

(12) 发明专利

(10) 授权公告号 CN 101263493 B

(45) 授权公告日 2011.08.03

(21) 申请号 200680030192.4

US 6285999 B1, 2001.09.04, 全文.

(22) 申请日 2006.06.19

CN 1438595 A, 2003.08.27, 全文.

(30) 优先权数据

11/155,882 2005.06.20 US

US 6820075 B2, 2004.11.16, 全文.

(85) PCT申请进入国家阶段日

2008.02.19

WO 01/46870 A1, 2001.06.28, 全文.
US 6754873 B1, 2004.06.22, 说明书第6栏
第28行-第7栏第17行, 第8栏第29-40行、附图2, 11, 12A.

(86) PCT申请的申请数据

PCT/US2006/023792 2006.06.19

王胜明, 潘卫华, 孙淑淇. 一种搜索引擎的新体系结构. 计算机工程第29卷 第11期. 2003, 第29卷(第11期), 98-100.

(87) PCT申请的公布数据

W02007/001974 EN 2007.01.04

审查员 唐嫣

(73) 专利权人 谷歌公司

地址 美国加利福尼亚州

(72) 发明人 路易斯·卡斯特罗 瓦特·林

贝内迪克特·戈梅斯

(74) 专利代理机构 中原信达知识产权代理有限公司
责任公司 11219

代理人 周亚荣 安翔

(51) Int. Cl.

G06F 17/30 (2006.01)

(56) 对比文件

CN 1389811 A, 全文.

US 6631372 B1, 2003.10.07, 全文.

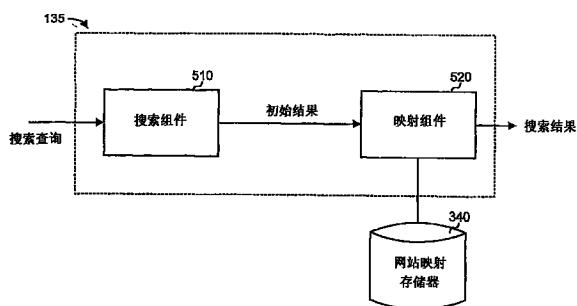
权利要求书 4 页 说明书 9 页 附图 7 页

(54) 发明名称

用于提供搜索结果的系统和方法

(57) 摘要

一种包括响应于用户查询生成搜索结果的方法, 其中至少一个搜索结果包括一组链接。该组链接可表示指向相同网站内的网页的链接, 并且可根据与所述链接相关联的至少一个因素对其进行识别。该方法还可以包括将搜索结果提供给用户。



1. 一种由一个或多个服务器设备执行的方法,包括:

收集与多个用户对与多个不同的网站相对应的多个网页的访问有关的信息;

在存储器中存储用于所述多个不同的网站中的每个特定网站的网页信息,对于所述多个网页中的每一个,所述网页信息包括质量因素,所述质量因素反映对应于所述特定网站的所述网页已被所述多个用户访问的次数;

经由客户端设备从用户处接收搜索查询;

根据所述搜索查询执行搜索,以生成搜索结果,其中所述搜索结果中的多个对应于与多个不同的域相关联的多个不同的网站;

识别与所述搜索结果中的所述多个中的一个相对应的所述多个不同的网站中的一个;

确定所述存储器是否存储有所识别的网站的网页信息;

在所述存储器存储有所识别的网站的网页信息的情况下,根据与一组网页相关联的质量因素从所述存储器识别与所识别的网站相关联的所述一组网页;以及

提供所述搜索结果用于在所述客户端设备的显示器上展示,其中提供所述搜索结果包括:

在所述存储器存储有所识别的网站的网页信息的情况下,提供所述搜索结果中的所述多个中的所述一个以及指向所述一组网页中的网页的多个链接,或

在所述存储器未存储有所识别的网站的网页信息的情况下,提供所述搜索结果中的所述多个中的所述一个,而不提供指向所述一组网页中的网页的链接。

2. 如权利要求 1 所述的方法,还包括:

对于所述多个网页中的每一个,将与所述网页相关联的链接信息进行存储。

3. 如权利要求 2 所述的方法,还包括:

根据所述多个网页已被访问的次数将用于所述多个网页的链接信息进行排序。

4. 如权利要求 1 所述的方法,其中所述质量因素进一步基于所述多个用户在访问所述多个网页的每一个时所花费的时间量。

5. 如权利要求 1 所述的方法,其中所述质量因素进一步基于带有指向所述多个网页中的一个的链接的多个网页。

6. 如权利要求 1 所述的方法,其中所述质量因素进一步基于通过所述多个网页的每一个购买的商品。

7. 如权利要求 1 所述的方法,其中所述质量因素进一步基于所述多个用户的关于所述多个网页的先前行为。

8. 如权利要求 1 所述的方法,其中提供所述搜索结果中的所述多个中的所述一个以及指向所述一组网页中的网页的多个链接包括:

将文本包括在各个链接的每一个中,所述文本包括位于对应于所述各个链接的网页上的文本。

9. 如权利要求 1 所述的方法,其中提供所述搜索结果中的所述多个中的所述一个以及指向所述一组网页中的网页的多个链接包括:

展示带有各个链接的每一个的文本。

10. 如权利要求 1 所述的方法,其中提供所述搜索结果中的所述多个中的所述一个以

及指向所述一组网页中的网页的多个链接包括：

提供所述多个链接用于在所述客户端设备的所述显示器上展示在所述客户端设备的所述显示器上的所述多个搜索结果中的所述一个的展示之下。

11. 如权利要求 1 所述的方法,还包括：

从所述多个不同的网站中的一个的提供者接收信息,其中所接收的信息识别与所述多个不同的网站中的所述一个相关联的所述多个网页。

12. 一种用于提供搜索结果的系统,包括：

用于存储多个不同网站中的每一个特定网站的网页信息的装置,对于与所述特定网站相关联的多个网页,所述网页信息包括质量因素,所述质量因素反映对应于所述特定网站的所述网页已被多个用户访问的次数；

用于经由客户端设备从用户处接收搜索查询的装置；

用于基于所述查询执行搜索以生成搜索结果的装置,其中所述搜索结果中的多个对应于与多个不同的域相关联的多个网站；

用于识别对应于所述搜索结果中的所述多个中的一个的网站的装置；

用于确定所识别的网站的网页信息已被存储的装置；

用于在所识别的网站的网页信息已被存储的情况下根据与一组网页相关联的质量因素从所述网页信息识别与所识别的网站相关联的所述一组网页的装置；以及

用于提供所述搜索结果用于在所述客户端设备的显示器上展示的装置,其中提供所述搜索结果的所述装置包括:在所识别的网站的网页信息已被存储的情况下,提供所述搜索结果中的所述多个中的所述一个以及指向所述一组网页中的网页的多个链接。

13. 如权利要求 12 所述的系统,还包括：

用于当所述多个网页中的一个被访问时接收信息的装置；以及

用于递增与所述多个网页中的所述一个相关联的计数值的装置。

14. 如权利要求 13 所述的系统,还包括：

用于根据所述计数值将顺序赋予所述多个网页的装置；以及

用于根据所赋予的顺序展示指向所述一组网页中的所述网页的所述多个链接的装置。

15. 如权利要求 12 所述的系统,还包括：

用于根据以下中的至少一个:与所述多个网页相关联的多个先前访问、带有指向所述多个网页中的一个的链接的多个网页、所述搜索查询与所述多个网页中的每一个的信息的匹配的接近程度,或者与访问所述多个网页中的每一个相关联的用户历史,来将所述多个网页进行排名的装置。

16. 一种用于提供搜索结果的系统,包括：

用于从客户端设备接收搜索查询的装置；

用于根据所述搜索查询执行搜索以识别搜索结果的装置,其中所述搜索结果中的多个对应于与多个不同的网站相关联的网页,其中所述多个不同的网站与不同的域相关联；

用于识别与所述搜索结果中的所述多个中的一个相对应的所述网页中的一个的装置；

用于识别所述多个不同的网站中的特定网站的装置,其中所识别的网页与所述特定网站相关联；

用于从所述特定网站识别多个网页的装置,其中识别所述多个网页基于至少一个因素,所述至少一个因素包括以下中的一个或多个:所述多个网页中的每一个已被用户访问的次数,或所述用户访问所述多个网页中的每一个所花费的时间量;以及

