

[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 200510103667.5

[43] 公开日 2006年4月12日

[11] 公开号 CN 1758248A

[22] 申请日 2005.9.5

[21] 申请号 200510103667.5

[30] 优先权

[32] 2004.10.5 [33] US [31] 10/958,560

[71] 申请人 微软公司

地址 美国华盛顿州

[72] 发明人 E·J·霍维茨 J·B·蒂凡

S·T·杜迈斯

[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

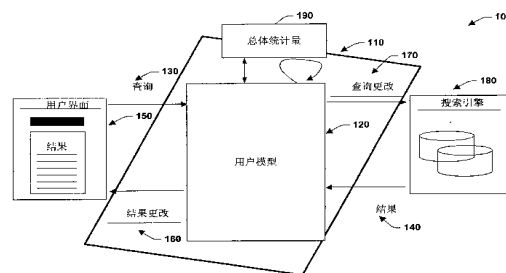
权利要求书 5 页 说明书 12 页 附图 15 页

[54] 发明名称

用于提供个性化搜索和信息访问的系统、方法和接口

[57] 摘要

本发明涉及采用用户模型根据与相应用户特征相关的信息来个性化一般查询和/或搜索结果的方法和系统。提供了便于产生信息的个性化搜索的系统。该系统包括一用户模型来确定用户的特征。用户模型可通过分析用户的内容、动作和整体环境来组合。个性化组件根据该用户模型自动更改查询和/或搜索结果，以便个性化用户的信息搜索。用户界面接收查询并显示来自一个或多个本地和/或远程搜索引擎的搜索结果，其中该界面可在较个性化的搜索到较一般化的搜索的范围内进行调整。



1. 一种便于产生信息的个性化搜索的系统，其特征在于，包括：
用户模型，用以确定用户的特征；
个性化组件，用以根据所述用户模型自动更改至少一个查询组件或至少一个搜索结果；以及
界面组件，用以接收所述查询并显示所述搜索结果。
2. 如权利要求 1 所述的系统，其特征在于，还包括一个或多个搜索引擎，用以接收所述查询并返回所述结果。
3. 如权利要求 1 所述的系统，其特征在于，还包括用户统计量数据的整体数据库，以便于更新所述用户模型。
4. 如权利要求 1 所述的系统，其特征在于，所述个性化组件对初始输入查询采用查询更改过程，通过所述用户模型来更改或重新产生所述查询，以从搜索引擎生成个性化结果。
5. 如权利要求 4 所述的系统，其特征在于，所述个性化组件采用相关性反馈，其中查询产生导向经有关初始结果集的显式或隐式判定而更改的查询的结果，以生成个性化结果。
6. 如权利要求 1 所述的系统，其特征在于，所述个性化组件采用结果更改，所述结果更改利用用户的输入原样来产生查询以生成结果，所述结果然后通过所述用户模型更改以产生个性化的结果。
7. 如权利要求 6 所述的系统，其特征在于，所述结果更改通常包括重新排序或从较大结果选择集中进行选择。
8. 如权利要求 6 所述的系统，其特征在于，所述结果更改包括所有或结果分集的聚集或归纳。
9. 如权利要求 1 所述的系统，其特征在于，所述个性化组件采用统计量相似性匹配，其中用户的兴趣和内容被表示为向量并对结果更改匹配。
10. 如权利要求 9 所述的系统，其特征在于，所述个性化组件采用类别匹配，其中用户的兴趣和内容使用较小的描述符集表示。
11. 如权利要求 1 所述的系统，其特征在于，所述个性化组件组合查询更改或结果更改，其中在所述两个更改中引入相关性并发挥作用。
12. 如权利要求 1 所述的系统，其特征在于，所述用户模型部分地基于计算环

境的历史，它可从本地、移动、或远程源中获取。

13. 如权利要求 12 所述的系统，其特征在于，所述计算环境包括应用程序打开、所述应用程序的内容、以及与所述应用程序的交互的详细历史的至少之一。

14. 如权利要求 1 所述的系统，其特征在于，所述用户模型至少部分地基于先前相遇的内容的索引，包括文档、Web 网页、电子邮件、即时消息、短信、以及日程安排预约的至少之一。

15. 如权利要求 1 所述的系统，其特征在于，所述用户模型至少部分地基于客户机交互，包括最近或频繁的联系、从关键字导出的兴趣标题、组织图中的关系、以及预约的至少之一。

16. 如权利要求 1 所述的系统，其特征在于，所述用户模型至少部分地基于先前 Web 网页或曾访问的本地/远程数据站点的历史或日志，包括先前搜索查询的历史。

17. 如权利要求 1 所述的系统，其特征在于，所述用户模型至少部分地基于用户在一段时间内访问的并由确定有关用户位置的信息的装置监视的位置历史或日志。

18. 如权利要求 17 所述的系统，其特征在于，所述装置包括确定用户位置的全球定位系统 GPS 或电子日程表。

19. 如权利要求 18 所述的系统，其特征在于，所述装置产生转换成文本城市名和邮编的空间信息。

20. 如权利要求 19 所述的系统，其特征在于，所述空间信息被转换成用户曾停留、居住或引起 GPS 信号丢失的位置的文本城市名和邮编。

21. 如权利要求 20 所述的系统，其特征在于，用户曾停留、居住或引起 GPS 信号丢失的所述位置可通过商务和兴趣点的数据库标识和转换成文本标签。

22. 如权利要求 21 所述的系统，其特征在于，所述位置可从一天中的时间或一周内的日期来确定。

23. 如权利要求 1 所述的系统，其特征在于，所述用户模型至少部分地基于可显式或隐式指定的用户兴趣的概况文件。

24. 如权利要求 1 所述的系统，其特征在于，所述用户模型至少部分地基于人口统计信息，包括位置、性别、年龄、背景和工作类别的至少之一。

25. 如权利要求 1 所述的系统，其特征在于，所述用户模型至少部分地基于协作性过滤和机器学习算法的至少之一。

26. 如权利要求 25 所述的系统，其特征在于，所述机器学习算法包括贝叶斯网络、自然贝叶斯分类器、支持向量机器、神经网络和隐藏马尔可夫模型的至少之一。

27. 如权利要求 1 所述的系统，其特征在于，所述个性化组件包括对控制结果或查询个性化的调整。

28. 一种具有存储其上的计算机可读指令的计算机可读介质，所述计算机可读指令用于执行如权利要求 1 所述的组件。

29. 一种客户机组件包括如权利要求 1 所述的系统。

30. 一种信息检索系统，其特征在于，包括：

用于建模用户的特征的装置；

用于查询并显示用户的搜索的搜索结果的装置；以及

用于至少部分地基于所述用户特征来更改所述搜索结果。

31. 如权利要求 30 所述的系统，其特征在于，还包括用于与至少一个搜索引擎交互的一装置。

32. 一种便于用户界面上信息搜索的方法，其特征在于，包括：

定义自动确定用户感兴趣参数的至少一个用户模型；

至少部分地基于所述用户模型来自动提炼查询或查询的结果；以及

在向用户显示更改的结果之前，根据所述用户模型自动格式化所述查询或结果。

33. 如权利要求 32 所述的方法，其特征在于，所述用户模型包括用户先前查看的条目的索引，包括电子邮件、文档、Web 网页、日程安排预约、短信、即时消息、和 blog 的至少之一。

34. 如权利要求 33 所述的方法，其特征在于，还包括用元数据来标记所述条目，所述元数据包括访问/创建/更改的时间、条目的类型、条目的作者的至少之一，所述条目可用来有选择地包括/排除用于比较的条目。

35. 如权利要求 33 所述的方法，其特征在于，还包括计算所述结果与用户索引的相似性，用以标识用户更感兴趣的结果。

36. 如权利要求 35 所述的方法，其特征在于，还包括确定相似性的以下等式：
个性化相似性

$$psim = \sum score_i$$

其中个性化相似性在所有感兴趣条目上叠加，对于每个项，结果的相似性与

置于该项出现上的值 $score_i$ 成正比。

37. 如权利要求 36 所述的方法, 其特征在于, $score_i = (tf_i / df_i) * pdf_i$, 与所述项在结果中的出现频率 tf_i 成正比, 与所述项出现在结果中的次数(df_i)成反比, 并与所述项出现在用户索引中的多少条目 pdf_i 成正比。

38. 如权利要求 36 所述的方法, 其特征在于, 所述感兴趣的项包括结果标题中的项、结果摘要中的项、扩展结果摘要中的项、全 Web 网页中的项、所述项的子集的至少之一。

39. 如权利要求 38 所述的方法, 其特征在于, 还包括标识来自以标题或结果摘要中每个查询项的多个词组成的窗口中的各项。

40. 如权利要求 35 所述的方法, 其特征在于, 还包括将条目的标准相似性与所述条目的个性化相似性相组合。

41. 如权利要求 40 所述的方法, 其特征在于, 还包括采用原始结果列表中条目的秩与每个条目的个性化相似性分数的标准化版本的线性组合。

42. 如权利要求 36 所述的方法, 其特征在于, 还包括采用相关性反馈算法来确定相似性 $score_i$ 。

43. 如权利要求 42 所述的方法, 其特征在于, 所述相关性反馈算法是 BM25 算法。

44. 一种执行信息检索的图形用户界面, 其特征在于, 包括:

输入组件, 用于接收查询;

显示组件, 用以显示查询的结果; 以及

个性化组件, 用以根据确定用户喜好的用户模型来更改所述查询或所述结果。

45. 如权利要求 44 所述的图形用户界面, 其特征在于, 还包括一控件, 用于根据从标准化搜索到个性化搜索的范围来提炼所述查询或所述结果。

46. 如权利要求 45 所述的图形用户界面, 其特征在于, 所述个性化搜索与具有文本或色彩增强的显示相关联。

47. 一种便于产生信息的个性化搜索的系统, 其特征在于, 包括:

用户模型, 用于确定用户特征;

与所述用户模型相关联的个性化组件; 以及

参数组件, 用以控制所述用户模型的数据全集。

48. 如权利要求 47 所述的系统, 其特征在于, 所述数据全集与用户预约、用户对文档的查看、用户动作或用户位置相关。

49. 如权利要求 47 所述的系统，其特征在于，所述参数组件确定所述数据全集的子集或确定匹配过程中的加权差异，用于至少部分地基于类型或年龄进行数据的个性化。

50. 如权利要求 47 所述的系统，其特征在于，所述参数组件通过最优化过程或通过用户界面提供的指令来更改一个或多个参数。

51. 如权利要求 50 所述的系统，其特征在于，所述参数是查询本质、一天内的时间、一周内的日期、基于环境的观察、或基于动作的观察的函数。

用于提供个性化搜索和信息访问的系统、方法和接口

技术领域

- 5 本发明一般涉及计算机系统，尤其涉及根据个性化用户模型自动提纯并聚焦搜索查询和/或结果。

背景技术

10 由于万维网和因特网极为流行，用户可从大量信息源获取有关几乎任何主题的信息。为了寻找信息，用户一般将各种搜索引擎应用于信息检索任务。搜索引擎使用户能找到包含信息的 Web 网页和包含特定字或短语的因特网上其它材料。例如，如果想要寻找有关美国第一任总统 George Washington 的信息，他们可键入“George Washington First President”，点击搜索按钮，然后搜索引擎将返回包含有关该著名总统的信息的 Web 网页列表。然而如果进行更一般的搜索，诸如仅键入
15 “George Washington”，将返回更多结果，诸如有关与同一名字相关联的地区或机构。

