



(12) 发明专利

(10) 授权公告号 CN 103116609 B

(45) 授权公告日 2015. 11. 25

(21) 申请号 201310020113. 3

(22) 申请日 2013. 01. 18

(73) 专利权人 北京奇虎科技有限公司

地址 100088 北京市西城区新街口外大街
28号D座112室(德胜园区)

专利权人 奇智软件(北京)有限公司

(72) 发明人 文哲

(74) 专利代理机构 北京市隆安律师事务所
11323

代理人 权鲜枝 董垚

(51) Int. Cl.

G06F 17/30(2006. 01)

H04L 29/08(2006. 01)

(56) 对比文件

CN 102646135 A, 2012. 08. 22,

US 7747749 B1, 2010. 06. 29,

US 8103742 B1, 2012. 01. 24,

审查员 张莹

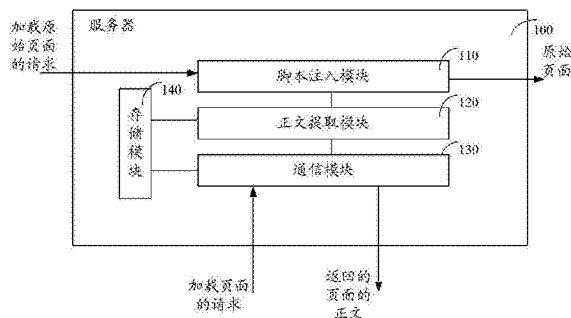
权利要求书3页 说明书12页 附图6页

(54) 发明名称

实现浏览器中页面加载的服务器、客户端、系统和方法

(57) 摘要

本发明涉及网络数据通信领域,公开了实现浏览器中页面加载的服务器、客户端、系统和方法,该服务器包括:脚本注入模块,适于在接收到客户端发送的加载原始页面的请求后,向发送给客户端的原始页面中注入正文加载脚本;正文提取模块,适于从客户端当前浏览的页面中解析出用于表示页面间正文的相互关系的关键字,提取关键字对应的链接,从网站下载所述链接对应的页面,提取下载的页面中正文,将提取的正文存储到存储模块中;存储模块,适于存储提取的正文;通信模块,适于接收客户端发送的加载关键字对应的链接的页面的请求,将存储模块中存储的该链接对应的正文发送给客户端。本发明能够解决在客户端造成无用的下载流量,浪费网络流量的技术问题。



1. 一种实现浏览器中页面加载的服务器,该服务器包括:

脚本注入模块,适于在接收到客户端发送的加载原始页面的请求后,向发送给客户端的原始页面中注入正文加载脚本,所述原始页面为初始访问网站时所浏览的页面;

正文提取模块,适于从客户端当前浏览的页面中解析出用于表示页面间正文的相互关系的关键字,提取关键字对应的链接,从网站下载所述链接对应的页面,提取下载的页面中正文,将提取的正文存储到存储模块中;

存储模块,适于存储提取的正文;

通信模块,适于接收客户端发送的加载关键字对应的链接的页面的请求,将存储模块中存储的该链接对应的正文发送给客户端,使客户端通过正文加载脚本将该正文添加到当前浏览的页面的正文中。

2. 根据权利要求 1 所述的服务器,其中,

所述脚本注入模块,还适于向发送给客户端的原始页面中注入内存释放脚本,以使客户端在接收到加载关键字对应的链接的指令后,判断当前浏览的页面所占内存是否超过预设阈值,如果超过,则释放该页面中之前浏览的正文所占内存。

3. 根据权利要求 1 或 2 所述的服务器,其中,

所述脚本注入模块,还适于向发送给客户端的原始页面中注入阅读标签功能脚本和/或页面选择功能脚本,以使客户端在当前浏览的页面中添加阅读标签功能键和/或选择页面功能键。

4. 根据权利要求 1 或 2 所述的服务器,其中,

所述正文提取模块,进一步适于根据页面所属网站中页面结构从下载的页面中提取正文。

5. 根据权利要求 1 或 2 所述的服务器,其中,

所述正文提取模块,进一步适于对提取的正文进行压缩处理,将生成的压缩包存储到存储模块中。

6. 根据权利要求 1 或 2 所述的服务器,其中,

所述服务器还包括:

页面预加载模块,适于提取客户端当前浏览的页面中非关键字对应的链接,从网站下载所述链接对应的页面,对下载的页面进行压缩处理,生成链接对应的页面压缩包,将该页面压缩包存储到存储模块,所述非关键字对应的链接为页面中除关键字对应的链接之外的链接;

所述存储模块,还适于存储页面压缩包;

所述通信模块,还适于接收客户端发送的加载当前浏览的页面中非关键字对应的链接的请求,将存储模块中该请求中链接对应的页面压缩包发送给客户端。

7. 根据权利要求 6 所述的服务器,其中,

所述服务器还包括:

存储更新模块,适于以发送给客户端的页面压缩包中页面为客户端当前浏览的页面,将该当前浏览的页面所包含的非关键字对应的链接与已存储的页面压缩包对应的链接比较,根据比较结果更新存储的页面压缩包。

8. 根据权利要求 7 所述的服务器,其中,

所述存储更新模块,进一步适于对于没有被该当前浏览的页面所包含的链接,将该链接对应的页面压缩包从存储模块中删除;对于该当前浏览的页面中所包含的、且对应的页面压缩包没有被存储的非关键字对应的链接,从网站下载该链接对应的页面,对下载的页面进行压缩处理,生成该链接对应的页面压缩包,将该页面压缩包存储到存储模块中。

9. 一种实现浏览器中页面加载的客户端,该客户端包括:

请求发送模块,适于在接收到加载原始网页的指令后,向服务器发送加载原始页面的请求,所述原始页面为初始访问网站时所浏览的页面,以及在接收到加载关键字对应的链接指令后,向服务器发送加载关键字对应的链接的页面的请求;

页面加载模块,适于接收服务器发送的注入有正文加载脚本的原始页面,加载该原始页面,以及接收服务器发送的关键字对应的链接的页面中正文,通过正文加载脚本将该正文添加到当前浏览的页面的正文中。

10. 根据权利要求 9 所述的客户端,其中,

所述原始页面中还注入有内存释放脚本;

所述页面加载模块,还适于在接收到加载关键字对应的链接的页面的指令后,通过内存释放脚本判断当前浏览的页面所占内存是否超过预设阈值,如果超过,则释放该页面中之前浏览的正文所占内存。

11. 根据权利要求 9 或 10 所述的客户端,其中,

所述原始页面中还注入有阅读标签功能脚本和 / 或页面选择功能脚本,

所述页面加载模块,还适于通过阅读标签功能脚本和 / 或页面选择功能脚本,在当前浏览的页面中添加阅读标签功能键和 / 或选择页面功能键。

12. 根据权利要求 9 或 10 所述的客户端,其中,

所述客户端还包括:解压模块,适于对页面压缩包进行解压缩;

所述请求发送模块,还适于在接收到加载非关键字对应的链接的指令后,向服务器发送加载当前浏览的页面中非关键字对应的链接的页面的请求;所述非关键字对应的链接为页面中除关键字对应的链接之外的链接;

所述页面加载模块,还适于接收服务器发送的指令中链接对应的页面压缩包,指示解压模块将该页面压缩包解压缩,加载该页面压缩包中页面。

13. 一种实现浏览器中页面加载的系统,该系统包括:

如权利要求 1 至 8 任一所述的服务器,以及

如权利要求 9 至 12 任一所述的客户端。

14. 一种实现浏览器中页面加载的方法,该方法包括:

在接收到客户端发送的加载原始页面的请求后,向发送给客户端的原始页面中注入正文加载脚本,所述原始页面为初始访问网站时所浏览的页面;

从客户端当前浏览的页面中解析出用于表示页面间正文的相互关系的关键字,提取关键字对应的链接,从网站下载所述链接对应的页面,提取下载的页面中正文,存储提取的正文;

当接收到客户端发送的加载关键字对应的链接的请求后,将提取的正文发送给客户端,使客户端通过正文加载脚本将该正文添加到当前浏览的页面的正文中。

15. 根据权利要求 14 所述的方法,其中,

所述方法还包括：

向发送给客户端的原始页面中注入内存释放脚本，以使客户端在接收到加载关键字对应的链接的指令后，判断当前浏览的页面所占内存是否超过预设阈值，如果超过，则释放该页面中之前浏览的正文所占内存。

16. 根据权利要求 14 或 15 所述的方法，其中，

所述方法还包括：

向发送给客户端的原始页面中注入阅读标签功能脚本和 / 或页面选择功能脚本，以使客户端在当前浏览的页面中添加阅读标签功能键和 / 或选择页面功能键。

17. 根据权利要求 14 或 15 所述的方法，其中，

所述提取下载的页面中正文进一步包括：

根据页面所属网站中页面结构从下载的页面中提取正文。

18. 根据权利要求 14 或 15 所述的方法，其中，

所述存储提取的正文进一步包括：

对提取的正文进行压缩处理，存储生成的压缩包。

19. 根据权利要求 14 或 15 所述的方法，其中，

所述方法还包括：

提取客户端当前浏览的页面中非关键字对应的链接，从网站下载所述链接对应的页面，对下载的页面进行压缩处理，生成链接对应的页面压缩包，存储该页面压缩包，所述非关键字对应的链接为除关键字对应的链接之外的链接；