用于提供所述搜索结果以及与所述多个网页有关信息以在所述客户端设备的显示器上展示的装置。

17. 如权利要求 16 所述的系统,其中用于识别所述多个网页的所述装置包括:

用于根据所述至少一个因素确定顺序的装置,其中与所述多个网页有关的所述信息以所述顺序被展示在所述客户端设备的所述显示器上。

18. 如权利要求 16 所述的系统,其中与所述多个网页有关的所述信息包括指向所述多个网页的链接。

19. 如权利要求 16 所述的系统,其中用于识别所述多个网页的所述装置包括:

用于从与所识别的网页相关联的信息确定网站标识符的装置,所述网站标识符对应于所述特定网站的主页,以及

用于将所述网站标识符存储为指向与所述特定网站相关联的其它网页的索引的装置。

20. 一种用于提供搜索结果的系统,包括:

用于将多个链接存储为已存储的多个链接的装置;从用户处接收搜索查询的装置;

根据所述搜索查询执行搜索以生成搜索结果的装置,其中所述搜索结果中的多个搜索结果对应于与多个不同的网站相关联的网页,其中所述多个不同的网站与不同的域相关联;

识别与所述多个搜索结果中的一个相对应的所述网页中的一个的装置;

识别所述多个不同的网站中的特定网站的装置,其中所识别的网页与所述特定网站相关联;

访问所述已存储的多个链接以识别与所述多个不同的网站中的所述特定网站相关联的所述已存储的多个链接中的一组链接的装置;

根据各个网页中的每一个已被访问的次数确定所述一组链接的顺序的装置;以及将所述搜索结果和所述一组链接按所确定的顺序提供给所述用户的装置。

21. 如权利要求 20 所述的系统,其中确定所述一组链接的顺序的所述装置包括:根据关于所述用户是否将访问对应于所述一组链接中的每一个的相应网页的估计将所述一组链接进行排序的装置。

22. 如权利要求 20 所述的系统,其中所述搜索结果和所述一组链接通过用户界面被提供给所述用户用于输出,所述一组链接的所述顺序进一步基于所述一组链接中的每一个链接的质量量度。

23. 如权利要求 20 所述的系统,其中识别所述多个不同的网站中的所述特定网站的所述装置包括:

从与所识别的网页相关联的信息确定网站标识符的装置,所述网站标识符对应于所述多个不同的网站中的所述特定网站的主页,以及

将所述网站标识符存储为指向与所述多个不同的网站中的所述特定网站相关联的其它网页的索引的装置。

24. 一种由一个或多个服务器设备执行的方法,包括:

- 在存储器中存储与第一多个不同的网站相对应的第一多个网页的网页信息；
从客户端设备接收搜索查询；
基于所述搜索查询执行搜索，以识别匹配所述搜索查询的搜索结果，其中所述搜索结果中的多个搜索结果对应于与第二多个不同域的不同网站相关联的第二多个网页；
从所述第二多个网页中识别与所述多个搜索结果中的一个相对应的网页；
识别与所识别的网页相关联的所述第二多个不同的网站中的特定网站；
当所述第二多个不同的网站中的所述特定网站与所述第一多个不同的网站中的所述一个网站相匹配时，从与所述第一多个不同的网站中的一个网站相关联的所述第一多个网页中识别第三多个网页；
根据与所述第三多个网页相关联的至少一个因素确定所述第三多个网页的顺序，所述至少一个因素包括所述第三多个网页中的每一个先前已被访问的次数；以及
按所确定的顺序提供所述多个搜索结果以及指向所述第三多个网页的链接用于在所述客户端设备的显示器上展示。
25. 如权利要求 24 所述的方法，其中所述至少一个因素进一步包括：有关用户是否在访问所述第三多个网页时滚读了所述第三多个网页或在没有滚读所述第三多个网页的情况下激励了所述第三多个网页的信息。
26. 如权利要求 24 所述的方法，其中识别所述第二多个不同的网站中的所述特定网站包括：
从与所识别的网页相关联的信息确定网站标识符，所述网站标识符对应于所述第二多个不同的网站中的所述特定网站的主页，以及
将所述网站标识符存储为指向与所述第二多个不同的网站中的所述特定网站相关联的其它网页的索引。

用于提供搜索结果的系统和方法

技术领域

[0001] 此处描述的实施方式一般涉及信息检索，尤其是涉及提供与搜索查询相关联的信息。

背景技术

[0002] 万维网（“web”）包括极大量的信息。然而，找到期望的信息部分可以是富有挑战性的。由于万维网上的信息量以及对万维网搜索没有经验的新用户的数量增长迅速，该问题被复杂化了。

[0003] 搜索引擎系统力图返回用户感兴趣网页的超链接。通常，搜索引擎系统依据用户输入的搜索项（称为搜索查询）来确定用户的兴趣。搜索引擎系统的目地是根据搜索查询，向用户提供高质量的、相关的搜索结果（如网页）的链接。典型地，搜索引擎系统是通过将搜索查询中的项目与一组预存储的网页进行匹配来执行搜索。包含用户搜索项的网页是“命中页”，其以链接的形式被返回给用户。

[0004] 当现有的搜索引擎系统返回搜索结果时，搜索结果通常包括指向不同网站的网页的链接。接着，用户可以选择一个指向特定网站的链接来尝试找到感兴趣的商品。然而，感兴趣的商品可能并不在与所选择的链接相关联的网页上。在这种情况下，用户可以通过从该第一网页上导航以找到具有需要信息的有效网页。这可能包括在一系列的网页点击，直到用户能够发现感兴趣的商品为止。

发明内容

[0005] 根据一个方面，方法可包括从用户处接收搜索查询，然后根据搜索查询生成搜索结果。方法还可包括根据与网页相关联的质量因素来识别与至少一个搜索结果相关联的多个网页，其中所述网页和与至少一个搜索结果相关联的网页都在一个网站内。方法还可以包括向用户提供搜索结果和与多个网页相关联的多个链接。

[0006] 在另一个方面，可以提供存储可由至少一个处理器执行的指令的计算机可读介质。指令可以使处理器执行方法，该方法可包括从用户处接收搜索查询并且根据搜索查询生成第一搜索结果。由处理器执行的方法还可包括根据与链接相关联的至少一个因素，来确定与第一搜索结果中的至少一个相关联的多个链接。处理器执行的方法还可包括将第二搜索结果发送给用户，其中第二搜索结果包括第一搜索结果和与第一搜索结果中的至少一个相关联的链接，其中链接对应于在相同网站上的网页。

[0007] 在另一个方面，可提供包括存储器和搜索引擎的系统。存储器可存储多个链接。搜索引擎可从用户处接收搜索查询并根据搜索查询生成搜索结果。搜索引擎也可访问存储器以确定与第一个搜索结果相关联的第一组链接，其中第一组链接与相同的网站相关联。搜索引擎还可将搜索结果和第一组链接提供给用户。

[0008] 在另一个方面，方法可包括响应于用户查询生成多个搜索结果，其中至少一个搜索结果包括按照根据与该组链接相关联的至少一个因素的顺序排列的一组链接。方法还可

以包括将多个搜索结果提供给用户。

附图说明

- [0009] 附图包含在本说明书中并构成本说明书的一部份，示出了本发明的具体实施例，并且与说明书一起解释本发明。在附图中：
- [0010] 图 1 是网络的示例图，在其中可实施与本发明原理相一致的系统和方法。
- [0011] 图 2 是与本发明原理相一致的实施方式中的在图 1 的网络中的服务器的示例图。
- [0012] 图 3 是与本发明原理相一致的实施方式中的图解了图 1 的日志处理系统的示例功能框图。
- [0013] 图 4 图解了与本发明原理相一致的实施方式中的图 3 的网站映射存储器的示例部分。
- [0014] 图 5 是与本发明原理相一致的实施方式中的图解了图 1 的搜索引擎系统的示例功能框图。
- [0015] 图 6 是图解了与生成与本发明原理相一致的网站映射信息相关联的示例处理的流程图。
- [0016] 图 7 是图解了与本发明原理相一致的实施方式中与提供搜索结果相关联的示例处理的流程图。
- [0017] 图 8 图解了在与本发明原理相一致的实施方式中的提供给用户的示例界面屏幕。

