

[19]中华人民共和国国家知识产权局

[51]Int. Cl⁶

G06F 15/163

[12] 发明专利申请公开说明书

[21] 申请号 98115681.9

[43]公开日 1999年8月18日

[11]公开号 CN 1226030A

[22]申请日 98.7.7 [21]申请号 98115681.9

[30]优先权

[32]97.8.8 [33]US [31]907,606

[71]申请人 国际商业机器公司

地址 美国纽约州

[72]发明人 C·H·布罗布斯特

P·B·蒙地

[74]专利代理机构 中国专利代理(香港)有限公司

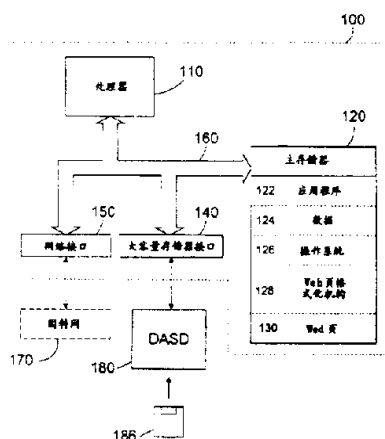
代理人 王 勇 王 岳

权利要求书 4 页 说明书 13 页 附图页数 8 页

[54]发明名称 格式化一个 Web 页的装置和方法

[57]摘要

把一组指定的相关的 Web 页格式化成单一 Web 页的装置和方法,使得用户可以定义若干被选页面和与每个被选页面相关的相关标准。一个格式化装置收集被选页面和根据相关标准的那些相关页面的 URL,并把这些 URL 储存在 URL 容器中。格式化装置进一步调用与储存在 URL 容器中的这些 URL 相关的每一个 Web 页,并产生一个集成页面。这个集成 Web 页面可以包括插入到一个或多个被选页面中的数据在其中引用的数据。然后,这个集成 Web 页可以用一个标准的浏览器打印功能打印出来。



ISSN 1008-4274

权 利 要 求 书

1. 一种装置，其特征在于包括：

至少一个处理器；

一个耦连到该至少一个处理器上的存储器；以及

5 一个位于该存储器中的并由该至少一个处理器执行的格式化装置，该格式化装置从至少一个被选页面和从多个与该至少一个被选页面有关的页面中产生一个集成页面。

2. 权利要求 1 的装置，其特征在于：该格式化装置包括：

一个选择至少一个页面的装置；

10 一个储存该至少一个页面和至少一个相关页面的装置；以及

一个从这些所储存的页面中产生该集成页面的装置。

3. 权利要求 2 的装置，其特征在于用来选择该至少一个页面的装置包含从用户方面判断这个被选的至少一个页面和至少一个用来把该相关页面中的至少一个页面关连到该至少一个被选页面上的相关标准的装置。

15

4. 权利要求 2 的装置，其特征在于用来储存该至少一个被选页面和至少一个相关页面的装置包括一个用来从该至少一个被选页面中判断该至少一个相关页面。

5. 权利要求 2 的装置，其特征在于：产生该集成页面的装置包括：

20 调用该至少一个被选页面和多个相关页面搜索至少一个嵌入代码的一个装置；以及

根据该嵌入代码把被引用的页面的一部分嵌入该集成页面的一部分中的嵌入装置。

6. 权利要求 5 的装置，其特征在于被嵌入的部分包括源代码。

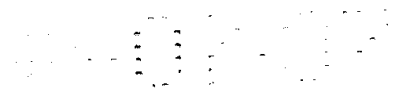
25 7. 权利要求 5 的装置，其特征在于被嵌入的部分包括一个 HTML 部分。

8. 权利要求 1 的装置，其特征在于用统一资源定位器（URL）选择该被选页面。

9. 权利要求 1 的装置，其特征在于该被选页面是一个超文本标记语言（HTML）页面。

30

10. 权利要求 1 的装置，其特征在于如果两个页面中的任何一个与另一个在一个预定的嵌套层内，则这两个页面相关。



11. 权利要求 1 的装置, 其特征在于如果该两个页面中的任何一个具有一个通向另一个的链接, 则这两个页面相关。

12. 权利要求 1 的装置, 其特征在于如果该两个页面存在于该同一个服务器上, 则这两个页面相关。

5 13. 权利要求 1 的装置, 其特征在于, 如果该两个页面具有相同的基地址, 则这两个页面相关。

14. 权利要求 1 的装置, 其特征在于, 还包括一个打印该集成 Web 网页的装置。

15. 一个装置, 它包括:

10 至少一个处理器;
一个耦连到该至少一个处理器上的存储器;
驻留在这个存储器中的多个页面; 以及
一个驻留在该存储器中的并由该至少一个处理器执行的页面格式化装置, 该格式化装置包括:

15 一个从该多个页面中选择至少一个页面的装置;
一个储存该至少一个页面和至少一个相关页面的装置; 以及
一个从所储存的页面中产生该集成页面的装置。

16. 权利要求 15 的装置, 其特征在于该多个页面中至少有一个页面包括至少一个嵌入属性。

20 17. 权利要求 16 的装置, 其特征在于该至少一个嵌入属性包括至少一个属性, 该属性至少部分地判断是否跟随一个通向该多个页面中的另一个页面的链接。

18. 权利要求 15 的装置, 其特征在于该多个页面中至少一个页面包括至少一个属性, 该属性决定该至少一个页面的至少一个部分的结束。

25 19. 一种程序产品, 包括:

- (A) 一个格式化装置, 该格式化装置从一个被选页面和从与这个被选页面相关的多个页面中产生一个集成页面; 以及
- (B) 携带这个格式化装置的信号载体介质。

30 20. 权利要求 19 的程序产品, 其特征在于该信号载体介质包括可记录介质。

21. 权利要求 19 的程序产品, 其特征在于该信号载体介质包括传输介质。

22. 权利要求 19 的程序产品, 其特征在于如果两个页面中的任何一个具有通向另一个的链接, 则这两个页面相关。

23. 权利要求 19 的程序产品, 其特征在于如果两个页面驻留在同一服务器上, 则这两个页面相关。

5 24. 权利要求 19 的程序产品, 其特征在于如果两个页面具有同样的基地址, 则这两个页面相关。

25. 一个程序产品, 包括:

(A) 一个 Web 页格式化装置, 这个 Web 页格式化装置包括:

一个用于选择至少一个页面的装置;

10 一个用于储存该至少一个页面和至少一个相关页面的装置; 以及
一个从所储存的页面中产生该集成页面的装置,

(B) 携带该 Web 页格式化装置的信号载体介质。

26. 权利要求 25 的程序产品, 其特征在于该信号载体介质包括可记录介质。

15 27. 权利要求 25 的程序产品, 其特征在于该信号载体介质包括传输介质。

28. 权利要求 25 的程序产品, 其特征在于该 Web 页格式化装置还包括一个打印该集成页面的装置。

20 29. 权利要求 25 的程序产品, 其特征在于使用统一资源定位器 (URL) 选择该被选页面。

30. 权利要求 25 的程序产品, 其特征在于该被选页面是一个超文本标记语言 (HTML) 页面。

31. 一种把若干相关的 Web 页格式化成集成页面的方法, 其特征在于包括以下步骤:

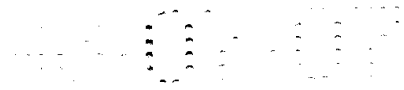
25 (A) 选择至少一个 Web 页并为该至少一个 Web 页选择至少一个相关标准;

(B) 存储该至少一个被选 Web 页并根据该相关标准储存与该至少一个被选 Web 页相关的至少一个 Web 页;

(C) 从所储存的页面中产生该集成页面。

30 32. 一种对现有 Web 页重新格式化的方法, 其特征在于包括下列步骤:

从该现有 Web 页中确定在现有 Web 页中对数据的至少一个引用以包



括在该现有 Web 页中;

通过把至少一个嵌入代码插入该现有 Web 页来修改该现有的 Web 页, 该嵌入代码标识现有 Web 页中数据的位置并标识要被并入该现有 Web 页中的数据。

5 33. 权利要求 32 的方法, 其特征在于在该现有 Web 页中该至少一个引用包括一个 URL。

34. 权利要求 32 的方法, 其特征在于要被并入现有 Web 页的数据包括 MIME 数据。

10 35. 将若干相关的 Web 页格式化为一个单一文件并打印的方法, 其特征在于包括以下步骤:

(A) 选择对应于至少一个被选 Web 页的至少一个 URL;

(B) 为每一个被选的 URL 选择一个相关标准;

(C) 根据该对应的相关标准为每一个被选的 URL 递归地收集所有的相关的 URL;

15 (D) 调用该至少一个被选 Web 页和对应于该相关 URL 的相关的 Web 页;

(E) 把该至少一个被选 Web 页和该相关的 Web 页储存在该单一文件中, 该至少一个被选 Web 页和该相关的 Web 页包括至少一个嵌入属性, 该属性指示要被嵌入的至少一个被引用的 Web 页的至少一部分;

