[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
G06F 17/30 (2006.01)

[12] 发明专利说明书

专利号 ZL 200510056263.5

[45] 授权公告日 2009 年 6 月 24 日

[11] 授权公告号 CN 100504861C

[22] 申请日 2005.3.29
[21] 申请号 200510056263.5
[30] 优先权
[32] 2004.3.29 [33] US [31] 10/811,653
[73] 专利权人 微软公司
地址 美国华盛顿州
[72] 发明人 C·马克加恩
[56] 参考文献
US2002097278A1 2002.7.25
审查员 姚天宇

[74] 专利代理机构 上海专利商标事务所有限公司
代理人 李 玲

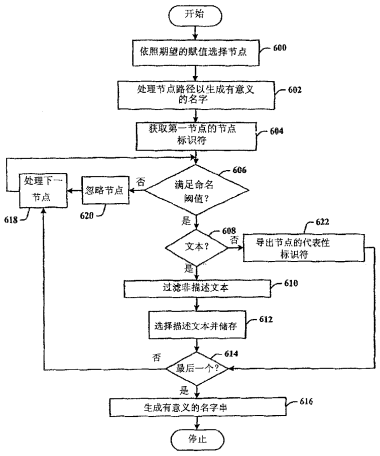
权利要求书 4 页 说明书 12 页 附图 8 页

[54] 发明名称

用于在展平的分层结构中有意义的名字的创建的方法和系统

[57] 摘要

一种用于为展平的数据结构生成有意义的名字的体系结构。该体系结构包括一种系统，它观察与分层数据结构相关联的用户活动，然后生成从文件和文件夹名或更容易由用户导航并理解的其它字符串数据导出的有意义的名字。该系统包括一赋值组件，它便于向数据结构的每一节点应用赋值度量。一选择组件选择一个或多个节点用于命名过程。一命名组件接收并处理来自选择组件的所选择的节点信息，并为给定用户生成一个或多个节点的名字。命名组件的输出是数据结构的有意义的名字格式的展平表示。



1．一种用于为展平的数据结构生成使用户能够理解和使用所述数据结构的描述的系统，其特征在于，包括：

一赋值组件，所述赋值组件依照一预定度量向被包括在分层数据结构中的多个数据节点中的一个或多个分配一赋值，其中所述度量是用于确定所述赋值的多个度量中的一个，其中，所述多个度量包括创建时间、访问的节点、访问节点的时间、修改的节点、何时修改节点、复制的节点、何时复制节点、访问频率、以及访问节点的单独用户的数量；以及

一描述组件，所述描述组件生成表示依照所述度量选择的所述一个或多个数据节点的至少一个的描述。

2．如权利要求 1 所述的系统，其特征在于，还包括一选择组件，所述选择组件根据期望的赋值数据选择要处理所述一个或多个数据节点中的哪一个用于察看。

3．如权利要求 1 所述的系统，其特征在于，所述度量与观察的用户活动相关联，其中所述用户活动是由所述用户针对被包括在所述分层数据结构中的数据节点做出的活动。

4．如权利要求 1 所述的系统，其特征在于，所选择的数据节点是由所述描述表示的，所述描述是一维视图。

5．如权利要求 1 所述的系统，其特征在于，所述数据结构与操作系统和数据储存库的至少一个相关联。

6．如权利要求 1 所述的系统，其特征在于，所选择的数据节点与播放列表相关联。

7．如权利要求 1 所述的系统，其特征在于，所述赋值被呈现为数字、图像数据、音频数据、着色和字符串的至少一个。

8．如权利要求 1 所述的系统，其特征在于，所述一个或多个数据节点是对第一计算系统本地的、远离所述第一计算系统定位的、以及跨多个计算系统分布的数据节点中的至少一个。

9．如权利要求 1 所述的系统，其特征在于，通过将所述描述超链接到对应的数据节点，所述描述能被用于导航到所述相关联的数据节点。

10．如权利要求 1 所述的系统，其特征在于，所选择的数据节点具有最高赋

值。

11．一种采用权利要求 1 所述的系统的计算机。

12．如权利要求 1 所述的系统，其特征在于，还包括一分类器，所述分类器通过至少基于与所述一个或多个数据节点相关联的度量作出推断，以实现用于描述的生成的所选择的操作的自动化。

13．一种用于为数据结构生成使用户能够理解和使用所述数据结构的名字的系统，其特征在于，包括：

一赋值组件，所述赋值组件依照至少基于观察到的活动的一个或多个预定度量向被包括在分层数据结构中多个数据节点中的一个或多个分配赋值，其中所述活动是针对被包括在所述分层数据结构中的数据节点做出的活动，其中所述度量是用于确定所述赋值的多个度量中的一个，其中，所述多个度量包括创建时间、访问的节点、访问节点的时间、修改的节点、何时修改节点、复制的节点、何时复制节点、访问频率、以及访问节点的单独用户的数量；

一选择组件，所述选择组件选择具有期望赋值的一个或多个数据节点中的至少一个；以及

一命名组件，所述命名组件生成表示依照所述一个或多个预定度量选择的数据节点的至少一个的一个或多个名字。

14．如权利要求 13 所述的系统，其特征在于，所述数据节点由以一维视图呈现的一个或多个名字来表示。

15．如权利要求 13 所述的系统，其特征在于，所述数据结构与操作系统和数据储存库的至少一个相关联。

16．如权利要求 13 所述的系统，其特征在于，所述赋值作为数字、图像数据、音频数据、着色和字符串的至少一个来呈现。

17．如权利要求 13 所述的系统，其特征在于，所选择的数据节点与一数据路径相关联，所述数据路径由多个数据节点定义，其每一个具有一个或多个节点指定，所述一个或多个节点指定由所述命名组件处理，来生成所选择的节点的一维表示。

18．如权利要求 13 所述的系统，其特征在于，所述命名组件生成对应于依照所分配的赋值选择的多个节点的多个展平的一维名字输出。

19．如权利要求 13 所述的系统，其特征在于，所述一个或多个名字包括表示位置数据、相对时间、新近性、再出现性和分类类型中的至少一个的元数据信息。

20．一种用于为数据结构生成使用户能够理解和使用所述数据结构的名字的

方法，其特征在于，所述方法包括：

接收具有多个节点的分层数据结构；

处理与所述多个节点相关联的观察到的用户活动，其中所述用户活动是由所述用户针对被包括在所述分层数据结构中的节点做出的活动；

向所述多个节点的每一个分配一赋值；

选择所述多个节点中与一预定赋值限制相关联的一个或多个；

提取与所述一个或多个选择的节点相关联的节点元数据；以及

基于所述相应的节点元数据，为所述一个或多个选择的节点的每一个生成名字。

21．如权利要求 20 所述的方法，其特征在于，还包括输出所述名字作为一维视图。

22．如权利要求 20 所述的方法，其特征在于，所述节点元数据是数字、图像数据、音频数据、字符串和单词的至少一个的格式。

23．如权利要求 20 所述的方法，其特征在于，还包括：

学习与所述多个节点相关联的新观察的用户活动；以及

向所述多个节点的每一个分配一新的赋值。

24．如权利要求 20 所述的方法，其特征在于，还包括滤除不重要的节点元数据。

25．如权利要求 20 所述的方法，其特征在于，还包括为第一用户生成第一组名字，并为第二用户生成第二组名字，其中，所述第一和第二组名字是从所述数据结构和相同和不同的数据结构之一生成的。

