



(12) 发明专利

(10) 授权公告号 CN 101276362 B

(45) 授权公告日 2011.05.11

(21) 申请号 200710088954.2

US 2007/0016577 A1, 2007.01.18, 全文.

(22) 申请日 2007.03.26

JP 2002-366479 A, 2002.12.20, 全文.

(73) 专利权人 国际商业机器公司

US 2006/0123042 A1, 2006.06.08, 全文.

地址 美国纽约

CN 1716191 A, 说明书第 2 页第 5-27 行, 第

4 页第 20 行至第 9 页第 4 行.

(72) 发明人 兰东俊 叶萌 程龙 陈滢

审查员 慈丽雁

李海萍

(74) 专利代理机构 中国国际贸易促进委员会专

利商标事务所 11038

代理人 赵冰

(51) Int. Cl.

G06F 17/30 (2006.01)

(56) 对比文件

JP 2002-236638 A, 2002.08.23, 全文.

CN 1529266 A, 2004.09.15, 全文.

US 6973483 B2, 2005.12.06, 全文.

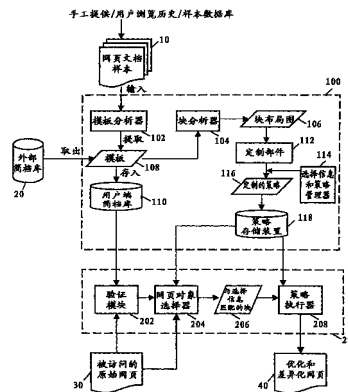
权利要求书 2 页 说明书 8 页 附图 3 页

(54) 发明名称

定制网页的装置和方法

(57) 摘要

本发明涉及优化和差异化网页浏览的装置和方法。定制网页的装置包括：块分析器，用于分析网页模板，得到构成网页模板的块元素；定制部件，用于选择要定制的块元素，并对所选择的块元素设定优化和/或差异化策略从而定制所选择的块元素；策略存储装置，用于与选择信息相关联地存储定制的策略。还提供了优化和差异化网页的装置，用于基于定制好的策略对网页进行优化和/或差异化，所述定制好的策略与选择信息和网页模板相关联地存储，该装置包括：网页对象选择器，用于比较被访问的原始网页和与定制的策略相关联的选择信息，确定网页中与选择信息相匹配的部分；以及策略执行器，针对所述匹配的部分执行相应的策略，从而显示优化和差异化的网页。



1. 一种定制网页的装置,包括:
 - 块分析器,用于分析网页模板,得到构成网页模板的块元素;
 - 定制部件,用于选择要定制的块元素,并对所选择的块元素设定优化和 / 或差异化策略从而定制所选择的块元素;
 - 策略存储装置,用于与选择信息相关联地存储定制的策略;
 - 模板分析器,用于分析要优化和差异化的网页的样本,从而提取网页模板,作为所述块分析器的输入。
2. 如权利要求 1 所述的装置,还包括:
 - 选择信息和策略管理器,用于管理用于选择定制对象的选择信息以及与选择信息相对应的可供选择的优化和 / 或差异化策略;
 - 其中,选择信息和策略管理器针对所述定制部件所选择的要定制的块元素,列出选择信息和相应的策略,所述定制部件在其中进行选择,从而完成定制。
3. 如权利要求 1 或 2 所述的装置,还包括:
 - 用户端简档库,用于存储用户定制的至少一个网页模板,
 - 其中,所述策略存储装置存储对应于所述至少一个网页模板的至少一个策略。
4. 如权利要求 1 或 2 所述的装置,其中,所述策略存储装置存储至少一个策略,并与其相关联地存储对应的网页模板。
5. 如权利要求 1 或 2 所述的装置,其中,所述块分析器被配置为通过检测网页模板脚本中的元素标记来得到构成网页模板的块元素。
6. 如权利要求 3 所述的装置,其中,所述块分析器被配置为通过检测网页模板脚本中的元素标记来得到构成网页模板的块元素。
7. 如权利要求 4 所述的装置,其中,所述块分析器被配置为通过检测网页模板脚本中的元素标记来得到构成网页模板的块元素。
8. 如权利要求 1 或 2 所述的装置,其中,所述模板分析器被配置为比较网页样本的脚本,在各网页样本之间相同的部分脚本构成模板。
9. 如权利要求 3 所述的装置,其中,所述模板分析器被配置为比较网页样本的脚本,在各网页样本之间相同的部分脚本构成模板。
10. 如权利要求 4 所述的装置,其中,所述模板分析器被配置为比较网页样本的脚本,在各网页样本之间相同的部分脚本构成模板。
11. 一种定制网页的方法,包括下述步骤:
 - 分析要优化和差异化的网页的样本,从而提取网页模板;
 - 分析网页模板,得到构成网页模板的块元素;
 - 选择要定制的块元素,并对所选择的块元素设定优化和 / 或差异化策略从而定制所选择的块元素;
 - 与选择信息相关联地存储定制的策略。
12. 如权利要求 11 所述的方法,其中:
 - 针对所选择的要定制的块元素,列出选择信息和相应的策略,在其中进行选择,从而完成定制。
13. 如权利要求 11 或 12 所述的方法,其中,相关联地存储网页模板和定制的策略。

