方法,其中所述转换程序和目标程序是合并到一个单独程序中的模块。
44. 如权利要求 1 所述的方法,其中所述数据流具有输出到计算机打印机的数据格式。

向数据的集合中添加与该数据相关联的元数据的方法

技术领域

[0001] 本发明一般而言涉及数据处理领域。

背景技术

[0002] 目前有成千上万的软件程序被安装在上百万个不能在彼此之间传递意义的计算机中。例如,具有很多部门或子公司的大公司经常发现一个部门或分公司所使用的会计或操作软件程序与其他部门或分公司所使用的软件或是与中央公司的程序不兼容。这需要相当大量的数据转换,并且经常导致大量的数据重输入以及伴随数据输入的附加成本和数据完整性问题。

[0003] 由于软件开发者当前所使用的程序、操作系统和软件标准是各种各样的,所以在供应商和他们的顾客之间存在大量不兼容性。这也需要相当大量的数据转换,并且经常导致大量的数据重输入及其牵涉的问题。当前计算机软件环境的未组织的和不明确的性质给管理组织(例如 SEC、FDIC、联邦和州立税务局、银行等和向它们提供报告的公司)施加了很大的负担和费用。

[0004] 为了克服这些问题已经形成并且正在形成很多标准组织,以建立供 XML(可扩展标记语言(eXtensible Markup Language))文件格式使用的定义的输入/输出词汇。XBRL(可扩展商业报告语言(eXtensible Business Reporting Language))是正被开发的一种 XML 语言格式。预期它能成为财务报表的全球标准。在该公开中我们将使用 XBRL 作为 XML 语言的例子。这并不是要将本发明限制于 XBRL 或 XML 语言。对于语义网(Semantic Web)我们找到很多相似性,在语义网中信息标记被用于促使计算机与计算机进行通话并且使该计算机由于通信而作出决定和采取行动。其他标准已经存在,而且将会开发出更多得益于本发明基本理论的标准。

[0005] 实际上现有软件应用程序都不能自动或是半自动地将常规文档或数据转换为附加有 XML 或其他标准委员会所要求的标准化信息标记的输出。在大多数情况下,这些标准本身仍处于开发中。为了确立 XML 和其他数据词典或商业词汇,需要现有应用程序和数据与这些标准词汇相关联或是附加有这些标准词汇。这个严峻的现实将会长期延迟这些标准的普遍使用,因为使公司转移到新的被设计来输出适当的信息标记的软件产品将花费几年的时间。在某些情况下这可能永远不会发生,因为实际上不可能替换传统软件系统。例如,更新所有目前使用的会计软件会是一项非常复杂的任务,不可能在任何短时期内完成。

[0006] 公认的使数据意义标准化的实用方法是把所定义的信息标记附加到正被传送的信息上。这样可以通过检查标记定义来确定数据意义。它还意味着计算机能够识别加标记的信息的“意义”,并根据该意义对其起作用。例如,具有相同“标签”的数据能够相加或相比较而不用担心将苹果和桔子相加或是比较。

[0007] 分类和它们的扩展被用来定义信息标记。例如在财务报表中,跟随有数字值的标记<销售额>指示该数字值与公司的销售额相关。跟随有数字值的<售出货物的成本>指示该值表示公司的售出货物的成本。由于总利润是销售额减去售出货物的成本,所以计算机能够

访问显示这些值的第三方报告,并且用简单的规则: < 销售额 > < 减去 > < 售出货物成本 > = < 总利润 > 容易地计算出总利润。

[0008] 因为不是所有的公司都使用相同的术语,所以标准组织所使用的分类还包括同义词和具有相同意义的替换短语。例如,销售额的同义词可能包括“收益”或“费用”。售出货物成本可以是“货物成本”或“销售成本”。这些信息标记还可以携带关于定义该标记的组织机构的信息。如果分类是由美国证券交易委员会 (US Securities&Exchange Commission) 创作的,则基于该分类的标记可能被标识为 USSEC 等等。

[0009] 因此,需要准确和有效地将数据转换成为 XML,特别是 XBRL 兼容的格式的方法和机制。例如,该转换可能包括将适当的标记添加到数据中,如在相关的 XBRL 分类中所定义的那样。还需要自动将 XML 和 XBRL 兼容的数据输入到非 XML 或非 XBRL 兼容的程序或应用程序中的方法和机制。

[0010] XBRL 基础 (XBRL Essentials),作者为 Charles Hoffman 和 Carolyn Strand,版权 2001 年由 XBRL Solutions, Inc. 所有,ISBN0-87051-353-2,在此引入作为参考。

发明内容

[0011] 在本发明的一个示例性实施例中,捕获一个数据流,识别在捕获的流中的数据,然后将识别出的数据映射到文件结构、模式 (schema) 或是分类。在本发明的示例性实施例中,输出数据流是到显示屏、存储器、硬盘驱动器、CD ROM 驱动器、软盘驱动器或打印机的数据流。该输出数据流可以通过串行或并行端口 (包括通用串行总线或“USB”、FireWire™),经由无线接口等传送,等等。在本发明的其他示例性实施例中,将识别出的数据映射到 XBRL (可扩展商业报告语言) 分类、电子表格、数据库或平面文件。

[0012] 在本发明的另一个示例性实施例中,向数据添加标记的方法包括:a) 在以电子方式表示的文件中识别数据,b) 根据一个将标记与文本串相关联的列表来选择与识别出的数据中的文本串对应的标记,以及 c) 将所选择的标记添加到以电子方式表示的文件中,以用于标记文本串和识别出的数据中与文本串相关的元素。该标记包括关于数据的信息并在一个或多个分类中被定义。如果该列表没有将标记与文本串相关联,则可以提示用户选择与识别出的数据中的文本串对应的标记。由该用户的选择所指示的关联可以被添加到所述将标记与文本串相关联的列表中。优选地,标记符合 XML (可扩展标记语言),并且也符合 XBRL (可扩展商业报告语言) 技术规范。该实施例可以通过一个可从目标程序接收以电子方式表示的文件的转换程序来实现。该转换程序可以:a) 执行识别、选择和添加步骤,以及 b) 被配置成对于目标程序看起来像是打印机驱动器。该转换程序可以独立于目标程序并与之分离。

[0013] 根据本发明的另一个实施例,提供了一种将 XBRL 兼容的数据集的至少一部分输入到非 XBRL 兼容的目标应用程序的方法。该方法包括以下步骤:从目标程序把数据输出到输出文件中,将该输出文件中的项与在一个或多个适当的 XBRL 分类中定义的标记进行用户关联,以及通过用来自数据集的具有对应标记的数据来替换输出文件中与特定标记相关联的项的数据,从而形成用于输入到目标程序的输入文件。由用户所作的关联被存储起来用于以后使用,以便可以根据所存储的关联,通过替换具有与原始输出的文件相同格式的文件中的数据来自动创建输入文件。

