



[12] 发明专利申请公开说明书

[21] 申请号 200410087493.3

[43] 公开日 2005 年 3 月 30 日

[11] 公开号 CN 1601532A

[22] 申请日 2004.9.16

[21] 申请号 200410087493.3

[30] 优先权

[32] 2003.9.16 [33] US [31] 10/663,933

[71] 申请人 微软公司

地址 美国华盛顿州

[72] 发明人 M·A·纳爵克

[74] 专利代理机构 上海专利商标事务所

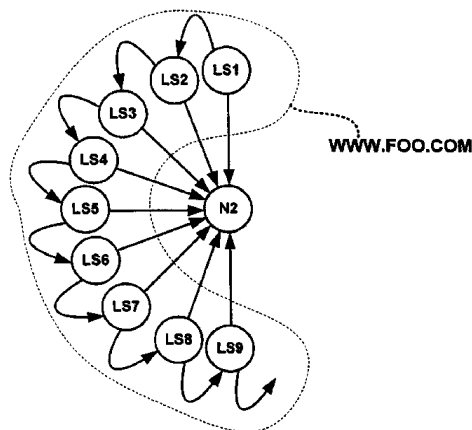
代理人 谢喜堂

权利要求书 4 页 说明书 21 页 附图 9 页

[54] 发明名称 根据结构相关信息排序文档的改进的系统和方法

[57] 摘要

提供了用于根据超链接信息以阻碍亲属链接的方式排序网页的系统和方法。在一个实施例中，提供一种返回质量查询结果的万维网搜索服务。本发明致力于用于网页的现有排序算法诸如页面排序的弱点，其中为了提高目标网页分值的单一目的而人工产生了这些网页。直觉上，本发明意识到，较到达具有很少网页的万维网服务器上的特定网页而言，经过随机的跳跃，到达具有许多网页的万维网服务器上的特定网页的可能性更小，这意味着减少了这种网页对通过链接到、或签注其它网页的另一个网页的影响。因此，在各种非限制性的实施例中，本发明为每个万维网服务器，而不是每个网页分配了一个所赋予的最小分值。然后，在该万维网服务器的所有网页中能够将分配给服务器的该最小分值进行细分。



1. 一种用于向结构上链接的多个文档的其中一个文档分配分值的方法，其特征在于，所述文档位于一 web 服务器上，所述 web 服务器通过以下的至少其中之一来定义：(A) 包括具有相同的符号主机名的多个网页的服务器、(B) 包括与相同的域相关联的多个网页的服务器、以及 (C) 具有与相同 IP 地址相关联的多个网页的服务器，并且所述文档具有来自所述结构上链接的多个文档中的至少一个其它文档的至少一个后链，而不论该文档位于何处，所述方法包括：
向所述文档分配一与位于所述 web 服务器上的文档数量成反比的分值。
2. 根据权利要求 1 所述的方法，其特征在于，它还包括：
向所述文档分配一与所述至少一个其它文档的数量成比例的分值。
3. 根据权利要求 1 所述的方法，其特征在于，它还包括：
分配与已分配给所述至少一个其它文档中的至少一个的至少一个分值成比例的分值。
4. 根据权利要求 1 所述的方法，其特征在于，它还包括：
分配与以下成比例的分值：(A) 所述至少一个其它文档的数量和 (B) 已分配给所述至少一个其它文档中的至少一个的至少一个分值。
5. 根据权利要求 2 所述的方法，其特征在于，它还包括：
向所述文档分配与所述至少一个其它文档中的至少一个的外链数量成反比的分值。
6. 根据权利要求 1 所述的方法，其特征在于，所述分配包括向所述文档分配与位于与所述文档相同的域上的文档数量成反比的分值。
7. 根据权利要求 1 所述的方法，其特征在于，所述分配包括向所述文档分配与具有与所述文档相同的符号主机名的文档数量成反比的分值。
8. 根据权利要求 1 所述的方法，其特征在于，所述分配包括向所述文档分配与关联到与所述文档相同的网际协议 (IP) 地址的文档数量成反比的分值。
9. 根据权利要求 1 所述的方法，其特征在于，它还包括：
基于链接到所述第一文档的至少一个其它文档的分值总和向所述文档分配分值。
10. 根据权利要求 1 所述的方法，其特征在于，所述结构上链接的多个文

档为具有超链接的网页，并且所述文档为一网页。

11. 根据权利要求1所述的方法，其特征在于，它还包括输出所述文档的分值到一 web 搜索服务的组件。

12. 根据权利要求1所述的方法，其特征在于，它还包括向一优选文档组
5 分配高于平均最小分值的分值。

13. 根据权利要求12所述的方法，其特征在于，所述优选文档组基于 Nielsen 等级、人所分配的等级、从 ISP 代理日志中提取的网页使用模式、从搜索引擎中提取的网页使用模式以及根据用户偏好所指定的文档的至少其中之一。

14. 根据权利要求1所述的方法，其特征在于，它还包括基于第二计分技
10 术改变所述文档的分值。

15. 根据权利要求1所述的方法，其特征在于，它还包括对照第二计分技术比较所述分值以发现反常结果。

16. 一种包括计算机可执行模块的应用编程接口，所述模块包含用于实现
权利要求1所述的方法的计算机可执行指令。

17. 一种包括用于实现权利要求1所述的方法的装置的计算设备。
15

18. 一种携带用于执行权利要求1所述方法的计算机可执行指令的已调制
数据信号。

19. 一种用于向结构上链接的多个文档中的一个文档分配分值的方法，其
特征在于，所述文档位于一 web 服务器上，所述 web 服务器通过以下其中至少
20 一个来定义：(A) 包括具有相同的符号主机名的多个网页的服务器、(B) 包括
与相同的域相关联的多个网页的服务器、以及 (C) 具有与相同的 IP 地址相关
联的多个网页的服务器，并且所述文档具有来自所述结构上链接的多个文档中的
至少一个源文档的至少一个后链，其中，所述文档的分值与同所述至少一个
源文档的至少一个相关联的至少一个分值成比例地计算，并且其中，所述分值
25 与位于所述 web 服务器上的所述至少一个源文档的数量成反比地计算。

20. 根据权利要求19所述的方法，其特征在于，所述分值与位于相同的
web 服务器上的所述至少一个源文档的数量成反比地计算。

21. 根据权利要求20所述的方法，其特征在于，所述分值与具有相同的符
号主机名的所述至少一个源文档的数量成反比地计算。

22. 根据权利要求20所述的方法，其特征在于，所述分值与同相同的域相
30

关联的所述至少一个源文档的数量成反比地计算。

23. 根据权利要求 20 所述的方法, 其特征在于, 所述分值与同相同的网际协议 (IP) 地址相关联的所述至少一个源文档的数量成反比地计算。

24. 根据权利要求 19 所述的方法, 其特征在于, 所述结构上链接的多个文档为具有超链接的网页, 并且所述文档为一网页。

25. 一种包括计算机可执行模块的应用编程接口, 所述模块包含用于实现权利要求 19 所述的方法的计算机可执行指令。

26. 一种包括用于实现权利要求 19 所述的方法的装置的计算设备。

27. 一种携带用于执行权利要求 19 所述的方法的计算机可执行指令的已调制数据信号。

28. 一种服务器对象, 其特征在于, 它包括:

一结合搜索引擎使用的应用程序接口, 它包括一查询机制, 用于基于一查询请求来查询网页信息数据库和相关联的分值, 籍此按照与每一结果相关联的分值排序从所述数据库中所检索的结果, 其中, 所述分值反映了满足所述查询条件的网页的质量; 以及

一生成与所述网页信息相关联的分值的计分对象, 其中, 对于具有到至少一个对应的源网页的至少一个后链的网页, 所述计分对象向所述网页分配一与同所述至少一个对应的源网页的至少一个相关联的至少一个分值成比例的分值, 并且其中, 所述分值与位于相同的 web 服务器上的所述至少一个对应的源网页的数量成反比地计算。

29. 根据权利要求 28 所述的服务器对象, 其特征在于, 所述计分对象独立于所述查询机制操作。

30. 根据权利要求 28 所述的服务器对象, 其特征在于, 基于一公共符号主机名定义一 web 服务器。

31. 根据权利要求 28 所述的服务器对象, 其特征在于, 基于一公共域定义一 web 服务器。

32. 根据权利要求 28 所述的服务器对象, 其特征在于, 基于一公共网际协议 (IP) 地址定义一 web 服务器。

33. 一种包括计算机可执行模块的计算机可读介质, 它包括用于向结构上链接的多个文档的其中一个文档分配分值的计算机可执行指令, 其特征在于,

所述文档位于一 web 服务器上, 并且具有一个来自所述结构上链接的多个文档中的至少一个其它文档的至少一个后链, 所述模块包括:

用于向所述文档分配与位于所述 web 服务器上文档数量成反比的分值的装置。

5 34. 根据权利要求 33 所述的计算机可读介质, 其特征在于, 它还包括:
用于向所述文档分配与所述至少一个其它文档的数量成比例的分值的装置。

35. 根据权利要求 33 所述的计算机可读介质, 其特征在于, 它还包括:
10 用于分配与分配给所述至少一个其它文档中的至少一个文档的至少一个分
值成比例的分值的装置。

36. 根据权利要求 33 所述的计算机可读介质, 其特征在于, 它还包括:
用于分配与以下成比例的分值的装置: (A) 所述至少一个其它文档的数量, 和 (B) 分配给所述至少一个其它文档的至少一个文档的至少一个分值。

37. 根据权利要求 34 所述的计算机可读介质, 其特征在于, 它还包括:
15 用于向所述文档分配与所述至少一个其它文档中的至少一个文档的外链数量成反比的分值的装置。

38. 根据权利要求 33 所述的计算机可读介质, 其特征在于, 所述分配装置包括向所述文档分配与位于具有与所述文档相同的符号主机名的 web 服务器上的文档的数量成反比的分值的装置。

20 39. 根据权利要求 33 所述的计算机可读介质, 其特征在于, 所述分配装置包括向所述文档分配与位于与所述文档相同的域上的文档数量成反比的的分值的装置。

40. 根据权利要求 33 所述的计算机可读介质, 其特征在于, 所述分配装置包括用于向所述文档分配与关联到与所述文档相同的网际协议 (IP) 地址的文档数量成反比的分值的装置。
25

根据结构相关信息排序文档的改进的系统和方法

5 版权通知和许可

本专利文献公开的一部分可能包含受到版权保护的内容。版权所有人并无意图反对任何人传真正如在专利商标局申请或记录的本专利文献或专利公开内容的副本，然而在别的方面却保留所有的版权权利。以下通知将应用于该文献：版权©2003，微软公司。

10 技术领域

本发明涉及根据结构相关的信息排序文档。更具体而言，本发明涉及根据超链接信息阻碍亲属或自服务链接的方式而排序网页。

背景技术

网页搜索服务，例如从用户或应用程序接受一个查询，并返回搜索结果的
15 列表，例如满足该查询条件的文档或与文档的链接。应该注意到，这里使用的术语“文档”指的是能够检索的任何内容，而不应该限制于诸如文字处理文档或网页的文档。为了提供满意的结果，当考虑到与用户最相关的这些文档应该首先出现时，应该对结果的列表进行排序。目前存在大量用于排序文档的算法，并且绝大多数的网页搜索引擎都应用若干的这种算法，并根据不同排序算法所
20 指定的排序组合，对查询的结果进行排序。