当接收到客户端发送的加载当前浏览的页面中非关键字对应的链接的请求后，将该请求中链接对应的页面压缩包发送给客户端。

20. 根据权利要求 19 所述的方法，其中，

所述方法还包括：

以发送给客户端的页面压缩包中页面为客户端当前浏览的页面，将该当前浏览的页面所包含的非关键字对应的链接与已存储的页面压缩包对应的链接比较，根据比较结果更新存储的页面压缩包。

21. 根据权利要求 20 所述的方法，其中，

所述根据比较结果更新存储的页面压缩包进一步包括：

对于没有被该当前浏览的页面所包含的链接，将该链接对应的页面压缩包删除；

对于该当前浏览的页面中所包含的、且对应的页面压缩包没有被存储的非关键字对应的链接，从网站下载该链接对应的页面，对下载的页面进行压缩处理，生成该链接对应的页面压缩包，存储该页面压缩包。

实现浏览器中页面加载的服务器、客户端、系统和方法

技术领域

[0001] 本发明涉及网络数据通信领域，具体涉及一种实现浏览器中页面加载的服务器、客户端、系统和方法。

背景技术

[0002] 现有技术中，根据浏览器的浏览页面中链接，对用户之后的浏览行为进行预测，对于页面中所有可能在之后被点击的链接都进行预先下载，通过将这些链接对应的页面预先下载的方式提升浏览器中页面的加载速度。但是，页面中多数链接在之后的浏览中没有被点击，因此采用现有技术虽然可以加快浏览页面的速度，但是，在客户端产生大量无用的下载流量，浪费了客户端的网络流量。

发明内容

[0003] 鉴于上述问题，提出了本发明以便提供一种克服上述问题或者至少部分地解决上述问题的实现浏览器中页面加载的服务器、客户端、系统和方法。

[0004] 依据本发明的一个方面，提供了一种实现浏览器中页面加载的服务器，该服务器包括：

[0005] 脚本注入模块，适于在接收到客户端发送的加载原始页面的请求后，向发送给客户端的原始页面中注入正文加载脚本，所述原始页面为初始访问网站时所浏览的页面；

[0006] 正文提取模块，适于从客户端当前浏览的页面中解析出用于表示页面间正文的相互关系的关键字，提取关键字对应的链接，从网站下载所述链接对应的页面，提取下载的页面中正文，将提取的正文存储到存储模块中；

[0007] 存储模块，适于存储提取的正文；

[0008] 通信模块，适于接收客户端发送的加载关键字对应的链接的页面的请求，将存储模块中存储的该链接对应的正文发送给客户端，使客户端通过正文加载脚本将该正文添加到当前浏览的页面的正文中。

[0009] 可选地，所述脚本注入模块，还适于向发送给客户端的原始页面中注入内存释放脚本，以使客户端在接收到加载关键字对应的链接的指令后，判断当前浏览的页面所占内存是否超过预设阈值，如果超过，则释放该页面中之前浏览的正文所占内存。

[0010] 可选地，所述脚本注入模块，还适于向发送给客户端的原始页面中注入阅读标签功能脚本和 / 或页面选择功能脚本，以使客户端在当前浏览的页面中添加阅读标签功能键和 / 或选择页面功能键。

[0011] 可选地，所述正文提取模块，进一步适于根据页面所属网站中页面结构从下载的页面中提取正文。

[0012] 可选地，所述正文提取模块，进一步适于对提取的正文进行压缩处理，将生成的压缩包存储到存储模块中。

[0013] 可选地，所述服务器还包括：

[0014] 页面预加载模块,适于提取客户端当前浏览的页面中非关键字对应的链接,从网站下载所述链接对应的页面,对下载的页面进行压缩处理,生成链接对应的页面压缩包,将该页面压缩包存储到存储模块,所述非关键字对应的链接为页面中除关键字对应的链接之外的链接;

[0015] 所述存储模块,还适于存储页面压缩包;

[0016] 所述通信模块,还适于接收客户端发送的加载当前浏览的页面中非关键字对应的链接的请求,将存储模块中该请求中链接对应的页面压缩包发送给客户端。

[0017] 可选地,所述服务器还包括:

[0018] 存储更新模块,适于以发送给客户端的页面压缩包中页面为客户端当前浏览的页面,将该当前浏览的页面所包含的非关键字对应的链接与已存储的页面压缩包对应的链接比较,根据比较结果更新存储的页面压缩包。

[0019] 可选地,所述存储更新模块,进一步适于对于没有被该当前浏览的页面所包含的链接,将该链接对应的页面压缩包从存储模块中删除;对于该当前浏览的页面中所包含的、且对应的页面压缩包没有被存储的非关键字对应的链接,从网站下载该链接对应的页面,对下载的页面进行压缩处理,生成该链接对应的页面压缩包,将该页面压缩包存储到存储模块中。

[0020] 根据本发明的一个方面,提供了一种实现浏览器中页面加载的客户端,该客户端包括:

[0021] 请求发送模块,适于在接收到加载原始网页的指令后,向服务器发送加载原始页面的请求,所述原始页面为初始访问网站时所浏览的页面,以及在接收到加载关键字对应的链接指令后,向服务器发送加载关键字对应的链接的页面的请求;

[0022] 页面加载模块,适于接收服务器发送的注入有正文加载脚本的原始页面,加载该原始页面,以及接收服务器发送的关键字对应的链接的页面中正文,通过正文加载脚本将该正文添加到当前浏览的页面的正文中。

[0023] 可选地,所述原始页面中还注入有内存释放脚本;

[0024] 所述页面加载模块,还适于在接收到加载关键字对应的链接的页面的指令后,通过内存释放脚本判断当前浏览的页面所占内存是否超过预设阈值,如果超过,则释放该页面中之前浏览的正文所占内存。

[0025] 可选地,所述原始页面中还注入有阅读标签功能脚本和/或页面选择功能脚本,

[0026] 所述页面加载模块,还适于通过阅读标签功能脚本和/或页面选择功能脚本,在当前浏览的页面中添加阅读标签功能键和/或选择页面功能键。

[0027] 可选地,所述客户端还包括:解压模块,适于对页面压缩包进行解压缩;

[0028] 所述请求发送模块,还适于在接收到加载非关键字对应的链接的指令后,向服务器发送加载当前浏览的页面中非关键字对应的链接的页面的请求;所述非关键字对应的链接为页面中除关键字对应的链接之外的链接;

[0029] 所述页面加载模块,还适于接收服务器发送的指令中链接对应的页面压缩包,指示解压模块将该页面压缩包解压缩,加载该页面压缩包中页面。

[0030] 根据本发明的另一个方面,提供了一种实现浏览器中页面加载的系统,该系统包括所述的服务器,以及所述的客户端。

[0031] 根据本发明的另一个方面,提供了一种实现浏览器中页面加载的方法,该方法包括:

[0032] 在接收到客户端发送的加载原始页面的请求后,向发送给客户端的原始页面中注入正文加载脚本,所述原始页面为初始访问网站时所浏览的页面;

[0033] 从客户端当前浏览的页面中解析出用于表示页面间正文的相互关系的关键字,提取关键字对应的链接,从网站下载所述链接对应的页面,提取下载的页面中正文,存储提取的正文;

[0034] 当接收到客户端发送的加载关键字对应的链接的请求后,将提取的正文发送给客户端,使客户端通过正文加载脚本将该正文添加到当前浏览的页面的正文中。

[0035] 可选地,所述方法还包括:

[0036] 向发送给客户端的原始页面中注入内存释放脚本,以使客户端在接收到加载关键字对应的链接的指令后,判断当前浏览的页面所占内存是否超过预设阈值,如果超过,则释放该页面中之前浏览的正文所占内存。

[0037] 可选地,所述方法还包括:

[0038] 向发送给客户端的原始页面中注入阅读标签功能脚本和/或页面选择功能脚本,以使客户端在当前浏览的页面中添加阅读标签功能键和/或选择页面功能键。

[0039] 可选地,所述提取下载的页面中正文进一步包括:

[0040] 根据页面所属网站中页面结构从下载的页面中提取正文。

[0041] 可选地,所述存储提取的正文进一步包括:

[0042] 对提取的正文进行压缩处理,存储生成的压缩包。

[0043] 可选地,所述方法还包括:

[0044] 提取客户端当前浏览的页面中非关键字对应的链接,从网站下载所述链接对应的页面,对下载的页面进行压缩处理,生成链接对应的页面压缩包,存储该页面压缩包,所述非关键字对应的链接为除关键字对应的链接之外的链接;

