



(12)发明专利申请

(10)申请公布号 CN 105745644 A

(43)申请公布日 2016.07.06

(21)申请号 201480059613.0

(22)申请日 2014.10.31

(30)优先权数据

PA2013-70640 2013.11.01 DK

(85)PCT国际申请进入国家阶段日

2016.04.29

(86)PCT国际申请的申请数据

PCT/EP2014/073434 2014.10.31

(87)PCT国际申请的公布数据

W02015/063260 EN 2015.05.07

(71)申请人 柯宝科技公司

地址 丹麦霍斯霍姆自治市

(72)发明人 莫滕·西尔韦斯特·奥尔森

(74)专利代理机构 北京市磐华律师事务所

11336

代理人 高伟 卜璐璐

(51)Int.Cl.

G06F 17/30(2006.01)

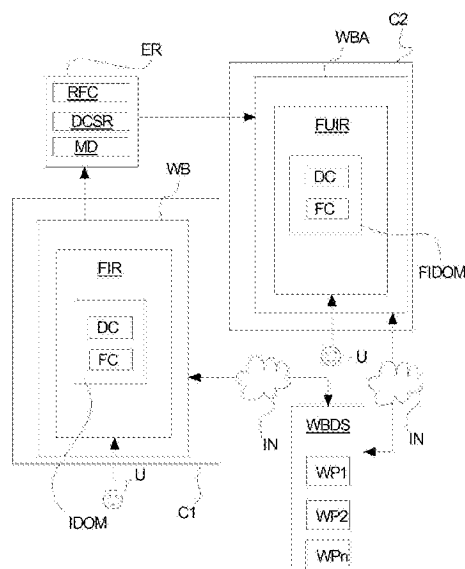
权利要求书4页 说明书26页 附图16页

(54)发明名称

表示在网页浏览器中的网页的状态表示的建立

(57)摘要

现在公开的发明性概念涉及建立表示在网页浏览器中的网页的状态表示。此概念包括处理从基于网页的数据源检索到的网页；建立网页的作为结果的第一内部浏览器状态表示；建立表示第一内部表示的状态的外部表示；建立第一内部浏览器状态表示的第一内容的表示；建立表示第一内部浏览器状态表示的动态内容的状态的动态内容状态表示；建立包括动态内容和第一内容之间的关系的映射数据；以及使得外部表示可用于网页浏览器应用以便在与第一内部浏览器状态表示的状态相对应的状态下在网页浏览器应用中建立网页的进一步的内部表示。



1. 建立表示在网页浏览器中的网页的状态表示的方法,所述方法包括:利用所述网页浏览器:

对从基于网页的数据源检索到的网页执行网页处理;

在所述网页浏览器中建立所述网页的作为结果的第一内部浏览器状态表示;

建立外部表示;所述外部表示表示所述第一内部表示的状态;

其中,建立所述状态的所述外部表示包括:

建立所述第一内部浏览器状态表示的第一内容的表示;

建立表示所述第一内部浏览器状态表示的动态内容的状态的动态内容状态表示;以及

建立包括映射所述动态内容和所述第一内容之间的关系的映射数据;以及

使得所述外部表示能够用于网页浏览器应用以便在与所述第一内部浏览器状态表示的状态相对应的状态下在所述网页浏览器应用中建立所述网页的进一步的内部表示。

2. 根据权利要求1所述的方法,其中建立外部状态表示包括串行化所述第一内部浏览器状态表示的至少一部分。

3. 根据权利要求2所述的方法,其中所述第一内部浏览器状态表示的所述部分包括所述动态内容和所述第一内容中的一个或多个的至少一部分。

4. 根据权利要求1所述的方法,其中建立所述进一步的内部表示包括解串行化所述外部表示的至少一部分。

5. 根据权利要求1所述的方法,其中建立所述进一步的内部表示包括将所述外部表示的至少一部分解析到所述网页浏览器应用中。

6. 根据权利要求1所述的方法,其中所述外部表示被建立成独立于一个或多个存储器位置,所述一个或多个存储器位置与所述第一内部表示相对应。

7. 根据权利要求1所述的方法,其中建立所述外部表示包括多个识别框架,所述多个识别框架包括所述网页的一个或多个内联框架。

8. 根据权利要求1所述的方法,其中对两个或多个框架的每一个建立外部框架表示,所述两个或多个框架的至少一个包括所述网页的至少一个内联框架。

9. 根据权利要求1所述的方法,其中建立所述外部表示包括建立框架相关表示。

10. 根据权利要求1所述的方法,其中所述外部表示基于所述网页浏览器中的所述网页的内部文档对象模型表示而建立。

11. 根据权利要求10所述的方法,其中所述外部表示通过串行化所述内部文档对象模型的内容的至少一部分而建立。

12. 根据权利要求1所述的方法,其中建立所述进一步的内部表示包括:

解串行化所述外部表示的至少一部分;以及

将所述外部表示的经解串行化的部分解析成所述网页浏览器应用的内部文档对象模型。

13. 根据权利要求1所述的方法,其中所述第一内部浏览器状态表示的所述第一内容包括标记内容,并且其中所述外部表示包括所述标记内容的表示。

14. 根据权利要求1所述的方法,其中所述内部浏览器状态表示的所述第一内容包括样式表内容的表示,并且其中所述外部表示包括所述样式表内容的表示。

15. 根据权利要求1所述的方法,其中所述第一内容包括关于所述第一内部浏览器状态

表示的cookie数据,并且其中所述外部表示包括所述cookie数据的表示。

16.根据权利要求1所述的方法,其中所述第一内容包括所述第一内部浏览器状态表示的形式内容,并且其中所述外部表示包括所述形式内容的表示。

17.根据权利要求1所述的方法,其中所述映射数据包括所述第一内部浏览器状态表示的所述动态内容和标记内容之间的一个或多个关系的映射。

18.根据权利要求1所述的方法,其中所述映射数据包括所述第一内部浏览器状态表示的所述动态内容和样式表内容之间的指针的映射。

19.根据权利要求1所述的方法,其中所述映射数据包括将一个或多个指针映射到所述第一内部浏览器状态表示的一个或多个对象的一个或多个存储器地址。

20.根据权利要求19所述的方法,其中建立所述映射数据包括:

处理所述第一内部表示;

至少部分基于所述处理识别所述对象之间的一个或多个指针;以及

将所述指针映射到所述存储器地址。

21.根据权利要求1所述的方法,其中建立所述外部表示包括将所述第一内容转化成一种或多种格式,所述一种或多种经转化的格式中的每一个都不同于所述内部浏览器状态表示的格式。

22.根据权利要求1所述的方法,其中建立所述外部表示包括将所述第一内部浏览器状态表示的至少一部分转化成所述第一内部表示的内部文档对象模型的外部表示。

23.根据权利要求1所述的方法,其中建立所述外部表示包括将所述第一内部浏览器状态表示的不同框架的多个内部文档对象模型转化成所述多个内部文档对象模型的一个或多个外部表示。

24.根据权利要求23所述的方法,其中所述第一内部浏览器状态表示包括标记内容,其中所述转化包括将所述标记内容的至少一些转化为标记语言格式。

25.根据权利要求23所述的方法,其中所述第一内部浏览器状态表示包括样式信息,并且

其中所述转化包括将所述样式信息的至少一些转化成层叠样式表格式。

26.根据权利要求1所述的方法,其中所述第一内部浏览器状态表示包括多个不同的内容元素,每个不同的内容元素包括下述内容的一个或多个中的至少一部分:

所述第一内容;以及

所述动态内容,

其中建立所述外部表示进一步包括建立多个个体表示,每个个体表示基于所述第一内部浏览器状态表示,并且

其中每个个体表示与所述多个不同的内容元素中的至少一个相对应。

27.根据权利要求1所述的方法,其中建立所述进一步的内部表示包括映射所述浏览器应用中的浏览器标记语言内容的表示和动态内容之间的至少一个联系,其中所述映射至少部分基于所述映射数据。

28.根据权利要求27所述的方法,其中建立所述进一步的内部表示包括建立所述进一步的内部表示的一个或多个对象之间的一个或多个指针,其中所述建立至少部分基于所述映射数据。

29. 根据权利要求1所述的方法, 其中所述第一内部表示和所述进一步的内部表示中的每一个都包含在下述内容的一个或多个中:

至少一个网页浏览器;

至少一个网页浏览器应用;

至少一个网页, 以及

其中所述第一内部表示和所述进一步的内部表示包含在下述内容的一个或两个中:

所述一个或多个网页浏览器、网页浏览器应用和网页的一个; 以及

所述一个或多个网页浏览器、网页浏览器应用和网页的不同的类型。

30. 根据权利要求1所述的方法, 进一步包括在多个网页浏览器应用中建立多个进一步的内部表示, 其中所述进一步的内部表示在其中建立的所述多个网页浏览器应用的每一个都不同于所述进一步的内部表示在其中建立的所有其他的所述网页浏览器应用。

31. 根据权利要求1所述的方法, 其中所述网页的所述网页处理包括在浏览操作期间处理所述网页的所述动态内容, 所述浏览操作对所述网页进行操作。

32. 根据权利要求1所述的方法, 其中所述网页的所述动态内容包括所述网页的脚本。

33. 根据权利要求1所述的方法, 其中包括所述第一内部表示的所述网页浏览器至少部分基于将所述第一内部表示的至少一部分串行化而将所述第一内部浏览器状态表示外在化成所述外部表示。

34. 根据权利要求1所述的方法, 其中用于所述进一步的内部表示的所述网页浏览器应用至少部分基于将所述外部表示的至少一部分解串行化而建立所述进一步的内部表示。

35. 根据权利要求1所述的方法, 进一步包括利用网络机器人:

确定是否满足一个或多个预定的标准;

至少部分基于确定满足一个或多个所述预定的标准发起建立所述外部表示,

其中所述网络机器人配置为基于外部表示建立一个或多个进一步的内部表示。

36. 一种系统, 所述系统用于建立网页浏览器的第一内部浏览器状态表示的外部表示并在此基础上建立进一步的内部表示, 所述系统包括:

网页浏览器, 配置为将所述第一内部浏览器状态表示外在化成所述状态的外部表示, 其中建立所述状态的所述外部表示包括:

建立所述第一内部浏览器状态表示的第一内容的表示;

建立表示所述第一内部浏览器状态表示的动态内容的状态的动态内容状态表示;

建立包括映射所述第一内部浏览器状态表示的所述动态内容和所述第一内容之间的关系的映射数据; 以及

其中网页浏览器应用配置为在与所述第一内部浏览器状态表示的状态相对应的状态下基于所述外部表示建立进一步的内部状态表示。

37. 根据权利要求36所述的系统, 其中所述系统配置为根据权利要求1到35的任一项的方法进行操作。

38. 结合一个或多个网络机器人使用根据权利要求1到37的任一项所述的方法和/或系统,

其中所述一个或多个网络机器人配置为在与第一内部浏览器状态表示的状态相对应的状态下在网络浏览器应用中发起建立网页的进一步的内部表示, 以及

其中所述进一步的内部表示基于外部状态表示而建立。

39. 根据权利要求1到38的任一项所述的方法、系统的使用和/或应用,用于在一个或多个不同的网页浏览器应用中建立多个浏览事件。

40. 使用根据权利要求39所述的方法,其中所述多个所述浏览事件在一个或多个服务器上执行。

41. 用于建立外部浏览器状态表示的网页浏览器,所述网页浏览器包括:

用于建立第一内部浏览器状态表示的第一内容的外部表示的装置,

用于建立表示所述第一内部表示的动态内容的状态的外部动态内容状态表示的装置,以及

用于建立包括映射所述第一内部浏览器状态表示的所述动态内容和所述第一内容之间的关系的映射数据的装置(MM)。

42. 根据权利要求41所述的网页浏览器,所述网页浏览器配置为根据权利要求1-35的任一项所述的方法进行操作。

43. 一种网页浏览器应用,所述网页浏览器应用用于基于预建立的外部浏览器状态表示建立进一步的内部浏览器状态表示,所述网页浏览器应用包括:

用于处理预建立的外部浏览器状态表示的装置,

用于将所述处理的结果解析成所述网页浏览器应用的内部文档对象模型的装置,以及

用于处理所述外部浏览器状态表示的映射数据、并基于所述映射数据实现在所述内部文档对象模型中所表示的动态内容和第一内容之间的关系的装置。

44. 根据权利要求43所述的网页浏览器,所述网页浏览器配置为根据权利要求1-35的任一项所述的方法进行操作。

45. 浏览包括多个网页的基于网页的数据源的方法,所述方法包括下列步骤:

利用网页浏览器浏览所述基于网页的数据源的一个或多个网页,

在所述浏览期间利用所述网页浏览器建立不同的第一内部浏览器状态的至少两个外部表示,

基于至少一个所述外部表示在一个或多个网页浏览器应用中建立一个或多个附加的浏览事件,

所述建立一个或多个附加的浏览事件包括处理所述外部表示的内容以及执行解析成所述一个或多个网页浏览器应用的内部状态表示。

46. 根据权利要求45所述的方法,其中所述至少两个外部表示根据权利要求1-35的任一项所述的方法和/或利用根据权利要求36-37的任一项所述的系统被建立和/或实现到一个或多个网页浏览器应用中。

47. 根据权利要求45或46所述的方法,其中所述方法由网络机器人实施。

48. 使用网络机器人用于利用网页浏览器浏览网页,其中所述网络机器人根据预定的规则集进行操作以便处理一个或多个网页,并且

其中所述网络机器人配置为响应于确定符合一个或多个预定的标准根据权利要求1-35的任一项发起建立一个或多个外部表示。

表示在网页浏览器中的网页的状态表示的建立

[0001] 优先权声明

[0002] 本申请涉及并要求于2013年11月1日提交的、名称为“ESTABLISHMENT OF STATE REPRESENTATION OF A WEB PAGE REPRESENTED IN A WEB BROWSER”的PA2013-70640号丹麦专利申请的优先权,其全文通过引用并入本文。

[0003] 发明的领域

[0004] 现在公开的发明性概念涉及配置为建立表示在网页浏览器中的网页的状态表示的方法、系统和计算机程序产品,用于建立网页浏览器的第一内部浏览器状态表示的外部表示以及在此基础上建立进一步的内部表示的系统,网页浏览器和浏览基于网页的数据源的方法。

背景技术

[0005] 近几年期间已经开发了网页的内容和基本建设。因此,网页的一大部分是动态网页,其包括动态内容诸如脚本内容连同标记语言内容和其他内容。网页的脚本内容帮助建立直观的以及多功能的网页,以及在浏览网页期间自动适于用户输入的网页。

