

[19] 中华人民共和国国家知识产权局

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



[12] 发明专利申请公开说明书

[21] 申请号 200480015212.1

[43] 公开日 2006 年 7 月 5 日

[11] 公开号 CN 1799050A

[22] 申请日 2004.4.12

[74] 专利代理机构 上海专利商标事务所有限公司
代理人 张政权

[21] 申请号 200480015212.1

[30] 优先权

[32] 2003.5.1 [33] US [31] 10/427,550

[86] 国际申请 PCT/US2004/011379 2004.4.12

[87] 国际公布 WO2004/099901 英 2004.11.18

[85] 进入国家阶段日期 2005.12.1

[71] 申请人 微软公司

地址 美国华盛顿州

[72] 发明人 Z·陈 S·刘 马维英

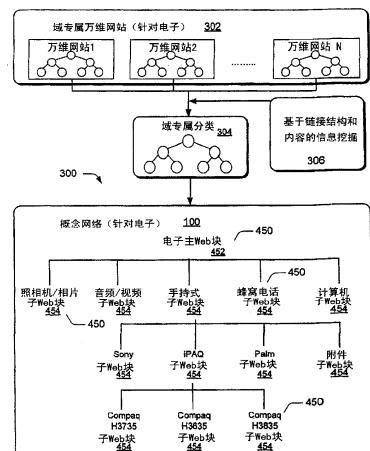
权利要求书 4 页 说明书 20 页 附图 6 页

[54] 发明名称

概念网络

[57] 摘要

一种可响应于用户查询而被生成的概念网络(100)。在一个实施例中，概念网络(100)配合搜索工具使用。该搜索工具搜索多个数据存储位置。每个数据存储位置都用一个节点(302)来排列。这些节点中的某几个至少用一个链接(306)相连。该概念网络(100)基于该链接(306)选择这些节点中的某几个中的一部分，其中这至少一个链接是用于内容目的(306)。



1. 一种装置，包括：

搜索多个数据存储位置的搜索工具，每个数据存储位置被安排成一个节点，

5 某些节点由至少一个链接相连；以及

概念网络，它基于所述链接考虑所述多个节点中的某几个的一部分，其中，所述至少一个链接是用于内容的目的。

2. 如权利要求 1 所述的装置，其特征在于，所述链接是用于导航的目的以及内容的目的。

10 3. 如权利要求 1 所述的装置，其特征在于，所述概念网络显示所述多个节点中的某几个。

4. 如权利要求 1 所述的装置，其特征在于，所述概念网络依靠域专属分类部分来选择所述多个节点中的某几个的一部分。

15 5. 如权利要求 4 所述的装置，其特征在于，所述域专属分类部分从多个域专属数据存储位置选择输入。

6. 如权利要求 4 所述的装置，其特征在于，所述域专属分类接收来自信息挖掘部分的输入。

7. 如权利要求 6 所述的装置，其特征在于，所述信息挖掘部分是基于链接结构。

20 8. 如权利要求 1 所述的装置，其特征在于，所述概念网络选择所述多个节点的某几个的一部分是依靠所述数据存储位置的结构。

9. 如权利要求 8 所述的装置，其特征在于，所述结构是基于统一资源定位器（URL）。

10. 如权利要求 9 所述的装置，其特征在于，所述 URL 获得多个数据存储位
25 置的结构。

11. 如权利要求 1 所述的装置，其特征在于，还包括合并数据存储位置。

12. 如权利要求 11 所述的装置，其特征在于，所述合并是由万维网站结构合
并工具执行的。

13. 如权利要求 11 所述的装置，其特征在于，还包括万维网站结构分析器。

30 14. 如权利要求 13 所述的装置，其特征在于，所述万维网站结构分析器包括

超链接队列。

15. 如权利要求 13 所述的装置，其特征在于，所述万维网站结构分析器包括万维网站爬寻器。

16. 如权利要求 13 所述的装置，其特征在于，所述万维网站结构分析器包括
5 HTML 语法分析器。

17. 如权利要求 13 所述的装置，其特征在于，所述万维网站结构分析器包括基于功能的对象模型（FOM）分析器。

18. 如权利要求 13 所述的装置，其特征在于，所述万维网站结构分析器包括一超链接分析。

10 19. 如权利要求 1 所述的装置，其特征在于，所述概念网络依赖于所述数据位置中已包含的结构。

20. 如权利要求 1 所述的装置，其特征在于，所述概念网络利用互信息。

21. 如权利要求 1 所述的装置，其特征在于，所述概念网络利用熵。

15 22. 如权利要求 1 所述的装置，其特征在于，所述概念网络利用不同单词对之间的相似性。

23. 如权利要求 1 所述的装置，其特征在于，所述数据存储位置包括万维网站。

24. 如权利要求 1 所述的装置，其特征在于，所述概念网络包括三种关系，即，后代、祖先和兄弟。

20 25. 一种装置，包括：

搜索多个数据存储位置的搜索工具，每个数据存储位置包括多个节点；以及概念网络，它基于所述多个节点的结构选择所述多个节点中的某几个的一部分。

25 26. 如权利要求 25 所述的装置，其特征在于，所述多个数据存储位置的结构是基于所述数据存储位置的统一资源定位器（URL）而导出的。

27. 如权利要求 25 所述的装置，其特征在于，所述概念网络基于 Web 合并选择所述多个节点中的某几个的一部分。

28. 如权利要求 25 所述的装置，其特征在于，所述概念网络显示一种包括多个 Web 块的结构。

30 29. 如权利要求 25 所述的装置，其特征在于，所述数据存储位置包括万维网页。

30. 如权利要求 25 所述的装置，其特征在于，所述概念网络包括三种关系，即，后代、祖先和兄弟。

31. 一种方法，包括：

接收对应于多个万维网页的多个统一资源定位器（URL）；

5 分析所述多个 URL 以确定所述多个 URL 的内容；以及

基于经分析的所述多个 URL，将多个 Web 块结构化为概念网络。

32. 如权利要求 31 所述的方法，其特征在于，还包括显示所述概念网络。

33. 如权利要求 32 所述的方法，其特征在于，还包括接收响应于所显示的概念网络的用户输入。

10 34. 如权利要求 31 所述的方法，其特征在于，还包括基于所述 URL 确定特定链接是向上链接、向下链接、兄弟链接、还是交叉链接。

35. 一种方法，包括：

考虑多个域专属万维网站；

通过分析所述多个域专属万维网站的相对内容导出域专属分类；以及

15 基于所述域专属分类规划概念网络。

36. 如权利要求 35 所述的方法，其特征在于，所述域专属分类是基于信息挖掘而导出的。

37. 如权利要求 36 所述的方法，其特征在于，所述信息挖掘是基于链接结构和内容。

20 38. 如权利要求 35 所述的方法，其特征在于，所述概念网络是基于熵来规划的。

39. 如权利要求 35 所述的方法，其特征在于，所述概念网络是基于互信息来规划的。

25 40. 如权利要求 35 所述的方法，其特征在于，所述概念网络是基于相似性来规划的。

41. 一种方法，包括：

生成概念网络，包括：

基于从用户提交的查询，分析关于多个万维网站的结构信息；

30 基于所述多个万维网站的结构信息，确定对“前 N 个”类型的查询的响应；以及

向用户返回涉及所确定的响应的信息。

42. 如权利要求 41 所述的方法，其特征在于，所述结构信息是基于统一资源定位器（URL）。

43. 如权利要求 41 所述的方法，其特征在于，所述结构信息是基于每个万维网页内的隐藏概念。

5 44. 一种方法，包括：

生成概念网络，包括：

基于从用户提交的查询，分析关于多个数据存储位置的结构信息；

基于所述多个数据存储位置的结构信息，确定对相关性类型的查询的响应；以及

10 向用户返回涉及所确定的响应的信息。

45. 如权利要求 44 所述的方法，其特征在于，所述结构信息是基于统一资源定位器（URL）。

46. 如权利要求 44 所述的方法，其特征在于，所述结构信息是基于每个数据存储位置内的隐藏信息。

15 47. 如权利要求 44 所述的方法，其特征在于，所述数据存储位置包括万维网页。

48. 一种具有用于生成概念网络的计算机可执行指令的计算机可读介质，包括：

基于从用户提交的查询，分析关于多个万维网站的结构信息；

20 基于所述多个万维网站的结构信息，确定对相关性类型的查询的响应；以及向用户返回涉及所确定的响应的信息。

49. 一种方法，包括：

通过以下动作自动导出域专属分类：

基于从用户提交的查询，分析关于多个数据存储位置的结构信息；

25 基于所述多个数据存储位置的结构信息，确定对相关性类型的查询的响应。

50. 如权利要求 49 所述的方法，其特征在于，还包括使用所述域专属分类生成概念网络。

概念网络

5 技术领域

本发明涉及搜索工具，尤其涉及所显示的搜索结果。

发明背景

随着诸如因特网等网络的迅速增长，搜索的准确性和质量变得越来越重要。

10 但是，许多用户发现使用搜索引擎的搜索产生大量（可能数以千计）的结果，其中许多并不严密地适用于他们所提交的查询。由此，许多用户对搜索结果不满意。一些用户还发现，大量返回的查询结果使因特网中包含的重要信息变得模糊不清。

许多现有技术的搜索引擎主要是基于关键词的比较。考虑询问世界上最顶尖的 N 个数码摄像机制造商的查询，其中 N 是整数。关键词比较搜索引擎将返回一些包含关键项“数字摄像机”的万维网页，以及其它包含关键项“制造商”的万维网页。因此，在关键词比较搜索引擎中返回的涉及数码摄像机制造商的总的返回结果的百分比相对很小。关键词比较搜索引擎还无法基于制造商的万维网页来比较一特定数码摄像机制造商是否比另一数码摄像机制造商更大或更有名（或其它某个可量化的比较）。因此，主要基于关键词比较的现有技术的搜索引擎常常导致大量结果，而其中许多与查询的相关度很低。此类关键词比较搜索引擎不能基于万维网站的结构来标识多个搜索所得的万维网站中最适用的那些。

