

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.

G06F 15/16 (2006.01)

G06F 9/06 (2006.01)



[12] 发明专利说明书

专利号 ZL 01142517.2

[45] 授权公告日 2006年2月8日

[11] 授权公告号 CN 1241131C

[22] 申请日 2001.11.29 [21] 申请号 01142517.2

[30] 优先权

[32] 2000.11.30 [33] US [31] 09/726,985

[71] 专利权人 国际商业机器公司

地址 美国纽约州

[72] 发明人 戴维·M·莫里茨

审查员 于平

[74] 专利代理机构 北京市柳沈律师事务所

代理人 黄小临 王志森

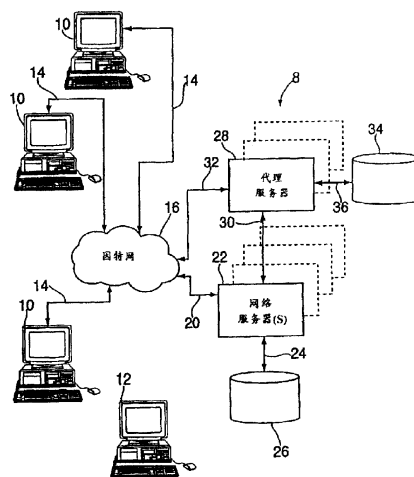
权利要求书 4 页 说明书 11 页 附图 5 页

[54] 发明名称

将资源从网络服务器提供到客户计算机的方法和系统

[57] 摘要

将 OFFLINE 请求从客户计算机发送到网络服务器。网络服务器确定由 <URL> 指定的网页和深度深到 <depth> 的所有链接(引用)网页。网络服务器还确定这些网页所需的所有图形、音频片段和其它资源,压缩这些资源,和创建包含压缩资源的档案文件。或者,网络服务器选择包括这些压缩资源的预封装档案文件。档案文件还包括网站的元数据和站点映像。档案文件由客户计算机来接收,并存储在以后可以在那里被解压缩的客户计算机的存储器中。档案文件使客户计算机能够脱机浏览网站和优化未来对资源的请求。



1. 一种将资源从网络服务器提供到客户计算机的方法，该方法包括：
从客户计算机接收单个请求，该单个请求识别所需网页；
5 将与所需网页相联系的多个资源包括在档案文件中；
在网络服务器中基于单个请求，生成站点映像，所述站点映像包括所需
网页；
限制该站点映像的范围；并且
响应单个请求，将包含站点映像的该档案文件发送到客户计算机。
- 10 2. 根据权利要求1所述的方法，其特征在于还包括：
将与所需网页相联系的多个资源压缩成档案文件。
3. 根据权利要求1所述的方法，其特征在于还包括：
从多个档案文件中选择一个档案文件。
4. 根据权利要求1所述的方法，其特征在于还包括：
15 将与附加网页相联系的多个资源包括在档案文件中。
5. 根据权利要求2所述的方法，其特征在于还包括：
从客户计算机接收深度值；
识别与所需网页相联系的多个附加网页；
利用深度值限制多个附加网页中的网页数量；和
20 将与有限数量的网页相联系的多个资源包括在档案文件中。
6. 根据权利要求1所述的方法，其特征在于还包括：
从客户计算机接收长度值；和
将档案文件的长度限制在该长度值上。
7. 根据权利要求1所述的方法，其特征在于还包括：
25 将来自所需网页的元数据包括在档案文件中。
8. 根据权利要求7所述的方法，其特征在于，元数据是从包括下列的一
组中选择出来的：
在所需网页中找到的关键词、所需网页的母网页、所需网页的子网页、
在所需网页中找到的链接、所需网页的管理联络点和在所需网页中找到的元
30 标记。
9. 根据权利要求1所述的方法，其特征在于还包括：

- 验证声明文件；和
将声明文件包括在档案文件中。
10. 根据权利要求 1 所述的方法，其中限制该站点映像的范围包括：
限制该站点映像的范围到包括含有所需网页 URL 的子串的 URL 的网页。
- 5 11. 根据权利要求 1 所述的方法，其特征在于还包括：
从客户计算机接收一个值；和
将站点映像中的网页数限制在该值上。
12. 根据权利要求 2 所述的方法，该方法进一步包括：
建立与网络服务器的连接；
- 10 将单个请求发送到网络服务器，该单个请求识别所需网页；
接收档案文件，档案文件包含与所需网页相联系的多个资源；
切断与网络服务器的连接；
解压缩与所需网页相联系的多个资源；
在所述切断连接之后，显示网页。
- 15 13. 根据权利要求 12 所述的方法，其特征在于，档案文件包含与附加网页相联系的多个资源，该附加网页与所需网页相链接，和其中，所述方法还包括：
在所述切断连接之后，显示附加网页。
14. 根据权利要求 12 所述的方法，其特征在于还包括：
- 20 在单个请求中指示长度值，该长度值指示档案文件的最大长度。
15. 根据权利要求 12 所述的方法，其特征在于，在所述接收档案文件中的档案文件包含关于所述网页的元数据，和其中，所述方法还包括：
在所述切断连接之后，搜索元数据。
16. 根据权利要求 15 所述的方法，其特征在于，元数据是从包括下列的
- 25 一组中选择出来的：
在所需网页中找到的关键词、所需网页的母网页、所需网页的子网页、
在所需网页中找到的链接、所需网页的管理联络点和在所需网页中找到的元
标记。
17. 根据权利要求 12 所述的方法，其特征在于，在所述接收档案文件中的
- 30 的档案文件包含包括所需网页的站点映像，和其中，所述方法还包括：
在所述切断连接之后，搜索站点映像。

18. 根据权利要求 1 所述的方法, 该方法进一步包括:
搜索站点映像。
19. 根据权利要求 18 所述的方法, 其特征在于还包括:
在单个请求中指示档案文件的最大长度。
- 5 20. 根据权利要求 18 所述的方法, 其特征在于还包括:
在单个请求中指示站点映像中最大网页数。
21. 根据权利要求 7 所述的方法, 该方法进一步包括:
搜索该元数据。
22. 一种将信息从网络服务器提供到客户计算机的系统, 该系统包括:
10 网络服务器;
存储设备, 与所述网络服务器耦合;
网站, 存储在所述存储设备中, 所述网站包括多个 HTML 网页和被所述多个 HTML 网页引用的多个资源;
网络, 与所述网络服务器相连接;
- 15 客户计算机, 与所述网络相连接, 所述客户计算机被配置成将单个 HTTP 请求提供给所述网络服务器, 所述单个 HTTP 请求识别所述网站中的所需 HTML 网页, 所述网络服务器被配置成基于单个 HTTP 请求生成站点映像, 所述站点映像包括所需 HTML 网页, 并且响应所述单个 HTTP 请求, 将包含站点映像的档案文件发送到所述客户计算机, 所述站点映像被限制到一定范围。
- 20 23. 根据权利要求 22 所述的系统, 其特征在于, 所述网络服务器被配置成将与所述所需 HTML 网页相联系的所述多个资源压缩成所述档案文件。
24. 根据权利要求 22 所述的系统, 其特征在于, 所述网络服务器被配置成从存储在所述存储设备中的多个档案文件中选择所述档案文件。
- 25 25. 根据权利要求 22 所述的系统, 其特征在于, 所述网络服务器被配置成将被所述 HTML 网页引用的所述多个资源包括在所述档案文件中。
26. 根据权利要求 23 所述的系统, 其特征在于, 所述网络服务器被配置成从所述客户计算机接收一个值, 识别从所述多个 HTML 网页中选择出来的一组 HTML 网页, 利用所述的值限制所述一组 HTML 网页中的 HTML 网页数量, 和将与所述一组 HTML 网页相联系的一组资源包括在所述档案文件中。
- 30 27. 根据权利要求 22 所述的系统, 其特征在于, 所述网络服务器被配置成从客户计算机接收长度值, 和将所述档案文件的长度限制在所述长度值上。