[0006] 其结果是,当动态网页由网页浏览器处理时,可导致在不同的网页浏览器中的网页基于例如用户动作、网页浏览器的位置、脚本内容本身等而具有独特表示。但是这导致了关于确定网页浏览器中动态网页的结果表示的问题。

[0007] 现在公开的发明性概念等提供了上述问题的解决方案。

发明内容

[0008] 现在公开的发明性概念包含系统、方法、计算机程序产品等,其配置为帮助建立表示在网页浏览器中的网页的状态表示。在多个实施例中,这些概念包括如下。

[0009] 在一种方法中,建立表示在网页浏览器中的网页的状态表示的方法包括利用网页浏览器实施下述操作:对从基于网页的数据源检索到的网页执行网页处理;在所述网页浏览器中建立所述网页的作为结果的第一内部浏览器状态表示;建立外部表示;所述外部表示表示所述第一内部表示的状态;其中,建立所述状态的所述外部表示包括:建立所述第一内部浏览器状态表示的第一内容的表示;建立表示所述第一内部浏览器状态表示的动态内容的状态的动态内容状态表示;以及建立包括映射所述动态内容和所述第一内容之间的关系的映射数据;以及使得所述外部表示能够用于网页浏览器应用以便在与所述第一内部浏览器状态表示的状态相对应的状态下在所述网页浏览器应用中建立所述网页的进一步的内部表示。

[0010] 在另一个实施例中,系统配置为建立网页浏览器的第一内部浏览器状态表示的外部表示并且在此基础上建立进一步的内部表示。该系统包括:网页浏览器,其配置为将第一内部浏览器状态表示外在化到该状态的外部表示,以及网页浏览器应用,其配置为在与第一内部浏览器状态表示的状态相对应的浏览器状态下基于外部状态表示建立进一步的内部状态表示。建立状态的外部表示包括:建立第一内部浏览器状态表示的第一内容的表示;

建立动态内容状态表示,其表示第一内部浏览器状态表示的动态内容的状态;以及建立映射数据,其包括映射第一内部浏览器状态表示的动态内容和第一内容之间的关系。

[0011] 在又一个实施例中,网页浏览器配置为通过下述操作建立外部浏览器状态表示:建立第一内部浏览器状态表示的第一内容的外部表示,建立外部动态内容状态表示,其表示第一内部表示的动态内容的状态,以及建立映射数据,其包括映射第一内部浏览器状态表示的动态内容和第一内容之间的关系。

[0012] 在附加的方法中,网页浏览器配置为基于预建立的外部浏览器状态表示通过下述操作建立进一步的内部浏览器状态表示:处理预建立的外部浏览器状态表示,将处理的结果解析成网页浏览器应用的内部文档对象模型,以及处理外部浏览器状态表示的映射数据,以及基于映射数据实现表示在内部文档对象模型中的动态内容和第一内容之间的关系。

[0013] 在再一个实施例中,方法用于浏览包括多个网页的基于网页的数据源。该方法包括,利用网页浏览器:浏览基于网页的数据源的一个或多个网页;在浏览期间建立不同的第一内部浏览器状态的至少两个外部表示;至少部分基于处理外部表示的内容建立一个或多个基于至少一个外部表示的附加浏览器事件;以及实施解析成一个或多个网页浏览器应用的内部状态表示。

附图说明

[0014] 下面将参考附图进一步详细解释现在公开的发明性概念。

[0015] 图1例示了根据现在公开的发明性概念的若干实施例的系统。

[0016] 图2例示了根据现在公开的发明性概念的实施例的、关于建立外部表示的实施例的流程图。

[0017] 图3例示了根据现在公开的发明性概念的实施例的、关于基于外部表示建立在网页浏览器中的进一步的内部状态表示的实施例的流程图。

[0018] 图4例示了根据现在公开的发明性概念的实施例的、相互关系的示意性映射,

[0019] 图5例示了现在公开的发明性概念的实施例的、建立外部表示的进一步的实施例,

[0020] 图6例示了根据现在公开的发明性概念的实施例的网页浏览器,

[0021] 图7例示了关于建立多个进一步的内部状态表示的现在公开的发明性概念的实施例

[0022] 图8例示了关于基于预建立的外部表示,例如利用服务器,建立多个进一步的内部状态表示的现在公开的发明性概念的进一步的实施例,

[0023] 图9例示了根据现在公开的发明性概念的实施例的客户端

[0024] 图10例示了根据现在公开的发明性概念的方面的、网页服务器的实施例,

[0025] 图11例示了现在公开的发明性概念的实施例,其中外部表示存储在客户端,

[0026] 图12例示了关于使用多个外部浏览器状态表示的现在公开的发明性概念的实施例,

[0027] 图13例示了根据现在公开的发明性概念的实施例的、外部表示的实施例,

[0028] 图14例示了根据现在公开的发明性概念的实施例的、通过网页浏览器在屏幕上可视化的网页的示例,以及

[0029] 图15-16例示了现在公开的发明性概念的实施例,其中计算机实现的网络机器人在浏览期间使用外部状态表示。

[0030] 当然,前述的附图要理解成仅仅是现在公开的发明性概念的范围内的示例性实施例的例示性说明。根据本领域普通技术人员阅读这些说明所获得的理解,附加的实施例、前述实施例的变化以及可替代的实施例也都认为完全落入本申请的范围。

具体实施方式

[0031] 本文的说明被呈现以使能本领域的任何技术人员制造和使用本发明并且本文的说明在本发明的特定应用和它们的要求的上下文中提供。对公开的实施例的多种修改对本领域的技术人员将是容易明显的,并且本文限定的一般原理可用于其他实施例和应用而不脱离本发明的精神和范围。因而,本发明并不意图限制为示出的实施例,而是给予与本文公开的原理和特征相一致的最宽的范围。

[0032] 尤其是,本文讨论的本发明的多个实施例利用互联网作为多个计算机系统之间通信的手段来实现。本领域技术人员将认识到,本发明并不限于使用互联网作为通信介质,并且本发明的替代性方法可适于使用专用网络、局域网(LAN)、广域网(WAN)或其他通信手段。此外,可以利用有线、无线(例如,射频)和光通信链路的多种组合。

[0033] 可以执行本发明的一个实施例的程序环境例示地包含一个或多个通用计算机或专用设备,诸如手持式计算机。这种设备(例如,处理器、存储器、数据存储、输入和输出设备)的细节是已知的并且为了清楚被省略了。

[0034] 还应该理解,本发明的技术可利用多种技术实现。例如,本文描述的方法可用在计算机系统上运行的软件实现,或用利用用于实施方法的操作的一个或多个处理器和逻辑(硬件和/或软件)的硬件、专用集成电路、可编程逻辑设备诸如现场可编程门阵列(FPGA)和/或其多种组合来实现。在一种例示性的方法中,本文描述的方法可以通过居于存储介质诸如物理的(例如,非暂态的)计算机可读介质上的一系列计算机可执行的指令来实现。此外,虽然本发明的特定实施例可利用面向对象的软件编程概念,但是本发明并不限于此并且容易适于利用指导计算机的操作的其他形式。

[0035] 本发明还可以计算机程序产品的形式提供,所述计算机程序产品包括其上具有计算机代码的计算机可读存储介质或信号介质,所述计算机代码可由计算设备(例如,处理器)和/或系统执行。计算机可读存储介质可包括能够在其上存储计算机代码由计算设备或系统使用的任何介质,包括光介质诸如只读或可写入CD以及DVD、磁存储器或介质(例如,硬盘驱动、磁带)、半导体存储器(例如,FLASH存储器和其他便携式存储卡等)、编码在芯片上的固件等。

[0036] 计算机可读信号介质是不适于前述存储介质类别内的介质。例如,例示性的计算机可读信号介质传达或转移系统内的、系统之间的暂态信号例如经由物理网络或虚拟网络等。

[0037] 根据一些方法,本文中描述的方法和系统可以用虚拟系统和/或诸如仿真MAC OS环境的UNIX系统、实际上主用MICROSOFT WINDOWS环境的UNIX系统、仿真MAC OS环境的MICROSOFT WINDOWS等仿真一个或多个其他系统的系统实现和/或在其上实现。在一些实施例中,此虚拟化和/或仿真可以通过使用VMWARE软件被加强。

[0038] 在更多的方法中,一个或多个网络可以表示通常称为“云”的系统集群。在云计算中,共享的资源,诸如处理功率、外围设备、软件、数据处理和/或存储、服务器等,被提供给云中的任何系统,优选地以按需关系,由此允许跨许多计算系统访问和分配服务。云计算典型地包含互联网或在云中操作的系统之间的其他高速连接(例如,4G LTE、光纤等),但还可以使用连接系统的其他技术。

[0039] 在一个实施例中,方法包括:利用网页浏览器对从基于网页的数据源检索到的网页执行网页处理,并且建立网页浏览器中的网页的作为结果的第一内部浏览器状态表示,在网页处理期间建立第一内部表示的外部表示,其中状态的外部表示的建立包括:建立第一内部浏览器状态表示的第一内容的表示;建立表示第一内部浏览器状态表示的动态内容的状态的动态内容状态表示;以及建立映射数据,其包括第一内部浏览器状态表示的动态内容和第一内容之间的相互关系的映射。

[0040] 在另一个实施例中,该方法进一步包括步骤:使得该外部表示可用于网页浏览器应用,以便在与第一内部浏览器状态表示的状态相对应的状态下建立网页浏览器应用中的网页的进一步内部表示。

[0041] 现在公开的发明性概念优选地帮助提取和存储一个或多个网页浏览器状态,例如,利用浏览器的两个统一资源定位符(URL)之间的访问。URL也称为网页地址。然后这些浏览器状态可在其他网页浏览器应用或在后面时间的相同网页浏览器应用中重新建立。因此,对于例如网络机器人开发者和/或网络机器人在后面的时间点返回先前的状态是可能的。这通过处理外部表示以便建立进一步的内部表示提供方便。因此,当在建立进一步的内部表示以后浏览网页时,由于在外在化第一内部浏览器状态表示时呈现网页,用户或网络机器人能够浏览网页。因此,促进了循环特征,用户或网络机器人可跟随不同于第一内部状态表示的浏览路径,而不用体验先前的用户或脚本动作已经修改状态。

[0042] 术语“网页浏览器”和“网页浏览器应用”涉及用于检索并在万维网(WWW)上呈现网页的软件应用。网页可包括数据诸如图像、文本、视频、脚本内容和/或其他内容部分。网页可包括超链接,以使能用户容易将他们的浏览器导航到相关资源。网页通常是具有基于例如由用户或计算机程序诸如脚本内容提供或操控的参数而改变的内容的动态网页。

[0043] 可理解,在现在公开的发明性概念的方面中,进一步的内部表示可在相同的网页浏览器中实现,所述网页浏览器已用于建立外部表示表示的第一初始内部浏览器状态表示。然而,如后面描述的,进一步的内部表示还可在进一步的网页浏览器应用中实现,以便此网页浏览器应用处于与第一内部表示的状态相对应的状态。

[0044] 由于进一步的内部表示可基于外部状态表示,进一步的内部表示可通过网页浏览器应用来建立以大致与第一内部表示相对应。这时,与导致第一内部表示的网页实施一系列交互的用户将能够识别在相同的或新的网页浏览器中的进一步的内部表示,并且从此点开始继续。由于映射,处于建立了外部表示的状态的内部表示的其他内容和动态内容的状态能够被复制,包括处于该状态的动态内容的状态。

[0045] 此外,现在公开的发明性概念导致相对于在动态网页的网页处理期间实施的脚本而言稳定的状态表示。由于用户触发的动作以及脚本触发的动作,这样的脚本可修改动态网页的内部表示。这样的动作动态修改网页的内部表示,以便网页看起来不同并且包含依赖于用户动作和脚本内容的状态的不同网页浏览器中的不同内容。然而,由于除了映射数

据以外,浏览器的状态可被仔细地并且广泛地在外部表示,以便提供随后修复与第一内部浏览器状态表示的状态相对应的状态的机会。

[0046] 而且,在现在公开的发明性概念的优选方面,第一内部浏览器状态表示的第一内容可以以第一格式建立,动态内容状态表示可以以第二格式建立,并且映射数据可以以第三格式建立。所以,在现在公开的发明性概念的方面,可以使用用于表示外部表示的不同部分的不同格式。这可以例如基于可以建立的外部表示提供更可靠的和/或高效的进一步内部表示的建立。由此,一种格式对于例如建立网页浏览器中的动态表示可能是有利的,但如果用于例如映射数据则可能不太有利。

[0047] 在现在公开的发明性概念的有利方面,外部状态表示的建立包括串行化第一内部浏览器状态表示的至少一部分。

[0048] 该串行化可导致第一内部表示的状态被转化/转换成可存储到网页浏览器外部(例如,在文档或存储缓冲器中)的格式,并且以后用相同的或另一网页浏览器重建。在串行化期间,根据预定的串行化格式,第一内部表示由计算机程序处理并且串行化。

[0049] 然后,经串行化的数据可用于创建第一内部表示的状态的进一步的内部表示。因此,如果内部表示在网页浏览器中,串行化可包括将数据结构和/或对象状态转换成可存储并反映浏览器状态的格式。

[0050] 在现在公开的发明性概念的方面,映射数据可在串行化期间根据串行化格式建立。

[0051] 而且,此外在某些方面,串行化可帮助将第一内部表示的部分转换成不同格式,所述第一内部表示例如网页浏览器的内部的文档对象模型。

[0052] 因此,串行化帮助提取第一内部表示到外部表示,以便有效地并且受控制地建立网页浏览器状态的“快照”。

[0053] 在现在公开的发明性概念的有利方面,包括第一内部表示的网页浏览器包括用于处理第一内部表示以建立外部表示的装置。

[0054] 在现在公开的发明性概念的方面,建立外部状态表示包括串行化动态内容和/或第一内部浏览器状态表示的第一内容的至少一部分。

[0055] 然后经串行化的动态内容和/或第一内容的部分被存储为外部表示。这例如提供了建立外部表示的有效途径。

[0056] 在现在公开的发明性概念的有利方面,建立进一步的内部表示包括解串行化外部表示的至少一部分。

[0057] 解串行化可通过意图用于进一步的内部表示的网页浏览器应用推动,并且可包括处理外部表示,以例如分配包括网页浏览器应用等的客户端或网页服务器上的数据存储器。例如,XML可认为对于提供例如内部表示的HTML内容或其他标记性内容的外部表示是尤其有利的,在其中其被严格/明确表示并提供可基于外部表示结合建立进一步的表示被精确解析的可靠表示。然而,任何其他格式也可在其他方面相关。