具体实施方式

[0018] 下列详细描述参考附图。不同图中的相同标记可指示相同或近似的要素。下列详细描述也不限于本发明。

[0019] 概述

[0020] 与本发明原理相一致的系统和方法可提供与一个或多个特定搜索结果相关联的多个链接。这些链接可以作为特定搜索结果表示相同网站中的网页，并且可以表示被确定为用户最可能访问网页的网页。这些链接还可包括帮助确定对应网页的文本。用户然后可选择这些链接中的一个而不用首先选择与搜索结果相关联的链接。

[0021] 在以下的描述中，将术语“站”(site) 或“网站”(website) 宽泛地解释为包括任何域或子域（如 :google.com, labs.google.com 等）或是域加统一资源定位符的前缀（如 :geocities.com/Athens/）。另外，下列描述集中于与网站相关联的搜索。然而，应当了解，与本发明相一致的实施方式也可应用于搜索文档或者文档库。术语“文档”也应被宽泛地解释为包括任何机器可读和机器可存储的工作成果。例如，文档可以包括网页、与新闻事件有关的信息、图片文件、与商品有关的信息、与新闻组网络 (USENET) 页面有关的信息、黄页条目、被扫描的书本、文件、文件的组合、带有指向其它文件的嵌入链接的一个或多个文件、网络日志 (BLOG)、网络广告、电子邮件等。文档通常包括文本信息并且可包括嵌入式信息（如元信息、超链接等）和 / 或嵌入式指令（如 Javascript 等）。在此所用的术语“链接 (link)”，应被宽泛地解释为包括指向 / 来自于一个文档而来自于 / 指向另一个文档或同一个文档的其它部分的指涉 (reference)。

[0022] 示例网络

[0023] 图 1 是网络 100 的示例图, 其中可实施与本发明原理相一致的系统和方法。网络 100 可包括通过网络 140 连接的客户端 110 和服务器 120 和 130。为简单起见, 在图 1 中将两个客户端 110、一个服务器 120 和一个服务器 130 图示为连接到网络 140。在实际中, 可有更多或更少的客户端和服务器。并且, 在一些情况下, 客户端可执行服务器的功能并且服务器可执行客户端的功能。另外, 在替代实施方式中, 由一个服务器执行的功能可以由另一个服务器或与由另一个服务器执行的功能相组合来执行。例如, 服务器 120 可以执行服务器 120 和 130 的功能, 或者反之亦然。

[0024] 客户端 110 可包括客户端实体。可将实体定义为装置 (例如个人计算机、无线电话、个人数字助理 (PDA)、笔记本电脑或其它类型的计算机或通信装置)、在这些装置中的一个上运行的线程或进程、和 / 或可由这些装置中的一个执行的对象。

[0025] 服务器 120 可包括一个或多个收集、处理、维护和 / 或管理信息的服务器实体。在与本发明原理相一致的示例实施方式中, 服务器 120 可包括日志处理系统 125, 该系统处理和存储与例如与客户端 110 访问网页相关联的网页访问 (如与点击有关的信息) 相关联的信息。

[0026] 服务器 130 可包括一个或多个服务器实体, 该服务器实体收集、处理、搜索、维护、管理文档和 / 或向用户提供搜索结果。在一个与本发明相一致的示例实施方式中, 服务器 130 可包括可由客户端 110 使用的搜索引擎系统 135。搜索引擎系统 135 可从客户端 110 接收搜索查询并将相关信息返回到客户端 110。

[0027] 网络 140 可包括局域网 (LAN)、广域网 (WAN)、电话网 (如公共交换电话网 PSTN)、企业内部互联网 (INTRANET)、因特网、不同类型的网络、或网络的组合。客户端 110 和服务器 120 和 130 可以通过有线、无线或者光线路连接到网络 140。在可替代的实施方式中, 图 1 中图示的一个或多个装置可以直接连接到另一个装置。例如, 在一个与本发明原理相一致的实施方式中, 服务器 120 可以直接连接到服务器 130, 如图 1 中的点划线所示。

[0028] 服务器结构示例

[0029] 图 2 是与本发明原理相一致的实施方式中的服务器 120 的示例图。服务器 130 也可以用类似方式进行配置。服务器 120/130 可包括总线 210、处理器 220、主存储器 230、只读存储器 (ROM) 240、存储装置 250、一个或多个输入装置 260、一个或多个输出装置 270、通信接口 280。总线 210 可包括允许服务器 120/230 的组件之间进行通信的路径。

[0030] 处理器 220 可包括解释和执行指令的任何类型的传统处理器、微处理器或处理逻辑。主存储器 230 可包括存储由处理器 220 执行的信息和指令的随机存取存储器 (RAM) 或其它类型的动态存储装置。ROM 240 可包括存储由处理器 220 使用的静态信息和指令的传统 ROM 装置或其它类型的静态存储装置。存储装置 250 可包括磁和 / 或光记录介质和对应的驱动。

[0031] 输入装置 260 可包括一个或多个传统的、允许用户将信息输入到服务器 120/130 的机构, 例如键盘、鼠标、笔、语音识别和 / 或生物特征识别机构等。输出装置 270 可包括一个或多个传统的将信息输出给用户的机构, 例如显示器、打印机、扬声器等。通信界面 280 还包括类似收发机的使服务器 120/130 可与其它装置和 / 或系统通信的机构。例如, 通信界面 280 可包括通过网络 (例如网络 140) 与其它装置或系统通信的机构。

[0032] 与本发明原理相一致, 服务器 120 可存储与特定网站内的个别网页相关联的数

据，并生成与网站相关联的网站映射信息。与本发明原理相一致，服务器 130 可根据来自客户端 110 的输入和存储在服务器 120 内的访问信息执行搜索。服务器 120/130 可以对处理器 220 执行包含在计算机可读介质例如存储器 230 中的软件指令作出响应执行这些操作。可将计算机可读介质定义为一个或多个存储装置和 / 或载波。

[0033] 可从其它计算机可读介质如数据存储装置 250 或通过通信接口 280 从其它装置将软件指令读入存储器 230。包含在存储器 230 中的软件指令使处理器 220 执行后面将描述的过程。可选地，可使用硬连接电路来代替或将其与软件指令组合在一起实施与本发明原理相一致的过程。因此，与本发明原理相一致的实施方式并不局限于任何特定的硬件电路和软件的组合。

[0034] 另外，在图 1 中将服务器 120 和服务器 130 图解为单个装置。然而，应当了解，服务器 120 和服务器 130 每一个都可以用计算机处理器和 / 或服务器的网络的形式来实施。

[0035] 图 3 是根据在与本发明原理相一致的实施方式的日志处理系统 125 的示例功能框图。日志处理系统 125 可包括前端 (FE) 310、日志数据存储器 320、处理组件 330 和网站映射存储器 340。图 3 中图示的逻辑块可用软件、硬件或硬件和软件的组合来实现。另外，每个功能块可以代表一个或多个计算机处理器、线程和 / 或对象。应当了解，在替代实施方式中，由图 3 中的一个逻辑块执行的功能可由其它逻辑块执行。另外，在替代实施方式中，单个逻辑块 / 处理装置可执行日志处理系统 125 的功能。

[0036] FE310 充当日志处理系统 125 的前端，并且接收与客户端 110 访问不同网页 / 站相关联的信息，例如点击信息。例如，当客户端 110 访问网页时，FE310 可接收超文本传输协议 (HTTP) 请求。FE310 可将该信息转发给日志数据存储器 320。