28. 根据权利要求 22 所述的系统, 其特征在于, 所述网络服务器被配置成将来自所述所需 HTML 网页的元数据包括在所述档案文件中。

29. 根据权利要求 28 所述的系统, 其特征在于, 所述元数据是从包括下列的一组中选择出来的:

5 在所需 HTML 网页中找到的关键词、所需 HTML 网页的母 HTML 网页、所需 HTML 网页的子 HTML 网页、在所需 HTML 网页中找到的链接、所需 HTML 网页的管理联络点和在所需 HTML 网页中找到的元标记。

30. 根据权利要求 28 所述的系统, 其特征在于, 所述网络服务器被配置成验证声明文件并将所述声明文件包括在所述档案文件中。

10 31. 根据权利要求 22 所述的系统, 其特征在于, 所述网络服务器被配置成从所述客户计算机接收 URL 的子串, 和生成所述站点映像以包括 HTML 网页, 该 HTML 网页含有包括所述子串的 URL。

15 32. 根据权利要求 22 所述的系统, 其特征在于, 所述网络服务器被配置成从所述客户计算机接收一个值, 和将所述站点映像中的 HTML 网页数限制在所述值上。

将资源从网络服务器提供到客户计算机的方法和系统

5

技术领域

本发明涉及计算机和计算机系统，尤其涉及将信息从网络（web）服务器提供到网络浏览器的方法和系统。

背景技术

10

因特网通过将网络服务器计算机与在客户计算机上运行的网络浏览器相连接，把全世界的计算机网络都汇集在一起。这种连接是通过称为传输控制协议/因特网协议（TCP/IP）的通信协议提供的。TCP/IP是因特网用来分割、路由和重构它管理的数据的分组交换方案（packet switching scheme）。

15

网络服务器是带有相关程序的计算机，它在因特网上将资源发送给相匹配的客户计算机，相匹配的客户计算机含有能够解释资源和作为网页显示它们的网络浏览器。资源是一些信息块，譬如，可以由统一资源定位器（URL）识别的图形或音频文件、资源在环球网（World Wide Web）上的全球地址。网页是超文本标记语言（HTML）文件和在 HTML 文件中识别的一个或多个资源。

20

在客户计算机上运行的网络浏览器通过将超文本传输协议（HTTP）请求发送到网络服务器获取资源。HTTP 定义如何格式化资源和在网络服务器与浏览器之间传输资源。HTTP 还定义网络服务器和浏览器为了响应各种命令，应该采取什么行为。HTTP 是一种无状态协议。也就是说，执行每条命令都与以前的和以后的命令无关。结果是，当用户访问网页时，用户的网络浏览器必须通过需要资源时作出的特定请求，请求每个网页、图形、内嵌项目、或其它资源。虽然可以在同一 TCP/IP 连接（利用 HTTP 1.1 中的持续性连接）上作出这些请求，但是，每个资源请求是独立的和分开的。例如，网络浏览器通常将发送众所周知的 HTTP 方法“GET”检索每个资源。GET（取）请求包括 URL 和正在使用的 HTTP 版本：

25
30

GET <URL> <使用的 HTTP 版本>

From: <用户 ID>

响应于 GET 请求，网络服务器返回由 URL 识别的资源。首标“From”用于识别作出请求的客户机。当使用 GET 请求时，必须在需要资源时利用单个请求，获取来自网络服务器的、客户计算机需要的每个资源。因此，网络浏览器必须发送多个 GET 请求，为单个网页检索所有必要的资源。产生的后果是，客户计算机和网络服务器二者要花很长时间管理消息。

在 HTTP 中，资源是严格地以按需方式获取的。还没有办法请求捆绑资源，或确定资源组之间是否是相关的。因此，如果客户计算机的用户试图脱机浏览（即，试图在断开因特网时，沿着链路继续浏览新网页），那么，客户机将无能为力，这是因为还没有获取未访问网页所需的资源。如果用户试图将网页存储在客户计算机的硬盘驱动器上，供脱机浏览用，那么，由于各条链路很可能会回指向网站，因此，HTML 文件和相关图形的大小占用大量的空间，往往不能在客户计算机上提供网页的可浏览副本。这对于手持计算机和蜂窝式电话之类，需要在没有与网络相连接的同时访问下载的资源的情况下“脱机”设备尤为麻烦。

发明内容

这些和其它缺点和缺陷可以通过把资源从网络服务器提供到客户计算机的一种方法来克服。根据本发明的一方面，提供一种将资源从网络（web）服务器提供到客户计算机的方法，该方法包括：从客户计算机接收单个请求，该单个请求识别所需网页；将与所需网页相联系的多个资源包括在档案文件中；在网络服务器中基于单个请求，生成站点映像，所述站点映像包括所需网页；限制该站点映像的范围；并且响应单个请求，将包含站点映像的该档案文件发送到客户计算机。

根据本发明的再另一方面，提供一种将资源从网络服务器提供到客户计算机的方法，该方法包括：从客户计算机接收单个请求，该单个请求识别所需网页；和响应单个请求，将包含元数据（metadata）的档案文件从所需网页发送到客户计算机。

根据本发明的再另一方面，提供一种将资源从网络服务器提供到客户计算机的方法，该方法包括：建立与网络服务器的连接；将单个请求发送到网络服务器，该单个请求识别所需网页；接收档案文件，档案文件包含与所需

网页相联系的多个资源；切断与网络服务器的连接；解压缩与所需网页相联系的多个资源；在所述切断连接之后，显示网页。

5 根据本发明的再另一方面，提供一种将资源从网络服务器提供到客户计算机的方法，该方法包括：将单个请求发送到网络服务器，该单个请求识别所需网页；接收包含站点映像的档案文件，该站点映像包括所需网页；和搜索站点映像。