14. 如权利要求 11 或 12 所述的方法,其中,所述分析网页模板的步骤包括通过检测网页模板脚本中的元素标记来得到构成网页模板的块元素。

15. 如权利要求 13 所述的装置,其中,所述分析网页模板的步骤包括通过检测网页模板脚本中的元素标记来得到构成网页模板的块元素。

16. 如权利要求 11 或 12 所述的方法,其中,所述提取网页模板的步骤包括比较网页样本的脚本,在各网页样本之间相同的部分脚本构成模板。

17. 如权利要求 13 所述的方法,其中,所述提取网页模板的步骤包括比较网页样本的脚本,在各网页样本之间相同的部分脚本构成模板。

定制网页的装置和方法

技术领域

[0001] 本发明涉及网页的定制,尤其涉及网页的优化和差异化。具体来说,本发明涉及优化和差异化网页浏览的装置和方法,以及实现该方法的程序产品。

背景技术

[0002] 在因特网上有数以百万计的站点。另外,越来越多的人的日常工作和生活依赖于某些网站。他们在这些网站上或许一天要浏览很多次,以浏览新闻、搜索信息、下载资源或者与他人通信等。如果用户能够根据自己的偏好来定制和优化他们经常光顾的网站,那将是很值钱的,这将提高速度、改善体验。考虑到这些网站 / 频道的内容,这种优化应当是语义方式的,也就是说能够从内容这个层面来有针对性地优化。

[0003] 当前的定制和优化主要是通过服务器端的用户信息管理。存在改善性能的网站优化服务,但它们是服务器端技术,不是以用户为中心的,因为在它们的优化处理中没有考虑最终用户的偏好。对于定制,在服务器端常常有用户帐户数据库和验证模块。一般,用户必须在服务器端建立自己的帐户。然后,用户必须使用网页应用程序所提供的不多的功能来定制网站并保存其定制。这些定制功能常常不令人满意。每一次用户都必须登录到网站上。只有在登录之后定制才会生效。这也给应用服务器带来巨大的压力,尤其是在高峰时段,许多用户同时访问的情况下。许多网站还不提供定制功能,比如某些新闻站点。在客户端,用户可以从客户机浏览器对某些部分加以修改,例如修改字体、文本颜色等。但是,这些功能要么是有限的,不涉及网站或者频道浏览的行为模式 (behaviour),要么只是面向熟悉 html 脚本语言的开发人员的。在这些情况下,尽管用户每天都访问这些网站,但是没有方便的手段来改变其某些行为方式。

发明内容

[0004] 本发明提供了能够实现以最终用户为中心的优化浏览和差异化浏览,从而改善整个网站性能和体验的方法和系统。简单地说其包括两个阶段。第一阶段是建立个性化的简档库。第二阶段时根据简档库优化和定制浏览,在此期间原始网页被转换为定制和优化网页,同时能够实现基于内容和用户偏好的差异化浏览。

[0005] 具体来说,本发明提供了一种定制网页的装置,包括:块分析器,用于分析网页模板,得到构成网页模板的块元素;定制部件,用于选择要定制的块元素,并对所选择的块元素设定优化和 / 或差异化策略从而定制所选择的块元素;策略存储装置,用于与选择信息相关联地存储定制的策略;模板分析器,用于分析要优化和差异化的网页的样本,从而提取网页模板,作为所述块分析器的输入。

[0006] 按照一个实施例,定制网页的装置还包括选择信息和策略管理器,用于管理用于选择定制对象的选择信息以及与选择信息相对应的可供选择的优化和 / 或差异化策略,其中,选择信息和策略管理器针对所述定制部件所选择的要定制的块元素,列出选择信息和相应的策略,所述定制部件在其中进行选择,从而完成定制。

[0007] 按照一个实施例,定制网页的装置还包括用户端简档库,用于存储用户定制的至少一个网页模板,其中,所述策略存储装置存储对应于所述至少一个网页模板的至少一个策略。

[0008] 按照一个实施例,所述策略存储装置存储至少一个策略,并与之相关联地存储对应的网页模板。

[0009] 按照一个实施例,所述块分析器被配置为通过检测网页模板脚本中的元素标记来得到构成网页模板的块元素。

[0010] 按照一个实施例,所述模板分析器被配置为比较网页样本的脚本,在各网页样本之间相同的部分脚本构成模板。

[0011] 本发明还提供了一种优化和差异化网页的装置,用于基于定制好的策略对网页进行优化和 / 或差异化,所述定制好的策略与选择信息和网页模板相关联地存储,该装置包括:网页对象选择器,用于比较被访问的原始网页和与定制的策略相关联的选择信息,确定网页中与选择信息相匹配的部分;以及策略执行器,针对所述匹配的部分执行相应的策略,从而显示优化和差异化的网页。

[0012] 本发明另外还提供了一种定制网页的方法,包括下述步骤:分析要优化和差异化的网页的样本,从而提取网页模板;分析网页模板,得到构成网页模板的块元素;选择要定制的块元素,并对所选择的块元素设定优化和 / 或差异化策略从而定制所选择的块元素;与选择信息相关联地存储定制的策略。

[0013] 在本发明一个实施例中,针对所选择的要定制的块元素,列出选择信息和相应的策略,在其中进行选择,从而完成定制。

[0014] 在本发明一个实施例中,相关联地存储网页模板和定制的策略。

[0015] 在本发明一个实施例中,所述分析网页模板的步骤包括通过检测网页模板脚本中的元素标记来得到构成网页模板的块元素。

[0016] 在本发明一个实施例中,所述提取网页模板的步骤包括比较网页样本的脚本,在各网页样本之间相同的部分脚本构成模板。

[0017] 本发明还提供了一种优化和差异化网页的方法,用于基于定制好的策略对网页进行优化和 / 或差异化,所述定制好的策略与选择信息和网页模板相关联地存储,该方法包括下述步骤:比较被访问的原始网页和与定制的策略相关联的选择信息,确定网页中与选择信息相匹配的部分;针对所述匹配的部分执行相应的策略,从而显示优化和差异化的网页。