20 (F) 把由该至少一个嵌入代码指定的该至少一个被引用的 Web 页的至少一部分插入在该单一文件的该对应的 Web 页处; 以及

(G) 打印该单一文件。

25 36. 权利要求 35 的方法, 其特征在于该至少一个嵌入属性包括至少一个属性, 它至少部分地决定是否跟随一个通向该多个页面中的另一个页面的链接。

37. 权利要求 35 的方法, 其特征在于该至少一个嵌入属性确定该至少一个页面的至少一部分的结束。

30

说明书

格式化一个 Web 页的装置和方法

5 本发明总的来说涉及诸如因特网这样的计算机网络。更具体地说，
本发明涉及格式化 Web 页的装置和方法。

1948 年 EDVAC 计算机系统的开发通常被称为计算机时代的开始。从那时起，计算机系统已经演变成为非常高级的设备，而且在许多不同的装置中都可以看到计算机系统。计算机的广泛扩展推动了使各个计算机彼此间相互联系的计算机网络的发展。随着个人计算机（PC 机）的
10 引入，用计算机对很多人来说已成为可望又可及的事了。个人计算机网络的开发，使得各个用户之间可以彼此联络。用这种方法，同一个公司内的许多人可以在同一时间使用运行在一个计算机系统上的一个应用软件进行联络。最近颇受欢迎的一个重要的计算机网络就是因特网。该因特网由于计算机和网络的激剧扩展而发展起来，而且已经演变成一个
15 非常复杂的世界范围的计算机系统资源的网络，一般称之为“万维网”，或 WWW 网。一个用户在一台单个 PC 机（即：工作站）上想要访问因特网的典型做法是：使用一种叫做 Web 浏览器的应用软件。Web 浏览器使之通过因特网与其它被称为 Web 服务器的计算机相联系，并从 Web 服务器中接受信息显示在用户工作站上。从 Web 服务器发送到 Web 浏览器的
20 信息一般是使用一种特殊的叫作超文本标记语言（HTML）的语言来格式化的，而且典型的作法是组织成一些叫作 Web 页的页面。许多 Web 页包含一个或多个叫作“链接”的特定的引用地址，链接可调用其它 Web 页。链接可以使用户用一个鼠标或其它指示装置点按适当的链接很容易地进入其它感兴趣的 Web 站点。

25 一个 Web 用户经常希望将当前看到的 Web 页打印出来。Web 浏览器通常具有打印功能，可使用户打印当前页面。但是，随着 Web 站点复杂程度的增加，寻找所需信息变得越来越困难，而且打印若干有关的 Web 页的过程成为一项非常枯燥乏味的工作，它包括：调用 Web 页、打印该 Web 页，再调用下一个 Web 页、打印、调用、打印等等。换句话说，先
30 有技术的浏览器要求用户在打印前先调用页面。用这种先有技术的浏览器，如果用户需要打印 40 页有关的 Web 页，该用户就必须人工调用和打印这 40 个 Web 页的每一页。勿需多言，这一过程是非常费时的。

随着因特网用户、供应商、以及 Web 服务器的数量持续地快速增长，对于用户来说，不用人工调用和打印每一页就能打印出有关的 Web 页将变得越来越重要。如果不改进目前这种 Web 页的打印方式，那末 Web 页的打印将继续成为有效利用因特网上资源的障碍。

5 根据本发明，揭示了一种把一组指定的有关的 Web 页格式化成一单个 Web 页的装置和方法。用户定义被选页面的数量及每一被选页面有关的相关标准。一个格式化机构负责收集被选页面和那些根据相关标准而相关的页面的 URL，并把这些 URL 储存在 URL 容器中。该格式化机构进而将与包含在 URL 容器中的 URL 相关的每一个 Web 页面调出并生成一个
10 集成页。该集成页可以包含在一个或多个被选页面中所插入或引用的数据。然后，这个集成页可用标准浏览器的打印功能打印。

本发明的上文所述目的及其它目的、特性和优点将从下面对本发明的优选实施例的更详细的描述及结合附图说明而变得更加清楚明了。

下面将要结合附图描述本发明的优选实施例，图中相同的标号表示
15 相同的单元：

图 1 是根据本发明的计算机系统的方框图；

图 2 是典型的因特网连接方框图；

图 3 是根据优选实施例的如何格式化所选择的和所相关的 Web 页的方法步骤的流程图；

20 图 4 是根据优选实施例的较佳实施例用作相关标准的嵌套结构的方框图；

图 5 是根据优选实施例格式化所选择和所相关的 Web 页的计算机系统方框图；

25 图 6 是根据优选实施例收集和格式化被选和相关 Web 页的方法步骤的流程图；

图 7 是根据优选实施例收集被选和相关 Web 页的递归方法的伪码表示；

图 8 是根据优选实施例的 URL 容器的处理方法的伪码表示；以及

图 9 是根据优选实施例的平展处理的伪码表示。

30 概述

本发明的方法和装置对于格式化因特网上的 Web 页具有特别的实用性。对于那些不熟悉因特网的人们，在此对有关因特网的概念作一个

概述。

图 2 表示一个典型的因特网连接的例子。一个想要从因特网 170 上获取信息的用户通常拥有一个计算机工作站 200，该工作站执行一个称作 Web 浏览器 210 的应用程序。在 Web 浏览器 210 的控制下，工作站 200 通过因特网 170 发送一个对 Web 页的请求。Web 页数据可以是文本、图形，也可以是其它信息形式，统称为 MIME 数据。因特网上的每一个 Web 服务器有一个已知的地址，叫作“统一资源定位器”（URL），Web 浏览器用它来连接适当的 Web 服务器。由于 Web 服务器 220 可以包含一个以上的 Web 页，所以用户还需指定他想要在 Web 网服务器 220 上看的特定的 Web 页的地址。Web 服务器计算机系统 220 执行 Web 服务器应用程序 222，监视请求，并对其负责的请求服务。当一个请求指定了 Web 服务器 220 时，Web 服务器应用程序 222 一般是访问对应于该特定请求的一个 Web 页，并将其发送到用户的工作站 200 上。

Web 页

一个 Web 页可以包含各种形式的 MIME 数据。大多数 Web 页包含准备在用户工作站 200 的监视器上显示的可视数据。Web 页一般用“超文本标记语言”（HTML）编写。当 Web 服务器 220 收到一个 Web 页请求时，它将把这个所请求的页面以 HTML 形式通过因特网 170 送到发出请求的 Web 浏览器 210。Web 浏览器 210 理解 HTML 语言，将它翻译成 Web 页并送往用户工作站 200 的监视器。在用户屏幕上显示的 Web 页可能含有任何适当的 MIME 数据，包括文本、图形，以及链接（它引用其它 Web 页的地址）。这些其它的 Web 页（即：由链接代表的那些页面）可以在同一 Web 服务器上，也可能在不同的 Web 服务器上。用户可以使用鼠标器或其它指示设备敲击这些链接来调用这些其它的 Web 页。这个具有连接到其它服务器上的其它 Web 页链接遍布世界的整个 Web 页系统叫作“万维网”（WWW）。

说明书的剩余部分将要描述本发明如何提高格式化和打印相关 Web 页的方便程度，它是通过提供各种方式使用户不用进行通常所需的调用和打印每一 Web 页的反复操作就能格式化和打印有关的 Web 页而实现的。本领域的普通技术人员将认识到本发明可以同样用于任何有关数据的格式化和/或打印，无论这种数据是 Web 页的形式、数据库记录的形式还是可能相互有关的其它数据形式。

详细说明

参看图 1, 根据本发明的一个计算机系统 100 包括一个处理器 110、一个主存储器 120、一个大容量存储器接口 140 和一个网络接口 150。它们都由一个系统总线 160 连接起来。本领域普通技术人员应理解该系统包括了所有类型的计算机系统: 个人计算机、中等规模计算机、大型计算机等等。注意, 在本发明的范围内, 可以对这个计算机系统 100 做许多增加、修改和删减。可能增加的设备举例如包括: 计算机监视器、键盘、高速缓冲存储器, 以及打印机那样的外设。

处理器 110 可以由一个或多个微处理器和/或一些集成电路构成。处理器 110 执行储存在主存储器 120 中的程序指令。主存储器 120 储存程序和计算机可以访问的数据。当计算机系统 100 起动机, 处理器 110 开始执行构成操作系统 126 的程序指令。操作系统 126 是一个管理计算机系统 100 的资源的复杂程序。这些资源中的一部分有处理器 110、主存储器 120、大容量存储器接口 140、网络接口 150 以及系统总线 160。

主存储器 120 包括一个或多个应用程序 122、数据 124、操作系统 126、Web 页格式化机构 128 以及一个或多个 Web 页 130。应用程序 122 在操作系统 126 的控制下被处理器 110 执行。应用程序 122 可以用程序数据 124 作为输入来运行, 应用程序 122 还可以输出其结果作为主存中的程序数据 124。本发明中, 计算机系统 100 包括一个 Web 页格式化机构 128, 它能够把多个相关的 Web 页格式化成一单个页面, 然后可以打印出来、下载到磁盘上、放到因特网上或用于一个本领域普通技术人员所知的任何其它用途。