根据本发明的再另一方面，提供一种将资源从网络服务器提供到客户计算机的方法，该方法包括：将单个请求发送到网络服务器，该单个请求识别所需网页；接收包含关于所需网页的元数据的档案文件；和搜索该元数据。

10 根据本发明的再另一方面，提供一种将信息从网络服务器提供到客户计算机的系统，该系统包括：网络服务器；存储设备，与所述网络服务器耦合；网站，存储在所述存储设备中，所述网站包括多个 HTML 网页和被所述多个 HTML 网页引用的多个资源；网络，与所述网络服务器相连接；客户计算机，与所述网络相连接，所述客户计算机被配置成将单个 HTTP 请求提供给所述网络服务器，所述单个 HTTP 请求识别所述网站中的所需 HTML 网页，所述网络服务器被配置成基于单个 HTTP 请求生成站点映像，所述站点映像包括所需 HTML 网页，并且响应所述单个 HTTP 请求，将包含站点映像的档案文件发送到所述客户计算机，所述站点映像被限制到一定范围。

20

附图说明

现在参照附图，仅通过举例的方式描述本发明，在附图中：

图 1 是并入了本发明方法的分布式计算机系统的图解表示；

图 2 是存储在图 1 所示的网络服务器上的网站的图解表示；

25 图 3 是包括本发明的 OFFLINE（脱机）方法请求的、客户计算机与网络（web）服务器之间的 HTTP 请求和响应序列的图解表示；

图 4 是包括本发明的 SITEMAP（站点映像）方法请求的、客户计算机与网络服务器之间的 HTTP 请求和响应序列的图解表示；和

30 图 5 是包括本发明的 METADATA（元数据）方法请求的、客户计算机与网络服务器之间的 HTTP 请求和响应序列的图解表示。

具体实施方式

现在参照图 1，在图 1 中，用 8 总体显示并入了本发明方法的分布式计算机系统的图解表示。还显示了通过象征性箭头线 14 连接到因特网 16 的多个客户计算机 10。这些连接 14 通常是通过现有技术中众所周知的局域网 (LAN) 或电话设备实现的。示出的移动客户计算机 12 与因特网 16 断开。相连的客户计算机 10 和不相连的移动式客户计算机 12 的功能基本相同。移动式客户计算机可以是，例如，IBM THINKPAD™，它运行 LINUX 作为操作系统、大量的商业应用程序、诸如 Microsoft™ Internet Explorer 之类的因特网浏览器和诸如 IBM Global Network Dialer 之类的因特网连接软件。在这个优选实施例中，浏览器能够翻译含有 HTML 标记和 Java 脚本的网页。不相连的移动式客户计算机 12 具有当用户需要时通过 LAN 或电话设备连接到因特网 16 的能力。

因特网可以直接在 20，或通过经由连接 30 和 32 的一个或多个代理服务器 28 连接到至少一个网络服务器 22。网络服务器 22 在 24 连接到至少一个硬盘 26。硬盘 26 包含操作系统、配置文件、日志文件、网络应用程序、网页、公共网关接口 (CGI) 脚本和关于网页的各种资源。用于本实施例的网页包含 HTML 标记和 Java 脚本。网络服务器 22 的 CGI 脚本应该能够执行如现有技术中众所周知的那样的，非 HTML 任务的数据处理技术。

代理服务器 28 在 36 连接到至少一个硬盘 34。硬盘 34 包含操作系统、配置文件、日志文件和网络应用程序。硬盘 34 还包含 HTTP 代理程序，HTTP 代理程序起客户计算机 10 和 12 与网络服务器 22 之间的中介作用。代理服务器 28 共同用作网络服务器和它的局域网 (LAN) 的防火墙。数据通过因特网 16 在客户计算机 16 与网络服务器 22 之间的流动通过一系列 HTTP 请求和响应来实施。操作客户计算机 10 的用户利用安装在客户计算机 10 上的网络浏览器程序，接通从例如电话设备到因特网 16 的连接 14。一旦处在这种连接模式中，用户就将网络服务器 22 的 URL 输入到网络浏览器中。网络浏览器打开到该网络浏览器的连接，然后，通过因特网 16 和通过连接 20 将 HTTP 请求消息发送到特定的网络服务器 22。HTTP 请求消息包括识别网络服务器的、硬盘 26 中所需资源的路径的 URL。网络服务器 22 从硬盘 26 中获取资源，并通过因特网 16 将包含所请求资源的响应消息返回给客户计算机 10 的浏览器。然后，将资源存储在网络浏览器可以检索它的客户计算机 10 的存储器器件

(ROM 或 RAM) 中。如果连接是通过代理服务器 28 接通的, 那么, 代理服务器 28 接收来自客户机的请求, 并将这些请求前送到目标网络服务器 22。各种响应按相同方式, 通过代理服务器 28 往回传递。

在 HTTP 1.0 和更早版本中, 在每次请求和响应之后都切断客户计算机 10 与网络服务器 22 之间的连接, 致使要检索每个资源都要请求新的连接。在 HTTP 1.1 中, 连接是持续性的, 使客户计算机 10 能够发送网络服务器 22 将在切断连接之前作出响应的一系列请求 (称为流水线操作)。

图 2 是存储在图 1 所示的网络服务器 22 中的网站 50 的图解表示。网站 50 包括母网页 52 和多个子网页 54、56 和 58。因为在本例中, 母网页 52 是客户机用户首次建立与网络服务器 22 的连接之后, 向他或她呈现的第一个网页, 所以母网页 52 是网站 50 的主页。

子网页 54、56 和 58 通过链路 60、62 和 64 与母网页相联系。一旦客户计算机 10 已经接收到母网页 52, 客户机用户就可以通过选择链路 60、62 和 64 的任何一条浏览网站 50, 观看所需子网页 54、56 和 58。子网页 54、56 和 58 和母网页 52 每一个都包括多个资源 66、68、70、72、74、76、78、80 和 82。资源 66、72、76 和 80 代表 HTML 文件, 资源 68、70、74、78 和 82 代表图形、音频片段等, 它们形成网页 52、54、56 和 58 的一部分。参照图 1 和图 2, 在客户计算机 10 上运行的网络浏览器通过将由 HTTP 方法 (命令) 和首标 (关于请求的信息) 组成的请求发送到网络服务器 22, 获取网页 52、54、56 和 58 的资源 66、68、70、72、74、76、78、80 和 82。正如上面所讨论的, 借助于现有技术的 HTTP 方法, 必须利用需要资源时作出的特定请求, 获取来自网络服务器 22 的、客户计算机 10 需要的每个资源 66、68、70、72、74、76、78、80 或 82。其结果是, 客户计算机 10 必须发送 9 个请求来检索所有必要的资源, 以观看网页 52、54、56 和 58。

本发明提供了 “OFF-LINE” HTTP 方法, 它是网络服务器 22 所支持的现有 HTTP 命令集的一个扩充。OFF-LINE 方法使网络浏览器 10 能够容易地定位网站 50 内的信息, 以提供脱机浏览。该方法还使网络服务器 22 能够管理网站 50 的内容。