26．如权利要求 20 所述的方法，其特征在于，还包括：

检测所述节点元数据的格式；以及

以相同格式、不同格式和相同格式和不同格式的组合的至少一个生成有意的名字。

27．如权利要求 20 所述的方法，其特征在于，所述名字是使用与所述多个节点的每一节点相关联的文件夹名和文件名的至少一个来生成的。

28．如权利要求 20 所述的方法，其特征在于，还包括：

分析与所选择的节点相关联的文档的文本；以及

基于单词在所述文档的文本中使用的频率生成该所选择的节点的名字。

29．一种用于为数据结构生成使用户能够理解和使用所述数据结构的名字的

系统，其特征在于，包括：

用于接收具有多个节点的分层数据结构的装置；

用于处理与所述多个节点相关联的观察到的用户活动的装置，其中所述用户活动是由所述用户针对被包括在所述分层数据结构中的节点做出的活动；

用于向所述多个节点的每一个分配赋值的装置；

用于选择所述多个节点中与一预定赋值限制相关联的一个或多个节点的装置；

用于提取与所述一个或多个所选择的节点相关联的节点元数据的装置；以及

用于基于相应的节点元数据为所述一个或多个所选择的节点的每一个生成一维视图格式的名字的装置。

30．如权利要求 29 所述的系统，其特征在于，还包括：

用于分析与所选择的节点相关联的一个或多个文件和文件名的描述符的装置；以及

用于基于与所述文件和文件名相关联的描述符生成名字的装置。

31．如权利要求 30 所述的系统，其特征在于，所述描述符是数字、图像数据、音频数据、字符串和单词的至少一个的格式。

32．如权利要求 29 所述的系统，其特征在于，还包括：

用于学习与所述多个节点相关联的新观察的用户活动的装置；

用于用所述新观察的用户活动更新所观察的用户活动的装置；

用于基于所述新观察的用户活动向所述多个节点的每一个分配新赋值的装置。

33．如权利要求 29 所述的系统，其特征在于，所述名字包括表示位置数据、相对时间、新近性、再出现性和分类类型的至少一个的元数据信息。

用于在展平的分层结构中名字的创建的方法和系统

技术领域

本发明涉及数据存储结构，尤其涉及以更有意义的方式表示这类数据结构的机制。

背景技术

现有的网络文件共享使用了由嵌套的文件夹和文件表示的分层存储机制，它造成了导航问题。共享内容的组织在不同的共享之间，以及在管理同一共享分段的个人之间是不均匀的。由于此不一致性，详细地熟悉此共享通常对有效的导航是需要的。作为对这一问题的复杂化，文件夹命名约定通常在使用共享的各人之间不同。尽管共享如档案那样有用，然而其副作用是寻找当前更相关的某一内容的人通常必须费力地通过多个文件和文件夹。

用于计算机网络共享的新导航方法在共享活动上使用收集的数据，以呈现共享空间的更有意义的视图，如示出与特定各人相关联的更新近的文件或更相关的文件（近来由他们打开或创建）。即使在项目名称已知时，就算有了这一类型的改进，共享空间的导航仍是困难的。通常在复杂的文件夹结构中放置了大量的文件，其分段由对文件夹层次和文件使用各种命名约定的多个人来维护。

需要的是一种使用有意义的命名察看这一分层结构的改进的系统。

发明内容

以下提出了本发明的简化概述，以提供对本发明的某些方面的基本理解。本概述并非本发明的广泛综述。它并不旨在标识本发明的关键/决定性元素或描绘本发明的范围。其唯一的目的是以简化的形式提出本发明的某些概念，作为稍后提出的更详细描述的序言。

此处揭示并要求保护的本发明在其一个方面包括一种系统，它观察与分层数据结构相关联的用户活动，然后生成从文件和文件夹名字或其它字符串数据导出的便于用户的导航和理解的有意义的名字。本发明通过基于收集的共享活动数据创建

项目视图，简化了这一空间的浏览。一般而言，数据库可具有进一步基于不同字段的不同类型的层次。目录结构中的文件夹是项目的一种表示。可以有额外的字段，例如指定项目的文件、表示项目的图像、或对项目更专用的某些其它数据。通过查找这些字段，可使有意义的命名过程更智能。某些子目录可以是项目内文件的集合，而不是作为单独实体的项目，它然后可用于生成有意义的名字。任一形式的元数据可用于促进名字生成。可采用许多类型的度量，包括但不限于新近性、再出现性（它是回到同一数据位置的用户的数量）以及分类类型（与在同一时刻打开或编辑两个项目的发生时段相关联）。作为复杂查询的结果生成的平面视图为在如操作系统中快速理解并导航数据结构提供了重要手段。这些视图可从查询作为操作系统的一部分的丰富数据库而得到。

作为对其支持，本系统包括一赋值组件，它便于将赋值度量应用于数据结构的每一节点。赋值过程可以响应于若干预定条件而执行。本系统也包括一选择组件，它选择一个或多个节点用于命名过程。一命名组件接收并处理来自选择组件的所选择的节点信息，并为给定的用户生成一个或多个节点的名字。命名组件的输出是数据结构的有意义名字形式的展平的表示。这一输出可以简单地为从各个节点生成的有意义名字的表格式清单。向用户提供了这类共享空间的更直观的视图，而无需该结构的内部知识，从而允许用户快速地审阅并选择名字或名字串来导航到期望的数据位置。监视给定共享上的活动的组件用于使用数据的收集。在一个实现中，活动数据用于展平复杂分层结构的过程，以及在展平的空间中有意义名字的生成。

在本发明的另一方面，提供了人工智能来自动化并学习各种用户活动。此外，自动化可以基于数据类型以及其它预定的用户定义的度量，以及学到的度量。

为实现上述和相关的目的，此处结合以下描述和附图揭示了本发明的某些说明性方面。然而，这些方面仅表明了其中可采用本发明的原理的各种方法中的少数一些，本发明旨在包括所有这样的方面及其等效方面。当结合附图考虑以下本发明的详细描述时，可以清楚本发明的其它优点和新颖特征。

附图说明

图 1 示出了依照本发明的命名系统。

图 2 示出了本发明的赋值过程的流程图。

图 3 示出了用于分层文件结构的更具体实例的本发明的命名过程。

图 4 示出了依照本发明采用人工智能的命名系统。

图 5 示出了依照本发明使用人工智能学习度量信息的过程的流程图。

图 6 示出了依照本发明生成有意义名字的过程的流程图。

图 7 示出了可用于执行所揭示的体系结构的计算机的框图。

图 8 示出了依照本发明的示例性计算环境的示意性框图。

具体实施方式

现在参考附图描述本发明，贯穿附图，相同的标号指相同的元素。在以下描述中，为解释的目的，陈述了众多具体细节，以提供对本发明的彻底理解。然而，很明显，本发明可以不使用这些具体细节来实施。在其它实例中，以框图的形式示出了众所周知的结构和设备，以便于描述本发明。

如本申请中所使用的，术语“组件”和“系统”指的是计算机相关的实体，无论是硬件、硬件和软件的组合、软件还是执行中的软件。例如，组件可以是，但不限于，运行在处理器上的进程、处理器、对象、可执行码、执行线程、程序和/或计算机。作为说明，运行在服务器上的应用程序和服务器都可以是组件。一个或多个组件可驻留在进程和/或执行线程中，并且组件可位于一个计算机上和/或在两个或多个计算机之间分布。

如此处所使用的，术语“推断”一般指从一组通过事件和/或数据捕捉的观察中推出或推断系统、环境和/或用户的状态的过程。例如，推断可用于标识特定的上下文或行动，或可生成状态的概率分布。推断可以是概率性的－即，基于数据和事件的考虑计算感兴趣的状态的概率分布。推断也可以指用于从一组事件和/或数据组成更高级事件的技术。这类推断导致从一组观察的事件和/或储存的事件数据构造新的事件或行动，无论那些事件是否在相邻的时间上相关，也无论事件是否来自一个或若干个事件和数据源。

现在参考图 1，示出了依照本发明的命名系统 100。提供了要依照本发明展平并命名的常规分层数据结构 102 的部分示例。结构 102 包括多个父节点，表示为节点 1、……、节点 N，其每一个可具有从其分支出的一个或多个子节点，尽管这不是必需的。为此一般描述的目的，第一父节点 104 具有第一子节点 106（也表示为节点 11），其中第一子节点 106 也是第二子节点 108（也表示为节点 111）的父节点。

结构 102 的每一节点被分配一赋值标记 110（对第一父节点 104 被表示为 V_1，而对第一父节点 104 的一个可能路径的子节点被表示为 V_{11}、V_{111}……），它提供

了该节点的重要性的某种指示，用于最终获得对到该节点的路径的命名。这一赋值标记是依照若干不同的预定准则导出的。例如，准则可包括创建节点的时间、访问时间、修改时间、对与该节点相关联的数据执行复制过程的时间、访问该节点的频率、单独用户的数量、以及与该节点相关联的数据的类型等等。在任一情况下，一旦确定了赋值并分配给该节点，则可在准备生成有意义的名字时发生节点选择。

作为其支持，系统 100 包括赋值组件 112，它便于将赋值标记 110 应用到数据结构 102 的每一节点。赋值过程可以响应于多个预定条件而执行。例如，在数据结构只驻留在客户机计算机上的客户机计算机环境中，赋值过程可以在计算机的启动之后执行，或响应于用户与其的交互而执行。或者，赋值可以来自先前的会话，其更新依照用户的手动启动而发生。当数据结构 102 包含远程数据节点时，赋值过程可以响应于用户登录过程或网络而自动执行。在任一情况下，赋值过程在时间中的某一点并依照无论是手动还是自动生成的触发事件来执行。如上所述，可采用多个准则来获得每一节点的赋值，其中准则可以使用赋值组件 112 来采用。