大容量存储器接口 140 可使计算机系统 100 从辅助存储设备如磁盘 (硬盘、软盘) 和光盘 (CD-ROM) 中检索和储存数据。这些大容量存储设备一般称为“直接存取存储装置” (DASD), 其作用是永久储存信息。一个适当的 DASD 类型是一个从软盘 186 读取和向其写入数据的软盘驱动器 180。从 DASD 来的信息可以是多种形式。一般的形式是应用程序和程序数据。通过大容量存储器接口 140 检索的数据通常放在处理器 110 可以处理它的主存储器 120 中。

尽管主存储器 120 和 DASD 装置 180 在通常情况下是分离的存储设备, 但计算机 100 使用众所周知的虚拟寻址机制, 它使计算机系统 100 的程序运行起来好象不是访问多个、较小的存储实体 (例如: 主存 120



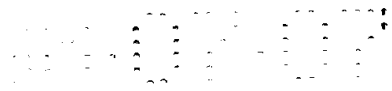
和 DASD 装置 185)，而是只访问一个大的、单一的存储实体。因此，尽管某些单元被表示为驻留在主存储器 120 中，将认识到这些单元未必在同一时间都完全包含在主存储器 120 中。应当注意的是在此使用的“存储器”一词一般是指计算机系统 100 的整个虚拟存储器。此外，根据本发明的装置包括含有本发明的各单元的任何可能的硬件和软件配置，不论这种装置是单个计算机系统还是由共同运行的多个计算机系统构成的系统。

网络接口 150 使计算机系统 100 可以对该计算机系统可能连接的任何网络发送和接收数据。这个网络可以是局域网 (LAN)，也可能是广域网 (WAN)，或者更特殊的说是因特网 170。连接到因特网的适当方法包括公知的模拟和/或数字技术，以及未来将要开发出来的连网络机制。许多不同的网络协议都可以用来实现一个网络。这些协议是一些能使计算机通过网络进行通信的特定的计算机程序。过去常用于通过因特网通信的 TCP/IP (传输控制协议/互连网协议) 就是一个适当的网络协议的例子。

系统总线 160 使得数据能在计算机系统 100 的各个组件之间传递。虽然图中所示的计算机系统 100 只包含一个主处理器和一条系统总线，但是本领域普通技术人员将认识到本发明可以用一个具有多个处理器和/或多条总线的计算机系统来实现。此外，在这个优选实施例中用到的接口还可以包括用来从处理器 110 卸载计算密集处理的分离的、全编程的微处理器，或者也可以包括完成同样功能的 I/O (输入/输出) 适配器。

在这一方面，注意到以下情况很重要：虽然本发明至今是以 (而且还将继续以) 一个全功能的计算机系统描述的，但本领域普通技术人员将认识到本发明也能作为程序产品以各种形式分布，而且不论实际上用来进行这种分布的信号载体的特定类型，本发明均可同样使用。合适的信号载体的例子包括：可记录型载体，如软盘 (如：图 1 中的 186) 和 CD ROM，以及传输类载体，如数字的和模拟的通信链路。

本说明书剩下的部分将描述 Web 页格式化机构 128 的优选实施例，该格式化机构能取得若干被选的 Web 页，收集其 URL，并建立一个可以称为“集成”或“展平”的 Web 页的单一文件。此处使用的“展平的 Web 页”一词是形象地表示在通常以交叉链接的树状结构中存在的几个



相关的页面全都被组合起来或“展平”成一个单一页面，因而取消交叉链接并将各页面按顺序排列。

5 现在参看图 3，为了格式化一个 Web 页，方法 300 从用户定义一个 URL 表并为每个被选的 URL 定义一个相关标准（步骤 310）开始。做到这一点可以有多种方式：由机械的方法由人类用户机械地输入到菜单屏幕，从 Web 浏览器的历史存储器中检索一个 URL 表，或任何其它的本领域普通技术人员公知的方法来确定 URL 数据。一旦 URL 表和每个 URL 的相关标准被定义后，用所有这些被选的和相关的 URL 建立 URL 容器（步骤 320）。最后，从此展平容器内的被选的和相关的 URL 中产生一个展平页面（步骤 330）。

15 在格式化过程中，相关标准是一个重要的因素，因为它定义了那些被认为“相关的”URL 间必须存在的必要的联系，从而确定哪些页面将包含在这一被展平的页面内。在本发明的范围内，关于两个 URL 是否“相关”的标准可能会改变。此处一个特定的相关标准是指“嵌套层”，将参照图 4 对此给予说明。

彼此间具有链接的页面可以会如图 4 所示那样排列成树状结构 400。嵌套结构 400 至少有一个所选择的 Web 页 411（如：第一所选择的 Web 页 411 和/或第二所选择的 Web 页 450），这些页面具有若干链接 421 至 439（即：链接 1，链接 2，链接 3）与其它页面 441 至 448 20 连结。每一个链接由一个调用一个 Web 页的机构构成，如 URL，该机构可以被用户激活。当一个用户定义了嵌套层时，他们就确定了该格式化机构寻找相关的 URL 所达到的嵌套树 400 的深度。例如，如果一个用户选择了第一被选的 Web 页 411 作为其被选的页面，而且确定了两级嵌套收集相关的 URL 的相关标准，这些相关的 URL 包括第一被选的 Web 页 411 的 URL 以及所有那些直接与该第一被选的 Web 页链接的 Web 页中所包含的所有链接的 URL，即：第一被选 Web 页 411、链接 1 Web 页 441、链接 2 Web 页 442、链接 3 Web 页 443、链接 A web 页 444、链接 B web 页 445、链接 D web 页 446、链接 F Web 页 447 以及链接 G Web 页 448 的 URL。如果嵌套等级设为三级，那么相关的 URL 除了包括上述两级嵌套的 URL 外，还另外包括链接 I 429、链接 II 430、链接 III 431、链接 IV 432、链接 V 433、链接 VI 434、链接 VII 435、链接 VIII 436、链接 IX 437、链接 X 438 以及链接 XI 439 的 URL。

其它相关 URL 的适当的相关标准包括: URL 是否在同一 Web 服务器上; 一个指定的搜索字是否出现在 Web URL 搜寻表中; URL 间是否有一个链接; 或者 URL 是否有相同的基地址。

下面是基地址相关标准的一个例子。一个主页的地址可以是:
5 WWW.corporation X.com/home.html, 那么任何有 WWW.corporation X.com 基地址的 URL 都与该主页有关。再举另外一个例子, 选择一个地址是 WWW.corporation X.com/support/index.html 的 URL, 那么共有基地址 WWW.corporation X.com/support 的任何 URL 都与这个被选 URL 有关, 而该站点上的其它 URL 与之无关。如下面将要更详细讨论的那
10 样, 不论使用什么样的特定相关标准, 相关的 URL 都被格式化成一个单一的 Web 页。

通过下述优选实施例的例子, 可以更好地理解上面描述的步骤。现在参看 5-6 图, 用于因特网的根据本优选实施例的装置 500 包括一个 Web 客户机 200 和一个通过因特网 170 与之相连的 Web 服务器 220。Web
15 客户机 200 包括一个 Web 浏览器应用程序 210 和一个 Web 页格式化机构 128。Web 浏览器应用程序 210 是一个业内人士公知的标准 Web 浏览器。Web 页格式化机构 128 包括一个 Web 页选择机构 540、Web 页储存机构 550, 以及 Web 页集成机构 560。虽然图 5 中所示的 Web 页格式化机构 128 是与 Web 浏览器 210 分开的, 但本发明的最佳模式是希望把 Web 页
20 格式化机构 128 并入 Web 浏览器应用程序, 从而提供一个对有关的 Web 页具有诸如打印、下载、文件传送等功能的高度格式化能力的浏览器。或者是, Web 页格式化机构 128 既可以是一个在 Web 客户机 200 上运行的单独的应用程序, 也可以是一个 Web 浏览器应用程序 210 的插入式的或 Java 小程序 (Java applet) / 应用程序。这里在描述 Web 页格式化
25 机构 128 的功能时不考虑机构 128 是在 Web 浏览器应用程序 210 内还是在 Web 浏览器应用程序 210 之外。

Web 页选择机构 540 的作用是建立一个用户选中的和相关的 Web 页表。Web 页储存机构 550 将这一被选的和相关的 Web 页表以 URL 形式储存。集成机构 560 取出被选的 URL 并将其格式化一个展平的 Web 页。因为
30 装置 500 把许多排成行列的 Web 页展平成单一的集成 Web 页, 所以任何浏览器提供的标准打印功能都能打印这个集成的 Web 页。联系图 6 的流程图可以最好地理解机构 540-560 的功能。