Web 上有很多搜索引擎。例如，AllTheWeb、AskJeeves、Google、HotBot、Lycos、MSN Search、Teoma、Yahoo 仅仅是许多示例的一部分。大部分这些引擎至少提供搜索信息的两种模式，诸如通过按用户要浏览的主题组织的它们自己的网站目录，
20 或通过执行经浏览器上用户界面入口输入的关键字搜索。一般而言，关键字搜索将尽计算机所能找到其中具有与指定的所有关键字和短语相关的任何信息的所有 Web 网站。搜索引擎网站将具有让用户输入关键字的框以及点击它以开始搜索的按钮。许多搜索引擎具有关于如何使用关键字来有效搜索的提示。提示通常有助于用户在更小范围内定义搜索词从而不返回扰乱信息检索过程的额外或不相关的信息。
25 因而，通过帮助减轻在寻找特定信息时接收数千个网站进行搜索的困难，手动缩小词的范围为用户节约了很多时间。

所有搜索技术的一个问题是需要手动聚焦搜索词或缩小搜索词的范围以便在短时间内生成所需结果。另一个问题是无论不同用户的需要和环境怎样搜索引擎对所有用户都进行一样的操作。因而，如果两个用户输入同样的搜索查询他们会得到
30 同样的结果，而不管他们的兴趣、先前的搜索历史、计算环境、或周围环境（例如

位置、所使用的机器、一天内的什么时间、一周内的哪一天)如何。不幸的是,现代搜索过程被设计成用于接收明确的关于搜索的命令,而不是考虑这些其它的人性化因素,这些因素可洞察用户的实际或需要的信息搜索目标。

5 发明内容

以下呈现了本发明的简化概述,以便提供对本发明某些方面的基本理解。该概述不是本发明的扩展性纵览。它并非旨在标识本发明的关键或主要元素,或旨在呈现本发明的范围。其唯一目的是以简化形式呈现了本发明的某些概念,作为以下提供的更详细描述的前言。

10 本发明涉及通过采用用户模型来增强信息检索方法的系统和方法,这些用户模型通过考虑信息如何与相应用户最相关来促成根据用户特征个性化信息搜索。模型可与传统的搜索算法相结合来更改搜索查询和/或更改搜索结果,以便自动地将信息检索方法聚焦到根据用户的个性特征更可能与用户相关的条目或结果。通过考虑诸如用户的内容(例如存储在用户计算机上的信息)、兴趣、专业、以及产生其
15 信息需要(例如搜索查询、计算事件)的特定环境等方面,提供了各种用于通过模型进行个性化搜索的技术来改进用户的搜索体验。通过根据用户的个性化喜好向用户提供更集中或经过滤的对感兴趣条目的搜索、移除不相关条目、和/或将返回的搜索结果重新排序,可看到这种改进。

20 用户模型可从多个源导出,包括考虑过去的用户事件、先前的客户机交互、搜索和历史日志、用户概况文件、统计数据的丰富索引,和/或可基于与其它用户的相似性(例如协作过滤)。此外,诸如机器信息的其它技术可用来监视用户的长期行为以确定和/或提炼用户模型。模型可与离线或在线搜索方法(或其组合)结合以更改搜索结果来产生最可能引起相应用户兴趣的信息检索输出。因而,用户模型被用来以自动化和有效的方式来区分个性化搜索与一般化搜索。

25 在一特定示例中,一般化搜索可包括词“weather”(天气)。因为模型可确定用户来自某城市(例如从电子邮件帐号、列出用户地址的存储文档、或通过位置的显式或隐式指定),可自动创建个性化搜索(例如通过自动化查询和/或结果更改),它返回有关用户所在城市的天气相关信息。在移动情形中,搜索环境可能不同从而可相应地更改查询和/或结果(例如从用户的移动计算机上进行的搜索,从最近的
30 机票订购或与朋友的即时消息中当前环境检测为不在该城市)。可提供各用户界面,它们返回个性化结果并使能个性化搜索算法从较一般的范围内搜索到更个性化的

搜索的调整。

为实现前述及相关结果，本方面的某些说明性方面结合以下说明书和附图在此说明。但这些方面仅仅示出了本发明诸原理可在其中采用的各种方式的其中几种，且本发明旨在包括所有这些方面及其等效方案。结合附图参阅本发明的以下详细
5 说明，本发明的其它优点和新颖特征会变得显而易见。

附图说明

图 1 是根据本发明一方面示出信息检索体系结构的示意框图。

图 2 是根据本发明一方面示出用户模型的框图。

10 图 3 是根据本发明一方面示出信息检索过程的流程图。

图 4-9 根据本发明一方面示出一示例用户界面。

图 10-13 根据本发明一方面示出一示例个性化算法。

图 14 是根据本发明一方面示出适当操作环境的示意框图。

图 15 是本发明可与之交互的样本计算环境的示意框图。

15

具体实施方式

本发明涉及采用用户模型来根据与相应用户相关的信息个性化一般查询和/或搜索结果的系统和方法。在一方面，提供了便于产生信息的个性化搜索的系统。该系统包括一用户模型来确定用户的特征。个性化组件根据该用户模型自动更改查询
20 和/或搜索结果，以便个性化用户的信息搜索。用户界面组件接收查询并显示来自一个或多个本地和/或远程搜索引擎的搜索结果，其中该界面可在较个性化的搜索到较一般化的搜索的范围内进行调整。

当用于本申请时，术语“组件”、“服务”、“模型”和“系统”旨在指计算机相关实体：硬件、硬件和软件的组合、软件、或执行中的软件。例如，组件可以是，
25 但不限于，运行于处理器上的过程、处理器、对象、可执行组件、执行线程、程序和/或计算机。作为说明，运行于服务器上的应用和服务器都是组件。一个或多个组件可驻留于执行的过程和/或线程中，且组件可在一台计算机上本地化和/或分布在两台或多台计算机之间。当在此使用时，术语“推断”一般指从通过事件和/或数据捕捉的一系列观察中推理和/或推断系统、环境、和/或用户的状态的过程。
30 推断可用来标识特定环境或动作，或可产生例如状态的或然性分布。

开始时参看图 1，系统 100 根据本发明一方面示出信息检索体系结构。系统

100 示出用于个性化搜索结果的一般示意图。个性化组件 110 包括用户模型 120 以及处理组件（例如根据用户模型更改的检索算法），该处理组件使用该模型来通过更改查询 120 或更改从搜索中返回的结果 140 影响搜索结果。用户界面 150 产生查询 130，并接收已基于查询更改 170 更改或个性化的结果和/或由个性化组件 110 提供的结果更改 160。当在此使用时，术语“查询更改”指有关查询 130 中术语的更改以及使查询 130 与文档匹配以便获取个性化结果 140 的算法的更改。已更改的查询和/或结果 140 从一个或多个本地和/或远程搜索引擎 180 返回。可维护用户统计数据的整体数据库，以便于更新用户模型 120。

一般而言，至少有两种基于用户模型 120 来调整搜索结果的方法。在一方面，查询更改处理初始输入查询，并更改或重新产生查询（通过用户模型）以生成个性化结果。以下所述的相关性反馈是该过程的双循环变体，其中查询产生结果，这些结果导向产生个性化结果的更改查询（使用有关初始结果集的显式或隐式判断），而这些个性化结果是基于查询和结果集根据短期模型来个性化的。长期用户模型也可在相关性反馈环境中使用。此外，如上所述，查询更改还指用来匹配查询和文档的算法的更改。在另一方面，结果更改取用户输入的原样来产生一查询，以生成（通过用户模型）更改而生成个性化结果的结果。注意，结果的更改通常包括某些形式的重新排序和/或更大选项集中的选择。更改或结果还可包括所有或结果子集的各类聚集和归纳。

用于更改结果的方法包括统计学相似性匹配（其中用户的兴趣和内容表示为向量并与条目匹配）、以及类别匹配（其中用户的兴趣和内容使用较小的描述符集表示为条目并与其匹配）。查询更改或结果更改的以上过程可独立地组合，或在将相关性引入两个过程之间并发挥作用的集中处理过程中组合。为了说明个性化搜索，提供了以下示例。

在一示例中，搜索者位于西雅图。对交通信息的搜索返回有关西雅图交通的信息，而不是一般性的交通。或者，对匹萨饼的搜索仅返回有关用户的适当邮编内的匹萨店。

在另一示例中，搜索者先前已搜索过词 Porsche（保时捷）。对 Jaguar（美洲虎）的搜索返回有关轿车美洲虎的结果，而不是动物或计算机游戏或手表；也可返回其它结果但优先考虑有关车的结果。

在另一情形中，搜索者查询“Bush”（布什）且大多数结果是有关总统的。然而，此人先前曾阅读过 Vannevar Bush 的文章并与 Susan Bush 通过电子邮件，则匹

配这些条目的结果将给予较高优先级。可以理解，假设数据由用户模型 120 存储和处理，可用多种不同方式来更改搜索，该用户模型在下面参照图 2 进行更详细的描述。

参照图 2，根据本发明一方面示出了用户模型 200。该用户模型 200 用来区分个性化搜索和一般化搜索。成功的个性化的一方面是要建立准确反映用户的兴趣且便于维护并适应有关长期和短期兴趣的变化的用户模型。该用户模型可从各种源中获取，这些源包括但不限于：

1)来自 210 的计算环境的丰富历史，可包含本地、移动、或远程源（例如应用程序的打开、那些应用程序的内容、以及这种交互的包括位置的详细历史）。

10 2)来自 220 的先前相遇内容的丰富索引（例如文档、Web 网页、电子邮件、即时消息、短信、日程安排预约等）。

3)来自 230 的监视客户机交互，包括最近或频繁的联系入、从关键字导出的兴趣标题、组织图中的关系、预约等。

15 4)来自 240 的先前 Web 网页或曾访问的本地/远程数据点的历史或日志，包括先前搜索查询的历史。

5)来自 250 的用户兴趣的概况文件，它可显式指定或通过背景监视隐式导出。

6)来自 260 的统计信息（例如位置、性别、年龄、背景、工作类别等）

从以上示例中，可以理解用户模型 200 可基于许多不同的信息源。例如，模型 200 可源自如由诸如全球定位系统（GPS）的装置监视的用户在一段时间内访问的位置的历史或日志。当用 GPS 监视时，原始的空间信息可转换成文本城市名和邮编。原始的空间信息可转换成用户曾停留、居住或引起 GPS 信号丢失的位置的文本城市名和邮编。用户曾停留、居住或引起 GPS 信号丢失的位置可通过商务和兴趣点的数据库标识和转换成文本标签。其它因素包括记录一天内的时间或一周内的日期以确定位置和兴趣点。

25 在本发明其它方面中，可提供组件来处理参数，用于控制如何将用户的信息群、预约、文档或文件的查看、动作、或位置分成子集或在匹配过程中如何不同地加权以基于类型、年龄、或其它组合个性化。例如，检索算法可限于有关该查询的用户全集的那些方面（例如包含查询项的文档）。类似地，可分析前一个月内的电子邮件、前 3 天内的 web 访问、以及过去一年内创建的用户内容。可能需要仅使用今天或其它时间段的 GPS 位置信息。可自动处理参数以创建子集（例如通过改变参数并测试来自用户或系统的响应的最优化过程）或者用户可通过用户界面更改

