



(12) 发明专利

(10) 授权公告号 CN 102713965 B

(45) 授权公告日 2015. 11. 25

(21) 申请号 201180006679. X

(22) 申请日 2011. 01. 05

(30) 优先权数据

12/691, 130 2010. 01. 21 US

(85) PCT国际申请进入国家阶段日

2012. 07. 20

(86) PCT国际申请的申请数据

PCT/US2011/020181 2011. 01. 05

(87) PCT国际申请的公布数据

W02011/090817 EN 2011. 07. 28

(73) 专利权人 微软技术许可有限责任公司

地址 美国华盛顿州

(72) 发明人 R·奥奇 E·佩蒂 B·兰伯特

G·莫罗米萨托

(74) 专利代理机构 上海专利商标事务所有限公

司 31100

代理人 蔡悦

(51) Int. Cl.

G06F 17/30(2006. 01)

(56) 对比文件

US 2008/0306913 A1, 2008. 12. 11, 摘要, 说明书第 0031 ~ 0033, 0043 段, 图 1.

CN 101021862 A, 2007. 08. 22, 全文.

CN 101583949 A, 2009. 11. 18, 全文.

审查员 黄捷

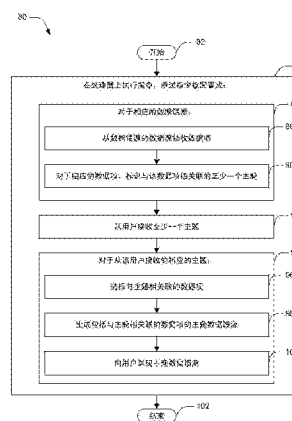
权利要求书3页 说明书20页 附图14页

(54) 发明名称

数据源的可缩放主题聚集

(57) 摘要

各个数据馈源的数据项(诸如张贴到网站的文章或 RSS 馈源中的条目)可与各个主题相关联, 但用户可能不能监视大量数据馈源, 在搜索新数据项时也不能避免先前已审阅的数据项。聚集服务可监视许多数据馈源, 执行对数据项的自动化主题评估, 并且生成主题数据馈源集。还可应用附加服务, 诸如按各种准则来过滤主题数据馈源, 将数据项从本机语言翻译成另一语言, 移除等效的数据项(诸如冗余地覆盖同一新闻故事的文章)。为改进的效率和价值可以对集中式或协作地分布的服务进行缩放, 因为可 s 代表许多用户来监视每一数据馈源并且接收和评估每一数据项。



1. 一种被配置成向用户呈现至少一个主题数据馈源的系统,所述至少一个主题数据馈源包括至少两个数据馈源的数据项,相应的数据馈源由至少一个数据源来提供,所述系统包括:

数据项接收组件,所述数据项接收组件被配置成:对于相应的数据馈源,从所述数据馈源的数据源接收数据项;

主题标识组件,所述主题标识组件被配置成:对于相应的数据项,标识与所述数据项相关联的至少一个主题;

以及

主题数据馈源呈现组件,所述主题数据馈源呈现组件被配置成:

从用户接收至少一个主题;以及

对于从所述用户接收到的相应主题:

选择所述数据馈源的引用所述主题的第一数据项;

标识所述数据馈源的与所述第一数据项相关联的第二数据项,所述第二数据馈源没有引用所述主题但因与所述第一数据项在上下文上相关而被标识为与所述主题相关;

生成包括所述第一数据项和所述第二数据项的主题数据馈源;以及

向所述用户呈现所述主题数据馈源。

2. 如权利要求 1 所述的系统,其特征在于,所述数据项接收组件包括:

至少一个数据源解析组件,所述至少一个数据源解析组件被配置成:对于数据源类型,解析具有所述数据源类型的数据源的数据馈源的数据项;以及

数据源接口,所述数据源接口被配置成:对于相应的数据源:

标识所述数据源的数据源类型,以及

调用与所述数据源的数据源类型对应的数据源解析组件,以解析所述数据源的数据项。

3. 如权利要求 1 所述的系统,其特征在于:

所述系统可访问数据存储,所述数据存储被配置成存储分别与至少一个主题相关联的数据项;以及

所述主题数据馈源呈现组件被配置成:

在从所述用户接收所述至少一个主题之后,将所述主题存储在与所述用户相关联的数据存储中;以及

在从所述用户接收对主题数据馈源的请求之后:

从所述数据存储检索与所述用户相关联的数据存储中的至少一个主题;

为相应的主题生成所述主题数据馈源;以及

向所述用户呈现所述主题数据馈源。

4. 一种向用户呈现至少一个主题数据馈源的方法,所述至少一个主题数据馈源包括至少两个数据馈源的数据项,相应的数据馈源由至少一个数据源来提供,所述方法是在具有处理器的设备上执行的且包括:

对于相应的数据馈源:

从所述数据馈源的数据源接收所述数据项,以及

对于相应的数据项,标识与所述数据项相关联的至少一个主题;

从用户接收至少一个主题 ;以及

对于从所述用户接收到的相应主题 :

选择所述数据馈源的引用所述主题的第一数据项 ;

标识所述数据馈源的与所述第一数据项相关联的第二数据项,所述第二数据项没有引用所述主题但因与所述第一数据项在上下文上相关而被标识为与所述主题相关 ;

生成包括所述第一数据项和所述第二数据项的主题数据馈源 ;以及

向所述用户呈现所述主题数据馈源。

5. 如权利要求 4 所述的方法,其特征在于,从数据馈源的数据源接收所述数据项包括 :
订阅所述数据源以接收所述数据馈源的数据项。

6. 如权利要求 4 所述的方法,其特征在于,

至少一个主题与至少一个主题元数据项相关联 ;以及

标识与数据项相关联的至少一个主题包括 :标识所述数据项的与所述主题的主题元数据相匹配的至少一个元数据项。

7. 如权利要求 4 所述的方法,其特征在于 :

所述主题具有主题标识符 ;以及

标识与数据项相关联的至少一个主题包括 :标识与所述数据项相关联的至少一个主题标识符,所述主题标识符指示所述数据项与具有所述主题标识符的主题相关联。

8. 如权利要求 4 所述的方法,其特征在于 :

至少一个数据项包括一表达 ;以及

标识与所述数据项相关联的至少一个主题包括 :用自然语言解析器来解析所述数据项的表达,以便标识由所述表达引用的至少一个主题。

9. 如权利要求 4 所述的方法,其特征在于,从用户接收主题包括 :

标识所述用户的至少一个用户描述符,以及

基于所述用户描述符,推断所述用户感兴趣的至少一个主题。

10. 如权利要求 4 所述的方法,其特征在于 :

至少一个主题与至少一个过滤器相关联,所述过滤器具有用于生成包括所述数据项的主题数据馈源的至少一个准则 ;以及

生成所述主题数据馈源包括 :生成包括与所述主题相关联的数据项的主题数据馈源,所述数据项是根据与所述主题相关联的过滤器来选择的。

11. 如权利要求 10 所述的方法,其特征在于 :

所述过滤器的至少一个准则指定可归因于与所述准则相匹配的数据项的分数 ;以及

生成所述主题数据馈源包括 :

对于相应的数据项,根据所述过滤器的至少一个准则来计算所述数据项的聚合分数 ;
以及

生成包括与所述主题相关联的数据项的主题数据馈源,所述数据项是根据所述相应的数据项的聚合分数来排序的。

12. 如权利要求 4 所述的方法,其特征在于 :

对于相应的主题,标识用户可访问的且与所述主题相关联的至少一个用户专用数据项 ;以及

生成所述主题数据馈源包括：生成包括与所述主题相关联的数据项以及与所述主题相关联的至少一个用户专用数据项的主题数据馈源。

13. 如权利要求 4 所述的方法，其特征在于，从所述主题数据馈源中移除与所述主题数据馈源的第二数据项等效的第一数据项。

14. 一种向用户呈现至少一个主题数据馈源的方法，所述至少一个主题数据馈源包括至少两个数据馈源的数据项，相应的数据馈源由至少一个数据源提供，所述方法在具有处理器的设备上执行且在可部署计算环境中表示，所述可部署计算环境包括被配置成存储用户专用数据项的第二设备，所述方法包括：

对于相应的数据馈源：

订阅所述数据源以接收所述数据馈源的数据项；

从所述数据馈源的数据源接收所述数据项，以及

对于相应的数据项，通过以下操作来标识与所述数据项相关联的至少一个主题：

标识所述数据项的与同所述主题相关联的主题元数据项相匹配的至少一个元数据项，

标识与所述数据项相关联的至少一个主题标识符，所述主题标识符指示所述数据项与所述主题的主题标识符相关联；

用自然语言解析器来解析所述数据项的表达，以便标识由所述表达引用的至少一个主题；以及

将与第二数据项相关联的至少一个主题标识成与所述数据项相关联，所述第二数据项与所述数据项相关联；

从用户接收至少一个主题；以及

对于相应的主题：

选择所述数据馈源的引用所述主题的第一数据项，所述第一数据项是根据与所述主题相关联的至少一个过滤器的至少一个准则来选择的，所述过滤器的至少一个准则指定可归因于与所述准则相匹配的数据项的分数；

标识所述数据馈源的与第一数据项相关联的第二数据项，所述第二数据项没有引用所述主题但因与所述第一数据项在上下文上相关而被标识为与所述主题相关；

对于针对所述主题数据馈源所选择的相应的第一数据项和第二数据项，根据所述过滤器的至少一个准则来计算所述数据项的聚合分数；

标识存储在所述第二设备上且与所述主题相关联的至少一个用户专用数据项；

通过所述可部署计算环境从所述第二设备检索至少一个用户专用数据项；

生成包括所述第一数据项和第二数据项以及与所述主题相关联的至少一个用户专用数据项的主题数据馈源，所述数据项是根据所述相应的数据项的聚合分数来排序的；

从所述主题数据馈源中移除与所述主题数据馈源的另一数据项等效的数据项；

根据所述主题的主题叙述来组织所述主题数据馈源；以及

向所述用户呈现所述主题叙述；以及

在向所述用户呈现所述主题叙述之后，在接收与主题数据馈源的主题相关联的至少一个数据项后，更新向用户所呈现的所述主题叙述。

数据源的可缩放主题聚集

[0001] 背景

[0002] 在许多计算场景中,用户可表达对与特定主题有关的数据馈源集感兴趣。例如,用户可能在网站上阅读关于特定主题的一组文章,或可能用订阅请求来向新闻网站订阅以向其通知关于特定主题的新文章。由于许多数据源(诸如网站或新闻组)可供应一个或多个数据馈源(诸如网站的频道或部分、新闻组内的子组),每一数据馈源供应许多数据项(诸如张贴在网站上的文章或评论、或张贴在新闻组的回复),因此用户可能需要频繁地访问各种数据源以便保持获悉关于该主题的评论。

[0003] 某些应用可被用于合并一个或多个数据馈源的数据项,且可在接收到新数据项时通知用户。然而,如果每一数据馈源包括许多数据项,则用户可能需要审阅大量数据项,其中许多数据项可能与用户感兴趣的主体不相关。此外,用户可能不愿意展开所跟随的数个数据馈源,且由此可能错过用户已选择不跟随的数据馈源的某些主题上相关的数据项。

[0004] 概述

[0005] 提供本概述以便以简化的形式介绍将在以下详细描述中进一步描述的一些概念。本概述并不旨在标识出所要求保护的主体关键因素或必要特征,也不旨在用于限定所要求保护的主体范围。

[0006] 为了监视许多数据馈源的数据项,用户可能尽力过滤所聚集的数据项集,诸如通过仅检查包括特定关键词的数据项、使用特定主题标识符(诸如 Twitter 散列标签)的数据项、或链接至特定网页或网站的数据项。然而,这一过滤机制在许多方面可能是低效的。首先,某些数据项可能以不被简单过滤器所捕捉的方式而与主题相关,例如,在网站上主题相关的帖子可能使用首字母缩写或同义词而不是关键词。尽管更具洞察力的上下文分析可能是可实现的(例如,通过自然语言解析和语言翻译算法),但这对于代表单个用户对大的数据项集合执行计算上密集的分析可能是低效的。第二,用户可能不被准许自动监视某数据源;例如,网站的可接受使用策略可能不允许对网站内容进行自动化轮询,且在轮询被检查到的情况下可能记入黑名单或拒绝向该用户提供服务。第三,即使对于准许用户自动监视的数据源,分配足够带宽来接收具有潜在大量数据项的潜在大量数据源可能是低效的或相当昂贵的。连续接收这种大的数据项集的网络传输成本对于为主题上相关的数据项监视数据馈源的场景可能是不可接受的。

[0007] 各技术可用于提高在各种数据源所供应的数据馈源集中标识面向主题的数据项、以及生成包括这种主题数据项的聚集的主题数据馈源的高效性和准确性。这些技术涉及聚集用户集(潜在的非常大的集合)的主题兴趣,并执行对数据项的主题分析和对主题数据馈源的生成。这一聚集可产生某些缩放的优点。具体而言,代表潜在大的用户集合以及潜在大的主题集合来执行评估(包括监视大量数据源和数据馈源、以及检索和分析大的数据项集合)可能更为有效。对这些技术的结果进行缩放以便向许多用户供应信息可以使得对许多数据源进行大规模监控的网络传输成本、和/或对相应数据项执行复杂的主题分析的计算资源成本变为可行。此外,在这些技术的上下文中,一个或多个数据源可能更愿意准许自动化轮询,因为这样的经轮询的数据项可被散布到大量用户而同时保留数据源的带宽使

用。通过将设备配置成接收各种数据馈源的数据项、标识与这些数据项相关联的主题、以及生成可从中向用户集提供的主题数据馈源,此处所讨论的技术相比于其他技术可实现这些以及其他优点。

[0008] 为实现上述内容和相关目的,以下描述和附图阐述了各个说明性方面和实现。这些方面和实现仅指示可以使用一个或多个方面的各种方式中的一些。结合附图阅读以下详细描述,则本公开的其他方面、优点、以及新颖特征将变得显而易见。

附图说明

[0009] 图 1 是表征用户检查各种数据源以标识与特定主题相关的数据项的示例性场景的图示。

[0010] 图 2 是表征用户订阅各种数据源以便接收与特定主题相关的数据项的示例性场景的图示。

[0011] 图 3 是根据此处所讨论的技术、表征用户指定主题集并接收与这些主题相关的主
题数据馈源的示例性场景的图示。

[0012] 图 4 是示出用于向用户呈现主题数据馈源的示例性系统的组件框图。

[0013] 图 5 是示出向用户呈现主题数据馈源的示例性方法的流程图。

[0014] 图 6 是包括被配置成实现本文中所述的措施中的一个或多个的处理器可执行指令的示例性计算机可读介质的图示。

[0015] 图 7 是其中可利用此处所讨论的技术的示例性可部署的计算环境的图示。

[0016] 图 8 是表征以包括按数据源解析组件的体系结构为特征的实施例的示例性场景的图示。

[0017] 图 9 是表征用于标识与数据项集相关联的主题的技术的示例性场景的图示。

[0018] 图 10 是表征从与各种主题相关联的数据项集中生成主题数据馈源的示例性场景的图示。

[0019] 图 11 是表征聚集包括数据项集和某些用户专用数据项的主题数据馈源的示例性场景的图示。

[0020] 图 12 是表征高效生成供应有针对性按用户来个性化主题数据馈源的用户专用数据项的主题数据馈源的示例性场景的图示。

[0021] 图 13 是表征以分层树视图来呈现主题数据馈源集的示例性场景的图示,该分层树视图具有两个数据项之间的相关性指示符。

[0022] 图 14 示出其中可实现本文中所述的措施中的一个或多个的示例性计算环境。

[0023] 详细描述

[0024] 现在参考附图来描述所要求保护的主体,所有附图中使用相同的附图标号来指代相同的元素。在以下描述中,为解释起见,阐明了众多具体细节以提供对所要求保护的主体全面理解。然而,很明显,所要求保护的主体可以在没有这些具体细节的情况下实施。在其他情况下,以框图形式示出了各个结构和设备以便于描述所要求保护的主体。

[0025] 在计算领域内,许多场景涉及这样的用户:该用户希望跟随由一个或多个数据源所供应的一个或多个数据馈源,以便保持获悉一个或多个主题。作为第一示例,该用户可访问一组网站,以便读取与各种主题有关的文章、论坛帖子、评论等,其中某些可能是用户感

兴趣的。作为第二示例,用户可订阅一个或多个真正简单聚合(RSS)馈源和/或新闻组,并且可检查接收到的数据项与感兴趣的主题的相关性。作为第三示例,用户可能希望检查本地数据源以寻找特定主题;例如,用户可监视与感兴趣的各種主题相关的电子邮件消息的电子邮件帐户,并且可检查接收到的电子邮件消息与其的相关性。在这些和其他场景中,一个或多个数据资源(诸如个体或个体组、公司或组织、以及自动化设备)可呈现一个或多个数据馈源(诸如网站部分、邮件发送列表、数据库、以及 RSS 或类似馈源)中的数据项(诸如创作的内容、文章、报道、照片、自动化的日志条目、以及数据集);每一数据项可与一个或多个主题相关,且用户可能希望呈现有与特定的主题子集相关的数据项。

[0026] 为了实现这一呈现,用户可以按各种方式来访问数据源和数据馈源。图 1 呈现了表征用户 12 的第一示例性场景 10,用户 12 希望保持获悉各种主题 20,它们可被各种数据源 14 所供应的数据馈源集 16 中所呈现的各种数据项 18 覆盖。该第一示例性场景 10 的熟悉的示例涉及 web 服务器集(作为数据源 14 来运作),其中每一个 web 服务器可供应一个或多个网页或网站的部分(每一网页或部分与数据馈源 16 相对应)。每一网页或部分供应各种文章、用户帖子或评论、照片、音频和/或视频记录、数据对象、数据集等(每一个包括数据项 18),且每一个这样的项可与用户 12 感兴趣的一个或多个主题 20 相关联。