[0045] 当接收到客户端发送的加载当前浏览的页面中非关键字对应的链接的请求后,将该请求中链接对应的页面压缩包发送给客户端。

[0046] 可选地,所述方法还包括:

[0047] 以发送给客户端的页面压缩包中页面为客户端当前浏览的页面,将该当前浏览的页面所包含的非关键字对应的链接与已存储的页面压缩包对应的链接比较,根据比较结果更新存储的页面压缩包。

[0048] 可选地,所述根据比较结果更新存储的页面压缩包进一步包括:

[0049] 对于没有被该当前浏览的页面所包含的链接,将该链接对应的页面压缩包删除;

[0050] 对于该当前浏览的页面中所包含的、且对应的页面压缩包没有被存储的非关键字对应的链接,从网站下载该链接对应的页面,对下载的页面进行压缩处理,生成该链接对应的页面压缩包,存储该页面压缩包。

[0051] 根据本发明的技术方案,在接收到客户端发送的加载原始页面的请求后,向发送给客户端的原始页面中注入正文加载脚本,从客户端当前浏览的页面中解析出用于表示页面间正文的相互关系的关键字,提取关键字对应的链接,从网站下载所述链接对应的页面,提取下载的页面中正文,存储提取的正文;当接收到客户端发送的加载关键字对应的链接

的请求后,将提取的正文发送给客户端,使客户端通过正文加载脚本将该正文添加到当前浏览的页面的正文中。

[0052] 采用本发明中技术方案,对于关键字连接对应的页面,仅提取其中正文,将正文下发给客户端,通过在原始页面中注入正文加载脚本的方式,使客户端将该正文添加到当前浏览的页面中。由于下发的正文已经预先提取,因此可以提高浏览页面的速度,同时下发内容仅为正文内容,减少了下发的数据量,减少客户端浏览网页所用网络流量。由此解决了在客户端造成大量无用的下载流量,浪费网络流量的问题,取得了既能够加快客户端中浏览速度又能够节约客户端的网络流量的有益效果。

[0053] 上述说明仅是本发明技术方案的概述,为了能够更清楚了解本发明的技术手段,而可依照说明书的内容予以实施,并且为了让本发明的上述和其它目的、特征和优点能够更明显易懂,以下特举本发明的具体实施方式。

附图说明

[0054] 通过阅读下文优选实施方式的详细描述,各种其他的优点和益处对于本领域普通技术人员将变得清楚明了。附图仅用于示出优选实施方式的目的,而并不认为是对本发明的限制。而且在整个附图中,用相同的参考符号表示相同的部件。在附图中:

[0055] 图 1 示出了根据本发明一个实施例的实现浏览器中页面加载的服务器的结构图;

[0056] 图 2 示出了根据本发明一个实施例的客户端预浏览的三个相关网页的示意图;

[0057] 图 3 示出了根据本发明一个实施例的客户端中完成正文添加后的完整页面的示意图;

[0058] 图 4 示出了根据本发明一个实施例的实现浏览器中页面加载的服务器的结构图;

[0059] 图 5 示出了根据本发明一个实施例的客户端当前浏览的页面的示意图;

[0060] 图 6 示出了根据本发明一个实施例的实现浏览器中页面加载的服务器的结构图;

[0061] 图 7 示出了根据本发明一个实施例的客户端链接点击后更新的页面的示意图;

[0062] 图 8 示出了根据本发明一个实施例的实现浏览器中页面加载的客户端的结构图;

[0063] 图 9 示出了根据本发明一个实施例的实现浏览器中页面加载的客户端的结构图;

[0064] 图 10 示出了根据本发明一个实施例的实现浏览器中页面加载的系统的结构图;

[0065] 图 11 示出了根据本发明一个实施例的实现浏览器中页面加载的方法的流程图;以及

[0066] 图 12 示出了根据本发明一个实施例的对于页面中非关键字对应的链接处理的方法的流程图。

具体实施方式

[0067] 下面将参照附图更详细地描述本公开的示例性实施例。虽然附图中显示了本公开的示例性实施例,然而应当理解,可以以各种形式实现本公开而不应被这里阐述的实施例所限制。相反,提供这些实施例是为了能够更透彻地理解本公开,并且能够将本公开的范围完整的传达给本领域的技术人员。

[0068] 参见图 1,示出了根据本发明一个实施例的实现浏览器中页面加载的服务器的结构图。服务器 100 包括:脚本注入模块 110、正文提取模块 120、存储模块 140 和通信模块

130。

[0069] 脚本注入模块 110,适于在接收到客户端 200 发送的加载原始页面的请求后,向发送给客户端 200 的原始页面中注入正文加载脚本。所述原始页面为初始访问网站时所浏览的页面。

[0070] 正文提取模块 120,适于从客户端 200 当前浏览的页面中解析出用于表示页面间正文的相互关系的关键字,提取关键字对应的链接,从网站下载该链接对应的页面,提取下载的页面中正文,将提取的正文存储到存储模块 140 中。

[0071] 存储模块 140,适于存储提取的正文。

[0072] 通信模块 130,适于接收客户端 200 发送的加载关键字对应的链接的页面的请求,将存储模块 140 中存储的该链接对应的正文发送给客户端 200,使客户端 200 通过正文加载脚本将该正文添加到当前浏览的页面的正文中。

[0073] 为进一步节约网络流量,正文提取模块 120 可以对提取的正文进行压缩处理,将生成的压缩包存储到存储模块 140 中。通信模块 130 在接收到客户端 200 发送的加载关键字对应的链接的页面的请求后,将存储模块 140 中存储的该链接对应的正文的压缩包发送给客户端 200。

[0074] 图 2 示出了根据本发明一个实施例的客户端预浏览的三个相关网页的示意图。图 2 中三个网页为一篇文章的第 1 页、第 2 页、第 3 页,三页中正文内容合并为一篇完整的文章。图 2 中方框所框部分为关键字。

[0075] 关键字用于表示页面间正文的相互关系,例如,关键字可以为“上一页”、“下一页”、“第 X 页”等。图 2 中标出的关键字仅为示例性说明,本发明中关键字不限于此。

[0076] 首先,脚本注入模块 110 在接收到客户端 200 发送的加载原始页面,即第 1 页,的请求后,向发送给客户端 200 的原始页面中注入正文加载脚本。客户端 200 可以通过该正文加载脚本将接收的正文添加到当前浏览的页面的正文中。

[0077] 正文提取模块 120 对于客户端 200 当前浏览的第 1 页中关键字“下一页”对应的链接,从网站中下载该链接对应的页面,提取该链接对应的页面中正文,即第 2 页中正文。具体地,正文提取模块 120 通过在第 2 页的 html 中查询,发现第三个<div>标签对<div class="w1 Text">...</div>中具有“Text”,则确定第三个<div>标签对中包含的为正文,提取该正文。

[0078] 当完成第 1 页浏览后,向下滑动页面,客户端 200 接收到加载当前浏览的页面中关键字“下一页”对应链接的页面的指令,发送加载关键字“下一页”对应的链接的页面的请求。通信模块 130 接收请求,将链接“下一页”,即第 2 页,对应的页面中正文发送给客户端 200。客户端 200 通过正文加载脚本将正文添加到当前浏览的页面的正文中。并且,客户端 200 接收到第 2 页正文后,自动将“下一页”对应链接中页码加 1,使得“下一页”对应第 3 页。当完成第 2 页浏览后,向下滑动,客户端 200 从服务器 100 获得第 3 页中正文,通过正文加载脚本将第 3 页中正文添加到当前浏览的页面的正文中。客户端 200 中完成正文添加后完整页面如图 3 所示。

[0079] 进一步,正文提取模块 120 根据页面所属网站中页面结构从下载的页面中提取正文。以图 2 所示的三个页面为例,经分析可知,三个页面的 html 中都包含“Text”,并且三个页面中包括“上一页”或“下一页”。由此可以确定,三个页面间具有关联,并且都包含有

正文。如上所述,正文提取模块 120 通过在页面的 html 中查询,发现第三个 <div> 标签对 <div class="w1Text">...</div> 中具有“Text”,则确定第三个 <div> 标签对中包含的为正文,提取该正文。

[0080] 进一步地,脚本注入模块 110,还适于向发送给客户端 200 的原始页面中注入内存释放脚本,以使客户端 200 在接收到加载关键字对应的链接的指令后,判断当前浏览的页面所占内存是否超过预设阈值,如果超过,则释放该页面中之前浏览的正文所占内存。

[0081] 例如,内存释放脚本中内容为:

[0082] parent.removeChild(e);

[0083] e=null;

[0084] delete e

[0085] 当完成第 2 页浏览后,向下滑动页面,客户端 200 接收到加载关键字“下一页”对应链接的页面的指令。客户端 200 判断当前浏览的页面所占内存是否超过预设阈值,如果超过,则释放第 1 页和 / 或第 2 页中正文所占内存。