[0018] 本发明还提供了用于使计算机执行上述方法的程序,以及存储有这样的程序的存储介质。

[0019] 与同用户简档相关的方法相比,本发明的系统不要求在服务器端数据库中对每一个用户建立用户帐户。用户能够通过客户机端的策略存储装置来定制其访问的网站或者频道。这降低了应用服务器的工作负荷,使其能够使用相同的基础设施来同时支持更多的用户。另外,本发明的方法和系统帮助用户按需优化其对网站的访问。通过用户预先定义的策略,优化的不仅是网页的视图,还包括网站的行为方式。其还帮助用户主动地保护自己免受恶意网页文件的侵害。因为本发明的方法和系统是基于模板和基于块的,通过该系统的自动执行的运行时模块,用户能够从其经常访问的站点 / 频道提取模板,并对其想访问和

不想访问的块进行定制。

附图说明

[0020] 下面结合附图描述本发明。附图中：

[0021] 图 1 是根据本发明的定制装置和定制应用装置优选实施方式组成的系统的框图；

[0022] 图 2 是根据本发明的定制方法的优选实施方式的流程图；

[0023] 图 3 是根据本发明的定制应用方法的优选实施方式的流程图。

具体实施方式

[0024] 首先,下面结合附图描述本发明的定制装置和定制应用装置的优选实施方式。

[0025] 图 1 为本发明的优化和差异化装置的系统图。该装置可以包括两个部分,一部分是定制装置 100,另一部分是定制应用装置 200。按照图 1 所示,定制装置 100 和定制应用装置 200 共同构成一个系统,基于网页文档样本 10 或者来自外部简档库 20 的模板 108,定制策略 116,并对被访问的原始网页 30 应用策略,从而得到优化和差异化的网页 40。

[0026] 但是,定制装置 100 和定制应用装置 200 可以分开实施。一部分用户可以利用定制装置 100 定制策略;另一部分用户可以利用定制应用装置 200 将其他人定制好的策略应用到其想访问的网页。可以想象到的一种情况是,由第三方服务提供商利用定制装置 100 针对各种网站、频道、网页定制各种策略,并将其提供给最终用户。最终用户则在访问网页时利用定制应用装置 200 将第三方服务提供商提供的策略应用于其访问的网页。

[0027] 下面进一步结合附图描述定制装置 100 和定制应用装置 200。需要说明的是,图 1 中图示出了根据本发明的一种优选实施方式的所有部件,但是,取决于具体情形,这些部件并不都是必需的。这将在下文对这些部件的详细说明中加以说明。

[0028] 如图 1 所示,定制装置 100 可以包括以下部件:模板分析器 102,用户端简档库 110,块分析器 104,定制部件 112,选择信息和策略管理器 114,以及策略存储装置 118。定制应用装置 200 可以包括以下部件:验证模块 202,文档对象选择器 204 以及策略执行器 208。

[0029] 下面对以上部件以及外部简档库 20 分别予以详细说明。

[0030] 如图 1 所示,在定制装置 100 中,可以网页文档样本 10 作为输入,模板分析器 102 从之提取站点或者网页文档样本所属特定频道的模板。然后将模板存入用户端简档库 110。块分析器 104 帮助用户指定用户要定制网页文档中的哪一个块,其以简档和网页文档作为输入。基于所选择的块,选择信息和策略管理器 114 控制用户能够规定的定制。其单独地记录作为选择信息的块上下文信息以及作为策略的定制信息。这些记录最终被存入策略存储装置 118 中。

[0031] 网页文档样本 10 是文档样本的原始数据集。网页文档样本为用户交互定制网站或者网页文档样本所属频道提供了一个起点。用户指定其想定制的网站或者频道,并提供网页文档样本作为例子。一般,为了从之提取模板,需要一个以上的样本。另外的样本或者是从用户的浏览历史中提取,要么是从网页文档样本数据库中提取,如果其 URL 与目标网站(频道)的 URL 匹配的话。如果不匹配,则用户需要手工提供另外的样本。网页文档样本用作下面的模板分析器和块分析器的输入。

[0032] 模板分析器 102 用来为网站/频道从网页文档样本 10 提取模板 108。网站或者频

道是网页的集合。它们具有其自己具体的模板,因此具有共同的外观和风格。模板是预先准备的主控网页(master webpage),用作编辑这些新网页的基础。当在浏览器上显示模板时,就是完整网页去除内容之后的框架,其由不同的块构成,例如其中填充文字的文字块,其中显示图像的图形块。换句话说,在网页或者网页模板中,“块”对应于在什么位置应当显示什么内容的标记。网页中的所有这些标记就构成了模板。对于多个网页样本来说,这些网页样本中相同的标记部分就构成了这些网页样本的模板。注意,在网页文档中存在两种模板。第一种是它们连接的公共层叠样式表(commonCascaded Style Sheet),其定义站点或者频道上的总体外观。另一种是网页内的模板,通过对所提供的样本进行比较过程来提取这样的模板。大多数网站两种模板都有。但是,某些旧样式的网站可能只有后一种模板。对于前一种模板,模板分析器 102 可以直接从网站提取,即下载 CSS(Cascaded Style Sheet,层叠样式表)文件。对于后一种模板,可以简单地通过比较至少两个网页文档样本来提取。对此可以参照下文对本发明的方法的描述。