根据是否这些现存的大量排序算法为依靠查询的（也称为动态的）或查询独立的（也称为静态的），从而能够对它们予以分类。依靠查询的排序算法在查询中使用术语，而查询独立的排序算法则不同于它；即，该查询的排序算法为万维网上的每个文档分配一个质量分值。因此，能够有利地提前执行查询独立
25 的排序算法，并且不论何时提交一个查询都不需要重新运行它。

广义上，还能够将排序算法分类为基于内容、基于使用和基于链接的排序算法。基于内容的排序算法使用文档中的单词来排序文档（例如，依靠查询的基于内容的排序算法可能提前给在文档中包含查询术语的这些文档较高的分值，或巨大的字体或黑体字体）。基于使用的排序算法根据对网页浏览频率的估计而排序该网页；通过检查万维网代理日志或者监测在搜索引擎结果页面上的
30

点击,从而能够产生这些估计。最后,基于链接的排序算法使用网页之间的超链接以排序网页。

例如,通过这种想法即来自其它指向一个页面的其它页面的链接“签注”该页面,基于非常自然的静态链接的排序算法可能给每个网页分配一个与指向该网页的链接(“后链”)数量成比例的分值。例如,如图 1A 中所示,网页 A、B、C 和 D 每个都包含三个与其它网页的链接(“外链”),如网页中黑色矩形所表示的。在该实例中,使用基于静态链接的排序算法,页面 D 接收低于页面 C 的分值,因为页面 D 没有任何后链,而页面 C 有一个来自页面 B 的后链 L2 和一个来自页面 A 的后链 L1。应该注意到,下载完页面 A、B、C 和 D 之后,由于能够阅读这些页面,从而确定每个页面具有多少外链,它们链接到哪里,但是仍然可能存在来自一些还未知位置的一些未知的后链,诸如后链 LU,这些后链不能够被作为算法中的因数。该自然解决途径的主要缺陷是,平等地对待每个“签注文档”,以采用一种简单的系统。

到目前为止,页面排序为最熟知的基于查询独立链接的排序算法,因此在这里将阐述其原理。页面排序建立在图 1A 所示的通过添加一个递归层到该系统而形成的基于自然静态链接的系统原理的基础之上。如图 1B 所示例,其中示例了四个网页,并示出了页面排序的直觉知识。利用页面排序,当给签注的页面分配一个分值时,考虑签注页面的分值。因此,来自网页 E(分值为 100)的签注权值对给网页 G 分配的分值,比来自网页 F(分值为 9)的签注值其影响更大。直觉上看,能够想到在其被签注者之间细分的签注页面的分值。

从数学上分析,页面排序算法的直觉知识能够作如下解释:假设该组已知网页及其之间的链接归纳为具有顶点集合 V (其中每个顶点对应于一个网页)和边缘集合 E (其中每个边缘 (u, v) 对应于从页面 u 到页面 v 的一个超级连接)的一幅图。并让 $|V|$ 表示集合 V 的大小,让 $O(u)$ 表示集合 u 的出度程度(即,嵌入网页 u 中的超链接数量),以及让 p 为 0 到 1 之间的一个数(如, 0.15)。网页 v 的页面排序 $R(v)$ 定义为:

$$R(v) = \frac{p}{|V|} + (1-p) \sum_{(u,v) \in E} \frac{R(u)}{O(u)}$$

页面排序公式通常如下解释。假想一个在万维网上进行随意行进的万维网冲浪者。在沿着行进的每一步,该冲浪者从一个网页移动到另一个网页,其使

用如下的算法：利用某概率 p ，冲浪者随机地均匀选择一个网页并跳到该网页；否则，该冲浪者在当前网页中随机地均匀选择一个出局超链接并跟随它。由于该比喻，数量 p 有时候被称作“跳跃概率”——该概率为冲浪者将跳到一个完全随机页面的概率。如果网页冲浪者利用概率 p 跳跃且有 $|V|$ 个网页，跳到特定页
5 面的概率将为 $p/|V|$ 。由于通过跳跃能够检索任何的页面，因此每个页面都被赋予一个至少为 $p/|V|$ 的分值。

页面排序分值能够用于排序查询结果。利用所有其它相同的因数，应用页面排序的搜索引擎将对高于具备较低分值页面的具备高页面排序分值的页面进行排序。由于搜索引擎的绝大多数用户只检查首先出现的少数结果，商业网站
10 的运营商具有既定的利益，即与其站点的链接则提前出现在检索结果列表中，也就是说，它们的网页接收高的页面排序分值。换言之，商业网站运营商具备人工增加其网站上页面的页面排序分值的动机。

通过分析页面排序公式，将变得显而易见的是，一种增加网页 v 的页面排序分值的方法为将许多的其它网页链接到该网页。这是因为网页能够经由其外
15 链签注其它网页的思想是页面排序的核心。如果所有链接到 v 的网页的页面排序分值较低，则每一个单独的网页将会贡献甚微。但是，由于每个网页被赋予具备最小的页面排序分值 $p/|V|$ ，因此，来自许多这些低质量网页的链接仍旧能够贡献可观的总量。这揭露了页面排序算法的弱点。

实际上，页面排序的弱点是由网站所开发的，这些网站包含了大量的页面
20 集合，其唯一的目的在于“签注”网站的主页。典型地，这些签注的页面包含了与将要签注页面的链接，以及与其它签注页面的其它链接。所有的签注页面在空闲时间中自动地创建。因此，一旦万维网浏览器已经在任何的签注页面上失败，它将继续下载更多的签注网页（这是因为签注页面链接到其它签注页面），从而累积了大量的页面。该大量的页面，其中所有的页面都签注了一个页面，
25 并人工地提高了被签注的该页面的页面排序分值。用于人工提高页面排序分值的
技术通俗地已知为“链接兜售信息”。

另外，还已知了个性化的页面排序分值能够根据特定的方法创建一个万维网的视图。例如，通过采用用户书签和提高用户书签中这些页面的页面排序分
30 值，可以实现个性化的页面排序计分系统。本质上，指定一个网页作为书签的用户，已经隐含地将该网页签注作为一个书签，因此用户愿意将该书签作为计

分系统的基础。当用户很少愿意选择一个“链接兜售信息”页面作为书签时，不管许多“链接兜售信息”页面如何，个性化的页面排序思想不能明确地处理链接兜售信息的问题，这是因为仍旧有一个最低分值与每个链接兜售信息网页相关联。

- 5 因此，当基本思想正确时，页面排序的结果受到了亲属关系链接所引起干扰的影响，即，为了自签注和促进的目的，能够创建页面家族，而不考虑签注者或被签注者的实际价值。尽管已知了链接兜售信息相对于页面排序分值而存在的问题，但是解决方案已经避开了现有技术。

- 10 因此，期望一种改进的基于查询独立链接的排序算法。更具体而言，期望一种显著减少亲属链接的影响而改进的排序系统和方法。而且，期望一种改进的排序系统和方法，以减少链接兜售者为了人工提高与目标网页被签注者相关联的页面排序分值而创建自签注网页家族的动机。

发明内容

- 15 考虑到上述现有技术的缺陷，本发明提供了用于根据有关文档的结构相关信息而排序文档的系统和方法。本发明的系统和方法能够用于根据超链接信息，以一种阻碍亲属关系链接的方式而排序网页。在各种实施例中，在网页搜索服务中实现本发明以返回质量查询结果。本发明致力于用于网页的现有排序算法诸如页面排序的弱点，其中为了提高目标页面分值的单一目的而人工产生了这些网页。直觉上，本发明意识到，较到达具有很少网页的万维网服务器上的特定网页而言，经过随机的跳跃，到达具有许多网页的万维网服务器上的特定网页的可能性更小，这意味着减少了这种网页对通过链接到、或签注其它网页的另一个网页的影响。因此，在各种非限制性的实施例中，本发明为每个万维网服务器，而不是每个网页分配了一个所赋予的最小分值。然后，在该万维网服务器的所有页面中能够将分配给服务器的该最小分值进行细分。

- 25 以下描述本发明的其它优点和特征。

附图说明

以下将参考附图进一步描述根据本发明用于提供改进的排序算法的系统和方法，其中：

- 30 图1A和1B分别示例了根据基于静态链接和页面排序的算法用于排序网页的现有技术；

图 2A 所示为包括各种计算机设备的示意性网络环境的方框图，其中在该计算机设备中可以实现本发明；

图 2B 所示为其中可以实现本发明的非限制性计算机设备的方框图；

图 3A 和 3B 示例了根据本发明在防止链接兜售信息之后的直觉知识；以及

5 图 4A 到 4C 示例了在搜索引擎应用程序中示意性地实现本发明的各种实施例。

具体实施方式

概述

如上所述，页面排序算法采取大胆的任务将万维网上的每个页面压缩到单
10 个数量，即页面的页面排序。页面排序为仅仅基于网页在万维网图形结构中的位置对所有网页的全局排序，而不考虑其内容如何。

使用页面排序，对搜索结果进行排序，从而给更重要和中心的网页分配优先权。在页面排序后的直觉知识就是，它使用网页本身之外的信息—提供同级评审的网页后链。而且，通过递归定义，考虑来自“重要”网页的后链比来自
15 普通链接的后链更重要。

另外，还已知了个性化页面排序分值能够根据特定的方法创建万维网的视图，例如通过采用用户书签和提高用户书签中这些网页的页面排序分值；但是，个性化页面排序不能明确地处理链接兜售信息的问题，这是因为仍旧存在一个与每个链接兜售信息网页相关联的最小分值。因此，链接兜售信息者仍然能够
20 在单个万维网服务器上（如果想要，自动地）创建大量的网页，每个网页具有其自己的最小页面排序分值，从而通过签注每个其它的目标被签注者网页，人工地提高了目标被签注者网页的分值。一个链接兜售信息者典型创建以下其中的任何一个或多个（A）将具有相同符号的主机名，（B）将与相同的域相关联，或（C）将与相同的 IP 地址相关联的大量网页。

25 本发明通过意识到较随机跳将到达具有很少页面的万维网服务器上的特定页面而言，它将会到达具有许多页面的万维网服务器上的特定页面的可能性更小，从而致力于该弱点。这意味着减少了这种页面对通过链接到、或签注其它页面的另一个页面的影响。因此，在各种非限制性的实施例中，本发明为每个万维网服务器，而不是每个网页分配了一个所赋予的最小分值。然后，在该万
30 维网服务器的所有页面中能够将该最小分值进行细分。

示意性联网和分布式的环境

本领域的普通技术人员能够理解，本发明能够结合任何计算机或任何客户机或服务器设备，或在分布式计算环境中予以实现，其中这些设备配置作为计算机网络的一部分。有关与此，本发明涉及具备任何数量存储器或存储单元的任何计算机系统或环境，在任何数量的存储单元或容量中发生的任何数量的应用程序和方法，这些程序可能结合根据本发明而排序文档的方法来使用。本发明可应用于具备在网络环境或分布式计算环境下而配置的服务器计算机和客户端计算机的环境中，其中上述计算机具有远程或本地存储器。本发明还可以应用于独立的计算设备中，该设备具有程序语言的功能、解释程序和产生、接收和发送与远程或本地服务有关信息的执行能力。下载和分析网页尤其与在网络中或分布式计算环境中运行的这些计算设备有关，因此，根据本发明的排序算法和技术能够以最高的效率应用于这些环境中。

分布式计算通过在计算设备和系统之间进行交换从而提供了计算机资源和服务的共享。这些资源和服务包括信息交换、用于文档的缓存存储器和磁存储器。分布式计算利用网络互连的优点，允许客户端平衡其集体的力量以使整个企业受益。为此，各种设备可能具有可以包含本发明排序算法和方法的应用程序、对象或资源。