在另一个方面，许多用户相信他们必须搜索大量查询以获得有用的搜索结果。因此，这些用户相信查询（以及就相关性对搜索结果所进行的检查）需要大量时间以确保所有有关响应都被考虑。即使在花费这么长时间以后，用户也常常相信最重要的搜索结果可能淹没在大量无关信息中。

在又一个方面，许多因特网应用程序使用诸如 WordNet®（在 George A. Miller 教授的指导下在普林斯顿大学开发）等词典工具来扩展用户的查询，以提高搜索引擎的精度。WordNet 是一种在线词典参考系统。使用 WordNet，名词、动词、形容词和副词被组织成同义词组，其中每一个都标识一个底层的词典概念。不同的关系 30 链接各同义词组。有了 WordNet，用户手动地输入其相对于万维网页的个人分类法。

因此，WordNet 并非被适当地配置成紧随因特网及其它联网计算机系统的迅速增长和动态改变。例如，Web 中出现的单词有一半以上没有出现在 WordNet 中。

发明概述

5 此发明涉及概念网络。概念网络可相应于用户查询而生成。在一个实施例中，概念网络配合搜索工具使用。搜索工具搜索多个数据存储位置。每一个数据存储位置都用一个节点来排列。这些节点中的某几个由至少一个链路连接。概念网络基于链路选择这某几个节点中的一部分，其中至少一个链路被用于内容目的。

10 附图简述

贯穿所有附图，相同的特征和组件引用相同的标号。

图 1 是一个包括概念网络的计算机环境的实施例的框图；

图 2 是开发概念网络的一个实施例的万维网站搜索部分的另一个实施例的框图；

15 图 3 示出用于建立概念网络的域的分类法构造的一个实施例的框图；

图 4 示出可在计算机环境的显示器上所显现的概念网络所显示的结果的透视图；

图 5 示出万维网站结构分析算法的一个实施例的流程图；以及

图 6 示出可用于推导或向用户显示概念网络的计算机环境的一个实施例的框图。

较佳实施例的详细描述

此发明提供概念网络的各种实施例。在概念网络中，查询等同于被搜索的概念。在概念网络中，多个搜索结果信息被结构化地组织为向用户显示的多个概念。25 如本文所揭示的概念网络根据搜索结果与查询的各个概念的相关性检索和/或显示搜索结果（搜索结果是基于概念来排列的）。概念网络可被配置成允许用户基于搜索结果与用户查询的相关性访问各个搜索结果、搜索结果的各个内容、或者搜索结果的各个部分。这一概念网络是基于查询在计算机环境中生成的。术语“概念网络”的一个方面涉及以概念可被用户理解和访问的方式将概念集合为概念网络。

30 概念网络的一个实施例特别针对于因特网，即使概念网络一般可被应用于任何计算机环境或计算机系统。在概念网络的因特网实施例中，用户可输入查询，而

所显示的输出概念网络可以是可由用户选择的概念列表。例如，如果用户查询“电子系统”，所显示的概念网络可包括诸如蜂窝电话、计算机、音频系统、视频系统等各种概念。由此用户可选择作为概念网络显示的这些概念中的一个，以显示更具体的搜索结果。

5 概念网络的一个实施例包括显示诸如相对于图 3 所描述等多个互连概念的大型连接图。如同来自现有技术的搜索引擎的搜索结果一样，概念网络是基于用户查询而得到的。与现有技术的搜索引擎相比，概念网络增加了对客户查询的相应的准确性。此外，概念网络限制大量无关的搜索结果，而这在依赖于关键词查询的现有技术的搜索引擎中间是普遍的。

10 图 1 示出被配置成生成并显示概念网络 100 的计算机环境 50 的实施例的框图。计算机环境 50 可包括可任选的网络部分 72（尽管该计算机也可以是独立计算机）。计算机环境 50 包括搜索工具 74 和显示工具 75。搜索工具 74 和显示工具 75 的部分包括概念网络 100 的一个实施例的部分。提供概念网络 100 作为用户界面，搜索结果的结构化表示通过该用户界面向用户显示，且搜索结果根据也可向用户示出以指示每个返回的搜索结果的相关性的概念被结构化或排列。现有技术的搜索引擎通常返回适用的万维网页的列表之类的结果，而如本发明中所揭示的所返回的概念网络包括，例如根据其内容结构化排列的多个万维网页。生成概念网络 100 的搜索依赖于被搜索的数据（例如，万维网页）内所包含的信息，如被搜索的数据的某个结构化特征所指示。由此，与现有技术的搜索引擎（基于关键词匹配）相比，概念网
15 络所返回的结果（基于所搜索的数据的结构）一般提供更准确的搜索结果。
20

包括概念网络 100 的图 1 的计算机环境 50 示出一般化的计算机环境。构想该概念网络高度适用于能够适用搜索引擎的任何类型的计算机环境，包括独立计算机、联网计算机、或大型计算机。但是，本发明被描述为适用于计算机环境的特定实施例。更特别地，构想概念网络 100 适用于网络。更特别地，构想包括概念网络 25 100 的计算机环境 50 的一个实施例包括排放在整个因特网上的若干万维网站上的各个万维网页。由此，概念网络的某些实施例依赖于服务器，它们在诸如因特网等联网计算机环境中向客户机提供构成搜索结果的数据。向用户显示的搜索结果的结构化表示的一个实施例是基于因特网用户一般已知的统一资源定位器（URL）。在此揭示中描述因特网的某些实施例中使用的 URL 的结构化方面。尽管此揭示描述 30 URL 为向概念网络中所依赖的数据提供结构，但是要强调的是，可包含能为概念网络所依赖的类似的结构化信息的任何其它机制都落入本发明的预期范围之内。

在计算机环境 50 的某些实施例中，提交查询的用户将导致有关的概念网络 100，其中所显示的结果被组织为一系列有关的概念。一般而言，基于各种用户查询可获得大量各种类型的搜索结果。由此，本揭示描述基于大量用户查询生成各种概念网络。概念网络的一个示例性例子是通过搜索“前 N 个”查询（寻求任何类别的前“N”个的查询，其中 N 是整数）产生的。概念网络的另一个示例性例子在诸如“解释电子”等复杂概念的搜索中产生有意义的结果。概念网络 100 是基于概念生成的（与基于用户查询所提供的实际关键词产生搜索结果的现有技术的搜索引擎不同）。概念网络 100 是响应于查询，以能向用户提供更详细和准确的信息的方式而生成的。例如，概念网络的某些实施例是考虑涉及查询所提出的概念的大量万维网站而生成的。概念网络考虑每个万维网站与查询所提供的概念的相关性，然后通过概念网络向用户示出各个万维网页的有关部分。

与依赖关键词查询结果的现有技术的搜索引擎相比，概念网络 100 能以更有组织和准确的方式显示来自各种查询的结果。包括了更为定向和准确的查询相应这一事实允许概念网络在向用户的显示中处理较少的万维网页。随即可根据概念网络 15 所生成的相对很少的万维网页的结构来组织这些万维网页。例如，涉及一类概念的万维网页可通过概念网络的一个部分来访问，而涉及另一类概念的万维网页可通过概念网络的另一个部分来访问。因此，概念网络 100 可包含涉及大量被检索的信息（诸如万维网站、万维网站内容信息、或万维网页的部分）的结构信息。

概念网络 100 提供大多数现有技术的搜索引擎所返回的万维网页列表上的若干改进。概念网络 100 的某些实施例改为提供并显示按（显示概念网络的）万维网页上的顺序排列的结构化信息。对用户而言，如图 4 所示的自动生成和显示的概念网络 100 是更易理解、解释和更为有用的形式。一般而言，与现有技术的搜索引擎相比，概念网络 100 提高搜索的精度和速度，并且提高在这些搜索期间所获信息的相关性。

一般而言，概念网络 100 显示基于数据的结构信息（例如，格式、节点之间的链接、等等）导出的相当大量的信息。在一个实施例中，此结构信息是基于统一资源定位器（URL）获得的，尽管可使用包含所检索的信息的结构信息的任何设备。在因特网中，URL 目前被用于导航目的，以允许浏览器通过因特网访问特定万维网页。URL 还可被用于提供此揭示中所描述的用于创建概念网络的结构信息（描述不同节点之间的关系）。此类结构信息的例子涉及，例如是另一个节点的祖先、后代、兄弟或其它关系的节点。此类结构信息被概念网络 100 的各种实施例用于结

构化地描述概念网络内部的不同节点之间的关系。

在概念网络 100 中此类结构信息被用于提供单词的分类或归类。概念网络的分类（如同现有技术的搜索引擎）涉及特定单词的含义。考虑到搜索引擎内已改变含义或者被添加或移除的大量单词，现有技术的手动搜索引擎难以维护现行的分类法。概念网络的某些实施例提供自动构造的分类法，基于在查询期间所访问的万维网站的结构，该分类法适应于域和用户。构想如本文所揭示的概念网络 100 可被应用于各种计算机系统，包括但不限于，数据库、在线购物、摄像机、个人计算机、手持式计算机、机器学习、以及计算机制造。

尽管此揭示描述概念网络 100 被应用于分析因特网上的万维网站，但是应当强调，这些概念适用于所有联网的、独立的、以及其它基于计算机的搜索引擎。由此，将概念网络应用于因特网或者任何其它网络或计算机系统都是在本发明的预期范围之内。

本揭示描述概念网络 100 及相关联组件的各种实施例。概念网络 100 被设计成自动对其自身进行更新，而无需代表用户进行任何更新。在查询之间，计算机环境的一个实施例以与关键词搜索高速缓存流行的搜索类似的方式（诸如通过使用万维网站爬寻器等）不断地进行搜索。概念网络 100 的一个实施例将爬寻所有涉及所收集的概念的万维网站以更新概念网络。构想此爬寻过程类似于常规搜索引擎所执行的那些过程。