[0037] 日志数据存储器 320 可存储与客户端 110 访问不同网页和 / 或网站相关联的数据。例如，日志数据存储器 320 可包括表示已由客户端 110 访问过的网页的信息，例如与网页相关联的超链接。日志数据存储器 320 还可存储指示客户端 110 已经访问过的对应于超链接的每个网页的总次数和 / 或客户端 110 已经在每个网页上停留的时间（例如平均时间）的信息。日志数据存储器 320 还可包括表示指向在日志数据存储器 320 中所识别的每个特定网页的链接数量的信息。可选地，或除了上述信息以外，日志数据存储器 320 可包括指示典型用户 110 是否滚读了在日志数据存储器 320 中所识别的网页或者没有滚读就链接出该网页的信息。在其它的替代中，或除了上述信息以外，日志数据存储器 320 可存储与在日志数据存储器 320 中所识别的每个网页相关联的信息检索分值，其中信息检索分值指示特定的搜索查询与网页上的信息的匹配有多接近。在其它的替代中或除了上述信息以外，日志数据存储器 320 可存储识别典型的客户端 110 作出与在网页上显示的商品相关联的购买的可行性的信息。做出购买的可能性可以由与特定网页相关联的实体（如公司）提供或者从用户的日志提供。在任一种情况下，可以使用存储在日志数据存储器 320 中的信息来确定与特定网站中的多个网页相关联的质量度量，如下面所详细描述。

[0038] 处理组件 330 可分析在日志数据存储器 320 中的信息并生成网站映射数据 / 表。例如，处理组件 330 可使用在日志数据存储器 320 中的信息，根据质量因素将单个网站中的特定网页进行排名。质量因素可以是根据上述多个度量中的任一个，或上述度量的组合。例如，在一个实施方式中，处理组件 330 可根据日志数据存储器 320 中所记录的每个网页已被访问的次数来将网页排名。在另一个实施方式中，处理组件 330 可根据存储在日志数据存

储器 320 中的,其它网页中指向各个网页的链接的数量来将网页排名。在每种情况下,处理组件 330 可将已排名或已排序的列表存储在网站映射存储器 340 中。

[0039] 如上所述,网站映射存储器 340,可包括网站和各个网站中网页的排名评定 (ranking) 的列表。例如,图 4 图示了网站映射存储器 340 的示例部分。参考图 4,网站映射存储器 340 可包括表目 400 和多个与表目 400 相关联的子表目 400-1 到 400-5。图 4 所图示的实例中,表目 400 存储表示与惠普相关联的网站 “HP. com”。子表目 400-1 到 400-5 存储支持和驱动、打印和多功能、笔记本电脑和平板个人电脑、家庭和家庭办公以及合作者 & 开发人员,等等。可按顺序将子表目 400-1 到 400-5 进行存储,其中该顺序表示所确定的与网站 www. hp. com 中对应的网页相关联的质量。例如,表目 500-1(即支持和驱动)可表示 www. hp. com 网站上最受欢迎的网页(例如,如存储在日志数据存储器 320 中的数据所指示的那样),接着是条目 400-2 中的打印和多功能,到可表示在区域 400 中所识别的网站上受欢迎度排第 5 位的网页的表目 400-5(即合作和开发)。网站映射存储器 340 可存储多个其它网站的类似信息。在本实例中,表目 400-1 到 400-5 可基于日志数据存储器 320 已接收的指示客户端 110 已访问过这些网页的信息的次数。然而,如上所述特定网站中网页的排名评定 / 排序可以基于任意数目的其它因素或因素的组合。在每种情况下,与特定网站中的网页相关联的网站映射信息可代表与网页相关联的质量度量。质量度量可表示与网页相关联的受欢迎程度,网页上的信息被用户访问的可能性,信息对提交搜索查询的用户有用的可能性,或者可表示与网页质量相关联的其它因素。

[0040] 搜索引擎系统 135(图 1)可使用网站映射存储器 340 中的信息,以提供带有与特定搜索结果相关联的一个或多个网页的链接的搜索结果。参考图 5,在示例实施方式中,搜索引擎系统 135 可包括搜索组件 510 和映射组件 520。搜索组件 510 可从客户端 110 处接收搜索查询。然后,搜索组件 510 可根据搜索查询生成识别相关文档的初始搜索结果。

[0041] 映射组件 520 可从搜索组件 510 处接收初始搜索结果,并访问网站映射存储器 340 来确定是否有搜索结果与存储在网站映射存储器 340 中的映射信息相关联。对于那些有映射信息存储在网站映射存储器 340 中的搜索结果来说,映射组件 520 可包括所有或一些与初始搜索结果相关联的映射信息。搜索引擎系统 135 然后可发送搜索结果和映射信息到提交原始搜索查询的客户端 110,更进一步的细节如下所述。

[0042] 示例处理

[0043] 图 6 是与本发明原理相一致的流程图,图示了生成网站映射信息的示例过程。以下描述集中在接收与用户 110 访问网页相关联的信息并向客户端 110 提供信息的一个或多个服务器。应当了解,这里描述的技术可被等同地应用于任何根据质量度量(例如受欢迎度,或者与网页或其它信息相关联的使用可能性的度量)将信息提供给用户的服务器,以及当然地应用于在其中提供给用户的信息的质量是重要的任何系统。例如,这里描述的技术可应用于根据与搜索存储在一个或多个数据库或存储库中的图书、期刊、图片数据或其它信息相关联的搜索查询将信息提供给用户。

[0044] 处理可由日志处理系统 125 通过网络 140 接收数据开始(步骤 610)例如,当客户端 110 访问不同网站时,前端 310 可接收数据。在一个实施方式中,假设用户已经在他们各自的客户端 110 上下载 / 安装了使得在搜索引擎如搜索引擎系统 135 上的网络搜索变得便利的工具条。在这种情况下,工具条可包括软件代码,该软件代码指令客户端 110,对客户端

110 访问的每个网页,将超文本传输协议 (HTTP) 请求发送给服务器 120。FE310 可使用 HTTP 请求中的信息来识别与客户端 110 访问过的网页相关联的特定的网页和网站。可选地,当客户端 110 点击由搜索引擎系统 135 提供的链接时,FE310 可接收类似的数据。另外,从客户端 110 接收的信息可使 FE310 识别与网站访问相关联的其它信息,例如客户端 110 访问特定网页的时间量、客户端 110 是否滚动了特定网页、是否通过特定网页进行购买,等等。

[0045] FE310 可将与所访问的网页相关联的网页信息如超链接或统一资源定位器 (URL) 以索引过的格式存储在日志数据存储器 320 中 (步骤 620)。例如,假设用户访问苹果电脑的网站并且点击了可通过苹果电脑主页访问的多个链接,例如与 PC 相关联的网页、与苹果在线商店相关联的网页,等等。对于每个被访问的主页,FE310 可以收到 HTTP 请求,识别与该请求相关联的超链接并存储该网页的超链接。FE310 可将超链接存储在与超链接相关联的主页 (即本例中的 Apple. com) 的表目下。FE310 也可以存储与超链接相关联的文本,例如与 HTTP 请求一同提供的网页名称。可选地,FE310 可将锚文本与超链接一起进行存储,例如在用来从网站上的其它网页例如主页访问本网页的超链接上显示的文本。在另一个替代中,FE310 可将位于网页本身的文本与对应的超链接 (例如,位于网页顶端、可能表示网页主题的信息) 一起存储。

[0046] 在一个示范实施方式中,任何客户端 110 每次访问存储在日志数据存储器 320 中的特定网页时,FE310 可增加与特定网页相关联的索引值 (步骤 630)。该索引值可表示特定网页被访问的次数并且可起到作为与特定网页相关联的质量度量的功能。也就是,被访问得更频繁的网页更可能包含对用户有用的信息。如上所述,应当理解,处理组件 330 可根据多个其它因素或因素的组合生成已排序的列表。