图 2A 提供了示意性联网或分布式计算环境的示意图。分布式计算环境包括计算对象 10a、10b 等等，以及计算对象或设备 110a、110b、110c 等等。这些对象可能包括程序、方法、数据存储器、可编程的逻辑等等。该对象可包括相同或不同设备的一部分，诸如 PDA、电视机、MP3 播放器、个人计算机等等。每个对象能够与另一个对象通过通信网络 14 的方式进行通信。该网络可能本身包括其它的计算对象和提供服务给图 2A 中所示系统的计算设备，以及本身表示多个互连的网络。根据本发明的一个方面，每个对象 10a、10b 等等或 110a、110b、110c 等等，可能包含一个应用程序，该程序可能使用 API，或其它对象、软件、固件和/或硬件，以请求使用根据本发明的排序方法。

而且还能够理解，一个对象诸如 110c，可能主管另一个计算设备 10a、10b 等等或 110a、110b 等等。因此，尽管所描绘的物理环境可能示出互连的设备，如计算机，但是这种示例仅仅是示意性的，并且可替换地描绘或描述物理环境，其包括各种数字设备，诸如 PDA、电视机、MP3 播放器等等，以及诸如接口、

COM 对象等等的软件对象。

存在各种支持分布式计算环境的系统、组件和网络结构。例如，可以通过有线或无线系统、本地网或广域分布的网络将这些计算系统连接在一起。目前，许多网络都耦接到因特网上，并且在因特网上提供了用于广域分布计算的基础设施并包含了许多不同的网络。任何的基础设施都可以用于示意性的根据本发

5 明的与排序文档相关联的通信，这些文档具有相关的链接。

在本地网络互连环境中，存在至少四个完全不同的每个都支持独特协议的网络传输媒体，诸如电源线、数据（无线和有线）、声音（例如电话）以及娱乐媒体。大多数本地控制设备诸如光开关和设备可以使用电源线来连接。数据服务作为宽带（例如，DSL 或电缆调制解调器）可进入到家庭中，以及可使用无线（例如，家庭射频或 802.11B）或有线（例如，家庭 PNA，Cat5、以太网、甚至电源线）连接在家庭内访问该数据服务。话音业务通过有线（例如，Cat3）或无线（例如，蜂窝电话）等方式可以进入到家庭，并且使用 Cat3 线路在家庭中进行布线。娱乐媒体或其它图形数据可以通过卫星或电缆进入家庭，并典型

10 地使用同轴电缆在家庭中布线。而且，IEEE1394 和 DVI 为用于媒体设备簇的数字互连。所有的这些网络环境和其它的可作为协议标准出现的网络环境都可以互连形成一个网络，诸如企业内部互联网，该网通过因特网而连接到外界。简而言之，为了数据存储和传输，存在各种完全不同的资源，因此，在将来，计算设备将需要共享数据诸如与程序对象关联的被访问或被使用数据的方法，这

15 些方法使用根据本发明的排序技术。

因特网通常指使用计算机网络互连领域中所熟知的 TCP/IP 协议组的网络和网关的集合。TCP/IP 为“传输控制协议/网际协议”的首字母缩写词。能够将因特网描述为通过执行联网协议的计算机互连的、在地理上分布的远程计算机网络的系统，联网协议允许用户相互作用并共享网络信息。由于这些广泛分布

25 的信息共享，诸如因特网的远程网络因此目前演化为开放式系统，开发人员可以为该系统设计各种执行具体操作或服务的、而基本上毫无限制的软件应用程序。

因此，网络基础设施允许网络拓扑，诸如客户机/服务器、点对点或混合体系结构的主机操作。“客户机”为一类或一组使用与其不相关的另一类或另一组

30 服务的一员。因此，在计算中，客户机为一个请求另一个程序所提供服务的进

程,即一般来说为一组指令或任务。该客户机进程使用所请求的服务而不必“知道”任何有关其它程序或服务本身的工作细节。在客户机/服务器体系结构中,尤其是在一个联网系统中,客户机通常为访问另一个计算机,例如服务器,而提供的共享网络资源的计算机。在图 2A 的实例中,能够将计算机 110a、110b 等等考虑作为客户机,以及将计算机 10a、10b 等等考虑作为服务器,这里服务器 10a、10b 等等保存接下来在客户机计算机 110a、110b 等等中复制的数据,尽管能够根据具体情况而将任何计算机考虑作为客户机、服务器或二者兼有。任何这些计算可以处理数据或请求可能包含本发明排序技术的服务或任务。

服务器典型为在远程或本地网络,诸如在因特网上可访问的远程计算机系统。在第一计算机系统中可能激活客户机进程,在第二计算机系统中可能激活服务器进程,它们在通信介质上相互通信,并因此提供分布式功能和允许多个客户机利用服务器的信息收集功能。任何被使用的按照本发明排序技术的软件对象可能分布于整个多计算设备或对象中。

客户机和服务器使用协议层提供的功能相互通信。例如,超文本传输协议 (HTTP) 为结合万维网 (WWW) 或“环球网”所使用的普通协议。典型地,诸如因特网协议 (IP) 地址或其它参照,诸如统一资源定位符 (URL) 的计算机网络地址能够用于相互识别服务器或客户端计算机。网络地址能够称之为 URL 地址。在通信介质上能够提供通信,例如,客户机和服务器可以经由 TCP/IP 连接彼此耦合用于高容量通信。

因此,图 2A 通过一个服务器经由网络/总线与客户端计算机通信,示例了示意性的联网或分布式的环境,以及本发明可以应用于该环境中。更具体而言,根据本发明,大量的服务器 10a、10b 等等经由通信网络/总线 14,例如 LAN、WAN、企业内部互联网、因特网等等,与大量的客户机或远程计算设备 110a、110b、110c、110d、110e 等等,诸如便携式计算机、掌上电脑、瘦型客户机、联网设备或其它设备,诸如 VCR、TV、烘箱、灯、加热器等等进行互连。因此应该理解,本发明可以应用于任何计算设备中,结合该计算设备,期望实现对具备结构相关链接的文档进行排序。

在一种网络环境中,通信网络/总线 14 为因特网,例如,服务器 10a、10b 等等能够为经由许多已知协议诸如 HTTP 与客户机 110a、110b、110c、110d、110e 等等通信的万维网服务器。服务器 10a、10b 等等还可以用作客户机 110a、

110b、110c、110d、110e 等等，这可以作为分布式计算环境的特征。

在合适的时候，通信可能为有线或无线的。客户机设备 110a、110b、110c、110d、110e 等等可以或不可以经由通信网络/总线 14 进行通信，并且可能具有与其相关的独立通信。例如，在 TV 或 VCR 的情况下，对于其控制而言，可能存在或不可能存在联网的方面。每个客户端计算机 110a、110b、110c、110d、110e 等等以及服务器计算机 10a、10b 等等可以装备有各种应用程序模块或对象 135，以及对各种类型存储单元或对象的连接或访问，在该存储单元或对象上可以存储文档或数据流，或可以下载、传输或移动文档或数据流的一部分到其上。任何一个或多个计算机 10a、10b、110a、110b 等等可负责维护和更新数据库 20 或其它存储单元，诸如数据库或存储器 20，其用于存储根据发明处理的数据。因此，本发明能够用于计算网络环境中，该网络环境包括客户端计算机 110a、110b 等等和服务器计算机 10a、10b 等等以及其它类似的设备和数据库 20，该客户端计算机能够访问和与计算机网络/总线 14 相互作用，服务器计算机可以与客户端计算机 110a、110b 等等相互作用。

15 示意性的计算设备

图 2B 和以下讨论旨在提供将本发明予以结合实现的合适计算环境的简要普遍描述。但是，应该理解，期待结合本发明而使用手持、便携式和其它计算设备和所有类型的计算对象，即在计算环境中与网页或其它结构相关文档接口的任何地方。尽管以下描述了通用目的的计算机，但是这仅仅作为一个实例，20 本发明也可以利用一个瘦型客户机实现，该客户机具备网络/总线互操作性和相互作用。因此，本发明可以在联网主机服务的环境中予以实现，在该服务中包含了几乎很少或最小的客户机资源，例如客户机设备仅仅用作与网络/总线的接口的联网环境，诸如位于设备中的一个对象。本质上，存储数据的任何地方或从任何地方检索数据或从其发送数据到其它计算机，为期望或适合的用于运行根据本发明排序技术的环境。

25 尽管不是所需要的，但是本发明能够经由一个操作系统而实现，该系统用于设备或对象的服务开发者所使用，和/或包括在结合本发明排序技术操作的应用软件中。软件通常可以在计算机可执行指令的环境中，诸如在一个或多个计算机诸如客户机工作站、服务器或其它设备所执行的程序模块中描述。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、对象、30

组件、数据结构等等。典型地，如在各种实施例中所所述，可以组合或分布程序模块的功能。而且，本领域的普通技术人员将会理解，可以利用其它计算机系统配置和协议来实践本发明。其它熟知的适于本发明所使用的计算系统、环境、和/或配置包括，但不限制于，个人计算机（PC）、自动柜员机、服务器计算机、手持或膝上型设备、多处理器系统、基于微处理器的系统、可编程的消费电子、网络 PC、家用电器、灯、环境控制元件、微计算机、大型计算机等等。本发明还可以在分布式计算环境中予以实践，在该环境中由远程处理设备执行任务，该设备通过通信网络/总线或其它数据传输介质链接在一起。在分布式计算环境中，程序模块可以位于包括存储器存储设备的本地和远程计算机存储介质中，以及客户机节点可轮流作为服务器节点。

图 2B 示例了可以实现发明的合适计算系统环境 100 的一个实例，尽管如上清楚所述，但是计算系统环境 100 仅仅作为合适计算环境的一个实例，并无意图建议作为本发明使用范围或功能的任何限制。因此，也不应该将计算环境 100 解释为具备任何独立性或与示意性操作环境 100 中所示例的任何一个组件或其组合有关的要求。

参考图 2B，用于实现本发明的示意性系统包括计算机 110 形式的通用目的的计算设备。计算机 110 的组件包括但不限于，处理单元 120、系统存储器 130、用于将各种系统组件包括系统存储器耦合到处理单元 120 的系统总线 121。该系统总线 121 可以为任何若干类型的总线结构，包括存储器总线或存储器控制器、外设总线、使用任何的各种总线结构的本地总线。通过实例而不是限制的方式，这些体系结构包括工业标准结构（ISA）总线、微通道结构（MCA）总线、增强 ISA（EISA）总线、视频电子标准协会（VESA）本地总线、和外设互连（PCI）总线（还已知为 Mezzanine 总线）。

计算机 110 典型包括各种计算机可读介质。计算机可读介质能够为计算机 110 可存取的任何可用介质，并且包括易失性和非易失性介质，可移动的和不可移动的介质。通过实例而不是限制的方式，计算机可读介质可以包括计算机存储介质和通信介质。计算机存储介质包括以任何信息存储方法或技术实现的易失性和非易失性介质，可移动和不可移动的介质，该信息诸如计算机可读指令、数据结构、程序模块或其它数据。计算机存储介质包括但不限于，RAM、ROM、EEPROM、闪存或其它存储技术、CDROM、数字光盘（DVD）或其它光盘存