[0033] 用户端简档库 110 用来存储所生成的模板 108。由于需要识别不同的模板,因此以简档的形式来存储模板。简档的每一条记录可以包括下述信息中的一个或者多个:名称,用户,站点,频道,模板,CSS。“名称”字段是唯一用来区分不同记录的。“用户”字段用来表示拥有该记录的用户帐户,这意味着网络浏览器可以针对不同用户维护各种简档。如果只有一个用户,则不需要“用户”字段。“站点”字段表示简档所属的网站。类似地,如果只有一个网站,则不需要该字段。同一站点可能具有多个频道,例如新闻、体育等。每一个频道具有不同的模板和样式,这是通过“频道”字段来表示的。同理,在只有一个频道的情况下,则不需要“频道”字段。“模板”和“CSS”是在站点和频道上共享的内容。

[0034] 上面描述了从网页文档样本 10 提取模板 108。但是,模板 108 也可以由第三方提供。在本发明中,用外部简档库 20 来表示第三方提供的模板的来源。外部简档库 20 类似于用户端简档库 110,存储网站/频道的模板。差别在于模板(简档)是由第三方提供者提供的。例如,某个第三方服务提供商提供用户想要定制的网站(频道)的简档记录。用户从第三方提供商下载简档而不是自己去生成它们。在某些情况下,网站所有者可能也想公开其站点的简档,使得其他人能够自由定制。在这些情况下,通过网站所有者提供的服务来查询简档。

[0035] 第三方可能提供了大量的模板,这些模板可能并非是每一个用户都全部需要的。另外,这些模板可能并不位于本地,而是在远程服务器中。因此,当用户从外部简档库 20 中获得自己所需的模板后,可以将其存入用户端简档库 110 以供以后使用。当然,如果方便的话,也可以将外部简档库当作用户端简档库或者用户端简档库的一部分来使用。

[0036] 获得模板 108 之后,由块分析器 104 对模板和网页进行分析,以获得网页的块布局图 106。网页模板中的块就是表示显示样式(blockdisplay style)的元素(element)所标记的部分。例如在 HTML 语言中,这样的元素包括 <div>, , <dl>, , <table>, <tr>, <td>, <p>, <h1 ~ 6>, <frame> 等。因此,检测模板中的块其实就是检测网页脚本中的这些元素标记。也就是,块分析器 104 提取模板中的各个组成部分(也就是块)的标记,从而获得这些部分的信息。所谓的块布局图,如前所述,就是相当于去掉内容之后的网页显示。当然,为了直观起见,可以对不同的块进行区别显示,也可以在显示时在块中保留某个网页样本的全部或者部分内容。用户定制的目标必须是网页中的块元素,而不是内部元素(inline

element) 或者文本。如上所述,基于网站/频道模板,可以将网页划分为模板信息和内容信息。用户可以定制模板中的每一个块,但是对于内容信息,用户只能将整个块作为一个整体来定制,因为内容信息在各页之间可能完全不同。

[0037] 在获得网页的块布局图 106 之后,就可以由定制部件 112 对感兴趣的块进行优化和差异化设定,或者说设定有关的“策略”,以改进网站性能和用户体验。这包括基本内容优化、图形和多媒体优化、脚本优化、控制优化和显示优化,等等。

[0038] * 基本内容优化:关于块可见或者不可见的选项,等等。

[0039] * 图形和多媒体优化:下载与否选项,播放与否选项,下载级别(下载的优先级)选项等。

[0040] * 脚本优化:下载与否选项,执行与否选项,下载级别选项等。

[0041] * 控制优化:下载级别选项,强制并行下载选项,等。

[0042] * 显示优化:显示级别(显示优先级)选项,保持在屏内(keepfocus)选项等。

[0043] 定制部件的设定可以完全手工进行,例如按照一定的语法规则直接输入。作为优选的实施方案,为了辅助定制部件 112,使得本发明的定制装置对用户友好,可以提供选择信息和策略管理器 114。该管理器控制和记录用户可能在网站/频道上制订的优化规则。这里,对每一个规则保存两种信息。第一种是选择信息,其定义该规则应用于网页内的什么标记(tag)或者元素(element)。另一种是策略,其定义支持什么样的优化和可以指定什么样的优化。选择信息可以是网页元素(web element)的类别或者 ID,或者网页文档内的上下文信息。可以定义的策略的例子包括“不下载块内的视频”、“不下载块内的图像”、“不显示块”等等。作为一种具体实施方式,可以在用户选中块布局图中的某一个块时,该管理器显示出可以定制的标记或者元素(例如列表方式或者下拉菜单方式)。与此同时,或者在用户选择要定制的标记或者元素之后,显示相应的可供选择的策略(例如列表方式或者下拉菜单方式)。策略可以分为两种,一种是差异化策略,其反映用户的偏好,包括下载级别或者显示级别等,还可以为块指定其他样式,比如背景颜色、字体。另一种是优化策略,这些策略与网页的视图或者控制优化有关。