系统 100 也包括选择组件 114，它选择一个或多个节点用于命名过程。描述组件 116 接收并处理来自选择组件 114 的所选择的节点信息，并对给定用户生成对一个或多个节点的描述。描述组件 116 的输出 118 是数据结构 102 的有意义描述形式的展平的表示。这一输出 118 可以简单地为从各个节点生成的描述的表格式清单，其每一个或任一个然后可以被超链接，以便于导航到对应的数据节点。用户然后可快速审阅并选择描述或描述字符串，以快速导航到期望的数据位置。本领域的技术人员可以理解，估值组件 112、选择组件 114 和描述组件 116 可以被组合到执行所有相应功能的单个组件中。

现在参考图 2，示出了本发明的赋值过程的流程图。尽管为解释的简明性的目的，此处例如以流程图的形式示出的一个或多个方法被示出并描述为一系列动作，然而可以理解，本发明不受动作的顺序所限制，因为依照本发明，某些动作可以不同的顺序和/或与此处示出并描述的其它动作一起发生。例如，本领域的技术人员将理解和明白，方法可替换地被表示为诸如状态图中一系列互相关的状态或事件。此外，并非需要所有示出的动作来实现依照本发明的方法。

在 200，确定赋值度量，并将其分配给数据层次的每一节点。这一准则可包括例如创建时间、访问时间、修改时间、数据是否已被修改、复制时间、访问频率、修改频率、单独用户的数量、与该节点相关联的数据的类型、节点的加权因子、一天中的时间。在 202，依照期望的赋值数据选择节点。在 204，使用所选择的节点

的数据来生成有意义的描述以呈现给用户。过程然后到达停止块。

现在参考图3，示出了本发明的一个命名过程，用于分层文件结构300的更具体的示例。本发明提供了基于观察到的用户活动展平分层结构并呈现简化的一维视图（如，以列表、播放列表……的形式）的有效方法。图3的示例在使用网络文件共享的环境中描述。计算机文件系统已经变得越来越数据库驱动，使软件设计者能够更灵活地组织嵌入在文件系统中的元数据。使用了观察分层结构中的活动的系统来导出展平的、便于用户结构的导航和理解的名字。

图3示出了具有某一所选择的测量（如，基于新近性和访问频率）的值的分层网络文件共享结构300的简化视图。名字生成系统302监视用于收集使用数据的给定共享结构300上的用户活动。系统302包括赋值组件112、选择组件114（先前已参考图1描述）和命名组件303。结构300中的某些节点可用于文档目的，而其它节点可用于最近的最新活动。系统302向用户提供了这类共享空间的更有意义的视图，而无需数据结构的内部知识。活动数据用于展平复杂分层结构的过程，以及用于生成展平的空间中的有意义的名字。

如图3所示，生成赋值数据（在括号中）并将其与结构300的每一节点相关联。例如，在从根节点开始的第一层，生成A节点，并将其与1399的赋值相关联。注意，在本示例中对节点使用大写字母用于描述本发明的一般操作的目的。然而，实际上，在每一节点处可用用户定义的名字来替换这些字母，如在系统302的输出之处可变得显而易见。然而，可以理解，用户可以将这些节点（即，文件夹、子文件夹等）命名为字母，如示例中所示，并且对其仍可生成有意义的名字作为展平的输出。这简单地可以是，为展平的结构生成的有意义的名字仅对该用户有意义。即，输出是对给定用户有意义的字母串，即使它不是一个或多个单词的形式。如此处所描述的，这也可以意味着命名系统302可用于从同一数据结构300中生成不同用户具有不同意义名字的展平的结构。

继续参考图3，如所见到的，A节点的最高赋值与沿A-AB-ABA-ABAA定义的路径相关联，它们具有相应的赋值1399、1340、1290和1234。由此，该表示的展平的输出是A-ABAA。类似地，对于B节点，最高的赋值定义了沿B-BA-BAA-BAAA的路径，它们具有相应的赋值2712、2370、2345和2290。该表示的展平的输出是B-BAAA。由此，对于这一小结构300，展平的输出是A-ABAA和B-BAAA的清单。

上述示例中的字母串是通过将由超过某一阈值的高赋值节点连接在一起而生

成的。实际上，名字可被进一步转换成更有意义的内容，如 Chicago-Design Specifications（芝加哥－设计规范）或 Company-Legal-Justice Dept.（公司－法律－审判部门），其任一个的选择可直接将用户路由到该位置。由此，命名组件 116 生成最终用于有意义的表示的名字。

如上所述，同一数据结构 300 的有意义的名字可依照用户的希望来生成。第一用户可选择将结构 300 表示为 A-ABAA 和 B-BAAA，因为这可以对第一用户是有意义的，然而看上去是神秘的。而“察看”由同样字母指定的带有相同节点的同一结构 300 的第二用户可选择以不同的方式来表示展平的输出 118。即，第二用户可选择令系统 302 对每一节点生成文本描述符，导致以类似于 Chicago-Design Specifications 或 Company-Legal-Justice Dept.的形式输出命名，因为这对第二用户将更有意义。察看同一数据结构的第三用户可请求对他们更有意义的命名输出，它是例如字母数字串（或统一代码（Unicode）字符和/或图像）和单词的组合的形式。由此，本发明不限于仅作为单词来输出有意义的名字，而是可采用对用户有意义的任何方法。

作为其支持，命名系统 302 提供了为希望察看展平的输出作为单词而非字符串的第二用户，对图 3 所示的加字母的节点提取单词的能力。例如，系统 302 可探查文件夹中与给定的单个或多个字母节点相关联的数据，并提取比其它单词（不包括冠词、代词等）更频繁使用的单词（或多个单词），并使用该单词来描述该节点。或者，系统可选择文件夹中出现的若干更频繁使用的单词，并向第二用户呈现这些单词，以供第二用户选择其中的一个或两个来描述该节点。系统 302 然后可基于那些用户选择的单词进一步生成展平的输出 118。

现在参考图 4，依照本发明示出了采用人工智能的命名系统 400。本发明（如，结合选择和命名生成）可采用各种基于人工智能的方案用于实现本发明的各方面。例如，用于确定要使用什么类型的有意义的命名方案的过程可通过自动分类器系统和过程 402 来得到方便。此外，当数据结构 102 不完全在用户的计算机上本地，而是通过若干位置分布的数据时，分类器 402 可用于例如基于结构 102 中的主要数据类型来确定要采用哪一命名方案。

分类器是将输入属性矢量 x=(x1, x2, x3, x4, xn)映射到输入属于一个类的置信度的函数，即 f(x)=*confidence*(*class*)。这一分类可采用基于概率和/或基于统计的分析（如，分解成分析效用和成本），来预测或推断用户希望自动执行的行动。

支持矢量机（SVM）是可采用的分类器的一个示例。SVM 通过找出可能输入

的空间中的超表面来运作，其中，超表面试图将触发准则从非触发事件中分离出来。直观上，这使分类对于测试接近但不等同于训练数据的数据是正确的。其它有向和无向模型分类方法包括，如单纯贝叶斯、贝叶斯网络、决策树，并且可采用提供不同的独立性模式的分类模型。此处所使用的分类也包括用于开发优先级模型的统计回归。

如从本说明书中可以理解的，本发明可采用明确训练的（如通过普通训练数据）以及隐含训练的（如通过观察用户行为、接收外来信息）分类器。例如，SVM通过分类器构造函数和特征选择模块中的学习或训练阶段来配置。由此，分类器可用于自动执行若干功能，包括但不限于以下：依照预定的准则确定如何向节点分配赋值；当作出赋值时包括什么度量；在为节点生成赋值时是否采用加权方案；基于用户采用特定的赋值方案；采用什么字符串或文本来生成名字输出；忽略一般学习的不重要的词语和字符（如，冠词、代词和串）；从节点上的若干词语确定使用哪一单个单词（或字符或图像）或单词（或字符或图像）的组合；当可采用字体或字符着色时，在单个字符和/或单词或其组合中可使用什么颜色或颜色的组合以向用户表示；什么声音可被链接到字符的有意义的名字词语，等等。