一个或多个这些参数，其中这种设置可以是查询本质、一天内的时间、一周内的日期、或其它环境或基于动作的观察的函数。

在 270 可诸如通过协作性过滤（如下所述）技术导出个人或个人分组的模型，该技术通过分析个人或个人分组之间的相似性来开发概况文件。相似性计算可基于
5 条目的内容和/或使用。注意，建模基础结构以及相关处理可驻留于客户机、多个客户机、一个或多个服务器、或服务器和客户机的组合。

在 280，可应用机器学习技术来学习一段时间内用户的特征和兴趣。学习模型基本上可包括任何类型的系统，诸如用于建模用户并确定喜好和兴趣的统计/数学模型和过程，包括使用可产生诸如贝叶斯网络、自然贝叶斯分类器、和/或包括例
10 如支持向量机器（SVM）的其它统计学分类方法的贝叶斯依从性模型的贝叶斯学习。其它类型的模型或系统可包括例如神经网络和隐藏马尔可夫模型。尽管根据本发明可采用具体的推理模型，可以理解也可使用其它方法。例如，不采用更彻底的或然性方法，而可采用确定性假设（例如根据规则最近对特定网站没有搜索 x 次可暗示用户对相关信息不再感兴趣）。因而，除了作不确定性推理之外，还可作有关
15 用户的状态、位置、环境、兴趣、焦点等的逻辑判定。

可从收集或聚集来自多个不同数据源的数据的用户事件数据库（未示出）中训练学习模型。这种源可包括记录或记下用户事件数据（例如蜂窝式电话、通过话
筒记录的声音动作、全球定位系统（GPS）、电子日程表、可视监视设备、桌面动作、网站交互等）的各种数据获取组件。注意，系统 100 基本上可用支持个性化查
20 询和结果处理的任何方式来实现。例如，系统可实现为服务器、服务器场、可在客户机应用程序内实现、或更一般地包括 web 服务或其它与诸如用户界面 150 和搜索引擎 180 的搜索功能交互的自动化应用程序。

在处理之前，在 270 应用的用户模型 200 的协作性过滤技术将更详细地进行描述。这些技术可包括采用协作性过滤来分析数据并确定用户的概况文件。协作性
25 过滤系统一般使用有关用户喜好的集中式数据库以预测用户可能想要的其它主题。根据本发明，协作性过滤应用于用户模型 200 以处理来自一组用户的先前用户动作，这些用户动作可指示给定用户的喜好，而这些喜好预测系统新用户的可能概况文件。可采用包括基于相关系数、基于向量的相似性计算、以及统计学贝叶斯方法的技术的若干算法。

30 图 3 根据本发明示出信息检索方法 300。尽管为了简单解释，方法被示为和描述为一系列动作，可以理解和了解本发明并未不受该动作顺序限制，因为根据本发

明某些动作可以不同顺序发生和/或与在此所示和所述的其它动作同时发生。例如，本领域技术人员将理解和了解方法可任选地表示为一系列相互关联的状态或事件，诸如状态图。此外，并非所有所示动作都是实现根据本发明方法所必须的。

5 显式或隐式获取的有关用户兴趣的信息可用各种方法采用，并可用查询特定方式来应用各类算法。许多算法考虑用户的个人内容和/或动作和/或查询和/或从就近的搜索引擎返回的结果，并考虑这种内容和整体内容之间的统计关系的度量或度量代理。

过程 300 示出可采取的两种基本路径，然而如上所述基于查询的更改和基于结果的更改的组合可用于个性化检索信息。在 310，一个或多个用户模型被确定为
10 如上参照图 2 所述。在 320，用户查询根据在 310 确定的模型进行更改。这可包括自动提炼查询或缩小查询范围，该查询对由模型确定的与用户兴趣相关的项进行。在 330，通过将已更改查询发送给一个或多个搜索引擎来由已更改查询执行搜索，其中来自已更改查询的结果在 340 返回。

在过程 300 的另一分支中，通过将用户的查询发送给一个或多个搜索引擎来
15 在 350 执行搜索。所返回的结果在 360 根据用户模型进行更改。这可包括基于所需搜索信息的某些结果更符合用户喜好的似然性来过滤或重新排序结果。在 370，已更改的结果通过用户界面显示器向用户呈现。

以下讨论描述已作为原型的个性化搜索系统的一个特定示例。用户模型可包括用户先前曾查看的所有条目的索引，包括电子邮件、文档、Web 网页、日程安
20 排预约、短信、即时消息、blog 等。条目可用元数据（例如访问/创建/更改的时间、条目的类型、条目的作者等）来标记，该元数据可用来有选择地包括/排除用于开发用户模型的条目。在此情形中，用户模型驻留在客户机上，其中基于利用搜索引擎从客户机内的数据存储中访问用户模型。

因为用户模型通常在客户机上运行，除非该客户机具有进行搜索的全集的本地索引，用于重新排序的全集范围的项统计量会难以计算或计算得太慢。因此，在
25 以下示例中，全集统计量由使用结果集来近似。

查询被导向搜索引擎（因特网或内联网）并返回结果。结果通过用户模型进行更改。更改还可在客户机上进行。对于每个结果，计算该条目与用户索引的相似性，以标识用户更感兴趣的结果。有若干种执行这种匹配的方法，诸如：

30 个性化相似性等式
$$psim = \sum_{t \in \text{terms_of_interest}} (tf_i / df_i) \cdot pdf_i$$

个性化相似性在所有感兴趣的项上叠加。对于每个项，结果的相似性与该项多长时间在结果中出现一次(tf_i)成正比，与在全集中搜索的出现该项的文档数量(df_i)成反比，并与该项出现在用户索引中的多少文档(pdf_i)成正比。感兴趣的项可包括，结果标题中的项、结果摘要中的项、扩展结果摘要中的项、全 Web 网页中的项、
5 或这些项的部分子集。全集中出现该项的文档数量可用结果集中出现该项的文档数量来近似，其中文档由文档的全文本或描述该文档的结果集片断来表示。

一个实现以标题或结果摘要中每个查询项组成的双字窗口来标识各项。一般而言，索引内的所有条目无论其类型或时间都被用来计算每个结果的个性化相似性度量。然后每个条目的标准相似性与每个条目的个性化相似性组合。一个实现采用
10 原始结果列表中条目的秩和每个条目的 $psim$ 分数的标准化版本的线性组合。其它实现包括组合来自原始和个性化列表的秩，或组合来自原始和个性化列表的分数。

现在参看图 4-9，根据本发明一方面示出用于个性化搜索的示例用户界面。注意，可在各种其它不同设置和环境中提供所述相应界面。作为一个示例，在此讨论的应用程序和/或模型可与例如桌面开发工具、邮件应用程序、日程安排应用程序、
15 和/或 web 浏览器相关联，尽管可使用其它类型的应用程序。这些应用程序可与图形用户界面 (GUI) 相关联，其中该 GUI 提供具有一个或多个显示对象 (未示出) 的显示，包括诸如可配置图标、按键、滑块、输入框、选项、菜单、选项卡等方面，它们具有多个可配置的尺寸、形状、色彩、文本、数据和声音以便于用应用程序和/或模型进行操作。此外，GUI 和/或模型还可包括多个其它输入或控件，用于调整
20 和配置本发明的一个或多个方面并如下详细描述。这可包括接收来自设备、键盘、语音输入、网站、远程 web 服务、和/或诸如照相机的其它装置的用户命令或视频输入，以影响或更改在此所述的 GUI 和/或模型的操作。

图 4 示出用于呈现个性化结果的界面 400。在此示例中，查询是“Bush”。标准的搜索结果在左侧 410 示出，而个性化结果在右边 400 示出。滑块 430 用于控制
25 组合标准和个性化结果的功能，其范围从无个性化到全个性化。

图 5 示出界面 500，其中个人感兴趣的结果通过按其 $psim$ 分数的比例增加其字号大小来进一步高亮；也可使用色彩或其它呈现线索。此外，将基本权重赋予 $psim$ 分数的项可在个人结果摘要中高亮。在左边 510 示出按尺寸增加排序的标准结果。在 500 界面示出个性化组合，再次对个人感兴趣条目用增加字体大小来显示。
30

图 6 示出界面 600 上提供个性化查询的过程。在此情形中，在 610 考虑从查询中返回的前 N 个结果。在 620 根据用户模型和返回结果计算相似性。在 630，组

合个性化和标准结果，并在 640 对这些结果重新排序，其中它们在 600 被示为个性化结果。

图 7-9 示出如上所述的个性化控制的效果。参照图 7，界面 700 通过个性化控件 710 来调整，其中采用搜索词“Eton”。在 720，Eton 学院的第一个结果被排序为 1/100。个性化控件 710 往右移，则部分个性化结果在列表中出现。出现在标准结果列表中位置 32 的结果在位置 4 显示。在图 8，个性化控件 810 略往右移，表示搜索的更为个性化。在此情形中，产生与 Eton 学校相关的第一排序，其中 Eton 学校与用户相关的人相关联。在此情形中，图 7 中的先前排序为 100 中的第 32。在图 9，个性化滑块被移到最右边 910，提供了更为个性化的结果排序，它与当天贴出的 Eaton 学校校服相关。

图 10-13 示出根据本发明一方面可用来个性化查询和/或结果的示例过程。图 10 在标号 1000-1020 示出各轴，表示包括查询、产生该查询的用户、以及从该查询接收的文档的标准信息检索维数。根据本发明，第四维或个性化维 1030 被视为基于用户模型根据用户的个人特征或兴趣来另外提炼、聚焦或更改查询和/或结果。

这种个性化信息可从与多种个人信息相关的元数据采样，这些个人信息可用于用户，诸如文档何时创建、查看或更改、时间标记信息、已存储或查看的信息、使用的应用程序、网站动作的日志（例如感兴趣的网站或主题）、诸如位置信息或最近活动、电子邮件活动、日程安排活动、诸如通过电子通信的个人交互的环境信息、统计信息、概况文件信息、情况相似的用户信息等。这些特征可进行采样并从先前所述的用户模型中导出。

参看图 11，维恩图 1100 示出了可从标准相关性反馈模型中导出的搜索条目的交集。外部圆 1110 示出表示可搜索文档总量的 N 。内部圆 n_i 表示与相关性反馈确定相关的文档，其中 n_i 和 R 之间的分段或交迭部分表示具有所需搜索特征的文档 r_i ，并通过该算法被视为相关。一般而言， R 由提供可变程度相关性的判定的用户来确定（例如用户分配分数）。根据本发明， R 通过分析前述用户模型来自动确定，以确定用户感兴趣的相关区域。 N 和 R 也可表示文档空间的一部分（例如与查询相关的文档子集，由查询项的显现所示）而不表示整个文档空间。此外，全集统计量 N 和 n_i 可使用结果集来近似， N 是结果集中的文档数量而 n_i 是具有给定搜索项的文档数量，其中文档由文档的全文本或描述该文档的结果集片断表示。