在合理的时间量内，概念网络 100 能够基于用概念网络生成的分类法来理解大量典型用途的关键词（包括它们的结构）。使用此分类法，概念网络以结构化的方式显示关键词。由此，概念网络能够被用作分类词汇汇编，因为概念网络能够基于分类来解释单词的含义。因此，概念网络的分类（即，字典）中越来越多的单词对搜索诸如网络或 Web 等计算机环境以寻找特定技术、法律或其它此类专用单词的用户特别有用。

几乎所有专业都有相当大量的专用单词，其中许多随时间不断更新。例如，诸如代理人、税务专家、工程师等专业或群体基于它们特定的使用和专长的领域，每一个都有其自己的分类法。由于每个领域都只有相对很少数量的用户的缘故，手动搜索引擎不更新其中的许多项。概念网络可自动更新其中专用的、不常用的或被频繁更新的许多项。

用于得出概念网络 100 的万维网站搜索部分 201 的一个实施例在图 2 中描述。万维网站搜索部分 201 的实施例包括入口页和爬寻器规则部分 202、万维网站结构

分析器 204、万维网页摘要部分 206、万维网站结构合并工具 208、以及概念网络 100。万维网站结构分析器 204 包括超链接队列 212、万维网站爬寻器 214、HTML 语法分析器 216、基于功能的对象模型 (FOM) 分析器 218、以及超链接分析 220。

5 为生成概念网络 100，万维网站结构分析器 204 分析万维网站的结构。然后图 2 的 Web 合并工具（本文中也称为 Web 站点结构合并工具 208）合并来自不同的结构化万维网站的内容以产生可使用概念网络来显示的搜索结果。

在常规万维网站中，链接被用于导航。为分析万维网站内容结构以创建每一个概念网络 100，链接从用于导航被转换为用于内容。为完成此转换，执行以下步骤：

10 a) 在 URL 中编码每个万维网站的结构化信息。由此，在 URL 中编码特定链接，而无论其是向上链接、向下链接、兄弟链接还是交叉链接。这在现有技术的搜索引擎中是没有的。在一个实施例中，区别链接的类型是由万维网站爬寻器 214 通过考虑万维网站爬寻器的访问顺序来执行的。

15 b) 执行聚合和关联分析。此聚合和关联分析包括确定各集线器和不同授权机构的位置。在一个实施例中，这可由 FOM 分析器 218 执行。

c) 然后区分信息链接和导航链接。此识别是使用基于功能的对象模式 (FOM) 来分析导航条、导航列表或独立链接来执行的。由此，页的布局被用来对万维网页进行分块。在一个实施例中，c) 可使用 FOM 分析器 218 来执行。

20 现有技术的搜索引擎在一次一个的基础上提供对多个万维网站的访问，而概念网络 100 被构造成同时包含从各个万维网站获得的结构信息。可在概念网络 100 上以可被读者轻易理解的方式组织来自这各个万维网页的信息。更特别地，可在概念网络 100 中以呈现来自多个万维网页（通常基于 URL）的结构信息的可量化的方式显示来自多个万维网站的类似地结构化的信息。随即可用导致不同万维网页的主题之间的比较的方式呈现来自多个万维网页的此类结构信息。例如，涉及特定工业或主题的多个公司或集团很可能在它们的万维网页中以相似的结构包含相似类型的信息。概念网络提供一种媒介以显示来自不同万维网页的此相似信息；或者以允许从同一概念网络轻易访问不同万维网页的方式来呈现要向用户呈现的不同但相关的万维网页。

25 在某些实施例中，万维网站结构分析器 204 从入口页和爬寻器规则部分 202 接受万维网站的进入点 URL 和一些万维网站爬寻器规则作为输入。这些 URL 包含涉及特定万维网页的各种结构信息（例如，链接的端点、万维网页的类型、等等）。

常规搜索引擎不使用由 URL 提供的这一结构来导出涉及万维网页的结构信息。万维网站结构分析器 204 分析万维网站结构，并向万维网页分配深度信息。结果是，万维网站结构分析器 204 的一个实施例生成万维网站的分层结构图，其节点包括概念。由概念网络所导出的概念可用如本揭示中所描述的关键词来表征。万维网站结构分析器 204 导致使用结构化的万维网站。

万维网站分析器 204 的一个实施例是基于 BFS（广度优先搜索）算法。万维网站结构分析器 204 维护超链接队列 212。万维网站爬寻器 214 从超链接队列 212 取 URL，然后使用万维网站爬寻器 214 爬寻来自因特网的超文本标记语言(HTML)源代码，然后将 HTML 源代码转发给 HTML 语法分析器 216。超文本队列 212 是 10 包括未经分析的超链接的队列。在分析开始以前，万维网站结构分析器 204 附接进入点 URL。在分析期间，只有万维网站爬寻器 214 从超链接队列 212 提取 URL。只有超链接分析器 220 应用新的未经分析的超链接。

万维网站的进入点 URL 从入口页和爬寻器规则部分 202 进入万维网站结构分析器 204 的超链接队列 212。当万维网站结构分析器 204 开始分析时，万维网站爬寻器 214 从超链接队列 212 提取 URL，然后万维网站爬寻器爬寻来自因特网的 HTML 源代码，并将 HTML 源代码转发给 HTML 语法分析器 216。HTML 语法分析器处理从因特网爬寻的 HTML 源代码。

HTML 语法分析器 216 接受从万维网站爬寻器 214 输入的 HTML 源代码。在一个实施例中，HTML 语法分析器 216 的活动包括 URL 提取、URL 统一、和 URL 分组。对于 URL 提取，HTML 语法分析器 216 根据输入的万维网站定义取所有指向一万维网页并且在该万维网页内部的 URL。用锚文本附接每一个 URL。对于图像链接，锚是包围文本。

对于 URL 统一，HTML 语法分析器 216 的一个实施例执行多个操作，包括：a)将相对 URL 地址转换为直接 URL 地址；b)将 IP 地址改为域名；以及 c)通过用最终目标 URL 地址替换该 URL 来解决重定向的 URL 的问题。对于 URL 分组，具有相同标签元数和相同外观的表格或列表中的超链接很可能被视为例如相关的节点。来自 HTML 语法分析器 216 的结果随即被转发给基于功能的对象模型 (FOM) 分析器 218。

基于功能的对象模型 (FOM) 分析器 218 使用 FOM 的基本概念和算法对超链接分配功能信息。此功能信息对分析每个万维网站的结构是非常有用的。FOM 表示万维网页的基于功能的对象模型。FOM 分析器 218 不是进行语义分析，而是试

图通过表示每个对象功能和类别来理解作者的意图。每个万维网页都可起到索引页或内容页的作用。一类导航对象是导航条。FOM 分析器 218 的一个实施例执行如以下 FOM 分析任务等索引/内容页识别和导航条检测。

对于索引/内容页识别，FOM 分析器 218 的一个实施例确定万维网页 URL 是 5 是否包括文本“Index”（索引）或“default”（默认），以及该 URL 是目录还是索引页。如果该页内部有对应于子目录的链接，则此链接是指向索引页。将超链接和内容单词的比值和阈值相比较。如果该比值大于阈值，则该万维网页是索引页。如果阈值大于该比值，则该万维网页是内容页。

FOM 分析器 218 的一个实施例提供导航条检测。导航条中的项是彼此互连的，
10 并且对应的链接拓扑结构是完全连通图。FOM 分析器 218 的输出包括多个超链接，它们被转发给超链接分析器 220。FOM 分析器 218 为万维网页提供分块。在一个实施例中，分块以后，基于其功能，万维网页被分成若干小单元，诸如内容块、导航块、广告块、等等。这些小单元可由用户单独访问。

超链接分析器 220 的一个实施例使用万维网站结构分析算法来处理 FOM 分析
15 器 218 所分析（并从该处发来）的每一个超链接。经语法分析的源代码被转发到 FOM 分析器 218 以执行功能分析。超链接分析器 220 根据万维网站结构分析规则分析每一个超链接，而新的未经分析的超链接被插入到超链接队列 212 中。超链接分析器 220 对每一个万维网页分配一个深度值（并维护万维网站的临时分层结构图）。该深度值可由万维网站爬寻器 214 输出。在一个实施例中，万维网站爬寻器
20 214 通过广度优先搜索访问万维网站。行进路径将被构成为树的格式，树的节点是万维网页，节点内部的链接是万维网页内部的超链接。因此树中节点的深度是我们想要获得的值。例如，进入点万维网页的深度（诸如由 URL <http://www.microsoft.com> 所标识的进入点页）是 0。相比之下，由 URL <http://www.microsoft.com/china> 所标识的万维网页的深度是 1。

25 万维网站结构分析器 204 构造了可被视为在超链接队列 212 处开始和结束的循环。万维网站爬寻器 214 从超链接队列 212 提取下一个 URI 以开始下一个循环。执行此动作直至超链接队列 212 中没有新的 URL 为止。分析过程完成，万维网站（称为结构化的万维网站）的分层结构图被构造出来。

30 每个万维网站的结构化信息是以可使用超链接分析器 220 检测的方式编码到 URL 中的。由此，无论特定链接是向上链接、向下链接、兄弟链接或交叉链接，它都被编码到 URL 中（并可使用超链接分析器 220 检测）。在一个实施例中，使

用基于 URL 块长度的启发式规则来检测向上链接和向下链接。URL 块长度被定义为块数，块是 URL 中被“/”或“?”分开的部分。例如，URL “http://www.sonystyle.com/digital/digital_camera.htm”的 URL 块长度是 3，包括“<http://www.sonystyle.com>”、“[digital](#)”和“[digital camera](#)”。在一个实施例中，