现在参考图 5，示出了依照本发明使用人工智能学习度量信息的过程的流程图。在 500，用户确定用于确定任一在给定节点处的赋值的初始度量组。在 502，将这些度量应用于数据结构的许多节点，以在 504 获得节点赋值。在 506，系统依照测量确定是否可采用新的度量。如果是，则在 508，系统学习观察到的赋值度量。在 510，系统相应地更新现有的赋值标准组。例如，这可包括一“超过时效（age out）”过程，它在预定时间段之后从赋值度量组中溢出不再被观察的一个或多个准则。或者，更新过程包括向现有的准则组添加一个或多个准则，以进一步增强有意义的命名过程。在 512，选择具有期望赋值的节点。在 514，然后对节点或节点路径生成有意义的名字（或单词的组合）。过程然后到达停止块。

如果系统未观察到新的度量，则过程从 506 流到 512，以选择具有期望赋值的节点，然后过程相应地获得有意义的名字。

本领域的技术人员可以理解，赋值不需要是最高值，但是可以是具有一个范围之内的赋值的节点－即，较高的范围、中间范围或甚至是较低的范围。例如，用户可能期望为数据结构中观察到最低活动的部分（低范围赋值）或观察到中等活动的部分（在中赋值范围）生成有意义的名字。

现在参考图 6，示出了依照本发明生成有意义的名字的过程的流程图。在 600，

选择一节点用于名字处理。这可包括单个高级节点，或其路径信息与需要被处理来生成所得的有意义名字的多个词语相关联的子节点。在 602，系统运作以开始处理路径信息。在 604，从第一节点获取节点标识符数据。在 606，确定节点赋值是否满足命名阈值。如果是，则在 608，确定节点标识符数据是否为单词文本或字符串。如果是文本，则过程流到 610，采用一过滤器来滤除诸如冠词或代词等不重要的文本。然而，注意，过滤的级别和所过滤的文本的类型可被调整，以包括或排除用户期望的任何单词。这一特征推断了单词字典的使用，系统通过该字典可测试节点标识符数据。这一字典可依照所得的有意义的命名信息来自动更新，如由分类器所执行的。在 612，对该节点选择描述性文本。这可包括单个单词或单词的组合。在 614，确定是否是要处理的最后一个节点。如果是，过程流到 616，生成并输出有意义的名字。过程然后到达停止块。

如果这不是要处理的最后一个节点，则过程从 614 流到 618，以处理下一节点。过程然后返回到 606，以确定下一节点是否满足阈值。如果节点不满足命名阈值，则过程从 606 流到 620 以忽略该节点。在 618，处理下一节点，然后过程相应地前进。

如果确定该节点标识符数据（或描述符数据）不是文本，而是字符串或可能是图像或音频数据，则过程从 608 流到 622，以导出该节点的代表性标识符串、音频和/或图像数据。在 614，再一次，确定这是否是最后一个节点。如果是，则可使用任何相关联的串数据、音频和/或图像数据来生成有意义的名字。

现在参考图 7，示出了用于执行所揭示的体系结构的计算机的框图。为向本发明的各方面提供额外的环境，图 7 及以下讨论旨在提供其中可实现本发明的各方面的合适的计算环境 700 的简要、概括描述。尽管上文在可运行于一个或多个计算机上的计算机可执行指令的一般上下文中描述了本发明，然而本领域的技术人员将认识到，本发明也可结合其它程序模块和/或作为硬件和软件的组合来实现。

一般而言，程序模块包括例程、程序、对象、组件、数据结构等等，执行特定的任务或实现特定的抽象数据类型。此外，本领域的技术人员可以理解，本发明的方法可以用其它计算机系统配置来实施，包括单处理器或多处理器计算机系统、小型机、大型机、以及个人计算机、手持式计算设备、基于微处理器或可编程消费者电子设备等等，其每一个可在操作上耦合至一个或多个相关联的设备。

所示的本发明的各方面也可在分布式计算环境中实施，其中，某些任务由通过通信网络连接的远程处理设备来执行。在分布式计算环境中，程序模块可以位于

本地和远程存储器存储设备中。

计算机通常包括各种计算机可读介质。计算机可读介质可以是可由计算机访问的任一可用介质，包括易失和非易失介质、可移动和不可移动介质。作为示例而非局限，计算机可读介质包括计算机存储介质和通信介质。计算机存储介质包括以用于储存诸如计算机可读指令、数据结构、程序模块或其它数据等信息的任一方法或技术实现的易失和非易失，可移动和不可移动介质。计算机存储介质包括但不限于，RAM、ROM、EEPROM、闪存或其它存储器技术、CD-ROM、数字视频盘（DVD）或其它光盘存储、磁盒、磁带、磁盘存储或其它磁存储设备、或可以用来储存所期望的信息并可由计算机访问的任一其它介质。

通信介质通常在诸如载波或其它传输机制的已调制数据信号中包含计算机可读指令、数据结构、程序模块或其它数据，并包括任一信息传送介质。术语“已调制数据信号”指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。作为示例而非局限，通信介质包括有线介质，如有线网络或直接连线连接，以及无线介质，如声学、RF、红外和其它无线介质。上述任一的组合也应当包括在计算机可读介质的范围之内。

再次参考图 7，示出了用于实现本发明的各方面的示例性环境 700，包括计算机 702，计算机 702 包括处理单元 704、系统存储器 706 和系统总线 708。系统总线 708 将包括但不限于系统存储器 706 的系统组件耦合至处理单元 704。处理单元 704 可以是各种可购买的处理器的任一种。双微处理器和其它多处理器体系结构也可用作处理单元 704。

系统总线 708 可以是若干种总线结构类型的任一种，它们可进一步互连到存储器总线（使用或不使用存储器控制器）、外围总线和使用各类可购买的总线体系结构的任一种的局部总线。系统存储器 706 包括只读存储器（ROM）710 和随机存取存储器（RAM）712。基本输入/输出系统（BIOS）包括如在启动时帮助在计算机 702 内的元件之间传输信息的基本例程，储存在如 ROM、EPROM、EEPROM 等非易失存储器 710 中。RAM 712 也包括如静态 RAM 的高速 RAM，用于高速缓存数据。

计算机 702 还包括内部硬盘驱动器（HDD）714（如，EIDE、SATA），该内部硬盘驱动器 714 也可被配置成在合适的机壳（未示出）中作外部使用、磁软盘驱动器（FDD）716（如，对可移动磁盘 718 进行读写）以及光盘驱动器 720（如，读 CD-ROM 盘 722，或对诸如 DVD 等其它高容量光介质进行读写）。硬盘驱动器

714、磁盘驱动器 716 和光盘驱动器 720 可分别通过硬盘驱动器接口 724、磁盘驱动器接口 726 和光驱接口 728 连接到系统总线 708。用于外部驱动器实现的接口 724 包括通用串行总线（USB）和 IEEE 1394 接口技术的至少一个或两者。

驱动器及其相关联的计算机可读介质提供了数据、数据结构、计算机可指令指令等的非易失存储。对于计算机 702，驱动器和介质容纳合适的数字格式的任何数据的存储。尽管上文对计算机可读介质的描述涉及 HDD、可移动磁盘和诸如 CD 或 DVD 的可移动光盘，然而本领域的技术人员可以理解，计算机可读的其它类型的介质，如 zip 驱动器、磁带盒、闪存卡、盒式磁带等等，也可用于该示例性操作环境，并且任何一种这类介质可包含用于执行本发明的方法的计算机可执行指令。

若干程序模块可储存在驱动器和 RAM 712 中，包括操作系统 730、一个或多个应用程序 732、其它程序模块 734 和程序数据 736。操作系统、应用程序、模块和/或数据的所有或部分也可被高速缓存在 RAM 712 中。

可以理解，本发明可用各种可购买的操作系统或操作系统的组合来实现。

用户可通过一个或多个有线/无线输入设备，如键盘 738 和诸如鼠标 740 等定位设备向计算机 702 输入命令和信息。其它输入设备（未示出）可包括麦克风、IR 遥控器、操纵杆、游戏垫、触针笔、触摸屏等等。这些和其它输入设备通常通过耦合到系统总线 708 的输入设备接口 742 连接到处理单元 704，但也可通过其它接口连接，如并行端口、IEEE 1394 串行端口、游戏端口、USB 端口、IR 接口等等。

监视器 744 或其它类型的显示设备也通过接口，如视频适配器 746 连接到系统总线 708。除监视器 744 之外，计算机通常包括其它外围输出设备（未示出），如扬声器、打印机等等。