[0027] 为了查看这样的数据项 18 并保持获悉这样的主题 20,用户 12 可选择按顺序访问每一数据馈源 16,并可以诸如在用户 12 所操作的设备的显示器 22(例如,台式计算机、诸如笔记本电脑之类的便携式计算机、或诸如智能手机之类的移动设备)上主动地审阅数据馈源 16 的呈现 24。在呈现 web 浏览的熟悉的示例中,用户 12 可选择查看 web 浏览器中的每一网页或部分。在接收到来自用户 12 的请求之后,web 服务器(作为数据源 14 来运作)可呈递包括当前数据项集 18 的所请求的网页或部分,并且可在监视器 22 上所显示的 web 浏览器内向用户 12 呈现这一呈递。例如,用户 12 可输入与第一 web 服务器相关联的 URL,该 URL 可供应第一部分(例如,诸如第一网络日志的作者之类的特定第一作者的帖子),该第一部分包括涉及第一主题 20 的第一数据项 18 以及涉及第二主题 20 的第二数据项 18。由第一 web 服务器生成的网页可在显示器 22 上作为第一呈现 24 来向用户 12 显示,第一呈现 24 可包括第一数据项 18 和第二数据项 18。接着,用户 12 可输入与第二 web 服务器相关联的 URL,该 URL 可供应表征第三数据项 18 和第四数据项 18 的第一网站部分,每一数据项 18 都与不同的主题集相关联。在显示器 22 上查看第二呈现 24 之后,用户 12 可输入与第二 web 服务器相关联的、标识同一网站不同部分(包括第三数据馈源 16)的另一 URL,并且第二 web 服务器可为用户 12 生成包括第三数据馈源 16 的数据项 18 的第三呈现 24。以此方式,用户 12 可以按自组织方式请求访问数据馈源 16,可以接收包括数据馈源 16 的当前数据项集 18 的呈现 24,并且可以标识和保持获悉与其相关的主题 20。

[0028] 第一示例性场景 10(尽管熟悉)展现了相对于其他场景的某些潜在缺点。作为第一示例,由于用户 12 标识并且主动地访问数据源 14,因此用户 12 可访问的数据项 18 的范围限于用户 12 所知且所查询的数据源集 14。例如,可能无法向用户 12 呈现由数据源 14 所呈现的用户 12 不知道的数据项 18、或用户 12 已知的但自这样的数据项 18 被呈现以来未被查询的数据项 18。相反,用户 12 必需查询相应的数据馈源 16,而不知道其中有用户 12 感兴趣的任何数据项被表征。例如,用户 12 可能重复地访问不包括任何自先前访问以来的新数据项 18、或仅包括用户 12 不感兴趣的新数据项 18 的数据馈源 16。作为第二示例,还可

能冗余地向用户 12 呈现不同数据馈源 16 中语义等同的数据项 12 (例如,两个新闻站点呈现或链接到同一文章)。例如,两次向用户 12 呈现第二数据项 18——一次是作为第一数据馈源 16 的一部分,而一次是作为第三数据馈源 16 的一部分。这可能造成用户 12 在重复地被呈现多个数据馈源 16 中的同一数据项 18 方面受挫。作为第三示例,对用户 12 而言标识数据项 18 的主题 20 以便选择和审阅那些用户 12 感兴趣的主体可能是困难的。例如,用户 12 可能必需查看特定数据馈源 16 中的所有数据项 18 以便标识那些感兴趣的数据项,包括每一数据项 18 的上下文(例如,呈现在网络日志上的第二评论可响应于第一评论,且可隐式地依赖于第一评论的上下文内容,但对于用户 12 而言在没有阅读第一评论的情况下进行标识可能是困难的)。这些冗余性和低效性可降低用户 12 (例如,经由在审阅旧的且不感兴趣的数据项方面损失的时间和带宽)和数据源 14 (例如,经由为用户 12 不太感兴趣或对用户 12 而言不太新奇的网页和项目提供冗余服务)两者的体验。例如,用户 12 可能对第二主题 20 不感兴趣,但可重复地接收并被呈现与第二主题 20 相关的数据项 18,从而徒劳地浪费计算资源(诸如带宽和存储器)以及用户 12 的注意。这些缺点不限于 web 浏览,而是可展现在其中用户 12 主动地寻求内容(诸如新闻组、事件日志、以及数据库查询)的任何场景中。

[0029] 这些缺点中某一些可通过订阅技术来实现。作为第一示例,用户 12 可订阅由数据源 14 所供应的数据馈源 16,以便从数据源 14 接收新数据项 18,而不必查询数据源 14。数据项 18 的递送可由数据源 14 来启动,如在涉及向用户 12 递送新电子邮件消息的邮件收发列表订阅中,或者可涉及数据源 14 所部署的用于监视数据馈源 16 的机制,诸如真正简单聚合(RSS)馈源,该 RSS 馈源以用户 12 可向其轮询以标识和接收新数据项 18 的可轮询数据馈源 16 来表征。这一订阅可在没有新数据项 18 的情况下减少对数据馈源 16 的主动的用户引导的请求和呈现,并且即使在没有用户引导的刷新的情况下还可向用户 12 通知新数据项 18。作为第二示例,用户 12 可将各种数据源 14 和数据馈源 16 所接收的数据项 18 进行合并,以便通过单个用户界面来访问。例如,用户 12 可通过单个电子邮件客户机来访问来自许多邮件收发列表的电子邮件消息,或者可通过 RSS 聚集器来访问 RSS 馈源集(例如,通过在用户 12 的计算机上本地执行以请求多个数据馈源 16 并且一起呈现这些数据馈源 16 的数据项 18 的聚集器应用,或通过聚集器服务,其中用户 12 可访问该聚集器服务以接收从多个数据馈源 16 所生成的由聚集器服务所聚集的数据馈源 16)。通过合并许多数据馈源 16 的数据项 18,用户 12 可更容易地标识与特定主题 20 相关联的数据项,例如,通过过滤电子邮件消息或 RSS 聚合项以便在该消息或项的主题或正文中寻找特定关键词。

[0030] 图 2 呈现了表征这些订阅技术中的某一些的第二示例性场景 30,这些订阅技术可促进用户在审阅与感兴趣的一个或多个主题 20 相关的数据项 18 方面的体验。在该第二示例性场景 30 中,用户 12 可利用订阅管理器 32 来请求对各种数据馈源 16 的一个或多个订阅 34,诸如对由第一数据源 14 供应的第一数据馈源 16 的第一订阅 34,以及对第二数据馈源 16 和第三数据馈源 16 (这两个数据馈源都由第二数据源 14 供应)的第二订阅 34 和第三订阅 34。数据源 14 可接受并且存储这些订阅 34,并且在接收到数据馈源 16 中对应于这些订阅 34 的数据项 18 之后,数据源 14 可向用户 12 递送数据项 18 (例如,通过将电子邮件消息发送到用户 12 可访问的电子邮件帐户)。在接收到这些数据项之后,用户 12 所操作的设备可向用户 12 呈现这些数据项 18,诸如通过在显示器 22 上呈递数据项 18 的聚集呈现 36。

在该示例性场景中,呈递了分层呈现,该分层呈现包括数据馈源 16 作为节点集来显示的树视图,其中每一节点包括作为子节点集的数据馈源 16 的数据项 18。用户 12 可由此以更方便的方式来浏览数据馈源 16,而无需为新数据项 18 请求数据源 14。另外,设备可向用户 12 呈现某些或所有数据馈源 16 中新数据项 18 的列表,由此降低向用户 12 冗余地呈现旧数据项 18 (与图 1 的第一示例性场景 10 相比较)。在该第二示例性场景 30 中,用户 12 还可能按一个或多个主题 20 来过滤整个数据项集 18,诸如通过请求呈现包括特定关键词的数据项 18。

[0031] 尽管图 2 的第二示例性场景 30 呈现了某些优点(尤其优于图 1 的第一示例性场景 30),但某些缺点可被标识出。作为第一示例(类似于图 1 的第一示例性场景 10),用户 12 仍旧受限用户 12 所知道的数据馈源 16 和数据源 14 的集合,且仍然不能查看涉及在用户 12 不知道的数据馈源 16 和数据源 14 上所张贴的、感兴趣的主题 20 的数据项 18。作为第二示例,这些技术可能仅可用于准许各种订阅形式的数据源 14,诸如供应 RSS 馈源的网站。许多数据源 14 可能不供应订阅服务;例如,网站管理员可拒绝提供 RSS 馈源以便迫使用户 12 访问网站来促进广告收入。尽管用户 12 可尽力实现针对不可订阅数据源 14 的类似订阅的服务(例如,通过轮询网站以寻找新内容),但数据源 14 可通过其可接受的使用策略来阻止这一访问,且可记入黑名单或以其他方式拒绝违反该策略的用户进行访问。作为第三示例,这些订阅技术涉及检索张贴到所有数据馈源 16 的所有数据项 18,且可能不能够请求将所检索的数据项 18 限制在与特定主题集 20 有关的那些数据项。该检索可能在计算上是昂贵的(例如,利用巨大的带宽、存储、和/或处理),且这可能对订阅和监视具有大量数据项 18 的许多数据馈源 16 是相当昂贵的(尤其在用户 12 仅对只被所接收的数据项 18 的一小部分所覆盖的小的主题集 20 感兴趣的情况下)。作为第四示例,标识与各种数据项 18 相关联的主题 20 以便限制向用户 12 呈现数据项 18 可能是困难的。诸如关键词检测之类的简单技术可能提供不准确的结果,尤其对于通过许多自然语言名称来讨论的主题 20。相反,复杂的分析技术可能是可用的,诸如对以外国语言呈现的文本数据项 18 的自然语言解析、语言翻译,对包括图像的数据项 18 的图像分析,以及上下文分析;然而,在这些技术的应用中所涉及的计算成本可能是相当昂贵的。例如,对大的数据项集 18 进行计算上密集的过滤以便向单个用户 12 呈现结果可能是相当昂贵的。

[0032] 鉴于这些示例性场景的潜在缺点,此处呈现了用于为用户集 12 生成主题数据馈源集的技术,其中从许多数据源 14 聚集的数据项 18 可经受复杂的主题分析以便标识与其相关联且被聚集成主题数据馈源集的主题 20。这些技术可涉及,例如,向潜在大的用户集供应潜在大的主题数据馈源集的集中式服务。这些技术的一个实施例可从数据馈源 16 和数据源 14 的广泛集合中接收数据项 18,且可对相应的数据项 18 执行计算上密集的分析以便标识与其相关联的主题 20。对于各种用户 12,服务可接收用户 12 感兴趣的主体集 20,且可生成包括与用户 12 标识出的主题 20 相关联的数据项 18 的主题数据馈源集。可响应于对主题 20 的说明来向用户 12 呈现这些主题数据馈源。

[0033] 图 3 呈现了根据此处讨论的技术、表征向用户 12 呈现主题数据馈源的示例性场景 40。在该示例性场景 40 中,主题数据馈源聚集服务 44 被配置成访问数据源集 14,其中每一数据源提供具有数据项集 18 的一个或多个数据馈源 16。主题数据馈源聚集服务 44 可监视这样的数据馈源 16 以寻求新数据项 18,且在接收到新数据项 18 之后可标识与每一数据项

18 相关联的一个或多个主题 20。主题数据馈源聚集服务 44 随后可从用户 12 接收主题集 42, 包括用户 12 感兴趣的主体集 20。对于用户 12 指定的每一主题 20, 主题数据馈源聚集服务 44 可选择与主题 20 相关联的数据项 18, 且可生成包括与主题 20 相关联的数据项 18 的主题数据馈源 48。例如, 可生成包括第一数据项 18 的第一主题数据馈源 48; 可生成包括与第三主题 20 相关联的第三、第四和第五数据项 18 的第二主题数据馈源 48; 以及可生成包括与第四主题 48 相关联的第四和第五数据项 18 的第三主题数据馈源 48。可向用户 12 递送这些主题数据馈源 48 包括主题数据馈源集 46, 以供呈现在例如用户 12 操作的设备的显示器 22 上。例如, 可显示树视图呈现 50, 其中每一主题 20 作为节点被包括在内, 该节点包括与主题 20 相关联的数据项 18 作为子节点。以此方式, 可响应于用户 12 感兴趣的主体 20 来生成主题数据馈源 48 并递送给用户 12。

[0034] 根据此呈现的技术(包括图 3 的示例性场景 40)来配置的服务可呈现某些优点和/或可减少相比于其他技术(包括图 1 和 2 的示例性场景)的缺点。以此方式配置的服务可通过代表潜在大的用户集 12 来聚集对数据项 18 的接收和分析, 从而实现比单个用户技术更高的规模经济。对这些技术进行足够广泛用户基础的缩放可使得呈递以下各项是可行的: 许多数据源 14 和数据馈源 16 的广泛调查, 对每一数据项 18 的主题分析进行的相对复杂的自动化(包括自然语言分析、外国语言中所呈现的文本数据项 18 的语言翻译、对基于图像的数据项 18 的图像分析、以及在有关数据项 18 之间的上下文分析)。这些技术中的某些实施例也为主题数据馈源 48 增加价值, 例如, 通过进一步按子主题来对主题数据馈源 48 中的数据项 18 分组; 通过按照预测值或可靠性来对数据项 18 排序; 通过删除旧的、冗余的、或不正确的数据项 18; 或通过将数据项 18 翻译成用户 12 的本机语言。另外, 如果服务被配置成为大量用户 12 提供服务, 则服务可具有足够优势来代理单个用户可能无法实现的带有特定数据源 12 的订阅机制; 例如, 服务可提出承担将数据项 18 从特定数据源 14 传播到大的用户集 12 的带宽花费, 且由此可迫使数据源 14 提供可能对单个用户 12 不可用的订阅机制(诸如, 在可接受的用户策略中对轮询限制的例外)。

[0035] 对于用户 12, 相比于代表单个用户所实现的技术(诸如图 1-2 的示例性场景), 使用这样的服务可呈现若干优点。可以向用户 12 呈现用户 12 不知道的或用户 12 无法定期监视的、来自数据源 14 和数据馈源 16 的数据项 18。这些技术还可减少用户以自组织方式主动轮询的数据源 14 的负担和低效性(例如, 必需访问许多网站以便保持知悉主题 20, 包括访问没有新的或感兴趣的数据项 18 的网站)。另外, 将监视大数据馈源 16 主体以及处理数据项 18 的计算负担从用户移除, 且用户 12 可仅接收用户 12 感兴趣的数据项 18。作为另一潜在的优点, 可向用户 12 呈现相比于可按其他方式实现的、对数据项 18 的更准确的主体聚集; 例如, 可由集中式服务实现的复杂主题分析可被认为比在用户设备上可能可行的较简单主题分析技术(诸如关键词检测)更准确。用户 12 还可受益于对数据馈源 16 的附加处理, 诸如通过将数据项 18 自动翻译成用户 12 的母语, 并按预测值来对数据项 18 排序, 这些附加处理可由服务来执行而对用户 12 没有成本。

[0036] 图 4 呈现了这些技术的第一示例性实施例, 它被示为被配置成向用户 12 呈现主题数据馈源 48 的示例性系统 66。示例性系统 66 在具有处理器 64 的计算机 62 上操作, 且可被实现为被配置执行此处讨论的技术的、例如处理器 64 上所执行的软件指令、诸如电路之类的硬件组件、和/或诸如现场可编程门阵列(FPGA)的混合设备。计算机 62 还可访问一

个或多个数据源 14, 其中每一数据源 14 提供包括至少一个数据项 18 的至少一个数据馈源 16。计算机 62 也可由这样的用户 12 访问: 可能对各种主题 20 感兴趣的, 以及可能向计算机 62 提交用户 12 希望监视的主题集 20 的用户。根据此处讨论的技术, 示例性系统 66 包括被配置成针对相应的数据馈源 16 从数据馈源 16 的数据源 14 接收数据项 18 的数据项接收组件 68。示例性系统 66 还包括被配置成针对相应的数据项 18 标识与数据项 18 相关联的至少一个主题 20 的主题标识组件 80。示例性系统 66 还包括主题数据馈源呈现组件 72, 该主题数据馈源呈现组件被配置成从用户 12 接收至少一个主题 20、并且针对相应的主题 20 选择与主题 20 相关联的数据项 18、生成包括与主题 20 相关联的数据项 18 的主题数据馈源 48、以及向用户 12 呈现主题数据馈源 48。生成并向用户 12 呈现包括与用户 12 所指定的主题 20 有关的数据项 18 的主题数据馈源集 48 之后, 示例性系统 66 由此可服务于用户 12 对于所指定的主题 20 的兴趣。

[0037] 图 5 呈现了这些技术的第二示例性实施例, 它被示为向用户 12 呈现包括至少两个数据馈源 16 的数据项 18 的至少一个主题数据馈源 48 的示例性方法 80, 其中相应的数据馈源 16 由至少一个数据源 14 来提供。可以对具有处理器 64 的设备 (诸如图 4 中示出的计算机 62) 执行示例性方法 80。该示例性方法 80 始于 82, 并且涉及在被配置成执行这里所提出的技术的处理器 64 指令上执行。具体而言, 指令可被配置成: 对于相应的数据馈源 86、从数据馈源 16 的数据源 14 接收数据项 18, 以及对于相应的数据项 18、标识与数据项 18 相关联的至少一个主题 20。指令还可被配置成: 从用户 12 接收 92 至少一个主题 20, 对于从用户 94 所接收的相应主题、选择与主题 20 相关联的数据项 18, 生成 98 包括与主题 20 相关联的数据项 18 的主题数据馈源 48, 以及向用户 12 呈现 100 主题数据馈源 48。基于从用户 12 所接收的主题 20 生成了主题数据馈源 48 并向用户 12 呈现主题数据馈源 48 之后, 示例性方法 80 由此针对用户 12 对所指定的主题 20 的兴趣提供服务, 因此在 102 结束。