用来格式化多个相关的 Web 页或 URL 的方法 600 从选择那些将明显包含在所展平的 Web 页中的 Web 页或 URL 开始 (步骤 610)。一旦 Web 页被选中, 则每一个被选的 Web 页的挖掘层就确定了 (步骤 620)。挖掘层是一个适当的相关标准的特定例子, 它相当于参照图 4 讨论的嵌套层。当所有这些 Web 页和它们的挖掘层被确定后, 一个这些 Web 页和相应的挖掘层的 URL 表就建立了 (步骤 630)。然后这个表处于一种可用于填充 URL 容器的格式。从步骤 640 开始处理这个表。如果 URL 表不是空的 (步骤 640 = NO), 则取出表中下一个 URL (步骤 650)。这个被选的 URL 然后被加入 URL 容器中准备处理 (步骤 660)。然后, 所有与该被选的 URL 相关的 URL 被递归收集, 收集方法是沿着相关的 URL 向下通过嵌套结构 400 直到到达所需挖掘层, 并把这些 URL 加进 URL 容器中待处理 (步骤 670)。步骤 650、660 和 670 连续进行直到 URL 表中所有的 URL 都被处理 (步骤 640 = YES)。一旦 URL 表中所有的 URL 都被处理了 (步骤 640 = YES), 则开始处理 URL 容器中的 URL, 处理方法是调用每一个 URL 并收集这些 Web 页 (步骤 680)。Web 客户机和 Web 服务器之间调用一个 Web 页的机制和相互作用是本领域中众所周知的。一旦这些 Web 页被调用, 就由这些被调用的 Web 页产生一个展平的 Web 页 (步骤 690), 最好是把这些相关的 Web 页一起添加进一单个 Web 页中。这个所展平的 (即: 集成的) Web 页处于这样一种形式: 它可以下载到磁盘, 可以向其他因特网用户传送, 可以在本地发送到其他网上用户, 可以被打印或者被用作本领域所知的其它用途。

为了更好地理解图 6 中一些步骤的细节, 我们现在参看图 7-9。图 7 的方法 700 表示一个用来把 URL 放入 URL 容器的适当的递归收集过程的一个特定例子 (步骤 670)。图 8 的方法 800 表示一个处理 URL 容器中的 URL 的一个特定例子 (步骤 680)。图 9 中的方法 900 表示产生一个展平的 Web 页的一个特定例子 (步骤 690)。

现在参看图 7, 方法 700 用递归伪码表示, 并且把被选的 URL 和由适当相关标准 (如: 挖掘或嵌套层) 相关的 URL 递归地储存在 URL 容器中。当收集 URL 方法 700 被调用时, URL 表、URL 容器以及挖掘层必需被确定。这个 URL 表就是用户选择的那个 URL 表 (如: 图 6 的步骤 630)。URL 容器参数指定了所用的 URL 容器的名字。而且挖掘参数为了联系这些 URL 指定了相关 URL 的挖掘或嵌套层。对于 URL 表中的每一个 URL (步

骤 710), 它要被加入 URL 容器 (步骤 712)。一旦 URL 被加入 URL 容器, 该 URL 就被用来调用对应的 Web 页 (步骤 714), 从这个页面中可收集相关的 URL (步骤 716)。如果用户确定的被选的 URL 的挖掘层大于零 (步骤 720), 那么那于每一个要收集的 URL, 必需按照相关标准 (即: 挖掘层) 进行递归访问 (即: 挖掘) (步骤 724) 进入嵌套树结构来收集其它相关的 URL 并将它们放进 URL 容器中。这一过程连续进行, 直到所有的被选的 URL 和所有相关的 URL 都被收集起来并放入 URL 容器中为止。例如: 如果被选的 URL 有连到其它页面的三个链接, 而且用户确定的挖掘层是三: 首先, 将这个被选的 URL 作为第一挖掘层被加入 URL 容器中; 其次, 调用被选的 URL 来产生这三个所链接的 Web 页的 URL, 并把这些 URL 收集进 URL 容器作为第二挖掘层; 第三, 分别调用这三个所链接的页面来产生它们所链接的 Web 页的 URL 并规定其为第三挖掘层, 再收集这些 URL 并放入 URL 容器以完成 URL 表中的 URL 的处理。

15 现在参看图 8, 方法 800 以伪码表示, 它为展平成一个集成页面准备 URL。对 URL 容器中的每一个 URL (步骤 810), 对应于每个 URL 的 Web 页被调用 (步骤 820)。所有在被调用的 Web 页内引用的 URL 被收集起来待处理 (步骤 830)。下一步, 步骤 830 中收集的每一个 URL 被处理 (步骤 840)。如果这个 URL 引用在 URL 容器内的一个 URL (步骤 842), 则使用 NAME 属性将该 URL 的第一次出现标记为目标 (步骤 846), 而该 URL 所有后续的出现都交叉引用回该目标引用, 这是通过修改 HREF 标记使其包含文本 “See Section__” 而进行的, 这里 “__” 是在目标 HREF 中所规定的文本 (步骤 848)。例如, 对于一个形式为 Chapter X 的 “锚” (anchor) 来说 (这里的 chapterX 包含在 URL 容器中), 插入到该文本的说明将可能是 “See Chapter X”。如果这个 URL 引用一个不在 URL 容器内的 URL, 则相关的 HREF 属性被修改成包括文本 “Section Not Included” (步骤 850)。这一过程持续进行, 直到所有 URL 容器中的 URL 都被标识为目标、交叉引用或标识为 “Not Included” 为止。

30 现在参看图 9, 方法 900 以伪码表示, 它在方法 800 (图 8) 结束时处理被修改的 URL 容器, 以便生成展平的或集成的 Web 页。对于 URL 容器中的每一个 URL (步骤 910), 每一个 URL 被调用来产生相关的 Web

5 页（步骤 920）。下一步，收集该页面中引用的 URL（步骤 930）。对于每一个收集到的 URL（步骤 940），如果 HREF 语句包括 EMBED 属性（步骤 942），则包括 EMBED 属性的 HREF 语句中指定的页面中的那些部分被插入该页面（步骤 944）。然后最好是通过把该新的页面添加到该展平的页面文件的末尾而将这个页面加到展平的页面文件中（步骤 950）。由本发明所定义的 EMBED 属性及其它新属性将在下面进行更详细的讨论。

10 从最基本上说，一个 Web 页是由被一些叫作“标记”的特别的定界符分开的各种信息字段组成的。标记告诉 Web 浏览器用这个特殊字段的信息做什么。例如，标记可以让 Web 浏览器显示一个图像、显示一个文本、播放一个声音消息、或者显示一个叫作超文本链接的特殊字段的文件。超文本链接是一个标识远程资源的引用机制，这个远程资源存放在该系统虚拟存储空间的任何地方，不论它存放在同一台计算机上、辅助存储装置上，还是网络上的远程计算机上。换句话说，链接标识的是地址（即：URL），这个地址是当用户选择该链接时（典型的作法是用鼠标器或其它指示装置敲击该链接）计算机将要调用的地址。在 HTML 一个超文本链接用“锚”（anchor）标记定义。定义锚开始的标记是<A>，定义锚终止的标记是。锚可以包括诸如 HREF 和 NAME 这样的属性。HREF 属性确定链接的超文本引用（即：URL）。NAME 属性把一个记号放在一个页面中，该记号可以被一个链接用于指定该页面中某个特殊的位置或部分。指定一个名字就是告诉浏览器从哪里开始显示数据。例如，NAME = “X” 用文本 X 来命名这个 HTML 页面的一部分来标记一个页面。本发明规定了一个定义被命名的字段或部分的结束的附加属性。这个结束属性是类似的，但写成 NAME = “X.end”，从而结束一个部分。重要的是要注意：浏览器将简单地忽略任何它不认识的标记或属性。这一特点使得 Web 页设计者可以加一些特殊浏览器能够识别和处理的特殊标记或属性（如：NAME = “X.end”），同时又保证该同一页面在现有的浏览器上显示没有问题。

30 格式化 HTML 各部分的另外一种方法是：假定该部分从一个 NAME 属性延伸到一个 NAME 属性，该属性在页面的某一点上定义一个新的部分。例如：一个目录可以包括以下 URL：

myUrl#a

myUrl#b
myUrl#c
myUrl#d

这个 myUrl 页面可以包括以下各部分:

5 (HTML 数据)
 (HTML 数据)
 (HTML 数据)
 (HTML 数据)
 (HTML 数据)
10 (HTML 数据)
 (HTML 数据)
 (HTML 数据)
 (HTML 数据)
 (HTML 数据)

15 假定整个文件将按目录中的各个表目打印, 如果部分 b 被引用了, 我们可以假定我们需要从 标记到 标记的全部 HTML。注意, 该部分的终止可以由所遇到的紧挨着的标记或如上面的例子那样由一个具有不同标签的标记来定义, 使得部分 b0 和 bn 被包括在对部分 b 的引用中, 而部分 c 则是作为不同的部分被标识的。用这种方式, 如果能够有把握地对 HTML 各部分的结束进行某种假定, 那么上面定义的 X.end 属性就不需要了。