图 3 是包括从客户计算机 10 发送到网络服务器 22 的 OFFLINE 方法请求的、客户计算机 10 与网络服务器 22 之间的 HTTP 请求和响应序列。当网络服务器 22 从客户计算机 10 接收请求 100 时, 它首先识别由 <URL> 指定的网页

和深度为 <depth> 的所有链接（引用）网页。深度是从 URL 引用的网页开始的先辈（ancestry）的级别。利用图 2 所示的网站 50 作为例子来说，深度 1 总是包括主页 52，深度 2 总是包括主页 52 和所有子网页 54、56 和 58，和深度 3 总是包括主页 52、所有子网页 54、56 和 58 以及网站 50 中与多个子网页直接链接的所有网页（孙网页）（未示出）。网络服务器还将确定 URL 指定的网页和到所指示深度的所有网页所需的所有图形、音频片段和其它资源。然后，网络服务器 22 创建包含所有必要资源的单个档案文件 102。例如，如果请求 100 中的 URL 引用了主页 52，并指示深度为 2，那么，档案文件 102 将包括网页 52、54、56 和 58 的资源 66、68、70、72、74、76、78、80 和 82。或者，网络服务器 22 将选择满足客户计算机请求的预封装档案文件 102。

档案文件 102 以已知的格式创建，譬如，用于创建 Java 档案（JAR）文件的格式，或者，最好是，譬如，美国专利第 5,937,411 号所描述的格式。这样的文件格式支持数据压缩，这会减少到客户计算机 10 的下载次数（times）。

除了网页和它们所需的资源之外，档案文件 102 还可以包含网站的元数据。元数据（关于网站数据的信息）可以包括象站点映像那样的信息，指示站点上网页之间的相互关系。可以存储在档案文件 102 中的其它信息包括：关键词、每个网页的母网页、当前网页中的所有链路、当前网页的引用资源、管理联络点（administrative contact）和元标记（meta-tag）。元标记是将有关网页的信息通常提供给搜索引擎的特殊 HTML 标记。存储在元标记中的信息通常包括谁创建了网页，每隔多久更新一次网页，该网页涉及到什么东西，和哪些关键词代表网页内容。

以 JAR 文件格式的档案文件 102 包括如现有技术中已知的那样的声明文件（manifest file）。声明文件包含关于 JAR 文件内其它文件的结构的信息。最好，声明文件还可以用于提供核实档案文件 102 的完整性的数字签名，以防止内嵌的内容受到篡改。

图 4 所示的 OFFLINE（脱机）请求消息 100 包括首标：OFFLINE - ACCEPT（接受）、OFFLINE - MAXSIZE（最大尺寸）和 OFFLINE - MAXPAGES（最大页数）。包括在这些首标中的信息通过驻留在客户计算机 10 中的网络浏览器输入，并由网络服务器 22 用来设置档案文件 102 的各种参数。OFFLINE - ACCEPT 首标包括在档案文件 102 中，要由网络服务器 22 提供的各种类型的多用途邮件扩

充部分 (multi-purpose mail extension, MIME)。MIME 的各种类型可以用逗号分开的格式列出。OFFLINE - MAXSIZE 首标设置档案文件 102 以千字节 (KB) 为单位的长度极限。OFFLINE - MAXPAGES 首标包括在档案文件 102 中的最大 HTML 网页数。

5 OFFLINE 请求消息 100 还可以包括如现有技术中已知的那样的 IF MODIFIED SINCE (此后是否修改) 首标。例如, 如客户计算机 10 以前已经请求了档案文件 102, 就使用 IF MODIFIED SINCE 首标。IF MODIFIED SINCE 首标让网络服务器 22 知道客户计算机 10 以前下载档案文件 102 的日期, 以便
10 自从那个日期之后, 如果没有对网站 50 作任何改变, 那么可以避免不必要的数据传输。另外, OFFLINE 请求消息 100 还可以包括如现有技术中已知的那样的 FROM (来自) 首标。FROM 首标识别客户计算机 10, 并且可以由网络服务器 22 用于安全的目的。例如, 网络服务器 22 可以拒绝某些客户计算机 10 的 OFFLINE 请求, 或者, 对于不同的用户可以响应不同的档案文件 102 (例如, 对于授权用户, 档案文件 102 可以包括更深的深度)。

15 在创建了档案文件 102 之后(或者, 在检索到预封装档案文件 102 之后), 网络服务器将包含档案文件 102 的响应消息发送给客户计算机 10。响应消息 104 包括, 例如, 响应的日期、包括在响应中的文件的类型 (例如, MIME 文件的类型)、内容长度 (以 KB 为单位) 和脚注。脚注可以用于提供数字签名, 以象现有技术中已知的那样, 核实响应的完整性和真实性。

20 档案文件 102 由客户计算机 10 来接收, 并存储在客户计算机的存储器中。同时, 客户计算机 10 可以切断它与因特网 16 的连接 14 并脱机工作 (从任何网络中退出来)。驻留在客户计算机 10 上的网络浏览器解压缩档案文件 102。解压的档案文件 102 包括客户计算机 10 的用户网页 52、54、56 和 58 所需的、对在 OFFLINE 请求消息 100 中规定的各种限制或由网络服务器 11 为 OFFLINE
25 请求规定的限制的所有资源 66、68、70、72、74、76、78、80 和 82。

元标记包括在档案文件 102 中使客户计算机 10 中的网络浏览器能够对网站 50 中的信息进行脱机搜索。例如, 档案文件 102 可以包括网站 50 中所有网页 52、54、56 和 58 的元标记和资源 66、68、70、72、74、76、78、80 和
30 82。但是, 由于可能受到客户计算机 10 中可用的存储器容量的限制, 因此, 解压缩档案文件 102 中每个资源 66、68、70、72、74、76、78、80 和 82 并不可行。元标记使客户计算机 10 中的浏览器起到搜索引擎的作用, 并识别包

含用户所需信息的特定网页 52、54、56 或 58。一旦找到网页，就可以与该网页所需的资源一道解压缩它。

元标记包括在档案文件 102 中还使客户计算机 10 中的网络浏览器优化未来从网络服务器计算机 22 的下载。客户计算机 10 中可用的存储器容量可能使请求包括网站 50 中所有网页 52、54、56 和 58 的资源的大档案文件 102 变得难以实现。但是，即使在档案文件中没有包括所有资源的情况下，档案文件也可以包括网站 50 中所有网页 52、54、56 和 58 的元标记。然后，浏览器可以搜索元标记，以识别包含所需信息的网页 52、54、56 或 58，如果该网页还没有在存储器中，则建立与网络服务器 22 的连接，请求包括所需资源的档案文件 102。