5 应用约束的规则来分析 URL。然后，对于未被规则覆盖的其余 URL，使用以上策略来进行分析。超链接检测规则的一个实施例根据两个规则来描述。第一规则是，如果 URL 块长度（超链接）小于或等于万维网页的 URL 块长度，则该超链接是向上链接。第二规则是，如果 URL 块长度（超链接）减去 URL 块长度（万维网页的 URL）大于或等于 2，则该超链接是前向链接。

10 假定当前万维网页节点为 B，它有到万维网页 C 的超链接。万维网站结构分析器 204 的超链接分析器部分 220 进行以下过程：

- I. 如果该超链接是向上链接，则将其丢弃（不再考虑）。
- II. 如果 B 和 C 属于导航条，则 B 和 C 是兄弟节点（如本文所讨论）。
- III. 如果 C 已被访问，且 B 的 URL 块长度大于或等于 C：

15 如果 B 是索引页；则 C 是 B 的子节点（如本文所讨论）；
 否则如果 B 是内容页，则 C 是 B 的兄弟节点。

- IV. 如果 C 未被访问，
 如果 B 是内容页，则 C 是 B 的兄弟节点，
 否则 C 是 B 的子节点。

20 否则如果 C 尚未被访问，则

 首先如果 B 是内容页或在若干页中被显示，则该链接是显式关联。
 否则此链接是聚合。

在分析超文本队列中的 URL 以后，使用万维网页摘要部分 206 得出万维网站
25 结构。例如，万维网页中所包含的某个量的数据可能与特定用户的查询有关，而其它数据则无关。万维网页摘要以可在概念网络 100 内的特定概念部分上显示的形式提供有关的信息。因为不在概念网络上示出每一个万维网页的整体，所以概念网络可提供可由用户访问的每个概念或万维网页的信息的更为定向的摘要。随即使用万维网站结构合并工具 208 将从万维网页摘要部分 206 导出的不同万维网页的各种各样的内容（或其它内容）合并到概念网络 100 中。万维网站结构用分层结构图表示。

概念网络 100 的某些实施例分析相关万维网站的结构，并据此将结果合并到

一起。在此揭示中，将来自多个万维网站的信息合并被称为 Web 合并，如图 2 所示由万维网站结构合并工具 208 执行。万维网站结构合并工具 208 所执行的 Web 合并提高概念网络的精度和速度，并被执行如下。

在每个万维网站都被结构化为“树状图”或“深度图”以后，接下来的问题是将这些图合并成一个网络。在该网络中，每个节点表示一个概念，而这些节点之间的链接表示这些概念之间的关系。基本关系可包括，但不限于，超义词、假义词、同义词、等等。因为每个万维网站都表示原始编辑者对有关主题的观点，所以将不同的观点合并为一个观点有点困难。因此在以下，我们给出将来自各种类型的资源的概念分层结构合并成一个可用的分层结构的解决方案。

为示出如何合并概念网络的分层结构的一个实施例，从两个不同的分层结构 H 合并给定概念 C 的一类关系 R 。解决此问题的详细算法如下：

以下技术表示可用于执行本体合并过程的一个实施例：

a) 对于每个 Web 块，使用如图 2 所示的万维网页摘要部分 206 为万维网页简述概念。这些概念被解释为关键词集合。

b) 然后对这些概念进行令牌化，通过此方法，要通过概念网络 100 来生成和显示的每个概念都由“令牌”短语或关键词表示。由此，建立关键词集合以表示和描述概念网络中所包含的概念。(1)被用于最终生成概念：

$$n_i = [w_{i1}, w_{i2}, \dots, w_{im}] \quad (1)$$

其中 $w_{i1}, w_{i2}, \dots, w_{im}$ 表示单词，而 n_i 表示单词数组， n_i 是概念网络中一个节点（万维网页）的摘要，它可被分解为若干单词/短语，即 $w_{i1}, w_{i2}, \dots, w_{im}$ 。

c) 在分层结构树上提供滑动窗口，以分别用(2)、(3)和(4)来生成后代 (offspring)、祖先 (ancestor) 和兄弟 (sibling) 的子树 ST。假定一些单词在不同的窗口中出现。

$$ST_i(offspring) = (n_i, sons_1(n_i), \dots, sons_d(n_i)) \quad (2)$$

$$ST_i(ancestor) = (n_i, parents_1(n_i), \dots, parents_d(n_i)) \quad (3)$$

$$ST_i(sibling) = (n_i, sibs_1(n_i), \dots, sibs_d(n_i)) \quad (4)$$

其中， $ST_i(offspring)$ 、 $ST_i(ancestor)$ 和 $ST_i(sibling)$ 是用于计算后代、祖先和兄弟关系的子树； $sons_d$ 、 $parents_d$ 和 $sibs_d$ 分别代表节点 n_i 的第 d 层子节点、父节点和兄弟节点。

d) 对于每一个所生成的子树（例如， $ST_i(ancestor)$ ），按等式(5)计算术语对的互信息。计算每个单词对 w_i, w_j 的互信息 MI 。具有高值的互信息指示该单词对

是相似的。

$$MI(w_i, w_j) = P_r(w_i, w_j) \log \frac{P_r(w_i, w_j)}{P_r(w_i)P_r(w_j)} \quad (5)$$

$$P_r(w_i, w_j) = \frac{C(w_i, w_j)}{\sum_k \sum_l C(w_k, w_l)} \quad (6)$$

$$P_r(w_i) = \frac{C(w_i)}{\sum_k C(w_k)} \quad (7)$$

$$P_r(w_j) = \frac{C(w_j)}{\sum_k C(w_k)} \quad (8)$$

其中, $MI(w_i, w_j)$ 是项 w_i 和 w_j 的互信息; $P_r(w_i, w_j)$ 表示项 w_i 和 w_j 一起在子树中出现的概率; $P_r(x)$ (x 可为 w_i 或 w_j) 表示项 x 在子树中出现的概率。

确定一对项的相关性的另一个因素是项对的分布。越多子树包含该项对, 则这两个项就越相似。在我们的实现中, 使用熵 (entropy) 来测量项对的分布, 如步骤(d)中所示。

d) 计算每个单词对 w_i, w_j 的熵。基于在(5)中确定的互信息实际上在所有万维网站中是共同的, 熵转换是对被确定为共同的单词对 w_i, w_j 的度量, 。熵越高, 概念网络可向用户提供的该单词在所有万维网站之间成对的信心就越高。

$$\text{entropy}(w_i, w_j) = -\sum_{k=1}^N P_r(w_i, w_j) \log P_r(w_i, w_j) \quad (9)$$

$$P_r(w_i, w_j) = \frac{C(w_i, w_j | ST_k)}{\sum_{l=1}^N C(w_i, w_j | ST_l)} \quad (10)$$

e) 根据(11)计算每个单词对的相似性 Sim :

$$Sim(w_i, w_j) = MI(w_i, w_j) \times \frac{\text{entropy}(w_i, w_j) + 1}{\alpha \log(N)} \quad (11)$$

如(11)中所阐述的相似性将互信息 $MI(w_i, w_j)$ 和 $\text{entropy}(w_i, w_j)$ 相结合。

为指示涉及(2)、(3)和(4)的有关概念 (后代、祖先和兄弟), 概念网络生成各种有关的类别。例如, 表 1 示出公知概念的各种示例性后代概念:

表 1: 后代概念

原始类别	后代
软件	实用程序、游戏、商务、一般、图形、数据库
视频	DVD、TV、投影、摄放像机
小说	故事、戏剧、恐怖、诗、科学、传奇
服装	衣服、女装、衬衫、鞋、饰品、男装、运动装、戏服、童装
鞋	靴子、高跟鞋、凉鞋、拖鞋、休闲鞋
宠物	照料、供应、鸟、猫、狗、鱼、食物、服务

表 2 示出各种示例性的祖先概念:

表 2: 祖先概念

原始类别	后代
软件	计算机
视频	电子、组件
小说	书、文学
服装	无
鞋	女装、男装、服装
宠物	无

表 3 示出各种示例性兄弟概念:

表 3: 兄弟概念

原始类别	后代
软件	硬件、网络、外观、存储、外设、存储器
视频	音频、摄影、摄像机、附件
小说	烹饪、历史、运动、旅游、作者、连环画
服装	时尚、软件、美丽、音乐、宠物
鞋	衣服、表、外套、衬衫、泳装、裤子
宠物	礼物、运动、玩具、珠宝、书

5 图 2 中所示的概念网络 100 的一个实施例被规定为图 3 中的结构形式、以及
图 4 中其向用户所呈现的形式中所示出的有向图。概念网络所基于的有向图(G) 300
由(12)描述:

$$G = (V, E) \quad (12)$$

其中 V 是节点的集合，而 E 是边或链接的集合。由此，如有向图所表示的概念网
络 100 包括多个节点以及链接这些节点的多个链接或边。节点表示概念。边或链接
表示概念之间的关系。如图 4 中所示的概念网络 100 的有向图 300 由此提供内容结
构。万维网页的内容结构被信息挖掘以产生用于生成概念网络的信息。

图 3 示出用于使用概念网络 100 构造特定域的分类的技术的一个实施例。图 3
以导出一个或多个域专属的万维网站 302 而开始。这可通过利用现有的元搜索引擎
来进行此工作来完成。例如，如果用户想要构造“数字摄像机”域的概念网络，用
户可向搜索引擎发送查询，并使用前 100 个万维网站来构造概念网络。每个域专属
的万维网站 302 包括对应于内容（由节点表示）的分析和链接结构（由链接结构表
示）的分析的结构。

产生概念网络 100 依赖于一个或多个万维网站的内容结构的有效挖掘。此挖
掘可通过分析链接类型来执行，链接类型确定链接是后代链接、祖先链接、还使兄
弟链接，诸如相对于图 2 的超链接分析器 220 所述。这些链接类型中的一个被分配
给每个链接。然后使用如图 2 中所示的万维网页摘要部分 206 来简述节点的语义。