[0014] 根据本发明的另一个实施例,提供了一种将数据集的至少一部分输入到目标应用程序的方法,其中该数据集包括指示关于数据集中数据的信息的标记,而且其中该标记是在一个或多个分类中定义的。例如,其中数据集是 XBRL 兼容的而标记是在一个或多个 XBRL 分类中定义的。该方法包括一个用于观察用户把与标记相关联的数据输入到目标应用程序并存储与每个不同标记的数据输入相关联的击键的数据输入程序。然后,当数据输入程序接收用于输入到目标应用程序(它可以是非 XBRL 和非 XML 兼容的)的 XBRL 兼容的数据集时,该数据输入程序能够通过执行所存储的对应于与数据集中的数据相关联的标记的击键,将数据从数据集输入到目标应用程序中。当数据输入程序是自动将数据输入到目标应用程序中,并遇到一个具有没有为其存储击键的标记的数据项时,该数据输入程序可以提示用户将该数据项输入到目标应用程序,然后观察并存储该用户的击键以便将来使用。

[0015] 根据本发明的另一个实施例,提供了一种将数据集的至少一部分输入到目标数据库中的方法。该方法包括将测试数据输入到目标数据库,然后搜索或扫描数据库中对应于该测试数据的型式(pattern)。独立于数据库的模式识别应用程序可以用于该目的。根据搜索结果对数据库结构进行建模。此后,可以使用该建模的结构直接访问数据库。特别地,该建模过程包括将数据库结构内的位置与标记相关联,其中该标记对应于在搜索步骤期间在该位置处找到的测试数据元素。然后,根据与位置和元素都相关联的标记,使用例如一个独立的软件应用程序,可以将数据元素直接输入到数据库内的特定位置。

[0016] 本发明的示例性实施例包括包含有已知标记或术语的同义词、或标记和/或术语之间的同义链接的同义词词典,以方便自动的或用户辅助的映射。该词典可以包括不是分类或模式(例如 XML 分类)的部分而是与分类、模式等中的术语同义相关的术语。在本发明的一个示例性实施例中,该同义词词典包括外语,从而使得标记或数据可以从一种语言映射到另一种语言。在本发明的一个示例性实施例中,货币值在数据流中被识别,并根据已知或指定的汇率被转换成不同货币中的对应值(例如从日元转换到美元)。根据本发明的实施例,映射过程将数据从一种标准转换到另一种标准,例如从美国 GAPP(公认会计准则)转换到国际 GAPP。根据本发明的实施例,映射过程包括在例如最小化文件大小是重要的情况下用其他标记来替换对应于识别出的数据的标记。

[0017] 根据本发明的实施例,从第一个计算机平台或系统输出的数据可以通过在第一平台上的软件模块从第一格式自动转换到中间格式,被传送到第二平台或系统,然后通过第二平台上的第二软件模块从中间格式转换到第二格式。例如,中间格式可以是 XML 分类而软件模块可以有效地“转换”,从而使得数据可以在两个平台之间被透明地交换而不管第一和第二格式是否兼容或对于两个平台的每个来说是否是已知的。中间格式还可以是加密的,例如用于安全传送。

[0018] 根据本发明的实施例,上述处理步骤和机制实时或非实时地以远程或分布式的方式执行。

[0019] 附图简述

[0020] 结合附图(其中相同的附图标记被指定到相同的元素),根据下面的优选实施例的详细描述,本发明的其他目标和优点对于本领域的技术人员来说将变得显而易见,其中:

[0021] 图 1A 示出根据本发明示例性实施例的流程图。

[0022] 图 1B 示出根据本发明示例性实施例的流程图。

[0023] 图 2 示出根据本发明示例性实施例的流程图。

[0024] 图 3 示出根据本发明示例性实施例的流程图。

[0025] 图 4 示出根据本发明示例性实施例的流程图。

[0026] 图 5 示出根据本发明实施例的目标程序和转换程序之间的关系。

[0027] 图 6 示出根据本发明实施例的目标模块和转换模块之间的关系。

[0028] 图 7 示出本发明示例性实施例中的软件层。

[0029] 优选实施例详述

[0030] 根据如图 1A 所示的本发明实施例,在步骤 150 捕获数据流,在步骤 152 识别在捕获到的数据流中的数据,然后在步骤 154 将识别出的数据映射到文件结构、模式或分类。输出数据流是到显示屏、存储器、硬盘驱动器、CD ROM 驱动器、软盘驱动器或打印机的数据流。该输出数据流可以通过串行或并行端口(包括通用串行总线或“USB”、FireWire™等),经由无线接口等在计算机内部传送,并可以经由复制或重定向,在传送的任何点上通过软件和/或硬件机制而被捕获。将识别出的数据映射到 XBRL(可扩展商业报告语言)分类、电子表格、数据库、XML(可扩展标记语言)分类、标准(例如美国 GAAP 或国际 GAAP)或平面文件。当将识别出的数据映射到平面文件时,也可以生成技术规范或“数据定义”文件来指示在平面文件的不同位置处(例如,在不同列中、在给定文本串内部的不同位置处等)的信息意义或特性,并且任选地指示平面文件中不连续信息元素之间或信息组之间的定界符(例如,制表符、逗号、空格、分号等)。例如,可以根据与平面文件相关的已知技术和格式来生成平面文件和伴随的数据定义。

[0031] 图 1A 所示的实施例可以如图 1B 所示的那样实现。根据图 1B 所示的本发明的示例性实施例,一种向数据添加标记的方法包括:a)在以电子方式表示的文件中识别数据,b)根据将标记与元数据相关联的列表选择对应于识别出的数据中的元数据的标记,以及c)将所选择的标记添加到以电子方式表示的文件中,用来标记元数据和/或在识别出的数据中与元数据相关联的元素。该标记包括关于数据的信息并在一个或多个分类中被定义。在本申请的上下文中,“元数据”或“元信息”是关于数据的数据,或是描述其他信息的信息。在这个例子中,识别出的数据中的元数据识别或描述在识别出的数据内部的其他数据元素,并且可以包括例如文本串、各种控制字符(例如,各种 ASCII 控制字符)等。例如,在捕获到的数据流或文件中的元数据可以被用来识别该元数据所涉及的数据,然后涉及识别出的数据的附加元数据可以被添加到捕获的数据流或文件中。例如,该列表可以包含来自多个分类、标准等的标记,包括来自多种语言、链接同义或相关的标记。当在捕获到的数据流或文件中识别出来自第一分类等的标记时,它所标记的数据元素还可以进一步用来自第二个不同分类、标准等的对应标记进行标记。因此,现在能够识别第二分类而不是第一分类的计算机程序将能够使用或识别以及组织该数据流或文件中的信息。可以通过添加第二分类的新标记,以及任选地从第一分类(或标准、模式等)除去旧标记来形成新的已转换的数据流或文件。