[0058] 在本公开的发明性概念的方面,建立进一步的内部表示包括将外部表示的至少一部分解析成网页浏览器应用。

[0059] 优选地该解析由网页浏览器应用提供帮助,其意图用于进一步的内部表示。该解析可将样式表内容、标记语言内容和其他产生于外部表示的内容解析成网页浏览器应用中

的内部的文档对象模型表示。由于此目的,网页浏览器应用可包括标记语言解析器诸如HTML解析器、XML解析器、样式表(CSS)解析器等等,以便将外部表示的内容解析成网页浏览器应用以在浏览器中建立内部DOM。

[0060] 在现在公开的发明性概念的优选方面,可以建立独立于第一内部表示的存储器位置的外部表示。

[0061] 这可以通过串行化和/或其他处理第一内部表示、通过映射如以后更加详细描述的第一内部表示的对象之间的关系等等提供帮助。因此外部表示可更加通用并且在进一步的浏览器中帮助容易/有利的实现方式。

[0062] 在现在公开的发明性概念的有利方面,建立外部表示可包括识别框架诸如网页的内联框架。

[0063] 动态网页通常包括多个框架。这样的框架可表示在第一内部表示的每个它们的不同的文档对象模型中,并且帮助随后的重建这种框架的状态,该框架对识别以提供合适的外部表示是有利的。框架的识别可例如通过使软件应用处理内部表示的DOM来完成。

[0064] 在方面,可建立两个或多个框架的外部表示,诸如网页的内联框架。

[0065] 因此第一内部表示的每个框架可包括外部表示中的它自己的外部框架表示,以便使能以后更加便利的建立第一内部状态表示。

[0066] 在方面,建立外部表示可包括建立框架关联表示。

[0067] 框架关联表示可例如包括第一内部表示的框架树的信息,以使得所表示的框架由此可在网页浏览器中的进一步的内部表示中建立。

[0068] 在本公开的发明性概念的有利方面,外部表示基于网页浏览器中的网页的内部的文档对象模型而建立。

[0069] 包括第一内部表示的网页浏览器的内部的文档对象模型连续地至少部分表示浏览器的状态,并且因此处理网页浏览器的内部文档对象模型是便利的。

[0070] 在有利的方面,外部表示可通过串行化内部的文档对象模型的内容的至少部分来建立。

[0071] 此串行化可例如帮助提供一致的并且可靠的外部状态表示。

[0072] 在本公开的发明性概念的方面,进一步的内部表示可通过解串行化外部表示的至少部分并且将解串行化的结果解析成网页浏览器应用的作为结果的内部文档对象模型来建立。

[0073] 解串行化可例如帮助提供一致的并且可靠的外部状态表示,其考虑实现外部表示所处的环境,例如进一步的内部表示将在其上建立的客户端的软件和/或硬件。

[0074] 在方面,第一内部浏览器状态表示的第一内容包括标记内容,并且外部表示包括标记内容的表示。

[0075] 因此,可以这么说网页浏览器的“静态”内容的状态被外在化成外部表示,优选地通过串行化标记语言内容。

[0076] 要理解,标记内容还可称为标记语言内容,包括网页的内部表示,其在将网页解析到网页浏览器期间已通过网页浏览器建立。

[0077] 有利地,在方面,内部浏览器状态表示的第一内容可包括样式表内容的表示,并且外部表示可包括样式表内容的表示。

[0078] 要理解,样式表内容包括网页的内部表示,其已在将网页的样式表内容解析到浏览器期间通过网页浏览器建立。

[0079] 在本公开的发明性概念的方面,第一内容可包括与第一内部浏览器状态表示相关的cookie数据,并且外部表示可包括cookie数据的表示。

[0080] cookie数据可例如关于帮助建立进一步的内部表示,其中在浏览期间以第一内部表示结束的与用户相关的数据可以是相关的。因此,cookie数据可从包含内部表示的浏览器应用中得到并进入外部表示。

[0081] 在本公开的发明性概念的方面,第一内容可包括第一内部浏览器状态表示的形式内容,并且外部表示可包括形式内容的表示。

[0082] 这在被填充形式以形成基础用于进一步浏览网页的方面是相关的,并且因此,通过外在化并且实现填充这些形式的内容,建立浏览器的状态是可能的,其使能浏览网页而不用填充形式。这尤其相关于网络机器人,其不能够填充形式。此外,用户可以在外部表示建立先前用限定网络机器人应如何在进一步的内部表示中处理数据的内容来填充形式。该形式内容可以从内部表示中得到,以便外部表示包括此信息用于后面使用。

[0083] 在本公开的发明性概念的有利方面,映射数据包括映射第一内部浏览器状态表示的动态内容和标记内容之间的关系。

[0084] 网页的动态内容可访问和/或修改内容诸如内部表示的标记内容的对象。因此,为了使能建立与第一内部状态相对应的浏览器状态,这样的关系对映射可是有利的。

[0085] 在本公开的发明性概念的有利方面,映射数据包括映射第一内部浏览器状态表示的样式表内容动态内容和之间的指针。

[0086] 在本公开的发明性概念的方面,映射数据包括将指针映射到第一内部浏览器状态表示的对象的存储器地址。

[0087] 例如动态内容诸如第一内部表示的脚本内容的指针可指向第一内部表示的例如样式表内容和/或标记内容的对象的存储器地址。然而,这些存储器地址不可用在进一步的内部表示中,其中存储器地址依赖于不同的处理事件随时间变化。例如,在建立外部表示以后,用户可继续浏览,并且这会对存储器地址产生影响,这样先前的对象被替代、修改甚至从存储器中删除。而且,如果进一步的内部表示位于另一客户端或网站服务器,而不是用于建立第一内部表示的那一个,则数据存储装置以及这样的内部建立可以是完全不同的。

[0088] 因此,这样的映射可以有利于识别指针所指向的存储器地址的对象,以便提供以后在进一步的内部表示中重建这种关系的可能性。然而,应理解,映射的其他或附加类型可以用在本公开的发明性概念的进一步方面中。

[0089] 在本公开的发明性概念的方面中,建立映射数据可包括处理第一内部表示,以便识别并映射对象之间的指针。

[0090] 因此,动态内容和第一内容诸如标记语言内容和/或第一内部表示的样式表内容之间的指针可以在进一步的表示中重建。这些指针可依赖于网页处理、用户输入、网页的脚本等等随时间改变。

[0091] 例如,指针可涉及/指向存储在处理网页以建立第一表示的客户端的数据存储器地址中的另一个值。因此这样的指针包含关于如何在进一步的表示中重建第一内部表示的信息,这样进一步表示的指针可以独立于第一表示的存储器分配而重建。即使客户端的指

定存储器位置的值可以在例如建立内部文档对象模型的表示中提取,当在进一步的表示中实现时,该值的指针可以是必须的,以便在进一步的表示中重建第一表示的状态。

[0092] 对象被认为是具有值的数据存储器中的位置并且通过标识该对象的标识符来引用。该对象可以是例如变量、函数或数据结构。

[0093] 在本公开的发明性概念的方面中,指针的映射可在内部表示的串行化期间由所谓的“取消混合”来实现。

[0094] 在本公开的发明性概念的方面中,建立外部表示包括将第一数据内容转化成不同于内部浏览器状态表示的格式的一个或多个形式。

[0095] 在优选的方面中,该转化可在串行化内部表示期间实施。

[0096] 在本公开的发明性概念的方面中,建立外部表示可包括将第一内部浏览器状态的部分表示转化成第一内部表示的内部的文档对象模型的外部表示,例如用可扩展标记语言。这可为确定以及重建内部表示提供益处。

[0097] 在本公开的发明性概念的方面中,建立外部表示包括将第一内部表示的不同框架的多个内部的文档对象模型转化成多个内部文档对象模型的一个或多个外部表示。

[0098] 因而,可以实现状态表示的更加深远的表示,其促进可靠的和/或灵活的外部表示,这对于以后的使用将更加容易和可靠。

[0099] 在方面,转化包括将第一内部浏览器状态表示的标记内容转化成标记语言格式诸如可扩展标记语言(XML)格式。

[0100] 这种转化促进了外部表示的后面的有利处理。在优选的方面中此内容可表示网页浏览器的内部DOM的至少一部分。

[0101] 在方面,转化包括将第一内部浏览器状态表示的样式信息转化成层叠样式表格式。

[0102] 此样式表形式可在解析样式表内容期间直接用于建立进一步的表示。

[0103] 在方面,建立外部表示可包括建立第一内部浏览器状态表示的不同内容的多个个体表示。

[0104] 因此,外部表示可包括关于此状态下的第一内部表示的样式表数据的样式表文件(例如,css)、关于此状态下的第一内部表示的样式表数据的标记语言文件(例如,XML)、包括此状态下的第一内部表示的动态内容的状态的文件(例如,用脚本语言)、包括关于此状态下的第一内部表示的cookie的表示的具有cookie数据的文本文件、此状态下的第一内部表示的形式内容的表示,以及包括如此文档中描述的映射信息的文件。所以外部表示可包括一起表示第一内部表示的状态的多个数据文件。

[0105] 在本公开的发明性概念的有利方面,建立进一步的内部表示可包括基于映射数据重新连接网页浏览器应用中的浏览器标记语言内容和动态内容。有利地,建立进一步的内部表示可包括基于映射数据重建进一步的内部表示的对象之间的指针。

[0106] 在将外部状态表示的内容解串行化并解析到网页浏览器应用之后,在方面,对象以及与第一内部表示的对象相对应的对象值也被建立。然而,以例如此类对象之间的指针形式的关系仍然需要被建立,并且因此映射数据被网页浏览器应用处理以重建此类对象之间的关系。

[0107] 在本公开的发明性概念的方面,重建指针可在解串行化期间通过所谓的“指针混

合”来实现,所述“指针混合”可例如基于对象名称/标识符实施参考的转化。

[0108] 在本公开的发明性概念的方面,第一内部表示和进一步的内部表示可并入/表示在不同的网页浏览器中。

[0109] 由此,例如,继续用不同的网页浏览器从相同的网页浏览器状态浏览是可能的。这对于在不同的用户和/或网页机器人之间委派浏览任务以及对于开发网页机器人是尤其有利的。

[0110] 在本公开的发明性概念的方面,方法包括基于外部表示在不同的网页浏览器应用中建立多个进一步的内部表示的步骤。

[0111] 以此方式,多个网络机器人和/或用户可随时间继续从相同的网页浏览器浏览一次或多次。

[0112] 而且,网页浏览器可制作得更加有效,其中它们可促进循环以从网页浏览器状态跟随不同的路径,并且当路径被网络机器人或用户充分检查时,可以再引入外部状态表示以重建先前的网页浏览器状态。

[0113] 在方面,网页的网页处理可包括在浏览网页期间处理网页的动态内容。

[0114] 在方面,在这样的方面,网页浏览器通常可包括脚本引擎,诸如java脚本引擎,其可配置为处理网页的脚本内容。在此期间,动态内容的对象到内部DOM中其他对象诸如与标记内容和/或样式表内容相关的对象之间的指针,可由网页浏览器建立和/或更新。

[0115] 在优选的方面,网页的动态内容包括网页的脚本。

[0116] 动态网页通常包括例如客户端侧要被网页浏览器执行的诸如例如脚本的多个动态内容。这样的动态内容在网页浏览器中表示期间随时间适应。

[0117] 例如,动态内容可以是JavaScript(或另一种脚本语言)内容,其响应于通过用户浏览网页实施的鼠标、键盘和/或触摸屏动作修改网页表示和/或表现,这可以在指定的定时事件下等等。例如,当用户“滚动”所呈现的网页时,脚本可设置成自动收集附加的内容和/或自动更新网页。

[0118] 在方面,网页的动态内容还可包括服务器端脚本的结果的内容。服务器端的脚本内容理解为网页的内容基于在网站服务器上运行的一个或多个程序被控制或更新,所述程序用于改变网页内容和/或网页上的外观。这些内嵌的脚本导致在服务器响应客户端的请求之前,由一个或多个在网站服务器上运行的脚本处理客户端对服务器网站的请求。例如,服务器端的脚本可接收用户已经输入到屏幕上呈现给用户的输入字段的数据和/或在用户浏览期间建立的数据。然后此数据通过网页的内嵌的脚本被传输到服务器端的脚本,其处理数据并根据结果将数据返回到浏览器。

[0119] 在有利的方面,包括第一内部表示的网页浏览器将第一内部浏览器状态表示外在化成外部表示,优选地利用串行化方式用于串行化第一内部表示的至少一部分。

[0120] 在方面,用于进一步的内部表示的网页浏览器应用可建立进一步的内部表示,优选地利用解串行化方式用于解串行化外部表示的至少一部分。

[0121] 因此解串行化的内容可在解串行化后利用网页浏览器应用被解析成内部DOM。

[0122] 网页浏览器包括网页的内部状态表示,其由网页、用户交互、动态内容等等指定。因此,通过将网页浏览器配置为建立外部表示和/或进一步的内部表示,网页浏览器可能访问其内部表示并且在外在化成外部状态表示的时候外在化其网页浏览器的精确状态。

[0123] 可替代地,在网页浏览器外部的装置能够拷贝网页浏览器状态诸如网页浏览器的内部DOM、形式数据、cookie数据等的拷贝,并且然后接着通过处理所拷贝的数据建立外部表示。

[0124] 在本公开的发明性概念的有利方面,当符合预定的标准时网络机器人可配置为开始建立一个或多个外部表示,和/或网络机器人可配置为基于外部表示建立一个或多个进一步的内部表示。

[0125] 而且现在公开的发明性概念涉及用于建立网页浏览器的第一内部浏览器状态表示的外部表示并且在此基础上建立进一步的内部表示的系统,该系统包括网页浏览器,其配置为将第一内部浏览器状态表示外在化成状态的外部表示,其中建立状态的外部表示包括建立第一内部浏览器状态表示的第一内容的表示,建立表示第一内部浏览器状态表示的动态内容的状态的动态内容状态表示,以及建立包括映射第一内部浏览器状态表示的动态内容和第一内容之间的相互关系的映射数据,并且其中网页浏览器应用配置为在与第一内部浏览器状态表示的状态相对应的浏览器状态下基于外部状态表示建立进一步的内部状态表示。