[0047] 然后,处理组件 330 可根据特定的质量因数生成已排序或已排名的列表 (步骤 640)。在以上讨论的实例中,质量因素可基于客户端 110 已访问特定网站内的网页的次数,如日志存储器 320 所记录的那样。这种度量可帮助确定那些最有可能包含对客户端 110 的用户有价值的信息的网页。

[0048] 处理组件 330 然后可将已排序 / 已排名的列表存储在网站映射存储器 340 中 (步骤 650)。处理组件 330 对于每个存储在日志数据存储器 320 中的其它网站生成类似的已排序的列表,并且将这些已排名的列表存储在网站映射存储器 340 中 (步骤 650)。然后,当向客户端 110 提供搜索结果时,搜索引擎系统 135 可使用网站映射信息。

[0049] 图 7 图示了与本发明原理相一致的与提供搜索结果相关联的示例处理。以下描述的处理使用通过网络例如网络 140 提供搜索结果的实例。然而,可将这里描述的技术应用于任何包括本地搜索系统的搜索系统。处理可从客户端 110 通过网络 140 访问服务器 130 开始,并接收用于输入搜索查询的用户界面 (UI)。可选地,如上所述,客户端 110 可具有装载在客户端装置 110 上的、与搜索引擎系统 135 相关联的便于输入搜索查询的工具条。

[0050] 假设用户通过客户端 110 输入搜索查询,并且通过网络 140 将搜索查询传输到服务器 130。搜索引擎系统 135 可接收搜索查询 (步骤 710)。在示例实施方式中,搜索组件 510 接收搜索查询、执行搜索并生成初始搜索结果列表 (步骤 720)。这些初始搜索结果根据该搜索查询可以包括广告。

[0051] 在搜索组件 510 已经生成初始搜索结果 (包括有关广告客户的信息) 之后,将初始结果转发到映射组件 520。然后,映射组件 520 可访问网站映射存储器 340 以确定是否

有搜索结果与存储在网站映射存储器 340 中的网页信息相关联（步骤 730）。例如，假设来自客户端 110 之一的查询是“hp”。在这种情况下，进一步假设网站映射存储器 340 包括与惠普网站相关联的网页列表，如图 4 中所示的已排序的列表。在这种情况下，映射组件 520 将该已排序的列表（即表目 400-1 到 400-5）识别为与对惠普的搜索结果相关联。对每个初始搜索结果，映射组件 520 可执行类似搜索，并且对每个包括存储在网站映射存储器 340 中的信息的搜索结果识别网站映射信息（步骤 740）。在初始搜索结果没有识别网站主页的情况下，可将网站标识符（例如图 4 中的表目 400）从初始搜索结果中确定并存储为指向位于相同网站中的其它网页的索引。

[0052] 在示例实施方式中，映射组件 520 仅可以从网站映射存储器 340 识别预定数量（例如 5 个或更少）的链接以包含在每个特定搜索的结果中。该预定数量可表示最高质量的链接。然而，应当理解，可将存储在网站映射存储器 340 中的任何数量的链接（例如超过 5 个）与特定搜索结果一起提供。另外，在一些实施方式中，可以将存储在网站映射存储器 340 的链接仅与预定数目（例如 10）的搜索结果一起提供。

[0053] 对于可能与具有多个网页的网站不相关联的搜索结果，对于该特定搜索结果，映射组件 520 可不识别存储在网站映射存储器 340 中的任何信息。在这种情况下，由搜索组件 510 生成的搜索结果将不包括任何其它的网页信息或超链接。另外，如果没有搜索结果包括网站映射存储器 340 中的信息，将把由搜索组件 510 生成的初始搜索结果提供给客户端 110。

[0054] 在映射组件 520 已识别了与初始搜索结果相关联的合适的网站映射信息之后，搜索引擎系统 135 可将搜索结果和所识别的网站映射信息转发给客户端 110（步骤 750）。也就是，转发给客户端 110 的搜索结果将包括在步骤 720 生成的初始搜索结果，以及在步骤 740 由映射组件 520 识别的其它信息。

[0055] 例如，图 8 图示了提供给客户端 110 的示例搜索结果屏幕 800。参考图 8，假设由客户端 110 提交给搜索引擎系统 135 的搜索查询是“hp”，如框 810 所示。在这种情况下，搜索引擎系统 135 可识别与搜索查询相关联的多个命中项。为简便，图 8 只示出了 3 个搜索结果。参考图 8，在屏幕 800 上的区域 820 中提供的第一命中项与 www.hp.com（即惠普主页）相关联，并包括指向 www.hp.com 的超链接。

[0056] 另外，与 HP.com 相关联的搜索结果包括在区域 830 中提供的五个其它链接。这五个其它链接中的前四个与步骤 740 中由映射组件 520 识别的网页 / 链接相对应。在一个示例实施方式中，这些链接可与 www.hp.com 中的、存储在网站映射存储器 340 内的四个被访问次数最多的网页相对应（例如支持和驱动、打印和多功能、笔记本电脑和平板个人电脑、家庭和家庭办公）。如前所述，这些链接可包括在网站 www.hp.com 上提供的锚文本，用来链接到这些网页（例如，支持和驱动、打印和多功能、笔记本电脑和平板个人电脑、家庭和家庭办公）。可选地，在区域 830 中提供的文本可包括与网页相关联的 URL 或位于网页本身的文本。在每种情况下，文本可使得用户较为容易地确定对应网页中将包含何种信息，并且可节省与访问不感兴趣的网页相关联的时间。文本和超链接允许用户选择特定链接，而不用首先前往与初始搜索结果（如主页）相关联的网页并从多个网页中导航直到用户发现特定的感兴趣的网页为止。换言之，在客户端 110 的用户可绕过区域 820 的链接，并在区域 830 中直接选择他 / 她相信提供了需要的信息的链接。这将大大节省与寻找感兴趣商品相关联

的时间,尤其是在第一访问页面令人迷惑并且难以从其导航以发现需要的感兴趣的网页的链接的情况下。

[0057] 在区域 830 的第五链接（即在该例中的“来自 www. hp. com 的更多结果”）可以用来指示可以获得与惠普网站相关联的其它链接。例如,如以上关于图 4 的描述,对于惠普网站,可将合作者和开发人员(表目 400-5)的链接存储在网站映射存储 340 中,在区域 830 中未显示该表目。在示例实施方式中,当选择“来自 www. hp. com 的更多结果”时,可将与搜索与区域 820 中的链接相关联的网站相关联的搜索页提供给客户端 110。用户然后可输入新查询,并且搜索引擎系统 135 将根据新查询动态生成新结果,其中每个结果都是指向与区域 820 中的搜索结果相关联的网站中的网页的链接。在替代实施方式中,选择“来自 www. hp. com 的更多结果”的链接,可简单地将存储在网站映射存储器 340(例如合作者和开发人员)中的附加链接提供给客户端 110。所述附加链接的数量可以设置为预定数量。

[0058] 在替代实施方式中,在步骤 750 提供给客户端 110 的搜索结果屏幕,可将显示术语“网站地图”或类似名称的图标或链接与搜索结果一起提供。例如,除了在区域 830 提供链接,搜索结果屏幕可以在区域 830 提供“网站地图”的图标或链接。当客户端 110 的用户选择该图标或链接或者将鼠标停留在该图标或链接上时,可通过覆盖窗口或通过显示附加链接的新界面屏幕将存储在网站映射存储器 340 的附加链接显示给用户。使用图标或链接而不是提供附加链接可以节省搜索结果页面的空间,并允许显示更多搜索结果。在其它的替代中,可在与客户端 110 相关联的工具条上提供与网站映射功能相关联的工具条按钮。工具条按钮可与使得将搜索查询输入到搜索引擎系统 135 变得便利的工具条相关联。当选择工具条按钮时,可类似地提供附加链接的列表。使用此种工具按钮可节省搜索结果页上的附加空间。

[0059] 结论

[0060] 与本发明原理相一致的实施方式可包括带有搜索结果的网站映射信息。这种附加信息可基于任何质量度量并可允许用户更快地找到感兴趣的信息。另外,向用户提供网站映射信息的过程可以用对用户透明的方式来执行。