[0038] 又一实施例涉及包括被配置成应用此处所呈现的技术的处理器可执行指令的计算机可读介质。可以这些方式设计的一种示例性计算机可读介质在图 6 中示出, 其中实现 110 包括其上编码有计算机可读数据 114 的计算机可读介质 112 (例如, CD-R、DVD-R、或硬盘驱动器的盘)。该计算机可读数据 114 进而包括被配置成根据此次阐述的原理来操作的一组计算机指令 116。在一个这样的实施例中, 处理器可执行指令 116 可被配置成执行一种向用户呈现主题数据馈源的方法, 诸如, 图 5 的示例性方法 80。在另一个这样的实施例中, 处理器可执行指令 116 可被配置成实现一种用于向用户呈现主题数据馈源的系统, 诸如, 图 4 的示例性系统 66。本领域普通技术人员可设计被配置成根据此处所呈现的技术来操作的许多此类计算机可读介质。

[0039] 此处所讨论的技术可被设计成在许多方面具有变型, 并且一些变型可呈现出相对于这些以及其他技术的其他变型的附加优点和 / 或减少缺点。此外, 一些变型可组合实现, 并且一些组合可表征通过相互协作而得到的附加优点和 / 或减少的缺点。各种变型可被合并到各个实施例中 (例如, 图 4 的示例性系统 66 和图 5 的示例性方法 80) 中, 以向此类实施例赋予单独和 / 或相互促进的优点。

[0040] 可在这些技术的各实施例之间变化的第一方面涉及其中可利用这些技术的情形。作为第一示例, 用户 12 (主题数据馈源 48 被呈现给该用户 12) 可包括个体、一组个体 (共同地动作和 / 或独立地动作)、诸如公司或组织之类的实体、设备、或自动化过程。作为第二示

例,与数据项 18 相关联的主题 20 可包括例如,个人、位置、对象、事件、或概念。此外,一个或多个主题 20 可按特定方式来呈现(例如,按文件、资源、或网站),和 / 或可按许多方式来标识(例如,按相关或区别名称、区别指示符、散列值、数据库键值、全局唯一标识符(GUID)、或统一资源标识符(URI)),诸如可通过代表性状态转移(REST)服务来访问的资源。作为第三示例,数据馈源 16 可包括许多类型的数据项 18,包括文本;诸如图像、声音、或视频之类的媒体;诸如文件或数据集之类的数字对象;数据库中的一项记录或记录集;或这些项的组合。作为第四示例,数据馈源 16 可包括许多类型的数据项 18 的集合,诸如人类可读和 / 或机器可解析列表、数据库或数据库的一部分、事件日志或文件、诸如可扩展标记语言(XML)文档之类的结构化文档、或诸如真正简单聚合(RSS)数据馈源之类的订阅格式。作为第五示例,数据馈源 16 可由许多类型的数据源 14 来提供,诸如文件服务器、web 服务器、数据库服务器、或具有输出的设备。可对其应用这些技术的数据源 14、数据馈源 16、以及数据项 18 的某些示例包括:供应包括新闻项集的新闻网站的新闻 web 服务器;供应包括论坛帖子集的 web 论坛的论坛 web 服务器;供应包括网络日志实体集的网络日志的网络日志服务器;供应包括数据馈源项集的数据馈源(诸如 RSS 馈源)的数据服务器;供应包括文件集的文件存储的文件服务器;以及供应包括数据记录集的数据库的数据库服务器。本领域普通技术人员可认识到其中可利用此处所讨论的技术的许多情形。

[0041] 可在这些技术的各实施例之间变化的第二方面涉及这些实施例的架构。图 4 示出一个实施例的示例性架构,其中各技术被实现为系统 66,系统 66 在计算机 62 的处理器 64 上运行,且包括数据项接收组件 68、主题标识组件 70 以及主题数据馈源呈现组件 72。然而,本领域技术人员可为这些技术的实施例选择许多架构,且某些架构可呈现相对于其他架构的优点和 / 或减少相对于其他架构的缺点。作为该第二方面的第一示例,各技术(诸如图 5 的示例性方法 80 中示出的逻辑)可被实现为包括可在通用计算机的通用处理器 64 上执行的一组指令的软件。指令可针对特定计算机 62 或特定处理器 64 部分地或全部地定制(例如,指令可利用特定计算机 62 的架构特征或图形处理器 64 的专用指令)。这样的指令可被存储在易失性存储器(例如,系统 RAM)中,或者存储在部分或全部非易失性存储器(例如,硬盘上所存储的高速缓存、只读光盘或可重写闪存)中。另选地或另外地,各技术可部分或全部以硬件实现,诸如被配置成执行图 5 的示例性方法 80 的电路、或被配置成实现图 4 的示例性系统 66 的现场可编程门阵列(FPGA)。

[0042] 作为该第二方面的第二示例,各技术可在许多类型的计算机 62 中实现,诸如服务器、工作站、如笔记本之类的便携式计算机、如移动电话之类的设备、或在单独设备中操作的电路。计算机 62 还可包括在另一设备上执行的虚拟机,诸如在 Java 中的软件实现,Java 是在设备的计算环境内所表示的 Java 虚拟机内执行的。计算机 62 还可包括不止一个设备,诸如协作(诸如通过网络)以接收数据项 18 并呈现主题数据馈源 48 的一组计算机。例如,在第一协作实施例中,对数据项 18 的接收可由服务器来执行,该服务器可通过将主题数据馈源 48 传输给用户 12 所操作的设备来呈现主题数据馈源 48 以供在与该设备耦合的显示器上进行呈递。另选地,在第二协作实施例中,服务器可向客户机设备呈现与主题 20 相关联的数据项集 18,且客户机设备可从其中生成主题数据馈源 48 以供向用户 12 呈现。在第三协作实施例中,处理的各个部分可被分布;例如,在对等网络中一起操作的一组客户机设备可协作以接收与其他对等体共享的数据项 18,和 / 或将数据项 18 与主题 20 进行关联中

所涉及的处理可分布在一组客户机设备上。例如,对于将这些技术的带宽和处理成本分布在一大组设备上(其中的某些或全部可由用户 12 来操作),和 / 或对于通过广泛地并行处理来提高这些技术的可靠性和速度而言,协作的对等网络可能是有利的。

[0043] 作为该第二方面的第三示例,与这些技术的各实施例的架构有关的特定场景涉及向可部署的计算环境中所表示的一组设备的用户 12 呈现主题数据馈源 48。最近,作出了尝试来开发用于以一致的、可部署的且可扩展的方式来提供对一批设备之中的计算环境的访问的技术。这些技术可设法在所有这些设备之间提供数据对象的自动化同步,且在协同操作的设备之间提供共同的应用集合的部署,以及用于管理应用在这些设备之间的取得、安装、使用以及卸载的集中式服务。数据对象和应用的集合在各个设备之间不必相同;例如,工作站可包含数据集的完整副本以及大量高性能应用(例如,照片编辑软件和图形密集的游戏),而蜂窝电话设备(具有较小的数据存储)可仅存储数据对象的子集,并且可表征与非便携式工作站不相关的便携性应用(例如,基于 GPS 的地图绘制软件)。然而,许多应用程序和与其相关的数据对象可以在这些设备之间共享(例如,被配置成管理用户日历对象的日历应用程序),并且该计算环境可以适应于能够在这些设备之间分发和同步该应用程序和数据对象。因此,可以明白,可以用允许在一组设备之间部署计算环境的方式来有利地表示计算机系统。

[0044] 在一种这样的技术中,计算环境一包括应用程序集合、应用程序资源、以及由此使用的数据对象—是用以下方式表示的:它可被传递到各设备以供根据该设备的能力来呈现。The objects include the data objects of the computer system, such as the user files and data created by the user, as well as representations of the myriad devices comprising the computing environment of the user. 以此方式表示的计算环境可被传递到任何设备并以适于该设备的能力的方式来呈现。例如,工作站可以将该信息呈现为稳健且通用的计算环境,而公共工作站可以通过 web 浏览器呈现不同的计算环境体验(例如,作为可在用户的会话结束时丢弃的虚拟机),并且蜂窝电话可以提供具有对蜂窝电话相关信息(例如,联系人、日历、以及导航数据)的较快访问的较瘦界面。此外,对该信息集的更新(例如,偏好改变和对其中包含的数据文件的更新)可被应用于该信息集的规范源,并且因此被传播到该信息集被传递到的所有其他设备。

[0045] 图 7 示出一个这样的场景 120,其中计算环境可由计算环境主机 122 主存,计算环境主机 122 可以存储并管理对象分层结构 124。计算环境主机 122 还可以代表各种设备(如蜂窝电话设备 126、个人笔记本计算机 130、以及公共工作站 134)并且还代表具有不同访问特权的不同类型的用户,以不同的方式呈现对象分层结构 124。对计算环境的更新可以传播回计算环境主机 122,并且可以与其他设备自动进行同步。因此,可将该计算环境设计并呈现为云计算体系结构,包括表达为跨形成同一计算环境的各协作门户(具有设备专用属性)的网格的所有设备(“客户机”)的一致呈现的设备无关表示(“云”)。

[0046] 针对这一示例性场景,此处所讨论的技术可按许多方式来利用可部署的计算环境。作为第一示例,对数据项 18 的接收、对数据项 18 与主题 20 的关联的评估、和 / 或对主题数据馈源 48 的生成可以跨对象分层结构 124 中所表示的设备集来分布,这可由计算环境主机 122 来协调。作为第二示例,可按更一致的方式通过设备中的每一个来向用户 12 呈现主题数据馈源 48。例如,如果用户 12 首先在蜂窝电话设备 126 上查看主题数据馈源 48,且

稍后在笔记本计算机 130 上查看主题数据馈源 48, 则笔记本计算机 130 上的呈现可排除先前在蜂窝电话设备 126 上向用户 12 所呈现的数据项 18。作为第三示例, 在某些实施例中, 主题数据馈源 48 可包括以对象分层结构 124 来表示的一个或多个数据对象, 诸如与主题数据馈源 48 的主题 20 有关的电子邮件消息。在图 7 的示例性场景 120 的上下文中, 呈现在第一设备 (诸如移动电话 126) 上的主题数据馈源 48 可包括至少一个数据对象, 该至少一个数据对象以对象分层结构 124 来表示且存储在另一设备 (诸如笔记本计算机 130) 上、和 / 或通过计算环境主机 122 来存储。

[0047] 作为该第二方面的第四示例, 与这些技术的各实施例的架构有关的特定场景涉及一实施例与各种数据源 14 的交互用于标识和提取数据源 16 的数据项 18。尽管某些数据源 14 可支持或甚至可提升接收和处理数据项 18 以及将数据项 18 潜在地重新分布到潜在大的用户集 12, 但其他数据源 14 可能不主动地支持这样的处理。作为第一示例, 数据源 14 可按并非出于自动化解析而是出于向用户 12 呈现来设计的方式, 来向计算机 62 呈现数据馈源 16, 例如, 呈现为表征嵌入式数据项 18 的网页。嵌入数据项 18 的方式可能难以解析; 例如, 将 HTML 文档的元素映射到数据项集 18 是困难的; 实际上, 某些数据源 14 可主动地使 HTML 文档的元素模糊以便减小对数据项 18 的自动化提取。作为第二示例, 数据源 14 可按有助于自动化解析的方式来呈现数据项 18, 但关于数据项 18 的有用元数据 (例如, 可有助于标识与数据项 18 相关联的主题 20, 和 / 或可向用户 12 呈现数据馈源 16 内的数据项 18) 可能更难以标识, 或按特定于数据源 14 和 / 或数据馈源 16 的方式是可标识的。例如, 数据源 14 可包括与数据项 18 在一起的相关联主题 20 的指示, 但可以按特定于数据源 14 和 / 或数据馈源 16 的方式 (诸如散列标签符号) 来这样做。

[0048] 鉴于这些细节, 某些实施例可在架构上被配置成便于从各种数据源 14 接收数据项 18。例如, 一实施例可利用各种试探来标识数据项 18 (例如, 通过标识在 HTML 元素中可指示项目列表的模式), 和 / 或提取与每一数据项 18 有关的数据源专用元数据 (例如, 通过标识信息的常见格式 (诸如电话号码和地址)、以及将所提取的元数据与其关联的数据项 18 相关联)。另选地或另外地, 一实施例可被配置成将特定处理逻辑应用于从特定数据源 14 接收到的特定数据馈源 16 的相应数据项 18。作为第一示例, 一实施例可包括利用特定解析协议来从特定数据源 14 提取数据项 18 的数据源专用逻辑。例如, 一实施例可基于 web 服务器所呈递的 HTML 中的可预测模式, 来从 web 服务器数据源 14 中对数据项 18 进行“web 抓取”。作为第二示例, 一实施例可尽力标识数据源 14 的数据源类型, 且可应用数据源类型专用逻辑以便提取数据项 18 和 / 或元数据。例如, 该实施例可标识特定数据源 14 具有特定数据源类型, 诸如由特定类型的 web 论坛软件驱动的 web 论坛, 且由此可基于这一数据源类型的许多数据源 14 所呈递的 HTML 元素的可预测模式、来对由数据源 14 所呈递的网页进行“web 抓取”。作为第三示例, 一实施例可利用数据馈源专用逻辑来从特定数据馈源 16 提取数据项 18 和 / 或元数据, 和 / 或利用数据馈源类型专用逻辑来从具有特定数据馈源类型的任何数据馈源 16 提取数据项 18 (例如, 可以认识到, 包括照片流的数据馈源 16 可例行地对作者姓名以及在表示图像的数据项 18 附近的图像的创建日期进行编码, 且因此被标识成照片流的所有数据馈源 16 可服从这样的逻辑: 该逻辑尽力标识并提取作者姓名以及创建日期作为相应数据项 18 的相关元数据)。作为第四示例, 一实施例可利用数据项类型专用逻辑来标识数据项 18 (例如, 如果数据馈源 16 包括图像, 则逻辑可被用来检查数据馈源 16

的数据对象、以及用来提取与图像的文件格式匹配的任何数据对象作为数据项 18), 和 / 或提取相关元数据 (例如, 图像的文件格式可包括作者姓名的元数据, 且当一实施例将特定数据项 18 标识成图像时, 它可尽力基于该文件格式来提取作者姓名)。

[0049] 图 8 示出示例性场景 140, 其中实现这一类型的架构的实施例 (被示为示例性系统 142) 可以在基于每一数据源类型的基础上评估从各种数据源 14 接收到的数据项 18。示例性系统 142 包括一组数据源解析组件, 该组数据源解析组件相应地被配置成: 对于特定数据源类型, 解析具有该数据源类型的数据源 14 的数据馈源 16 的数据项 18。例如, 第一数据源解析组件 146 可被配置成解析由第一数据源类型的各数据源所提供的数据项 18, 各数据源包括提供第一数据馈源 16 的第一数据源 14; 以及第二数据源解析组件 148 可被配置成解析由第二数据源类型的各数据源所提供的数据源 18, 各数据源包括提供第二数据馈源 16 的第二数据源 14。示例性系统 142 还可包括数据源接口 144, 该数据源接口 144 可被配置成: 对于相应的数据源 14, 标识该数据源类型的数据源 14, 并且调用与该数据源类型的数据源 14 对应的数据源解析组件, 以便解析数据源 14 的数据项 18。

[0050] 基于这一架构, 图 8 中示出的示例性系统 142 可按以下方式从各种数据源 14 接收数据项 18。为了从第一数据源 14 接收数据项 18, 示例性系统 142 可调用数据源接口 144 来标识该数据源类型的第一数据源 14。这一数据源类型可由第一数据源解析组件 146 支持, 第一数据源解析组件 146 可由数据源接口 144 调用以便从第一数据源 14 接收数据项 18 并且根据该数据源类型的第一数据源 14 的格式化来对它们进行解析。第一数据源解析组件 146 可由此接收由第一数据源 14 所供应的第一数据馈源 16 的第一数据项 18 和第二数据项 18, 可根据第一数据源 14 的该数据源类型的格式化来评估它们, 并且可生成经评估的数据项集 150。类似地, 为了从第二数据源 14 接收数据项 18, 示例性系统 142 可调用数据源接口 144 以便标识该数据源类型的第二数据源 14, 数据源接口 144 进而可调用第二数据源解析组件 148 以接收由第二数据源 14 所供应的第二数据馈源 14 的数据项 18、并且根据第二数据源 14 的该数据源类型的格式化来解析数据项 18 以便产生经评估的数据项 150。随后可将经评估的数据项 150 提供给示例性系统 142 的主题数据馈源呈现组件 72, 主题数据馈源呈现组件 72 可从中生成主题数据馈源 48。这一示例性系统 142 可以用可插入的或基于驱动程序的架构来实现, 使得当发现新的数据源类型时或当数据源类型改变时, 数据源解析组件可被添加或更改以便从新的或经更改的数据源类型的数据源 14 接收数据项 19。本领域普通技术人员可设计其中可利用本文所讨论的技术的许多场景。