以下等式示出基于文档项的某个子集的和将分数分配给给定文档的记分函数，其中文档中项 i 的频率(tf_i)倍乘表示该项稀有性的已定权重(w_i)。然后可采用该

记分函数来个性化结果。在此情形中采用 BM25 相关性反馈模型，但可以理解，根据本发明基本上任何信息检索算法可适用于个性化查询和/或结果更改。

$$\text{Score} = \sum \text{tf}_i * w_i$$

$$w_i = \log \frac{(r_i+0.5)(N-n_i-R+r_i+0.5)}{(n_i-r_i+0.5)(R-r_i+0.5)}$$

继续到图 12，个性化相关文档信息(R)在维恩图 1200 中被示为与集合信息(N)分开。在此情形中，引入项 N' 和 n_i' 以便于该分离，其中 $N' = N + R$ ， $n_i' = n_i + r_i'$ 且 w_i 被计算为：

$$w_i = \log \frac{(r_i+0.5)(N'-n_i'-R+r_i+0.5)}{(n_i'-r_i+0.5)(R-r_i+0.5)}$$

图 13 示出在 1300 分开的数据的个性化聚集，其中个性化条目和匹配搜索主题的条目都在 1310 示出。例如，圆 1320 可包括 web 上存在的所有文档，在 1320 表示的文档可包括与个人数据相关的文档(例如与从用户模型导出的对汽车的兴趣相关的文档)，而在 1310 的条目则是那些与搜索词相关的个人文档。可以理解，查询和结果可依赖于模型和感兴趣的查询根据多个词或条件进行更改。

参照图 14，实现本发明各方面的示例性环境 1410 具有计算机 1412。计算机 1412 具有处理单元 1414、系统存储器 1416、及系统总线 1418。系统总线 1418 耦合系统组件包括，但不限于将系统存储器 1416 耦合到处理单元 1414。处理单元 1414 可以是各种可用处理器的任一种。双微处理器和其它多处理器架构也可被用作处理单元 1414。

系统总线 1418 可以是若干类总线结构的任一种，包括存储器总线或存储器控制器、外围总线或外部总线、和/或使用各种可用总线结构任一种的本地总线，这些总线结构包括，但不限于，11 比特总线、工业标准结构(ISA)、微信道结构(MCA)、扩展 ISA (EISA)、智能驱动器设备 (IDE)、VESA 局部总线 (VLB)、外围部件互连 (PCI)、通用串行总线 (USB)、高级图形端口 (AGP)、个人计算机存储卡国际协会总线 (PCMCIA)、和小型计算机系统接口 (SCSI)。

系统存储器 1416 具有易失性存储器 1420 和非易失性存储器 1422。包含在计算机 1412 硬件间传送如起动机信息的基本例程的基本输入/输出系统 (BIOS)，存储在非易失性存储器 1422 上。作为说明且非限制，非易失性存储器 1422 可包括只读存储器 (ROM)、可编程 ROM (PROM)、电可编程 ROM (EPROM)、电可擦除 ROM (EEPROM) 或闪存。易失性存储器 1420 包括作为外部高速缓存的随机

存取存储器 (RAM)。作为说明且非限制, RAM 有多种形式可用, 如同步 RAM (SRAM)、动态 RAM(DRAM)、同步 DRAM(SRAM)、双数据数率 SDRAM(DDR SDRAM)、增强 DSRAM(ESDRAM)、同步链接 DRAM(SLDRAM)、和直接 Rambus RAM (DRRAM)。

5 计算机 1412 还具有可移动/不可移动、易失性/非易失性计算机存储介质。图 14 示出了例如磁盘存储器 1424。磁盘存储器 1424 包括, 但不限于, 象磁盘驱动器、软盘驱动器、磁带驱动器、Jaz 驱动器、Zip 驱动器、LS-100 驱动器、闪存卡、或存储棒。另外, 磁盘存储器 1424 可具有单独存储介质或结合其它存储介质包括, 但不限于, 诸如光盘 ROM 设备 (CD-ROM)、可记录光盘驱动器 (CD-R 驱动器)、
10 可重写光盘驱动器 (CD-RW 驱动器)、或数字化通用盘 ROM 驱动器 (DVD-ROM) 的光盘驱动器。为便于磁盘存储器 1424 与系统总线 1418 的连接, 通常可使用诸如接口 1426 的可移动或不可移动接口。

应当理解, 图 14 描述了作为用户和在适当操作环境 1410 中描述的基本计算机资源间中间体的软件。这种软件包括操作系统 1428。可存储在磁盘存储器 1424
15 中的操作系统 1428, 其作用是控制和分配计算机系统 1412 的资源。系统应用 1430 得益于操作系统 1428 通过存储在或系统存储器 1416 或磁盘存储器 1424 中的程序模块 1432 和程序数据 1434 对资源的管理。可以理解本发明可用各种操作系统或操作系统组合来实现。

用户通过输入设备 1436 将命令或信息输入到计算机 1412 中。输入设备 1436
20 包括, 但不限于, 诸如鼠标、跟踪球、铁笔、触板、键盘、麦克风、操纵杆、游戏垫、卫星接收器、扫描仪、电视调谐器卡、数字相机、数字录像照相机、网络相机等定位设备。这些和其它输入设备经系统总线 1418 通过接口端口 1438 与处理单元 1414 相连。接口端口 1438 包括例如串行端口、并行端口、游戏端口、和通用串行总线 (USB)。输出设备 1440 使用象输入设备 1436 的某些相同类型端口。因此,
25 例如 USB 端口可用来为计算机 1412 提供输入, 并将信息从计算机 1412 输出到输出设备 1440。输出适配器 1442 被提供用来说明象监视器、扬声器、和打印机的某些输出设备与其它输出设备 1440 之间需要特殊适配器。输出适配器 1442 包括, 作为说明且非限制, 在输出设备 1440 和系统总线 1418 间提供了连接方法的视频卡和声卡。应该注意的是诸如远程计算机 1444 的其它设备和/或设备系统提供输入和输出功能。
30

计算机 1412 使用与一台或多台远程计算机如远程计算机 1444 的逻辑连接在

网络化环境中操作。远程计算机 1444 可以是个人计算机、服务器、路由器、网络 PC、工作站、基于微处理器的电气设备、对等设备或其它公共网络节点等等，而且通常包括所述与计算机 1412 相关的许多或全部元件。为简化起见，仅存储器存储装置 1446 与远程计算机 1444 一起图示。远程计算机 1444 经网络接口 1448 与计算机 1412 逻辑连接，然后通过网络连接 1450 进行物理连接。网络接口 1448 包括诸如局域网（LAN）和广域网（WAN）的有线和/或无线通信网络。LAN 技术包括光纤分布式数据接口（FDDI）、铜线分布式数据接口（CDDI）、以太网/IEEE1402.3、令牌环/IEEE1402.5 等等。WAN 技术包括，但不限于，点对点链接、象综合服务数字网（ISDN）及其后变种的电路交换网络、分组交换网络、和数字用户线路（DSL）。

通信连接 1450 指向被用来将网络接口 1448 连接到总线 1418 的硬件/软件。尽管为了清晰地说明，通信连接 1450 被示为位于计算机 1412 内，但它也可在计算机 1412 之外。与网络接口 1448 连接所必需的硬件/软件具有，仅作为示例，诸如包括常用电话级调制解调器、电缆调制解调器和 DSL 调制解调器的调制解调器、ISDN 适配器、和以太网卡等内部和外部技术。

图 15 是本发明可与之交互的采样计算环境 1500 的示意方框图。系统 1500 包括一台或多台客户机 1510。客户机 1510 可以是硬件和/或软件（例如线程、过程、计算装置）。系统 1500 还可包括一台或多台服务器 1530。服务器 1530 也可以是硬件和/或软件（例如线程、过程、计算装置）。服务器 1530 可包含采用本发明来执行变换的线程。在客户机 1510 和服务器 1530 间的一可能通信可能是以适于在两个或多个计算机过程间传送的数据包形式进行。系统 1500 包括可用来便于客户机 1510 和服务器 1530 间通信的通信框架 1550。客户机 1510 可与用来存储客户机 1510 本地信息的一个或多个客户数据存储 1560 作可操作连接。类似地，服务器 1530 可与用来存储服务器 1530 本地信息的一个或多个客户数据存储 1540 作可操作连接。

以上所述包括本发明的各个示例。当然，为描述本发明而对每一能想到的组件或方法组合进行描述是不可能的，但本领域普通技术人员明白本发明的更多排列和组合是可能的。因此，本发明旨在包含所有这样的在所附权利要求书精神和范围内的变更、修改、和变化。此外，就用于具体实施方式或权利要求书的术语“包含”而言，这种术语意在以类似于术语“包括”在权利要求书中作及物词的方式作包括意义解。