存储器、磁性盒式磁带、磁带、磁盘存储器或其它磁盘存储设备、或任何其它能够用于存储想要信息并由计算机 110 存取的介质。通信介质典型地包含计算机可读指令、数据结构、程序模块或已调制数据信号，诸如载波或其它传输机制中的其它数据，并包括任何信息传送介质。术语“已调制数据信号”指的是，
5 具有以一种方式设置或改变的一个或更多特性的信号，以通过这种方式编码信号中的信息。通过实例而不是限制的方式，通信介质包括有线介质诸如有线网络或直接的有线连接，以及无线介质，诸如声音、射频、红外或其它无线介质。任何上述介质的组合也都应该包括在计算机可读介质的范围之内。

系统存储器 130 包括易失性和/或非易失性存储器，诸如只读存储器(ROM)
10 131 和随机访问存储器 (RAM) 132 形式的计算机存储介质。基本输入/输出系统 133 (BIOS) 包含帮助在计算机 110 的组件之间，诸如在启动过程中传送信息的基本例程，并且该系统被典型地存储在 ROM131 中。RAM132 典型包含处理单元 120 立即访问和/或不久由处理单元运行的数据和/或程序模块。通过实例而不是限制的方式，图 2B 示例了操作系统 134、应用程序 135、其它程序模块
15 136 和程序数据 137。

计算机 110 还可以包括其它可移动/不可移动的、易失性/非易失性计算机存储介质。仅仅通过实例的方式，图 2B 示例了硬盘驱动器 141，它从非移动、非易失性磁介质中读取或写入数据，和从可移动的、非易失性磁盘 152 中读取或写入数据的磁盘驱动器 151，以及从可移动的、非易失性光盘 156、诸如 CD-ROM
20 或其它光介质中读取或写入数据的光盘驱动器 155。其它能够用于示意性操作环境中的可移动/不可移动的、易失性/非易失性计算机存储介质包括，但不限制于，磁性盒式磁带、闪存卡、数字化视频光盘。数字视频磁带、固体 RAM、固体 ROM 等等。硬盘驱动器 141 典型地通过不可移动的存储器接口，诸如接口 140 与系统总线 121 相连接，以及光盘驱动器 155 典型通过可移动的存储器接口，
25 诸如接口 150 与系统总线 121 相连接。

上述讨论和在图 2B 中所示例的驱动器及其相关联的计算机存储介质提供了计算机可读指令、数据结构、程序模块和其它用于计算机 110 的数据的存储器。在图 2B 中，例如，硬盘驱动器 141 示例为存储操作系统 144、应用程序 145、其它程序模块 146 和程序数据 147。应该注意到，这些组件能够与用于操作系统
30 134、应用程序 135、其它程序模块 136 和程序数据 137 相同或不同于后者。操

作系统 144、应用程序 145、其它程序模块 146 和程序数据 147 在这里给定了不同的数量以便通过最小值来示例它们为不同的复制件。用户可以通过输入设备，诸如键盘 162 和定位设备 161，通常称之为鼠标、轨迹球或触摸屏，以在计算机 110 中输入命令和信息。其它输入设备（未示出）可以包括麦克风、操纵杆、游戏垫、圆盘式卫星电视天线、扫描仪等等。这些和其它的输入设备通常通过系统总线 121 相耦合的用户输入接口 160 连接到处理单元 120，但是也可以通过其它接口和总线结构，诸如并行端口、游戏端口或通用串行总线（USB）予以连接。图形接口 182，诸如北桥，还可以与系统总线 121 相连接。北桥为与 CPU 或主机处理单元 120 通信的芯片组，并假定它负责加速图形端口（AGP）通信。

一个或多个图形处理单元（GPU）184 可以与图形接口 182 通信。有关这方面，GPU184 通常包括芯片内存储存储器，诸如寄存存储器，以及 GPU184 与视频存储器 186 通信，其中本发明的应用变型可以具有影响。但是，GPU184 仅作为协处理器的一个实例，因此各种的协处理设备可以包括在计算机 110 中，并且可以包括各种程序补偿器，诸如像素和峰值补偿器。监视器 191 或其它类型的显示设备还通过接口，诸如视频接口 190 与系统总线 121 连接，该接口依次与视频存储器 186 通信。除了监视器 190 以外，计算机还可以包括其它外围输出设备，诸如通过输出外围接口 195 相连接的扬声器 197 和打印机 196。

计算机 110 可以使用与一个或多个远程计算机，诸如远程计算机 180 的逻辑连接在联网或分布式环境中操作。远程计算机 180 可以为个人计算机、服务器、路由器、网络 PC、点对点设备或其它公共网络节点，并且虽然在图 2B 中已经仅仅示例了一个存储器设备 181，但是它典型包括部分或全部上述与计算机 110 相关的元件。图 2B 中描述的逻辑连接包括局域网（LAN）171 和广域网（WAN）173，但是还可以包括其它的网络/总线。这种网络互连的环境常见于家庭、办公室、企业广域计算机网络、企业内部局域网以及因特网。

当在 LAN 联网环境中使用计算机 110 时，它通过网络接口或适配器 170 与 LAN171 相连接。当在 WAN 联网环境中使用计算机 110 时，它典型地包括调制解调器 172 或其它用于在 WAN 173 诸如因特网上建立通信的装置。该调制解调器 172 可以为内置或外置的，并通过用户输入接口 160，或其它适当的机制连接到系统总线 121。在联网环境中，上述与计算机 110 有关的程序模块或其部分，可以存储在远端存储设备中。通过实例而不是限制的方式，图 2B 示例了如

驻留在存储设备 181 中的远程应用程序 185。将会理解，所示的网络连接为示意性的，并且可以使用其它装置在计算机之间建立通信链接。

示意性的分布式计算框架或结构

根据个人计算和因特网集中的教导，已经并正在开发各种分布式计算框架。

- 5 个体和商务用户配有用于应用程序和计算设备的无缝协同工作和基于 Web 的接口，使得计算活动逐渐面向 Web 浏览器或网络。

- 例如，微软®的管理代码平台，即.NET，包括服务器、构造块服务，诸如基于 Web 的数据存储和可下载的设备软件。一般而言，.NET 平台提供了（1）使得整个范围内的计算设备一起工作并自动更新以及同步用户信息的能力，
- 10 （2）增加了除了 HTML 以外而大量使用 XML 而使能的网页交互作用能力，（3）以从用于管理诸如电子邮件的各种应用程序，或诸如办公.NET 的软件的中央开始点定制访问和传送产品以及服务到用户为特征的在线服务，（4）集中数据存储，增强了效率和访问信息以及在用户和设备之间信息同步的便捷性，（5）集成各种通信媒体，诸如电子邮件、传真和电话的能力，（6）为开发者而言，增
- 15 加了可再用模块的能力，以此增加了生产率并减少了程序错误的数量以及（7）以许多其它的交叉平台和语言集成为特征。

- 尽管在这里结合驻留在计算机设备中的软件描述了一些示意性实施例，但是本发明的一部分或更多部分还可以通过操作系统、应用程序接口（API）或“中间人”对象、控制对象、硬件、固件、中间语言指令或对象等等来实现，以便
- 20 在所有的语言和管理代码，诸如.NET 代码所使能的服务，以及在其它分布式计算框架中可以包括、支持这些方法或通过上述来访问该方法。

用于排序网页的系统和方法

- 如以上在背景技术中所述，本发明针对现有排序算法，诸如页面排序的弱点，即单单为了增加目标网页的分值而人工产生网页。假定万维网冲浪者将要
- 25 随机地跳到一个特定的服务器，本发明意识到该冲浪者将在具有许多页面的万维网服务器的特定页面上“着陆”的可能性，较他将在具有较少页面的万维网服务器的特定页面上“着陆”的可能性较小。应用该原理，当应用一个排序算法时，当签注来自相同的服务器时，本发明减少了签注的影响。为了实现这种减少签注的影响，在各种非限制性的实施例中，本发明为每个万维网服务器，
- 30 而不是每个万维网网页分配一个已赋予的最小分值。然后，在万维网服务器的

所有网页之间细分该分配给服务器的最小分值。因此，通过万维网服务器的许多页面创建任意大数量的亲属链接，与在万维网服务器的较少页面上创建相对少数量的亲属链接相比较而言，获得了几乎相等的“签注值”。通常，任意大数量的亲属链接 (A) 将具有相同符号的主机名，(B) 与相同的域相关联，或 (C) 与相同的 IP 地址相关联。在一个实施例中，为了给用户返回一个质量查询结果，在万维网搜索服务中实现本发明。

应该注意到，存在若干可能的构成万维网服务器的定义：万维网服务器可能通过符号主机名（例如，www.google.com）、公共域来定义，或者可能通过一个（或若干）IP 地址（例如，207.46.134.222）定义。当利用万维网服务器地址时，实施将针对链接兜售信息问题，后者的两个定义更适于本发明的目的，并作为对链接兜售信息者的更好威慑，这是因为有可能配置 DNS 服务器来解析主机名的几乎无限的数量（RFC1035 限制主机名最大为 255 个字符长，每个字符为一个字母、数字或连字号。因此，有 37^{255} 个可能的主机名，对于所有实际目的而言这是“几乎无限的”），正如有可能配置万维网服务器来用作网页的无限数量。以下说明轮流使用用于两个可替换实施例的每个定义。

在根据本发明排序度量的第一实施例中，万维网服务器通过它的符号主机名来定义。 $h(u)$ 表示 URL u 的主机名组成部分，以及 H 为一组所有的主机，即：

$$H = \{h(v) : v \in V\}$$

$$V^H(h) \text{ 为主机 } h \text{ 所服务的一组 URL，即}$$

$$V^H(h) = \{v : v \in V \wedge h(v) = h\}$$

然后根据本发明的第一实施例，在第一新排序度量 R^H 的定义中使用这两种定义，其如下定义：

$$R^H(v) = \frac{p}{|H| |V^H(h(v))|} + (1-p) \sum_{(u,v) \in E} \frac{R^H(u)}{O(u)}$$

在根据本发明的排序度量的第二个实施例中，通过其域名定义万维网服务器。 $d(u)$ 表示 URL u 的主机名组成部分以及 D 为一组所有的域，即：

$$D = \{d(v) : v \in V\}$$

$$V^D(d) \text{ 为在域 } d \text{ 中万维网服务器所服务的该组 URL，即：}$$

$$V^D(d) = \{v : v \in V \wedge d(v) = d\}$$

然后, 在根据本发明第二实施例的第二新排序度量 R^D 的定义中使用这两个定义, 第二排序度量如下表示:

$$R^D(v) = \frac{p}{|D|} \frac{1}{|V^D(d(v))|} + (1-p) \sum_{(u,v) \in E} \frac{R^D(u)}{O(u)}$$

- 5 在本发明的第三实施例中, 通过它的一组 IP 地址来定义万维网服务器。A(u) 表示该组 IP 地址, 根据该地址来服务 URL u (即, h(u) 解析的该组 IP 地址)。A 为一组所有的 IP 地址, 即:

$$A = \bigcup_{v \in V} A(v)$$

$V^A(a)$ 为 IP 地址 a 所服务的 URL 组, 即:

10 $V^A(a) = \{v : v \in V \wedge a \in A(v)\}$

然后, 在根据本发明第三实施例的第三新排序度量 R^A 的定义中使用这两个定义, 第三排序度量如下表示:

$$R^A(v) = \frac{p}{|A|} \sum_{a \in A(v)} \frac{1}{|V^A(a)|} + (1-p) \sum_{(u,v) \in E} \frac{R^A(u)}{O(u)}$$