[0061] 上述本发明示例实施例的描述提供了图示和描述,但并非是要穷尽或将本发明限定在所公开的特定形式。根据上述教导是可以进行修改和改变的,并且可以从本发明的应用中获得这种修改和变化。例如,尽管关于图 6 和 7 描述了一系列步骤,这些步骤的顺序可以在与本发明一致的其它实施方式中被改变。另外,非从属性的步骤可以被并行实现。

[0062] 并且,已就收集和存储与客户端 110 访问网站相关联的信息描述了与本发明原理一致的实施方式。在其它实施方式中,不是考虑一个网站中的所有网页,而是可以将网站映射存储器 340 中存储的网站映射信息限定为链接到与初始搜索结果相关联的网页的网页。

[0063] 在另外实施方式中,网站提供者可将他们自己的网站内最受欢迎的网页列表提供给搜索引擎系统 135。例如,当搜索引擎生成识别网站提供者的网站的搜索结果时,网站提供者可以指令搜索引擎提供者(例如,与搜索引擎系统 135 相关联的提供者)提供指向受欢迎网页的链接的列表。该链接列表可基于网站提供者的特殊偏好和对于他们自己网站的了解。例如,网站提供者可能希望用户访问那些更可能发生购买的网页。由网站提供者提供链接 / 网页列表可以消除由日志处理系统 125 执行的一些跟踪的需求。

[0064] 另外,在替代实施方式中,日志处理系统 125 可就每个特定客户端 110 访问不同网

站的历史记录去跟踪该特定客户端 110 执行的动作。在这种情况下，网站映射存储器 340 可根据特定客户端 110 的用户历史记录将一个网站内的网页进行排名，并且对于每个特定的客户端 110 的网站映射信息可以是不同的。

[0065] 在再另外的实施方式中，与搜索结果中的特定个体（如图 8 中的区域 830）一起提供的链接，可能最初没有与搜索结果一起被提供。在这种实施方式中，在已点击搜索结果之后可以提供链接。这在目的站点并不包括网站地图或所提供的网站地图是令人迷惑的或难于导航的情况下可能是有用的。另外，在一些实施方式中，在已经点击搜索结果之后，可将单独的窗口自动弹出给用户。该窗口可包括附件链接的列表。在再另外的实施方式中，在提供附加链接之前可以向在客户端 110 的用户询问他 / 她是否想看网站地图（即，来自网站映射存储器 340 的链接列表）。

[0066] 如上所述，在附图中图示的实施方式中，对本发明所属技术领域的普通技术人员显而易见的是，可以用不同形式的软件、固件、硬件来实现本发明的各个方面。这些用来实现与本发明原理相一致的所述各个方面的实际软件代码或专用控制硬件并不限于本发明。所以，没有引用具体的软件代码来描述所述各个方面的运行和行为一应了解，本领域的普通技术人员能够根据这里描述设计软件和控制硬件来实现所述各个方面。

[0067] 除非特别指明，在本应用的描述中使用的单元、步骤或指令都不应当被解释为对于本发明是至关重要或是必需的。并且，如在此使用的，不加数量限定的项用于指一个或多个项。在只想表示一个项的地方，使用了术语“一个”或类似的语言。另外，除非明确说明，短语“根据”意为“至少部分根据”。