计算机 702 可以在通过有线和/或无线通信使用到一个或多个远程计算机，如远程计算机 748 的逻辑连接的网络化环境中操作。远程计算机 748 可以是工作站、服务器计算机、路由器、个人计算机、便携式计算机、基于微处理器的娱乐设备、对等设备或其它公用网络节点，并通常包括许多或所有上述与计算机 702 相关的元件，尽管为简明起见，仅示出了存储器存储设备 750。描述的逻辑连接包括到局域网（LAN）752 和/或更大的网络，如广域网（WAN）754 的有线/无线连接。这类 LAN 和 WAN 网络环境常见于办公室和企业，并方便了企业范围计算机网络，如内联网，它们所有都可连接到全球通信网络，如因特网。

当在 LAN 网络环境中使用时，计算机 702 通过有线和/或无线通信网络接口或适配器 756 连接至局域网 752。适配器 756 可便于到 LAN 752 的有线或无线通信，

它可包括在其上布置的无线接入点，用于与无线适配器 756 通信。当在 WAN 网络环境中使用时，计算机 702 可包括调制解调器 758，或通过 LAN 连接到通信服务器，或具有用于通过 WAN 754，如因特网建立通信的其它装置。调制解调器 758 可以是内置或外置的，并且可以是有线或无线设备，通过串行端口接口 742 连接至系统总线 708。在网络化环境中，描述的与计算机 702 相关的程序模块或其部分可储存在远程存储器/存储设备 750 中。可以理解，示出的网络连接是示例性的，也可以使用在计算机之间建立通信链路的其它装置。

计算机 702 可用于与操作上布置在无线通信中的任一无线设备或实体通信，如打印机、扫描仪、台式和/或便携式计算机、便携式数据助理、通信卫星、与无线可检测标签相关联的任一个设备或位置（如，公用电话亭、报亭、休息室）和电话。这至少包括 Wi-Fi 和 Bluetooth™（蓝牙）无线技术。由此，通信可以是如常规网络那样的预定义结构，或者简单地为至少两个设备之间的特别（ad hoc）通信。

Wi-Fi，或无线保真，允许从家里的沙发、酒店房间的床或工作时的会议室连接到因特网，而不需要线缆。Wi-Fi 是一种类似于蜂窝电话的无线技术，它使这类设备，如计算机，能够在室内和室外发送和接收信息；在基站的范围之内的任何地方。Wi-Fi 网络使用称为 IEEE 802.11（a、b、g 等）无线电技术来提供安全、可靠、快速的无线连接。Wi-Fi 网络可用于将计算机彼此连接、连接到因特网、并连接到有线网络（使用 IEEE 802.3 或以太网）。Wi-Fi 网络在未许可的 2.4 和 5 GHz 无线电频带中运作，它具有 11 Mbps（802.11b）或 54 Mbps（802.11a）的数据速率，或具有包含两种频带（双频带）的产品，因此网络可提供类似于许多办公室中使用的基本 10BaseT 有线以太网的真实世界的性能。

现在参考图 8，示出了依照本发明的示例性计算环境 800 的示意框图。系统 800 包括一个或多个客户机 802。客户机 802 可以是硬件和/或软件（如，线程、进程、计算装置）。例如，客户机 802 通过采用本发明可容纳 cookie 和/或相关联的上下文信息。系统 800 也可包括一个或多个服务器 804。服务器 804 也可以是硬件和/或软件（如，线程、进程、计算装置）。例如，服务器 804 通过使用本发明可容纳线程以执行变换。客户机 802 和服务器 804 之间的一个可能的通信可以是适用于在两个或多个计算机进程之间传输的数据分组的形式。例如，数据分组可包括 cookie 和/或相关联的上下文信息。系统 800 包括可用于便于在客户机 802 和服务器 804 之间通信的通信框架 806（如，诸如因特网等全球通信网络）。

可通过有线（包括光纤）和/或无线技术来方便通信。客户机 802 操作上连接

至可用于储存对客户机 802 本地的信息（如，cookie 和/或相关联的上下文信息）的一个或多个客户机数据存储 808。类似地，服务器 804 操作上连接至可用于储存对服务器 804 本地的信息的一个或多个服务器数据存储 810。

上文所描述的包括本发明的示例。当然，不可能为了描述本发明而描述组件或方法的每一可想象的组合，但是本领域的普通技术人员可以认识到，本发明的许多另外的组合和置换是可能的。因此，本发明旨在包含落入所附权利要求书的精神和范围中的所有这样的改变、修改和变化。此外，在术语“包括”用于详细描述和权利要求书的意义上，这类术语旨在以与术语“包含”用作权利要求书中的过渡词语解释成“包含”的类似方式为包含性的。

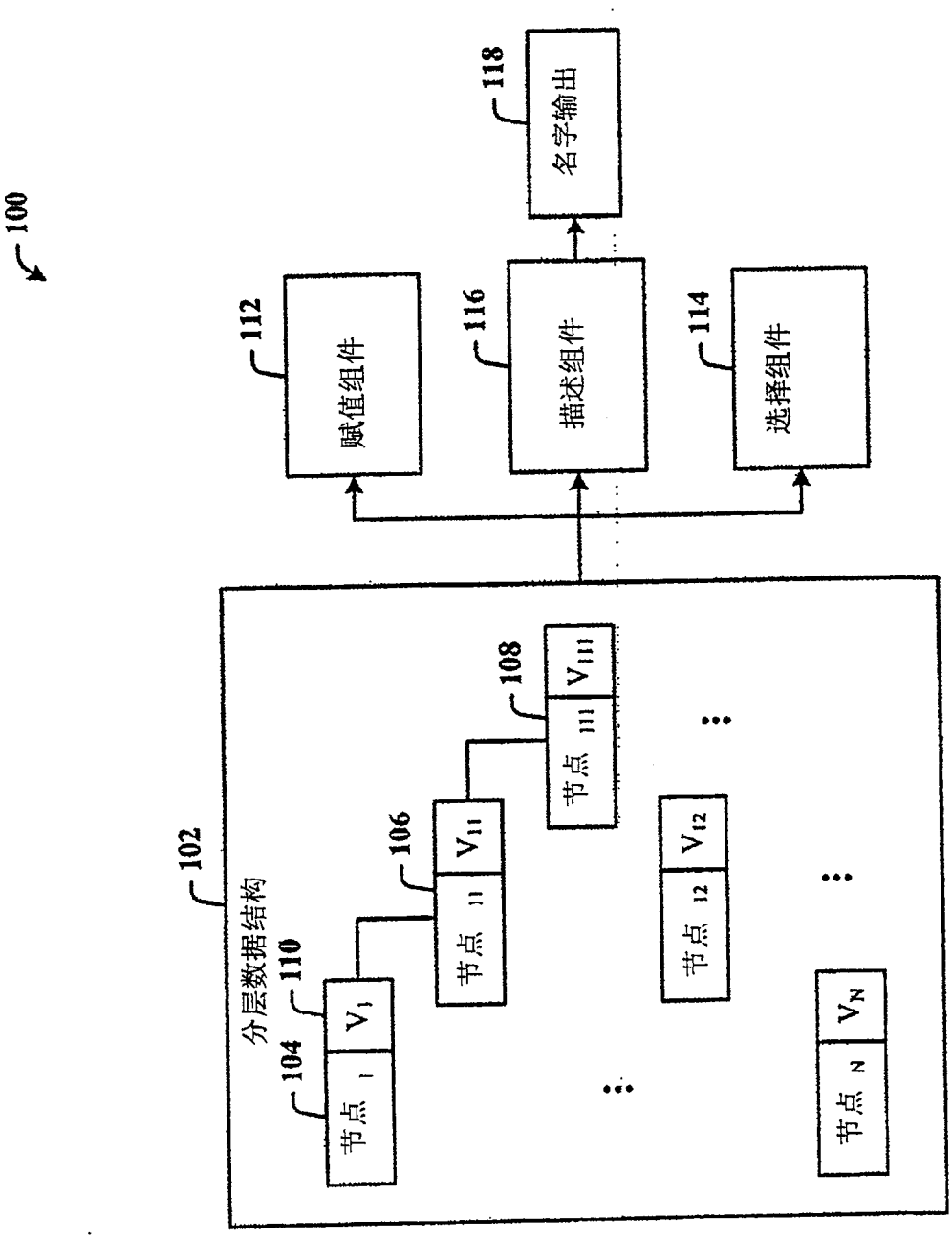

100
102
分层数据结构
104
110
106
108
节点 I
V_I
节点 II
V_{II}
节点 III
V_{III}
节点 I2
V_{I2}
节点 N
V_N
112
赋值组件
116
描述组件
114
选择组件
118
名字输出