25 本发明定义的另一个属性是 EMBED 属性。在本优选实施例中, 另一个属性 EMBEDSRC 是一个具有特殊目的的 EMBED 属性的特例。为了说明本发明的这一概念, EMBED 属性被用来将普通文本嵌入一个 HTML 页面中, 此页面中由 NAME = "X" 属性和 NAME = "X.end" 属性或 NAME = "X" 属性和下一个 NAME 属性隔开, 这些属性决定下一部分的开始 (因而也决定前一部分的结束)。EMBEDSRC 属性用来嵌入代表源码表一部分的文本。把源码用不同的可视特性嵌入, 使得程序员可以用一种特殊形式 (如: 颜色、字型、字型大小等等) 在计算机程序的文件页面中显示源码。这些属性被放在一个 HREF 语句中。例如, 这个嵌入源码的 HREF
30 语句可以写成: See method, 而嵌入 HTML 部分的 HREF 写成 <A HREF=

“mysource.html#section” EMBED>See method。这些新的属性标识在格式化 Web 页面期间可以插入所链接的信息的嵌入位置，这些在方法 900 和图 9 中已描述了。

5 本发明定义的另一个新的锚属性是 FOLLOW、SHOULD FOLLOW 和 NO FOLLOW。这些属性又放在 HREF 语句中分别指出一个链接是否必须被跟随、可能被跟随或必须不被跟随。下面是使用 FOLLOW 属性的一个可能的例子：See method。如果这个被引用的 URL 包含一种即使嵌套层规定不包括而应包含在这个集成 Web 页中的信息（如：关键信息）的话，那么，一个具有 FOLLOW 属性的 URL 就会被包括进去。NO FOLLOW 属性作用相反，即使嵌套层可能已经包括它，也不会包括这个被引用的 URL。SHOULD FOLLOW 属性也被规定为：如果相关标准指示跟随用 SHOULD FOLLOW 属性标记的链接，则把所引用的材料包括进集成 Web 页中。SHOULD FOLLOW 和 NO FOLLOW 属性以与 FOUOW 属性同样的方式在 HREF
10 语句中使用。应当注意：另外的标记或属性可能被建立去进行其它操作，如象压缩或扩展数据、把数据重新格式化另一种源形式或本领域普通技术人员公知的其它操作。

20 这些属性如何被处理取决于用户确定的相关标准。例如，用户可以确定严格按照嵌套层，而忽略所遇到的任何 FLOOLW、SHOULD FOLLOW 或 NO FOLLOW 属性。另一种选择是：用户可以确定一种相关标准，它包括所有具有 FOLLOW 属性的 URL，而不包括具有 NO FOLLOW 属性和 SHOULD FOLLOW 属性的 URL。还有一种选择是：将具有 FOLLOW 或 SHOULD FOLLOW 属性的 URL 包括在集成 Web 页面中，而特别不包括具有 NO FOLLOW 属性的 URL。由本发明规定的这些属性使用户可以确定一个可能更容易改变以满足某种需要的更复杂的相关标准。注意：在这里确定的所有的新属性统称为“嵌入属性”。

30 对于图 9 的方法 900 来说，每一个相关的 Web 页都被扫描以判断上面讨论过的那些新属性即：EMBED、EMBEDSRC、NAME = “X. end”、FOLLOW、SHOULD FOLLOW、NO FOLLOW 等等位于什么地方（步骤 940 和 942）。定位所有那些用来指定已命名部分的终止的 NAME 标记以标识数据部分。一旦数据部分被确定，任何一个 EMBED 和 EMBEDSRC 标记都被定位，而且对应的相关的部分就被插入到相应的相关的 Web 页中的对应的 EMBED

或 EMBEDSRC 的位置处（步骤 944）。这一过程持续到 URL 容器中所有那些具有 EMBED 或 EMBEDSRC 属性的 URL 都被插入到集成页面为止。所有被选页面都被收集起来并放进新的集成页面来完成 Web 页面的格式化过程。

5 应当注意的是：Web 页面格式化机构 128 的这个优选实施例使用 HTML 标记来标识各部分和插入点，但是，本领域的普通技术人员会认识到使用其它的语言和系统如 Java 和 Java Script 也能进行同样的操作。

10 在本发明的这个优选实施例中，上面讨论过的 HTML 属性已经包含在被选的和相关的 Web 页面的代码中。但是，根据本发明的一种方法使用户可以手动地或动态地把上面讨论的属性插入现有的 Web 页面以使所有现有的 Web 页在本发明的格式化过程中能被使用。这种方法最好是一个交互过程，在这里，现有页面的 HTML 被扫描来找出 HREF 语句，而且给用户一个机会适当地插入上面新近定义的属性中的任何一个。用于
15 进行现有 HTML 的这种转换的一个工具使用户可以迅速地把页面转换成根据本发明的格式，它使格式化机构 128 在处理这些页面时就好像这些页面原来就是用本文定义的那些属性开发的。

20 虽然参照其优选实施例已经对本发明进行专门的展示和描述，但本领域普通技术人员将会理解到在不脱离本发明的精神和范围的情况下，在形式和细节上可以做各种各样的变化。例如：虽然本文的优选实施例是按照 HTML 页面讨论的，但其它的页面格式和数据格式也同样包括在本发明中。本文使用的术语页面意在包括任何数量的可以被处理或显示的数据。此外，虽然为了示范目的，相对于通过因特网通信的 Web 客户机和 Web 服务器来展示本发明，但本发明可以用于任何适当的网络
25 上的任何类型的客户机服务器的场合。而且，使用 URL 把若干页面收集起来并格式化成一个单一 Web 页面仅仅是本发明范围内收集所需数据的一种方法。还有，本文所称作的用户可以执行某种功能。本说明书和权利要求书中使用的术语“用户”一词显然包括可以执行用户功能的任何媒体，它包括：无限的人类用户、计算机功能，以及任何形式的软件
30 程序。