[0032] 如果该列表没有将标记与识别出的数据中的元数据相关联,则可以提示用户选择对应于该元数据的标记。由该用户选择所指示的关联可以被添加到将标记与元数据相关联的列表中。优选地,该标记符合 XML(可扩展标记语言),并且也符合 XBRL(可扩展商业报告

语言)技术规范。当然,该标记还可以符合电子表格、关系数据库以及其他文件结构或模式或标准的数据格式。

[0033] 该实施例可以通过从目标程序接收以电子方式表示的文件的转换程序来实现。该转换程序可以:a) 执行识别、选择和添加步骤,以及 b) 可以被配置成对于目标程序看起来像是对于目标程序已知的软件类型。例如,该转换程序可以对于目标程序看起来像是打印机驱动器。

[0034] 该转换程序可以独立于目标程序并与之分离。该转换程序还可以全部驻留在与目标程序相同的计算机或系统上,或者可以远程地位于不同系统中,或者分布在不同系统之间。转换模块可以是单个模块、或是多个合作模块。转换程序或模块用来识别元数据并添加对应的元数据的列表和/或同义词词典可以被存储为与程序或模块分离的数据文件,而且可以经由例如因特网服务器被远程存储或访问。

[0035] 例如,数据流可以在信息供应者的地点被捕获,被传送(作为实时数据流或作为包含捕获的数据流内容的数据文件)到另一个位置(例如中间位置或信息接收者的地点),然后被提供到信息接收者的地点。信息供应者的计算机可能具有例如模拟打印机驱动器的转换程序,当信息被输出用于映射时选择该转换程序。输出会被提供到转换程序,然后(通过电子邮件、调制解调器、软盘上的文件等)被传送到信息接收机。在信息接收机上的转换程序然后会打开或接收数据,并将其映射到目标输入程序可用的批处理文件格式,或映射到被写入以用于更新数据库的程序可用的文件格式。

[0036] 在供应者和接收机上的转换程序可以是相同的,并且都能够接收、发送和映射数据,或者可以具有不同的性能。例如,转换程序可以被配置成处理中间格式,使得在信息供应者处的转换程序将数据映射到中间格式,并将中间格式的数据发送到在接收机上的转换程序。接收机将数据从中间格式映射到在接收机上有用(或者如用户所要求)的另一种格式。程序可以是不同版本,从而使得该转换程序比在接收机处的转换程序识别更多的格式,由此能够将更多的格式映射到中间格式,或是从中间格式映射到更多的格式。此外或者作为替换,在接收机上的转换程序可以被配置成(或者其特点为)只能将数据从中间格式映射出来到另一种格式,而不能将数据映射到中间格式,这与 Adobe Acrobat™ 阅读器可以打开和查看但是不能创建.pdf 文件几乎相同。转换程序还可以被配置成自动操作而不需要用户干预。例如,根据例如在供应机上适当的规则或要求(例如,用户允许公众访问供应机上的信息的预先核准),在供应机上的转换程序能够响应来自接收机上转换程序的请求而自动发送数据。供应机和接收机可以经由因特网进行通信。例如,供应机可以与因特网连接或者起 web 服务器的作用,而接收机可以与因特网连接或者起 web 浏览器的作用。此外,中间格式可以被加密,并且在接收机上以对于接收机用户透明的方式来解密。例如,加密/解密机制可以是转换程序的专有功能。

[0037] 转换程序可以通过例如向文件添加适当的 XBRL 标记来改变或转换其从目标程序接收的文件。可替换地,转换程序可以将来自从目标程序接收的文件中的数据与所选择的标记相结合,以生成并输出一个新的已转换的文件。作为另外的替换,例如在从一种标准或语言转换到另一种标准或语言时,转换程序可以用新添加的标记来替换文件中的标记。这在希望将已转换的文件或已转换的数据流的大小和复杂性最小化时是有利的。

[0038] 如图 1B 中所示,在第一个步骤 102 中,以电子方式表示的文件中的数据被识别。接

下来,在步骤 104 中,根据将标记与文本串相关联的列表来选择对应于元数据(例如识别出的数据中的文本串)的标记。尽管在图 1B 中特别地提到“文本串”,但是可以用“元数据”代替每个出现的“文本串”。换句话说,图 1B 所示的构思还可以用于所有其他的元数据形式,而不仅仅是文本串。这也适用于在此描述的其他实施例。

[0039] 从步骤 104 控制进行到步骤 106,在其中判断是否已经遇到未识别的文本串或者在列表上不具有相关联的标记的文本串。如果是,则控制进行到步骤 108,在其中提示用户选择对应于该文本串的标记。例如,可以在弹出窗口中或作为对话框的部分向用户提供一个或多个分类,从而使得可以迅速并且容易地选择正确的标记。

[0040] 从步骤 108 控制进行到步骤 110。在步骤 110 中,存储用户响应于上述提示而选择的关联以便将来使用。从步骤 110 控制进行到步骤 112。如果在步骤 106 中判断是否定的,则控制从步骤 106 进行到步骤 112。

[0041] 在步骤 112 中,判断是否已经(例如,使用列表)为识别出的数据中的所有相关文本串都选择了标记。这里的假设是存在以某种形式与每个数据相关联的标记,它可以被用来将该数据映射到(例如)XBRL 分类中的适当标记。执行这种功能的软件应用程序可以运用一定程度的智能,以滤出无关或多余的文本,并用来正确解释文本和附近的数据。例如,在会计系统的输出(比方说,资产负债表)中,输出可以包含一个报告页眉(Header)和一个报告页脚(Footer),根据情况它们中的一个或两个都不需要被翻译。而且,可能正被解释并与 XBRL 标记相关的文本可以复盖不止一条线,但涉及文本的数据将只在一条线上。在这种情况下,软件应用程序会适当地合并多条线。此外,还有可能文本是涉及或适用于多个数据项的文本串,例如财务报表具有名为“现存现金”的文本标记和另一个用于报告时间“2000”的标记。数据在文件中的布置或位置还可以帮助指示对于数据来说哪个 XBRL 标记是合适的。关于与有助于对其进行标记的其他数据有关系的的数据的位置的任何信息(例如,文档中的一种安排,它能显示嵌套在像分级结构那样的另一项中的特定位置的数据项),可以被用来帮助确定该数据的适当 XBRL。