在图 3 中，域专属分类是基于此信息挖掘而导出的。注意，与诸如 WordNet® 等要求手动的编辑输入以进行分类的现有技术的工具相比，在本揭示中，域专属分类的导出是自动执行的。Wordnet 是针对一般域的手动构造分类。此分类是由编辑者而不是端点用户构造的。信息挖掘依赖于域专属万维网站的链接结构和内容。这与某些现有技术的自动分类词汇汇编构造不同，在后者中，信息是从内容而不是链接结构挖掘的。

然后使用本体学习来构造概念网络 100。基于本体学习，自动构造的概念网络开发其自身的分类。本体学习是基于统计框架，并且能够产生多个编辑者的视图。统计框架容易被应用于许多统计应用程序。如图 3 中所示地构造的概念网络 100 描述电子的各种概念网络。概念网络 100 包括各种 Web 块 450，其中每个 Web 块表示一个不同的电子类别（例如，电子产品、电子类别、以及电子设备制造商、等等）。

每个 Web 块由用户可识别的关键词描述。每个子 Web 块 454 可被视为与主 Web 块相关。例如，在图 3 中，单词“电子”表示主 Web 块 452。术语“电子”表示好的主 Web 块 452，因为此术语出现在许多涉及各种产品（各种产品中的每一种都可被视为一个子 Web 块）的万维网站中。例如，在图 3 中，在电子主 Web 块下示出各种子 Web 块 454（包括照像机和相片、音频和视频、手持式、蜂窝电话、计算机、Sony®、iPAQ®、Palm®、外设、以及各种 Compaq® 产品）。在此揭示中，每个 Web 块都被视为包含同构信息的一个概念。因此术语“概念网络”描述多个概念，或 Web 块的网络。

每个 Web 块都可由一个关键词（诸如照相机、计算机、以及“Sony”，如图 3 中所示）简述。图 3 中每个子 Web 块的主题极大地涉及主 Web 块电子，因此可被宽泛地归类到概念“电子”之下。基于 Web 块的结构、挖掘、以及概念网络 100 的域专属分类，如图 3 中所示的电子的概念网络包含这些术语中的许多。图 3 中所示的所生成的概念网络 100 可被视为自动构造的最终结果。

图 4 中示出示例性概念网络 100 的一个实施例，如其在诸如平面显示器或 CRT 监视器等计算机显示器 200 上所显现的。由此，图 4 示出所生成的包括各种概念 402 的概念网络 100（使用图 2 和 3 中所示出的技术）。每个概念 402 包括涉及以此方式所生成的至少一个 Web 块 450，其中某些实施例在图 3 中示出。因此图 4 中示出的概念网络 100 包含覆盖在显示器上的若干概念 402。就所关注的领域（在此实例中为“电子”）相对地详述了概念网络的细节。例如，这些概念中的几个如

果被用户选择，则会将用户带到比当前所显示的概念网络更狭隘或更宽泛的另一个概念网络。例如，用户可从电子概念网络转移到计算机概念网络。

通过搜索所有各种万维网站执行概念网络的分析。分析指出在概念网络的某些实现中，被正确定位的万维网站的百分比有所提高（增至 75%）。就准确性而言，这表示现有技术上的显著进步。

考虑示例性查询“数字摄像机制造商”。典型的现有技术的搜索引擎搜索整个 Web 并返回包含关键项“数字”和/或“摄像机”和/或“制造商”的万维网页。因而此类现有技术的搜索引擎将返回相当大量的无关万维网页。

概念网络 100 仅需搜索从节点“数字摄像机”扩展的子图。因此，概念网络速度更快，并且所返回的无关万维网页的数量显著减少。

概念网络 100 提高对查询的期望响应的简单程度、速度和可靠性。首先，在概念网络 100 中定位项“数字摄像机”。所有指向节点“数字摄像机”或其所指向的节点都被提取。然后属性为“制造商”的节点被选择并分级（例如，基于点击数）。如此，便可搜索对任何类别的万维网页的前 N 个的查询（最大的公司、最大的制造商、最多的办公室、最近的位置、等等），而捕获合理数量的准确点击的概率大大提高。

由概念网络所表征的此类改良的搜索得以发生是因为查询是定向到被搜索的万维网站的结构（如 URL 内所包含）。如相对于图 5 所描述的概念网络 100 的某些实施例可提供各种搜索服务，它们可搜索如因特网上的万维网站上所列出的前 N 个（其中“N”是某个正数）组织、公司、项目、群组、产品等等的某个可量化的阐述。例如，某些实施例生成提供找出世界上前五个数字摄像机制造商的查询的搜索结果的概念网络 100。另一个查询提供诸如指示欧洲前五个钢铁制造公司等另一个复杂查询的搜索结果。预期概念网络高度有益的一类查询依赖于基于万维网站的结构（例如，基于 URL 所提供的结构）来访问数据。“前 N”类型的查询基于多个万维网站的结构来分析和返回信息。例如，一种确定谁是美国前三名的汽车生产商的技术涉及访问所有可能的汽车生产商的万维网站，从每个万维网站导出类似的生产信息，然后比较从不同的万维网站导出的生产信息。由此，概念网络 100 的某些实施例可搜索万维网页内的详细特征。

数据挖掘针对于此类万维网站分析。一般而言，数据挖掘（有时称为数据或知识发现）是基于查询从不同角度分析数据并向用户将数据总结为有用信息的过程。数据挖掘软件是用于分析数据的许多分析工具中的一种。它允许用户从许多不

同的方面或角度分析数据、归类数据、以及总结所标识出来的关系。技术上，数据挖掘是在大型关系型数据库中的几十个字段之间寻找相关性或模式的过程，并且通常在查询中是公知的。由此，概念网络的某些实施例可使用如图 3 所提供的数据挖掘 306 来导出域专属的分类 304。

5 图 5 示出导致生成概念网络的过程 600 的一个实施例。过程 600 包括用户将查询输入到计算机环境 50 (如图 1 中所示) 中的 602。该查询将导致生成并向用户显示概念网络。在 604，该查询被提交给如相对于图 3 所描述的多个域专属的万维网站 302。这些万维网站由常见的元搜索引擎或人类构建的 Web 分层结构返回。在 606，计算机环境通过诸如考虑对相关联万维网站的 URL 结构和内容等来分析 10 万维网站结构。在 610，被挖掘的信息被用于生成域专属分类 (如相对于图 3 中的 304 所描述)。过程 600 前进至 612，生成并向用户显示概念网络 100。

概念网络 100 能够被生成以返回对诸如“解释单词：电子”等查询的响应 (这是现有技术的搜索引擎所不能执行的)。此类概念网络也是通过分析各种万维网站和万维网页的结构来生成的 (如图 5 中的查询的情形)。概念网络的一个实施例保 15 存万维网站的结构信息，此信息标识编辑者对概念的分层结构的观点。在概念网络 100 中，不同编辑者的观点被合并到一起，从而用户可确定什么是最常见的解释。

概念网络 100 的某些其它实施例可提供确定用户执行诸如解释单词“电子”等任务的最佳万维网站的查询。此类查询可被视为解释和/或比较的查询。由此，必须由概念网络对若千万维网站进行评估和比较。生成此类概念网络 (诸如能够解 20 释复杂问题的概念网络等) 所涉及的一种机制牵涉到考虑涉及查询所提出的问题的大量万维网站；在某种程度上如现有技术的搜索引擎所可测量地考虑每个万维网站的相关性，然后向概念网络的用户显示万维网页的有关部分。图 5 的过程 600 的实施例也可被用于执行此类查询。

为响应这些类型的相对复杂的查询 (前 N 类型的查询，或者必须评估和比较 25 多个万维网站的查询，等等)，通过评估所考虑的每个万维网页或万维网站的结构来构造概念网络 100。现有技术的搜索引擎不能够从万维网站导出该结构以执行这些分析 (因此不能响应于此类查询)。例如，相对于电子的例子，概念网络考虑被结构化的万维网页，以提供定位到准确描述电子主题处足够的信息。

概念网络 100 在查询扩展中也非常有用。目前，许多因特网应用程序使用现 30 有技术的 WordNet 来扩展用户的查询以提高现有搜索引擎的精度。但是，WordNet 是手动构造 (分类词汇汇编) 的高强度劳动的工作。几乎没有哪个万维网站愿意手

动构造分类词汇汇编。万维网站操作者更喜欢自动分类词汇汇编构造。用户的手动分类词汇汇编构造不适合于因特网的迅速增长。诸如因特网等网络中的文档数量不断增长。越来越多的新单词和概念不断出现，这使得如在本揭示中所描述的概念网络的有用性得以突出。与依赖于关键词比较的现有技术的搜索引擎相比，概念网络
5 返回较少的、但更为定向的结果。因此，用户较易评估概念网络所返回的每一个结果。此外，用户较易评估查询是否未能返回期望类型的结果，从而用户将能够修改原始查询使之更为定向。

现用分类词汇汇编（概念网络可发挥此作用）对因特网和其它网络搜索是有用的。此外，概念网络 100 不但包含概念的分层结构，而且还包含这些概念的统计
10 信息。因此它可被轻易地应用于诸如调查等某些关于普及性的特定问题。

因为概念网络 100 的一个实施例为因特网和其它网络环境合并来自所有作者的单词和概念的观点，所以概念网络 100 可被视为向网络用户提供替换的分类词汇汇编。概念网络 100 可被适应于客户方以作为个人分类词汇汇编。用户的浏览路径将生成 Web 的子空间。类似的方法可被应用来分析 Web 的子空间，以生成个人频
15 繁使用的概念的关系。

因此概念网络提供万维网页的摘要。超链接和页标题上的文本可被用作万维网页的摘要。在另一个实施例中，自然语言语法分析（NLP）技术可被集成到万维网站搜索部分 201 中（可作为 HTML 语法分析器 216 的一部分），以使用某些主导关键词来总结文档。