[0126] 在上述系统的方面,该系统配置为根据权利要求1到35的任一项的方法进行操作。

[0127] 此外,现在公开的发明性概念涉及使用根据权利要求1到37的任一项的方法和/或系统连同—个或多个网络机器人,其中—个或多个网络机器人配置为在与第一内部浏览器状态表示的状态相对应的状态下开始建立网页浏览器应用中网页的进一步的内部表示,并且其中进一步的内部表示基于外部状态表示被建立。

[0128] 而且,现在公开的发明性概念涉及进一步使用方法、系统和/或使用根据权利要求1到38的任一项的用于在不同的网页浏览器应用中建立多个浏览事件。

[0129] 在进一步使用的方面,多个浏览事件可以在—个或多个服务器上执行。因此在方面服务器可承载服务器执行的网络机器人。这提供了用于提供快速并且高效浏览网页的高效并且有效的解决方案,优选地通过—个或多个网络机器人大致同时进行。

[0130] 此外,现在公开的发明性概念涉及用于建立外部浏览器状态表示的第一网页浏览器,所述网页浏览器包括用于建立第一内部浏览器状态表示的第一内容的外部表示的装置,用于建立表示第一内部表示的动态内容的状态的外部动态内容状态表示的装置,以及用于建立包括映射第一内部浏览器状态表示的动态内容和第一内容之间的相互关系的映射数据的装置。

[0131] 在方面术语“装置”通常也可视为布置或应用。

[0132] 在方面,第一网页浏览器可配置为根据权利要求1到35的任一项的方法进行操作。

[0133] 而且,现在公开的发明性概念涉及用于基于预建立的外部浏览器状态表示建立进一步的内部浏览器状态表示的第二网页浏览器应用,该网页浏览器应用包括:用于处理预建立的外部浏览器状态表示的装置,用于将处理结果解析成网页浏览器应用的内部的文档对象模型的装置,以及用于处理外部浏览器状态表示的映射数据并基于映射数据实现表示在内部的文档对象模型中的动态内容和第一内容之间的相互关系的装置。

[0134] 在第二网页浏览器的方面,第二网页浏览器可配置为根据权利要求1-35的任一项的方法进行操作。

[0135] 上面公开的网页浏览器可视为包括用于处理网页的计算机可读的计算机程序产

品。用于建立并操控外部表示的上述装置以外的网页浏览器包括用于处理网页的附加的传统装置,诸如解析器、布局引擎、数据通信层等,以帮助浏览网页。

[0136] 此外,现在公开的发明性概念涉及浏览基于网页的数据源的方法,所述数据源包括多个网页,该方法包括如下步骤:利用网页浏览器浏览基于网页的数据源的一个或多个网页,在浏览期间利用网页浏览器建立不同的第一内部浏览器状态的至少两个外部表示,基于至少一个外部表示在一个或多个网页浏览器应用中建立一个或多个附加的浏览事件,建立一个或多个附加的浏览事件包括处理外部表示的内容并且实施解析成一个或多个网页浏览器应用的内部状态表示。

[0137] 在浏览基于网页的数据源的方法方面,根据权利要求1-35的任一项的方法和/或利用根据权利要求36-37的任一项的系统,至少两个外部表示可建立和/或实现到一个或多个网页浏览器应用中。

[0138] 在浏览基于网页的数据源的方法方面,该方法通过网页机器人实施。

[0139] 可替代地,在现在公开的发明性概念的方面,用户可实施浏览的至少一部分。在机器人实施浏览的一部分的情况下,这部分浏览不必在计算机屏幕上监测。

[0140] 此外,现在公开的发明性概念涉及使用网络机器人用于利用网页浏览器浏览网页,其中网络机器人按照预定的规则集进行操作,以便处理一个或多个网页,并且其中网络机器人配置为当符合预定标准时根据权利要求1-35的任一项开始建立一个或多个外部表示。

[0141] 在方面,预定标准可涉及关于例如处理网页的标准,例如当网页的网页浏览器处理满足一个或多个标准时,诸如关于所执行的脚本内容、所执行的经解析内容的状态,样式表信息的状态等等的标准,这些标准可触发建立外部表示,例如根据在此文件中所描述的一个或多个方法。

[0142] 而且现在公开的发明性概念可涉及建立表示在网页浏览器中的网页的状态表示的进一步方法,该方法包括下列步骤:利用网页浏览器对从基于网页的数据源检索到的网页执行网页处理,以及在网页浏览器中建立网页的作为结果的第一内部浏览器状态表示,在网页处理期间建立第一内部表示的状态的外部表示,其中建立状态的外部表示包括:建立第一内部浏览器状态表示的第一内容的表示,建立表示第一内部浏览器状态表示的动态内容的状态的动态内容状态表示,以及建立包括映射第一内部浏览器状态表示的动态内容和第一内容之间的相互关系的映射数据,以及将外部表示存储在数据存储中。

[0143] 上面的进一步方法的此实施例可结合关于例如图1-11所描述的任何合适的实施例。

[0144] 通常理解为,建立外部表示和/或使用外部表示用于重建浏览器状态可以是计算机实现的,以便大体上自动生成外部状态表示或根据来自例如用户或网络机器人的请求使用/实现外部状态表示。

[0145] 在一个一般方法中,建立表示在网页浏览器中的网页的状态表示的方法包括利用网页浏览器实施下列操作:对从基于网页的数据源检索到的网页执行网页处理;在网页浏览器中建立作为结果的网页的第一内部浏览器状态表示;建立外部表示;外部表示表示第一内部表示的状态;其中建立状态的外部表示包括:建立第一内部浏览器状态表示的第一内容的表示;建立表示第一内部浏览器状态表示的动态内容的状态的动态内容状态表示;

以及建立包括映射动态内容和第一内容之间的相互关系的映射数据;以及使得外部表示可用于网页浏览器应用,以便在与第一内部浏览器状态表示的状态相对应的状态下在网页浏览器应用中建立网页的进一步的内部表示。

[0146] 在另一个一般实施例中,系统配置为建立网页浏览器的第一内部浏览器状态表示的外部表示以及在此基础上建立进一步的内部表示。该系统包括:网页浏览器,配置为将第一内部浏览器状态表示外在化成状态的外部表示,以及网页浏览器应用,配置为在与第一内部浏览器状态表示的状态相对应的浏览器状态下基于外部状态表示建立进一步的内部状态表示。建立状态的外部表示包括:建立第一内部浏览器状态表示的第一内容的表示;建立表示第一内部浏览器状态表示的动态内容的状态的动态内容状态表示;以及建立包括映射第一内部浏览器状态表示的动态内容和第一内容之间的相互关系的映射数据。

[0147] 在又一个一般实施例中,网页浏览器配置为通过如下建立外部浏览器状态表示:建立第一内部浏览器状态表示的第一内容的外部表示;建立表示第一内部表示的动态内容的状态的动态内容状态表示;以及建立包括映射第一内部浏览器状态表示的动态内容和第一内容之间的相互关系的映射数据。

[0148] 在附加的一般方法中,网页浏览器配置为基于预建立的外部浏览器状态表示通过如下方法建立进一步的内部浏览器状态表示:处理预建立的外部浏览器状态表示,将处理结果解析成网页浏览器应用的内部的文档对象模型,以及处理外部浏览器状态表示的映射数据,以及基于映射数据实现表示在内部的文档对象模型中的动态内容和第一内容之间的相互关系。

[0149] 在再一个一般实施例中,方法是针对浏览包括多个网页的基于网页的数据源。该方法包括,利用网页浏览器:浏览基于网页的数据源的一个或多个网页;在浏览期间建立不同的第一内部浏览器状态的至少两个外部表示;至少部分基于处理外部表示的内容在一个或多个网页浏览器应用中建立基于至少一个外部表示的一个或多个附加的浏览事件;以及实施解析成一个或多个网页浏览器应用的内部状态表示。

[0150] 在优选的方法中,本文公开的发明性实施例可包括任何数目的下列功能、特征和/或组件:

[0151] 建立表示在网页浏览器中的网页的状态表示可包括,利用网页浏览器:对从基于网页的数据源检索到的网页执行网页处理;在网页浏览器中建立作为结果的网页的第一内部浏览器状态表示;建立外部表示;外部表示表示第一内部表示的状态;其中建立状态的外部表示包括:建立第一内部浏览器状态表示的第一内容的表示;建立表示第一内部浏览器状态表示的动态内容的状态的动态内容状态表示;以及建立包括映射动态内容和第一内容之间的关系的映射数据;以及使得外部表示可用于网页浏览器应用,以便在与第一内部浏览器状态表示的状态相对应的状态下在网页浏览器应用中建立网页的进一步的内部表示。建立外部状态表示可包括串行化第一内部浏览器状态表示的至少一部分。第一内部浏览器状态表示的该部分可包括动态内容和第一内容中的一个或多个的至少一部分。建立进一步的内部表示可包括解串行化外部表示的至少一部分。建立进一步的内部表示可包括将至少一部分的外部表示解析到网页浏览器应用。可建立外部表示以独立于一个或多个存储器位置,该一个或多个存储器位置与第一内部表示相对应。建立外部表示可包括多个识别框架,该多个识别框架包括网页的一个或多个内联框架。可以建立两个或多个框架的每一个的外

部框架表示,两个或多个框架的至少一个包括至少一个网页的内联框架。建立外部表示可包括建立框架关联表示。外部表示可基于网页的内部文档对象模型表示在网页浏览器中建立。外部表示可通过串行化内部文档对象模型的内容的至少一部分来建立。建立进一步的内部表示可包括:解串行化外部表示的至少一部分;并且将所解串行化的该部分外部表示解析成网页浏览器应用的内部文档对象模型。第一内部浏览器状态表示的第一内容可包括标记内容,并且外部表示可包括标记内容的表示。内部浏览器状态表示的第一内容可包括样式表内容的表示,并且外部表示可包括该样式表内容的表示。第一内容可包括关于第一内部浏览器状态表示的cookie数据,并且外部表示可包括cookie数据的表示。第一内容可包括第一内部状态表示的形式内容,并且外部表示可包括该形式内容的表示。映射数据可包括第一内部浏览器状态表示的动态内容和标记内容之间的一个或多个关系的映射。映射数据可包括第一内部浏览器状态表示的动态内容和样式表内容之间的指针的映射。映射数据可包括映射一个或多个指针到第一内部浏览器状态表示的一个或多个对象的一个或多个存储器地址。建立映射数据可包括:处理第一内部表示;至少部分基于该处理识别对象之间的一个或多个指针;以及将指针映射到存储器地址。建立外部表示可包括将第一内容转化成一种或多种格式,一种或多种经转化的格式中的每一个都不同于内部浏览器状态表示的格式。建立外部表示可包括将第一内部浏览器状态表示的至少一部分转化成第一内部表示的内部文档对象模型的外部表示。建立外部表示可包括将第一内部浏览器状态表示的不同框架的多个内部文档对象模型转化成多个内部文档对象模型的一个或多个外部表示。第一内部浏览器状态表示可包括标记内容,并且该转换可包括将标记内容的至少一些规定为标记语言格式。第一内部浏览器状态表示可包括样式信息,并且该转换可包括将样式信息的至少一些规定为层叠样式表格式。第一内部浏览器状态表示可包括多个不同内容要素,每个不同的内容要素包括下述内容的一个或多个中的至少一部分:第一内容;以及动态内容。建立外部表示进一步可包括建立多个个体表示,每个个体表示都基于第一内部浏览器状态表示,并且每个个体表示可与多个不同内容要素的至少一个相对应。建立进一步的内部表示可包括映射网页浏览器应用中的浏览器标记语言内容和动态内容的表示之间的至少一个联系,并且该映射可至少部分基于映射数据。建立进一步的内部表示可包括建立进一步的内部表示的一个或多个对象之间的一个或多个指针,其中该建立可至少部分基于映射数据。第一内部表示和进一步的内部表示的每一个都包含在下述内容的一个或多个:至少一个网页浏览器;至少一个网页浏览器应用;以及至少一个网页。第一内部表示和进一步的内部表示的每一个都包含在下述两者或其中一个中:一个或多个网页浏览器、网页浏览器应用和网页的不同的一个;以及一个或多个网页浏览器、网页浏览器应用和网页的不同的类型。在多个网页浏览器应用中建立多个进一步的内部表示,其中进一步的内部表示可建立在其中的多个网页浏览器应用的每一个都可不同于进一步的内部表示可建立在其中的所有其他网页浏览器应用。网页的网页处理可包括在浏览操作期间处理网页的动态内容,该浏览操作在网页上操作。网页的动态内容可包括网页的脚本。包括第一内部表示的网页浏览器至少部分基于串行化至少一部分第一内部表示将第一内部浏览器状态表示外在化成外部表示。用于进一步的内部表示的网页浏览器应用可至少部分基于解串行化至少部分外部表示建立进一步的内部表示。附加地和/或可替代地,并且优选地利用网络机器人,帮助以下步骤是有利的:确定是否满足一个或多个预定的标准;以及至少部分基于确定满

是一个或多个的预定标准开始建立外部表示。网络机器人可配置为基于外部表示建立一个或多个进一步的内部表示。系统可配置为建立网页浏览器的第一内部浏览器状态表示的外部表示以及在此基础上建立进一步的内部表示。该系统包括：网页浏览器，配置为将第一内部浏览器状态表示外在化成状态的外部表示，其中建立状态的外部表示可包括：建立第一内部浏览器状态表示的第一内容的表示；建立表示第一内部浏览器状态表示的动态内容的状态的动态内容状态表示；以及建立包括映射第一内部浏览器状态表示的动态内容和第一内容之间的相互关系的映射数据。在系统内，网页浏览器应用可配置为在与第一内部浏览器状态表示的状态相对应的浏览器状态下基于外部状态表示建立进一步的内部状态表示。该系统可配置为按照前述方法进行操作，可选地结合一个或多个网络机器人，并且一个或多个网络机器人配置为在与第一内部浏览器状态表示的状态相对应的浏览器状态下在网页浏览器中开始建立网页的进一步的内部表示。进一步的内部表示可基于外部状态表示而建立。这些概念对于在一个或多个不同的网页浏览器应用中建立多个浏览事件是有用的，即使多个浏览事件在一个或多个服务器上执行。网页浏览器可配置为建立外部浏览器状态表示。网页浏览器包括：用于建立第一内部浏览器状态表示的第一内容的外部表示的装置，用于建立表示第一内部表示的动态内容的状态的外部动态内容状态表示的装置，以及用于建立包括映射第一内部浏览器状态表示的动态内容和第一内容之间的相互关系的映射数据的装置。网页浏览器应用可配置为基于预建立的外部浏览器状态表示建立进一步的内部浏览器状态表示。该网页浏览器应用包括：用于处理预建立的外部浏览器状态表示的装置，用于将处理结果解析成网页浏览器应用的内部的文档对象模型的装置，以及用于处理外部浏览器状态表示的映射数据并基于映射数据实现表示在内部的文档对象模型中的动态内容和第一内容之间的相互关系的装置。浏览包括网页的基于网页的数据源的方法包括：利用网页浏览器浏览基于网页的数据源的一个或多个网页，在浏览期间建立不同的第一内部浏览器状态的至少两个外部表示，基于至少一个外部表示在一个或多个网页浏览器应用中建立一个或多个附加的浏览事件，建立一个或多个附加的浏览事件，其包括处理外部表示的内容并且实施解析成一个或多个网页浏览器应用的内部状态表示。该至少两个外部表示可按照前述方法和/或系统建立和/或实现到一个或多个网页浏览器应用中。再者，该方法可由网络机器人实施，尤其是利用网页浏览器完成浏览网页。网络机器人按照预定的规则集进行操作，以便处理一个或多个网页，并且网络机器人配置为响应于确定符合一个或多个预定标准根据前述方法和/或系统的任一个开始建立一个或多个外部表示。