站点映像包括在档案文件 102 中进一步提供了客户计算机 10 的脱机搜索能力。站点映像使客户计算机的网络浏览器能够透过网站寻找包含所需内容的网页的位置。另外，网络浏览器可以利用站点映像优化未来对档案文件 102 的请求。例如，如果用户请求包括在多于一个网页中的信息，那么，网络浏览器可以参照站点映像确定这些网页是否是相互关联的(即，共享共同先辈)。如果这些网页是相互关联的，网络浏览器可以改变 OFFLINE 请求参数(例如，<URL>、<depth>、OFFLINE MAXSIZE、OFFLINE MAXPAGES)，以保证所有相关网页都包括在档案文件中。

应该认识到，包括站点映像或元标记的档案文件 102 对于联机浏览，同样是有用的。当联机浏览时，在客户计算机 10 上运行的网络浏览器可以参照存储的站点映像和/或元标记，容易地定位网站 10 中网页 52、54、56 或 58 内的信息，从而减少获取信息所需的 HTTP 请求次数。

除了 OFFLINE 方法(命令)之外，本发明还提供了“SITEMAP”和“METADATA”方法。SITEMAP 和 METADATA 方法使网络浏览器能够在 OFFLINE 命令将导致不必要的数据或 OFFLINE 命令受到网络服务器的限制的情况下，利用站点映像和元标记的长处。

图 4 是包括从客户计算机 10 发送到网络服务器 22 的 SITEMAP 方法请求 120 的、客户计算机 10 与网络服务器 22 之间的 HTTP 请求和响应序列。当网络服务器 22 从客户计算机 10 接收这个请求时，它首先确定由 <URL> 指定的网页、指定网页直到 <maxparents> 高度的所有母网页和指定网页直到 <maxchild> 深度的所有子网页。例如，如果 URL 识别出图 2 中的网页 72 具有二

者都被设成 1 的 <maxparents> 和 <max child>，那么，所得的站点映像将识别出母网页 52 和与网页 72 链接的直到深度 1 的所有子网页（未示出）。应该注意，网页 72 的兄弟姐妹网页不包括在这个站点映像中。在创建了站点映像之后，网络服务器 22 接着创建包含站点映像的单个档案文件 122。或者，

5 网络服务器 22 选择满足客户计算机的请求的预封装档案文件 122。

图 4 的 SITEMAP 请求消息包括首标 SITEMAP - ACCEPT、SITEMAP - MAXSIZE、SITEMAP - MAXPAGES。包括在这些首标中的信息是通过客户计算机 10 中的网络浏览器输入的，并由网络服务器 22 用来设置档案文件 122 的各种参数。SITEMAP - ACCEPT 首标包括 URL 要包括在站点映像中的、必须存在的子串，

10 从而限制了站点映像的范围。例如，如果 SITEMAP - ACCEPT 首标包括子串“www.uspto.gov/web”，那么，站点映像将包括含有包括这个子串的 URL 的所有网页（例如，“www.uspto.gov/web/offices.html”）。SITEMAP - MAXSIZE 首标设置档案文件 122 的长度极限（以 KB 为单位）。SITEMAP - MAXPAGES 首标包括要包括在站点映像中的最大网页数。SITEMAP 请求消息还可以包括上

15 面参照 OFFLINE 方法所述的 IF MODIFIED SINCE 首标和 FROM 首标。

在创建了档案文件 122 之后（或者在搜索了预封装档案文件 122 之后），网络服务器 22 将如上面参照图 3 所述那样的、包含档案文件 122 的响应消息 124 发送到客户计算机 10。档案文件 122 由客户计算机 10 来接收，并存储在存储器中。

图 5 是包括从客户计算机 10 发送到网络服务器 22 的 METADATA 方法请求 140 的、客户计算机 10 与网络服务器 22 之间的 HTTP 请求和响应序列。当网络服务器 22 从客户计算机 10 接收这个请求 140 时，它首先确定由 <URL> 指定的网页，然后复制它的元标记、或诸如关键词、母层、链路、引用资源等之类的其它信息。在复制了元数据之后，网络服务器 22 接着创建包含元数据的单个档案文件 142。或者，网络服务器 22 选择满足客户计算机请求的预封装档案文件 142。

20

25

图 5 所示的 METADATA 请求消息包括首标 METADATA - ACCEPT、METADATA - MAXSIZE、IF MODIFIED SINCE 和 FROM。包括在这些首标中的信息是通过客户计算机 10 中的网络浏览器输入的，并由网络服务器用来设置档案文件 142 的各种参数。METADATA - ACCEPT 首标包括 URL 要包括在档案文件 142 中的、必须存在的子串，从而限制了元数据的范围。METADATA - MAXSIZE 首标设置

30

档案文件 142 的长度极限 (以 KB 为单位)。IF MODIFIED SINCE 和 FROM 首标如上面参照 OFFLINE 方法所描述的那样。

在创建了档案文件 142 之后 (或者在搜索了预封装档案文件 142 之后), 网络服务器 22 将如上面参照图 3 所述那样的、包含档案文件 142 的响应消息 5 144 发送到客户计算机 10。档案文件 142 由客户计算机 10 来接收, 并存储在存储器中。

本发明扩充了网络服务器支持的当前 HTTP 命令集 (即各种方法和各种首标)。这些方法和首标提供了更容易定位网站中的信息的方法, 并且提供了脱机浏览。除了这些好处之外, 本发明还提供了在包括多个网络服务器或代理 10 服务器的网络中管理网站内容的简易办法。

再参照图 1、图 2 和图 3, 如果多个网络服务器 22 的每一个都包括它们自己的硬盘驱动器 26, 以及每个硬盘驱动器 26 存储了网站 50 它自己的副本, 那么, 每当对网站 50 加以改变时, 必须更新网站 50 的每个副本。本发明的 OFFLINE 方法提供了更新网站 50 的每个副本的办法。例如, 可以将一个网络 15 服务器 22 维持为主网络服务器, 以及把网站 50 的更新直接做成主网络服务器硬盘驱动器中网站 50 的副本。其它每一个网络服务器 22 可以在预定时间间隔向主网络服务器 22 发送 OFFLINE 请求 100, 以获取网站 50 的任何改变。或者, 任何网络服务器 22 都可以接收它存储的网站 50 的各种改变。如果对 20 存储在一个网络服务器 22 上的网站 50 作出改变, 那么, 那个网络访问服务器可以把档案文件 100 推入其它网络服务器 22, 以更新网络服务器 50 的所有副本。

在带有代理服务器 28 的系统中, 本发明的 OFFLINE 方法可以用于向代理服务器 28 提供元数据和各种资源, 从而使代理服务器 28 能够在本机上管理 HTTP 请求, 而不是将其传递给网络服务器 22。一旦从客户计算机 10 接收到 25 OFFLINE 请求 100, 代理服务器 28 将会重新映射 URL, 和向客户计算机 10 响应预封装档案文件 102。使代理服务器 28 能够在本机上管理各种请求将增加对代理器 28 的用户的响应时间, 并将减轻原网络服务器 22 上的负担, 同时保证内容是正确的和安全的。