20 此揭示描述各种概念网络 100。概念网络可被视为通过分析多个万维网站的结构并合并分析结果而从万维网站构建的因特网概念网络。概念网络 100 在提高搜索引擎的精度和速度方面特别有用。概念网络从万维网站结构、而不是单纯地从万维网站内所包含的明文文本中提取知识。概念网络提供域的自动构造。来自概念网络的统计结果揭示各个万维网站中所包含的一般知识。

25 由此，概念网络不仅从特定万维网站获得信息，而且还将从网络上大量各种各样的万维网站获得知识。概念网络可使用本体学习来维护涉及万维网站的结构信息。因此，当新的万维网页和概念被应用于因特网时，本体允许来自万维网页的结构信息被自动集成到概念网络中。此外，概念网络 100 可提供普通搜索引擎所不能提供的一些服务，诸如“找出世界上前 N 个数字摄像机制造商”以及“解释单词：
30 电子”等。概念网络还可对查询扩展起到现用因特网分类词汇汇编的作用，因为它提供如此多样的子 Web 块，它们通过主 Web 块相互关联，如图 3 中所示。

图 6 示出合适的计算机环境或网络 500 的示例，其中包括可生成概念网络的用户界面。计算机环境 500 表示图 1 中所示的计算机环境的一个实施例。类似的资源可使用本文中所描述的计算机环境和过程。

图 6 中所示出的计算机环境 500 是一般计算机环境，它可用于实现本文中所 5 描述的概念网络技术。计算机环境 500 只是计算机环境的一个例子，它并不试图对计算机和网络体系结构的使用范围或功能提出任何限制。也不应将计算机环境 100 解释为具有涉及在示例性计算机环境中所示出的任一组件或其组合的任何依赖性或要求。

计算机环境 100 包括计算机 502 形式的通用计算设备。计算机 502 可包括， 10 例如以下组中的一个或多个，该组包括独立计算机、联网计算机、大型计算机、PDA、电话、微型计算机或微处理器、或与存储器结合使用处理器的任何其它计算机设备。计算机 502 的组件可包括，但不限于，一个或多个处理器或处理单元 504（可选地包括加密处理器或协处理器）、系统存储器 506、以及耦合包括处理器 504 和系统存储器 506 在内的各个系统组件的系统总线 508。

15 系统总线 508 表示若干类型的总线结构中的一种或数种，包括存储器总线或存储器控制器、外围总线、加速图形端口、以及使用各种总线体系结构中的任何一种的处理器或局部总线。作为示例，此类体系统结构可包括工业标准体系结构 (ISA) 总线、微通道体系结构 (MCA) 总线、增强型 ISA (EISA) 总线、视频电子标准协会 (VESA) 局部总线、以及也称为 Mezzanine 总线的外围组件互连 (PCI) 总 20 线。

计算机 502 通常包括各种计算机可读介质。这些介质可以是可由计算机 502 访问的任何可用介质，并包括易失性和非易失性、可移动和不可移动介质。

25 系统存储器 506 包括诸如只读存储器 (ROM) 512 等非易失性存储器和/或诸如随机存取存储器 (RAM) 510 等易失性存储器形式的计算机可读介质。包含诸如在启动期间帮助在计算机 502 内部各元件之间传递信息的基本例程的基本输入/输出系统 (BIOS) 514 存储在 ROM 512 中。RAM 510 通常包含可由处理单元 504 即时访问和/或正由其操作的数据和/或程序模块。

计算机 502 还可包括其它可移动/不可移动、易失性/非易失性计算机存储介质。 30 作为示例，图 6 示出用于读和写不可移动、非易失性磁介质 (未示出) 的硬盘驱动器 515，用于读和写可移动、非易失性磁盘 520 (例如，“软盘”) 的磁盘驱动器 518，以及用于读和/或写诸如 CD-ROM、DVD-ROM 或其它光介质等可移动、非

易失性光盘 524 的光盘驱动器 522。硬盘驱动器 515、磁盘驱动器 518 和光盘驱动器 522 每一个都由一个或多个数据介质接口 527 连接到系统总线 508。或者，硬盘驱动器 515、磁盘驱动器 518、以及光盘驱动器 522 可由一个或多个接口（未示出）连接到系统总线 508。

5 各盘驱动器及其相关联的计算机可读介质为计算机 502 提供计算机可读指令、控制节点数据结构、程序模块以及其它数据的非易失性存储。尽管该例示出硬盘驱动器 515 内的硬盘、可移动磁盘 520 和非易失性光盘 524，应当认识到，能存储可由计算机访问的数据的其它类型的计算机可读介质也可被用于实现示例性计算机环境 500，诸如磁带盒或其它磁存储设备，闪存卡、CD-ROM、数字多功能盘（DVD）
10 或其它光存储，随机存取存储器（RAM）、只读存储器（ROM）、电可擦除可编程只读存储器（EEPROM），等等。

可在硬盘驱动器 515 内所包含的硬盘、磁盘 520、非易失性光盘 524、ROM 512、和/或 RAM 510 上存储任何数量的程序模块，例如包括，OS 526、一个或多个应用程序 528、其它程序模块 530、以及程序数据 532。OS 526、一个或多个应用程序
15 528、其它程序模块 530、以及程序数据 532 中的每一个（或其组合）可实现支持分布式文件系统的所有或部分常驻组件。

用户可经由诸如键盘 534 和定位设备 536（例如，“鼠标”）等输入设备将命令和信息输入到计算机 502 中。其它输入设备 538（未具体示出）可包括话筒、操纵杆、游戏垫、圆盘式卫星天线、串行端口、扫描仪和/其它。这些及其它输入
20 设备经由耦合到系统总线 508 的输入/输出接口 540 连接到处理单元 504，但也可由诸如并行端口、游戏端口或通用串行总线（USB）等其它接口和总线结构连接。

监视器、平面显示器、或其它类型的计算机显示器 200 可经由诸如视频适配器 544 等接口连接到系统总线 508。除了计算机显示器 200 以外，其它输出外围设备可包括诸如扬声器（未示出）和打印机 546 等组件，它们可经由输入/输出接口
25 540 连接到计算机 502。

计算机 502 可使用到诸如远程计算机 548 等一个或多个远程计算机的逻辑连接在联网环境中工作。例如，远程计算机设备 548 可以是个人计算机、便携式计算机、服务器、路由器、网络计算机、对等设备、或其它普通网络节点、游戏控制台、等等。图示远程计算机设备 548 为可包括在本文中相对于计算机 502 所描述的许多
30 或所有元件的便携式计算机。

计算机 502 和远程计算设备 548 之间的逻辑连接被示为局域网（LAN）550

和一般广域网 (WAN) 552。此类网络环境常见于办公室、企业范围的计算机网络、内联网和因特网。

当在 LAN 网络环境中实现时，计算机 502 经由网络接口或适配器 554 连接到局域网 550。当在 WAN 网络环境中实现时，计算机 502 通常包括调制解调器 556 或用于通过广域网 552 建立通信的其它装置。可以是内置或外置于计算机 502 的调制解调器 556 可经由输入/输出接口 540 或其它适当机制连接到系统总线 508。应当认识到，所示网络连接是示例性的，并且可以使用在计算机 502 和 548 之间建立通信链路的其它装置。

在诸如以计算机环境 500 所示的联网环境中，相对于计算机 502 所描述的程序模块或其部分可存储在远程记忆存储设备中。作为示例，远程应用程序 558 驻留在远程计算机 548 的记忆设备上。为说明的目的，应用程序及诸如操作系统等其它可执行程序组件在本文中被示为离散的 Web 块，尽管可以认识到，这些程序和组件在各个时间驻留在计算机 502 的不同存储组件中，并由计算机 502 的数据处理器执行。可以认识到，所示出并描述的网络连接是示例性的，并且可以使用建立计算机之间的通信链路的其它装置。

本文中在一个或多个计算机或其它设备所执行的诸如程序模块等计算机可执行指令的通用上下文中描述各种模块和技术。一般而言，程序模块包括执行特定任务或实现特定抽象数据类型的例程、程序、控制对象 650、组件、控制节点数据结构 654、等等。通常，在各实施例中可按需组合或分布各程序模块的功能。

这些模块和技术的一个实现可被存储在某种形式的计算机可读介质上，或可通过其发送。计算机可读介质可以是可由计算机访问的任何可用介质。作为示例，而非限制，计算机可读介质可包括“计算机存储介质”和“通信介质”、

“计算机存储介质”包括以用于存储诸如计算机可读指令、控制节点数据结构、程序模块或其它数据等信息的任何过程或技术实现的易失性和非易失性、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、数字多功能盘 (DVD) 或其它光存储，磁带盒、磁带、磁盘存储或其它磁存储设备，或可用于存储所需信息并可由计算机访问的任何其它介质。

“通信介质”通常具体化为诸如载波或其它传输机制等已调制数据信号中的计算机可读指令、控制节点数据结构、程序模块或其它数据。通信介质还包括任何信息传递介质。术语“已调制数据信号”指以在信号中以将信息编码的方式设置或

改变其一个或多个特征的信号。作为示例，而非限制，通信介质包括诸如线网络或直接连线连接等有线介质，以及诸如声学、RF、红外和其它无线介质等无线介质。以上任何组合也被包括在计算机可读介质的范围之内。

尽管使用专属于结构和功能特征和/或方法的语言描述了系统、介质、方法、
5 方式、过程等，但是应当理解，所附权利要求书中所定义的本发明不必被限制于所
描述的具体特征或方法。相反，揭示这些具体特征和方法是将其作为实现要求保护
的发明的示例性形式。