图 1

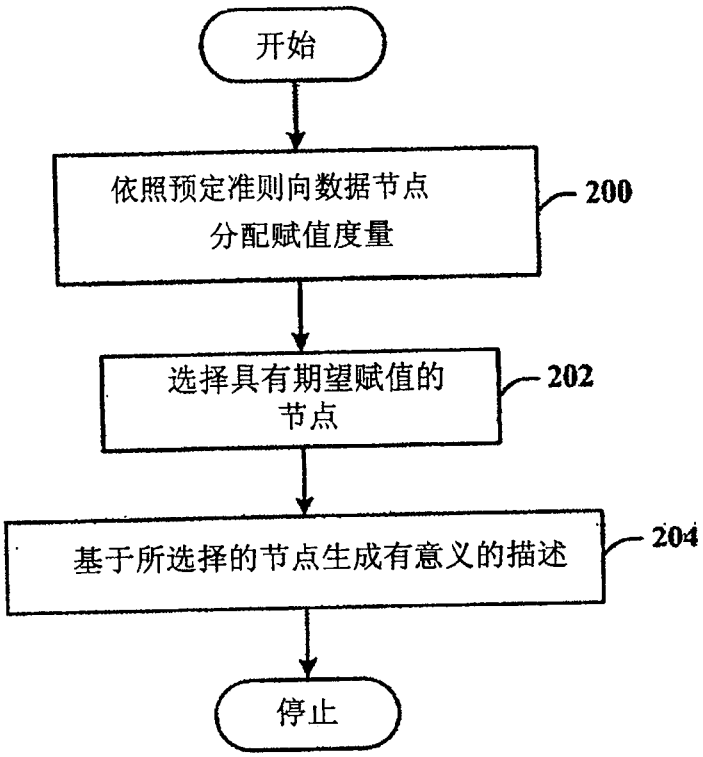

开始
依照预定准则向数据节点分配赋值度量
200
选择具有期望赋值的节点
202
基于所选择的节点生成有意义的描述
204
停止

图 2

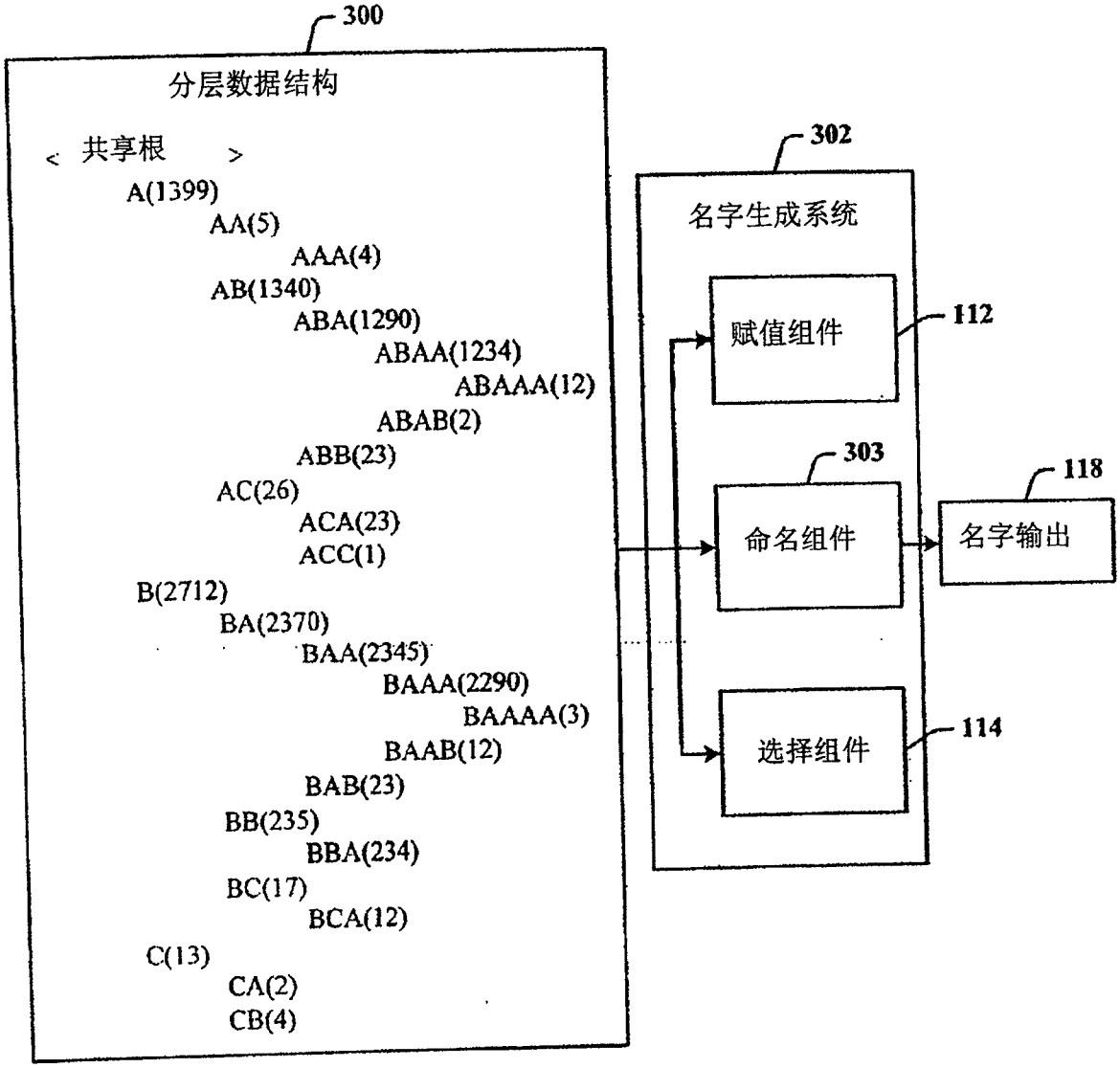

300
分层数据结构
< 共享根 >
A(1399)
AA(5)
AAA(4)
AB(1340)
ABA(1290)
ABAA(1234)
ABAAA(12)
ABAB(2)
ABB(23)
AC(26)
ACA(23)
ACC(1)
B(2712)
BA(2370)
BAA(2345)
BAAA(2290)
BAAAA(3)
BAAB(12)
BAB(23)
BB(235)
BBA(234)
BC(17)
BCA(12)
C(13)
CA(2)
CB(4)
302
名字生成系统
赋值组件
112
303
命名组件
118
名字输出
选择组件
114