[0086] 由上可见,通过向发送给客户端的原始页面中注入内存释放脚本,解决由于添加的正文过多导致内存溢出的问题,取得了防止内存溢出的有益效果。

[0087] 进一步地,脚本注入模块 110,还适于向发送给客户端 200 的原始页面中注入阅读标签功能脚本和 / 或页面选择功能脚本,以使客户端 200 在当前浏览的页面中添加阅读标签功能键和 / 或选择页面功能键。

[0088] 在此,通过向原始页面中注入功能脚本的方式,可以在客户端 200 的页面中添加功能键,由此达到扩展页面中功能的有益效果。

[0089] 参见图 4,示出了根据本发明一个实施例的实现浏览器中页面加载的服务器的结构图。该实施例中服务器 100 除了包括脚本注入模块 110、正文提取模块 120、存储模块 140 和通信模块 130 外,还包括页面预加载模块 150。

[0090] 页面预加载模块 150,适于提取客户端 200 当前浏览的页面中非关键字对应的链接,从网站下载该链接对应的页面,对下载的页面进行压缩处理,生成链接对应的页面压缩包,将该页面压缩包存储到存储模块 140。

[0091] 所述非关键字对应的链接为页面中除关键字对应的链接之外的链接;

[0092] 存储模块 140,还适于存储页面压缩包。

[0093] 通信模块 130,还适于接收客户端 200 发送的加载当前浏览的页面中非关键字对应的链接的请求,将存储模块 140 中该请求中链接对应的页面压缩包发送给客户端 200。

[0094] 参见图 5,示出了根据本发明一个实施例的客户端当前浏览的页面的示意图。以图 5 中所示页面为例,对实施例进行举例说明。

[0095] 在该页面包含有链接:“登录”、“注册”、“首页”、“科技”、“IT”、“互联网”、“国际”、“上一页”等。其中,“上一页”为关键字对应的链接,其余的为非关键字对应的链接。页面预加载模块 150 提取该页面中所包含非关键字对应的链接。由于客户端 200 从服务器 100 获得浏览的页面,因此服务器 100 能够获知客户端 200 当前浏览的页面。页面预加载模块 150 从各个网站下载各个非关键字对应链接对应的页面。页面预加载模块 150 对下载的页面进行压缩处理,生成链接对应的页面压缩包,将页面压缩包存储到存储模块 140。存储模块 140 可以为服务器 100 的内存,其中存储链接对应的页面压缩包。当页面中某个非关键字

对应的链接被点击后,客户端 200 就会发送加载该被点击的链接的页面的请求。通信模块 130 接收客户端 200 发送的请求,并将该请求中链接对应的页面压缩包发送给客户端 200。由此,客户端 200 获得点击的链接对应的页面压缩包,将该压缩包解压缩后在客户端 200 显示该页面。

[0096] 本实施例中服务器可以为云服务器,该云服务器对各个客户端浏览的页面中链接对应的页面进行统一的下载和保存,并将该页面压缩成压缩包。当客户端中因为链接点击等原因,实际产生下载需要时,将对应的压缩包下发给客户端。由此,能够向客户端提供非关键字对应的链接的压缩数据包,提高浏览非关键字对应的链接的页面的速度,取得了既能够加快客户端中浏览速度又能够节约客户端的网络流量,减少客户端的工作负担的有益效果。

[0097] 参见图 6,示出了根据本发明一个实施例的实现浏览器中页面加载的服务器的结构图。该实施例中服务器 100 除了包括脚本注入模块 110、正文提取模块 120、存储模块 140、通信模块 130 和页面预加载模块 150 外,还包括存储更新模块 160。

[0098] 存储更新模块 160,适于以发送给客户端 200 的页面压缩包中页面为客户端 200 当前浏览的页面,将该当前浏览的页面所包含的非关键字对应的链接与已存储的页面压缩包对应的链接比较,根据比较结果更新存储的页面压缩包。

[0099] 进一步地,存储更新模块 160 对于没有被该当前浏览的页面所包含的链接,将该链接对应的页面压缩包从存储模块 140 中删除;对于该当前浏览的页面中所包含的、且对应的页面压缩包没有被存储的非关键字对应的链接,从网站下载该链接对应的页面,对下载的页面进行压缩处理,生成该链接对应的页面压缩包,将该页面压缩包存储到存储模块 140 中。

[0100] 参见图 7,示出了根据本发明一个实施例的客户端链接点击后更新的页面。以图 7 所示的在客户端链接点击图 5 中链接“国际”后更新的页面为例,进行示例性说明。该页面中包含链接:“登录”、“注册”、“首页”、“科技”、“IT”、“互联网”、“国际”,以及新闻条目链接,例如:“百度下周股价走高的 4 个理由”、“亚马逊称 Kindle Fire 创纪录”等。

[0101] 存储更新模块 160 将该页面所包含的链接与已存储的页面压缩包对应的链接比较。其中,“登录”、“注册”、“首页”、“科技”、“IT”、“互联网”、“国际”被该更新后的页面所包含,且这些链接对应的页面压缩包已经被存储,所以这些链接对应的页面压缩包仍然被存储,没有变化。该更新后的页面中没有包含链接“上一页”,因此,将“上一页”对应的正文或正文压缩包从存储模块 140 中删除。该更新后的页面中包含的新闻条目链接对应的页面压缩包没有被存储模块 140 存储,因此从网站下载新闻条目链接对应的页面,对下载的页面进行压缩处理,生成该链接对应的页面压缩包,将该页面压缩包存储到存储模块 140 中。

[0102] 由上可见,通过在客户端中浏览的页面变化时,对应更新存储的链接对应的页面压缩包,使得已存储的页面压缩包中有用的页面压缩包被保留,仅增加变化后页面中新增链接对应的页面压缩包,因此能够进一步减少服务器端的网络流量,以及服务器进行页面压缩的工作负担。

[0103] 图 8 示出了根据本发明一个实施例的实现浏览器中页面加载的客户端的结构图。该客户端 200 包括:请求发送模块 210 和页面加载模块 220。

[0104] 请求发送模块 210,适于在接收到加载原始网页的指令后,向服务器 100 发送加载

原始页面的请求,所述原始页面为初始访问网站时所浏览的页面,以及在接收到加载关键字对应的链接指令后,向服务器 100 发送加载关键字对应的链接的页面的请求。

[0105] 页面加载模块 220,适于接收服务器 100 发送的注入有正文加载脚本的原始页面,加载该原始页面,以及接收服务器 100 发送的关键字对应的链接的页面中正文,通过正文加载脚本将该正文添加到当前浏览的页面的正文中。

[0106] 由于下发的正文已经预先提取,因此可以提高浏览页面的速度,同时下发内容仅为正文内容,减少了下发的数据量,减少客户端浏览网页所用网络流量。

[0107] 进一步地,原始页面中还注入有内存释放脚本。

[0108] 页面加载模块 220,还适于在接收到加载关键字对应的链接的页面的指令后,通过内存释放脚本判断当前浏览的页面所占内存是否超过预设阈值,如果超过,则释放该页面中之前浏览的正文所占内存。

[0109] 通过执行原始页面中注入内存释放脚本,解决由于添加的正文过多导致内存溢出的问题,达到了防止内存溢出的有益效果。

[0110] 进一步地,原始页面中还注入有阅读标签功能脚本和 / 或页面选择功能脚本。

[0111] 页面加载模块 220,还适于通过阅读标签功能脚本和 / 或页面选择功能脚本,在当前浏览的页面中添加阅读标签功能键和 / 或选择页面功能键。

[0112] 通过向原始页面中注入功能脚本的方式,可以在客户端的页面中添加功能键,由此达到扩展页面中功能的有益效果。

[0113] 图 9 示出了根据本发明一个实施例的实现浏览器中页面加载的客户端的结构图。该客户端 200 除了包括请求发送模块 210 和页面加载模块 220 为,还包括解压模块 230。

[0114] 解压模块 230,适于对页面压缩包进行解压缩;

[0115] 请求发送模块 210,还适于在接收到加载非关键字对应的链接的指令后,向服务器 100 发送加载当前浏览的页面中非关键字对应的链接的页面的请求;所述非关键字对应的链接为页面中除关键字对应的链接之外的链接;

[0116] 页面加载模块 220,还适于接收服务器 100 发送的指令中链接对应的页面压缩包,指示解压模块 230 将该页面压缩包解压缩,加载该页面压缩包中页面。

[0117] 实施例中,客户端仅接收要浏览的链接对应的页面压缩包,不需要将页面中所有链接的页面都做预下载,另一方面,在服务器侧已经预先生成了页面压缩包,通过从服务器侧下载该页面压缩包,能够提高浏览速度。因此,进一步取得了既能够加快客户端中浏览速度又能够节约客户端的网络流量,减少客户端的工作负担的有益效果。

[0118] 参见图 10,示出了根据本发明一个实施例的实现浏览器中页面加载的系统的结构图。系统包括前述服务器 100,以及多个前述客户端 200。服务器 100 和客户端 200 间可以直接连接也可以通过网络连接,在此没有特别限制。