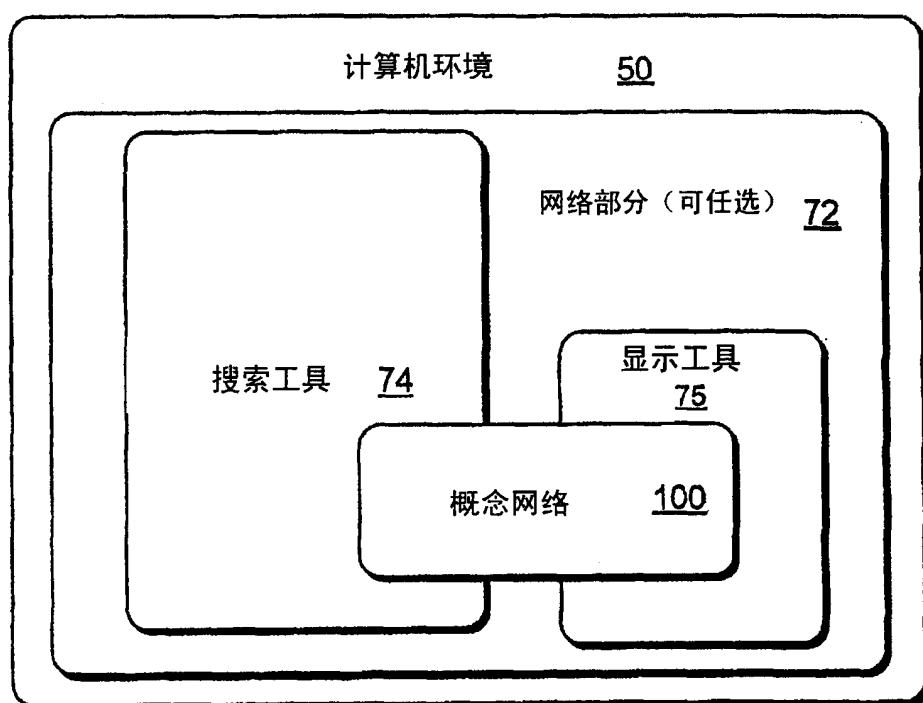


图 1

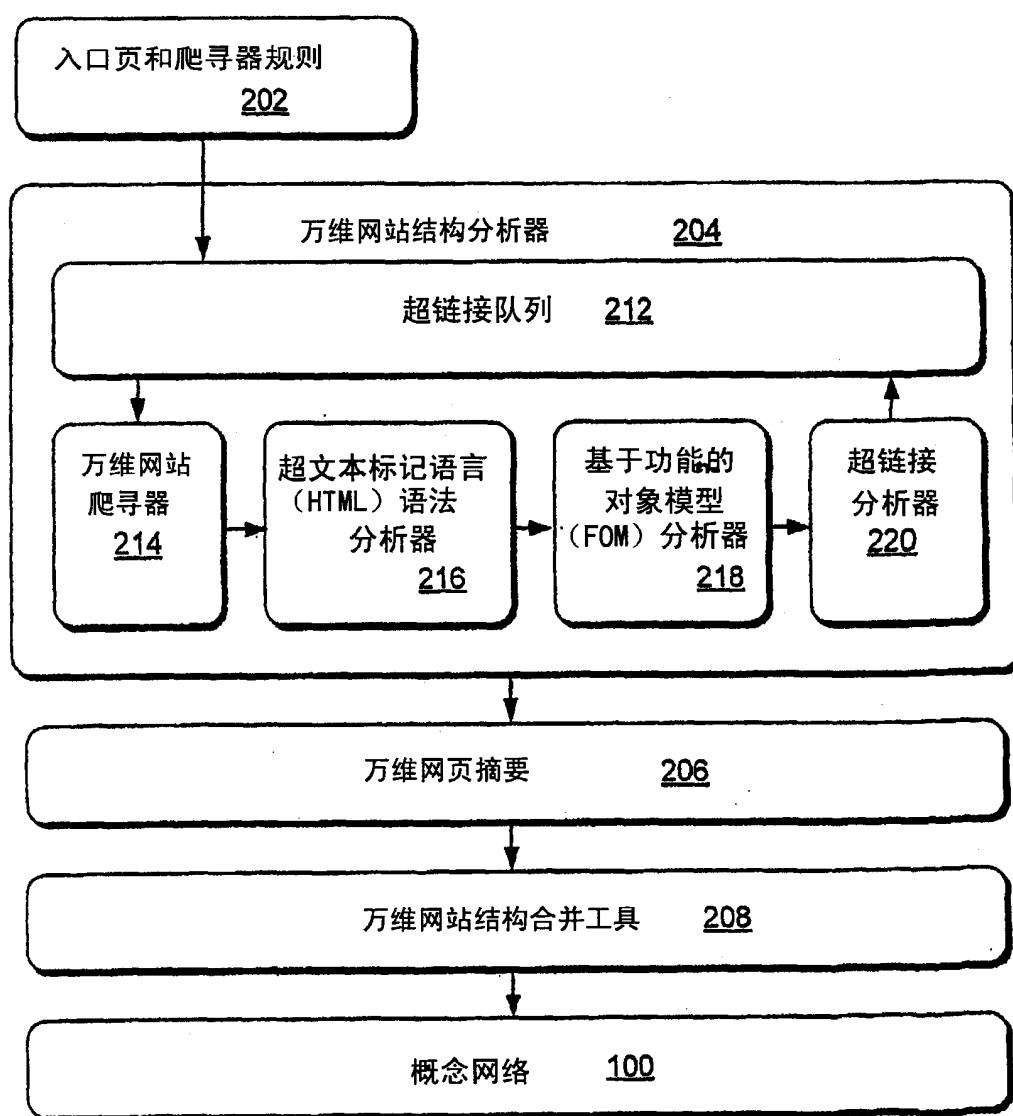


图 2

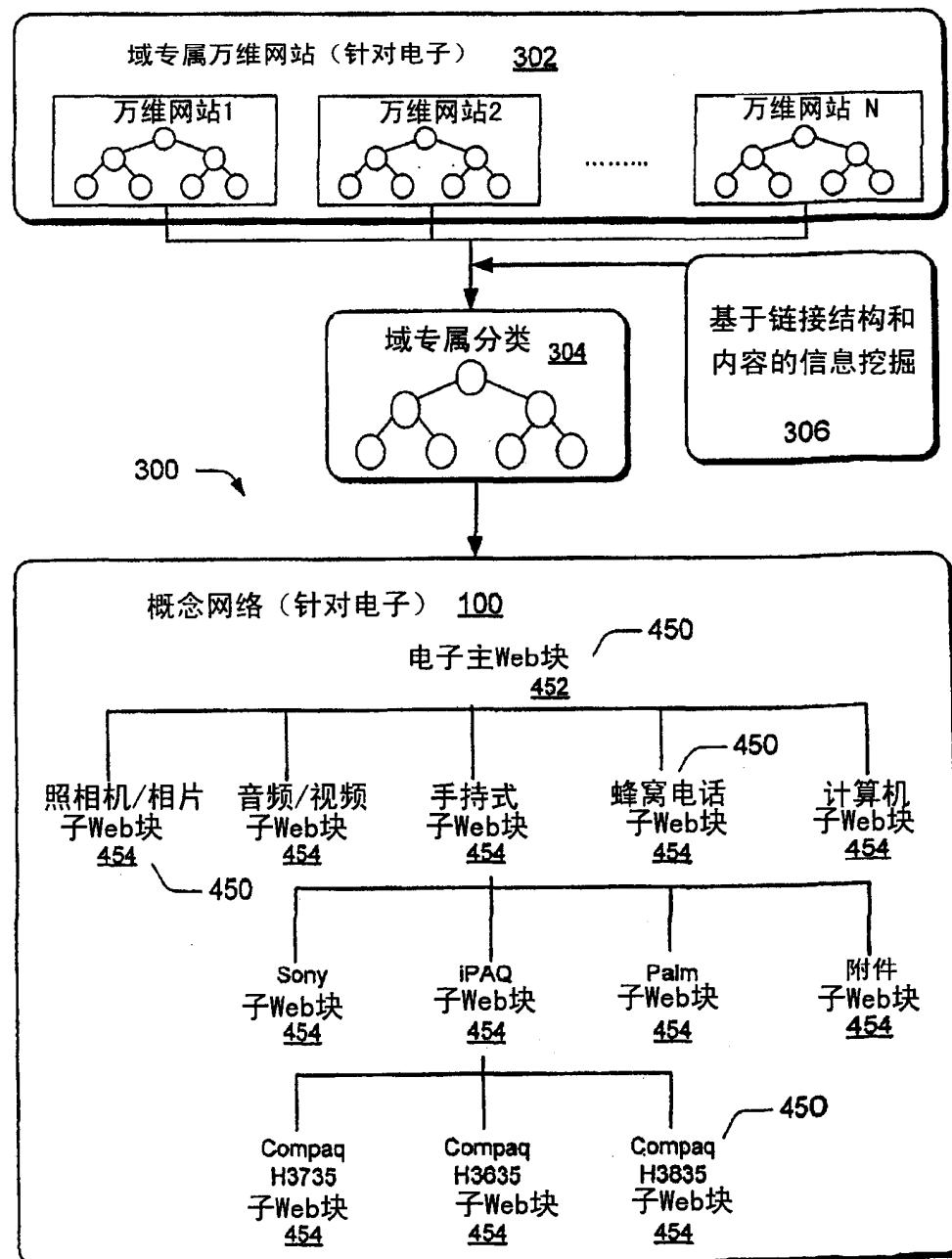


图 3