- 现在参考图 3A 到图 3G 描述本发明。图 3A 示例了链接到第二节点 N2 的
15 第一节点 N1 (例如, 文档或网页)。页面排序给人的直觉知识为, N2 为比不具备入局链接的节点更好的页面, 这完全是因为 N1 通过链接到 N2 而签注或“验证”了 N2 的存在。通过比较图 3B 和图 3A 可进一步说明页面排序的直觉知识。由于在图 3A 中, N1 只链接到一个节点 N2, 而在图 3B 中, 节点 N1 链接到九个节点 N2 到 N10, 由于 N1 看似不加选择地签注其它与图 3A 的节点 N1 相关的
20 页面, 因此页面排序降低了 N1 的签注值。由于图 3A 的节点 N1 更有选择地签注其它的节点, 从而作为 N1 链接到 N2 的结果 (所有的其它因数都相同), 在图 3A 中相应地给 N2 分配一个较高的分值。通过比较图 3C 和图 3A 可进一步说明页面排序的直觉知识。在该比较中, 在图 3C 中分配给签注节点 N1 的分值为在图 3A 中分配给签注节点 N1 的分值的 10 倍。因此, 由于签注节点的质量 (分值) 高于图 3C 中所示的节点的分值, 因此在图 3C 中分配给节点 N2 的分值
25 高于在图 3A 中分配给节点 N2 的分值。组合这些直觉的知识产生了页面排序算法。

- 但是, 如在背景技术中所述, 由于无论它签注多少其它的节点, 或无论它自己的页面排序分值多小, 每个节点都接收一个最小的分值, 因此对于图 3D 中
30 所示例的场景来说, 该页面排序存在弱点。当该最小分值较小时, 以巨大的数量与小分值相乘能变得显著。因此, 节点 N2 的所有者, 通过手动或自动创建大

量的签注页面，每个签注节点 N2 和每个其它的某种方式或安排的节点例如，诸如图 3D 所示例的节点，能够人工增加节点 N2 的分值。为了以更经济有效的方式实现上述操作，一般而言，链接兜售信息者将通常在相同的主机（例如，www.foo.com）上放置每个链接兜售节点 LS1 到 LS9，因为在相同的主机诸如

5 www.foo.com/页面 1，www.foo.com/页面 2，www.foo.com/页面 3 等等上产生附加的网页几乎毫无成本。因此，聪明的链接兜售者能够通过这种技术在万维网上增加不受欢迎的网页显示。由于存在某些动态创建和服务无限数量页面的万维网服务器，网页的数量是无限的。很明显，增加创建另一个页面的成本相对低廉。

10 因此，本发明处罚对这种链接兜售信息技术的使用。假定每个链接兜售信息节点 LS1 到 LS9 都位于相同的主机 www.foo.com，在本发明的一个实施例中，代替给每个 LS1 到 LS9 分配一个最小分值，本发明给每个主机分配了一个最小分值，并在该主机的每个节点之间分配该最小分值。从而，不论在主机 www.foo.com 上存在 1000 链接兜售信息节点 LS1 到 LS1000，还是一个链接兜售

15 信息节点 LS1，分配给节点 N2 的签注都是相同的（假定所有的其它因数相同）。因此，如图 3E 所示例，本发明给主机 www.foo.com 分配了一个最小分值，并在页面 LS1 到 LS9 之间细分该最小值。

但是，被确定的链接兜售信息者能够瞒骗本发明的该变型。可以确定该链接兜售信息者，从而以相对小的量来增加与他或她的网页相关联的质量分值，

20 以使得分超过其它具有相似质量分值的网页。例如，如果在排序中页面被列表为第 5 到第 1 位，因此引起了典型搜索引擎排序中网页的增加，那么质量分值中的相对小的改变可能显著地受益于链接兜售信息者。在因特网域名系统的当前形式中，用户为每个域名付年费（例如 25 美元），并且能够在该域内创建一个任意数量的符号主机名。因此，链接兜售信息者能够获得小数量的域，并配

25 置 DNS 服务器来解析这些域内任何可能的主机名。兜售信息者然后能够提供链接兜售信息页面，该页面似乎来自这些域内许多不同的主机，籍此通过累积巨大数量页面的最小分值，而重新获得签注一个页面的能力。在图 3F 中示例了这种情形，其中通过被确定的链接兜售信息者已经产生了许多网页，每个网页来自唯一的符号主机名 www1.foo.com、www2.foo.com、www3.foo.com 等等，但

30 是，每个网页还来自于公共域 foo.com。因此，在本发明的第二实施例中，不是给每个网页或每个独立的主机名，而是给每个独立的域名分配最小分值。

正如所预见的，尽管域名花费了金钱，但是它们并不额外高昂。尽管存在无限数量的网页并且存在潜在的大量域名（37²⁵⁵），但是，实际上，只存在大约1千5百万个域名。（通过配置DNS服务器解析域内的任何可能主机名）创建一个新的主机名为免费的，而创建一个新的域名大约花费25美元。因此，由于用于大约25美元就能获得一个域名，因此对于被确定的链接兜售信息者实施图3G所示例的该系统是可行的。在图3G中，被确定的链接兜售信息者已经创建了链接兜售信息节点LS1到LS9等等，但是这时，链接兜售信息者已经在其域的每个节点上放置了其自己的符号主机名，如www.foo.com、www.goo.com、www.hoo.com等等。但是，类似地，链接兜售信息者已经在相同的IP地址放置了每个www.foo.com、www.goo.com、www.hoo.com等等。大约有四十亿个不同的IP地址。尽管IP地址的获取相对低廉，但是链接兜售信息者不可能实施一个系统，这是由于所提供的不同IP地址不是无限的，因此该系统给每个链接兜售信息网页分配一个不同的IP地址。相应地，在本发明的第三实施例中，不给每个网页、每个主机名或每个域，而是给每个不同的IP地址分配一个最小分值，以防御被确定的链接兜售信息者，因为不论链接兜售信息者在服务器上放置了具有链接兜售节点的100个不同域，还是该链接兜售信息者在该服务器上放置了具有一个链接兜售信息节点的1个不同域，就签注节点N2而言，效果是相同的（再次，保持所有的其它因数恒定）。

尽管本发明可应用于排序任何结构相关的文档，但是图4A（系统图）和4B和4C(流程图)示例了本发明示意性应用于为搜索引擎应用程序排序网页。由于本发明的算法独立于查询，因此用于根据本发明排序文档的过程可独立于应用程序诸如搜索引擎而发生，该过程请求的文档是基于分配给文档的分值。例如，如图4A所示，一个或更多的浏览器410能够浏览结构相关文档400，诸如因特网上的网页的来源，并提取文档或有关该文档的相关信息以存储在资料档案库420中。有关文档的相关信息还可以来自其它的来源415。

浏览器是一种访问万维网站点并阅读其网页和其它信息以为搜索引擎索引创建输入条目的程序。浏览器通过以下从服务器到服务器的超文本链接和基于搜索原则的索引信息，来定位新的文档和新的站点。

万维网上的主要搜索引擎都具有这样的一种程序，它还已知为“蜘蛛”、“蚂蚁”、“机器人”（“马蝇幼虫”）或“智能代理”。浏览器被典型地编程用于当其

被更新时访问其拥有者已经提交的站点。可以通过选择性地访问和索引输入的站点或具体页面。由于浏览器每次通过一个站点来浏览一个页面，因此它们明显获得了接下来链接到站点上的其它页面的名字，直到已经阅读完所有的页面为止。典型地，浏览器同时浏览许多的万维网站点。通常，对于万维网浏览器而言，浏览器遵守礼让的原则，即在 Robot 排他运算标准 (SRE) 中所规定的原则。

图 4B 借助流程图示例了收集文档的处理。在 450，发现网页并通过浏览器收集网页和/或有关网页的信息。如箭头所述，该收集可以为交互或正在进行的处理。在 460，在已知文档收集结构的时间的任何给定点上，可应用本发明分配一个分值给每个被收集的文档，以便存在网页（或其链接）的资料档案库，每个库具有使用 R^A 度量、 R^D 度量或 R^H 度量的相关联的质量分值。通过 API 432，对象 430 能够在分配一个分值给资料库 420 中的每个文档之后执行智能。如以下详细所述，还应该理解，能够将 R^A 度量、 R^D 度量和/或 R^H 度量于其它度量进行组合以增加分配给有关特定应用或用户的文档的质量分值。

一旦产生初始的文档和分值资料库 420，该库可被交互地、不断地、或周期性地更新，那么将能够应用搜索引擎的示意性应用。例如，搜索引擎（或其它应用）对象 440 可以在 470 接收用户的一个输入查询。在 480，基于该查询，能够根据其相关联的分值检索并排序包括符合该查询原则的查询项的网页，以便首先或主要地将最大质量分值的文档显示给用户。在示意性的实现中，如图 4A 所示，通过搜索引擎服务器对象 440 接收该查询，该对象通过 API 434 与对象 430 进行接口。对象 430 基于该查询通过 API 432，从资料档案库 420 检索并排序相关的网页（或与网页的链接）。然后，对象 430 返回结果的排序列表给应用程序对象 440，以用于将其显示给用户。可替换地，对象 440 能够根据相关联的分值对结果执行排序。

正如应该澄清的，还可以组合本发明的度量，或与整个度量的其它改进组合。例如，利用“信任”文档的基本组，能够改进利用任何度量所实现的分值的总质量。这种改进包括考虑 Nielsen 等级。例如，最高等级的 Nielsen 网页应该用作网页信任的基础，并根据该基础测量其它分值。而且，Nielsen 等级应用组合本发明的度量来使用以分配一个加权的分值。或者将本发明度量的结果与 Nielsen 等级比较以发现反常的结果。

本质上，有关值得信任文档诸如网页的信息的任何外部来源，能够组合本发明的度量结果使用，或者用于检验本发明的度量结果。其它例子包括使用基于从一个 ISP 所收集的用户信息的信息。例如，ISP 能够收集类似于 Nielsen 信息的有关有关用户访问特定网页频率的直接信息，并相应地分配一个质量、
5 声望或信任度、基于使用模式的分值。为此可以检查 ISP 代理日志。另一个例子包括在搜索引擎上观测人们。仅仅因为网页接收最高的分值并不意味着该用户最有可能选择该网页。因此，搜索引擎上的用户行为用于验证网页的质量。另外，能够分配人编辑器该验证例如 10000 个优秀网页的工作。当正在进行该工作时，能够增加这 10000 个网页的分值，以便通过本发明的度量将会增加其
10 签注能力，因为对于信任的网页而言它们是已知的。有关质量网页的另一个信息源来自用户的选项，诸如用户的书签。简而言之，能够将本发明的度量与任何其它已知的质量度量组合用于保证，提供最好的用户经历。有利的是，包括本发明度量的任何组合都将阻碍链接兜售信息者的努力。

存在多种实现本发明的方法，例如合适的 API、工具箱、驱动器代码、操作系统、控制、独立和可下载的软件对象，等等，以使应用程序和服务能够使用本发明的排序系统和方法。本发明打算根据 API（或其它软件对象）的立场、以及接收网页或与网页有关的结构信息的软件或硬件对象来使用本发明，以应用根据本发明的该排序技术。因此，这里描述的本发明的各种实现可能具有的方面是，本发明完全在硬件中，部分在硬件和部分在软件中，以及在软件中予以实现。
20