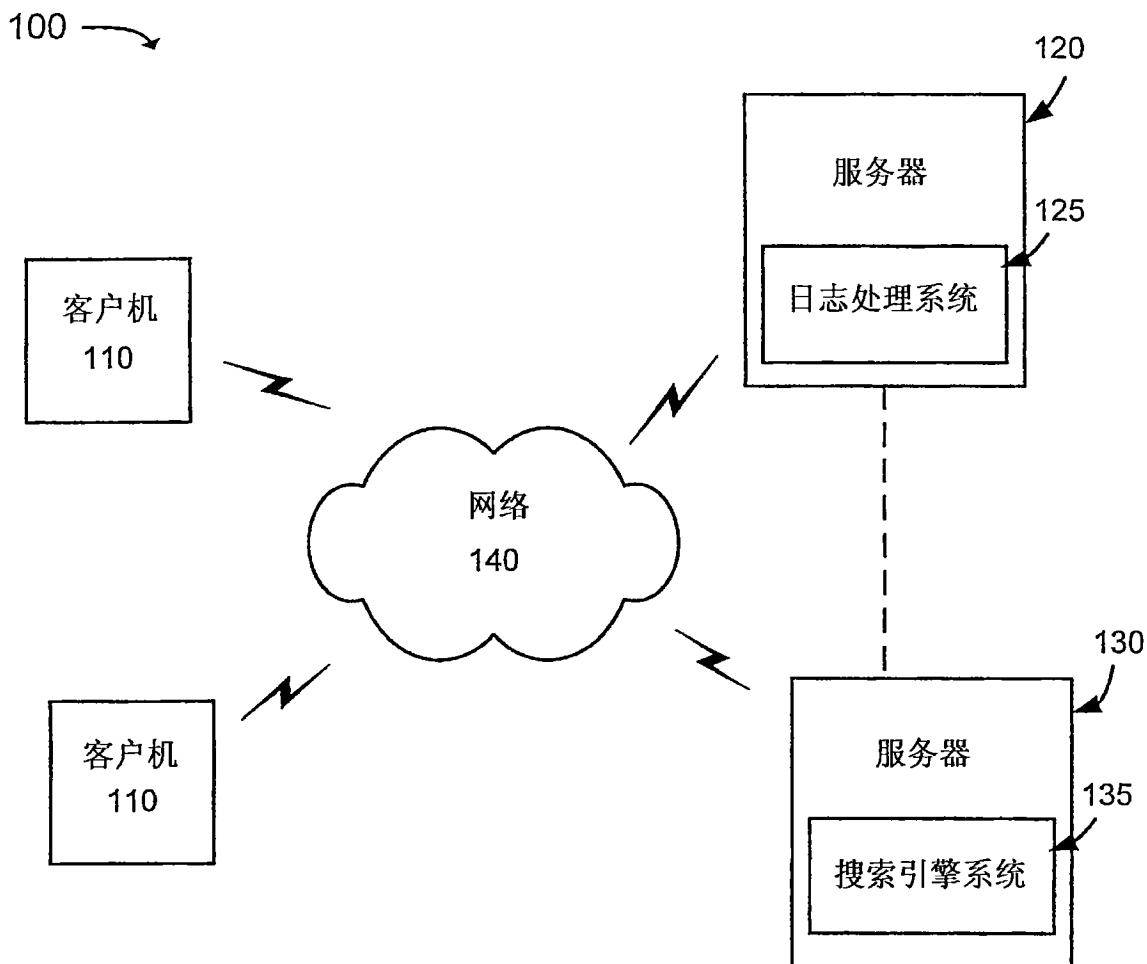


图1

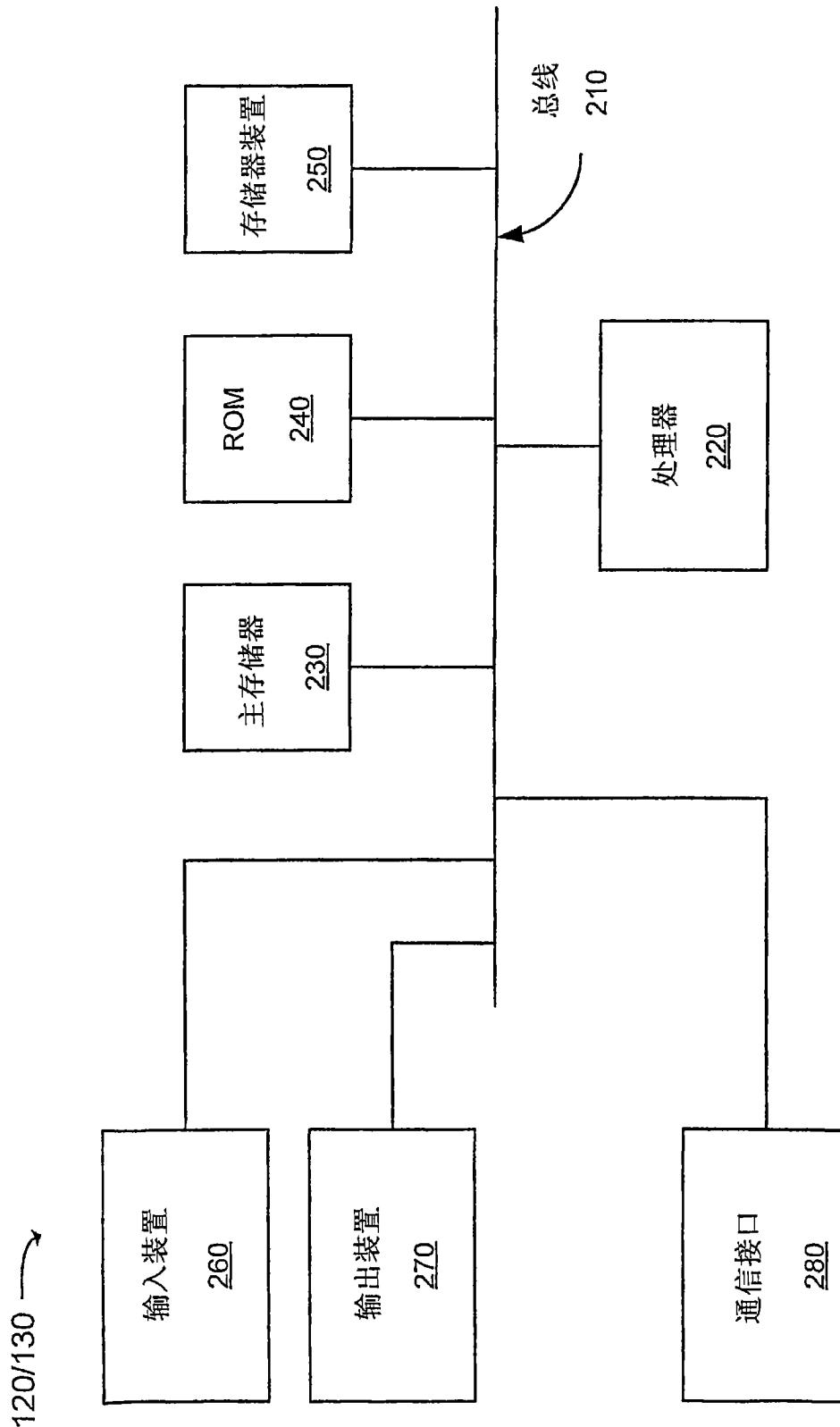


图2

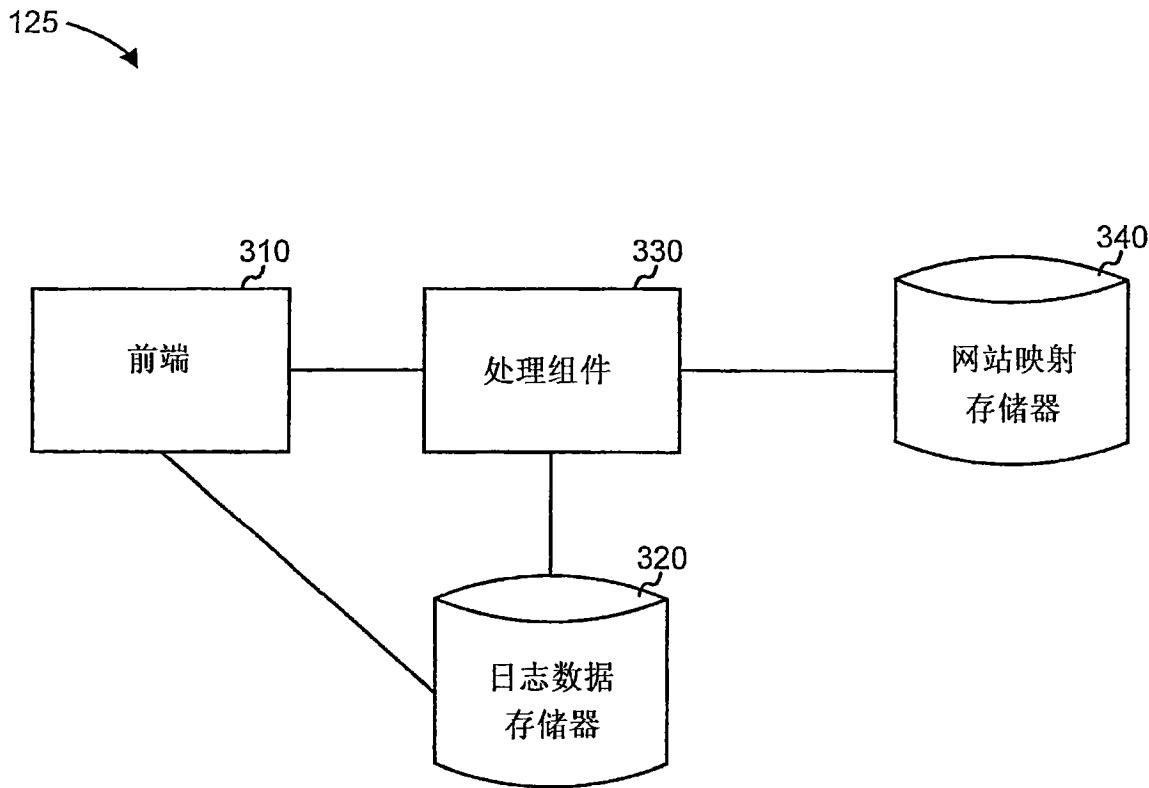


图3

340

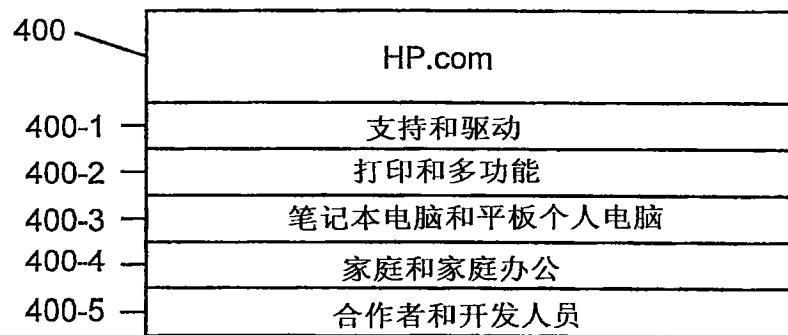


图4

135

搜索组件

510