[0051] 可在这些技术的各实施例之间变化的第三方面涉及从对应数据源 14 接收数据馈源 16 的数据项 18 的方式。作为第三方面的第一示例, 可以对可访问数据馈源 16 的计算机 62 执行这些技术, 诸如为数据馈源 16 服务的 web 服务器或可访问包含数据馈源 16 的数据项 18 的数据库的计算机 62。作为该第三方面的第二示例, 计算机 62 可经由网络 (诸如局域网、蜂窝式网络或因特网) 来请求数据馈源 16。这些技术 (诸如数据项接收组件 68) 的一实施例可由此例如通过在周期性的基础上或应用户 12 的请求来轮询数据源 14 从而监视数据馈源 16。另选地或另外地, 该实施例可能能够请求数据源 14 来在接收到一个或多个数据项 18 之后通知该实施例, 诸如通过向数据源 14 订阅以接收数据馈源 16 的数据项 18。数据源 14 由此可在接收到一个或多个新数据项 18 之后主动地通知该实施例, 和 / 或可周期性地生成新接收的数据项 18 的摘要并将其发送给该实施例。

[0052] 作为该第三方面的第三示例,某些数据源 14 可能不只是无法支持从一个或多个数据馈源 16 自动化提取数据项 18 或阻止自动化提取(例如,通过使用验证码或 HTML 结构的混淆),而是可能主动地禁止自动化处理,诸如通过在表征数据馈源 16 的网站的可接受使用策略中指定禁止。该策略可基于数据源 14 和 / 或馈源 16 的管理员的各种兴趣,例如,限制人类拜访者对网站的访问以便促进广告收入,或减少可能低效的带宽和处理密集的自动化访问。由此,对于单个用户 12 或用户集 12,可能不允许按自动化方式来接收数据项 18,并且数据源 12 的管理员可尽力检测违反者对数据馈源 16 的访问并将其记入黑名单或以其他方式阻止违反者对数据馈源 16 的访问。然而,此处所呈现的各技术的某些实施例可能在被用于为大的用户集 12 服务时尤其有利,例如,通过代表用户的大主体来接收和评估一次每一数据项 18。这一可缩放性也促进了数据源 12 的管理员的兴趣;例如,这些技术可使得数据馈源 16 能够到达用户 12 的大得多的主体,而不会导致数据源 14 的附加带宽成本。由于这些技术可提供可缩放性优点,因此各种数据源 14 的管理员对于准许该实施例自动监视数据馈源可能感兴趣,由此促进该实施例对数据项 18 的接收。本领域的普通技术人员可设计接收数据项 18 的各种方式而同时实现本文所讨论的各技术。

[0053] 可在这些技术的各实施例之间变化的第四方面涉及对数据项 18 可与其相关联的主题 20 进行标识的方式(诸如可由图 4 的示例性系统 66 中的主题标识组件 70 来执行)。可以理解,各种各样的数据源 14 所供应的各种各样的数据馈源 16 可呈现各种各样的数据项 18,诸如写在许多不同的域中的文本(例如,新闻文章、技术演示、统计和事实概要、从网络日志和社交联网站点所检索的轶事和对话、图像和视频记录、以及各种类型的数据对象)。此外,这样的数据项 18 可按各种语言、样式(例如,技术的、学术的、或通俗的)以及上下文来创作。对与如此各种各样的数据项 18 相关联的主题 20 进行自动化标识可能是困难的,而尽管标识某些这样的主题 20 可涉及人类参与(结合“Mechanical Turk (机械土耳其人)技术”),但接收对大量数据项 18 的准确的人类注意力的成本可能不可行。

[0054] 因此,为了标识与相应数据项 18 相关联的主题 20,这些技术的各实施例可表征各种形式的自动化分析,包括大的且复杂的自动化技术集。某些这样的技术可涉及检查数据项 18 的元数据以标识相关联的主题 20。作为该第四方面的第一示例,可将同数据项 18 相关联的各种类型的元数据与常常同各种主题 20 相关联的主题元数据项作比较,以便标识与数据项 18 相关联的主题 20。例如,可将在用地理编码的位置处捕捉的照片图像与各种陆标的已知地理编码作比较,以便标识该照片可能的议题。作为该第四方面的第二示例,具有特定主题标识符的主题 20 以及数据项 18 可包括主题 20 的主题标识符来作为数据项 18 与主题 20 的关联的指示,特定主题标识符诸如个体的区别性标识符(例如,社交网络中所表示的个体的简档的数字标识符、或与描述特定事件的资源相关联的 URI)。此外,这些主题标识符中的某些可专用于特定数据源 14、特定数据馈源 16、和 / 或特定类型的数据项 18;例如,特定网站可采用指示性约定(诸如特定格式化的散列标签)以指示与数据项 18 相关联的主题 20。此处所讨论的架构变型(诸如图 8 的示例性场景 140 中所示)可由此检测这些约定以供在标识与数据项 18 相关联的主题 20 中使用。

[0055] 该第四方面的其他变型可利用基于内容的形式的分析。作为该第四方面的第三示例,可利用各种试探来在数据项 18 中标识主题 20,诸如使用已知主题 20 的词汇表来检测关键词、或检测可能与数据项 18 的主题 20 相对应的专有名称。作为该第四方面的第四示例,

可利用语义分析。例如,对于包括一表达的数据项 18 (诸如用户 12 创作的文档或评论),可应用自然语言解析器来解析该表达并且基于诸如语言学结构之类的概念来标识由该表达引用的一个或多个主题 20 ;以及可利用图像分析技术和生物测定来标识照片图像中存在的特定对象和个体。作为该第四方面的第五示例,可利用各种形式的上下文分析;例如,可假设从特定数据馈源 16 或数据源 14 检索的所有数据项 18 都与特定主题 20 有关;或者,当主题 20 被标识为与第一数据项 18 相关联时,与第一数据项 18 有关的第二数据项 18 推测起来可能与同一主题 20 相关联。作为该第四方面的第六示例,数据项 18 之间的上下文关系可用于标识与第一数据项 18 相关联的主题 20 也与上下文有关的第二数据项 18 相关联。例如,可将两个个体之间的对话表示为有关数据项集 18,该有关数据项集 18 包括响应于彼此的评论。由此可将数据项 18 内被引用的主题 20 标识成与其他数据项 18 所表示的响应性评论在上下文有关。可例如通过重复提及该主题来显式地遵循这一关系,或者可例如通过假设主题 20 的后续评论的上下文关系(例如在特定时间段内在引用主题的评论之后的评论)来推断这一关系。更复杂的实施例可包括这些技术中的若干技术,且可使用各种形式的人工智能技术来结合这样的分析,诸如模糊逻辑、贝叶斯分析、专家系统、以及人工神经网络。

[0056] 图 9 呈现了表征由利用这些技术中的若干技术的主题标识组件 70 来对主题进行自动化标识的示例性场景 160。在这一示例性场景 160 中,第一数据源 14 提供包括两个用户之间的对话的第一数据馈源 16(诸如 web 论坛),第一数据馈源 16 被表示成具有关系 162 的一组两个数据项 18,以及第二数据源 14 (名为 Mark Smith 的个体的 web 服务器)提供包括表示由该个体创作的评论的第三数据项 18 的第二数据馈源 16 (诸如该个体的网络日志)。可应用主题标识组件 70 来标识可与这些数据项 18 相关联主题集 20,且该标识中可利用各种技术。首先,主题标识组件 70 可访问第一数据项 18,且可标识评论中所使用的首字母缩写“PDC”。主题标识组件 70 可尽力通过参考主题词汇表 164 用更大的特殊性来标识该主题,这可包括将首字母缩写“PDC”映射成被称为“专业开发者会议”的事件;因此,主题标识组件 70 可将第一数据项 18 与这一事件的主题 20 相关联。主题标识组件 70 随后可检查第二数据项 18,并且通过标识第一数据项 18 与第二数据项 18 (第二数据项被表示成对第一数据项的响应性评论)之间的关系 162 以及它们之间的语义关系(其中第二数据项 18 中的“我也是!”评论是指第一数据项 18 所引用的事件),来将第二数据项 18 与同第一数据项 18 相关联的同一主题 20 进行关联。主题标识组件 70 还可标识嵌在引用第二数据源 18 的第二数据项 18 中的 URL,第二数据源 18 由被称为特定主题 20 的已知个体(例如被频繁引用的个体)来管理。基于这一主题标识符,主题标识组件 70 由此可将第二数据项 18 与同该个体相关联的主题 20 进行关联。

[0057] 转向图 9 的示例性场景 160 中所表示的第三数据项 18,主题标识组件 70 可首先评估第三数据项 18 的域作为主题标识符,可将第三数据项 18 与第二数据项 18 中所引用的同一个体的主题 20 相关联。主题标识组件 70 也可在第三数据项 18 的内容中标识“社交联网”关键词,并且可将第三数据项 18 与表示社交联网的主题 20 相关联。主题标识组件还可将第三数据项 18 的语义内容与第一和第二数据项 18 的语义内容进行匹配,并且可标识要在“专家开发者会议”上给出所引用的与社交联网有关的演示。因此,主题标识组件 70 可将第三数据项 18 与其中给出该演示的“专家开发者会议”主题 20 相关联,并且可将第二数

据项 18 与针对其来给出所引用的演示的“社交联网”主题 20 相关联。以此方式,主题标识组件 70 可使用涉及每一数据项 18 (包括其中的关系 162) 的元数据以及每一数据项 18 的内容两者的许多技术,以便将数据项 18 与主题 20 相关联。在实现此处所讨论的技术时,本领域普通技术人员可以设计将数据项 18 与主题 20 进行关联的许多方式。

[0058] 可在这些技术的各实施例之间变化的第五方面涉及从用户 12 接收主题 20 的方式。作为该第五实施例的第一示例,用户 12 可简单地指定用户 12 感兴趣的主题集 20,诸如通过提供关键词集或从可用主题 20 的列表中选择,其中这些技术的一实施例被配置成针对可用主题来提供主题数据馈源 48。另选地或另外地,可基于关于用户 12 的信息来推断一个或多个主题 20。作为该第五方面的第二示例,可监视用户 12 的通信来标识用户 12 有可能讨论的主题 20,诸如通过标识用户 12 所写的对应于各种主题 20 的关键词。作为该第五方面的第三示例,可标识用户 12 的至少一个用户描述符,诸如用户简档中存储的关于用户 12 的信息(例如,用户 12 的年龄、地理位置、职业、或所描述的兴趣),并且基于用户描述符,至少一个主题 20 可被推断成用户 12 可能对其感兴趣。作为该第五方面的第四示例,可监视用户 12 的各种用户动作,且可推断用户 12 可能感兴趣的至少一个主题 20。例如,这些技术的一实施例可监视用户 12 所访问的网站,可标识与这些网站相关联的各种主题 20,并且可根据用户 12 与网站的交互程度(诸如浏览网站花费的时间量、返回访问网站的频率、以及用户 12 与网站交互所耗费的努力,诸如针对特定主题 20 被动地浏览 web 论坛相比于主动参与 web 论坛)来测量用户 12 对这些主题 20 的兴趣。

[0059] 可在这些技术的各实施例之间变化的第六方面涉及基于与主题 20 相关联的数据项 18 来生成主题数据馈源 48 的方式。作为该第六方面的第一示例,一实施例可选择与特定主题 20 相关联的数据项 18,并将数据项 18 聚集成主题 20 的主题数据馈源 48。该聚集可按许多方式来执行,例如,通过将数据项 18 作为无序集来呈现、或以许多方式有序地呈现(例如,按时间顺序,通过测量诸如与主题 20 的关系程度等显著性,通过测量诸如每一数据项 18 的数据源 14 的可信度或权威度等优先级,或这些因素的组合)。作为该第六方面的第二示例,可按许多方式将数据项 18 集成为主题数据馈源 48,例如,通过包括每一数据项 18 的完整内容;通过包括每一数据项 18 的概要或摘录,诸如文本数据项 18 的前几个单词或语句或图像数据项 18 的缩略图版本;通过包括每一数据项 18 的描述,诸如文件名、对象类型、大小、和源;或通过包括对数据项 18 的引用,诸如藉此可访问数据项 18 的 URI。

[0060] 该第六方面的某些变型涉及各种过滤器,这些过滤器可由用户 12 指定以供应用于数据项 18 同时生成主题数据馈源 48。用户 12 可指定各种准则来调整将与主题 20 相关联的数据项 18 聚集成主题数据馈源 48。这些过滤器可指定各种准则,诸如可能出现在数据项 18 中的关键词;与数据项 18 相关联的各种元数据项的各个值或值的范围,诸如数据项 18 的公开日期的日期范围;或一组可信的或不可信的身份,包括数据源 14 或可作为各种数据项 18 的作者或议题的个体的身份。此外,对于与过滤器的准则相匹配的数据项 18,该过滤器可指定:可将数据项 18 包括在主题数据馈源 48 中或排除在外;可按不同方式在主题数据馈源 48 中表示数据项 18(例如,主题数据馈源 48 可能包括更感兴趣的数据项 18 的完整版本,较不感兴趣的数据项 18 的摘录或描述);可在主题数据馈源 48 中对数据项 18 加以强调或不再强调,诸如通过增加或减小数据项 18 的缩略图表示的大小以反映优先级或显著性;可按特定方式将数据项 18 放置在主题数据馈源 48 内,诸如在其他数据项 18 之前或

之后 ;等等。另外,这样的过滤器集可按许多方式来组合,例如,使用布尔逻辑或模糊逻辑的分层结构,诸如贝叶斯网络或专家系统,或组合成打分或加权系统以标识期望阈值(数据项 18 超过该阈值则可能是用户 12 感兴趣的)或主题数据项 48 中数据项 18 之间的期望排序。在一个这样的实施例中,过滤器的准则可指定可归因于与准则相匹配的数据项 28 的分数;且当主题数据源 48 被生成时,可根据过滤器的准则针对相应的数据项 18 计算聚集分数,且主题数据源 48 的数据项 18 可根据聚集分数来排序。可以涉及许多类型的准则,且具有各种效果和相互关系的许多类型的过滤器可将这样的准则应用于数据项 18 同时生成主题数据源 48。

[0061] 该第六方面的附加变型涉及可以应用各种类型的服务同时从与主题 20 相关联的数据项 18 生成主题数据源 48。作为该第六方面的第三示例,一实施例可尽力减少将冗余数据项 18 包括在一个或多个主题数据源 48 中。在一个这样的场景中,可将与同一主题 20 相关联的两个数据项标识成等同的(诸如描述同一事件或表征实质相似的内容的两篇新闻文章),并且可从主题数据源 48 中移除一个这样的数据项 18。在另一这样的场景中,可将包括在第一主题数据源 48 中的数据项 18 标识成已经包括在第二主题数据源 48 中、或标识成已经包括在第一主题数据源 48 的先前呈现中,并且由此可将数据项 18 从第一主题数据源 48 中移除以减少在不止一个场合中向用户呈现同一数据项 18。作为该第六方面的第四示例,可变换和 / 或补充数据项 18 以供包括在主题数据源 48 中。例如,可强调对数据项 18 的内容中主题 20 的引用;可将数据项 18 从原始语言翻译成本机语言;或可将数据项 18 从各种表示转换成在主题数据源 48 的数据项 18 之间更一致的标准化表示(例如,包括图像的数据项集 18 可按诸如 GIF、JPG 和 PNG 等各种图像格式来接收,且可被转换成诸如 TIFF 等标准化图像格式,以供包括在主题数据源 48 中)。作为该第六方面的第五示例,可对数据项 18 进一步分类;例如,可将特定主题 20 (诸如演员)分成各子主题(诸如该演员在不同的表演或作品中担任的角色),并且可在主题数据源 48 内按子主题或相关主题 20 的交叉引用来对数据项 18 进行子分类。可以设计许多类型的服务,并且将其应用于主题数据源 48 的数据项 18。

[0062] 图 10 呈现了表征从先前已接收且与各种主题 20 相关联的数据项集 18 中生成主题数据源集 48 的示例性图示。在该示例性场景中的第一时间点 170,所接收的数据项 18 可由聚集组件 172 处理,聚集组件 172 可生成包括与第三主题 20 相关联的数据项 18 的第三主题数据源 48 以及包括与第四主题 20 相关联的数据项 18 的第四主题数据源 48。尽管可将这些主题数据源 48 呈现给用户 12,但可通过应用过滤器和服务集来将附加值添加到主题数据源 48 中。因此,在这一示例性场景中的第二时间点 174,可将过滤器集 176 应用于主题数据源 48 以根据各个过滤器 178 的准则和效果来调整包括每一数据源 48 的数据项集 18,诸如通过根据用户 12 指定的各种准则来移除、重新排序、强调、或不再强调数据项 18。例如,可以应用选择过滤器来选择相比于其他数据项 18 更感兴趣的特定数据项 18;例如,在第三主题数据源 48 中,可将第三数据项 18 标识成对用户 12 具有较高的优先级或用户 12 可能对其感兴趣,且可以扩展第三数据项 18 以便强调,而可将第四数据项 18 和第五数据项 18 标识成具有较低的优先级并且可减少第四数据项 18 和第五数据项 18 以便不再强调。还可将时间过滤器和关键词过滤器应用于主题数据源 48 以便移除陈旧的或用户 12 较不感兴趣的数据项 18;例如,在第四数据源 48 中,若第四数据项 18 包括与第

四主题 20 分歧的关键词则可移除第四数据项 18 (例如,第四主题 20 可涉及“web 开发”,以便减少包括涉及蜘蛛网的发展的不相关昆虫学文章,可将“蜘蛛类动物”的关键词过滤器应用于主题数据馈源 48)。在该示例性场景中的第三时间点 180,可将服务集 182 应用于主题数据馈源 48 以根据各种服务 184 来进一步调整数据项 18。例如,可以应用去重复服务以减少重复数据项 18 的冗余呈现;例如,该服务可标识出:第三主题数据馈源 48 和第四主题数据馈源 48 中都包括第五数据项 18,并且因此可将第五数据项 18 从第三主题数据馈源 48 中移除。还可以应用转换服务以将数据项 18 转换成更一致或更兼容的格式;例如,可将第五数据项 18 转换成新的表示以供包括在第四主题数据馈源 48 中。通过将这些过滤器 178 和服务 184 应用于主题数据馈源 48 的数据项 18,图 10 的示例性场景中所呈现的这些技术的实施例可由此生成用户 12 可能感兴趣的改进的主题数据馈源集 48。