[0119] 参见图 11,示出了根据本发明一个实施例的实现浏览器中页面加载的方法的流程图。该方法包括如下步骤。

[0120] 步骤 S1110,在接收到客户端发送的加载原始页面的请求后,向发送给客户端的原始页面中注入正文加载脚本。原始页面为初始访问网站时所浏览的页面;

[0121] 步骤 S1120,从客户端当前浏览的页面中解析出用于表示页面间正文的相互关系的关键字,提取关键字对应的链接,从网站下载所述链接对应的页面,提取下载的页面中正

文,存储提取的正文。

[0122] 步骤 S1130,当接收到客户端发送的加载关键字对应的链接的请求后,将提取的正文发送给客户端,使客户端通过正文加载脚本将该正文添加到当前浏览的页面的正文中。

[0123] 为进一步节约网络流量,在步骤 S1120 中,可以对提取的正文进行压缩处理,存储生成的压缩包。步骤 S1130 中,接收到客户端发送的加载关键字对应的链接的页面的请求后,将存储的该链接对应的正文的压缩包发送给客户端。

[0124] 以图 2 中所示的客户端预浏览的三个相关网页为例,进行示例性说明。图 2 中三个网页为一篇文章的第 1 页、第 2 页、第 3 页,三页中正文内容合并为一篇完整的文章。图 2 中方框所框部分为关键字。

[0125] 关键字用于表示页面间正文的相互关系,例如,关键字可以为“上一页”、“下一页”、“第 X 页”等。图 2 中标出的关键字仅为示例性说明,本发明中关键字不限于此。

[0126] 首先,在步骤 S1110 中,在接收到客户端发送的加载原始页面,即第 1 页,的请求后,向发送给客户端的原始页面中注入正文加载脚本。客户端可以通过该正文加载脚本将接收的正文添加到当前浏览的页面的正文中。

[0127] 在步骤 S1120 中,对于客户端 200 当前浏览的第 1 页中关键字“下一页”对应的链接,从网站中下载该链接对应的页面,提取该链接对应的页面中正文,即第 2 页中正文。具体地,通过在第 2 页的 html 中查询,发现第三个 <div> 标签对 <div class="w1 Text">...</div> 中具有“Text”,则确定第三个 <div> 标签对中包含的为正文,提取该正文。

[0128] 当完成第 1 页浏览后,向下滑动页面,客户端接收到加载当前浏览的页面中关键字“下一页”对应链接的页面的指令,发送加载关键字“下一页”对应的链接的页面的请求。在步骤 S1130 中,接收请求,将链接“下一页”,即第 2 页,对应的页面中正文发送给客户端。客户端通过正文加载脚本将正文添加到当前浏览的页面的正文中。

[0129] 并且,客户端接收到第 2 页正文后,自动将“下一页”对应链接中页码加 1,使得“下一页”对应第 3 页。当完成第 2 页浏览后,向下滑动,客户端获得第 3 页中正文,通过正文加载脚本将第 3 页中正文添加到当前浏览的页面的正文中。客户端中完成正文添加后完整页面如图 3 所示。

[0130] 进一步,在步骤 S1120 中根据页面所属网站中页面结构从下载的页面中提取正文。以图 2 所示的三个页面为例,经分析可知,三个页面的 html 中都包含“Text”,并且三个页面中包括“上一页”或“下一页”。由此可以确定,三个页面间具有关联,并且都包含有正文。如上所述,通过在页面的 html 中查询,发现第三个 <div> 标签对 <div class="w1 Text">...</div> 中具有“Text”,则确定第三个 <div> 标签对中包含的为正文,提取该正文。

[0131] 进一步地,方法还包括:向发送给客户端的原始页面中注入内存释放脚本,以使客户端在接收到加载关键字对应的链接的指令后,判断当前浏览的页面所占内存是否超过预设阈值,如果超过,则释放该页面中之前浏览的正文所占内存。

[0132] 例如,内存释放脚本中内容为:

[0133] parent.removeChild(e);

[0134] e=null;

[0135] delete e

[0136] 当完成第 2 页浏览后,向下滑动页面,客户端接收到加载关键字“下一页”对应链

接的页面的指令。客户端判断当前浏览的页面所占内存是否超过预设阈值,如果超过,则释放第 1 页和 / 或第 2 页中正文所占内存。

[0137] 由上可见,通过向发送给客户端的原始页面中注入内存释放脚本,解决由于添加的正文过多导致内存溢出的问题,取得了防止内存溢出的有益效果。

[0138] 进一步地,方法还包括:向发送给客户端的原始页面中注入阅读标签功能脚本和 / 或页面选择功能脚本,以使客户端在当前浏览的页面中添加阅读标签功能键和 / 或选择页面功能键。

[0139] 在此,通过向原始页面中注入功能脚本的方式,可以在客户端的页面中添加功能键,由此达到扩展页面中功能的有益效果。

[0140] 参见图 12,示出了根据本发明一个实施例的对于页面中非关键字对应的链接处理的方法的流程图。在该实施例中,方法除包括步骤 S1110 至步骤 S1130 外,还包括如下步骤。

[0141] 步骤 S1210,提取客户端当前浏览的页面中非关键字对应的链接,从网站下载所述链接对应的页面。所述非关键字对应的链接为除关键字对应的链接之外的链接。

[0142] 步骤 S1220,对下载的页面进行压缩处理,生成链接对应的页面压缩包,存储该页面压缩包。

[0143] 步骤 S1230,当接收到客户端发送的加载当前浏览的页面中非关键字对应的链接的请求后,将该请求中链接对应的页面压缩包发送给客户端。

[0144] 以图 5 中所示页面为例,对实施例进行举例说明。

[0145] 在该页面包含有链接:“登录”、“注册”、“首页”、“科技”、“IT”、“互联网”、“国际”、“上一页”等。其中,“上一页”为关键字对应的链接,其余的为非关键字对应的链接。步骤 S1210 中,提取该页面中所包含非关键字对应的链接。由于客户端从服务器获得浏览的页面,因此服务器能够获知客户端当前浏览的页面。并且,从各个网站下载各个非关键字对应链接对应的页面。步骤 S1220 中,对下载的页面进行压缩处理,生成链接对应的页面压缩包,存储页面压缩包。当页面中某个非关键字对应的链接被点击后,客户端就会发送加载该被点击的链接的页面的请求。步骤 S1230 中,接收客户端发送的请求,并将该请求中链接对应的页面压缩包发送给客户端。由此,客户端获得点击的链接对应的页面压缩包,将该压缩包解压缩后在客户端 2 示该页面。

[0146] 本实施例中对各个客户端浏览的页面中链接对应的页面进行统一的下载和保存,并将该页面压缩成压缩包。当客户端中因为链接点击等原因,实际产生下载需要时,将对应的压缩包下发给客户端。由此,能够向客户端提供非关键字对应的链接的压缩数据包,提高浏览非关键字对应的链接的页面的速度,取得了既能够加快客户端中浏览速度又能够节约客户端的网络流量,减少客户端的工作负担的有益效果。

[0147] 在一实施例中,方法还包括:以发送给客户端的页面压缩包中页面为客户端当前浏览的页面,将该当前浏览的页面所包含的非关键字对应的链接与已存储的页面压缩包对应的链接比较,根据比较结果更新存储的页面压缩包。

[0148] 进一步地,对于没有被该当前浏览的页面所包含的链接,将该链接对应的页面压缩包删除;

[0149] 对于该当前浏览的页面中所包含的、且对应的页面压缩包没有被存储的非关键字对应的链接,从网站下载该链接对应的页面,对下载的页面进行压缩处理,生成该链接对应

的页面压缩包,存储该页面压缩包。

[0150] 以图 7 所示的在客户端链接点击图 5 中链接“国际”后更新的页面为例,进行示例性说明。该页面中包含链接:“登录”、“注册”、“首页”、“科技”、“IT”、“互联网”、“国际”,以及新闻条目链接,例如:“百度下周股价走高的 4 个理由”、“亚马逊称 Kindle Fire 创纪录”等。

[0151] 将该页面所包含的链接与已存储的页面压缩包对应的链接比较。其中,“登录”、“注册”、“首页”、“科技”、“IT”、“互联网”、“国际”被该更新后的页面所包含,且这些链接对应的页面压缩包已经被存储,所以这些链接对应的页面压缩包仍然被存储,没有变化。该更新后的页面中没有包含链接“上一页”,因此,将“上一页”对应的正文或正文压缩包删除。该更新后的页面中包含的新闻条目链接对应的页面压缩包没有被存储,因此从网站下载新闻条目链接对应的页面,对下载的页面进行压缩处理,生成该链接对应的页面压缩包,存储该页面压缩包。

[0152] 由上可见,通过在客户端中浏览的页面变化时,对应更新存储的链接对应的页面压缩包,使得已存储的页面压缩包中有用的页面压缩包被保留,仅增加变化后页面中新增链接对应的页面压缩包,因此能够进一步减少服务器端的网络流量,以及服务器进行页面压缩的工作负担。