[0044] 策略存储装置 118 用来记录用户进行的定制,以供以后使用。策略存储装置中的每一条记录链接到用户端简档库 110 中的用户端简档记录。每一条记录可以包含以下字段:名称,用户端简档名称,选择信息,策略。“名称”是规则的唯一标识符。“用户端简档名称”规定该规则将应用于哪个网站(频道)。“选择信息”定义策略将应用于网页内的什么标记(tag)/元素(element)。“策略”表示规则的详细信息。注意,在同一记录中可以存在多个策略。多个个性化简档可以对应于同一个用户端简档记录,按照特定顺序应用。

[0045] 从上面的描述可以看到,选择信息、策略和用户端简档记录(即网页模板)是相互关联的关系。因此,可以将用户端简档库 110 和策略存储装置 118 合二为一(图 1 未图示),存储到在一个数据库中。

[0046] 在以上各部件完成策略的定制以后,当用户要访问某个原始网页 30 时,验证模块 202 获取原始网页 30 的 URL,然后使用 URL 从用户端简档库 110 中进行查询,看该原始网页是否被定制过。如果在用户端简档库 110 中存在所述 URL,则表示该原始网页被定制过。如前所述,如果在用户简档库 110 中区分用户,则该查询还应包括用户信息,即在用户端简档库 110 中查询同时包括该 URL 和相应用户信息的条目。若查询到,则表示相应用户对该原

始网页进行过定制。如果原始网页被定制过,则调用后面的部件,执行对应的定制策略。验证模块还可以检查网页文档的修订日期,比较原始网页与存储的网页模板,看在模板生成之后原始网页是否有变化。如果有,则验证模块更新模板,并验证基于原始模板的策略是否仍然有效。例如被定制的块是否仍然存在,或者性质是否有变化。验证模块 202 给用户关于这些变化的对应信息。

[0047] 如前所述,对于从第三方(包括被访问的网站本身)获取的模板,也可以不存储在用户本地而是仍然在第三方的外部简档库中。这个时候,验证工作是针对外部简档库进行的,如果验证结果是肯定的,则需要将模板下载到本地(未图示)。但是,由于外部简档库可能包含很多模板,某个特定的用户可能并未对其中所有的模板进行定制,因此,在这种情况下,验证操作还需要访问策略存储装置,看被访问的网页是否有对应的策略定制信息。

[0048] 文档对象选择器 204 首先获取策略存储装置的对应于网页模板(如前所述,网页模板可存储于策略存储装置中,或者存储于单独的客户端简档库中)的所有记录。然后,在网页解析过程中,文档对象选择器将这些记录的选择信息与解析的网页对象进行匹配。只对匹配的部分应用相应的规则。

[0049] 策略执行器 208 进行控制,以按照定制策略来取出和显示网页。通过策略执行器,原始网页被转换为定制网页,包括根据网站或者频道的预定定制规则进行网页的取出和显示。例如,策略“不显示图像内的图像”被转译,然后浏览器就不会启动新的获取该块内的图像的请求,而其他块比如内容块中的图像仍会被下载和显示。又如,在策略中为块指定的其他样式,比如背景颜色、字体等,这些新的样式具有最高优先级,优先于原始网页的样式。

[0050] 下面结合图 2 和图 3 描述本发明的定制方法和定制应用方法的优选实施方式。需要注意的是,本发明的方法当涉及与上面说明的装置类似的技术问题时,上面的说明和这里对方法的说明可以相互参照。

[0051] 本发明的方法包括定制方法和定制应用方法,它们可以一并实施,也可以分开实施。也就是说,一部分用户可以使用本发明的定制方法(如图 2 所示)定制策略;另一部分用户可以使用本发明的定制应用方法(如图 3 所示)将其他人定制好的策略应用到其想访问的网页。可以想象到的一种情况是,由第三方服务提供商针对各种网站、频道、网页定制各种策略,并将其提供给最终用户。最终用户则在访问网页时利用本发明的定制应用方法将第三方服务提供商提供的策略应用于其访问的网页。

[0052] 同样,需要说明的是,图 2、图 3 中图示出了根据本发明的优选实施方式的所有步骤,但是,取决于具体情形,这些步骤并不都是必需的。这将在下文对这些步骤的详细说明中加以说明。

[0053] 下面首先结合图 2 描述本发明的定制方法的优选实施方式。

[0054] 如图 2 所示,本发明的定制方法始于步骤 S101。首先要提供网页文档样本(步骤 S102)。网页文档样本可以由用户提供,或者可以从用户的浏览历史记录中获取,或者可以从样本数据库中获取。然后,在步骤 S103 中,从网页文档样本提取模板。

[0055] 模板有两类。第一种是各网页文档样本连接的公共级联样式表(common Cascaded Style Sheet),其定义站点或者频道上的总体外观。对于这种模板,可以直接从网站提取(步骤 S201)。另一种是网页内的模板,可以通过对所提供的样本进行比较过程来提取这样的模板,即,各样本中相同的部分即为模板。例如,比较网页脚本的由标记构成的框架,相同

部分即可视为模板。大多数网站两种模板都有。但是,某些旧样式的网站可能只有后一种模板。此时,就没有图 2 中的步骤 S201。

[0056] 为提取模板而进行的比较过程属于常规的比较,可以有多种实现方式。例如,可以先比较两个样本(例如最前两个样本)(步骤 S202),从而得到初步的模板。然后再将该初步的模板与更多的样本进行进一步比较,从而使模板进一步精确化(步骤 S203)。在优选的实施方式中,为了方便在以后应用定制的策略时判断当时的模板是否仍然是定制策略时的模板,从而判断定制的策略是否仍然有效(可参见下文对定制应用方法的说明),可以记录模板的修订日期(S204)。另外需注意,步骤 S201 不一定在步骤 S202 和 S203 之前。