说明书附图

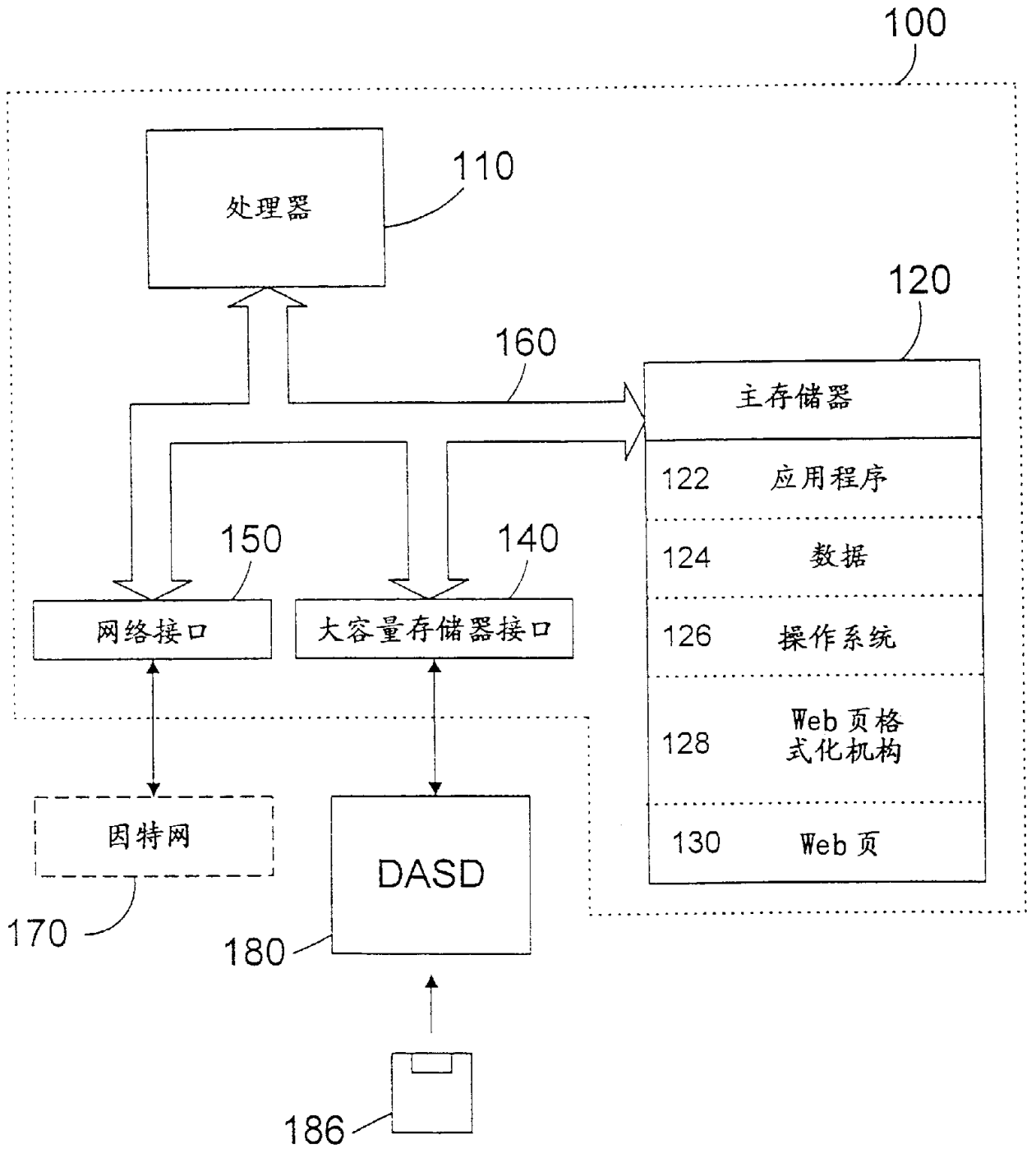


图 1

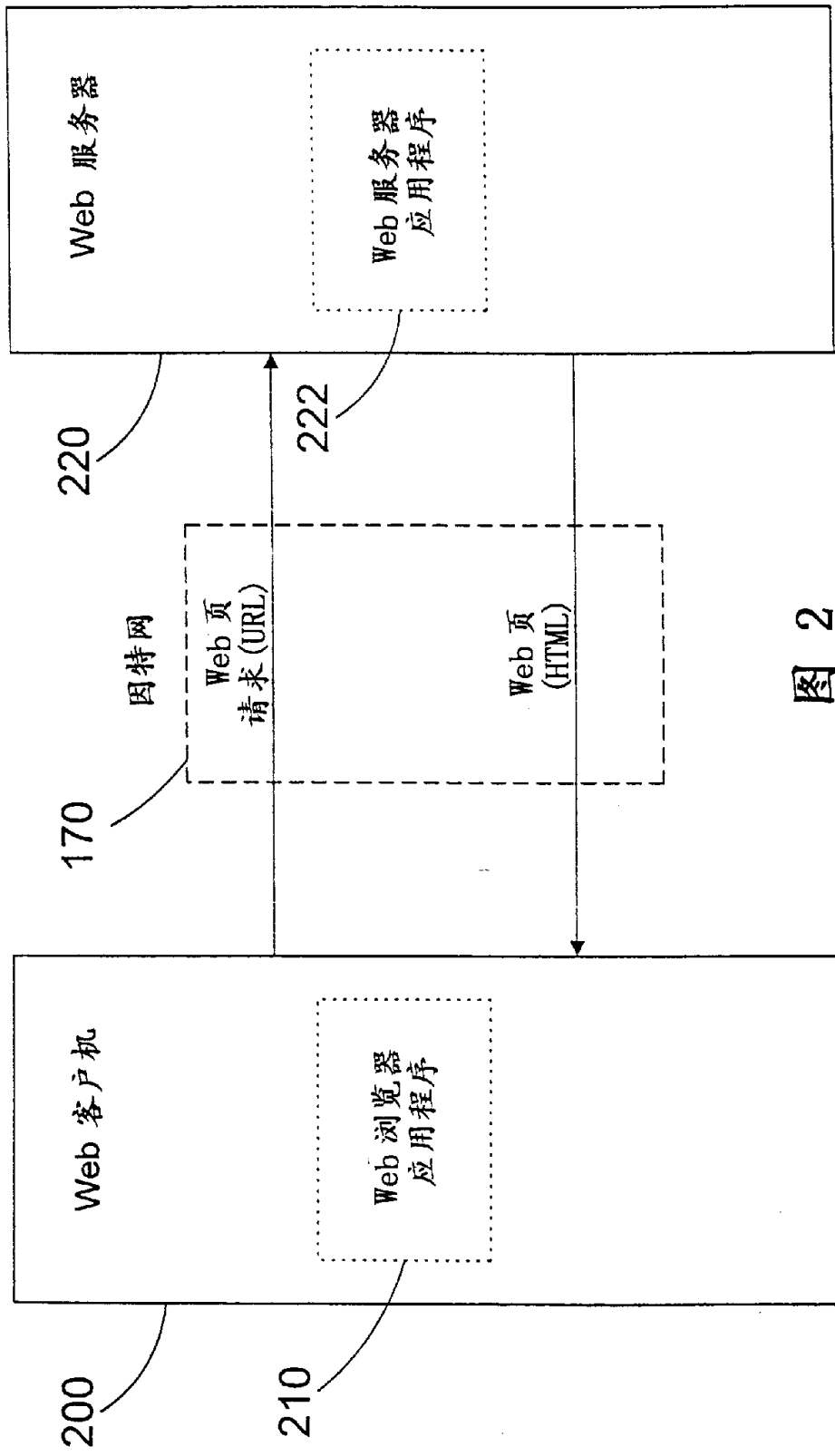


图 2

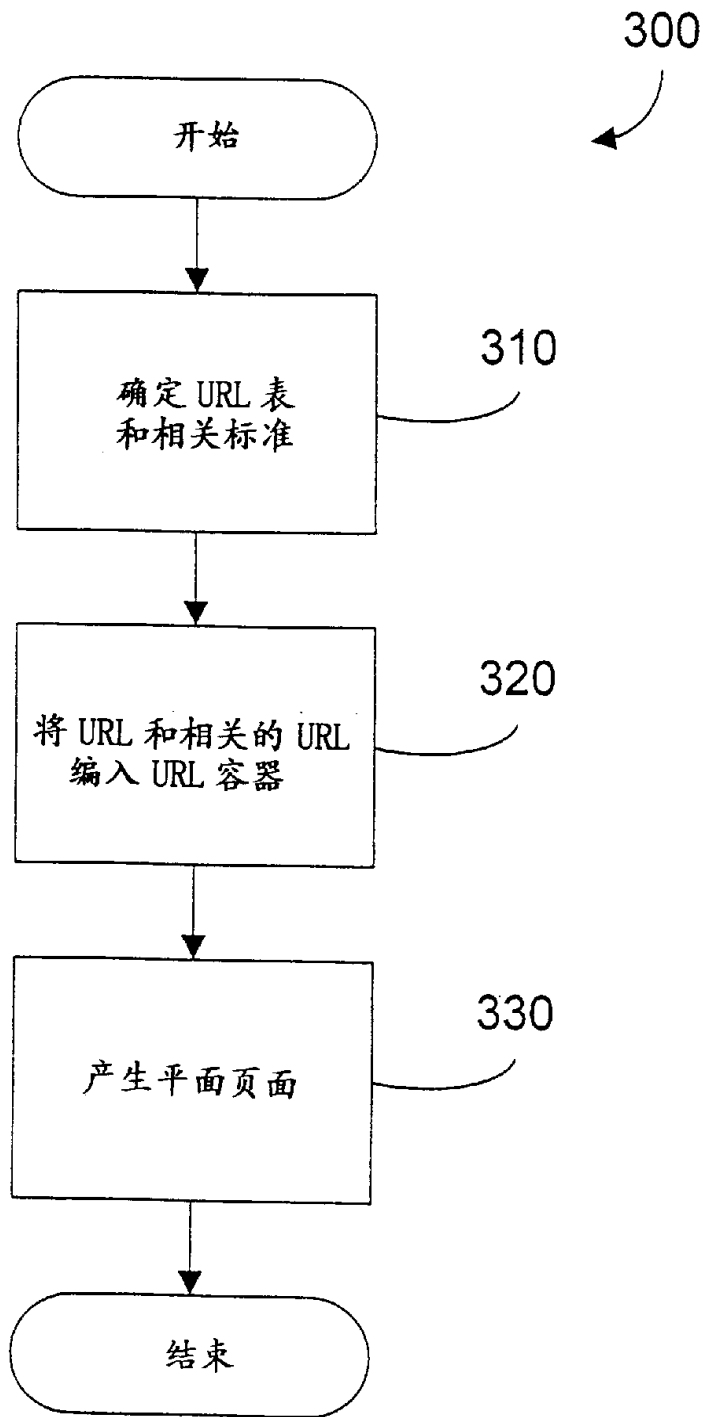


图 3