如上所述，尽管结合各种计算设备和网络结构已经描述了本发明的示意性实施例，但是基本的理论可应用于任何计算设备或系统中，其中在应用中期望排序结构相同的文档。例如，可将本发明的算法和硬件实现应用于计算设备的操作系统，提供作为设备上的单独对象、另一个对象的一部分、可再用的控制、
25 从服务器可下载的对象、设备或对象以及网络之间的“中间人”、分布式对象、硬件、提供在存储器中，或上述的任何组合等等。尽管这里选择了示意性的编程语言、名称和实例作为各种选择的代表，但是这些语言、名称和实例并不是限制性的。本领域的普通技术人员将会理解，有大量的提供对象代码和术语的方法，该术语实现了通过本发明的各种实施例所实现的相同、相似或等同的功能。
30

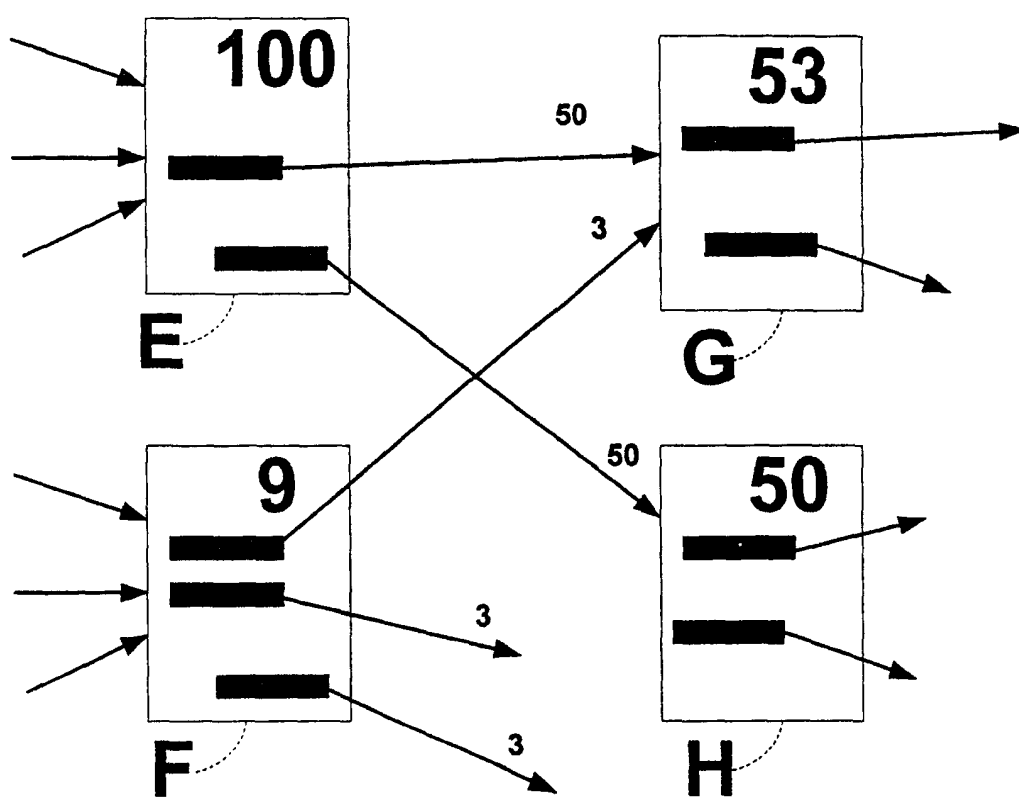
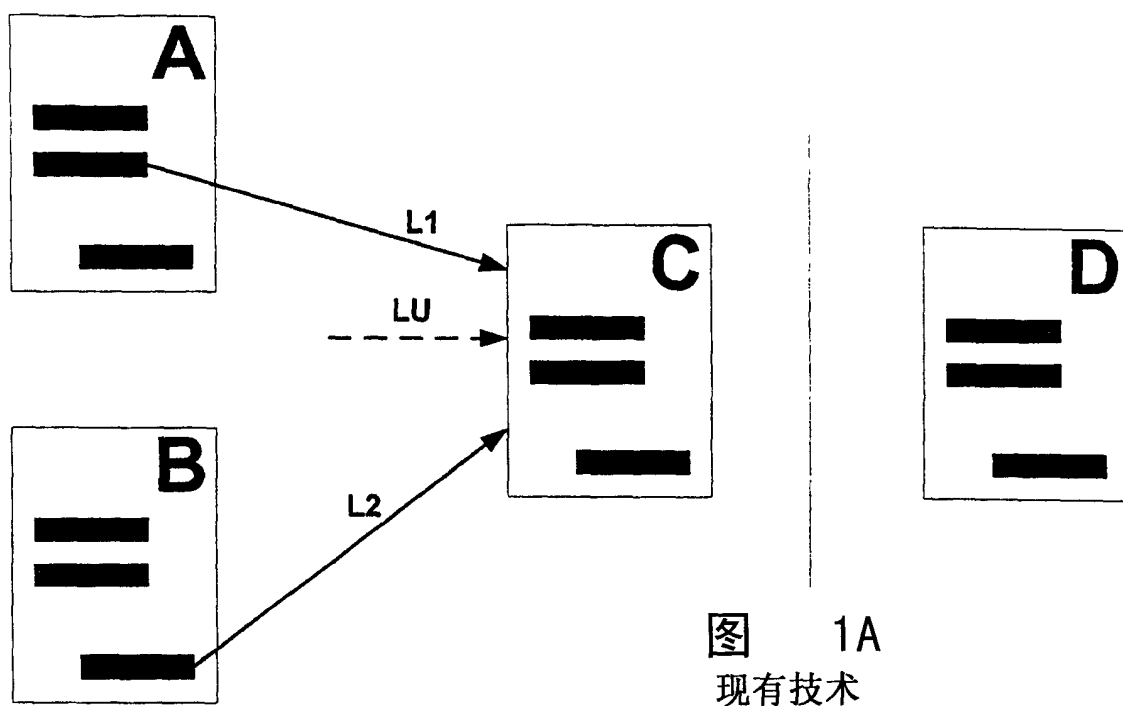
正如所述，这里描述的各种技术可结合硬件或软件或在适当的时候结合其组合而予以实现。因此，本发明的方法和设备，或本发明的特定方面或部分，可以采取程序代码（即，指令）的形式，这些代码嵌入到有形的介质中，诸如软盘、CD-ROM、硬盘、或任何其它的机器可读存储介质中，其中，当程序代码载入机器诸如计算机并由它执行时，该机器变为实践本发明的设备。在扩编程的计算机上执行该程序代码的情况下，计算设备通常包括一个处理器、处理器可读的存储介质（包括易失性和非易失性存储器和/或存储单元）、至少一个输入设备和至少一个输出设备。可以实现或使用本发明排序技术的一个或更多的程序，例如通过使用数据处理 API、可再用控制等等，优选地在高级程序或面向对象的编程语言中实现，以与计算机系统通信。但是，如果想要的话，这些程序可在汇编或机器语言中予以实现。在任何情况下，语言可以为编译或翻译语言，以及与硬件实现相组合。

本发明的方法和设备还可以通过以程序代码形式嵌入的通信来实践，该代码在某些传输介质上诸如电线或电缆，通过光纤或任何其它形式的传输介质传送，其中当接收程序代码并载入到机器中和由机器执行时，该机器诸如 EPROM、门阵可编程逻辑器件（PLD），客户端计算机等等，它变为实践本发明的设备。当在通用目的的处理器的实现本发明时，程序代码与处理器组合以提供唯一的操作用于调用本发明功能的设备。另外，结合本发明所使用的任何存储技术可以不变地作为硬件和软件的组合。

当结合各种附图的优选实施例已描述了本发明时，应该明白可以使用其它类似的实施例，或对上述用于执行本发明相同功能的实施例进行修改和添加，而不会偏离本发明。例如，当以联网环境诸如点对点的联网环境为背景描述本发明的示意性网络环境时，本领域的普通技术人员将会意识到，本发明并不限制于此，正如本申请所述的方法可以应用于任何计算设备或环境中，诸如游戏控制台、掌上电脑、便携式计算机等等，无论有线还是无线的，本申请可用于任何数量的通过通信网络连接的这种计算设备，这些计算设备在该网络上可交互作用。而且，应该强调指出，各种包括手持设备操作系统和其它专用操作系统的计算机平台，都意图特别地作为无线互连设备而继续扩展。

尽管示意性的实施例参照使用在万维网搜索服务背景下的本发明，但是本发明并不限制于万维网搜索服务，而是可以被实现用来提供一种对任何组文档

或内容的质量度量，其在某种方式相互参照。例如，用户可能在他的或她的计算机上（或在多个计算设备上）存储了一组图片、电影、音乐等等，它们以某种方式（通过人、地点、时间、事件、艺术家、选集、标题、演员等等）在结构上相互关联，以及根据内容的结构相互关系，本发明的分值可应用于这些图片。例如，第一内容和第二内容之间的相似性或其它的关系可以考虑为从第一内容到第二内容的链接。而且，当结合分配分值处理在以上描述 API 432 时，应该注意到为此可以实现单独的 API，即，基于一个不需要限制于分值处理的查询使用 API 432 进行检索和排序。而且，如这里使用的术语“与...成比例”指的是两个实体之间的任何数学关系，其中当一个实体增加时，另一个实体则根据所有已知的数学关系包括但不限于，几何、线性、指数、对数和其它关系而增加。相同的应用该术语“反比例”或“反比于...”即当一个实体增加时，另一个实体减小。另外，由于术语服务器指在各种背景下的各种事物，如这里各种使用的术语“万维网服务器”旨在至少指服务器计算机和/或服务器对象，该对象包括以下其中的任何一个或多个（A）具有相同符号主机名的许多网页，（B）与相同域相关联的许多网页，以及（C）与相同 IP 地址相关联的许多网页。进一步，可以在许多处理芯片或设备中或上实现本发明，以及在许多的设备上可类似地影响存储器。所以，本发明不应该限制于任何的单个实施例，而是应该限定在根据附属权利要求所规定的宽度和范围内。