[0063] 该第六方面的附加变型涉及用用户专用数据项来补充主题数据馈源 48。可以标识许多类型的用户专用数据项,将许多类型的用户专用数据项与主题 20 相关联,以及将许多类型的用户专用数据项包括在主题数据馈源 48 中。作为该第六方面的第六示例,用户专用的数据项可包括访问受控的数据项,该访问受控的数据项仅可由包括用户 12 的特定个体集访问(并且可能仅用户 12 可访问)。例如,用户专用数据项可包括用户 12 个人的文件、电子邮件消息、或数据库记录,或在基于订阅的服务中的一组文章(对其的访问限于一组特定的订阅者)。作为该第六方面的第七示例,该访问受控的数据项可包括以对象分层结构 124 (诸如图 7 的示例性场景 120 中)来表示的数据对象,且可将该访问受控的数据项存储在可通过网格来访问的另一设备上或存储在计算环境主机 122 上。在这些示例中,这些技术的一实施例可检查用户 12 可访问的各个数据对象,且可标识这些数据对象与之相关联的主题 20;以及在为主题 20 生成主题数据馈源 48 时,该实施例可包括这些数据对象中的一个或多个,作为与主题 20 相关联的用户专用数据项。例如,该实施例可通过利用用户 12 的凭证来代表用户 12 获取用户专用的数据项(诸如使用用户 12 的凭证来登录访问受控的订阅服务);或在第一设备上运行的该实施例可联系第二设备或可部署的计算环境的计算环境主机 122,以便获取以对象分层结构 124 表示的主题相关的数据对象。

[0064] 图 11 呈现了在生成主题数据馈源 48 时表征包括用户专用的数据项的示例性场景 190。这些技术的一实施例可从数据源 14 接收数据馈源 16 的数据项集 18,数据源 14 诸如表征包括 web 帖子集的 web 论坛的网站。可评估数据项 18 以供与特定主题 20 相关联(诸如“专家开发者会议”),并且可从与主题 20 相关联的数据项 18 中生成主题数据馈源 48。另外,用户 12 可访问特定的用户专用数据项,诸如文件系统 192 中的一组文件以及电子邮件帐户 196 中的一组电子邮件消息。可将来自文件系统 192 的文件 194 标识成与主题 20 相关联,并且可将来自电子邮件帐户 196 的电子邮件消息 198 标识成与主题 20 相关联。因此,可将文件 194 和电子邮件消息 198 连同从数据源 14 检索到的数据项 18 一起包括在主题数据馈源 48 中,以便向用户 12 呈现与同一主题 20 有关的定制且综合的数据项集。在实现这里讨论的技术时,本领域的普通技术人员可以设计生成主题数据馈源许多方式。

[0065] 可在这些技术的各实施例之间变化的第七方面涉及向用户 12 呈现主题数据馈源 48 的方式。作为该第七方面的第一示例,呈现 100 可涉及在显示器上呈递包括一个或多个主题数据馈源 48 的数据项 18,或者可涉及向用户 12 所操作的设备递送主题数据馈源 48。作为该第七方面的第二示例,可响应于许多事件来执行呈现 100。在第一个这样的场景中,

用户 12 可指定感兴趣的主体集 20, 并且一实施例可迅速生成或检索与所指定的主体 20 相关联的主体数据源 48, 且可向用户 12 呈现主体数据源 48。在第二个这样的场景中, 在从用户 12 接收到至少一个主体 20 之后, 主体数据源呈现组件 72 可被配置成将主体 20 存储在数据存储中, 该数据存储被配置成存储与各个主体 20 相关联的数据项 18。稍后, 在从用户 12 接收到呈现主体数据源 48 (基于先前指定的主体 20) 的请求之后, 主体数据源呈现组件 48 可从数据存储检索这些主体 20, 可为这些主体 20 生成主体数据源 48, 且可向用户 12 呈现主体数据源 48。此外, 包括数据项 18 和 / 或主体数据源 48 的数据存储可准许对这样的主体数据源 48 进行存档, 且可准许用户 12 请求对旧主体 48 的存档主体数据源 48, 诸如对若干年前的事件的主体叙述。在第三个这样的场景中, 一实施例可存储主体 20, 并可周期性地向用户 12 提供一系列刷新的主体数据源集 48, 主体数据源集 48 包括 (或限于) 自前次呈现以来已经接收的、与主体 20 相关联的新接收的数据项 18。在第四个这样的场景中, 在呈现 100 第一主体数据源集 48 之后, 一实施例可在接收到与主体数据源 48 的主体 20 有关的新数据项 18 后, 向用户 12 发送新数据项 18, 诸如通过电子邮件通知新接收的数据项 18、或通过刷新当前呈现的主体数据源 48 以包括最近的数据项 18。

[0066] 作为该第七方面的第三示例, 可按各种方式来分配计算资源, 以代表 (可能是大的) 用户集 12 来生成主体数据源 48。在第一个这样的场景中, 可代表所有用户接收数据项 18 并将其与主体 20 相关联, 但对主体数据源 48 的生成可以在每个用户和每个请求的基础上执行; 例如, 当用户 12 指定感兴趣的主体 20 时, 一实施例可基于与其相关联的数据项 18 来生成并呈现主体数据源 48。在第二个这样的场景中, 可针对各个主体 20 生成主体数据源集 48, 且当用户指定或选择主体 20 时, 主体 20 的主体数据源 48 可被检索且呈现给用户 12。例如, 该实施例可维护流行主体 20 的大批主体数据源 48, 以便防止为每一用户 12 和 / 或用户请求重复地生成同一主体数据源 48, 由此提供效率增益 (尤其在许多用户频繁地请求主体 20 的情况下, 和 / 或在对主体数据源 48 的生成是计算密集的情况下 (诸如涉及对数据项 18 的复杂语言翻译))。然而, 该第二场景可能不包括为特定用户 12 对主体数据源 48 进行个性化的某些方面, 诸如图 11 的示例性场景 190 中示出的。因此, 某些场景可涉及用用户专用数据项来补充先前生成的主体数据源 48, 以便促进为相应的用户对主体数据源 48 进行个性化。

[0067] 图 12 呈现了表征代表各个用户 12 来个性化主体数据源集 202 的示例性场景 200。这些技术的一实施例可为三个主体集 20 生成相应的主体数据源 48, 三个主体集 20 可由各个用户 12 来请求。例如, 第一用户 12 可将第一主题和第二主题包括在感兴趣的主体集内, 并且先前为这两个主体 20 所生成的主体数据源 48 可被呈现。然而, 作为呈现的一部分, 一实施例可标识、选择与第一主题 20 有关的任何用户专用数据项 204 并将其包括在第一主体数据源 48 中, 以及可标识、选择与第二主题 20 有关的任何用户专用数据项 204 并且将其包括在第二主体数据源 48 中。相似地, 第二用户 12 可指定感兴趣的第一主题 20 和第三主题 20, 且一实施例可检索每一主体数据源 48, 用与同一主题 20 相关联的用户专用数据项 204 来补充每一主体数据源 48, 并且向第二用户 12 递送所补充的主体数据源 48。因此, 尽管第一主体数据源 48 曾经是从与第一主题 20 相关联的 (非个性化) 数据项 18 中生成的, 但对这一主体数据源 48 的呈现可针对相应用户 12 来个性化, 而用户专

用数据项 204 也与第一主题 20 相关联,由此实现改进的效率和个性化两者。

[0068] 作为该第七方面的第四示例,呈递主题数据馈源 48 可按许多方式来实现。可将主题数据馈源 48 的数据项 18 可视地呈现为列表、拼贴图、或示出主题 20 和子主题的分层结构的嵌套节点集。在第一个这样的场景中,可将数据项 18 组织成可向用户呈现的主题叙述;例如,可按时间顺序来对关于事件的评论进行排序并作为该事件的叙述性报道来呈现。在第二个这样的场景中,数据项 18 可根据以下各项来组织:优先级(例如,“兴趣高的项”、“兴趣中等的项”、以及“兴趣低的项”)、主题(例如,包括被组织成“子主题 1”和“子主题 2”的数据项 18 的“主题 1”)、可靠性(例如,“来自权威性的源的项”以及“来自其他源的项”)、或个性化(例如“个性化的项和来自熟人的项”以及“非个性化的项”)。在第三个这样的场景中,可将数据项 18 之间的相互关系可视地包括在该呈现中。例如,主题数据馈源 48 的第一数据项 18 可具有与同一主题数据馈源 48 或另一主题数据馈源 48 中的第二数据项的响应性关系(诸如对评论的回复),并且可在主题数据馈源 48 中的数据项 18 的表示中包括响应性指示符(诸如箭头或分层组织)。在第四个这样的场景中,可按能查询的格式来表示主题数据馈源 48,使得用户 12 可在与特定准则相匹配的主题数据馈源 48 内搜索数据项 18。

[0069] 图 13 呈现了表征对包括各个主题 20 的主题数据馈源 48 的数据项 18 的呈现 218 的示例性场景 210。这些技术的一实施例(诸如系统 216)可访问数据存储 212(诸如文件系统或数据库),且可将数据项 18 存储在与各个主题 20 相关联的数据存储 212 中。用户 12 可指定特别感兴趣的两个主题 20,且系统 216 可将这两个主题 20 存储在与用户 12 相关联的数据存储 212 中。稍后,用户 12 可向系统 216 提交呈现主题数据馈源集 48 的请求 214。系统 216 可通过以下操作来满足该请求:从数据存储 212 检索由用户 12 先前指定的两个主题 20,检索与相应主题 20 相关联的数据项 18,从其中生成两个主题数据馈源(有可能涉及应用诸如图 10 的示例性场景中的各种过滤器和/或服务),以及在显示器 22 上呈递主题数据馈源 48 的呈现 218。在该示例性场景 210 中,主题数据馈源 48 作为树视图来呈现,其中相应的主题 20 作为第一节点集来呈现,该第一节点集包括作为与每一主题 20 相关联的数据项 18 的子节点。此外,在两个数据项 18 具有响应性关系(诸如对评论的回复)的情况下,这一关系可作为响应性指示符来呈现,诸如将响应性数据项 18 呈现为数据项 18 的嵌套子节点作为该响应的目标。以此方式,系统 216 可通过向用户 12 呈递各种主题 20 的数据项 18 的表示,来满足用户 12 对呈现主题数据馈源 48 的请求,该表示带有相关性的视觉指示符。本领域的普通技术人员可以在实现在此所讨论的技术的同时设想许多方式来向用户 12 呈现主题数据馈源 48。

[0070] 尽管用结构特征和/或方法动作专用的语言描述了本主题,但可以理解,所附权利要求书中定义的主题不必限于上述具体特征或动作。更确切而言,上述具体特征和动作是作为实现权利要求的示例形式公开的。

[0071] 如在本申请中所使用的,术语“组件”、“模块”、“系统”、“接口”等一般旨在表示计算机相关的实体,该实体可以是硬件、硬件和软件的组合、软件、或者执行中的软件。例如,组件可以是,但不限于是,在处理器上运行的进程、处理器、对象、可执行码、执行的线程、程序和/或计算机。作为说明,运行在控制器上的应用程序和控制器都可以是组件。一个或多个组件可以驻留在进程和/或执行线程中,并且组件可以位于一个计算机内和/或分布在两个或更多计算机之间。

[0072] 此外,所要求保护的主体可使用产生控制计算机以实现所公开的主体的软件、固件、硬件、或其任意组合的标准编程和 / 或工程技术来实现为方法、装置、或制品。如此处所使用的术语“制品”旨在涵盖可从任何计算机可读设备、载体、或介质进行访问的计算机程序。当然,本领域技术人员应当认识到,在不背离所要求保护的主体范围或精神的情况下可对该配置作出许多修改。

[0073] 图 14 和以下讨论提供了对实现本文所阐述的一个或多个原理的实施例的合适计算环境的简要、概括描述。图 14 的操作环境只是合适的操作环境的一个示例,而不旨在对该操作环境的使用范围或功能提出任何限制。示例计算设备包括但不限于,个人计算机、服务器计算机、手持式或膝上型设备、移动设备(诸如移动电话、个人数字助理(PDA)、媒体播放器等)、多处理器系统、消费电子产品、小型计算机、大型计算机、包括任何以上系统或设备中的任一个的分布式计算环境等。

[0074] 虽然并非必需,但是实施例在由一个或多个计算设备执行的“计算机可读指令”的一般上下文中进行描述。计算机可读指令可经由计算机可读介质来分发(在下文中讨论)。计算机可读指令可被实现为执行特定任务或实现特定抽象数据类型的程序模块,诸如函数、对象、应用程序编程接口(API)、数据结构等。通常,计算机可读指令的功能可按需在各种环境中组合或分布。

[0075] 图 14 示出包括被配置成实现此处所提供的一个或多个实施例的计算设备 222 的系统 220 的示例。在一种配置中,计算设备 222 包括至少一个处理单元 226 和存储器 228。取决于计算设备的确切配置和类型,存储器 228 可以是易失性的(例如诸如 RAM)、非易失性的(例如诸如 ROM、闪存等)、或两者的一些组合。该配置在图 14 中由虚线 224 示出。

[0076] 在其他实施例中,设备 222 可包括附加特征和 / 或功能。例如,设备 222 还可包括附加存储(例如,可移动和 / 或不可移动),包括但不限于磁存储、光存储等。此类附加存储在图 14 中由存储 230 示出。在一个实施例中,实现此处所提供的一个或多个实施例的计算机可读指令可位于存储 230 中。存储 230 还可储存实现操作系统、应用程序等其他计算机可读指令。可将计算机可读指令加载到存储器 228 中,以供例如处理单元 226 执行。

[0077] 如此处所使用的术语“计算机可读介质”包括计算机存储介质。计算机存储介质包括以用于存储诸如计算机可读指令或其他数据之类的信息的任何方法或技术来实现的易失性和非易失性、可移动和不可移动介质。存储器 228 和存储 230 都是计算机存储介质的示例。计算机存储介质包括但不限于, RAM、ROM、EEPROM、闪存或其他存储器技术、CD-ROM、数字多功能盘(DVD)或其他光存储、磁带盒、磁带、磁盘存储或其他磁存储设备、或者可用于存储所需信息且可由设备 222 访问的任何其他介质。任何这样的计算机存储介质都可以是设备 222 的一部分。

[0078] 设备 222 还可包括允许该设备 222 与其他设备进行通信的通信连接 236。通信连接 236 可包括但不限于,调制解调器、网络接口卡(NIC)、集成网络接口、射频发射机 / 接收机、红外线端口、USB 连接、或用于将计算设备 222 连接到其他计算设备的其他接口。通信连接 236 可包括有线连接或无线连接。通信连接 236 可发送和 / 或接收通信媒体。

[0079] 术语“计算机可读介质”可包括通信介质。通信介质通常以诸如载波或其他传输机制之类的“已调制数据信号”来体现计算机可读指令或其他数据,并且包括任何信息传送介质。术语“已调制数据信号”可包括以对信号中的信息进行编码的方式设置或改变其一

个或多个特征的信号。

[0080] 设备 222 可包括输入设备 234, 诸如键盘、鼠标、笔、语音输入设备、触摸输入设备、红外照相机、视频输入设备、和 / 或任何其他输入设备。在设备 222 中还可包括输出设备 232, 诸如一个或多个显示器、扬声器、打印机、和 / 或任何其他输出设备。输入设备 234 和输出设备 232 可经由有线连接、无线连接、或其任何组合连接到设备 222。在一个实施例中, 来自另一计算设备的输入设备或输出设备可用作计算设备 222 的输入设备 234 或输出设备 232。

[0081] 计算设备 222 的组件可通过多种互连来连接, 例如总线。这样的互连可以包括诸如 PCI Express 之类的外围部件互连 (PCI)、通用串行总线 (USB)、火线 (IEEE 1394)、光学总线结构等等。在另一实施例中, 计算设备 222 的组件可通过网络来互连。例如, 存储器 228 可包括位于通过网络互连的不同物理位置的多个物理存储器单元。

[0082] 本领域技术人员应当认识到, 用于存储计算机可读指令的存储设备可分布在网络上。例如, 可经由网络 238 访问的计算设备 240 可存储实现此处所提供的一个或多个实施例的计算机可读指令。计算设备 222 可访问计算设备 240, 并且下载部分或全部计算机可读指令以供执行。替换地, 计算设备 222 可按需下载计算机可读指令的片断, 或者一些指令可在计算设备 222 处执行, 而一些指令则可在计算设备 240 处执行。

[0083] 本文提供了各实施例的各种操作。在一个实施例中, 所描述的操作中的一个或多个可以组成存储在一个或多个计算机可读介质上的计算机可读指令, 这些指令如果由计算设备执行则使得计算设备执行所描述的操作。描述一些或全部操作的顺序不被解释为暗示了这些操作必须要依赖于顺序。从本说明书获益的本领域技术人员将认识到替换顺序。此外, 应该理解, 并非所有的操作都一定存在于本文所提供的每一个实施例中。

[0084] 此外, 本文中所使用的词语“示例性”意指用作示例、实例或说明。在本文中描述为“示例性”的任何方面或设计并不一定被解释为比其他方面或设计有利。相反, 使用词语“示例性”旨在以具体的方式呈现各个概念。如本申请中所使用的, 术语“或”意指包括性“或”而非互斥性“或”。即, 除非另有指定或从上下文中清晰可见, 否则“X 使用 A 或 B”意指任何自然的包括性排列。即, 如果 X 使用 A ; X 使用 B ; 或 X 使用 A 和 B 两者, 则在任何以上情况下, 都满足“X 使用 A 或 B”。另外, 本申请中和所附权利要求书中所使用的冠词“一”和“一个”一般可被解释为意指“一个或多个”, 除非另有指定或从上下文中清晰可见是指单数形式。