[0042] 如果在步骤 112 中判断是否定的,则控制返回到步骤 104。如果在步骤 112 中判断是肯定的,则控制从步骤 112 进行到步骤 114。在步骤 114 中,依照所选择的标记重新格式化数据。换句话说,根据数据同在一个或多个 XBRL 分类中定义的标记之间被确定的对应关系来重新格式化数据。这个重新格式化可以包括添加对应的 XBRL 标记到数据中。如在步骤 116 中所指示的,重新格式化还可以包括依照所选择的 / 对应的 XBRL 标记的分级结构对数据重新排序。

[0043] 总之,转换程序可以以各种方式转换数据,包括插入和 / 或翻译用于描述、表征和 / 或组织数据的信息标记或标签,以便使得数据更可用。转换程序可以与各种操作系统适当兼容,包括(但不局限于)MSWindows、Unix、Mac OS、Solaris、Linux 等。转换程序可以获得将以任何不同格式转换的数据文件,包括作为数据库文件、平面文件、EDI、屏面数据,或任何其他可以以数字格式分析的数据集合或数据流。转换程序还可以以任何适当的格式输出包括已转换数据的转换文件。例如,输出文件可以是 XBRL 兼容的任何格式。

[0044] 转换程序还可以起动或调用应用程序或子模块来验证输出文件,并且可以起动比较程序,用以通过将文件中的文本串与标准化的 XBRL 分类进行比较来分析接收到的文件。然后,转换程序可以将文件中的文本串与适当的 XBRL 分类(包括同义词)进行比较。可以

通过分析数据或利用 Rev-Gen 模式识别扫描技术来进行该比较。也可以检查任何先前的 XBRL 信息标记到数据的用户映射。

[0045] 转换程序还可以将适当的 XBRL 信息标记链接到相关的信息,只要这种链接在没有用户干预的情况下可以被清楚地建立。任何不能自动被识别并与 XBRL 分类信息标记链接的文本串将在第一次出现时呈现给用户。用户将使用拖放 (drag and drop) 或者任何其他方便的映射技术将所讨论的信息与适当的 XBRL 信息标记 (标签) 链接起来。

[0046] 例如,在公司使用这种技术第一次公布财政报表时,公司名称可能没有被识别为 < 公司名称 > 数据。为了将 < 公司名称 > 标记与公司名称数据链接起来,用户只要将 < 公司名称 > 信息标记拖动到公司的名称,该链接就被建立了。这个链接将会保留在转换程序中用于随后的报告,所以用户只需要进行一次这种连接。

[0047] 转换程序还可以创建新的 XBRL 输出文件,它包括所有适当信息标记、式样信息和 XBRL 兼容的合适的 XML 文件扩展。一旦 XBRL 信息标记已经链接到适当的数据,在产生随后的报告时可以绕过一些步骤,除非应用程序中的术语已经改变或已经添加新术语到报告中。

[0048] 本发明的示例性实施例包括包含已知标记或术语的同义词或标记和 / 或术语之间的同义链接的同义词词典,以方便自动或用户辅助的映射。例如,在捕获到的数据流或文件正被映射到的标准、模式或分类中的已知标记是“销售额”的情况下,该词典可以包括例如“费用”和“收益”的同义词,使得当在捕获到的数据流中识别出同义词时,它们所指示的数据将会被适当地映射到标记“销售额”(或用标记“销售额”来标记)。同义词词典可以被并入到将数据和元数据相关联的列表中。该词典可以包括不是例如 XML 分类的分类或模式的部分,但它们与分类、模式等中的术语同义相关。在本发明的一个示例性实施例中,同义词词典包括外语,从而使得标记或数据可以从一种语言映射到另一种语言。

[0049] 例如,转换程序还可以被用于将文档中的术语从一种语言翻译成另一种语言。例如,把数据和元数据相关联的列表 (转换程序使用它来识别数据并选择附加或替代的标记),可以包括各种语言或部分语言以及指示各种语言之中的同义词的链接。语言部分可以是:例如,美国 GAAP 中出现的英语描述性术语和法语、德语、西班牙语等中的对应同义词、以及可能出现在其他标准 (例如国际 GAAP) 中的相似术语。因此,用户可以向转换程序提供包含符合美国 GAAP 的金融信息的文档,并指定转换程序用法语词汇代替英语词汇来输出该文档。用户还可以要求转换程序用德语词汇代替英语词汇来将美国 GAAP 文档转换成国际 GAAP 文档等等。用户可以指定所希望的输出语言,并且可以任意地指定原始语言。例如当转换程序在捕获到的数据中找到与它的列表中的标记相对应并且它知道属于一种特定语言的标记时,该转换程序能够自动识别原始语言。

[0050] 此外,在本发明的示例性实施例中,转换程序可以被用来识别在捕获到的数据流或文件中识别出的货币值,然后根据已知或指定的汇率把识别出的货币值转换到不同货币中的对应值 (例如,从日元转换到美元)。可以使用默认的汇率,例如,在 a) 创建原始数据, b) 捕获数据流或文件, c) 执行转换,或 d) 用户指示的日期时有有效的汇率。用户还可以指定汇率。

[0051] 根据图 2 中说明的本发明的另一个实施例,提供了一种用于将至少部分的 XBRL 兼容的数据集输入到非 XBRL 兼容的目标应用程序中的方法。该方法包括以下步骤:从目标程

序输出数据到输出文件中；将输出文件中的项与在一个或多个适当的 XBRL 分类中定义的标记进行用户关联；以及通过利用来自数据集的具有对应标记的数据来代替输出文件中与特定标记相关联的项的数据，从而形成用于输入到目标程序中的输入文件。存储用户所作的关联用于以后使用，因而，可以根据所存储的关联、通过替换文件中具有与原始输出的文件相同格式的数据来自动地创建输入文件。可以根据输出文件的结构和用户所作的关联来生成输入文件模板，然后可以通过根据与正在被输入的数据和输入文件模板中的项都相关联的标记来输入数据的方法，用数据填充输入文件模板，从而形成输入文件。当然可以再次使用该模板以便输入不同的数据集。用户可以以任何适当或适合的方式指示输出/输入文件格式中的项之间的关联。例如，当用数据填充一个用于输入到目标程序或目标应用程序中的（空的）输入文件模板时，用户可以将与标记相关联的数据插入到输出文件的各种项中，然后软件可以扫描输出文件中的项，根据新输入的数据识别相关联的标记，然后存储该关联用于以后使用。输出文件的结构连同所存储的关联可以表示一个输入文件模板。新输入的数据可以包括标记本身。可替换地，软件可以为输出文件中的每项提供标记列表，用户可以从该列表中选择一个或多个合适的标记来指示该关联，随后存储该关联。例如，通过识别与一个要被输入的数据相关联的标记、在模板中定位与该相同标记相关联的项、把数据输入到模板中定位的项中、以及对要被输入的数据集中的所有数据重复这些步骤，可以用用于输入的数据来填充模板。