[0153] 在此提供的算法和显示不与任何特定计算机、虚拟系统或者其它设备固有相关。各种通用系统也可以与基于在此的示教一起使用。根据上面的描述,构造这类系统所要求的结构是显而易见的。此外,本发明也不针对任何特定编程语言。应当明白,可以利用各种编程语言实现在此描述的本发明的内容,并且上面对特定语言所做的描述是为了披露本发明的最佳实施方式。

[0154] 在此处所提供的说明书中,说明了大量具体细节。然而,能够理解,本发明的实施例可以在没有这些具体细节的情况下实践。在一些实例中,并未详细示出公知的方法、结构和技术,以便不模糊对本说明书的理解。

[0155] 类似地,应当理解,为了精简本公开并帮助理解各个发明方面中的一个或多个,在上面对本发明的示例性实施例的描述中,本发明的各个特征有时被一起分组到单个实施例、图、或者对其的描述中。然而,并不应将该公开的方法解释成反映如下意图:即所要求保护的本发明要求比在每个权利要求中所明确记载的特征更多的特征。更确切地说,如下面的权利要求书所反映的那样,发明方面在于少于前面公开的单个实施例的所有特征。因此,遵循具体实施方式的权利要求书由此明确地并入该具体实施方式,其中每个权利要求本身都作为本发明的单独实施例。

[0156] 本领域那些技术人员可以理解,可以对实施例中的设备中的模块进行自适应性地改变并且把它们设置在与该实施例不同的一个或多个设备中。可以把实施例中的模块或单元或组件组合成一个模块或单元或组件,以及此外可以把它分成多个子模块或子单元或子组件。除了这样的特征和/或过程或者单元中的至少一些是相互排斥之外,可以采用任何组合对本说明书(包括伴随的权利要求、摘要和附图)中公开的所有特征以及如此公开的任何方法或者设备的所有过程或单元进行组合。除非另外明确陈述,本说明书(包括伴随的权利要求、摘要和附图)中公开的每个特征可以由提供相同、等同或相似目的的替代特征来代替。

[0157] 此外,本领域的技术人员能够理解,尽管在此所述的一些实施例包括其它实施例

中所包括的某些特征而不是其它特征,但是不同实施例的特征的组合意味着处于本发明的范围之内并且形成不同的实施例。例如,在下面的权利要求书中,所要求保护的实施例的任意之一都可以以任意的组合方式来使用。

[0158] 应该注意的是上述实施例对本发明进行说明而不是对本发明进行限制,并且本领域技术人员在不脱离所附权利要求的范围的情况下可设计出替换实施例。在权利要求中,不应将位于括号之间的任何参考符号构造成对权利要求的限制。单词“包含”不排除存在未列在权利要求中的元件或步骤。位于元件之前的单词“一”或“一个”不排除存在多个这样的元件。本发明可以借助于包括有若干不同元件的硬件以及借助于适当编程的计算机来实现。在列举了若干装置的单元权利要求中,这些装置中的若干个可以通过同一个硬件项来具体体现。单词第一、第二、以及第三等的使用不表示任何顺序。可将这些单词解释为名称。