[0085] 同样, 虽然参考一个或多个实现示出并描述了本公开, 但本领域技术人员基于对本说明书和附图阅读和理解, 可以想到各种等效更改和修改。本发明包括所有这样的修改和更改, 并且仅由所附权利要求的范围来限定。具体来说, 对于由上述组件 (例如, 元素、资源等等) 执行的各种功能, 除非另外指明, 否则用于描述这些组件的术语旨在对应于执行所描述的执行此处在本公开的示例性实现中所示的功能的组件的指定功能 (例如, 功能上等效) 的任何组件, 即使这些组件在结构上不等效于所公开的结构。此外, 尽管可相对于若干实现中的仅一个实现来公开本公开的一个特定特征, 但这一特征可以如对任何给定或特定应用所需且有利地与其他实现的一个或多个其他特征相组合。此外, 就在说明书或权利要求书中使用术语“包含”、“具有”、“含有”和“带有”及其变体而言, 此类术语旨在以与术语“包括”相似的方式为包含性的。

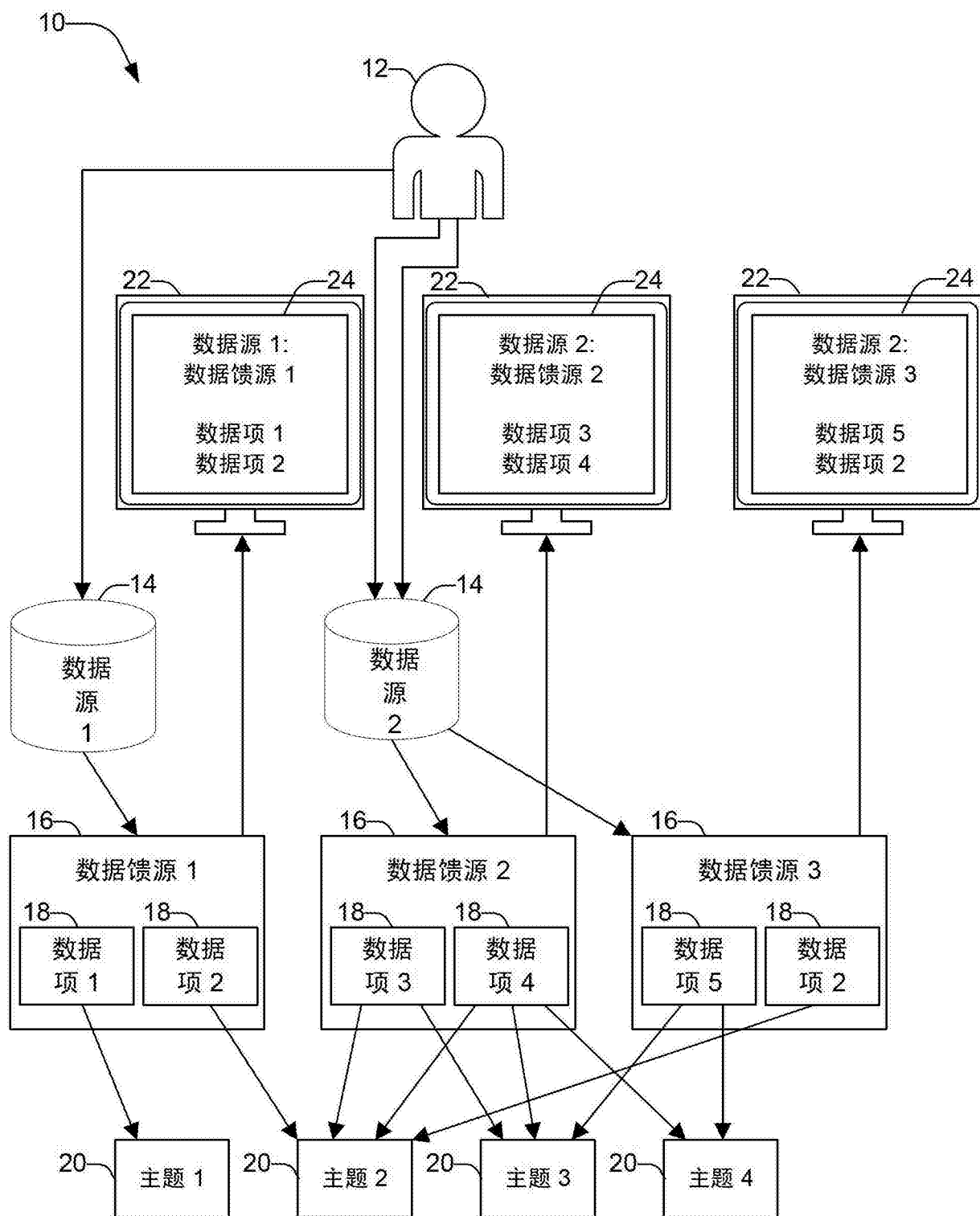


图 1

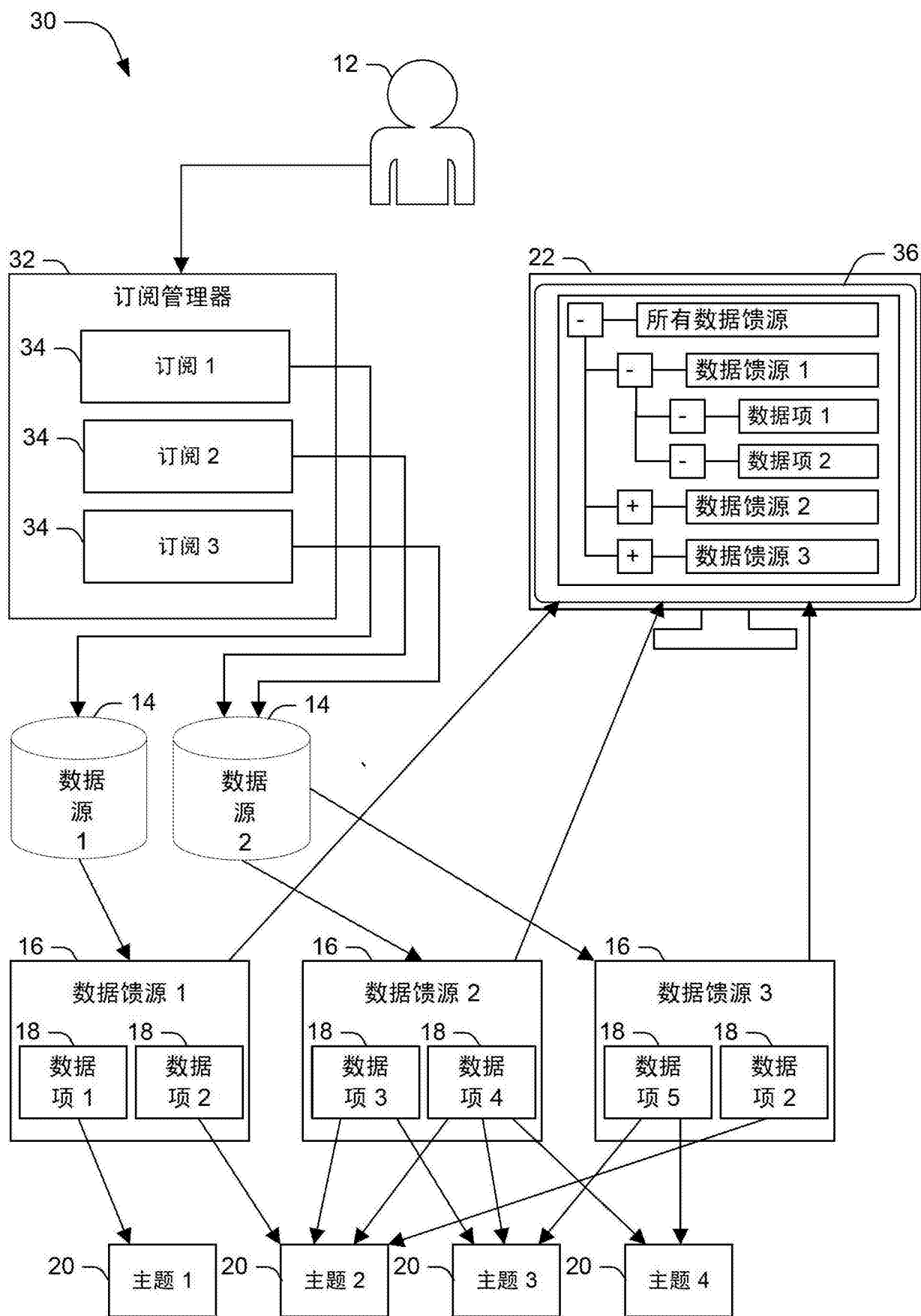


图 2

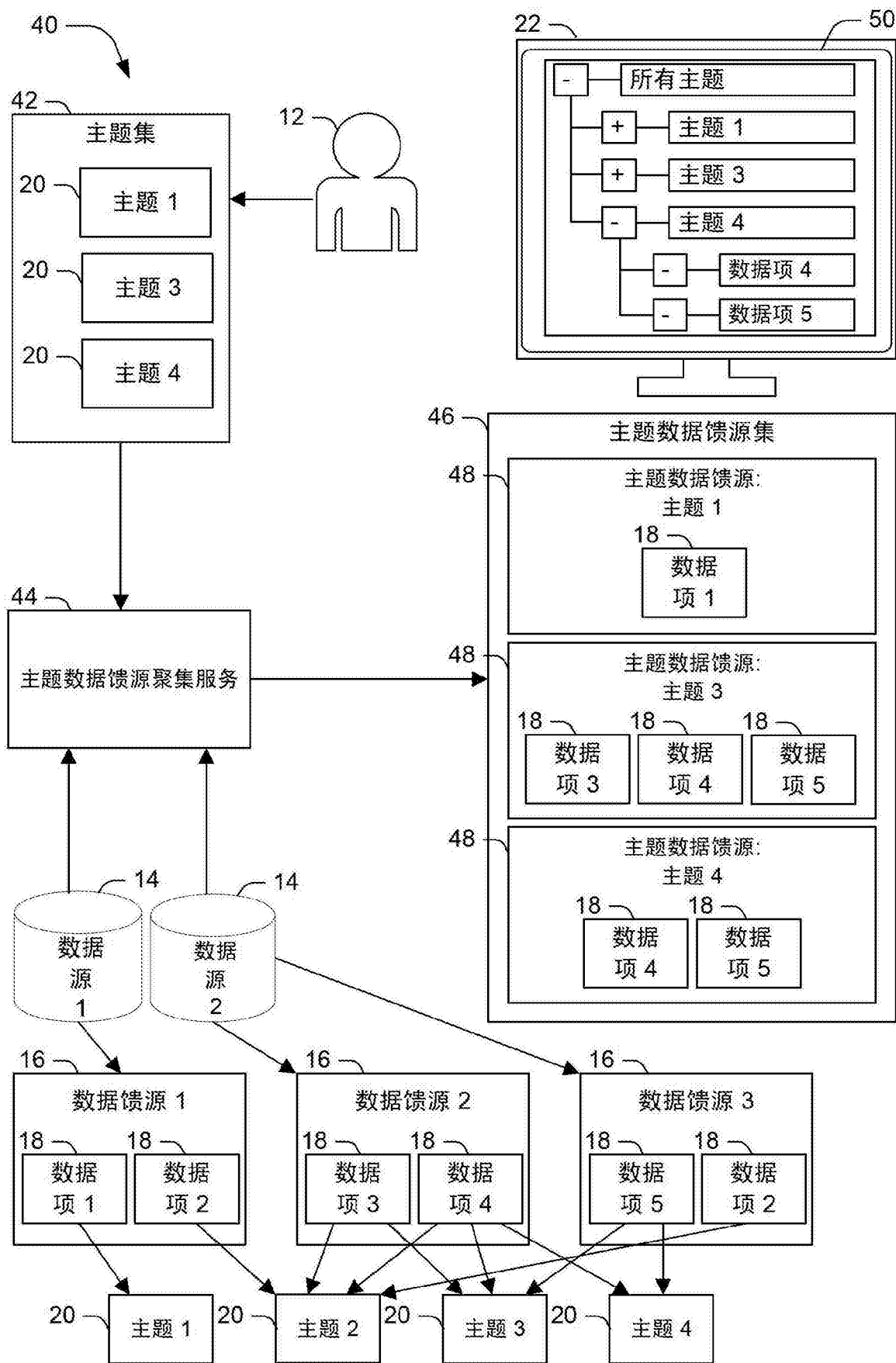


图 3

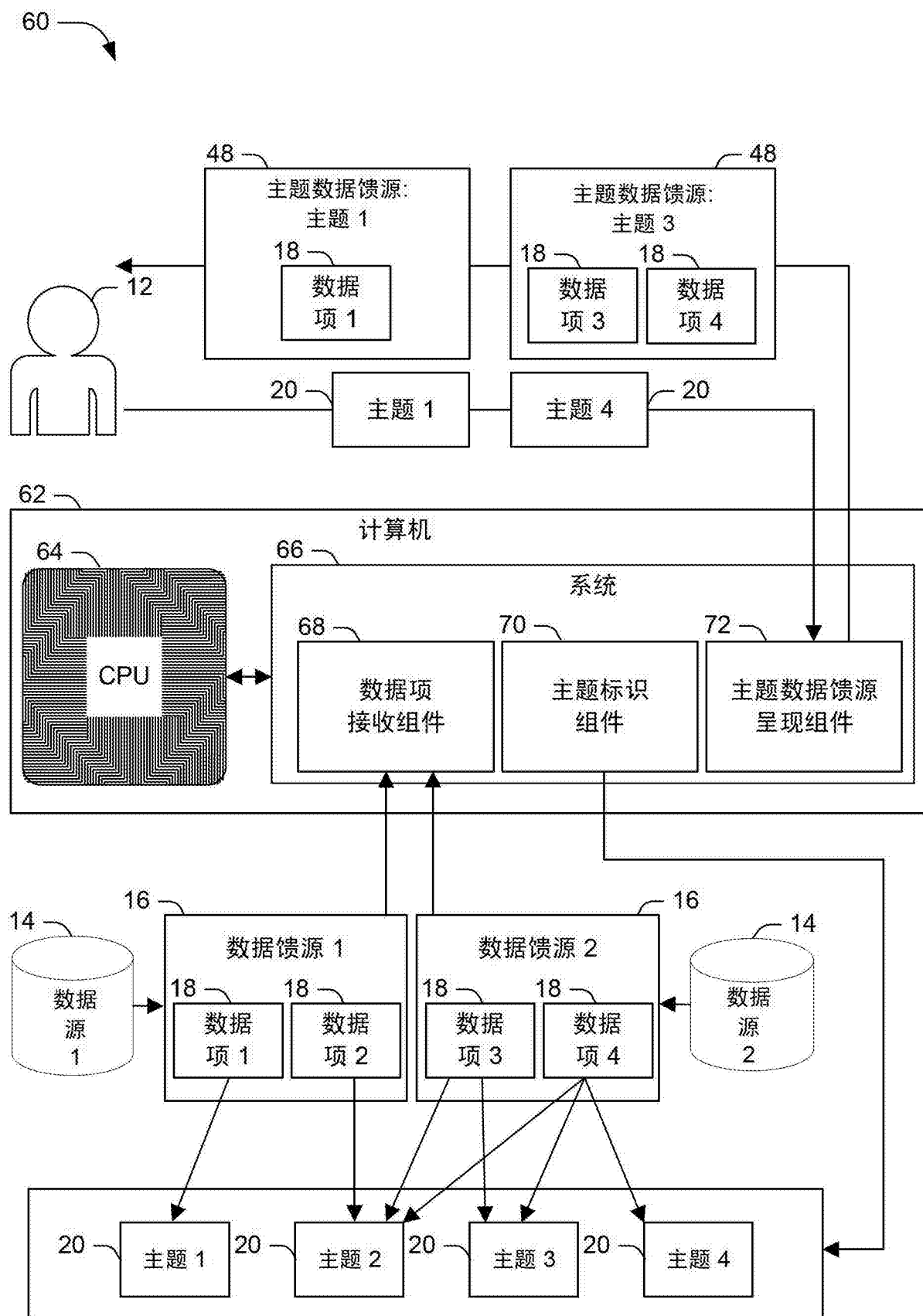


图 4

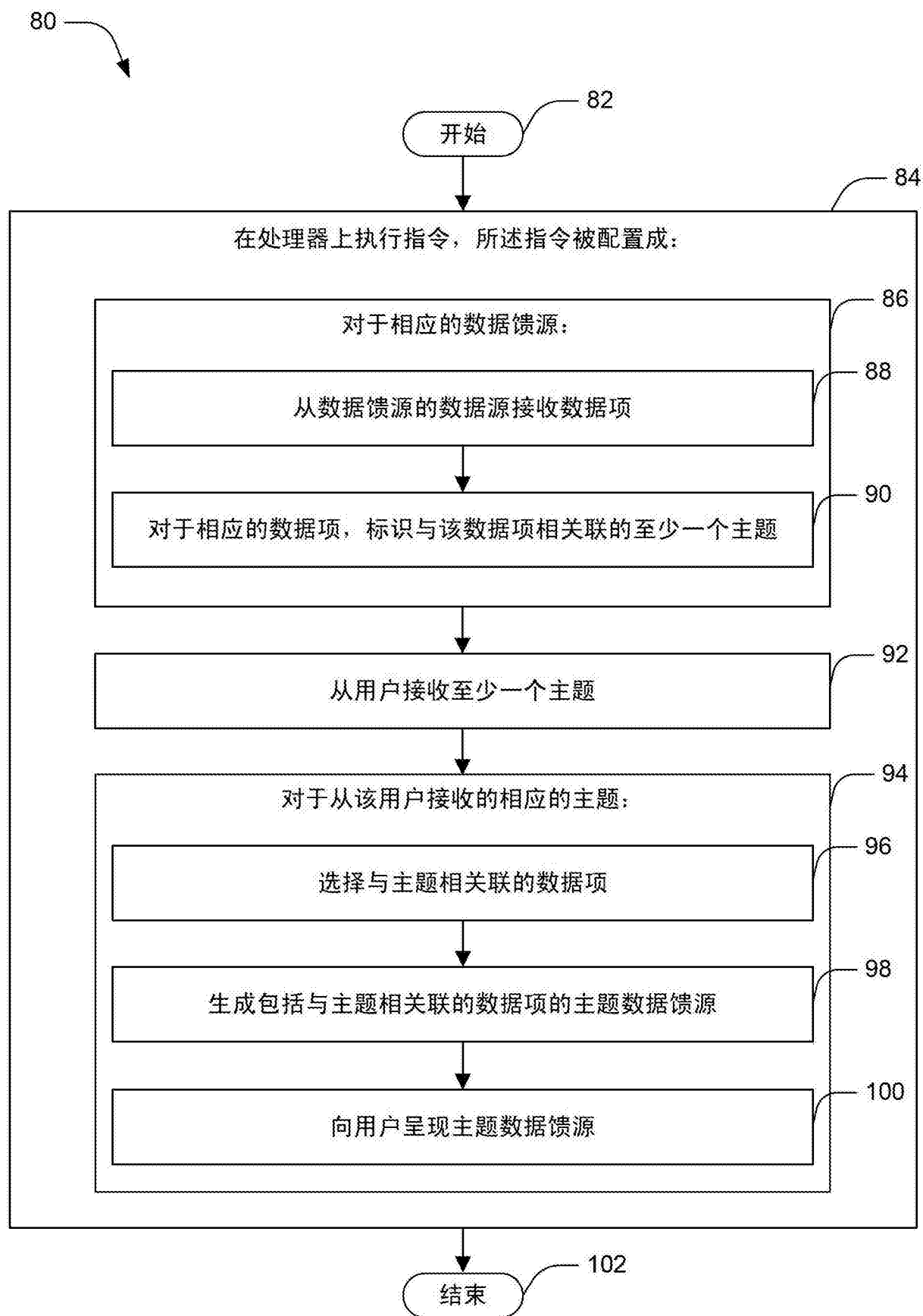


图 5

110

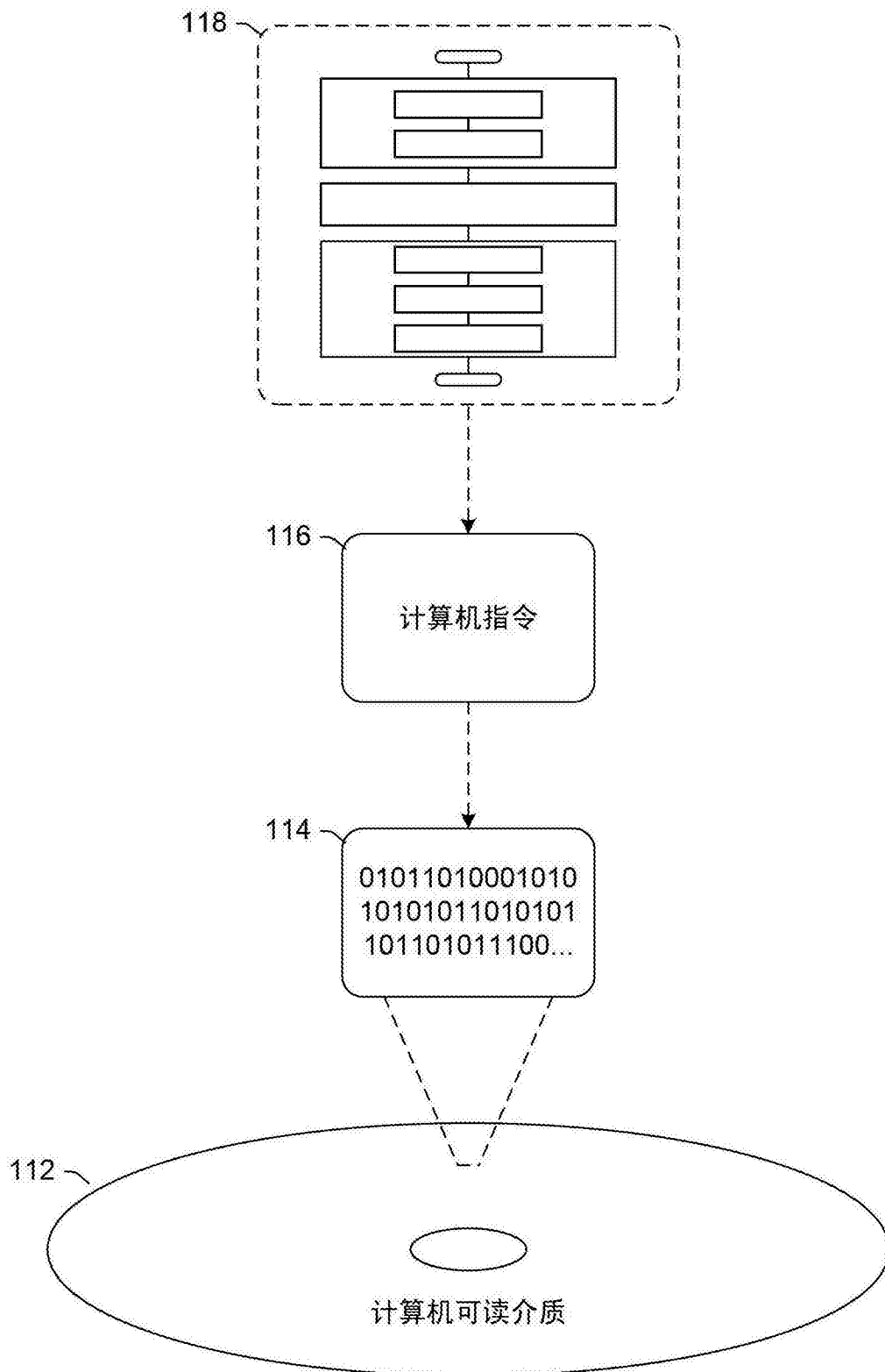


图 6

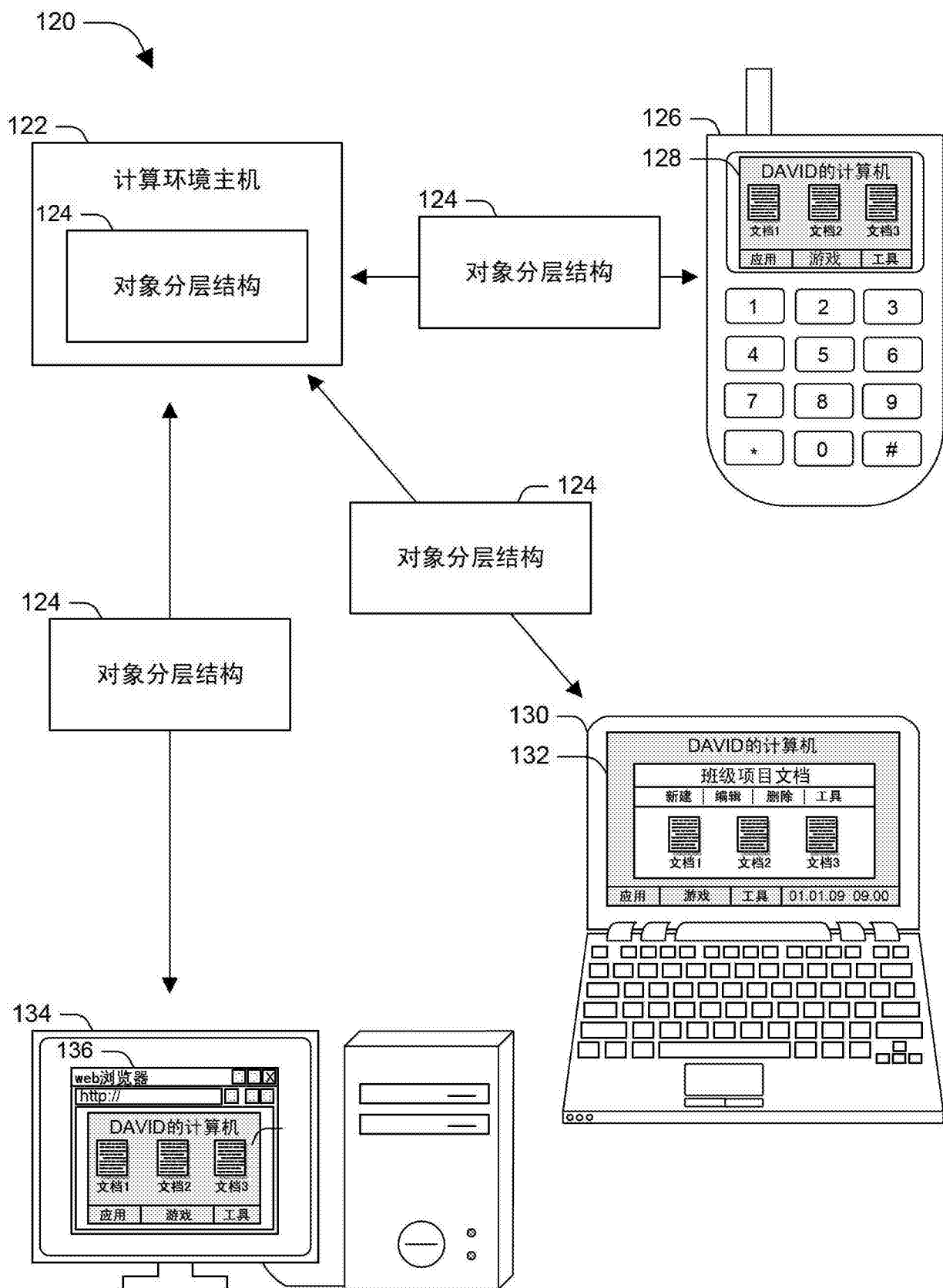