[0052] 如图 2 所示，在步骤 202，从目标应用程序或程序输出数据输出到文件中。控制从步骤 202 进行到步骤 204，在这里用户将输出文件中的项与标记（例如在 XBRL 分类中定义的标记）相关联。控制从步骤 204 进行到步骤 206，在这里存储用户所作的关联。控制从步骤 206 进行到步骤 208，在这里通过用具有对应标记的新数据代替输出文件中与（例如 XBRL）标记相关联的项或位置的数据，从而生成输入文件。

[0053] 在图 3 中说明的本发明的另一个实施例中，提供了一种用于向目标应用程序输入至少部分数据集的方法，其中该数据集包括指示关于数据集中的数据的信息的标记，并且其中在一个或多个分类中定义这些标记。例如，在这里数据集是 XBRL 兼容的并且这些标记是在一个或多个 XBRL 分类中定义的。该方法包括一个程序，它观察用户将与这些标记相关联的数据输入到目标应用程序中，并存储每个不同标记的与数据输入相关联的击键。因此，当数据输入程序接收到用于输入到目标应用程序（它可以是非 XBRL 和非 XML 兼容的）中的 XBRL 兼容的数据集时，该程序或一个不同的程序可以通过执行所存储的对应于与数据集中的数据相关联的标记的击键，把来自数据集的数据输入到目标应用程序。如果自动把数据输入到目标应用程序中的程序遇到具有没有为其存储击键的标记的数据项，则该程序可以提示用户把数据项输入到目标应用程序中，并随后观察和存储用户的击键以便将来使用。

[0054] 如图 3 所示，在第一个步骤 302 中，第一软件应用程序观察用户将与标记相关联的数据输入到目标应用程序中。控制从步骤 302 进行到步骤 304，其中第一应用程序存储观察到的击键，这些击键与每个不同标记（例如 XBRL 标记）的数据输入相关联。从步骤 304 控制进行到步骤 306，在这里第一应用程序接收用于输入到目标应用程序的数据集。从步骤 306 控制进行到步骤 308，在这里第一应用程序通过执行所存储的对应于与数据集中的数据相关联的标记的击键，把来自数据集的数据输入到目标应用程序中。从步骤 308 控制进行到步骤 310，在这里第一应用程序判断它是否已经遇到数据集中没有为其存储击键的

数据项。换句话说,数据集中是否存在任何这样的数据,该数据具有一个第一应用程序没有为其存储或观察任何击键的标记。如果是的话,则控制进行到步骤 312,在这里第一应用程序提示用户将数据项输入到目标应用程序中,否则提供合适的击键顺序以便将该数据项输入到目标应用程序中。例如,可以从预先记录的击键顺序菜单或是一组预先记录的击键顺序中选择一个合适的顺序。从步骤 312 控制进行到步骤 314,在这里由第一应用程序存储所提供的击键顺序以便将来使用。从步骤 314,控制进行到步骤 316。

[0055] 在步骤 310 中,如果判断是否定的,则控制进行到步骤 316。

[0056] 在步骤 316 中,第一应用程序判断是否数据集中的所有相关数据都已经被输入到目标应用程序中。如果是的话,则控制进行到步骤 318,在这里结束该过程。如果否的话,则控制返回到步骤 308。“相关”数据可以依据关于图 1 的步骤 112 的上述考虑进行确定或处理。

[0057] 根据图 4 中说明的本发明的另一个实施例,提供了一种用于把至少部分数据集输入或输进目标数据库的方法。该方法包括将测试数据输入到目标数据库中,然后搜索或扫描数据库寻找对应于该测试数据的模式。为此目的可以使用一种独立于该数据库的模式识别应用程序。根据搜索结果对数据库的结构进行建模。此后,可以使用该建模的结构来直接访问数据库。特别地,该建模过程包括将在数据库结构内部的位置与标记相关联,其中这些标记对应于在搜索步骤期间在这些位置处找到的测试数据元素。然后,可以使用例如一个独立的软件应用程序根据与位置和元素都相关联的标记将数据元素直接插入到数据库中的特定位置。

[0058] 如图 4 所示,在步骤 402 中,将测试数据集输入或输进目标数据库中。优选地,以一种常规的方式将该测试数据集输入到数据库中,例如借助于通过管理该数据库的应用程序接口的按键输入。数据库可以从管理应用程序分离出来,或者可以嵌入在管理应用程序内。从步骤 402 控制进行到步骤 404,在这里由独立的软件应用程序(例如就像由英国公司 RevGen Plc. 制造并且由它们的美国分支机构 Generos 公司经销的那些模式识别应用程序)扫描数据库。独立的应用程序搜索或扫描数据库中对应于测试数据集的模式。

[0059] 从步骤 404 控制进行到步骤 406,在这里独立的应用程序(例如,模式识别应用程序或另外的单独的应用程序)根据搜索/扫描结果来构造一个数据库结构的模型。从步骤 406 控制进行到步骤 408,在这里数据库结构中的位置与标记(例如在一个或多个 XBRL 分类中定义的标记)相关联。这些标记对应于在搜索/扫描期间在数据库结构中的那些位置处所找到的测试数据元素。从步骤 408 控制进行到步骤 410,在这里根据与位置和元素都相关联的标记,将来自数据集的元素直接输入到数据库中。

[0060] 图 5 显示符合图 1 中所描述的实施例的转换程序。如图 5 所示,转换程序可以独立于目标程序或是与目标程序分开。特别地,目标程序 502 提供一个输出文件(例如打印文件)到转换程序 504。转换程序 504 被配置以便使它对于目标程序看起来象打印机驱动器。转换程序 504 不需要或不执行目标程序 502 的任何修改或改变,并且可以被设计或配置成与市场上可以买到的目标程序(例如电子表格、会计程序、文字处理程序等)相兼容。

[0061] 图 6 显示图 1 中所描述的实施例的转换程序和目标程序,它们可以备选地被分别实现为共同处于应用程序 606 内部的模块 604 和目标程序模块 602。模块 604 看起来象是模块 602 的打印机驱动器。例如,转换程序模块 604 可以被实现为 DLL、OCX、Active X 控制