[0152] 目前公开的发明性概念所呈现的附加的特征、功能、组件和优点将参考附图呈现。下面的说明要理解为示例性的，并且所呈现的任何特征都可以以技术人员在阅读本公开后将领会的任何合适的方式进行组合、替换。

[0153] 图1例示了根据现在公开的发明性概念的实施例的系统。该系统包括第一客户端C1和第二客户端C2。此示例中的客户端是计算单元，诸如台式电脑、膝上型电脑、平板电脑等等，但要理解，在现在公开的发明性概念的进一步实施例中，包括一个或多个网络机器人的服务器也可视为客户端。客户端C1、C2包括一个或多个数据处理器/CPU、数据存储、电路板、屏幕，并且可包括输入装置，诸如键盘（硬件或软件（经由触摸屏））、计算机鼠标等等。

[0154] 而且，客户端C1、C2包括网页浏览器装置WB，WBA。这些网页浏览器装置是适于访问、处理和呈现万维网上的网页WP1-WPn的软件应用，也叫做互联网IN。这些网页WP1-WPn位

于基于网页的数据源WBDS上。

[0155] 因此网页浏览器WB、WBA可包括任何使能网页浏览器可从基于网页的数据源WBDS访问、检索并解释网页的合适的装置。这些装置可例如包括专用计算机软件,其解释并执行JavaScript,例如java脚本引擎,它可包括一个或多个布局/渲染引擎,用于操控标记内容诸如HTML、XML、图像文件等等,以及格式化信息诸如层叠样式表(CSS)、可扩展样式表语言(ESL)等等。

[0156] 基于网页的数据源WBDS也可称为网页服务器。因此,这些网页可通过互联网IN访问,包括网页浏览器应用的客户端或网页服务器包括合适的数据通信装置,用于通过互联网与例如基于网页的数据源WBDS通信。

[0157] 客户端的用户U(或网络机器人)利用网页浏览器WB访问网页WP1-WPn,例如通过进入浏览器中指代网页的网页地址/统一资源定位符。然后该网页浏览器WB在URL中检索网页WP1并且在客户端C1、C2的屏幕上可视化网页。这优选地通过对从基于网页的数据源WBDS检索到的网页WP1实施网页处理、并且在网页浏览器中建立网页WP1例如网页WP1的内部浏览器表示FIR、FUIR来完成。

[0158] 这通过处理检索到的网页WP1并且将网页的内容解析成网页浏览器WB中的内部的文档对象模型IDOM来完成。用户随后可利用浏览器来浏览网页WP1。

[0159] 网页WP1包括例如标记语言内容和样式表内容形式的静态内容。标记语言内容可用任何合适的标记语言来写,但是通常用超文本标记语言(HTML)。

[0160] 样式表内容可包括层叠样式表(CSS),其包括关于应如何在客户端C1、C2的屏幕上将标记语言内容可视地呈现给用户的信息。

[0161] 而且,多数网页包括例如脚本内容诸如JavaScript内容或类似内容形式的动态内容。此脚本内容可响应于鼠标或键盘动作或用户和/或网页本身实施的其他动作而改变网页的行为。该动态内容可包括客户端侧的脚本和/或用于帮助服务器端的脚本的装置。

[0162] 网页的动态和“静态”内容由网页浏览器处理并且利用网页浏览器WB中除了内部的文档对象模型表示IDOM以外的导致网页浏览器中的上述作为结果的第一内部浏览器状态表示FIR。内部的文档对象模型表示IDOM在下面也可称为“内部DOM”。

[0163] 然而由于用户U实施的动作和/或由于利用IDOM自动修改网页表示的脚本内容,内部DOM随时间被修改。所以由于与网页的不同用户交互和/或网页的脚本随时间变化使得在不同时间呈现的网页看起来和/或表现的不同,因此在不同的客户端但基于相同的网页WP所建立的内部DOM IDOM看起来是不同的。因此,当根据现在公开的发明性概念的实施例建立表示在网页浏览器WB中的网页WP1-WPn的状态表示时,必须考虑此点。根据现在公开的发明性概念的实施例,这可由例如下面的方法完成。

[0164] 用户U(或网络机器人)可希望促进稍后在另一网页浏览器(WBA)或相同网页浏览器中继续网页WP1的第一内部表示的特定状态。根据现在公开的发明性概念的实施例,用户U(或网络机器人)随后开始在网页处理期间的某点建立和存储第一内部表示FIR的状态的外部表示ER。这可通过利用计算机鼠标或在特定位置触摸触摸屏点击客户端屏幕上的按钮、利用键盘输入命令等来实施。

[0165] 网络机器人可限定为通过互联网运行自动任务的软件应用。典型地,网络机器人以比人类单独执行所可能的更高的速率执行任务。网络机器人包括预定的规则集和/或指

令集,其被设计以便从特定网页或某类型网页访问、处理和检索内容。因此,当被实现并运行时,机器人可设计为利用浏览器访问网页,并且从网页检索数据以例如存储和/或索引通过网页可访问的数据库内容,从而实施网页(或利用网页可访问的数据库)内容的分析等等。因而机器人可由用户基于网页或某类型的网页创建,这样机器人将能够自动访问网页并访问意图由设计机器人的用户所访问的数据。

[0166] 因此可设计网络机器人,以便当在浏览器中符合特定标准例如达到浏览器的特定状态时,机器人开始建立外部表示。例如,当执行了网页的脚本,并且处理了网页时,这样机器人所要求的数据是可访问的,可建立外部表示,以便可在其他或另外的网页浏览器中建立一个或多个进一步的内部表示,或以便机器人稍后可返回内部表示,其是在与建立外部表示的时间的浏览器状态相对应的状态。而且这稍后在此文档中描述。

[0167] 根据现在公开的发明性概念的实施例在利用网络机器人的情况下,网络机器人可配置为按照规则集进行操作,其中,当网页浏览器符合标准集、达到预定的浏览器状态时,可调用一个或多个网络机器人API方法,以便建立如现在公开的发明性概念的不同实施例所描述的外部表示ER和/或处理网页数据。因而在现在公开的发明性概念的方面,机器人可利用不同的常规方法以便在网页中导航,但可另外包括使能机器人以建立如本文档的不同实施例公开的外部表示的整合功能。因而机器人能够利用网页浏览器例如加载网页、提取URL信息、提取图像、提取文本等。

[0168] 开始建立外部表示ER开始了处理网页浏览器WB的内部DOM表示IDOM,以便建立外部表示ER,外部表示ER包括第一内部表示FIR的第一内容FC的表示RFC。第一内容FC包括例如标记内容的内部浏览器表示以及样式表内容的内部浏览器表示。而且在实施例中,其包括cookie的内部浏览器表示和/或关于网页浏览器中所呈现的网页的形式内容的内部浏览器表示。

[0169] 而且,建立外部动态内容状态表示DCSR,其表示第一内部表示(FIR)的动态内容DC的状态。建立动态内容状态的上述外部表示DCSR以及第一内容RFC可包括处理各自的内部表示并且将该内容转化成预定的数据格式。

[0170] 例如,内部浏览器表示的标记内容可以以标记语言格式诸如可扩展标记语言(XML)或类似语言存储在外部表示中。

[0171] 样式表内容可以以样式表格式诸如层叠样式表(CSS)或类似样式表存储。

[0172] cookie信息可存储在文本文档或类似包括表示cookie的文本串的文档中,并且形式内容也可以文档格式诸如标记语言例如XML来存储。

[0173] 动态内容的表示可以以二进制格式存储,并且包含动态内容的状态,并且优选地也包含动态内容本身。

[0174] 一般理解为,在现在公开的发明性概念的实施例中,外部表示可建立在网页浏览器外面。因此,它可存储在客户端C1的数据存储器中(后面会更加详细解释),可存储在客户端C1外面的数据存储器中等等。

[0175] 而且,第一内部状态表示FIR可包括从动态内容到包含内部表示的标记语言内容和/或样式表内容的变量值的计算机存储器地址的多个指针。这些指针对于映射是必需的,以便帮助在另一网页浏览器应用WBA中建立相应的浏览器状态。因此,根据现在公开的发明性概念,建立映射数据MD,其包括动态内容DC和第一内容FI的至少一部分诸如第一内部表

示FIR的标记语言内容之间的相互关系的映射。因此,此映射数据MD包括关于动态内容DC和第一内部表示FIR的第一(静态)内容FC如何相关以及因此相应的内容在进一步的内部表示FUIR中应如何相关的信息。

[0176] 外部表示ER存储在客户端C1、客户端外面的网页服务器或任何其他合适的位置的数据存储中。外部表示ER甚至可拷贝到多个外部表示(未在图1中例示)中并且这些外部表示的每一个都可用于在多个网页浏览器应用WBR中建立一个或多个进一步的内部表示。

[0177] 因而外部表示ER可用于(如后面更加详细描述)在与网页浏览器WB的第一内部状态相对应的状态下建立网页浏览器应用WBA的进一步的内部浏览器表示FUIR。

[0178] 根据现在公开的发明性概念的实施例,处理外部表示并在此基础上在网页浏览器应用中建立进一步的内部DOM表示FIDOM。这可通过处理外部表示ER来实施,所述外部表示ER以如下形式:第一内容FC的外部表示RFC、外部动态内容状态表示DCSR、以及映射数据。

[0179] 当在网页浏览器应用WBA中建立了进一步的内部浏览器表示FUIR时,用户可以随后从与第一内部表示的状态相对应的状态开始继续浏览网页WP。

[0180] 图2例示了根据现在公开的发明性概念的实施例的、关于建立外部表示ER的实施例的流程图。

[0181] 在步骤21(检索基于网页的数据源的网页),网页浏览器从基于网页的数据源WBDS检索网页WP,例如由于用户在浏览器中输入URL或激活浏览器中的链接或文档中的超链接。

[0182] 在步骤22(将网页解析到网页浏览器),浏览器处理检索到的网页并且分析将网页解析成网页浏览器WB的内部DOM表示。

[0183] 现在用户U可通过点击链接和网页的标签来浏览网页WP(步骤S23),用户U可输入网页的形式内容等等。在浏览网页WP期间的某点处,用户U可希望在另一浏览器或在相同网页浏览器的标签内建立网页浏览器WB的状态。

[0184] 因此,用户U可如上描述的发起外部表示ER的建立。浏览器(或另一工具)因此将浏览器WB的状态外在化。通过将网页浏览器的内部状态表示串行化成共同表示浏览器状态的一个或多个外部表示(步骤S24)为其提供帮助。因此浏览器WB将内部表示FIR的第一内容FC串行化为外部表示ER。该串行化包括将内部表示FIR的数据结构转化成要存储在数据存储器中的文件等等中的格式。可替代地,在方面,可立即传输经串行化的内容以用于在另一网页浏览器应用或标签中建立进一步的内部表示FUIR,而不用立即存储经外在化的数据。

[0185] 串行化优选地还包括在建立外部表示ER时在外部表示ER中获得内部表示的对象的值。

[0186] 该对象表示一段数据并且通常在指派给对象的存储器位置指派给其值。此/这些值优选地拷贝给外部表示ER以便在后面进一步的内部表示中重新插入该值。

[0187] 然后经串行化的内容稍后可在相同的客户端C1或在另一客户端或网页服务器上重建。

[0188] 而且内部表示FIR被处理(步骤S25)以便确定内部表示FIR的动态内容DC和第一内容FC之间的引用/关系。这通过建立映射数据MD实施。在现在公开的发明性概念的优选实施例中,如后面更加详细描述,该映射数据MD可包括关于将指针或引用映射到在客户端C1或网页服务器处的网页浏览器WB的第一内部状态表示FIR中使用的变量的存储器地址的信息。用此方式,确定动态内容DC和第一内容FC之间的引用/关系是可能的,所述第一内容FC

诸如内部表示FIR的标记语言相关内容和/或样式表相关内容。

[0189] 然后结合步骤S24和S25建立的外在化的数据被存储在数据存储中S26,用于后面建立进一步的内部浏览器状态表示FUIR使用。

[0190] 然后用户可继续浏览网页,已知建立网页浏览器的较早的状态是可能的,并且此外可在进一步浏览网页期间建立网页浏览器的进一步的第一状态表示FIR。

[0191] 图3例示了根据现在公开的发明性概念的实施例的、关于基于外部表示ER在网页浏览器中建立进一步的内部状态表示FUIR的实施例的流程图。

[0192] 在步骤S31,检索表示网页浏览器WB的先前状态的外部表示ER。

[0193] 在步骤S32,外部表示ER的至少一部分被解串行化并被转化成意图被用于进一步的内部表示的网页浏览器WBA的解析器能够解析(步骤S33)成网页浏览器WBA的进一步的内部文档对象模型FIDOM的形式。

[0194] 对象在外在化第一内部表示FIR时呈现的值优选地也读到进一步的内部表示FUIR中。

[0195] 在步骤34(处理并实现映射数据),处理并实现外部表示ER的映射数据MD以完成重建第一内部表示FIR的状态。以此方式,在进一步的内部表示FUIR中重建动态内容DC和第一内容FC之间的关系,该关系基于外部表示ER的信息。这可例如包括建立和/或修正指针,使得进一步的内部表示FUIR的动态内容DC和第一内容FC之间的指针被设置为大致与第一内部表示FIR的指针相对应。然而,由于存在差异,进一步的内部表示FUIR的指针/引用被设置为指向相对于第一内部表示FIR的存储器地址的其他存储器地址。