[0057] 模板也可以由第三方或者被访问的网站本身提供。在这种情况下,上述步骤 S102 和 S103 被代之以从所述第三方或者被访问的网站获取模板的步骤 S104。

[0058] 提取的模板可以保存起来(步骤 S105)以供以后使用。从第三方(包括被访问的网站本身)获取模板也可以保存在用户本地以供以后使用,但也可以不保存在本地,这样以后使用模板时仍需要从第三方下载。

[0059] 下一步是检测模板中的块(步骤 S106)。如前所述,网页模板中的块就是表示显示样式(block display style)的元素(element)所标记的部分。例如在 HTML 语言中,这样的元素包括 <div>, , <dl>, , <table>, <tr>, <td>, <p>, <h1 ~ 6>, <frame> 等。因此,检测模板中的块其实就是检测网页脚本中的这些元素标记。

[0060] 定制策略的步骤 S107 是在块的基础上进行的。因为,对于要长期适用的策略,必定是针对网页的架构而不是针对其中的具体内容,因为具体内容会改变。

[0061] 图 2 右侧显示了步骤 S107 的一种优选实施方式。为了方便用户的定义,首先可以生成网页内所有块的布局图(步骤 303),用户在显示的决布局图上选择要定制的块(步骤 S304)。为了方便用户,可以突出显示所选中的块(步骤 S309)。然后对选中的块指定策略(步骤 S310)。指定策略的方式可以是直接输入,也可以提供一个可能策略的列表,从中进行选择。选择块的方式可以采用任何常规的选择对象或者文本的方式。例如,当点击一个块内的某处时,就选中了该块。或者可以通过选择内容的方式来选择块,当某个块内被选中的内容超过一定比例时,即可认为选定了该块,等等。

[0062] 在定制了策略之后,需要将其保存起来(步骤 S109)。策略和网页模板可以保存在不同的位置或者相同的位置,但是它们保存应当是相互关联的。

[0063] 上面结合图 2 描述了本发明的定制方法。下面结合图 3 描述本发明的定制应用方法。

[0064] 如图所示,本发明的定制应用方法始于步骤 S401。首先,用户会请求访问一个网页,我们称之为原始网页(步骤 S402)。这时,首先需要验证该网页是否被定制过(步骤 S403)。与本发明的定制方法相应,可以搜索本地是否保存了该网页的模板。如果保存了,则意味着对该模板进行过定制。但是,如前所述,模板有可能是第三方提供的,或者,也有可能用户提取或者下载并保存了模板,但是并未进行定制。

[0065] 如果验证结果是没有定制过,则按照原始网页进行解析(步骤 S407)和显示(步骤 S413)。如果验证结果表明进行过定制,则解析网页,同时将文档中的各对象与所存储的策略所对应的选择信息进行匹配(步骤 S409)。如果某文档对象与某选择信息匹配(判断步骤 S410),则对匹配的文档对象执行所述选择信息对应的策略(步骤 S412)从而按照定制

的策略处理（下载、显示）该对象，否则直接处理该对象。

[0066] 如本领域的普通技术人员所能理解的，本发明的方法和装置的全部或者任何步骤或者部件，可以在任何计算设备（包括处理器、存储介质等）或者计算设备的网络中，以硬件、固件、软件或者它们的组合加以实现，这是本领域普通技术人员在了解本发明的内容的情况下运用他们的基本编程技能就能实现的，因此不需在此具体说明。

[0067] 此外，显而易见的是，在上面的说明中涉及到选择、指定等动作的时候，无疑要使用与任何计算设备相连的任何显示设备和任何输入设备、相应的接口和控制程序。总而言之，计算机、计算机系统或者计算机网络中的相关硬件、软件和实现本发明的前述方法中的各种操作的硬件、固件、软件或者它们的组合，即构成本发明的数据分析设备及其各组成部件。

[0068] 因此，基于上述理解，本发明的目的还可以通过在任何信息处理设备中运行一个程序或者一组程序来实现。所述信息处理设备可以是公知的通用设备。因此，本发明的目的也可以仅仅通过提供包含实现所述方法或者设备的程序代码的程序产品来实现。也就是说，这样的程序产品也构成本发明，并且存储有这样的程序产品的存储介质也构成本发明。显然，所述存储介质可以是本领域技术人员已知的，或者将来所开发出来的任何类型的存储介质，因此也没有必要在此对各种存储介质一一列举。

[0069] 在本发明的设备和方法中，显然，各部件或各步骤是可以分解和 / 或重新组合的。这些分解和 / 或重新组合应视为本发明的等效方案。

[0070] 以上描述了本发明的优选实施方式。本领域的普通技术人员能够理解，本发明的保护范围并不局限于这里所公开的具体细节。这些具体方式可以在本发明的实质精神所及的范围内进行各种修改和等同替换。