程序、或者可以销售给软件供应商以便集成到独立开发的应用程序中的任何其他形式。

[0062] 图 7 显示本发明的各种实施例关于计算机的软件层的示例性结构。特别地,图 7 显示一个应用层 702,在该层任何 Windows™ 应用程序 701 都可以在计算机上执行。在层 702 下面是 Windows™ OS(操作系统)层 704,在该层可以找到 GDI(图形接口设备)703。在层 704 下面是低级 OS 接口层 706,在该层可以找到本发明的 XBRL 打印机驱动器 705。在层 706 下面是具有 XML 映射代理 707 的 XML 映射代理层 708。在层 708 下面是数据转换层 710,该层包括以各种格式(包括 712-722 所示的格式(分别为 HTML、Excel™、XML、SQL、xBase 和 ASCII))输出或可以输出数据的数据转换器 709。根据本发明的各种示例性实施例,执行本发明的各种功能的转换程序包括 XBRL 打印机驱动器 705、XBRL 映射代理 707 和数据转换器 709。尽管元素 705、707 显示为与 XBRL 相关,但是元素 705、707 也可以与任何或所有上述的格式、分类、协议、标准等以及它们的等同物相关。此外,格式 712-722 是示例性的而不是限制性的。

[0063] 关于每个上述的实施例,可以使用拖放、剪贴、从所提供的菜单选择选项、键盘输入或任何其他适合的技术来产生由用户提供的信息(例如来自目标应用程序的数据与 XBRL 标记或标签之间的关联)。此外,所述的实施例可以以各种方式组合。从目标程序或目标应用程序提取数据可以包括:直接从文件或从数据库中访问数据而不需要运行或起动本源(parent)(目标)应用程序,从显示屏或窗口擦除数据等等,这些可以作为对获得打印文件的补充或是用来代替该方式。

[0064] 本领域的技术人员应该理解:这里所描述的软件功能可以以各种方式实现为:a)在例如具有中央微处理器的台式计算机的硬件机器上运行的软件指令、b)适当配置的现场可编程门阵列(FPGA)、c)专用集成电路(ASIC)、或任何其他等同或合适的计算设备。

[0065] 本领域的技术人员应该理解:在不偏离本发明的精神或本质特征的情况下,可以以其他特定形式实现本发明,并且本发明并不局限于在此所述的特定实施例。因此,从各方面来考虑,目前公开的实施例都是示例性的而不是限制性的。本发明的范围是由所附的权利要求而不是上述说明书指示的,并且在本发明的意义和范围以及等同物范围之内的所有变化都包含在其中。