[0196] 在步骤35(从浏览器浏览网页),用户、网络机器人等等可因此从与第一内部表示FIR的状态相对应的浏览器状态使用进一步的内部表示FUIR来浏览网页WP。

[0197] 图4示意性例示了根据现在公开的发明性概念的映射指针的实施例。

[0198] 网页浏览器WB的第一内部表示FIR包括多个引用,其以网页浏览器WB的第一内部表示的动态内容的对象DOB1-DOBn到包含第一内容FC的对象OB1-OBn的值的存储器地址M_{ADR}之间的指针的形式。

[0199] 动态内容的这些对象和指针通过JavaScript引擎用内部二进制表示表示。此java脚本引擎可以用编程语言实现,诸如例如C++等,其与以指针形式的参考对象一起操作,所述参考对象是其值利用存储器地址直接引用/指向存储在计算机存储器的其他地方的另一值的数据类型。

[0200] 动态内容的对象DOB1-DOBn可以是脚本对象,诸如JavaScript对象或其他类型的对象。在本示例中,这些对象相比第一内容OB1-OBn的对象(这些被分配到存储器的0xXXZZ1-0xXXZZn处)被自然放置在其他存储器地址(0xXXZZ10-0xXXZZz)处。

[0201] 此实施例中的第一内容FC包括第一内部表示的标记语言内容。该标记语言内容MLC优选地是在标记语言部分例如网页WP的HTML内容被解析到网页浏览器WB期间已建立的内容。然而在进一步的实施例中,第一内容FC还可包括在网页被解析到网页浏览器WB期间已建立的样式表内容。

[0202] 如箭头所指示的,动态内容的第一对象DOB1指向/引用存储器地址0xXXZZ1并且以此获得第一内容FC的第一对象OB1的值,动态内容的第二对象DOB2指向/引用存储器地址0xXXZZ2并且以此获得第一内容FC的第二对象OB2的值等等。注意到,没有指针从动态内容

指向第一内容FC的第五对象OB5。然而,在实施例中,处理此对象并以使能在进一步的内部表示中建立相应的对象为目的为其添加标签是有利的。然而,指派给进一步的内部表示中的这样的对象的存储器位置并不必是重要的。

[0203] 现在,当如上描述的要建立外部表示ER时,映射指针形式的上述关系。优选地这通过处理内部表示FIR的动态内容以便识别动态内容的对象DOB1-DOBn和标记语言内容MLC的对象OB1-OBn之间的指针来完成。如果指针被识别,则在各自的存储器地址M_ADR处具有值的对象OB1-OBn被识别,并且因此建立映射数据MD以便包含此信息。

[0204] 因此外部表示ER的映射数据MD将包括动态内容DC的第一对象DOB1指向第一内容FC的第一对象OBJ1的存储器地址的信息,动态内容DC的第二对象DOB2指向第一内容FC的第二对象OBJ2的存储器地址的信息等等。而且,第一内容和第二内容的对象优选地在串行化期间用标识符添加标签,以便后面在实现进一步的内部表示FUIR期间使能识别对象。这可以在例如如后面更加详细解释的串行化期间完成。

[0205] 现在,一旦建立进一步的内部表示FUIR,外部表示的第一内容RFC和动态内容DCSR的表示被实现到网页浏览器应用WBR中。这优选地通过解串行化并将表示RFC、DCSR的至少一部分解析成网页浏览器应用的内部DOM来完成,并且它还包括实现在建立外部表示期间从第一内部表示FIR中提取的进一步的内容,诸如例如浏览历史数据、cookie数据、形式数据等等(未在图4中例示)。

[0206] 在此期间,第一表示FIR的第一内容和动态内容的必要对象OB1-OBn DOB1-DOBn的存储器将被分配到客户端(未在图4中例示)的数据存储器的存储器地址FM_ADR,所述客户端包括用于进一步的内部表示FUIR的网页浏览器应用WBA。这些存储器地址FM_ADR可能会不同于如所指示的第一内部表示FIR的地址M_ADR。

[0207] 因此,第一内容FC的第一对象OB1被分配到存储器地址0xXXZZ30,第一内容FC的第二对象OB2被分配到存储器地址0xXXZZ31等等。动态内容的第一对象DOB1被分配到存储器地址0xXXZZ40,动态内容的第二对象DOB2被分配到存储器地址0xXXZZ50等等。

[0208] 此后,处理映射数据MD。基于此映射数据,重建动态内容DC和标记语言内容MLC之间的指针。在初始解析表示第一内部表示的动态内容以及标记语言内容和样式表内容的外部表示ER的内容以后,实施实现关于对象之间的关系/指针的经映射数据。这是因为,在此解析之后,对象的存储器被分配,并且因而重建用于进一步的内部表示的指向正确存储器地址的指针是可能的。

[0209] 例如,映射数据包括动态内容的第一对象DOB1应指向第一对象OB1的存储器地址的信息。因此,识别第一对象OB1的存储器地址(在此案中其为0xXXZZ30),并且然后在动态内容的第一对象DOB1和此存储器地址之间建立指针。对余下的对象DOB2-DOBn实施相应的处理,以便建立与第一表示FIR的指针相对应的指针,但是在解串行化并解析到网页浏览器应用WBA以后,存储器地址是基于进一步的内部表示的存储器地址。

[0210] 图5例示了通过识别指针以及给对象添加标签来建立外部表示的实施例。

[0211] 在步骤S51,用户U或网络机器人浏览主页。在某一点,用户U或网络机器人希望建立外部表示ER(EER=建立外部表示)。在感测应建立浏览器状态的外部表示中,如果检测TE51(EER?)是肯定的,则这发起包括多个步骤的串行化处理。

[0212] 因此,网页浏览器WB开始串行化第一内部表示FIR的标记语言内容和动态内容DC。

这通过处理网页浏览器的内部DOM IDOM依靠识别内部表示的对象(检测TE52(OBJ?))和它们的关系来实现。如果对象被识别,为了后面的识别(步骤S52)给其添加标签,拷贝其在相应从存储器位置的值等等。

[0213] 而且,如果第一内部表示FIR的动态内容(图4的DOB1-DOBn)的对象被识别,则检测(检测TE53(?))其是否包括指向第一内容FC的另一存储器地址的指针。如果包含,开始步骤S53,其中具有被指向/引用的存储器地址的第一内容(FC)的对象被识别并映射,这样映射数据MD包括应从动态内容DC的各自对象指向第一内容FC的对象建立指针或引用形式的关系的信息。在现在公开的发明性概念的实施例中,此信息可基于在串行化期间为对象加标签来建立,以便后面容易识别各自的对象。然后,网页浏览器如上所述的继续识别对象以识别对象并建立动态内容DC和第一内容FC的外部表示。如果对象不包括指针(TE53),网页浏览器继续识别进一步的对象(TE52)。

[0214] 进行此步骤,直到第一内部表示FIR的动态内容DC的对象(DOB1-DOBn)的所有相关对象和第一内容都被处理(测试TE54是否所有相关对象已被处理(AOP?))。

[0215] 在步骤S52和S53中的处理期间,建立内部文档对象模型IDOM的外部表示ER,这样作为结果的表示ER包括与网页浏览器的内部DOM结构相对应的数据。这可通过利用例如对象处理期间的xpath创建XML表示来促进。

[0216] 而且现在外部表示包括映射数据MD,其包括在外部表示建立时第一内部表示FIR中的对象之间的引用/指针的信息,以及表示网页浏览器中的第一内部表示FIR的标记语言内容MLC的数据。

[0217] 图6例示了根据现在公开的发明性概念的实施例的、促进建立外部表示ER以及实现外部表示ER以建立进一步的内部表示的网页浏览器。

[0218] 网页浏览器WB包括常规浏览器装置诸如例如用于将网页文档解析成以内部DOM(文档对象模型)IDOM表示(未例示)的内部表示的解析装置PAR。在此解析期间,对象都被分配到内部IDOM中。网页浏览器WB还包括其他常规功能诸如用户界面UI、浏览器引擎BE、用于推动网页浏览器WB可从/向网页服务器接收和传输数据等的通信设备DCF、以及脚本引擎SE,诸如JavaScript引擎,其用于解析并执行主页的脚本(例如JavaScript)。

[0219] 而且,网页浏览器包括串行化装置SM,用于如上描述的串行化内部表示。网页浏览器(WB)还包括解串行化装置/布置DSM,用于如上描述的解串行化外部表示ER。

[0220] 此外,网页浏览器包括映射装置MM,用于如上描述的映射内部表示FIR中的对象之间的关系并建立映射数据。

[0221] 而且,网页浏览器可包括映射数据解释器MDI装置,用于在进一步的内部表示FUIR中实现映射数据。

[0222] 在现在公开的发明性概念的进一步实施例中,如果网页浏览器WB仅配置为建立外部表示ER(未在图6中例示)并且不配置为创建进一步的内部表示FUIR,则解串行化装置DSM和映射数据解释器MDI可省略。

[0223] 而且,在现在公开的发明性概念的实施例中,如果网页浏览器WB仅配置为基于由另一网页浏览器建立的预建立的外部表示ER(未在图6中例示)建立进一步的内部表示FUIR,则串行化装置SM和映射装置MM可省略。

[0224] 图7例示了关于基于第一内部表示FIR的状态的一个预建立的外部表示ER来建立

多个进一步的内部状态表示FUIR1-FUIRn的本公开的发明性概念的实施例。

[0225] 已如上描述地建立了外部表示ER。外部表示用于在不同的客户端C2-Cn建立多个进一步的内部状态表示FUIR1-FUIRn。因此,这些客户端C2-Cn的每一个都包括网页浏览器应用WBA1-WBAn,当基于外部表示ER建立进一步的内部表示FUIR时,其可用于分别访问表示在第一内部表示FIR(未在图7中例示)中表示的基于网页的数据源WBDS的网页,如关于例如图1所描述的。因此不同的个体用户U1-U3(或网络机器人)可从与第一内部表示FIR相对应的状态访问网页。

[0226] 因此网页浏览器应用WBA1-WBAn基于外部表示ER分别建立与第一内部表示FIR(未在图7中例示)的状态相对应的浏览器状态表示FUIR1-FUIRn。

[0227] 在未例示的现在公开的发明性概念的实施例中,客户端C1-Cn可包括两个或多个网页浏览器应用用于其每个个体进一步的内部表示FUIR。

[0228] 在未例示的现在公开的发明性概念的进一步实施例中,客户端C1-Cn可包括一个网页浏览器应用,其包括两个或多个个体进一步的内部表示FUIR。这可通过包括“标签”功能的网页浏览器应用来促进,其中一个网页浏览器可在网页浏览器WB的不同标签中浏览若干网页(或一个网页的子页)。

[0229] 图8例示了关于基于一个预建立的外部表示ER建立多个进一步的内部状态表示FUIR1-FUIRn的现在公开的发明性概念的进一步实施例。

[0230] 在此实施例中,在位于服务器S上的网页浏览器应用WBA1-WBAn中建立第一内部表示的多个进一步的内部表示。该服务器S连接至互联网IN以便个体网页浏览器应用WBA1-WBAn可在建立内部表示以后从与第一内部表示FIR(未在图8中例示)的状态相对应的状态访问各自的网页。

[0231] 因此网页浏览器应用WBA1-WBAn基于外部表示ER分别建立与第一内部表示FIR(未在图8中例示)的状态相对应的浏览器状态表示FUIR1-FUIRn。

[0232] 因此多个个体网络机器人WR1-WRn和/或用户可利用在网页服务器上的个体网页浏览器应用从与第一内部表示FIR的状态相对应的状态浏览网页。

[0233] 理解为,在实施例中,网页浏览器应用WBA1-WBAn可在两个或多个不同的服务器S(未在图8中例示)之间分配。

[0234] 图9例示了根据现在公开的发明性概念的实施例的、包括装置的客户端C1(或服务器)。该客户端可用于建立并且外在化如上关于例如图1、2、4和5描述的第一内部表示FIR和/或建立如上关于例如图1、3、4和5描述的进一步的内部状态表示FUIR。

[0235] 客户端C1包括允许用户U看到被安装在客户端C1上的网页浏览器应用WB、WBA可视化的内容的屏幕SC。而且,客户端C1包括用于根据指令处理数据的一个或多个数据处理器DP。该数据处理器DP还可称为中央处理单元CPU。

[0236] 因此,客户端C1包括一个或多个数据存储DS。此数据存储DS用于存储网页浏览器以及由网页浏览器应用WBA生成的数据,诸如第一内部表示FIR和/或进一步的内部表示FUIR。该数据存储DS可包括随机存取存储器(RAM)硬盘驱动HDD和/或固态硬盘等等。

[0237] 而且客户端C1包括用于允许用户U利用客户端C1浏览互联网的键盘KB。在屏幕是触摸屏的方面,键盘可以是安装在客户端C1上并且由用户U1访问的软件键盘应用。

[0238] 图10例示了根据现在公开的发明性概念的实施例的服务器S的实施例。该服务器S

包括用于根据指令处理数据的一个或多个数据处理器DP。该数据处理器DP还可称为中央处理单元CPU。此外,该服务器WS包括一个或多个数据存储DS。这个/这些数据存储DS用于存储一个或多个网页浏览器应用WBA并且其还可用于存储由网页浏览器应用WBA生成的数据,诸如在网页浏览器应用中生成的第一内部表示FIR(未例示)的外部表示ER和/或进一步的内部表示FUIR。该数据存储DS可包括随机存取存储器(RAM)硬盘驱动HDD和/或固态硬盘等等。理解为,数据处理器DP、数据存储DS。

[0239] 在进一步的实施例中,服务器S可包括多个网页浏览器应用WBA,其由服务器执行的网络机器人(未例示)操作/操控,参见图8以及后面的图的实施例。

[0240] 此外,服务器S可包括网页浏览器WB,其包括第一内部表示,以及用于建立外部表示的装置。一般理解为,在现在公开的发明性概念的方面,其中网页服务器包括多个网页浏览器应用,多个这些可包括用于在此文档限定的浏览期间生成外部表示ER的装置,例如如在此文档中描述的由服务器实现的网络机器人触发的。