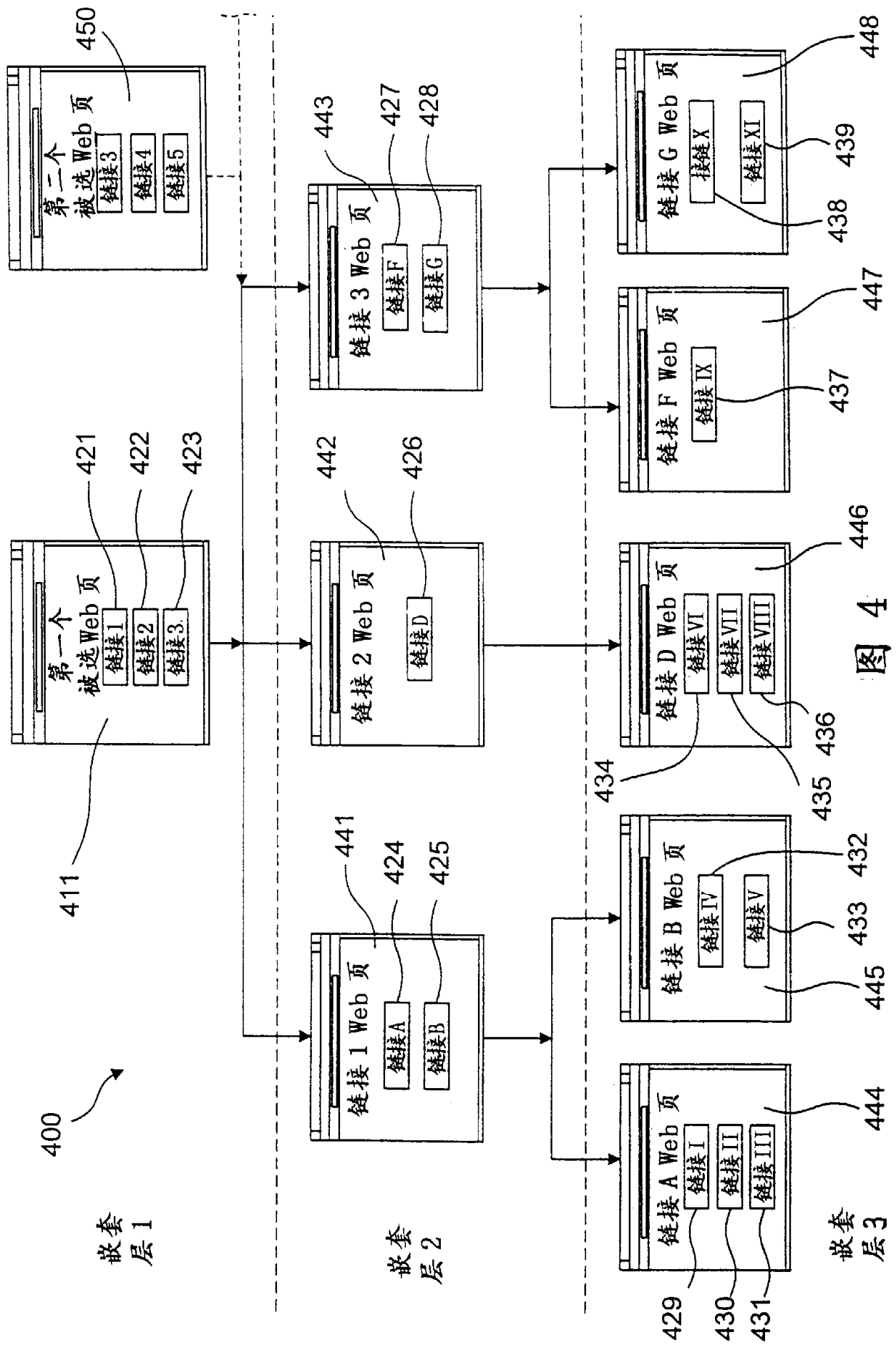


图 4

500

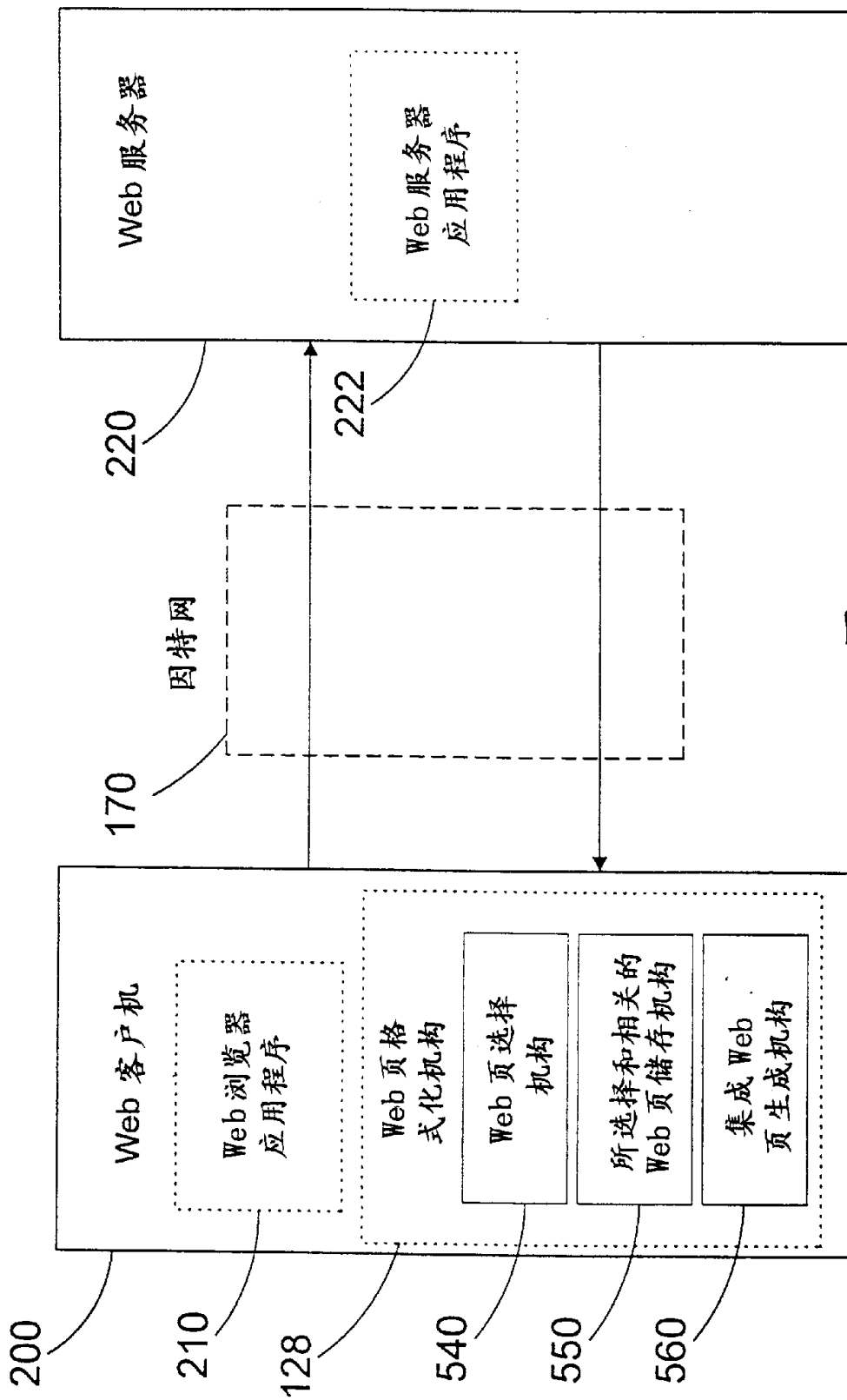


图 5

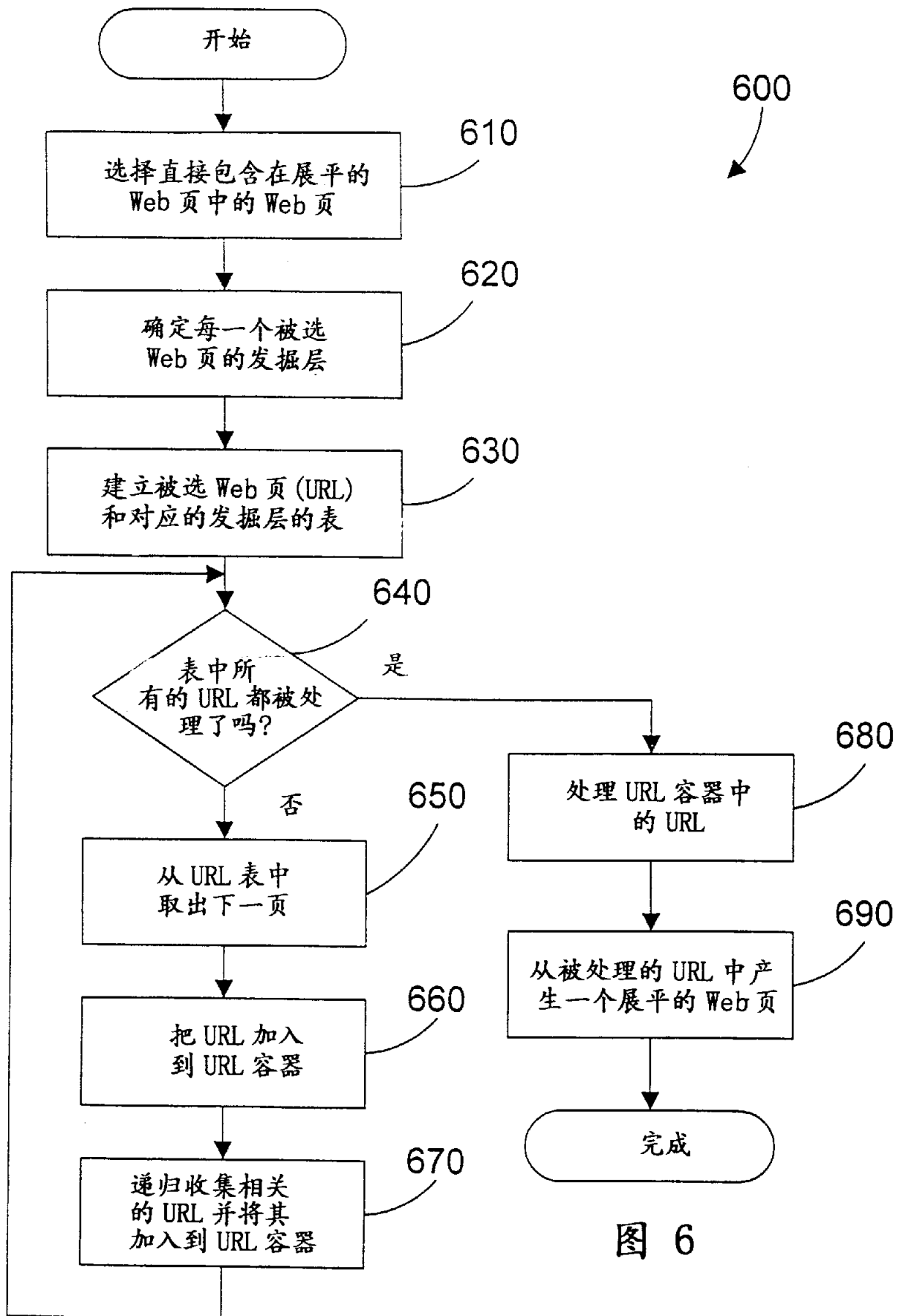