OFFLINE 方法还可用于信息在从网络服务器 22 提供信息的提供者与在客 30 户计算机 10 上接收信息的使用者之间的商业分配。信息的使用者可以通过将 OFFLINE 请求 100 发送到网络服务器 22, 或者让网络服务器 22 把档案文件

102 推向使用者的计算机 10 周期性接收信息。

本发明可以以计算机实施的处理和实践那些处理的设备的形式具体化。本发明还可以以体现在有形介质中的、包含指令的计算机程序代码的形式具体化，有形介质是，譬如，软盘、CD-ROM（只读光盘存储器）、硬盘驱动器、5 或任何其它计算机可读存储介质，其中，当将计算机程序代码装载到计算机中，由计算机来执行时，计算机成为实践本发明的设备。本发明还可以以，例如，无论是存储在存储介质中的、装载到计算机中和/或由计算机执行的、还是经由光纤、或通过电磁辐射，在一些传输介质上，譬如，在电线或电缆上传输的计算机程序代码的形式具体化，其中，当将计算机程序代码装载到10 计算机中，由计算机来执行时，计算机成为实践本发明的设备。当在通用微处理器上实施时，计算机程序代码段配置微处理器，以创建专用逻辑电路。

本领域的普通技术人员应该明白，可以在权利要求书的范围和含义内，对本文所示的优选实施例作各种修改。虽然对本发明的描述是通过本发明的特定实施例实现的，但是，本发明并不受特定实施例的限制，而是更宽地由15 权利要求书的范围和精神所涵盖。