[0241] 图11例示了现在公开的发明性概念的实施例,其中外部表示ER存储在包括网页浏览器WB的客户端C1上的数据存储中,所述网页浏览器WB包括第一内部表示FIR。然后外部表示ER可通过解串行化并实现外部状态表示ER的内容到网页浏览器WB,在后面时间点用于将网页浏览器WB恢复成与网页浏览器的第一内部装置相对应的浏览器状态以建立进一步的内部状态表示FUIR(未在图10中例示)。

[0242] 外部表示ER还可替代地存储在如关于图8所描述的网页服务器或另一服务器上,其中外部表示ER可被若干网页浏览器应用共享或可替代地外部表示可针对每个网页浏览器WB被存储,但外部表示可表示相同的第一内部表示FIR。

[0243] 图12例示了关于使用多个外部表示ER1-ER3的现在公开的发明性概念的实施例,其中外部表示ER1-ER3表示不同的第一内部浏览器表示FIR。图12公开了四个浏览事件BE_WB1、BE_WBA1、BE_WBA2、BE_WBA_n。这些浏览事件与浏览外部基于网页的数据源的网页相关,例如,如结合上面的说明所公开的。

[0244] 第一浏览器WB1由网络机器人或用户(未在图12中例示)使用用于浏览网页。在网页的浏览BE_WB1期间,用户或网络机器人希望在浏览事件BE_WB1期间在第一点P1建立外部表示ER1,以便帮助在后面点重建此第一内部浏览器状态FIR1。因此第一内部表示FIR1表示在特定浏览事件BE_WB1期间那点P1处的网页浏览器WB1的特定状态。因此用户或网络机器人开始建立表示点P1处的浏览器的第一内部表示的第一外部表示ER1。

[0245] 知道先前的第一内部表示FIR1后面可恢复,用户或网络机器人随后继续浏览网页。

[0246] 因此用户或网络机器人可在浏览事件BE_WB1期间在点P2,P3处建立进一步的外部表示ER2、ER3,并且因此这些中的每一个以后都可用于在点P2、P3处建立网页浏览器WB1的状态的进一步的第一内部表示FUIR2,FUIR3。因此进一步的外部表示ER2、ER3表示网页浏览器WB1的进一步的第一内部状态表示FIR2、FIR3的状态,其在浏览事件BE_WB1期间在后面点处建立。

[0247] 在点P4,用户或网页浏览器希望重建先前的第一内部浏览器状态FIR1以从先前的浏览器状态FIR1浏览网页。因此,作为对用户或网络机器人指令的响应,网页浏览器处理第一内部浏览器状态FIR1的外部表示ER1,并且因此网页浏览器被恢复为点P4的浏览器状态

FUIR1,然后用户或网络机器人可从其继续与第一内部表示FIR相对应的浏览器状态下的浏览事件。这可如上面在之前的实施例中描述的通过利用如结合例如图1、3和4所描述的解串行化并且使用映射数据实现外部表示ER1来促进。这个后面的方案也可称为循环。

[0248] 在现在公开的发明性概念的实施例中,另一网页浏览器应用WBA1用于从与第一网页浏览器WB1的第一内部浏览器状态FIR1相对应的浏览器状态开始进一步的个体浏览事件BE_WBA1。因而通过处理外部表示ER1来建立Si进一步的内部表示FUIR1。这还可如在此文档中先前描述的来完成。

[0249] 用相同的方式,进一步的内部浏览器状态表示FIR2、FIR3可在分别与网页浏览器状态FIR2、FIR3相对应的不同的进一步的网页浏览器应用WBA2、WBA_n中分别利用第二和第三表示ER2、ER3建立。因此这从各自的浏览器状态开始进一步的个体浏览事件BE_WBA2、BE_WBA_n。这些浏览器中的循环还可通过分别在浏览事件BE_WBA2、BE_WBA3中分别重建网页浏览器状态FIR2、FIR3来促进。

[0250] 理解为,在现在公开的发明性概念的实施例中,多个浏览事件BE_WB1、BE_WBA1、BE_WBA2、BE_WBA_n可用相同的浏览器应用WB、WBA促进,其在现在公开的发明性概念的实施例中可帮助建立并操控多个浏览事件。

[0251] 图13例示了外部表示ER的实施例以及在外在化第一内部状态表示FIR之后其所包括的内容。

[0252] 外部表示ER包括动态内容DC及其状态的外部表示DCSR。此外外部表示DCSR表示如之前描述的网页浏览器WB的状态的第一内部表示FIR的动态内容DC的状态。

[0253] 此外,外部表示ER包括描述动态内容DC和第一内容FC的至少一部分诸如标记语言内容、样式表内容(或类似的表示第一内部表示FIR的可视布局的内容)之间的相互关系的映射数据MD。

[0254] 而且,外部表示ER包括第一内部表示FIR的第一内容FC的状态的表示RFC。

[0255] 第一内部表示FIR的第一内容FC的状态的表示RFC优选地至少包括标记语言内容MLC的表示MLR,并且还可包括内部状态表示FIR的样式表内容SSC的表示SSR。

[0256] 而且在现在公开的发明性概念的实施例中,第一内部表示FIR的第一内容FC的状态的表示RFC包括表示第一内部表示FIR的cookie数据COD的cookie信息CODR。此数据可在建立外部浏览器状态表示ER期间从第一内部状态表示FIR识别并拷贝。因此,第一内部表示FIR的cookie可在进一步的内部表示FUIR中建立。

[0257] 在现在公开的发明性概念的进一步实施例中,可建立第一内部表示FIR的第一内容FC的状态的表示RFC以包括形式内容表示FCOR。该形式内容表示FOCR包括关于填充到网页的第一内部表示FIR的形式中的形式内容FOC的信息。

[0258] 而且,在现在公开的发明性概念的实施例中,第一内部表示FIR的第一内容FC的状态的表示RFC包括在第一内部表示FIR时候的网页浏览器WB的浏览历史数据BHD的表示BHDR。该表示可通过串行化浏览历史数据获得。因此,当此浏览历史表示BHDR随后在网页浏览器应用WBA中实现时,其将使能用户或网络机器人能够去先前的URLs,其在导向第一内部表示FIR的浏览事件期间被访问。因此,在基于外部表示ER建立了进一步的内部表示FUIR以后,用户可例如使用浏览器应用中的“后退按钮/功能”并访问外部状态表示ER建立之前通过网页浏览器WB访问的URL。

[0259] 一般理解为,第一内容FC和动态内容DC可依靠第一内部表示FIR以任何合适的方式提取。因此,在外部表示中样式表内容可被转化成层叠样式表形式,标记语言内容MLC可被转化成例如XML形式,动态内容DC可被转化或拷贝成二进制形式等。

[0260] 一般理解为,在有利的实施例中,建立外部表示基于网页浏览器WB的内部的文档对象模型IDOM表示(未在图13中例示)而建立。此内部的文档对象模型表示建立外部表示ER时候的浏览器WB的状态。

[0261] 网页浏览器WB一般可包括通过识别并且如果必要的话串行化内部表示FIR的相关部分使能建立外部表示ER的软件代码(未例示)。因而,可以实现软件代码,这样当用户或网络机器人发起外部表示ER的建立时,软件代码可处理内部DOM、动态内容DC、标记内容MLC、样式表内容SSC、cookie数据COD、动态内容DC和标记内容MLC和/或样式表内容SSC之间的指针、形式数据FOC、历史数据BHD等等。

[0262] 在其他实施例中,软件代码可视为对网页浏览器的补充,并且在这样的实施例中,网页浏览器(WB)可适于与此软件协作以提供对内部DOM、动态内容DC、标记内容MLC、样式表内容SSC、cookie数据COD、动态内容DC和标记内容MLC和/或样式表内容SSC之间的指针、形式数据FOC、历史数据BHD等等的访问。

[0263] 图14例示了通过网页浏览器WB在屏幕SC上可视化的网页的示例。

[0264] 网页包括多个(在此示例中为3个)框架FR1-FR3(诸如内联框架)。每个框架FR1-FR3将网页划分为不同的部分,其可显示不同的文档/数据。用户可在地址段AF中输入URL地址并且在此基础上,网页浏览器WB、WBA访问、检索并可可视化基于网页的数据源的网页。

[0265] 框架FR1-FR3的每一个表示不同内容并且包括网页浏览器WB中的它们自己的内部文档对象模型结构、样式表内容等等。因此,根据实施例,外部表示可包括建立多个框架表示FR1R-FR3R,其表示网页的第一内部表示FIR的不同框架。这些框架的每一个可包括如上描述的第一内容FC和动态内容DC两者,并且因此框架表示FR1R-FR3R包括网页浏览器中的每个框架FR1-FR3的内部状态的外部表示ER。

[0266] 而且,框架FR1-FR3在框架树结构中的网页浏览器中是相关的,并且而且此框架树结构在外部表示ER中被确定并外在化为框架相关表示FRAR,以便后面能够在如早前描述的相同的或另一网页浏览器中的进一步的内部表示FUIR中正确关联和/或可视化框架FR1-FR3。

[0267] 框架树和框架的内容优选地基于浏览器WB的内部文档对象模型IDOM被处理并串行化。

[0268] 连同每个框架,可建立映射数据MD,以便帮助关系的后续重建,所述关系诸如每个个体框架的对象和存储器地址之间的指针和/或如果存在的话框架之间的指针。

[0269] 一般理解为,网页可包括应外在化的一个、两个、三个、四个、五个或甚至更多的框架。

[0270] 因此框架FR1-FR3可通过处理网页浏览器的IDOM来识别,并且在现在公开的发明性概念的方面,因此外部表示ER可基于IDOM的配置来建立。例如,IDOM可确定标记语言内容表示以及每个框架的其他部分的外部表示的数目。

[0271] 例如,网页浏览器可处理内部IDOM并且确定DOM表示具有个体框架FR1-FR3的网页。因此,应建立三个动态内容状态表示DCSR,并且应建立关于每个它们的框架的第一内容

FC的三个表示。而且,应建立框架的三个表示/框架相关表示FRAR。所有这是外部表示的一部分,其一起描述网页浏览器WB的第一内部表示FIR的状态。

[0272] 所以,当后面基于外部表示ER在网页浏览器应用WBA中创建进一步的内部表示FUIR时,网页的作为结果的可视化将仅如其在第一内部表示中那样出现。

[0273] 可替代地,即使外部表示没有可视化,例如由于网络机器人使用外部表示,也可基于外部表示ER通过处理框架表示FR1R-FR3R和框架相关表示FRAR在网页浏览器应用中建立进一步的内部表示。

[0274] 图15例示了网络机器人的操作的实施例,例如,如在此文档中之前描述的,其中网络机器人利用外部状态表示ER。

[0275] 在步骤S151,网络机器人根据规则集和提供给网络机器人的指令利用网页浏览器加载网页LWP。

[0276] 在检测T151处,机器人检测是否符合一个或多个预定的标准(PCCP?)。这些标准可例如包括计时器,当期满时,允许机器人开始进行,计时器可包括对一个或多个网页加载参数的监督,所述网页加载参数诸如浏览器中的网页的经执行的脚本内容,它可包括可帮助识别网页何时被充分加载到网页浏览器的时间因数或任何其他相关状态指示符。

[0277] 在步骤S152,已符合预定的标准,并且然后机器人开始建立/创建外部表示(EER)。这可如在此文档中所描述的实现,例如关于图1-14的任一个。

[0278] 在步骤S153,机器人提取表示在网页浏览器(EX SUM)中的网页的Link/URL的概要。这可包括提取简介的信息,所述简介认为是告知如果点击/激活相关Link/URL则将预计有哪些信息的信息的简短概要或促销部分。

[0279] 在步骤S154,网络机器人通过激活链接来跟随Link/URL。如果步骤153是机器人的一部分,则优选地是指向所跟随的概要的相关链接。

[0280] 在步骤S155,机器人提取(或可替代地索引)在步骤154中激活链接时由浏览器处理的网页的数据。

[0281] 在步骤S156,当提取了步骤155中的数据时,机器人在于步骤152中的网页状态相对应的状态下恢复网页。因而网络机器人访问在步骤152中建立的外部表示,并且然后此外部表示ER用于恢复如当建立外部表示时具有的网页浏览器状态。这可例如如关于图1-14的任何一个所描述的完成。

[0282] 在检测T152处,机器人随后检测是否网页的所有(或所期望的)链接/URL都由机器人跟随/处理了。所以,系统/机器人因此继续追踪哪个链接/URL由机器人跟随/处理了。

[0283] 如果机器人尚未跟随/处理网页的所有(或期望的)链接/URL,则机器人开始新的链接/URL指定的新处理,例如,如关于步骤153-156所公开的。

[0284] 如果机器人已跟随/处理了网页的所有(或期望的)链接/URL,则机器人可如图15中所公开的终止。可替代地,机器人可如图15中所公开的继续处理新的网页,例如,基于由上述经处理的网页或新的网页的一个链接所指定的网页。

[0285] 图16例示了利用多个网络机器人的实施例。

[0286] 在步骤S161,将网页加载到浏览器以便在浏览器中表示。

[0287] 在步骤S162和163,输入用户名(EUN)和密码(EPW)以便允许访问另外的受限数据。这可输入到所表示的网页中配置为输入用户名(EUN)和密码(EPW)的字段,以便这些可通过

例如点击/激活“OK”或“继续”按钮进行处理。(在实施例中此激活还可由机器人促进)

[0288] 在步骤S164,实施建立/创建外部表示ER EER。这可如在此文档中所描述的完成,例如关于图1-15中的任一个。因而,外部表示ER将表示网页浏览器的状态,其包括密码和用户名。为了此目的,如果形式数据包括用户名和/或密码信息,例如形式数据可例如有利地被提取以表示为外部表示的一部分。此外或可替代地,包含所输入的密码和/或用户名信息的其他信息可被检索并提取到外部表示ER。

[0289] 步骤S161-164可如上所述的由机器人R1执行。可替代地,用户可通过操控网页浏览器执行步骤S161-164来检索网页,用户可利用键盘或触摸屏或另一输入设备输入用户名和密码并且然后操控浏览器建立外部表示EER。

[0290] 然后,多个机器人R1-Rn访问在步骤S164中建立的外部表示,以便访问先前受限数据的内容,所述受限数据由于需要用户名和密码而受限。

[0291] 因而,机器人R2-Rn的每一个从步骤S164各自在其网页浏览器中恢复外部表示,以使状态与步骤S164中的网页浏览器的状态相对应。这可如关于图1-15中任一个所公开和描述的完成。