图 3

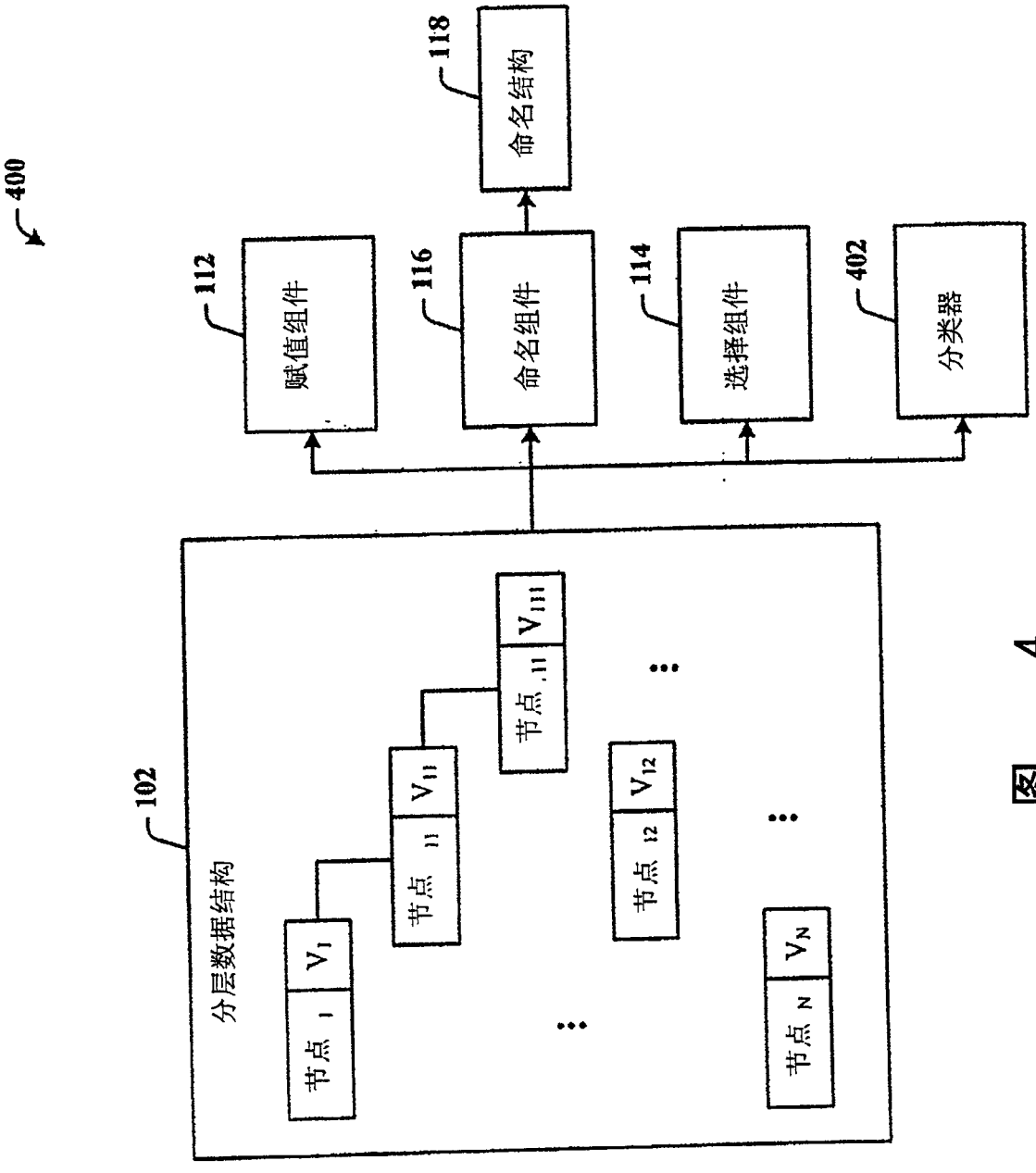

400
102
分层数据结构
节点 I
V_I
节点 II
V_{II}
节点 III
V_{III}
节点 I2
V_{I2}
节点 N
V_N
112
赋值组件
116
命名组件
118
命名结构
114
选择组件
402
分类器

图 4

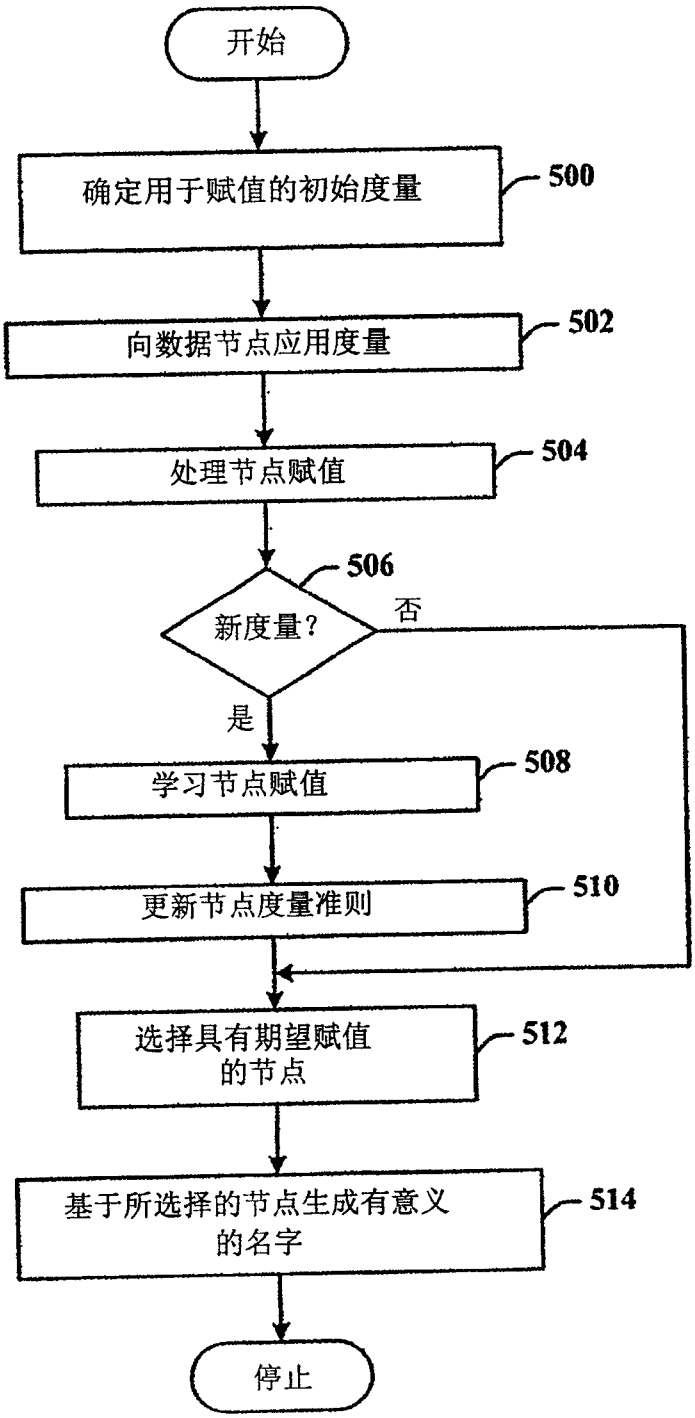

开始
确定用于赋值的初始度量
500
向数据节点应用度量
502
处理节点赋值
504
506
新度量？
否
是
学习节点赋值
508
更新节点度量准则
510
选择具有期望赋值的节点
512
基于所选择的节点生成有意义的名字
514
停止