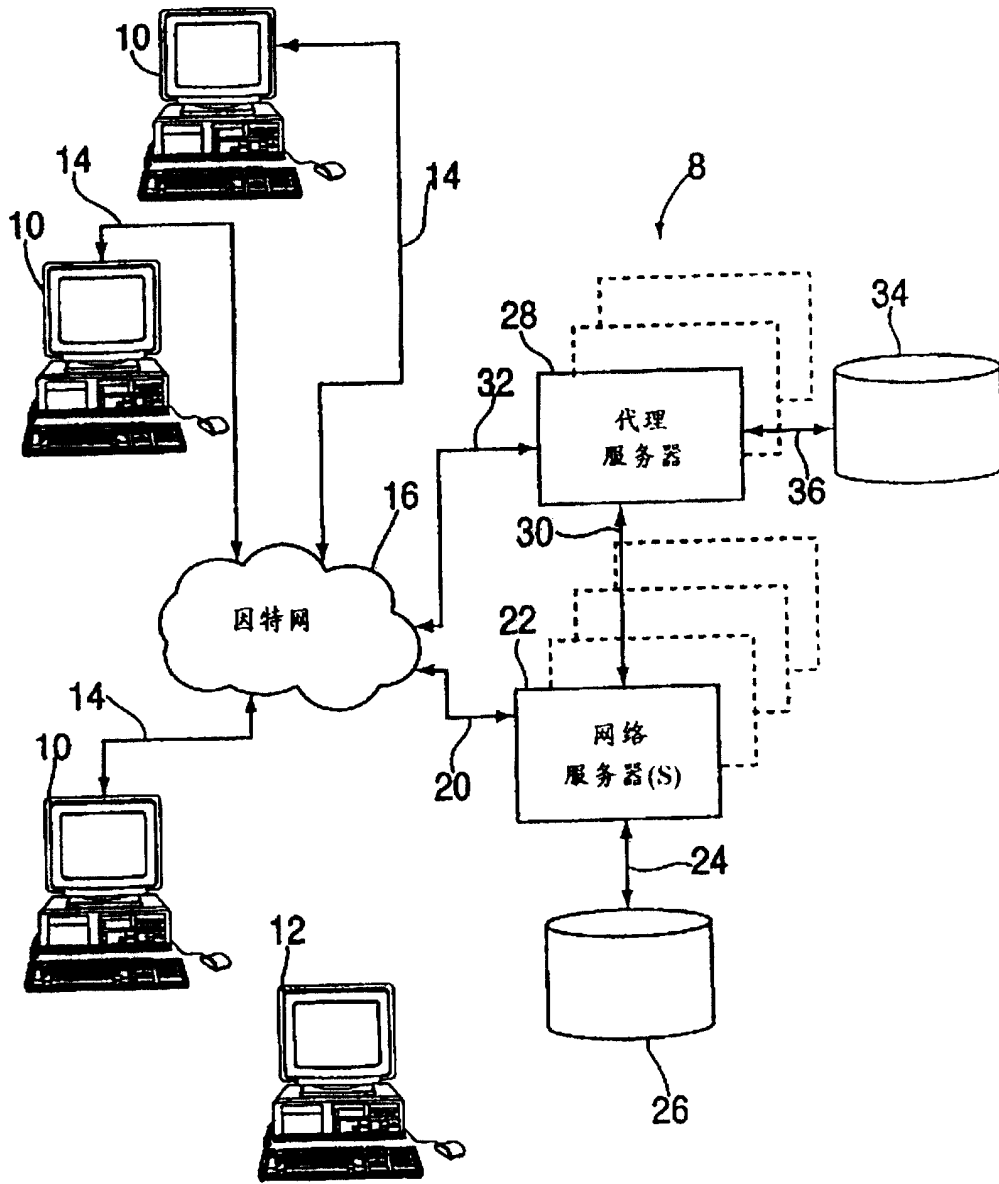


图 1

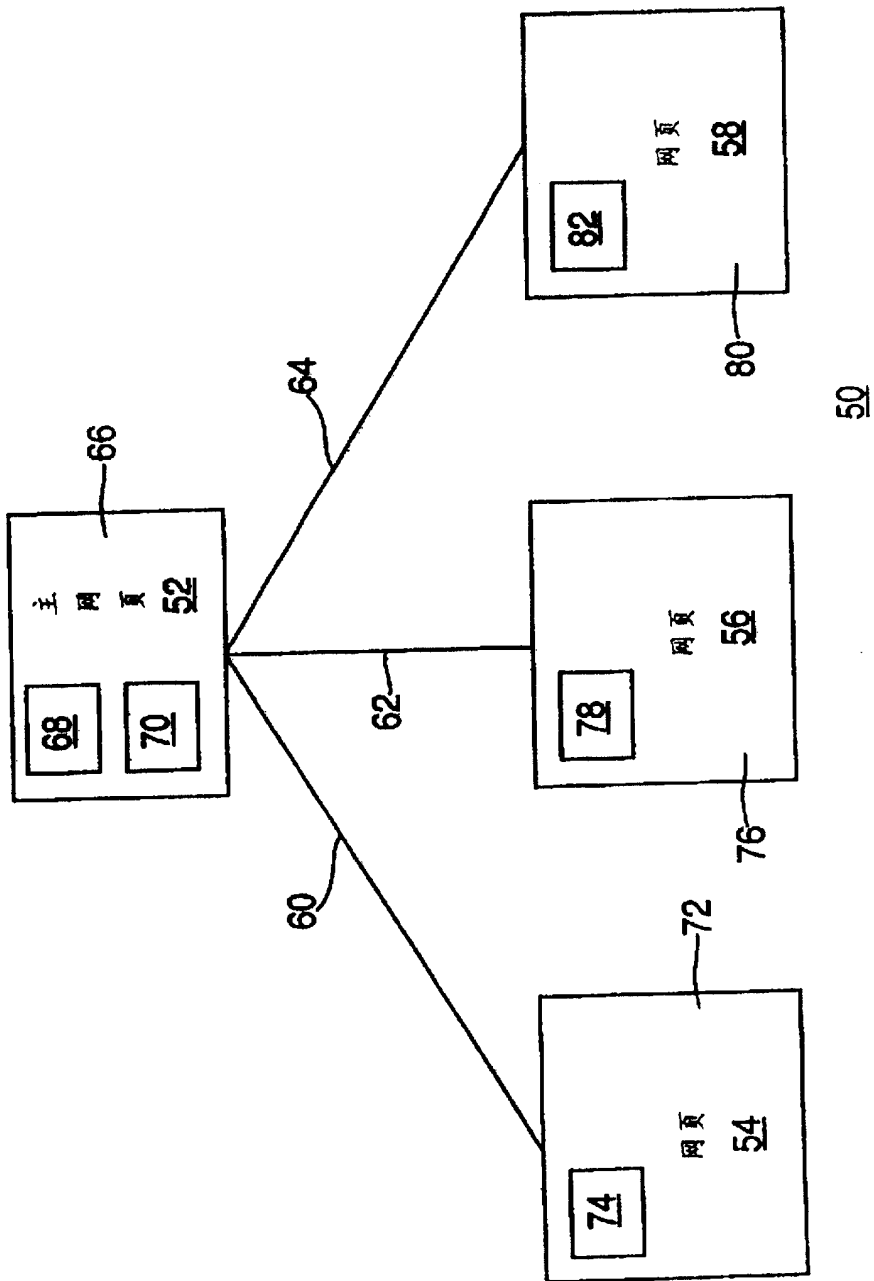


图 2

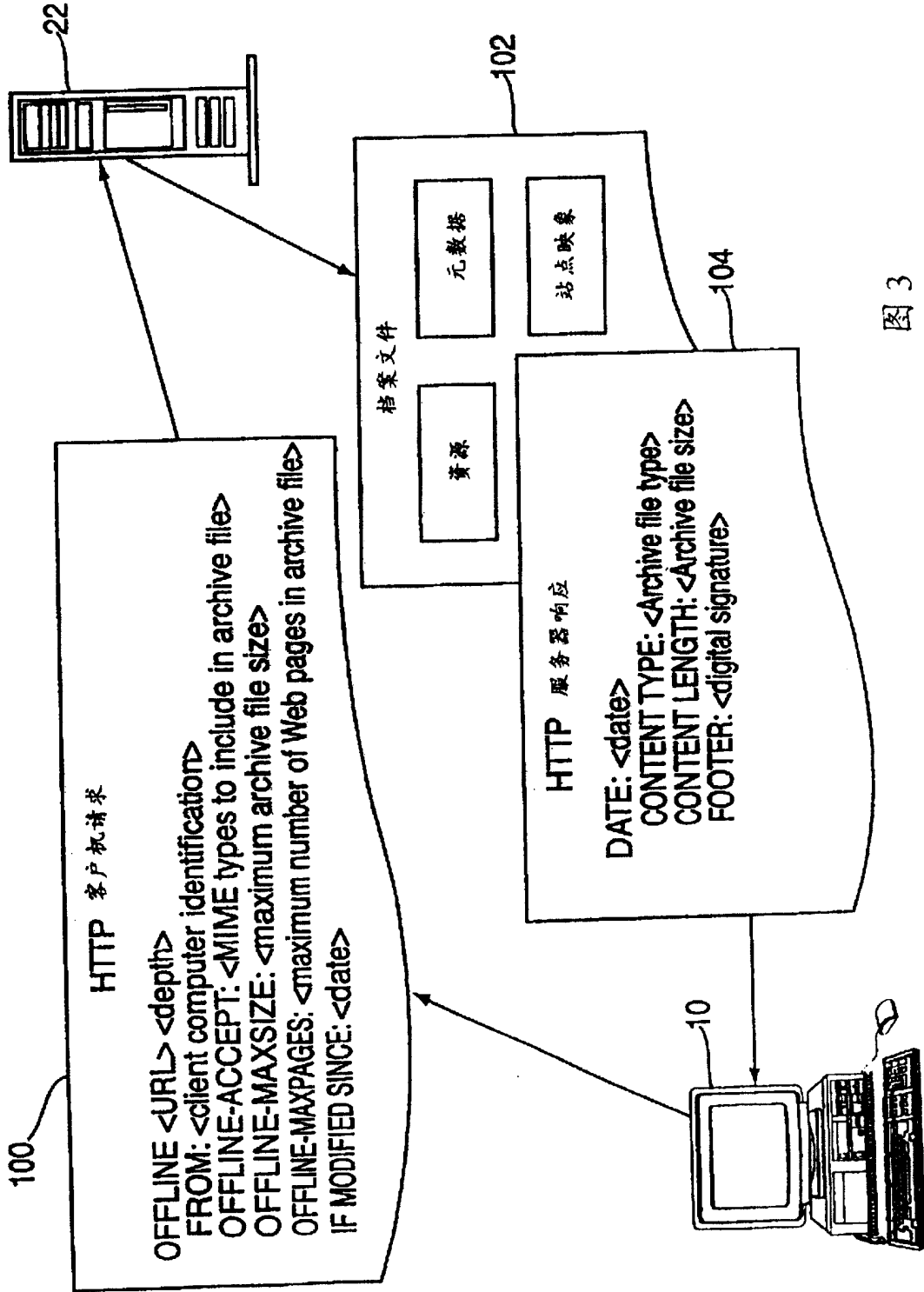