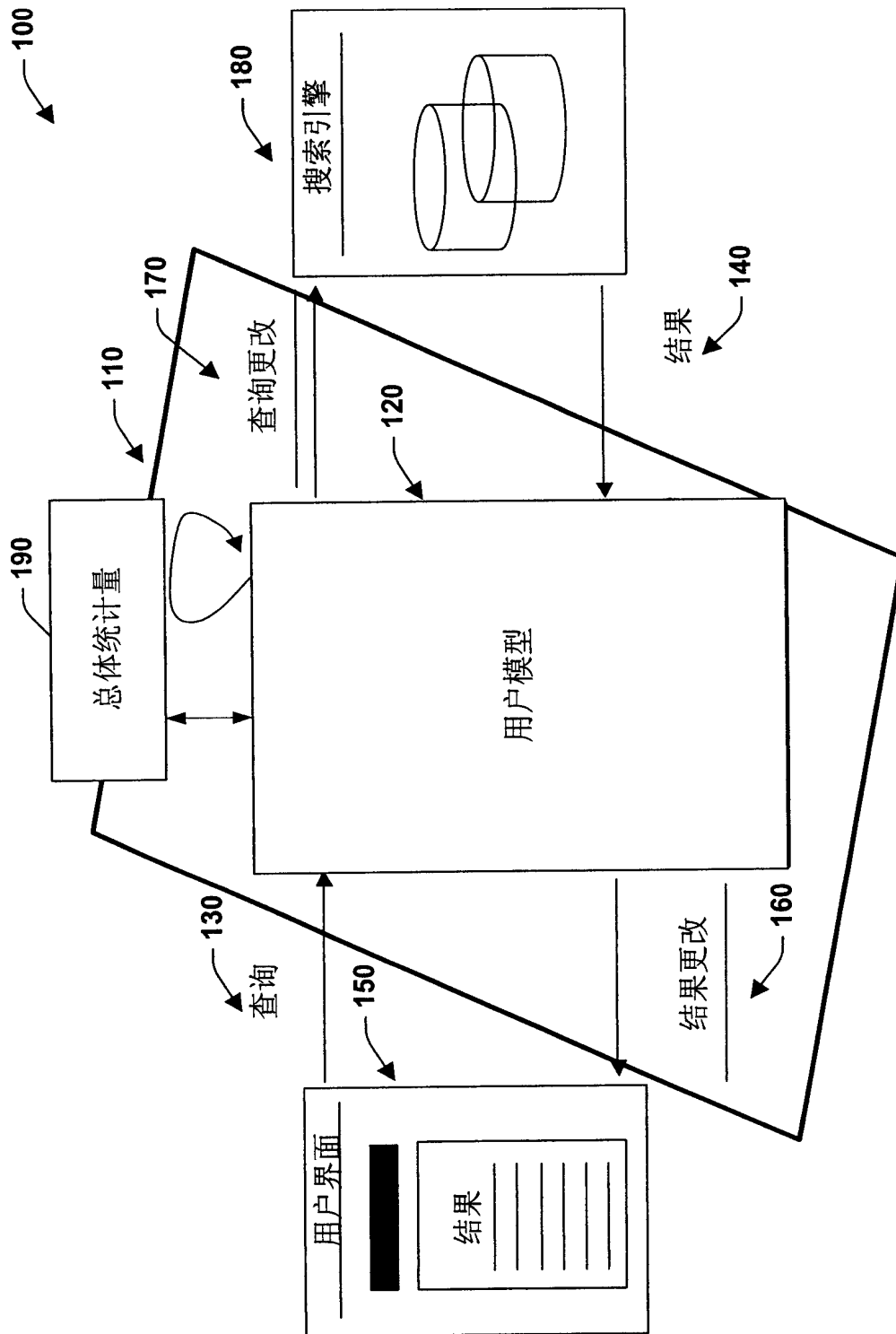


图 1

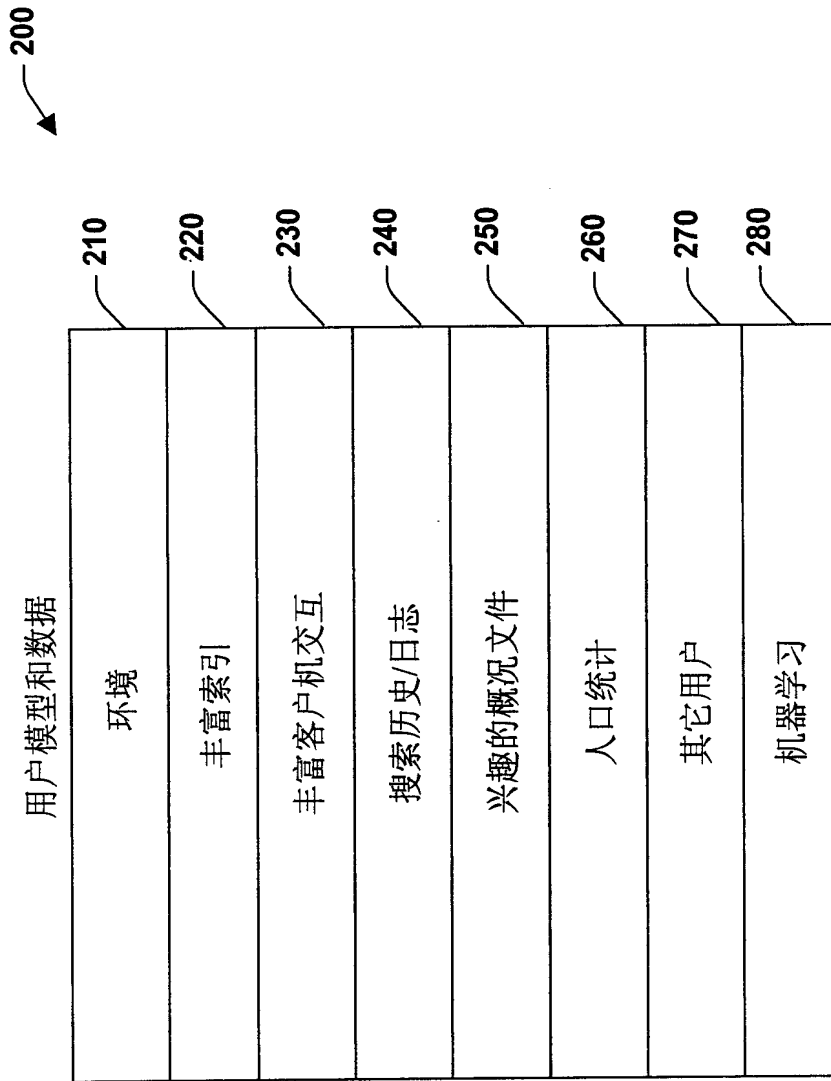


图 2

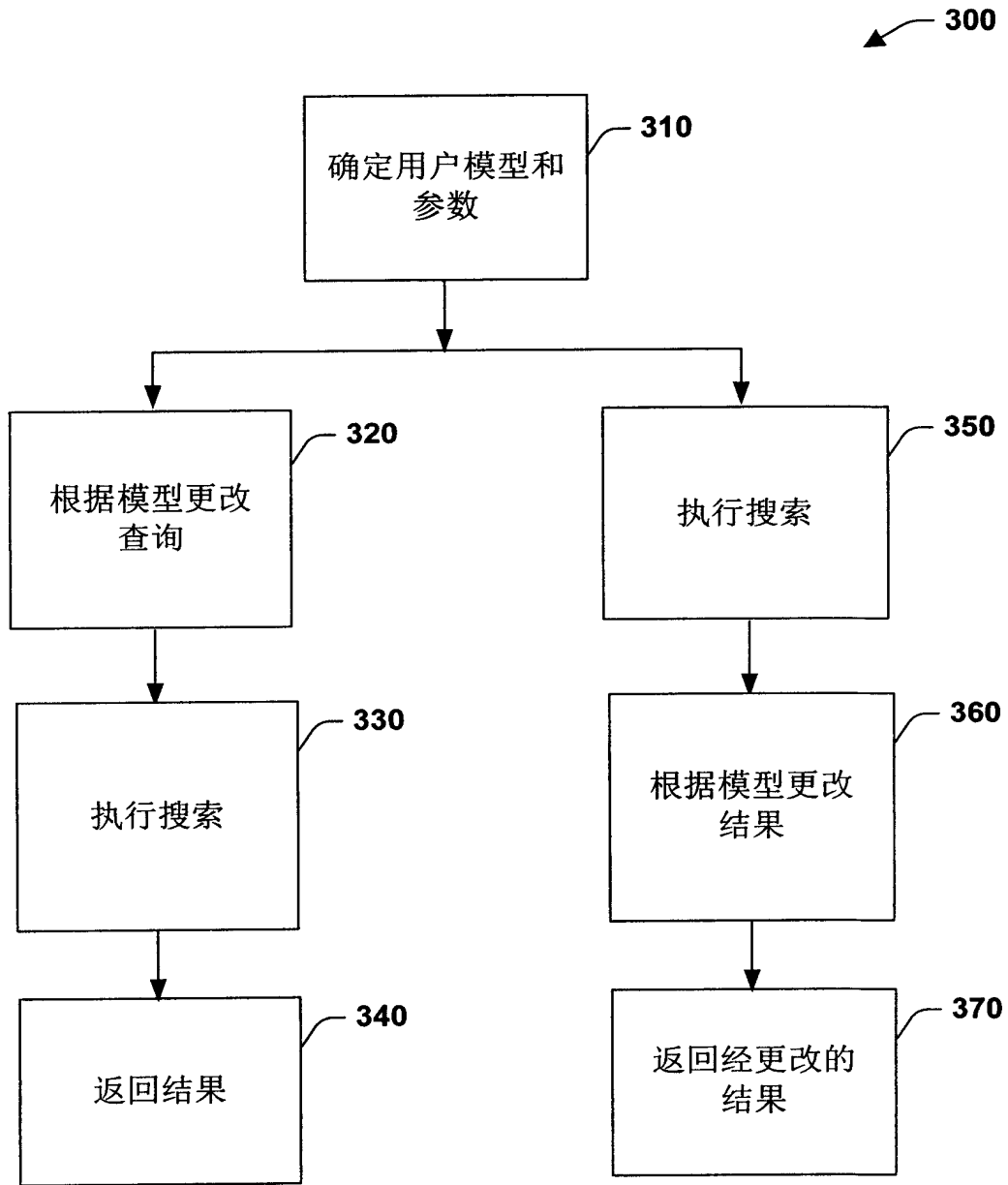
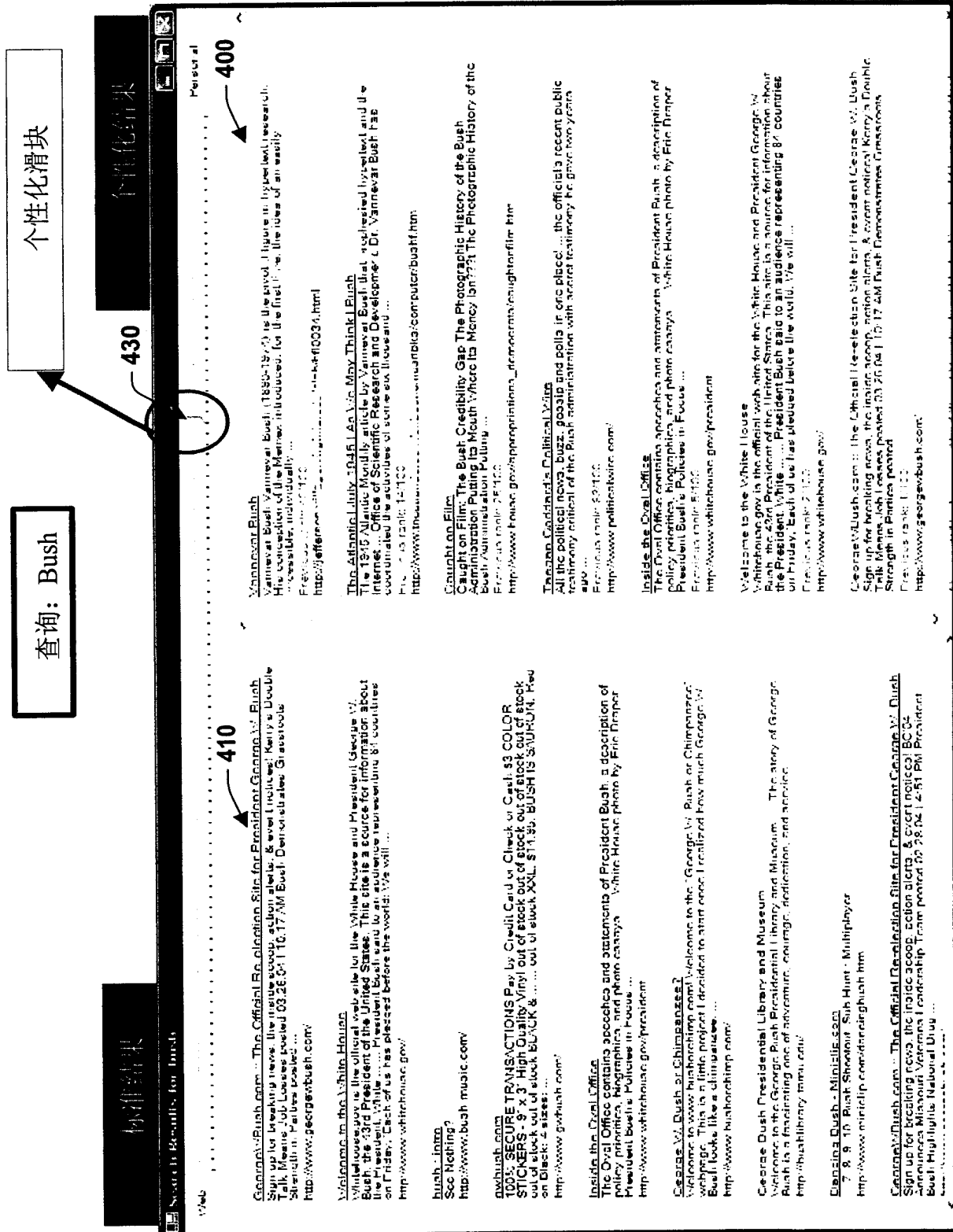
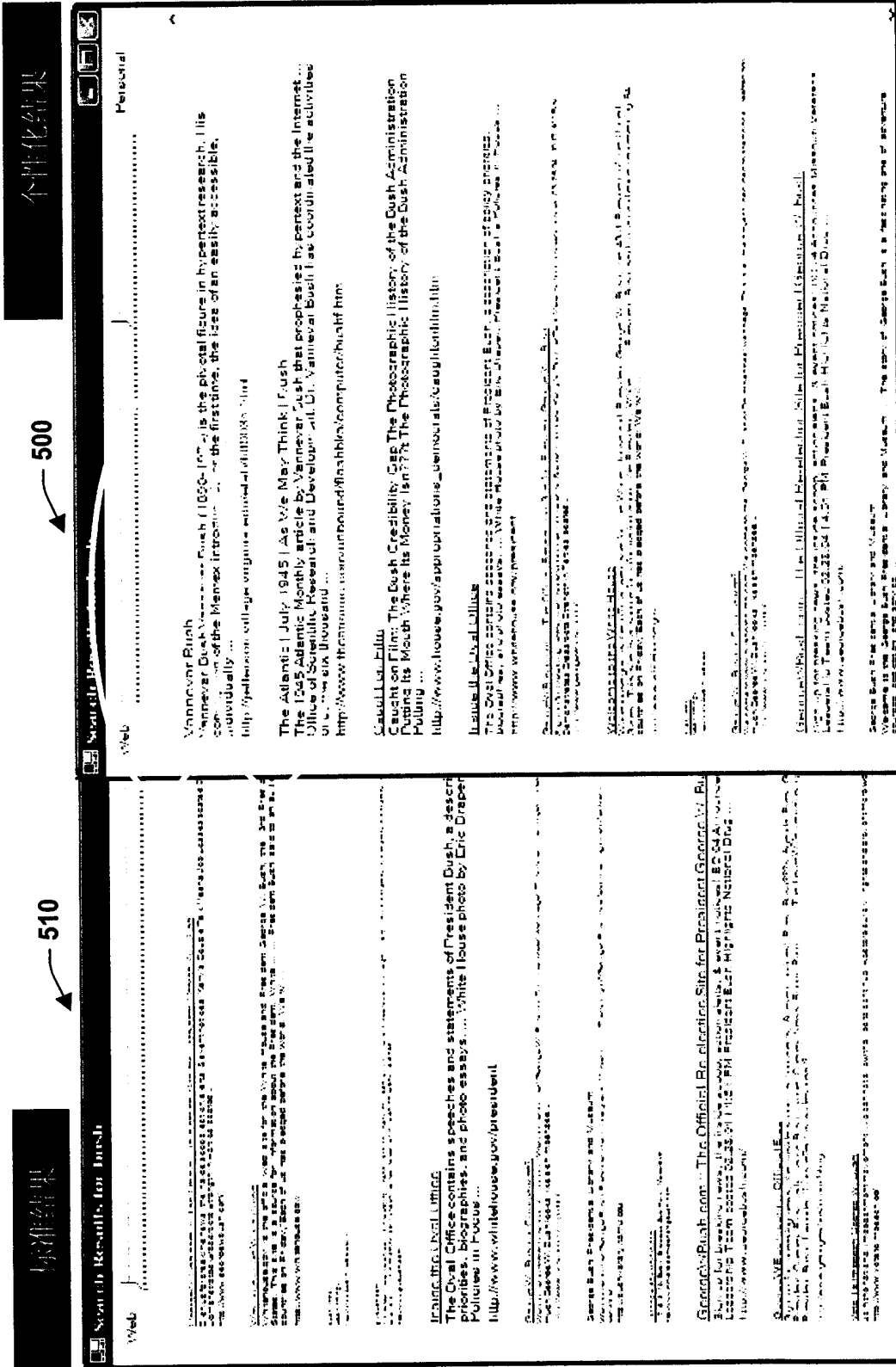