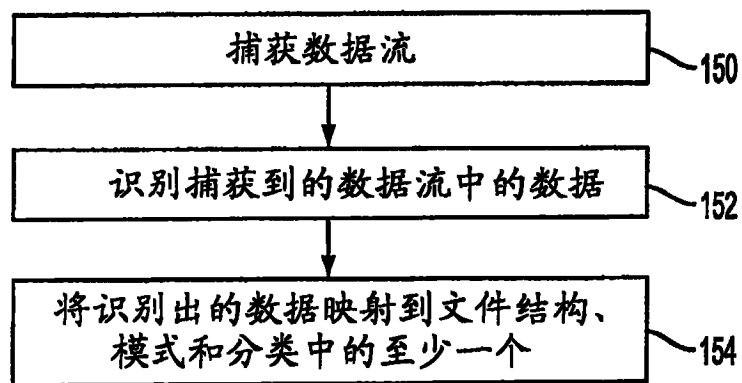


图 1A

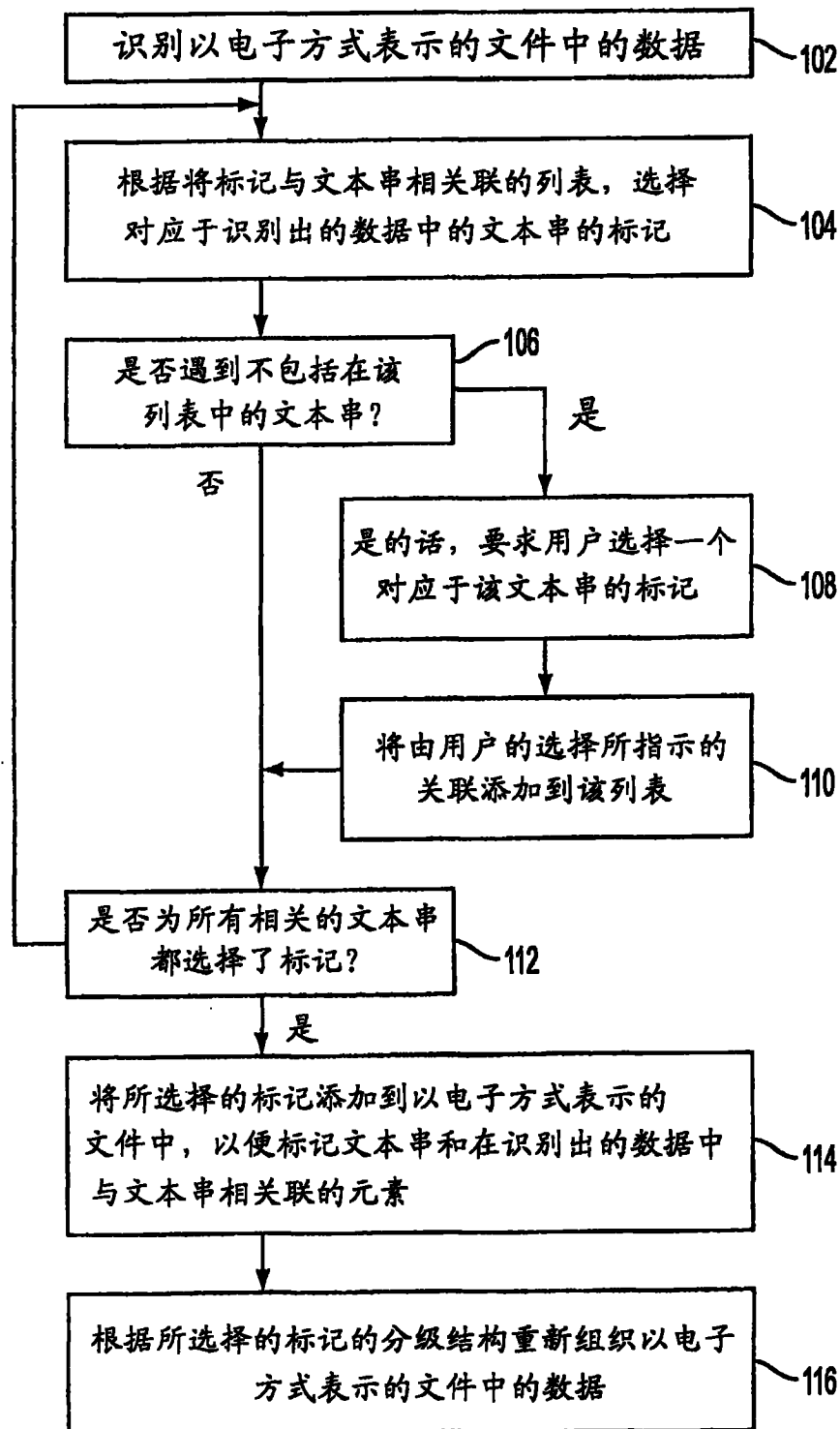


图 1B

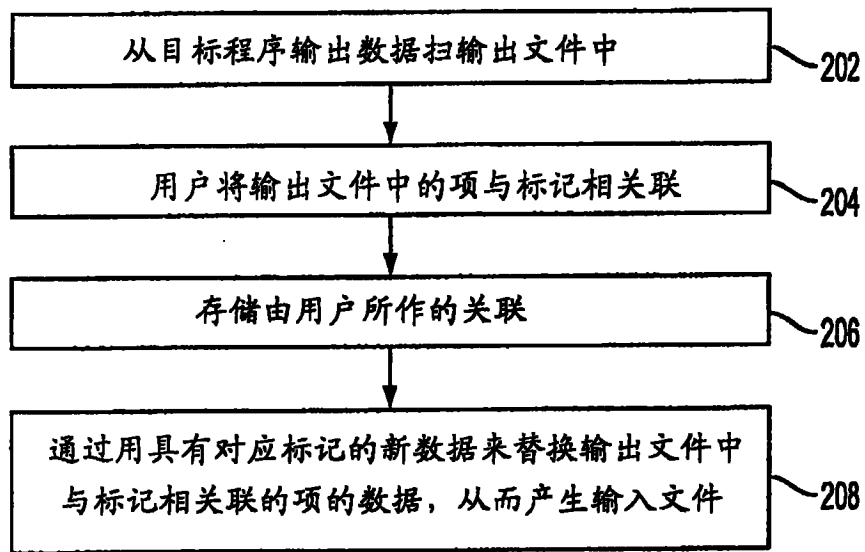


图 2

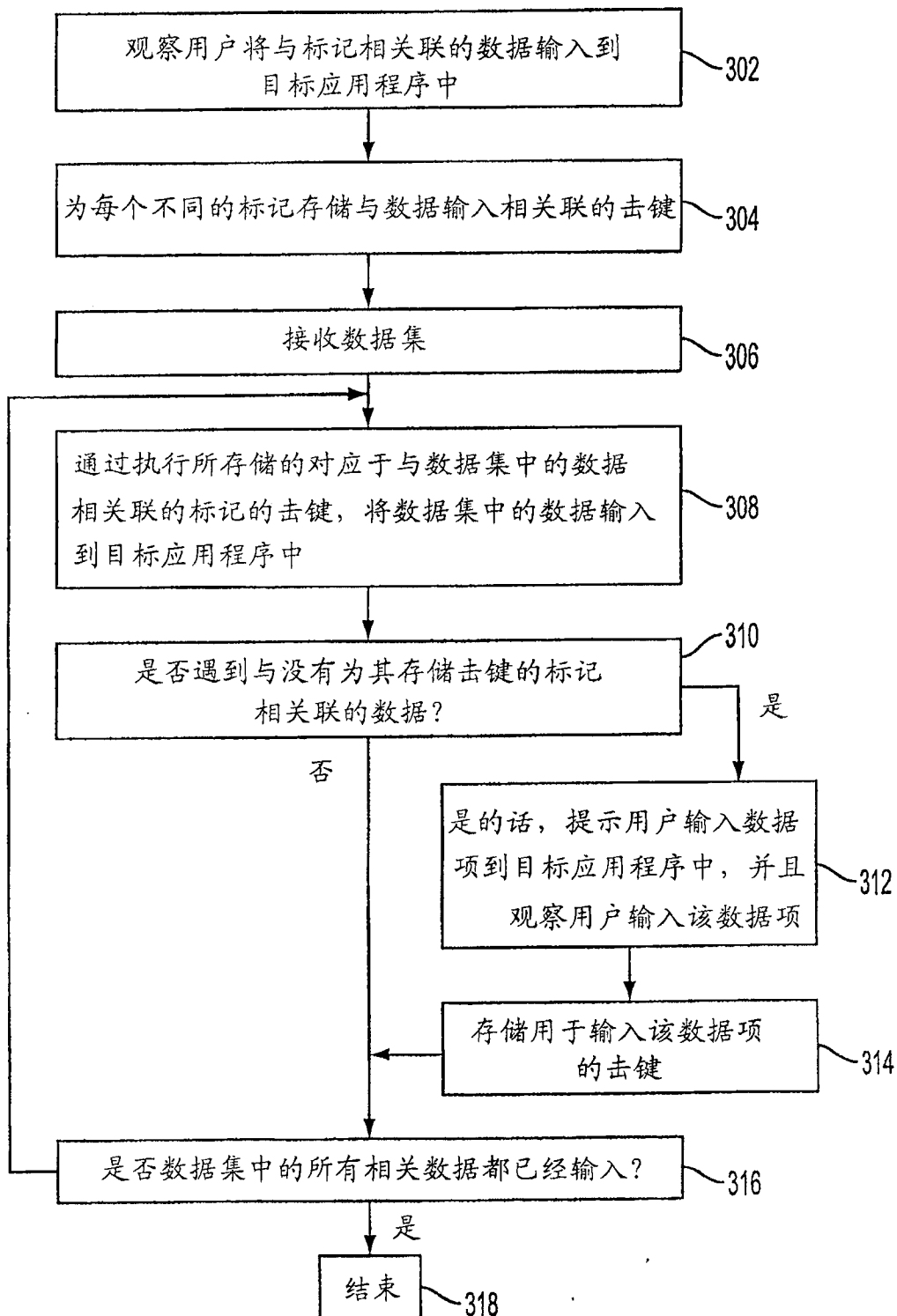


图 3