图 3

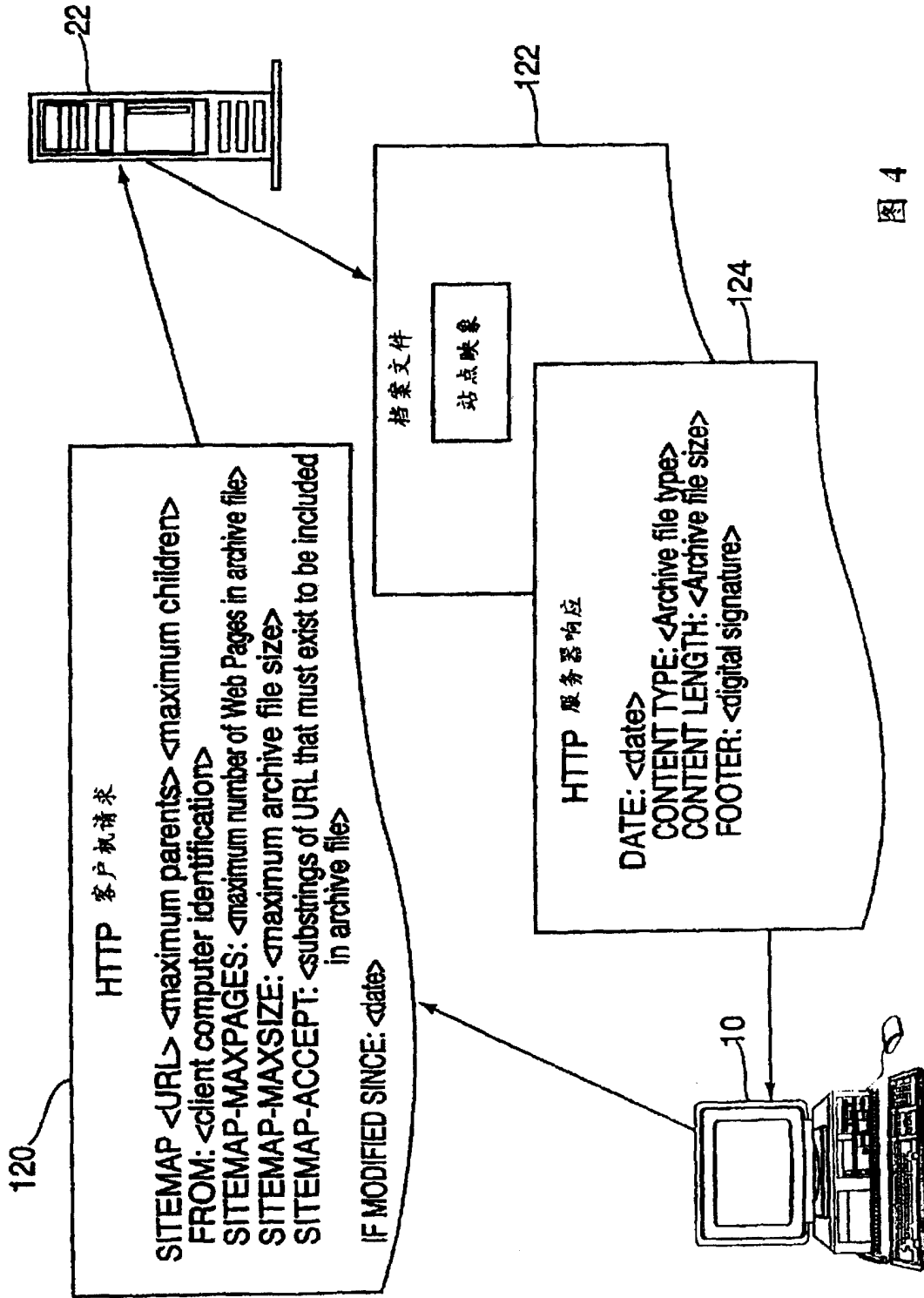


图 4

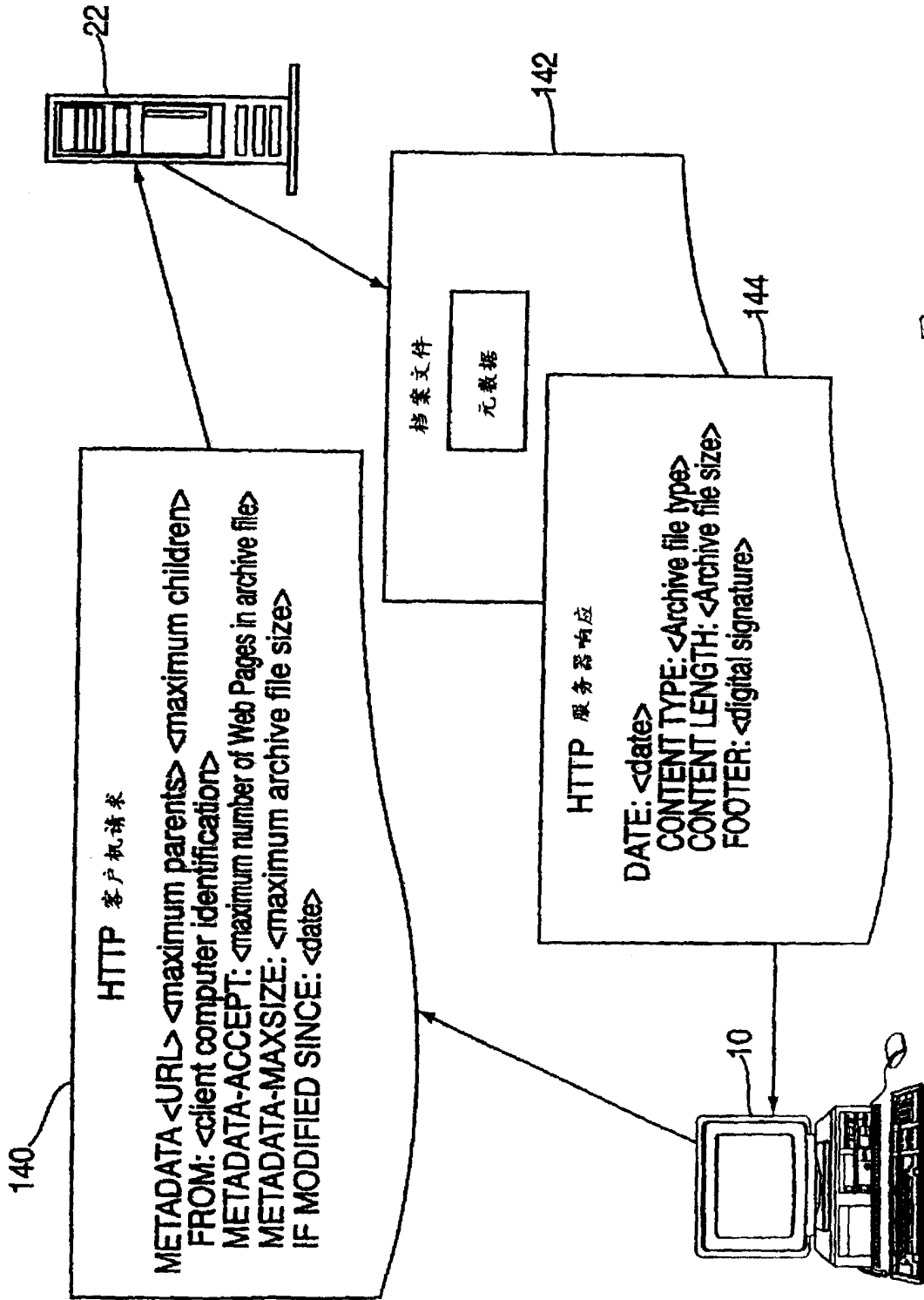


图 5