图 6

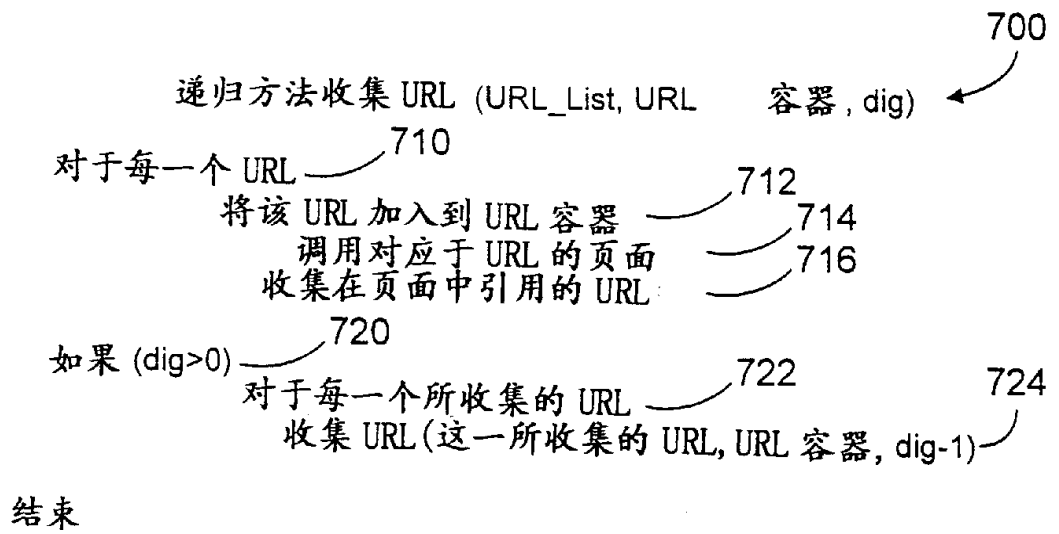


图 7

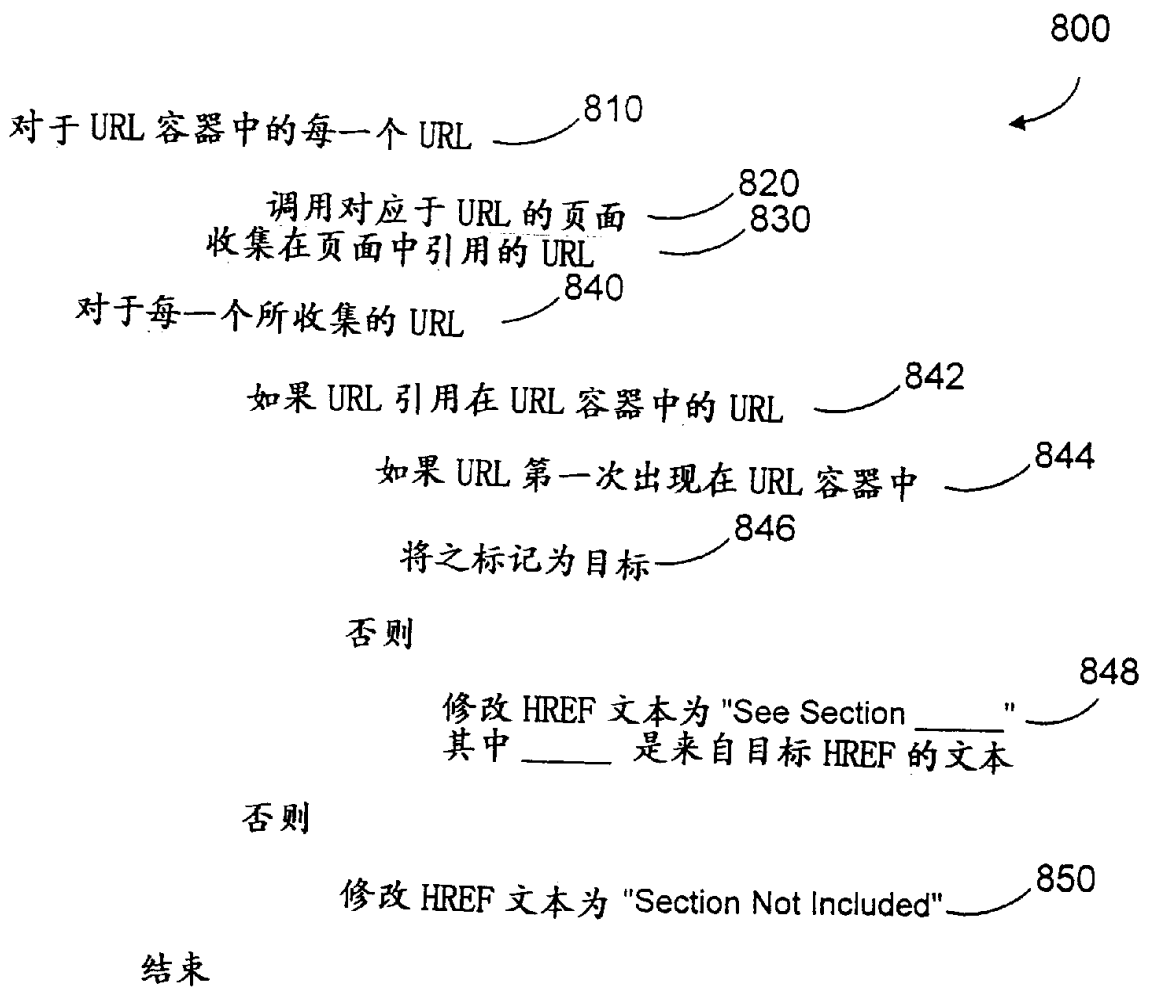


图 8

结束

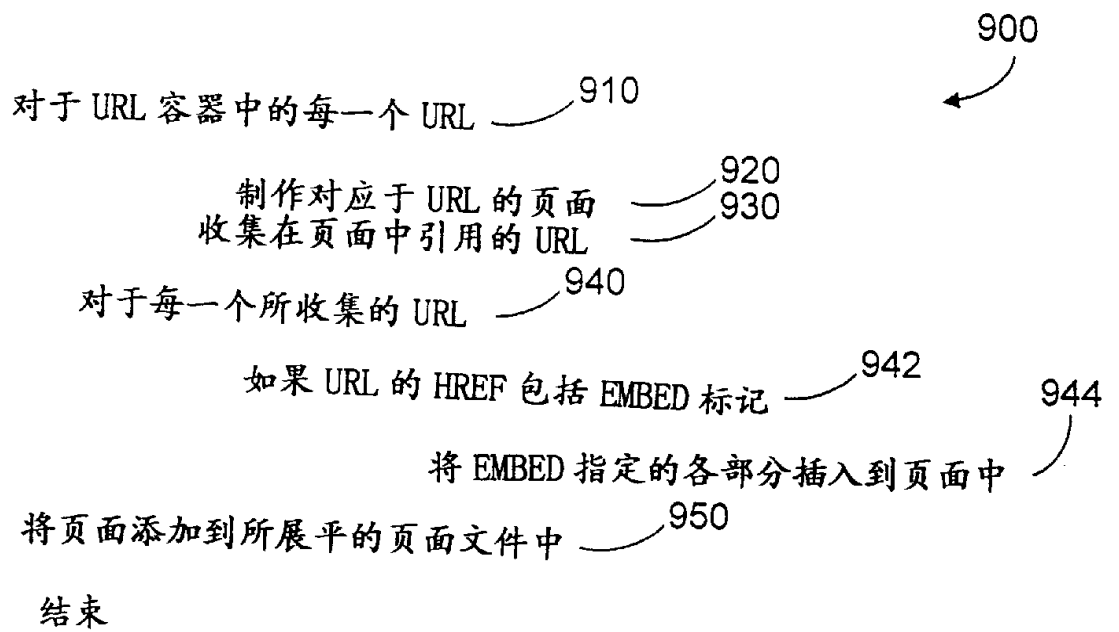


图 9