图 3





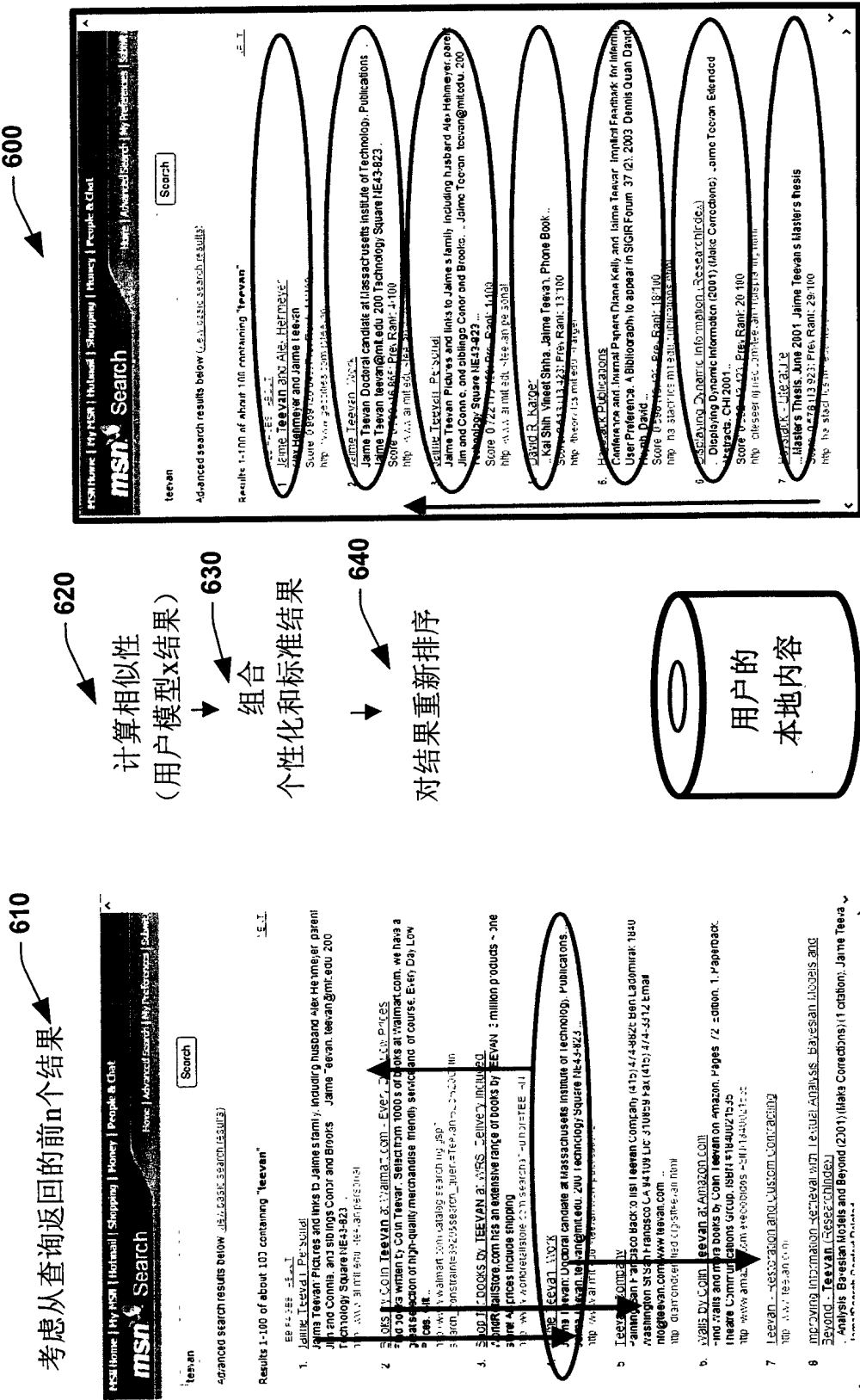


图 6

The screenshot shows a search engine interface with a search bar at the top containing the text '700' and a search button. Below the search bar, there are two columns of search results. The left column shows results for 'Eton College' with a rank of 720 and a URL 'http://www.etoncollege.com/'. The right column shows results for 'Eton' with a rank of 710 and a URL 'http://www.eton.co.uk/'. Both results include a brief description of Eton College. Below these, there are several other search results, including 'Eton Import Co. Ltd.' and 'E-Ton', each with a rank and a URL. The search results are displayed in a list format with various icons and symbols.

700

Personal

Web

710

Eton College
<http://www.etoncollege.com/>
 Eton: arts, antiques, jewellers, bookshops, tailors and gift shops
 Eton's historic half mile High Street, leading from the world famous college to Windsor Castle, enjoys an international reputation for the arts, antiques, jewellers, bookshops, tailors and gift shops. ... The Eton Traders Association welcomes you to the ancient and modern town of Eton. ...
<http://www.eton.co.uk/>

720

Eton College
<http://www.etoncollege.com/>
 Eton: arts, antiques, jewellers, bookshops, tailors and gift shops
 Eton's historic half mile High Street, leading from the world famous college to Windsor Castle, enjoys an international reputation for the arts, antiques, jewellers, bookshops, tailors and gift shops. ... The Eton Traders Association welcomes you to the ancient and modern town of Eton. ...
<http://www.eton.co.uk/>

ETON
<http://www.etonshirts.com/>
 Previous rank: 37100
<http://www.etonshirts.com/>

Eton Import Co. Ltd.
 (ETON EXCLUSIVE CARE) ??????????????????????????????????
 2 ??
 (???
 ?????????? ...
<http://www.eton-import.com/>

E-Ton
 120702 Welcome ... Redirecting to new Home
<http://www.etonamerica.com/>
 Welcome to Eton Systems Inventor of the Unit Production ...

3 ??
 (ETON EXCLUSIVE CARE) ??????????????????????????????????????

800



Personal

Search Results for eton

Web

Eton College

<http://www.etoncollege.com/>

Eton: arts, antiques, jewellers, bookshops, tailors and gift shops
Eton's historic half mile High Street, leading from the world famous college to Windsor Castle, enjoys an international reputation for the arts, antiques, jewellers, bookshops, tailors and gift shops. ... The Eton Traders Association welcomes you to the ancient and modern town of Eton. ...
<http://www.eton.co.uk/>

E10N

<http://www.etonshirts.com/>

Eton Import Co.,Ltd.

(ETON EXCLUSIVE CARE) ??????????????????
3 ??????????????????????????????????????
2 ??????????????????????????????????????
(?????????????????????????????????????
????????? ...
<http://www.eton-import.com/>

E-Ton

120702>Welcome ... Redirecting to new Home
<http://www.etonamerica.com/>

Welcome to Eton Customs Invention of the Unit Production

Welcome to Eton School - Today's date is 4/17/04

Eton School was founded in 1978 as an independent school with a Montessori philosophy located on the Eastside of Lake Washington. The goal of Eton School is to develop the child's intellectual ...
Previous rank: 32/100
<http://www.etonschool.com/>

Eton School Uniform

Eton School Uniform
Previous rank: 44/100
<http://www.histclo.hispeed.com/style/suit/eton/etonsu.html>

Eton College

Previous rank: 1/100
<http://www.etoncollege.com/>

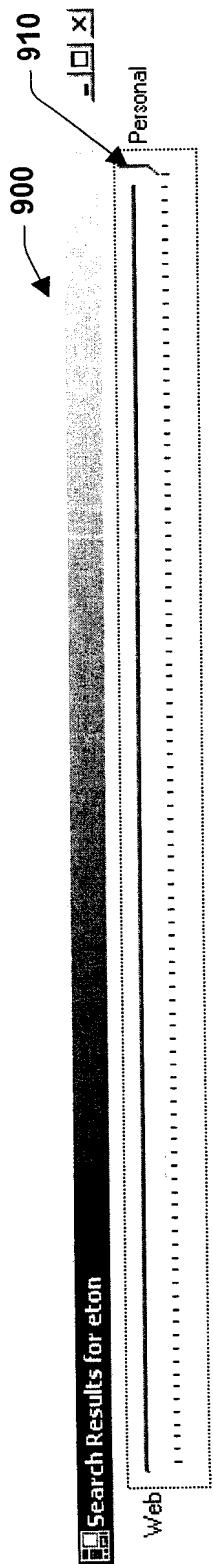
Eton: arts, antiques, jewellers, bookshops, tailors and gift shops