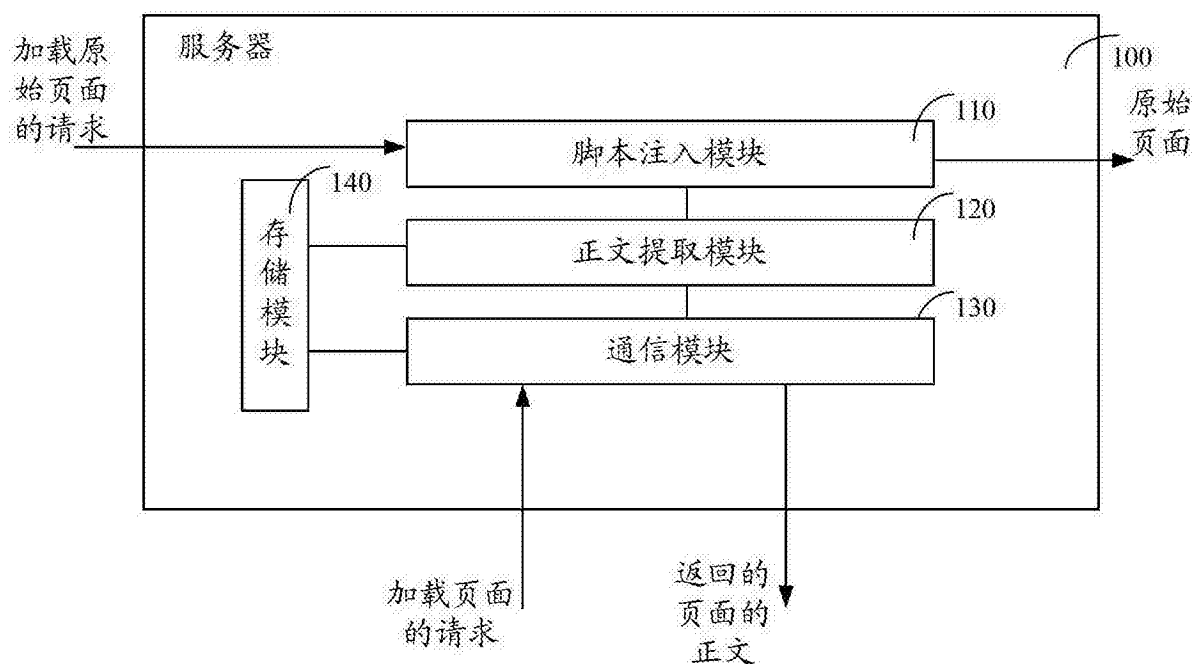


图 1

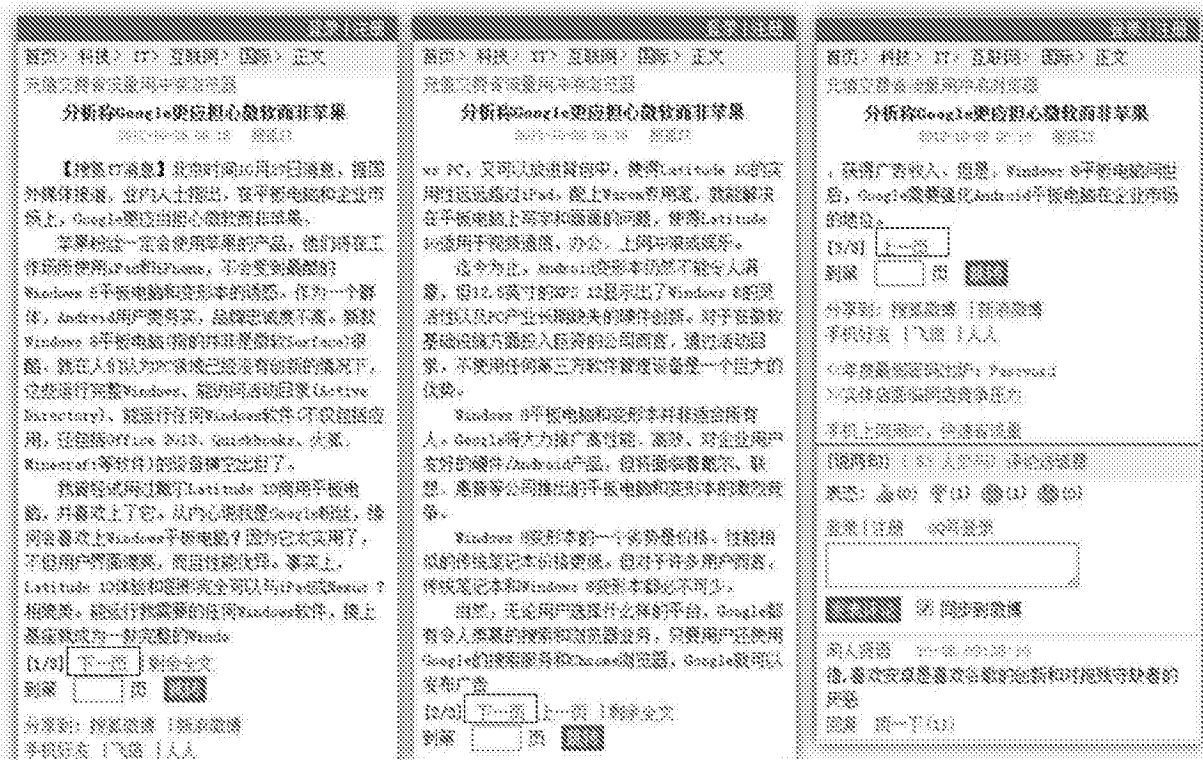


图 2

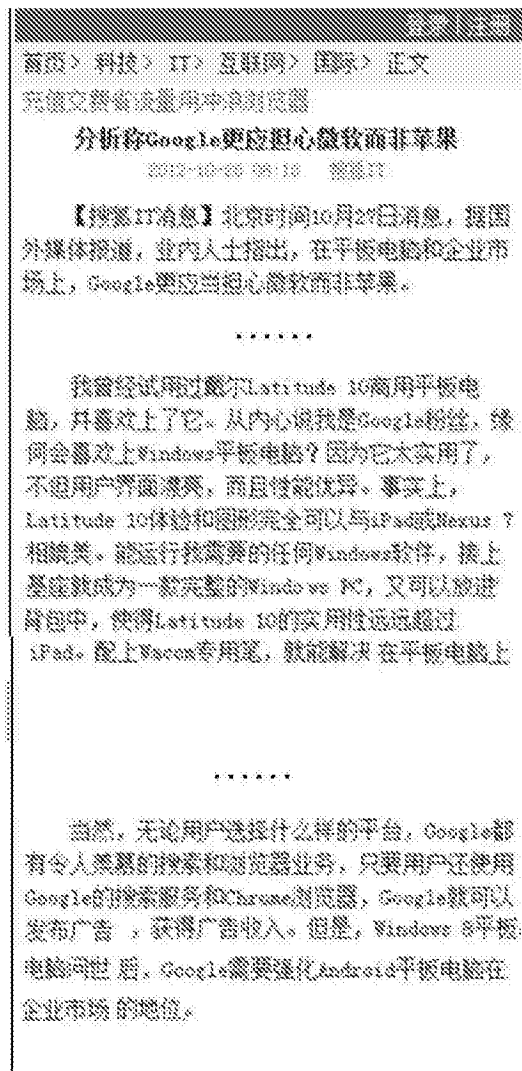


图 3

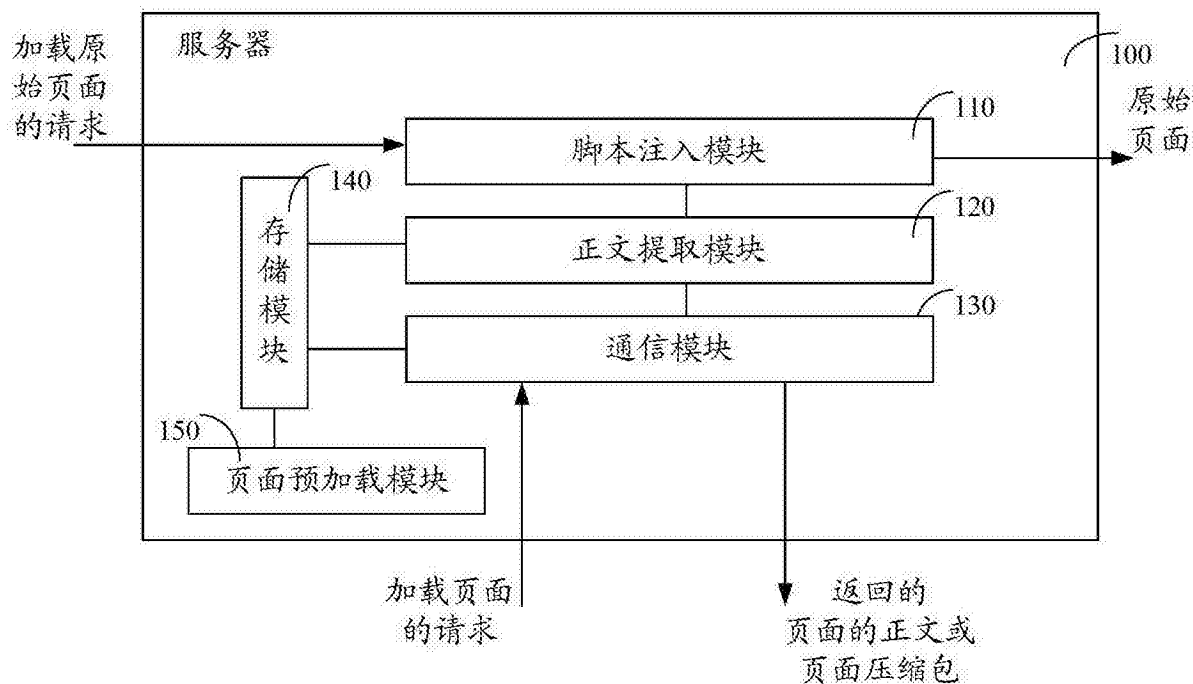


图 4

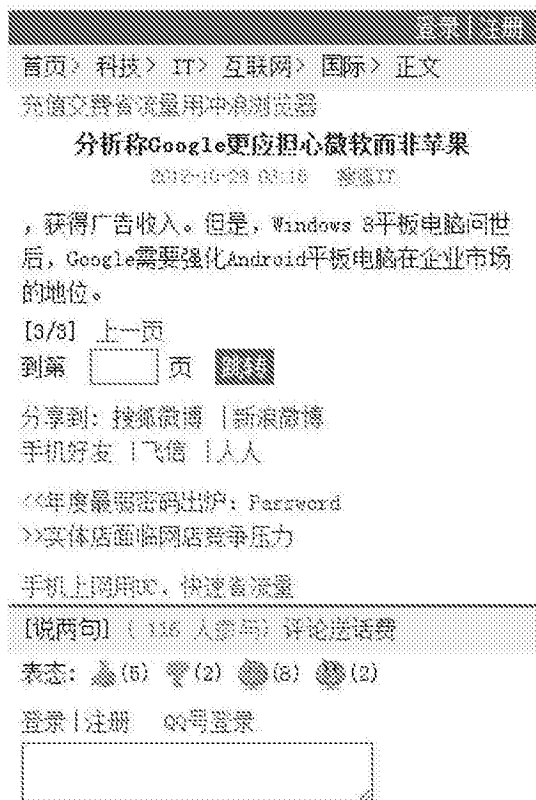


图 5

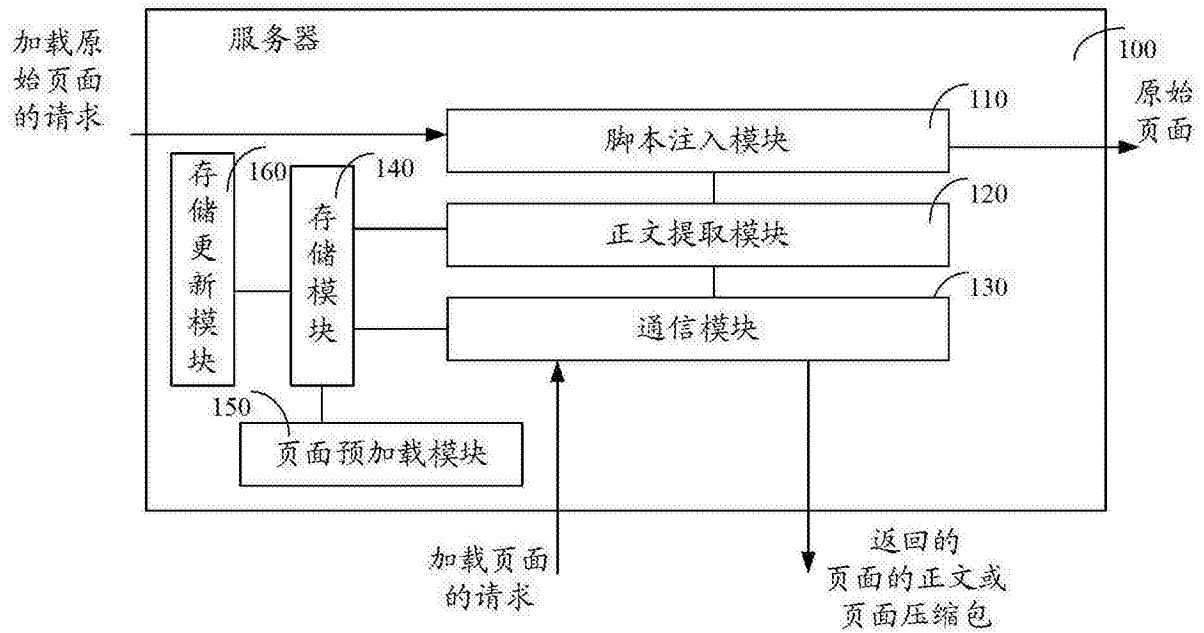


图 6