初始结果

映射组件

520

搜索结果

搜索查询

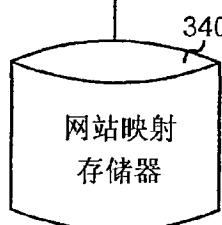


图5

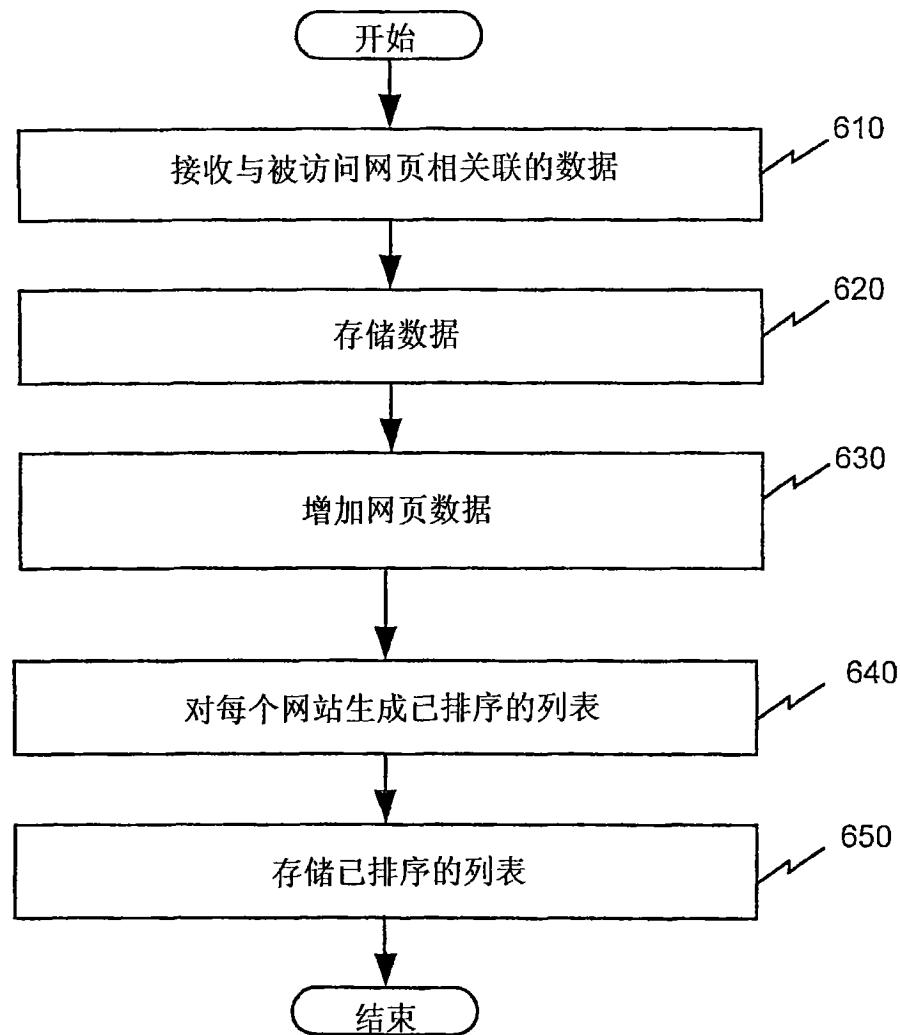


图6

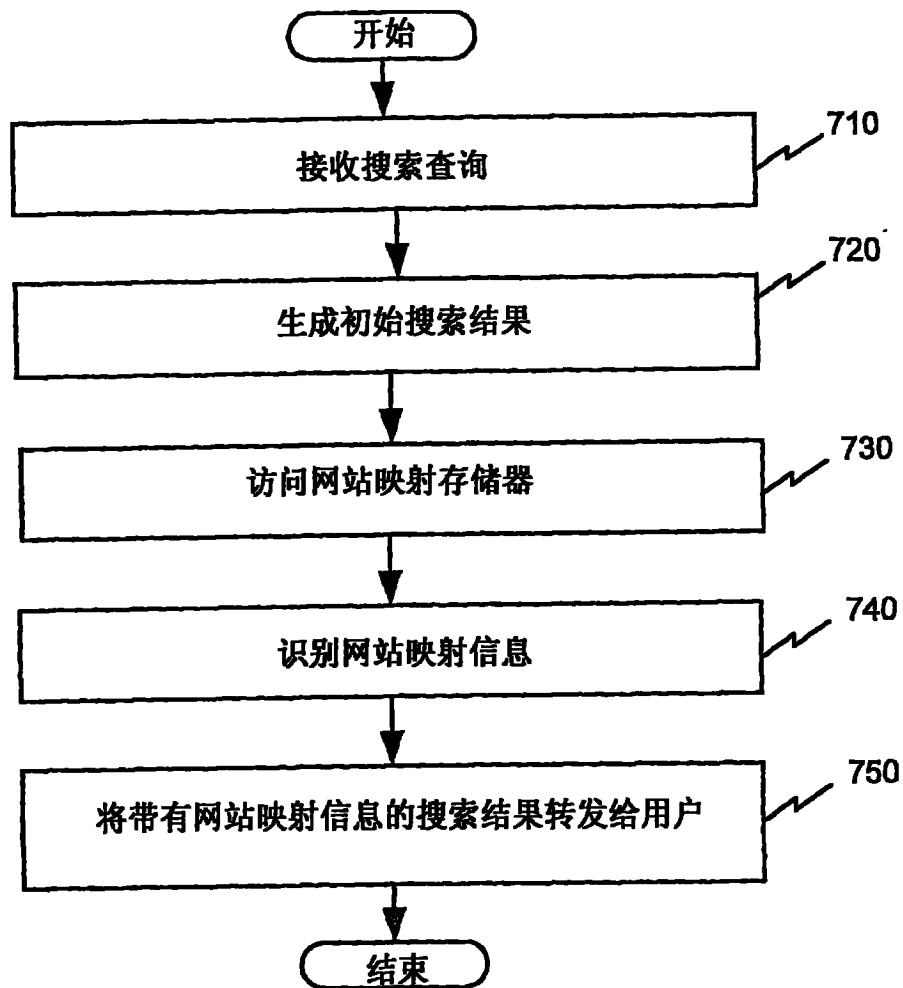


图7

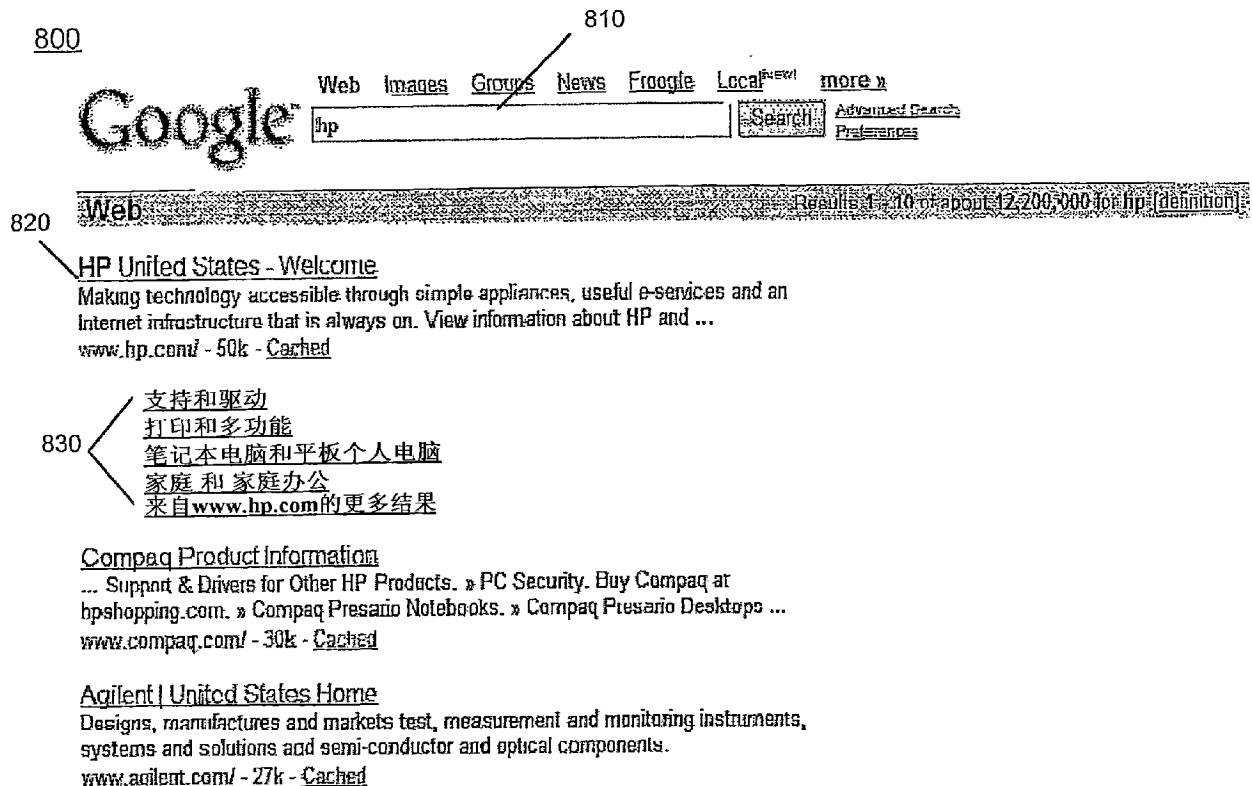


图8