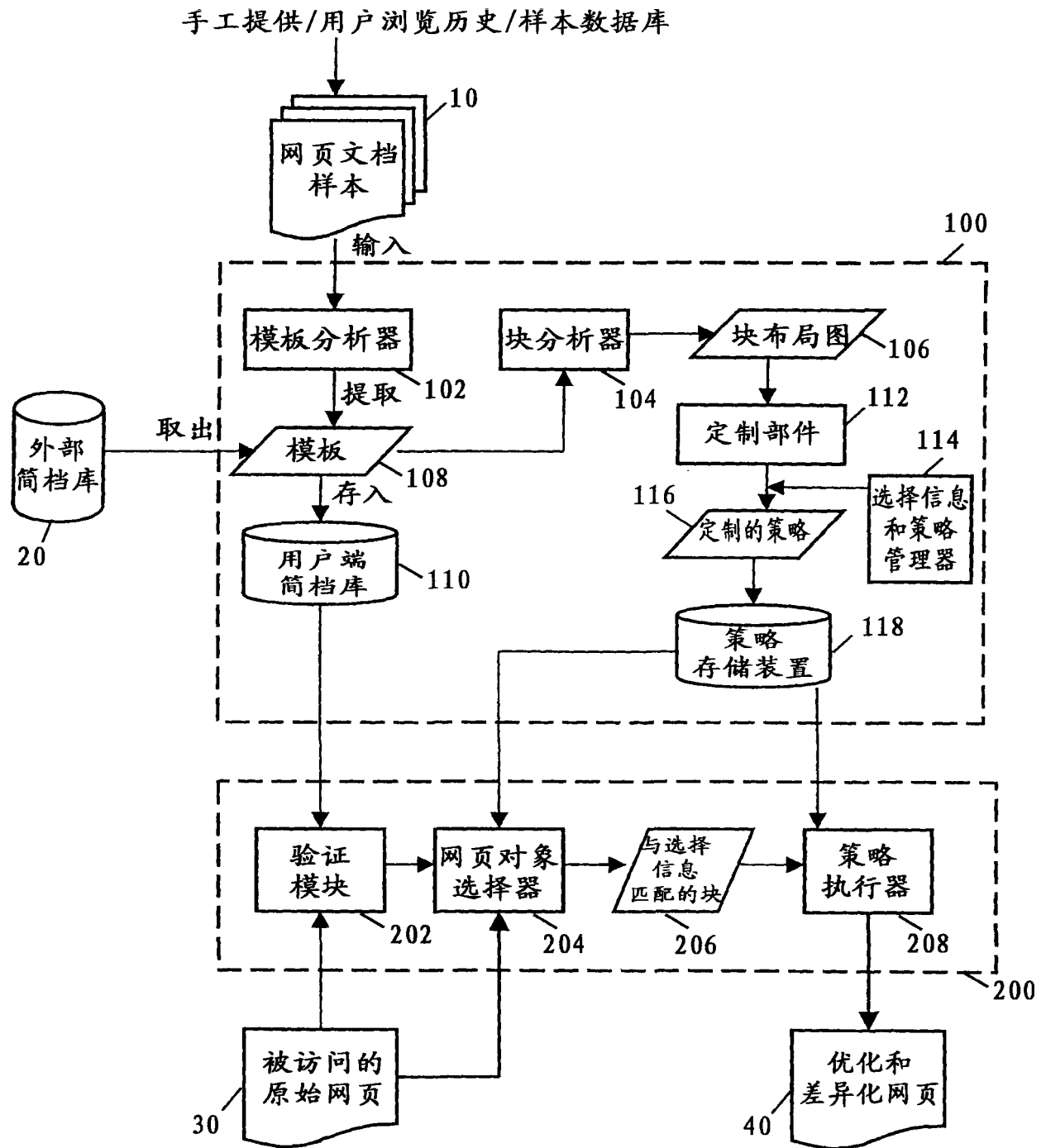


图 1

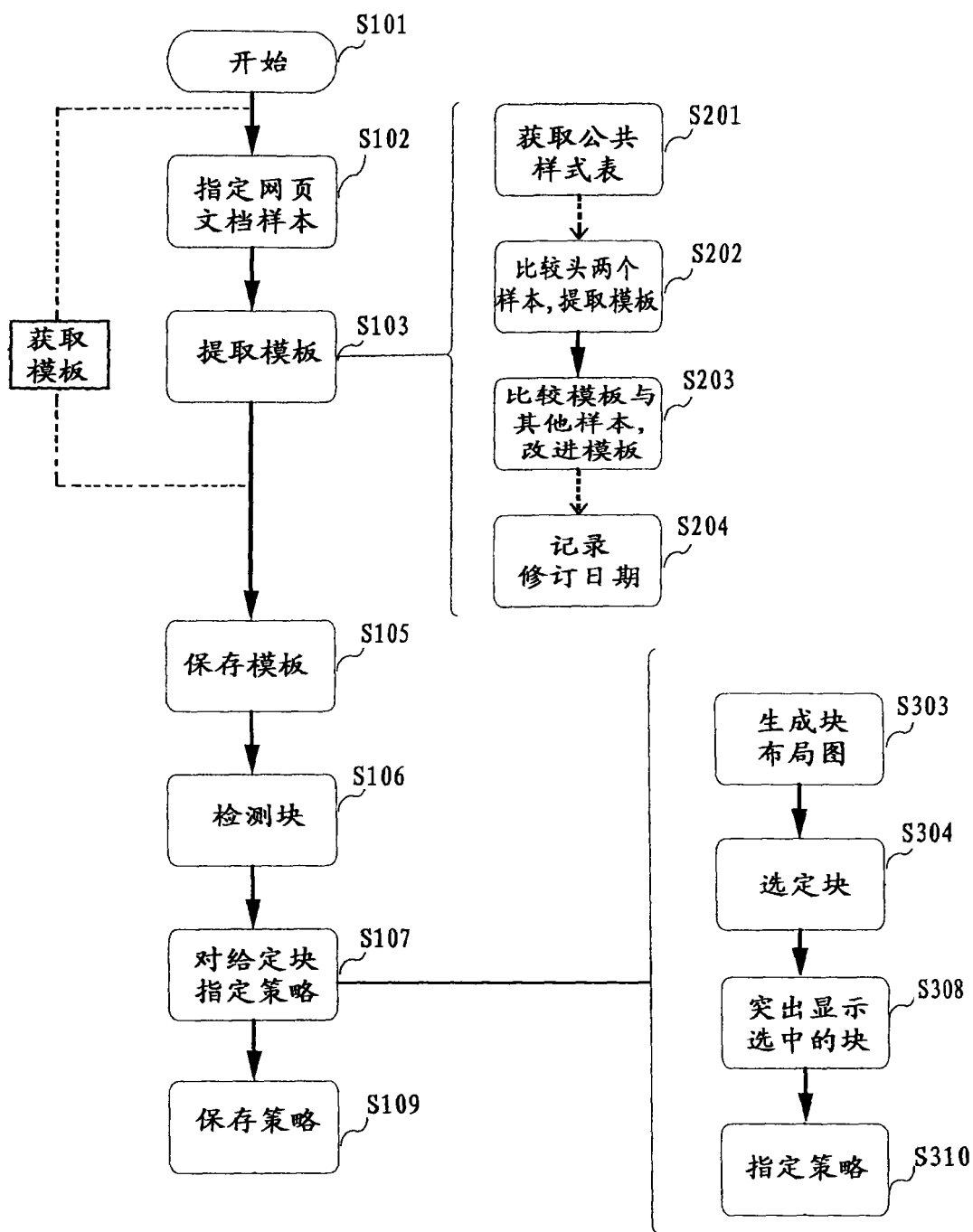