Eton's historic half mile High Street, leading from the world famous college to Windsor Castle, enjoys an international reputation for the arts, antiques, jewellers, bookshops, tailors and gift shops. ... The Eton Traders Association welcomes you to the ancient and modern town of Eton. ...
Previous rank: 2/100
<http://www.eton.co.uk/>

E10N

Previous rank: 3/100





Eton College
<http://www.etoncollege.com/>
Eton: arts, antiques, jewellers, bookshops, tailors and gift shops
 Eton's historic half mile High Street, leading from the world famous college to Windsor Castle, enjoys an international reputation for the arts, antiques, jewellers, bookshops, tailors and gift shops. ... The Eton Traders Association welcomes you to the ancient and modern town of Eton. ...
<http://www.eton.co.uk/>

Eton School Uniform
<http://www.histclo.hispeed.com/style/suit/eton/etonsu.html>
Welcome to Eton School - Today's date is 4/17/04
 Eton School was founded in 1978 as an independent school with a Montessori philosophy located on the Eastside of Lake Washington. The goal of Eton School is to develop the child's intellectual ...
 Previous rank: 32/100
<http://www.etonsschool.com/>

Eton
 ... The new line of ETON Youth ATV's has arrived. ...
 Previous rank: 65/100
<http://www.cyclemarine.com/eton1>

Eton Import Co., Ltd.
ETON EXCLUSIVE CARE ??????????????????????
 3 ???
 2 ???
 (???
 ?????????? ...
<http://www.eton-import.com/>

E-Ton
 120702 Welcome ... Redirecting to new Home
<http://www.etonamerica.com/>
 Welcome to Eton Systems Inventor of the Unit Production

Adoption Agencies & Services, Eton, Georgia, GA Superpages Yellow Pages
 SuperPages.com can help you find Adoption Agencies & Services business listings in our online Yellow Pages directory service. Here you can find yellow pages phone listings, directions, online ...
 Previous rank: 33/100
<http://phonebook.superpages.com/yellowpages/C-Adoption+Agencies>

WEFC On-Line - The Official Site Of Windsor & Eton FC Windsor & Eton Football Club The Official Web Site Home Latest News First Team 03/04 Fixtures 03/04 Tables Squad Info Squad Stats Sponsorship Reserve Team Reserve Home 03/04 Fixtures 03/04