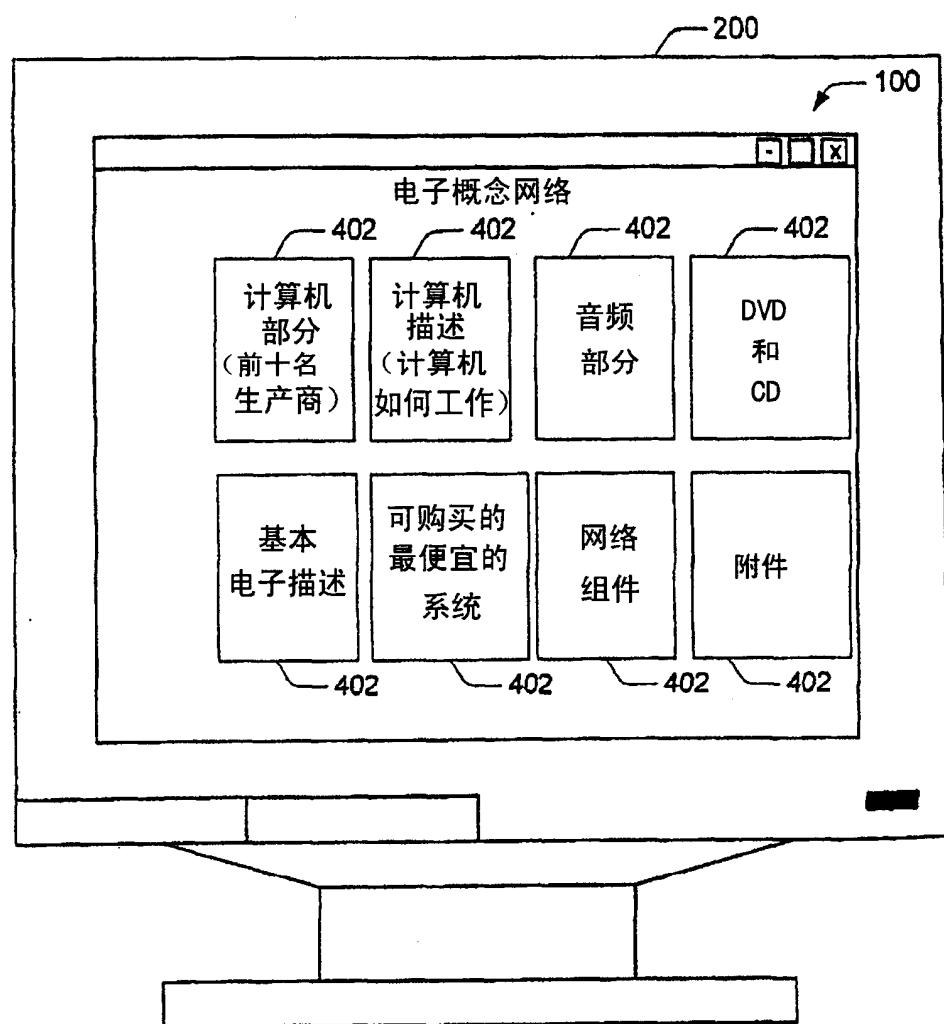


图 4

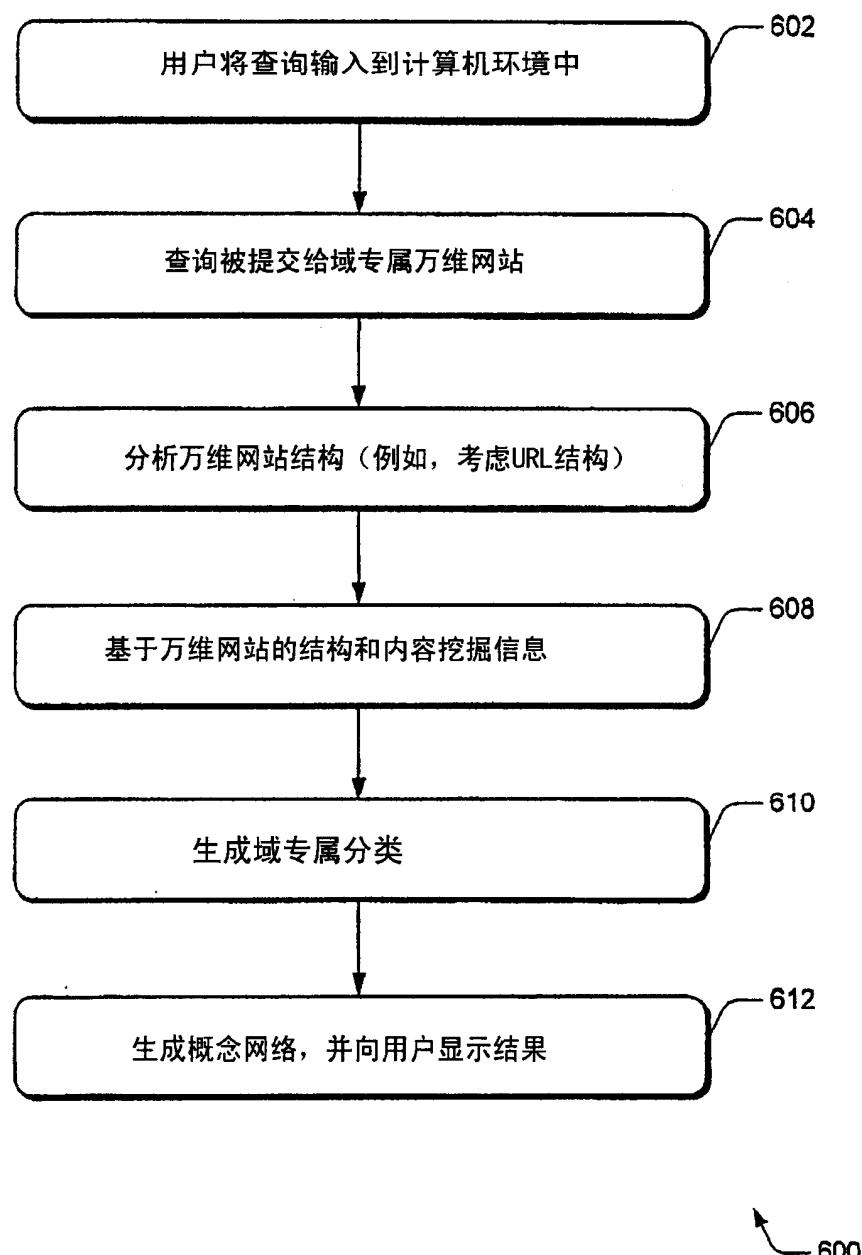


图 5

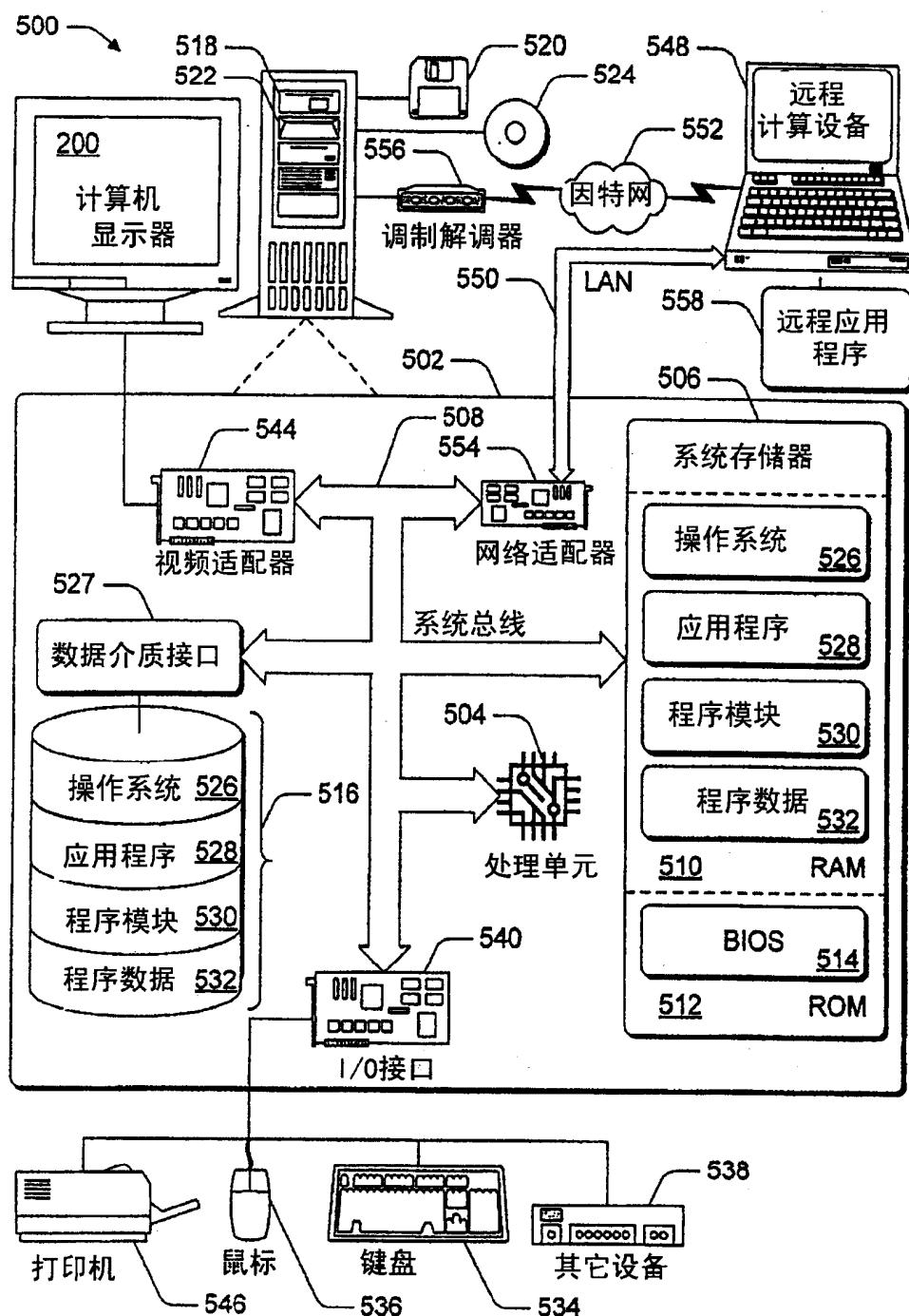


图 6