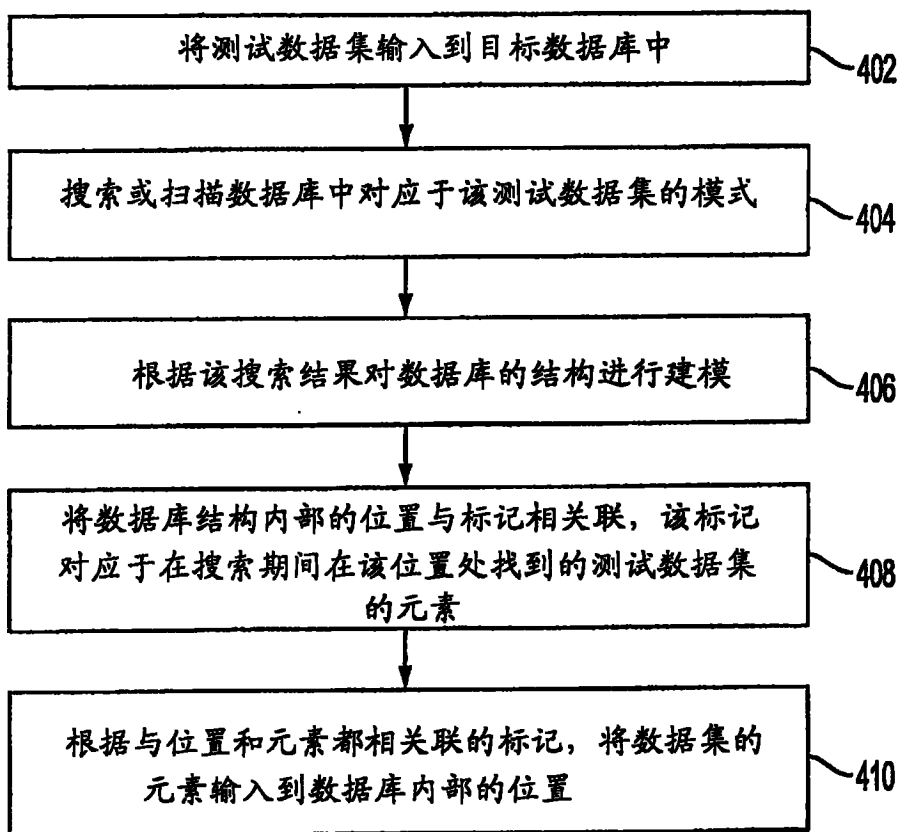


图 4

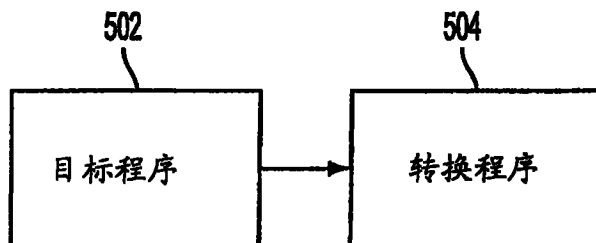


图 5

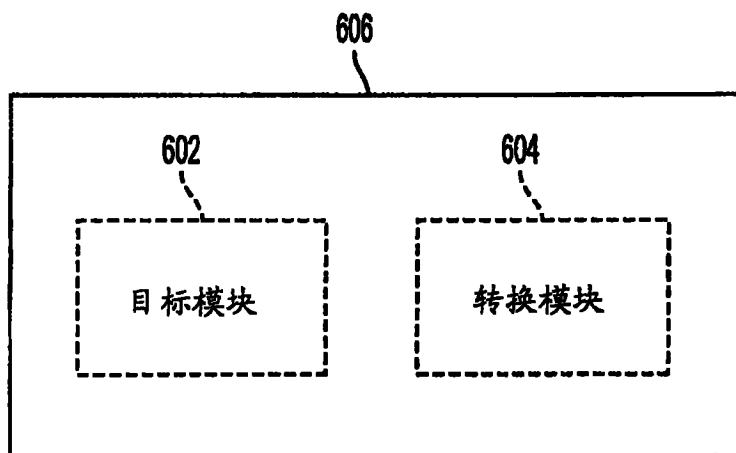


图 6

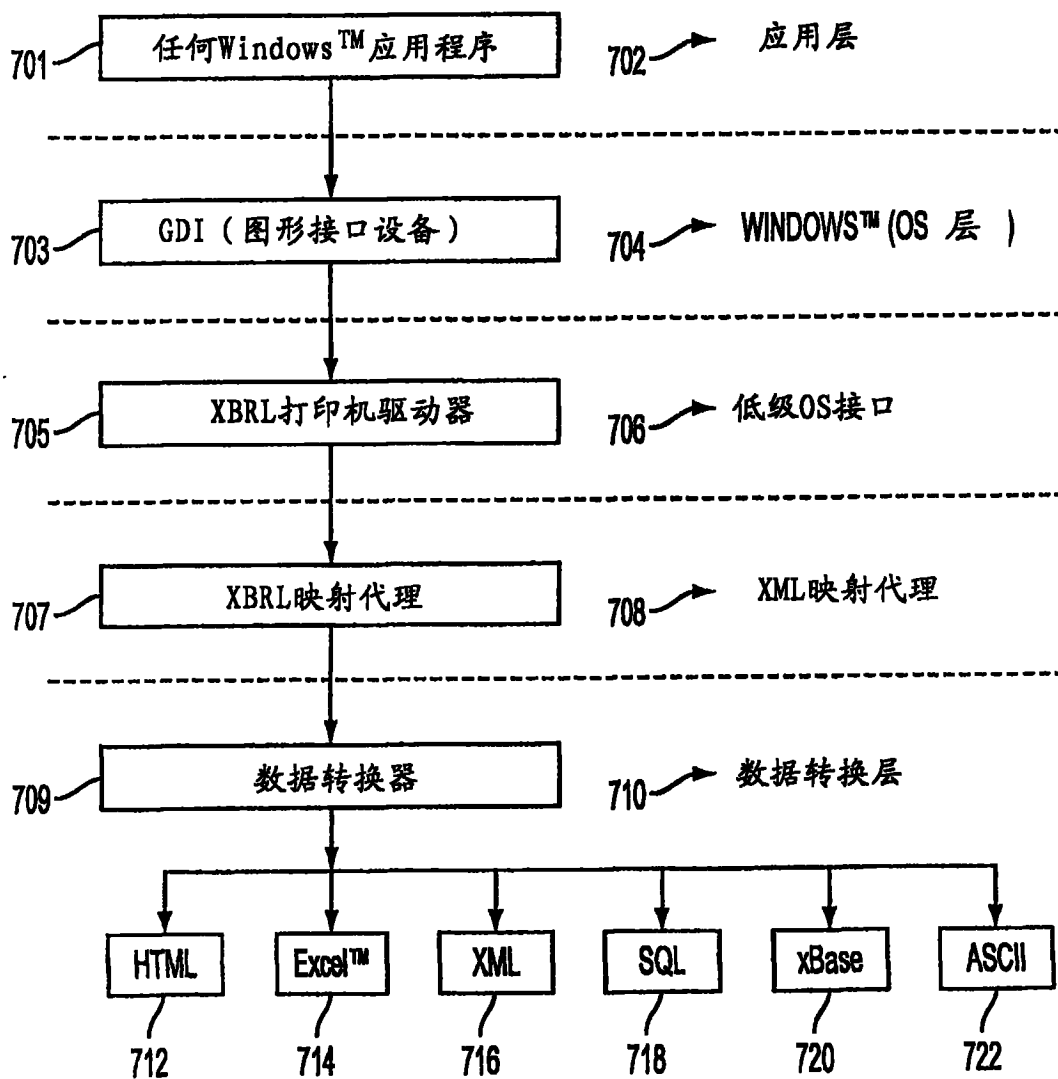


图 7