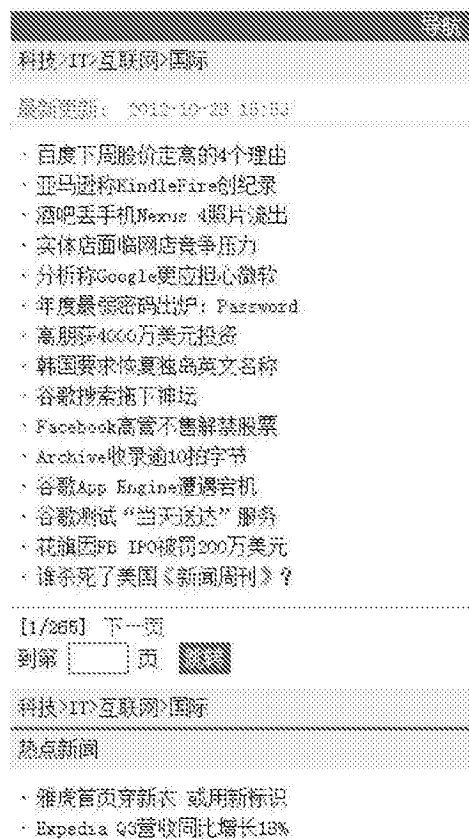


图 7

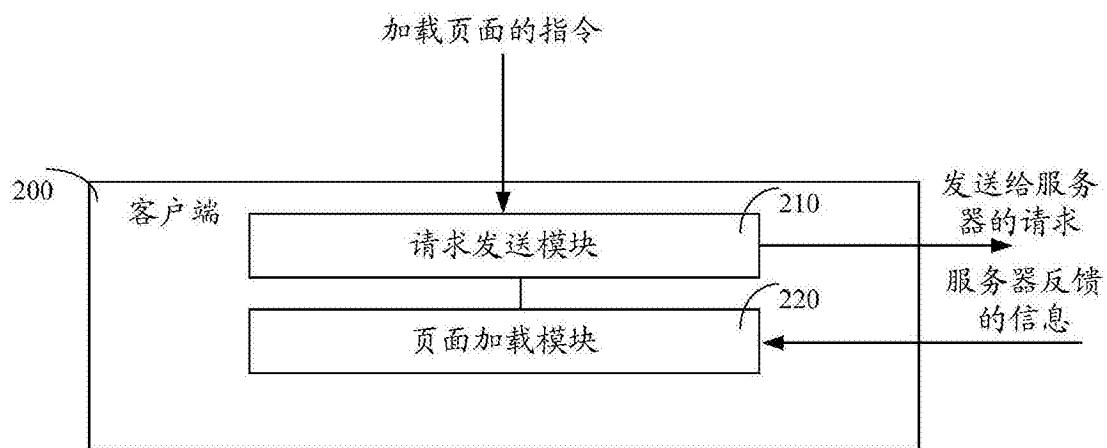


图 8

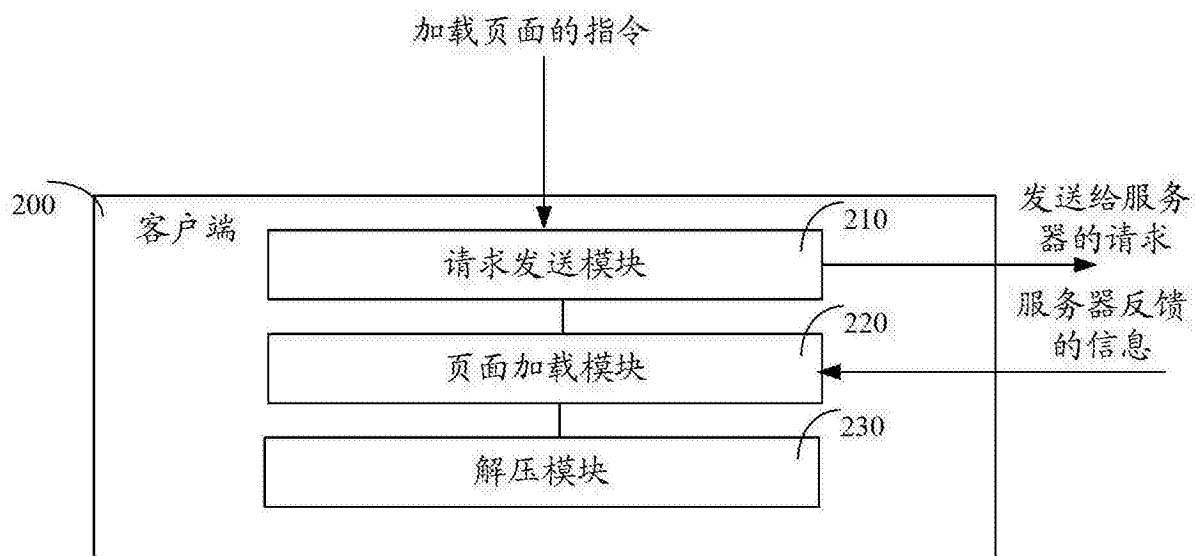


图 9

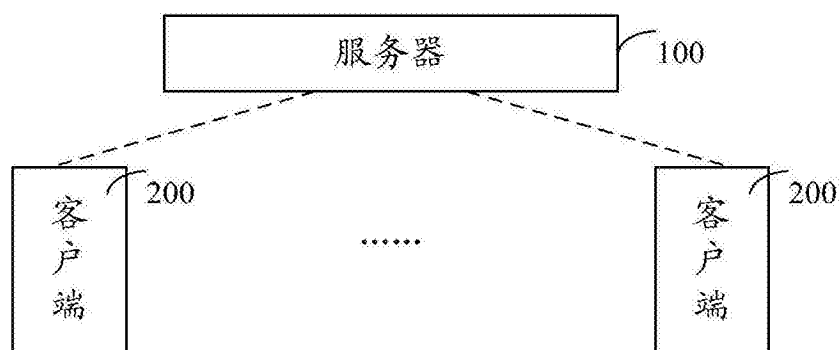


图 10

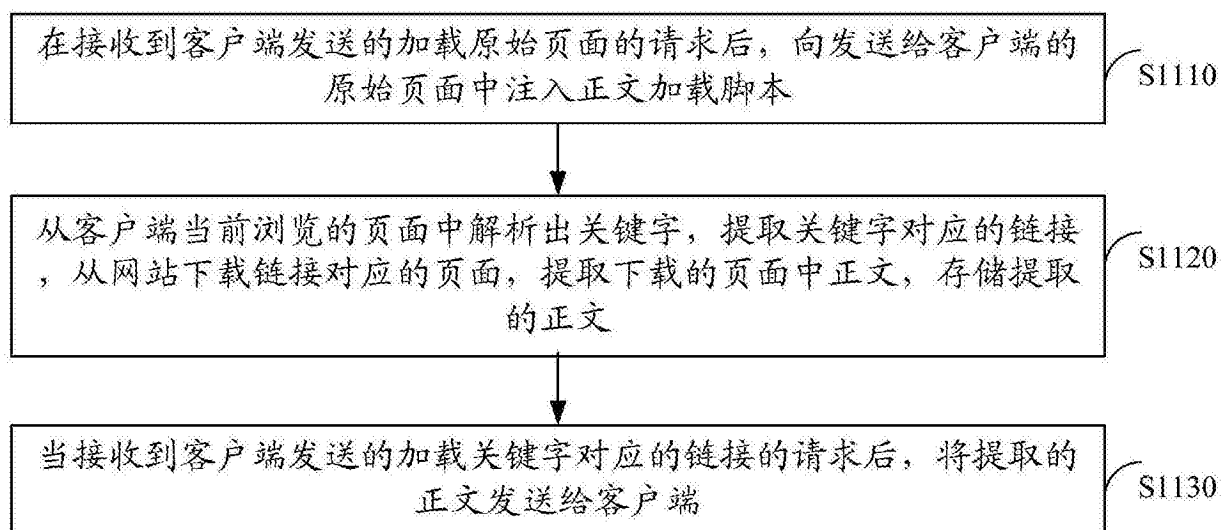


图 11

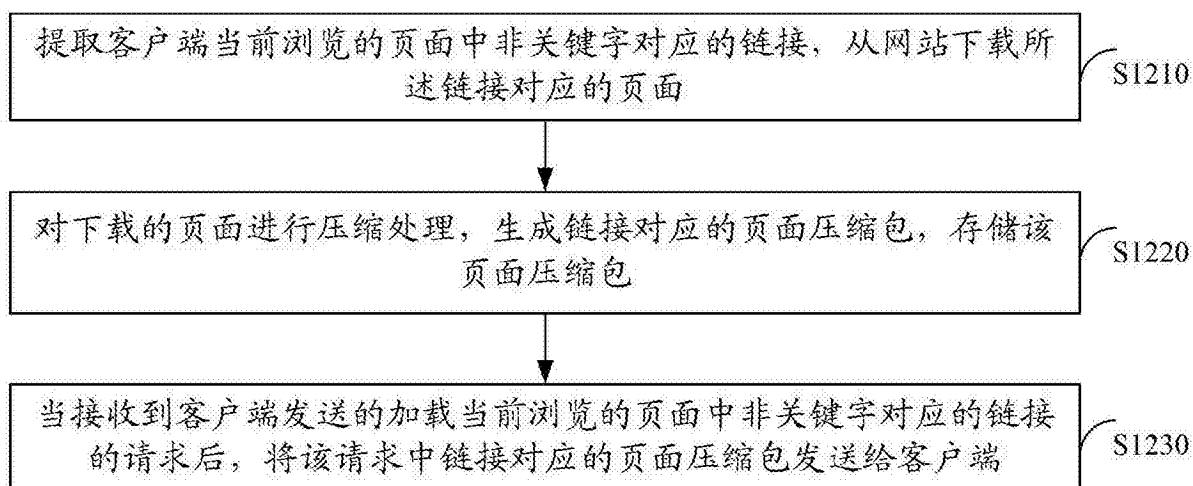


图 12