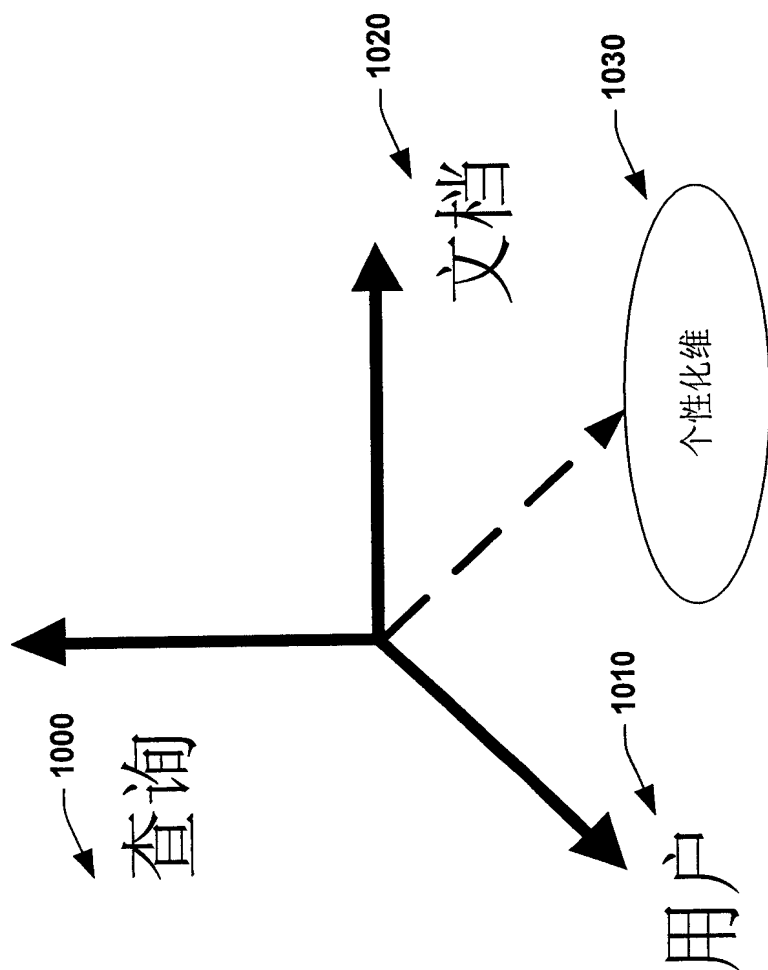


图 10

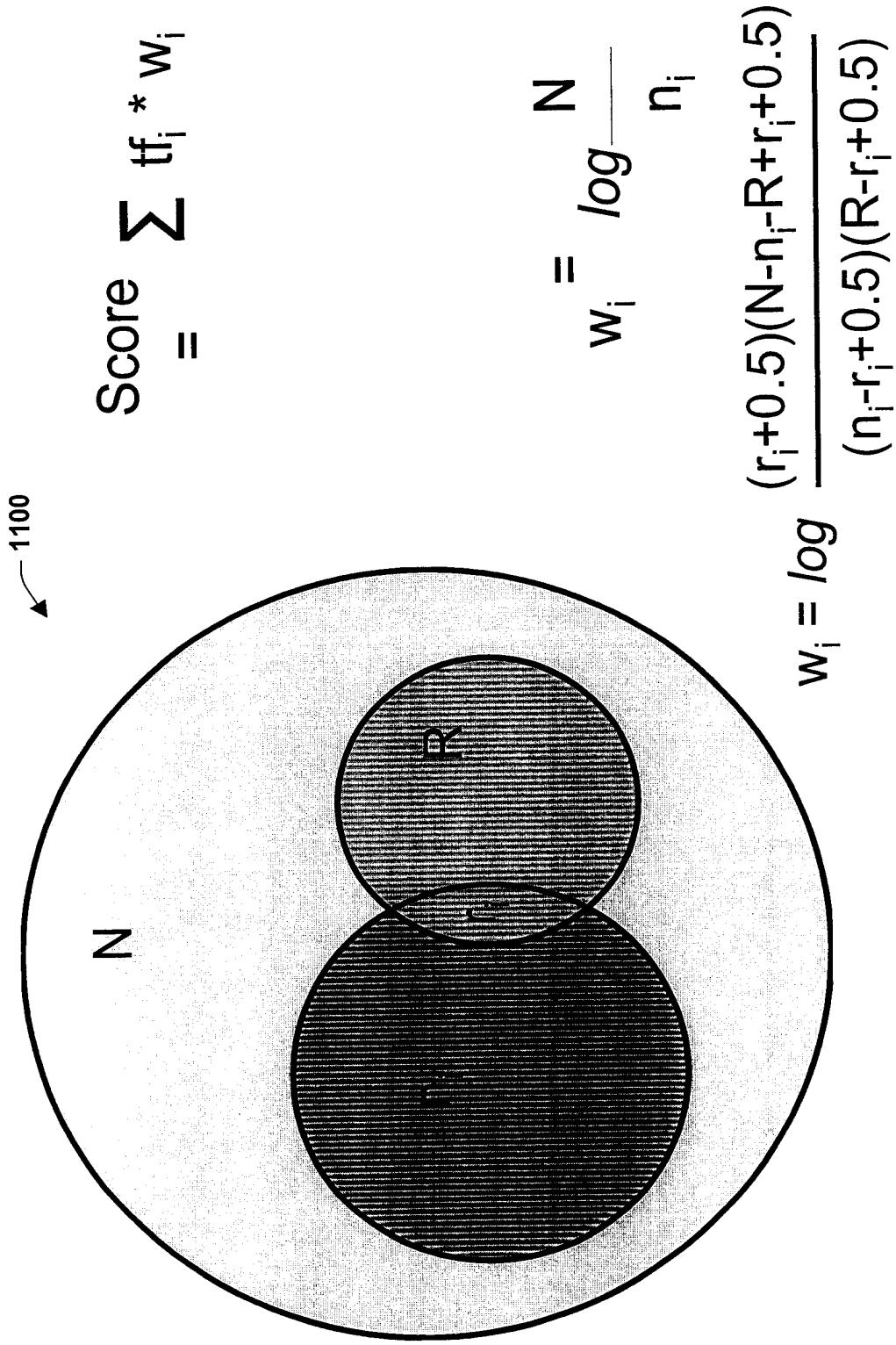


图 11

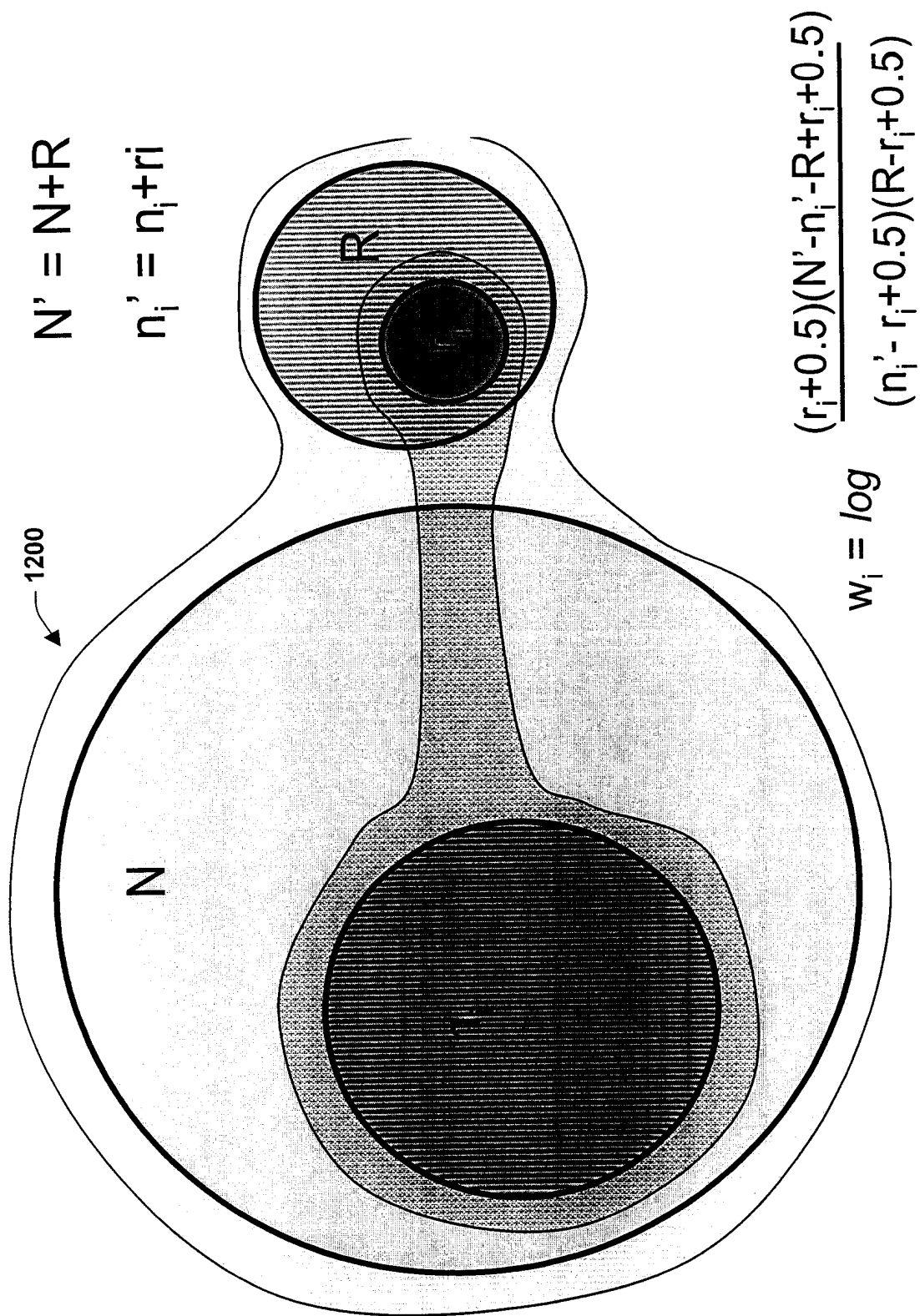


图 12

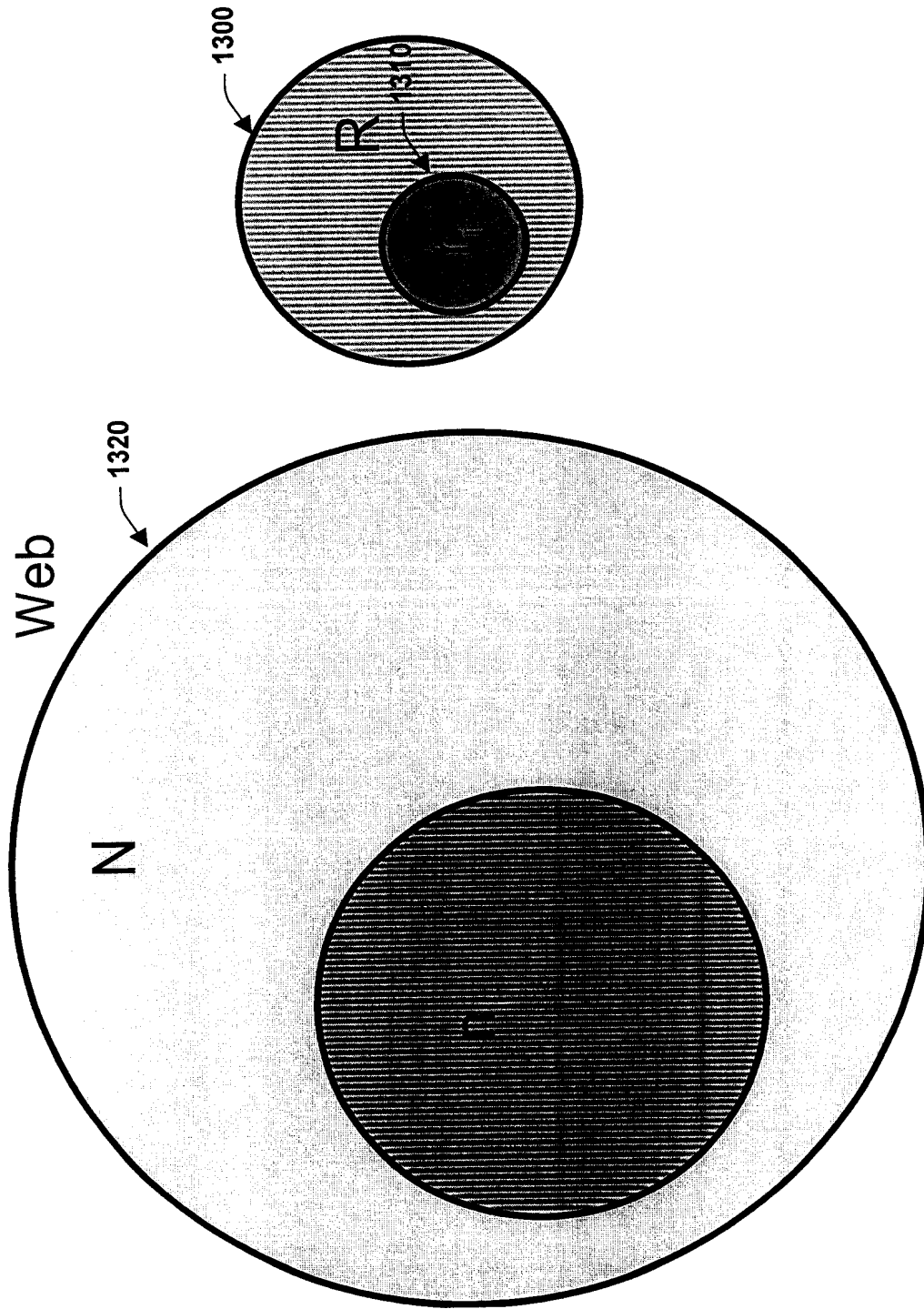


图 13

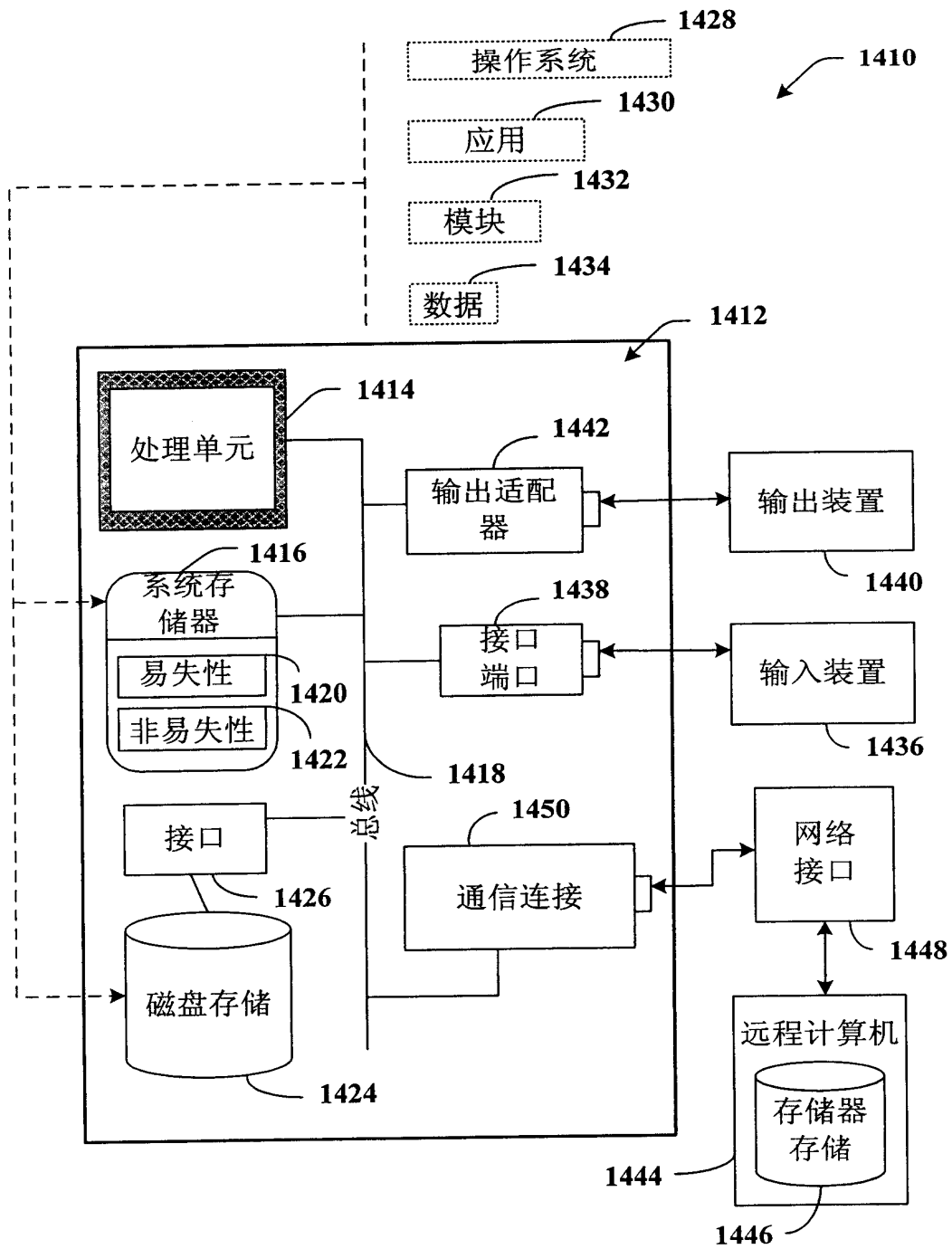


图 14

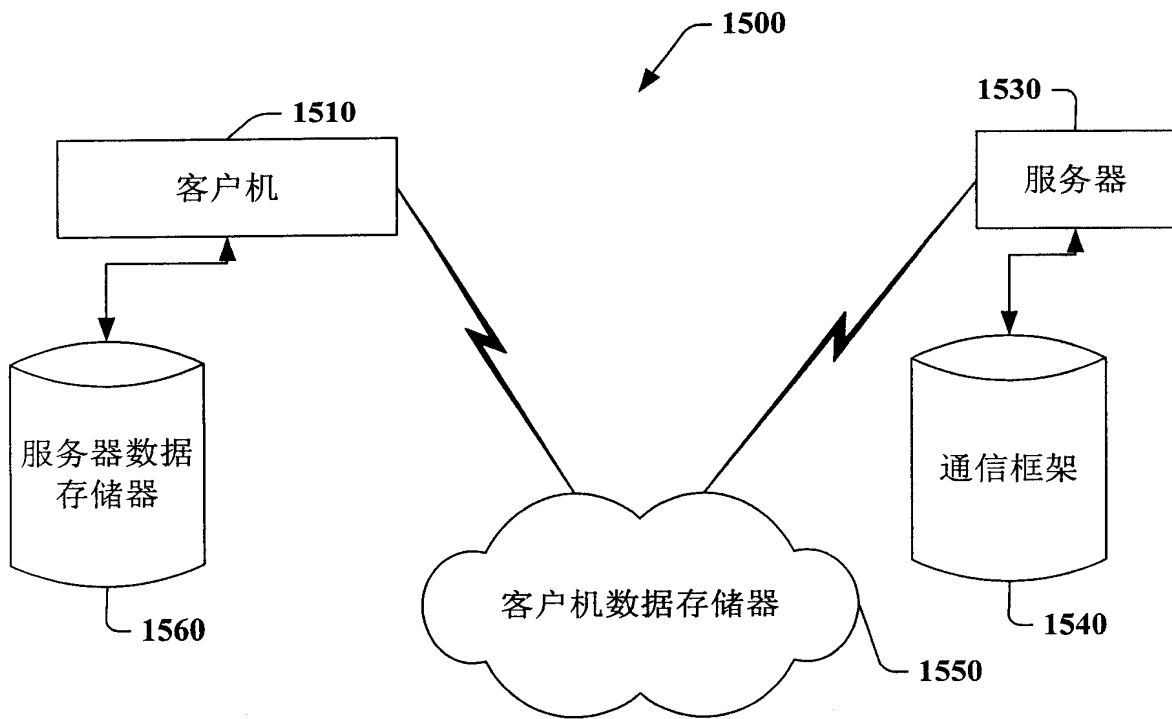


图 15