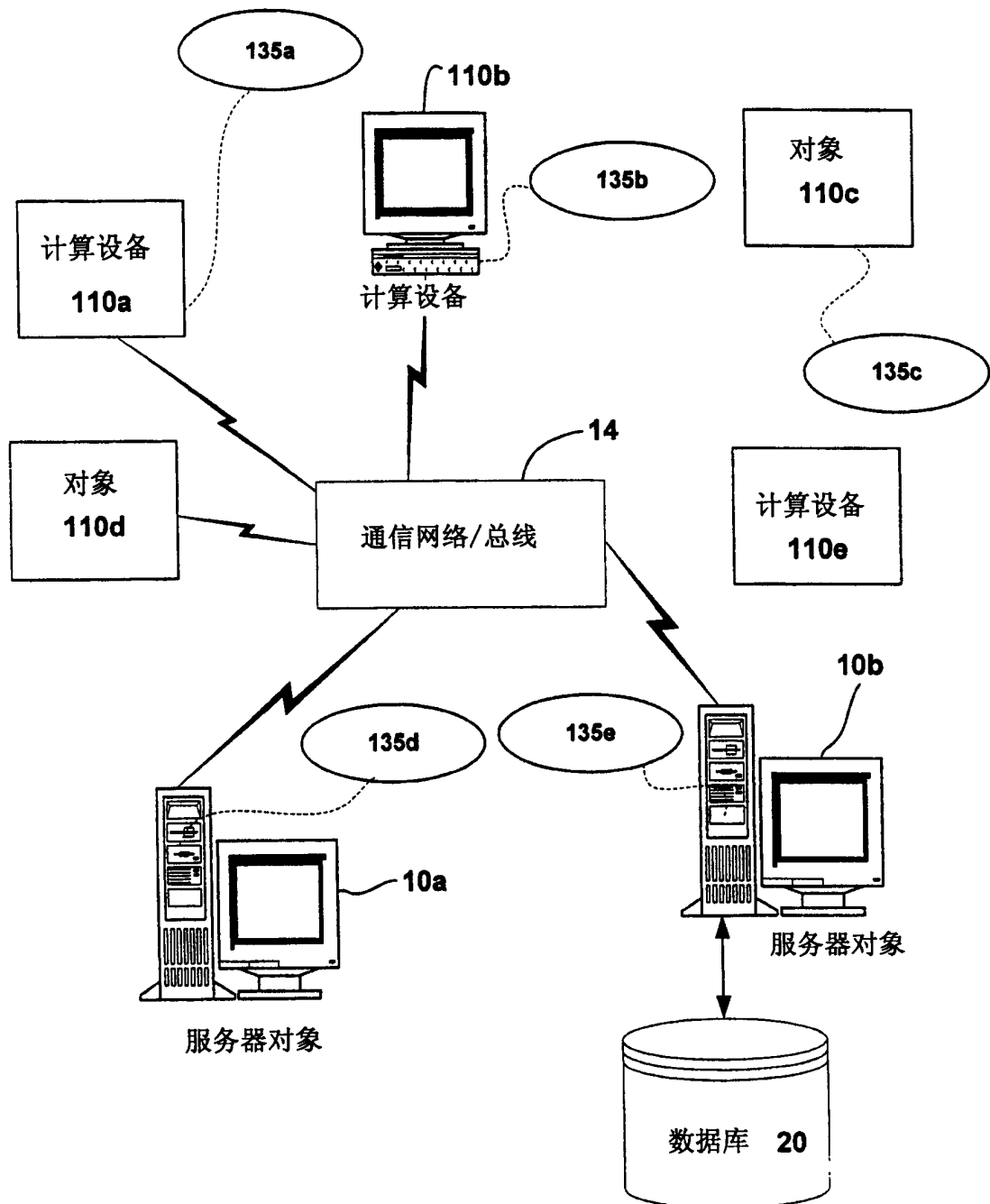


图 2A

计算环境 100

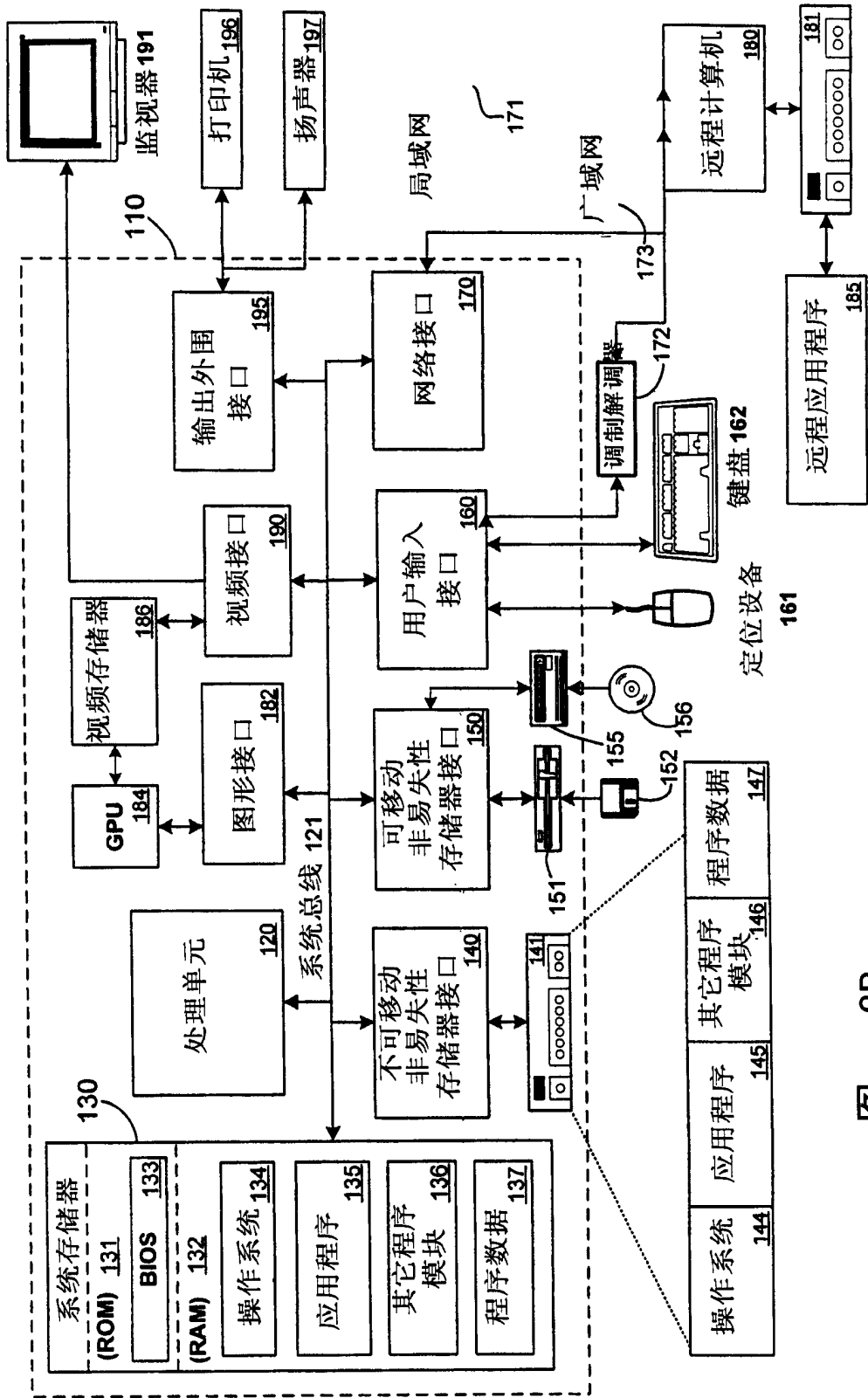


图 2B



图 3A

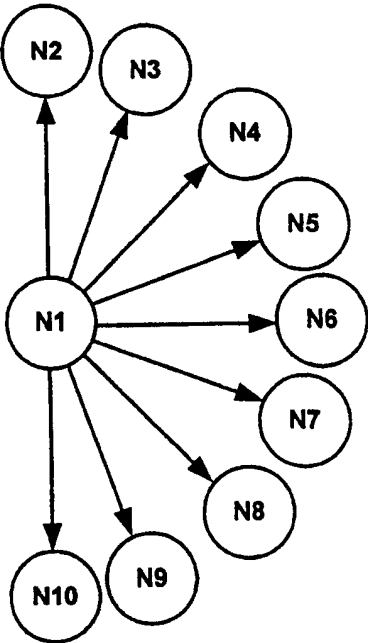


图 3B

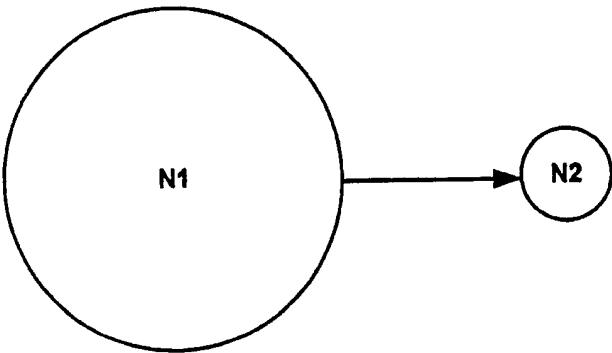


图 3C

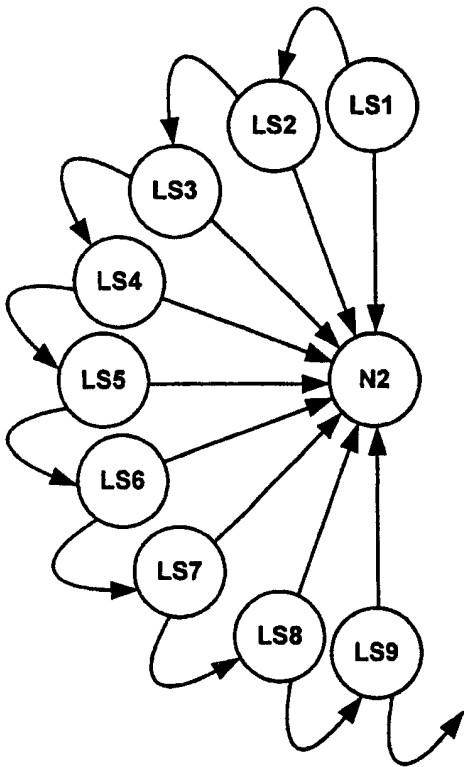


图 3D