图 5

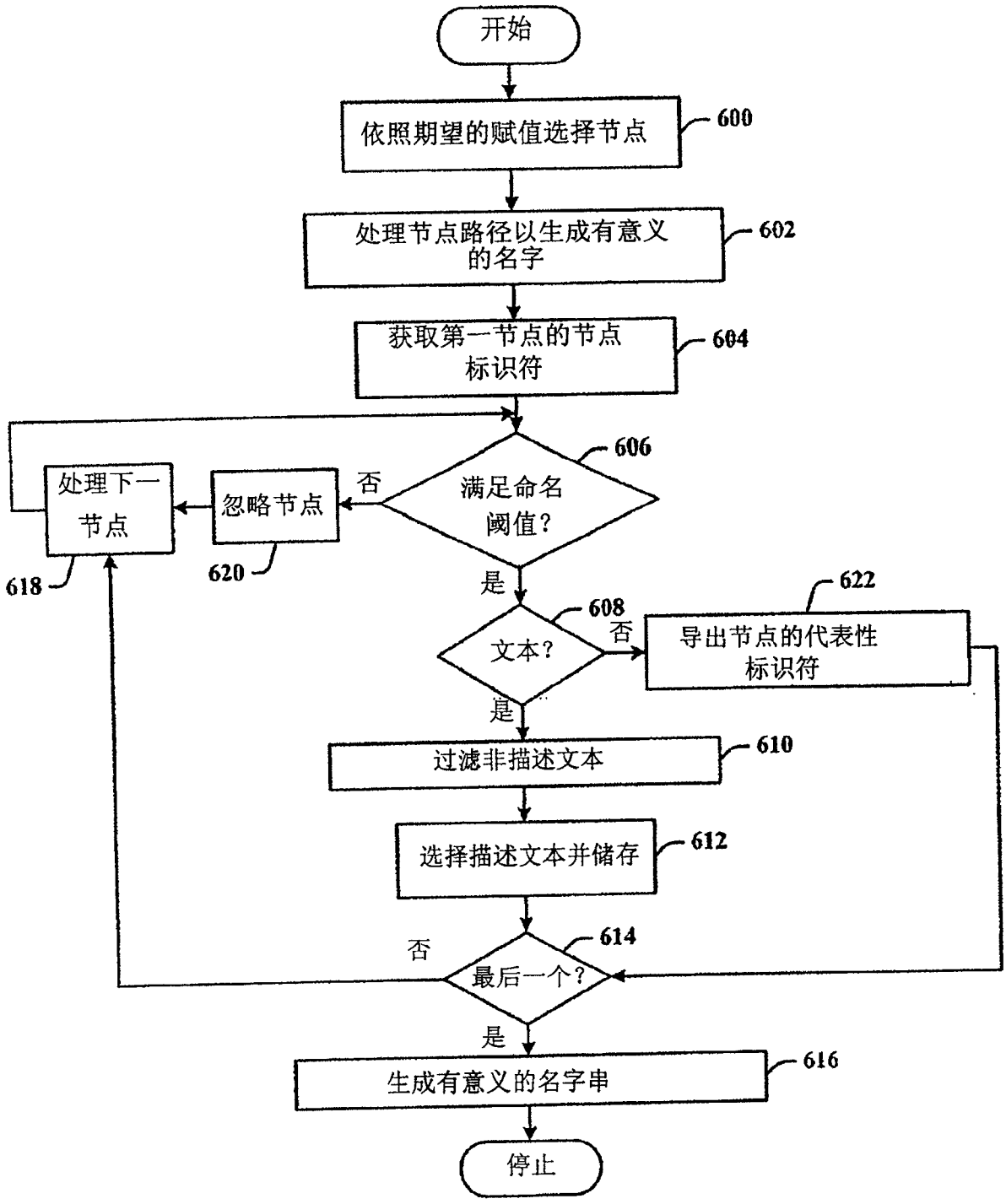

开始
依照期望的赋值选择节点
600
处理节点路径以生成有意义的名字
602
获取第一节点的节点标识符
604
606
满足命名阈值?
否
忽略节点
620
处理下一节点
618
是
608
文本?
否
622
导出节点的代表性标识符
是
过滤非描述文本
610
选择描述文本并储存
612
614
否
最后一个?
是
生成有意义的名字串
616
停止

图 6

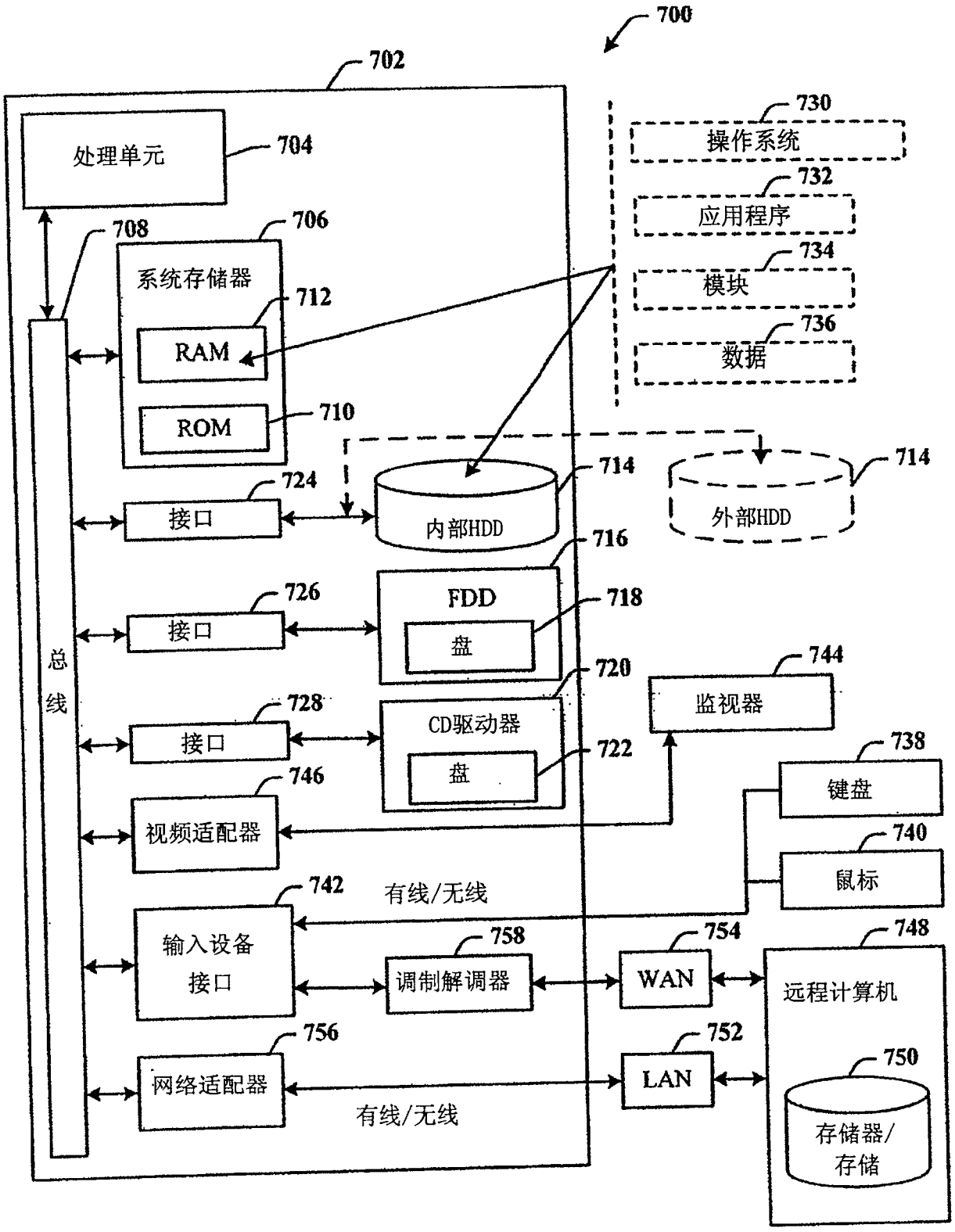

700
702
处理单元
704
730
操作系统
732
应用程序
734
模块
736
数据
708
706
系统存储器
712
RAM
ROM
710
714
内部HDD
714
外部HDD
724
接口
716
FDD
726
接口
718
盘
720
CD驱动器
744
监视器
728
接口
722
盘
738
键盘
746
视频适配器
740
鼠标
742
有线/无线
输入设备
接口
758
调制解调器
754
WAN
748
远程计算机
756
网络适配器
752
LAN
750
有线/无线
存储器/
存储
总
线

图 7

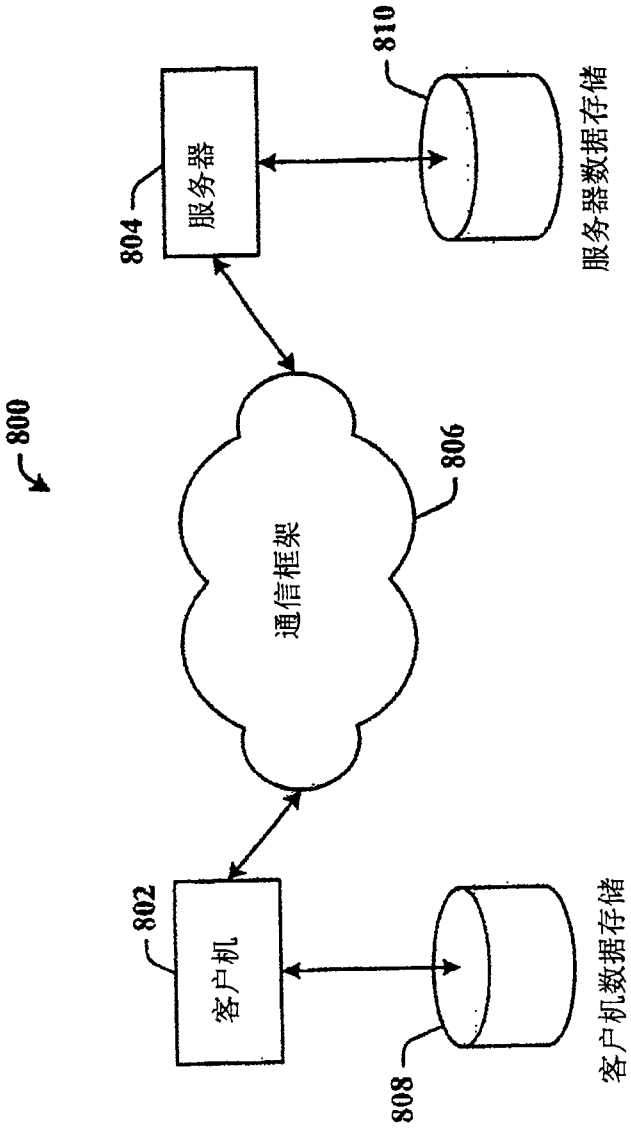


图 8