图 7

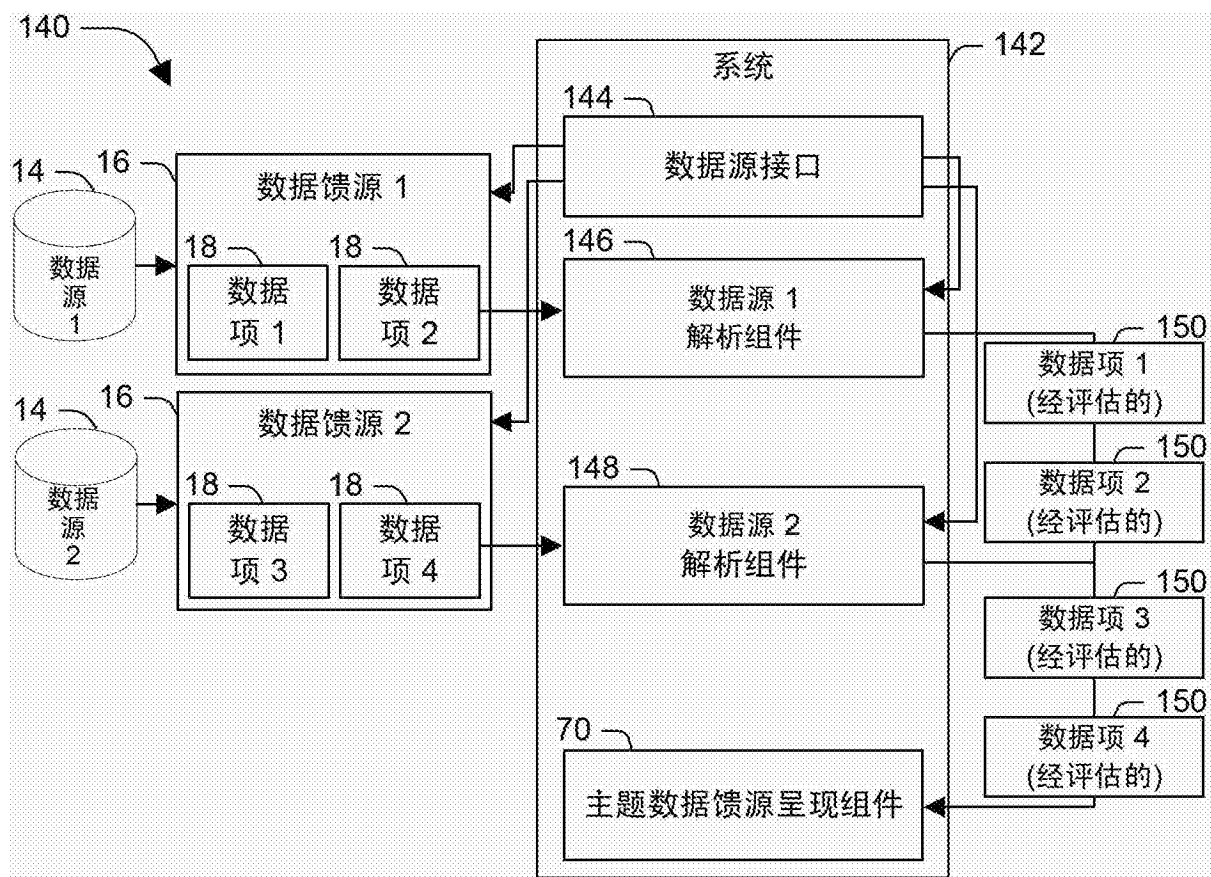


图 8

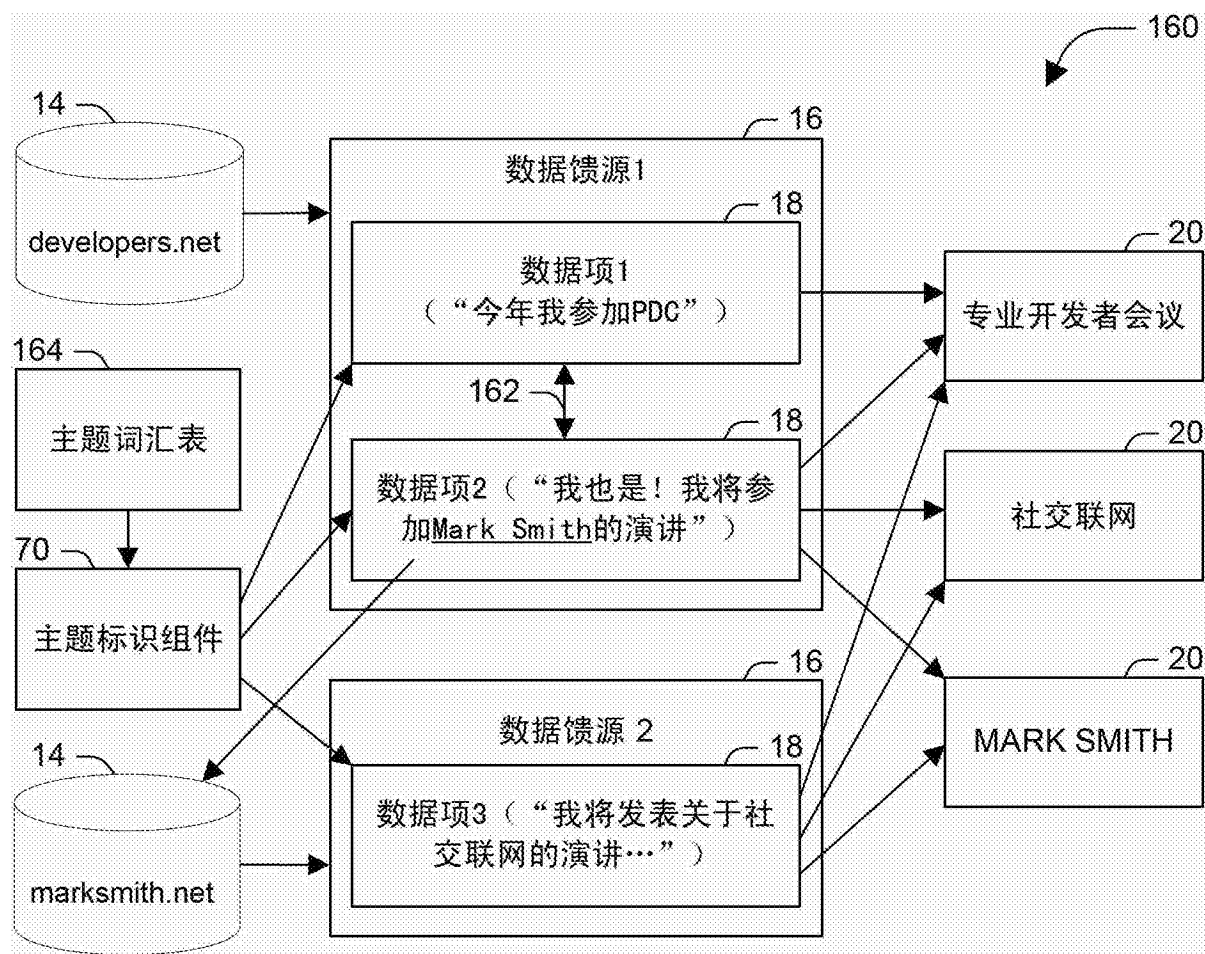


图9

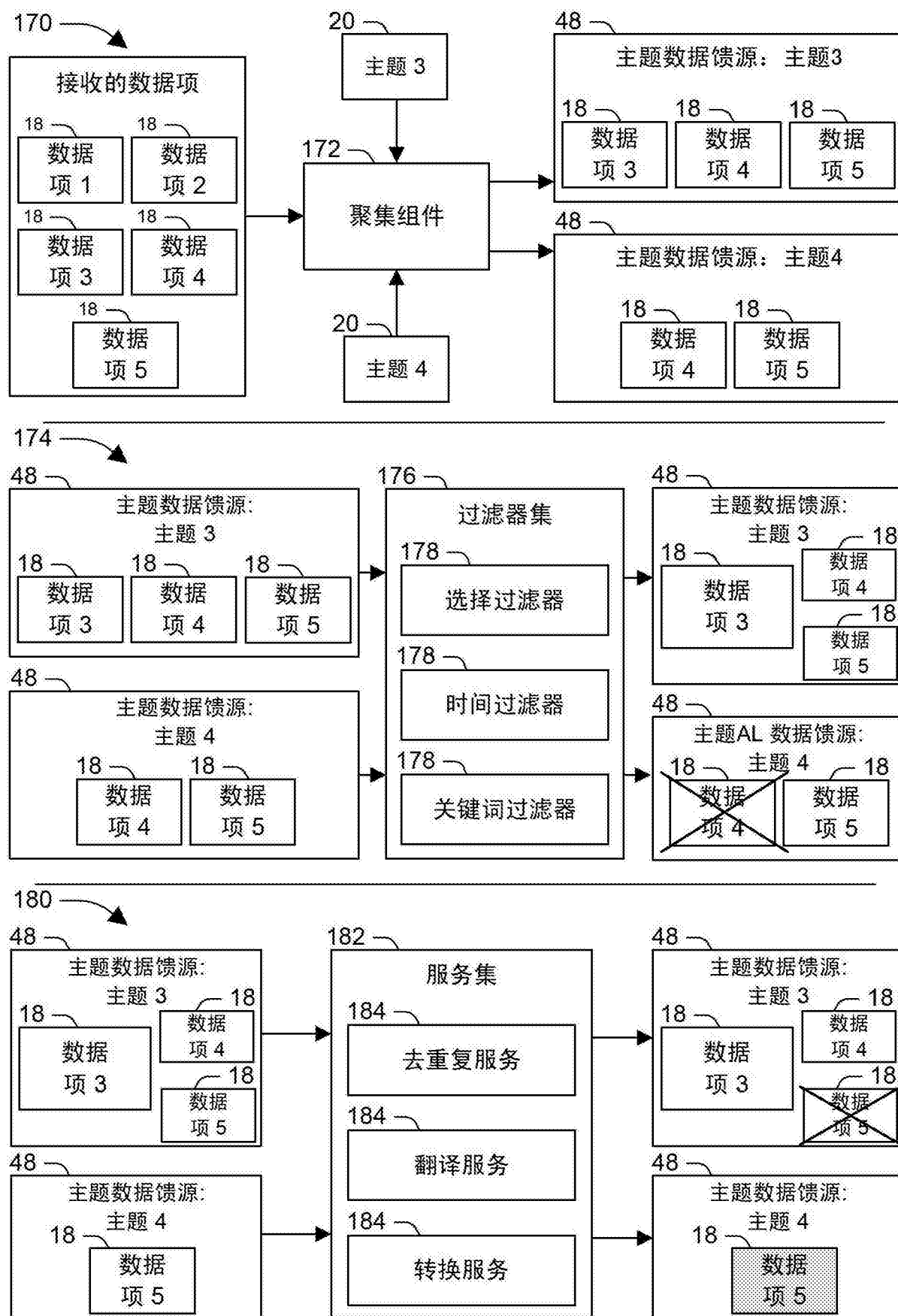


图 10

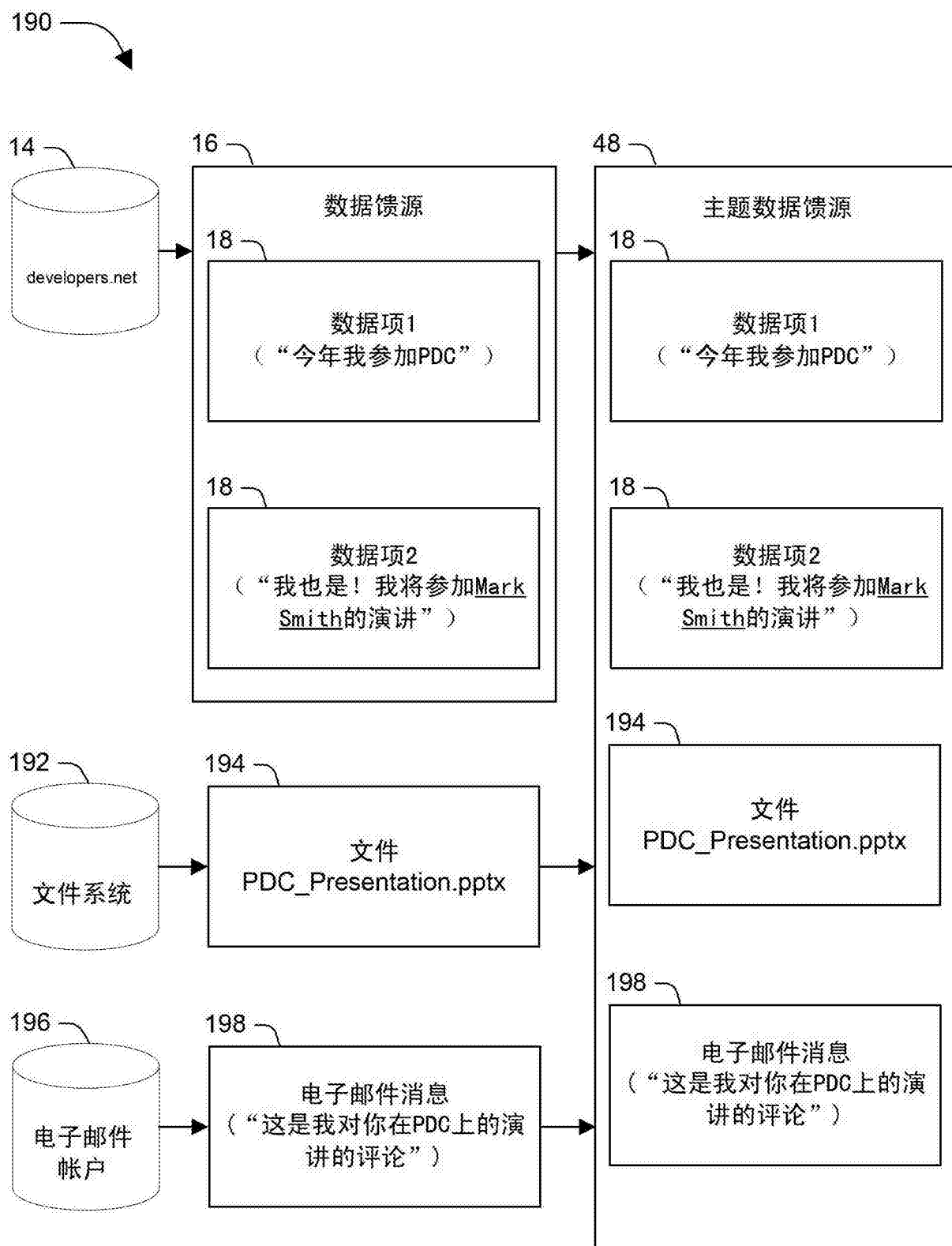


图 11

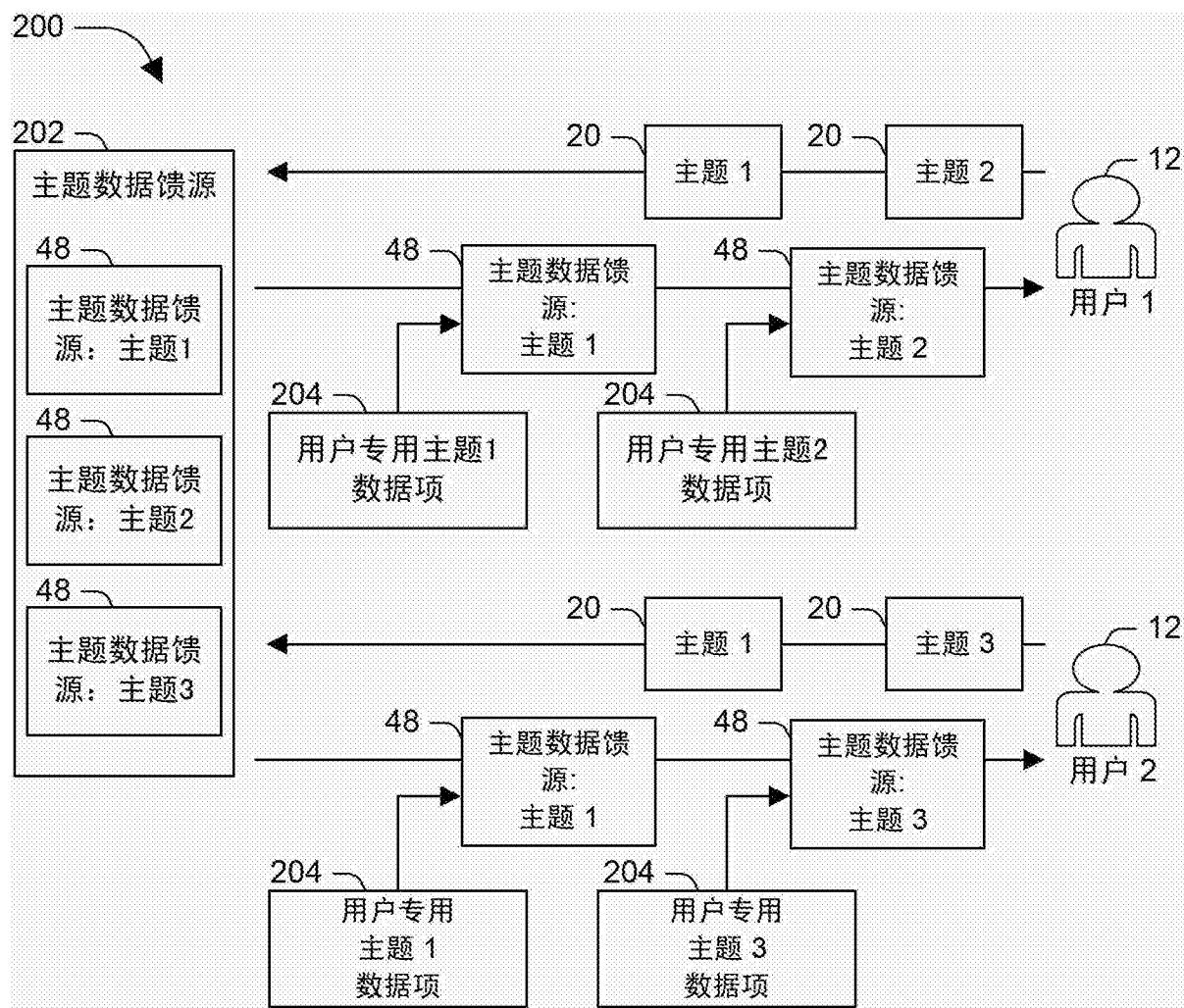


图 12

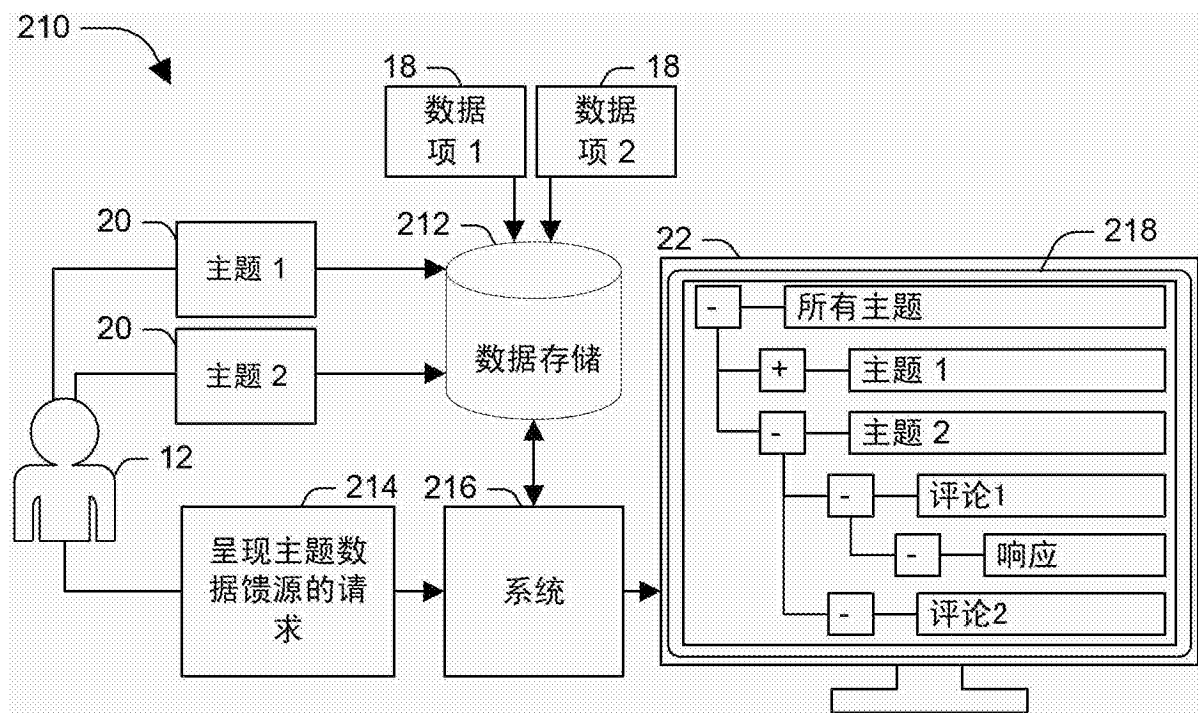


图 13

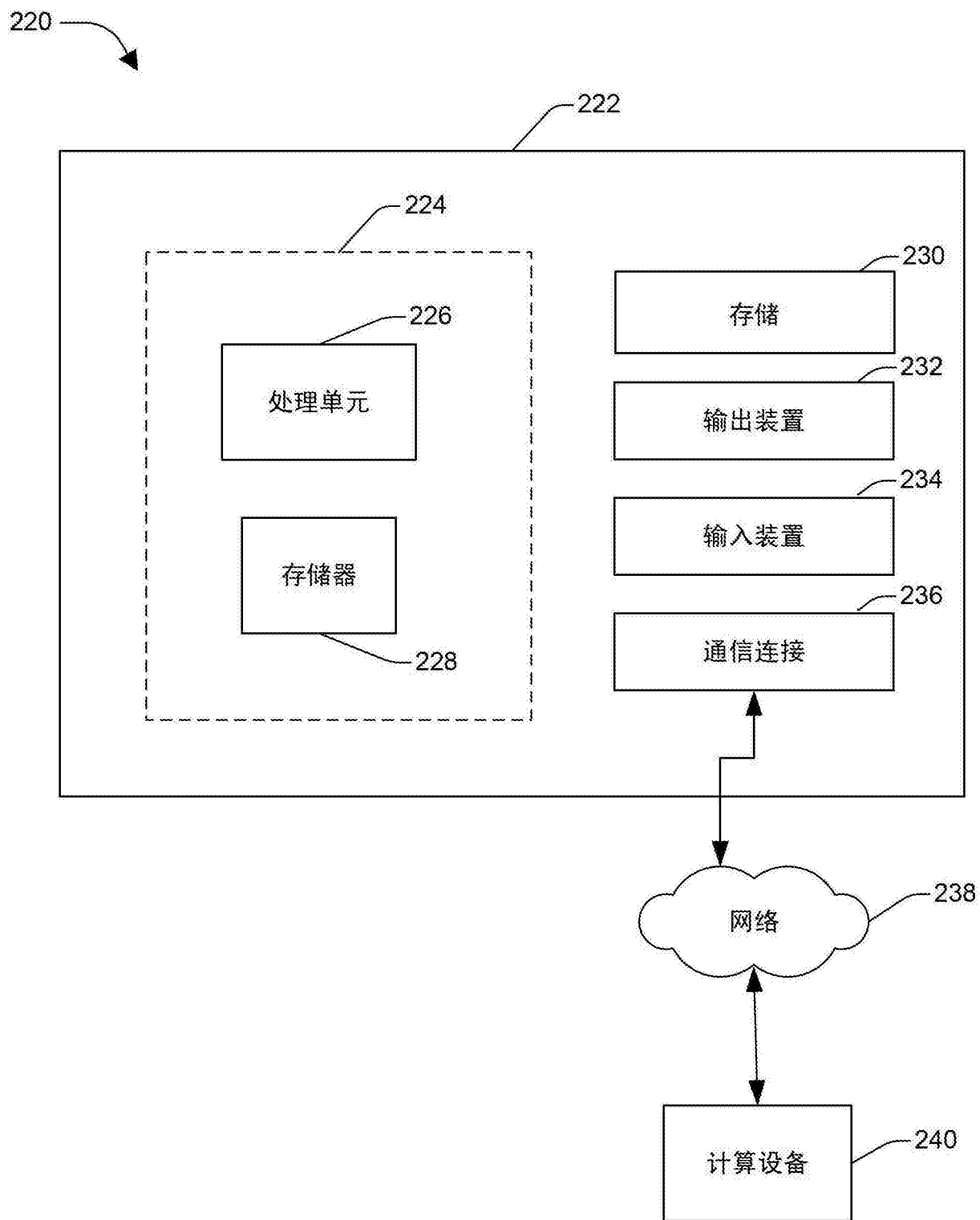


图 14