图 2

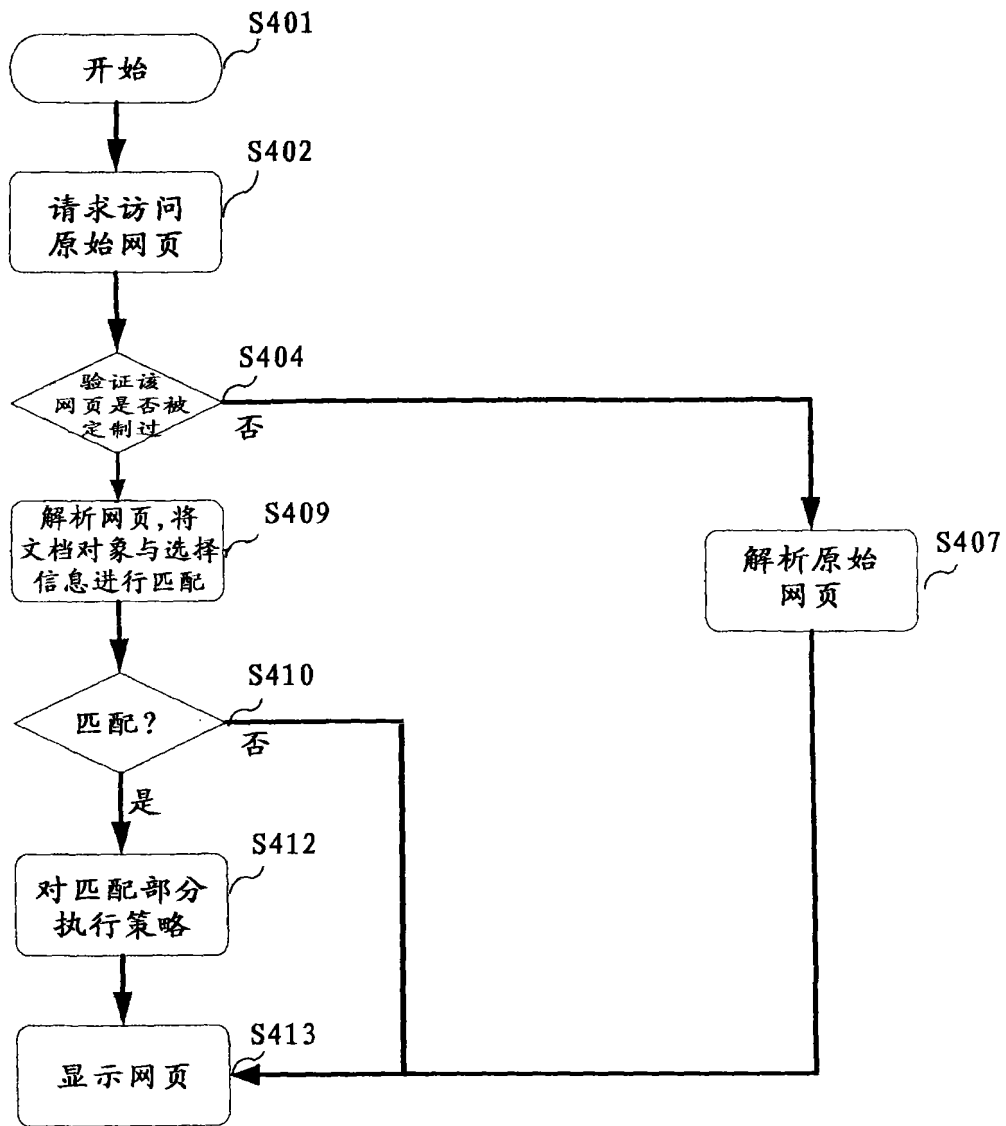


图 3