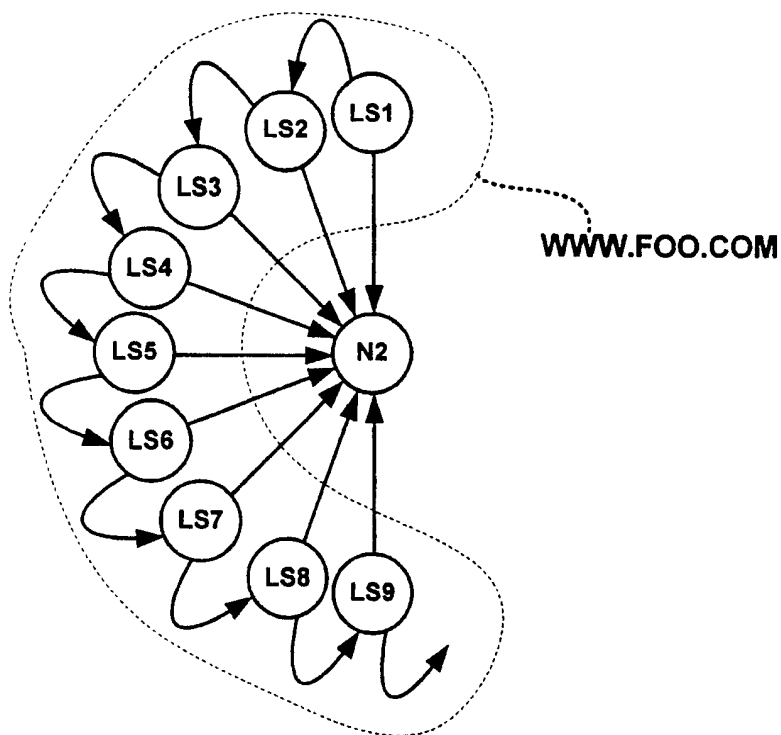


图 3E

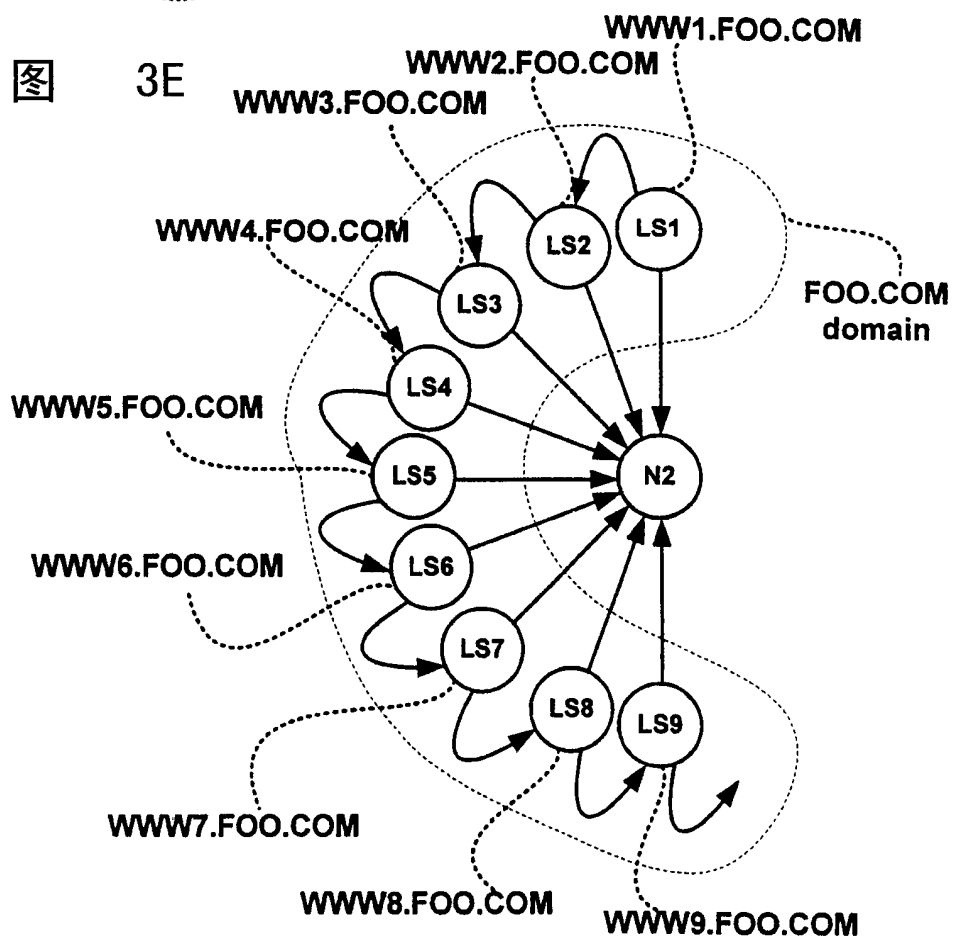


图 3F

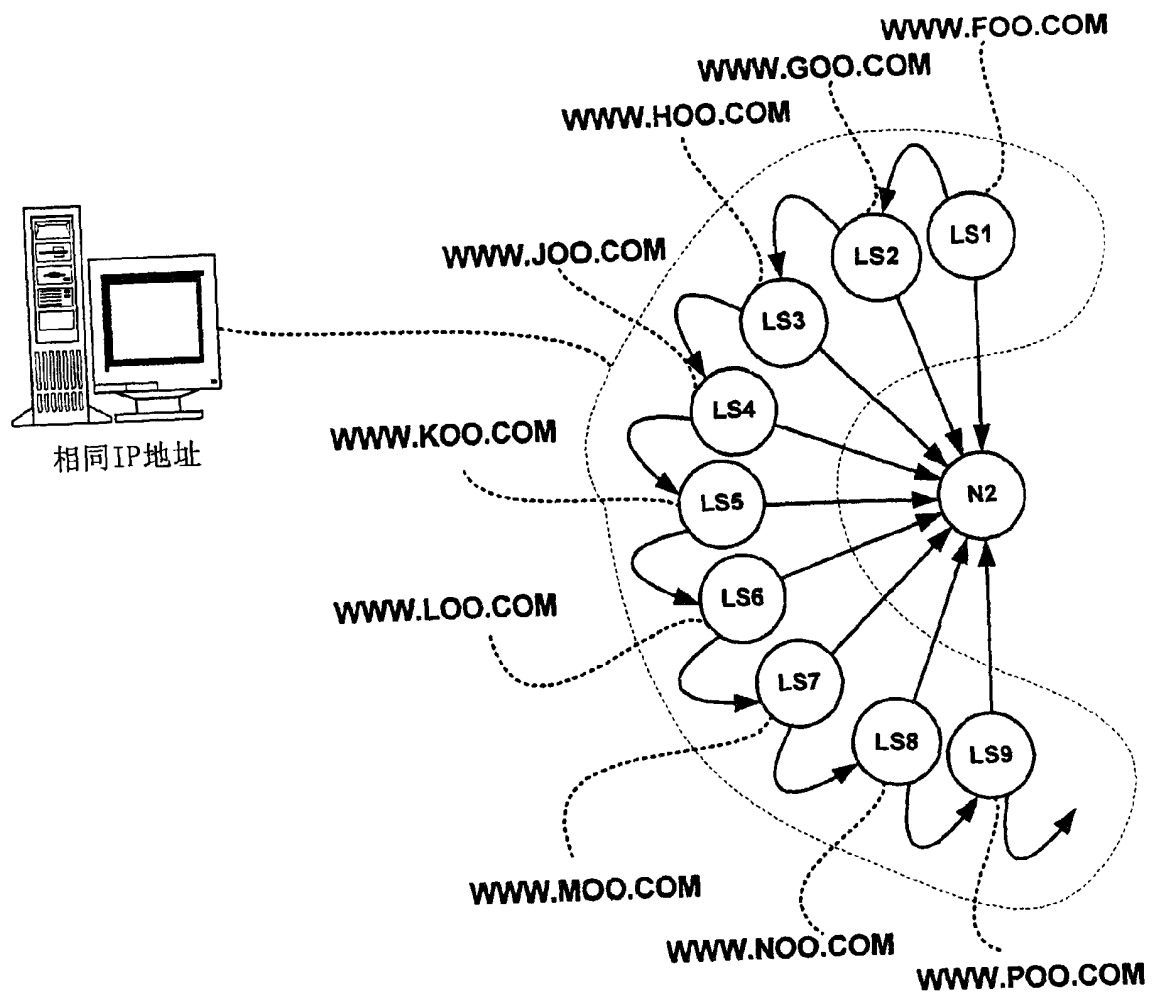


图 3G

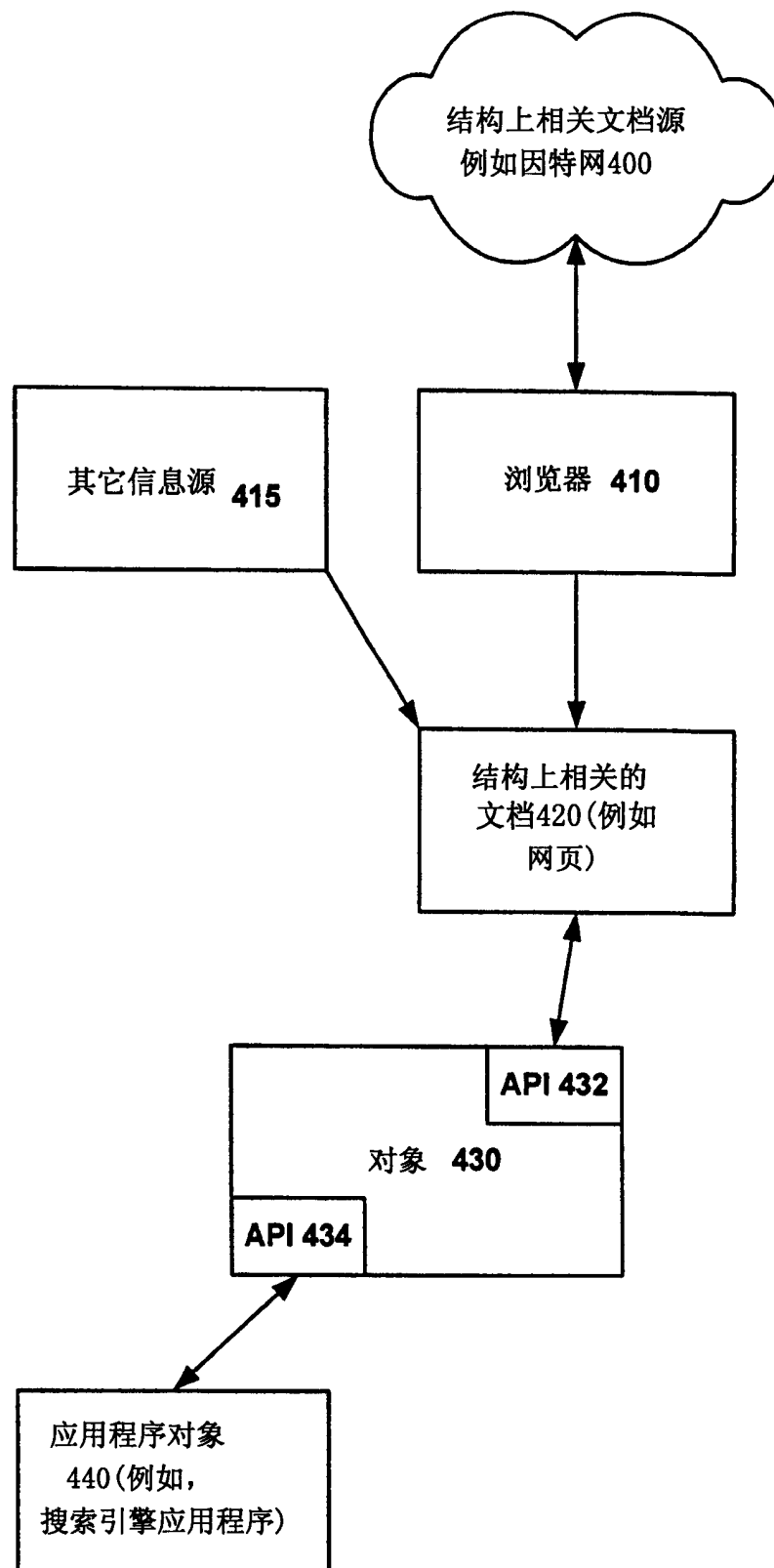


图 4A

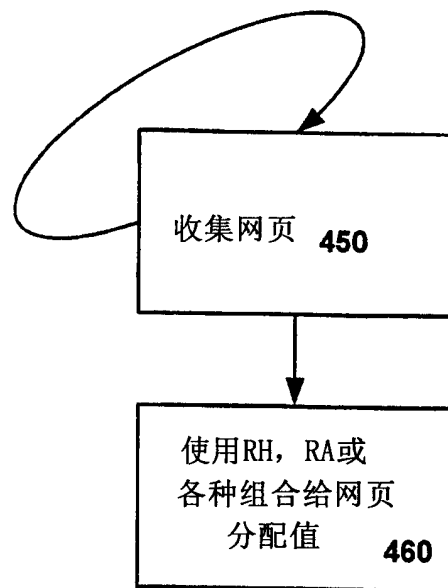


图 4B

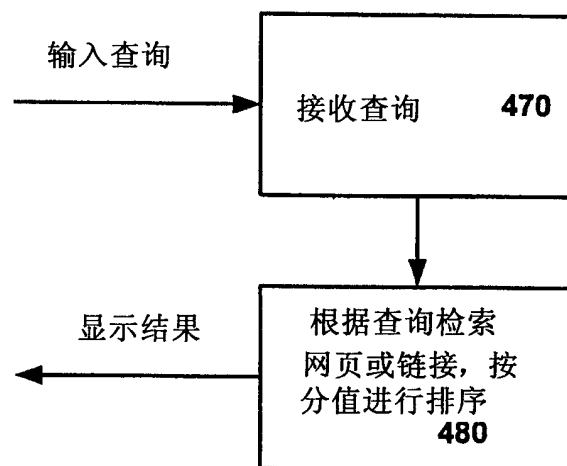


图 4C