[0292] 因而机器人执行网页的并行处理,优选地通过访问网页的不同内容、从网页提取所想要/所期望的数据(步骤S166),诸如HTML数据(EXHTML)或任何其他适合的数据,并且而且可将数据记录到数据存储(LTD)、将数据存储到数据库(STDB)、索引内容等等(步骤S167)。

[0293] 理解为,在进一步的实施例中,由机器人R1-Rn执行的步骤可与步骤S151-156组合和/或用其替代。

[0294] 一般理解为,关于图15和16所描述的机器人可以是包括一个或多个规则集和/或指令的软件代码,所述规则集和/或指令可被设计以便如此前在此文档中所描述的访问、处理和从特定网页或某些类型的网页检索内容。

[0295] 一般理解为,现在公开的发明性概念并不限于上述示例,但可以如例如在权利要求中指定的以众多变化加以组合。而且,理解为,关于图1-11所描述的实施例可结合网络机器人的使用而使用。

[0296] 当然,上述说明要理解为仅表征发明人所认为的要落入现在公开的发明性概念的范围内的若干示范性实施例。前述特征的任何组合、排列、变化或综合要理解为本申请的方面,即使特定组合、排列、变化、综合等并未在具有这些说明的任何单独实施例或附图中清楚提及。

[0297] 而且,根据本领域普通技术人员在阅读这些说明后达到的理解,附加的实施例、前述实施例的变化以及可替代的实施例也要认为完全落入本申请的范围。

[0298] 因此,上述说明应认为决不限制本申请的范围。而是,本文中呈现的主题被提供以示例本发明性概念的范围。根据本申请的知识产权的界线和边界要通过下面的权利要求及其所有等价形式进行限定,技术人员通过回顾前面的公开会理解这点。

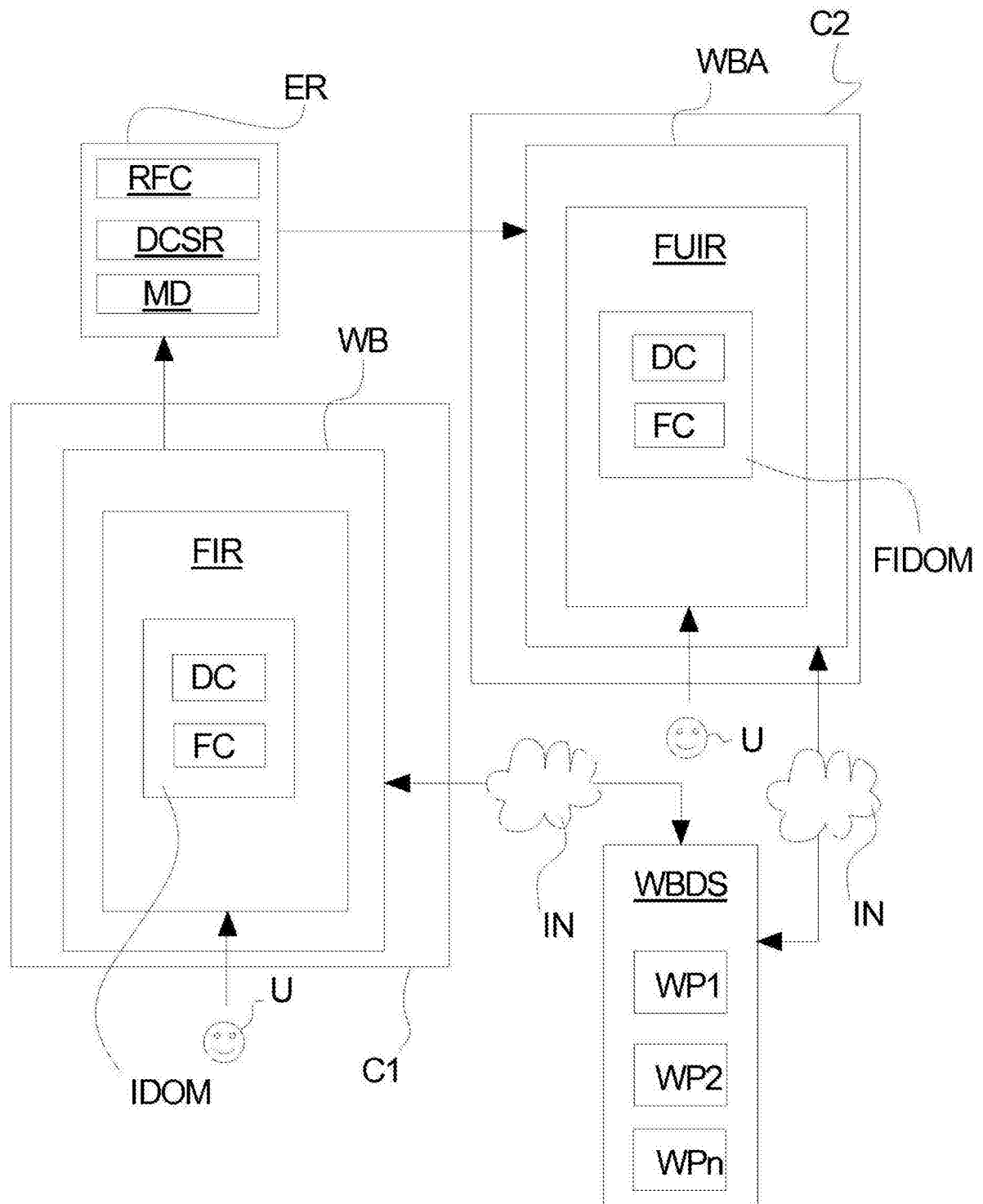


图1

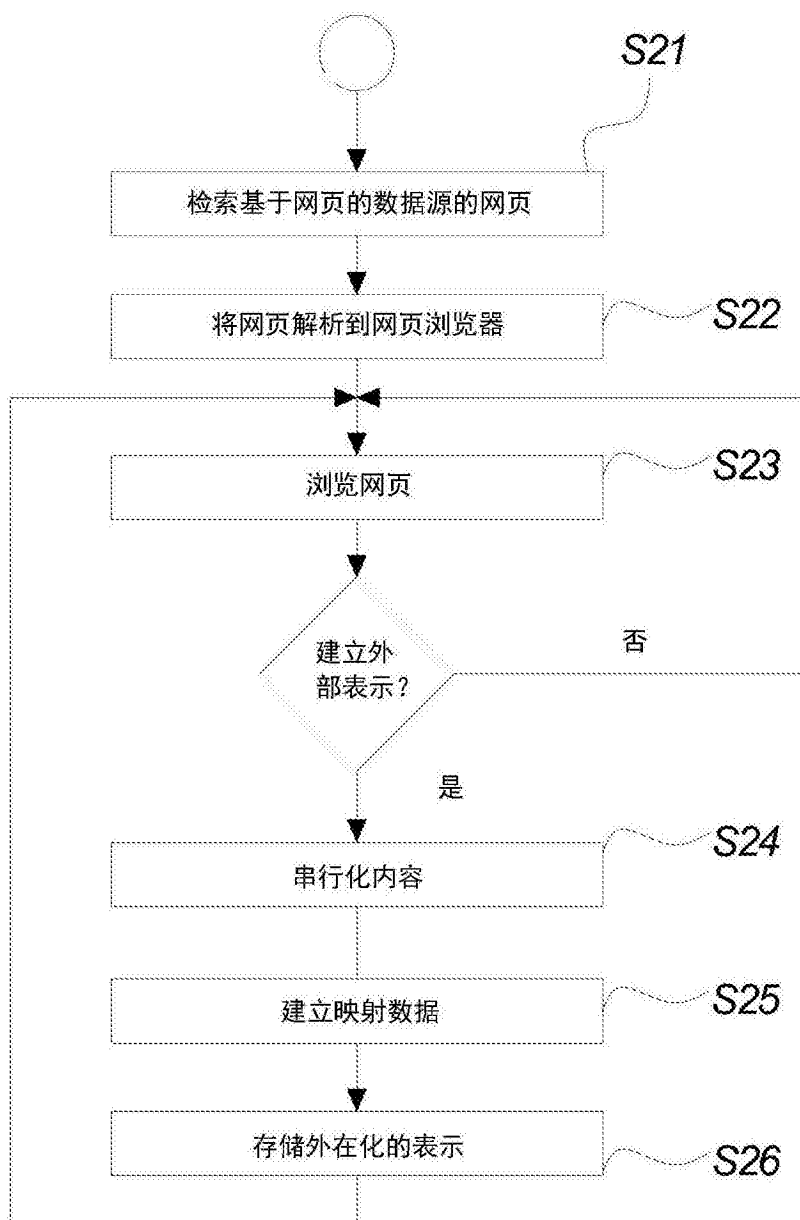


图2

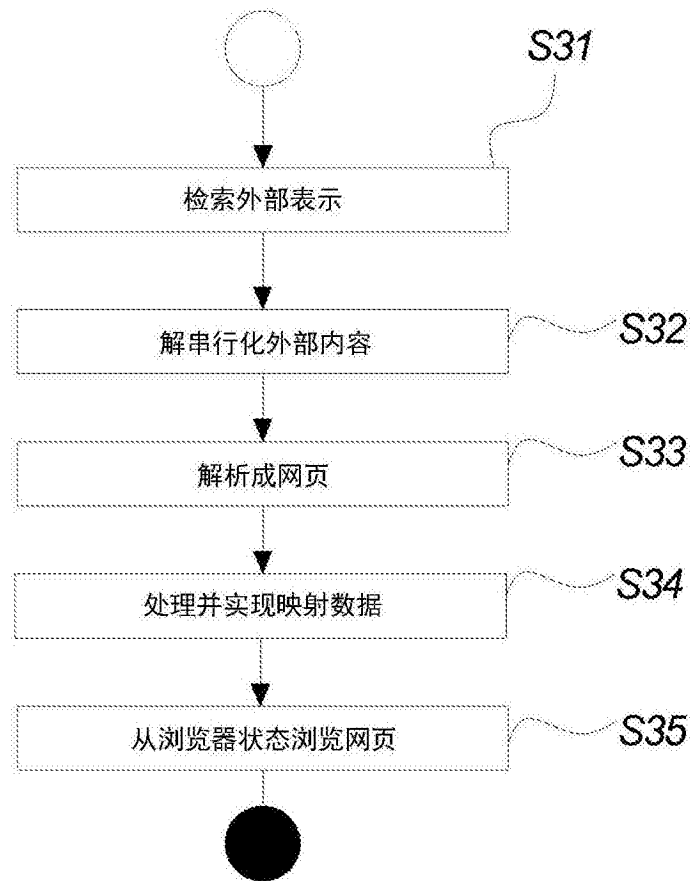


图3

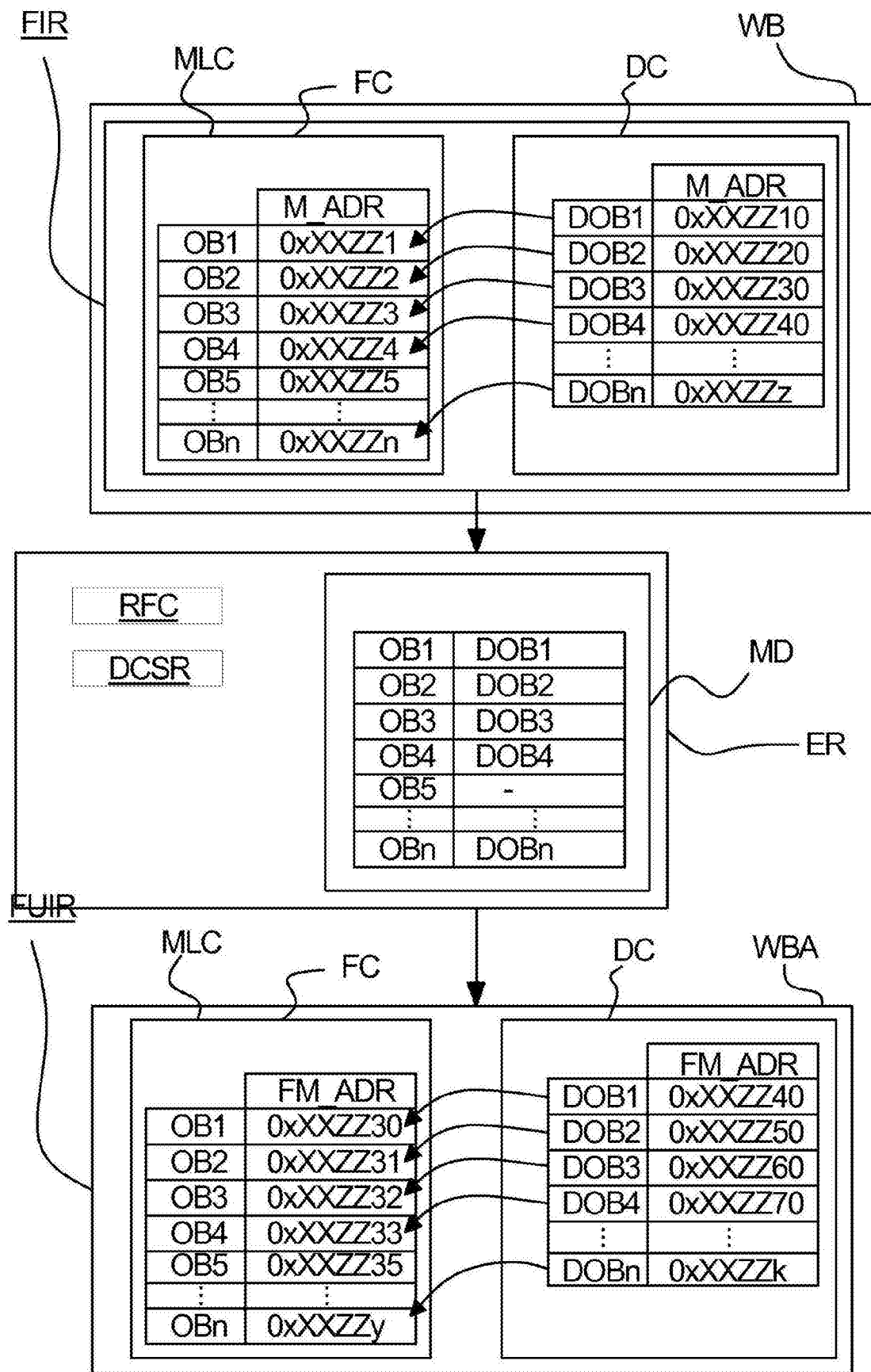


图4

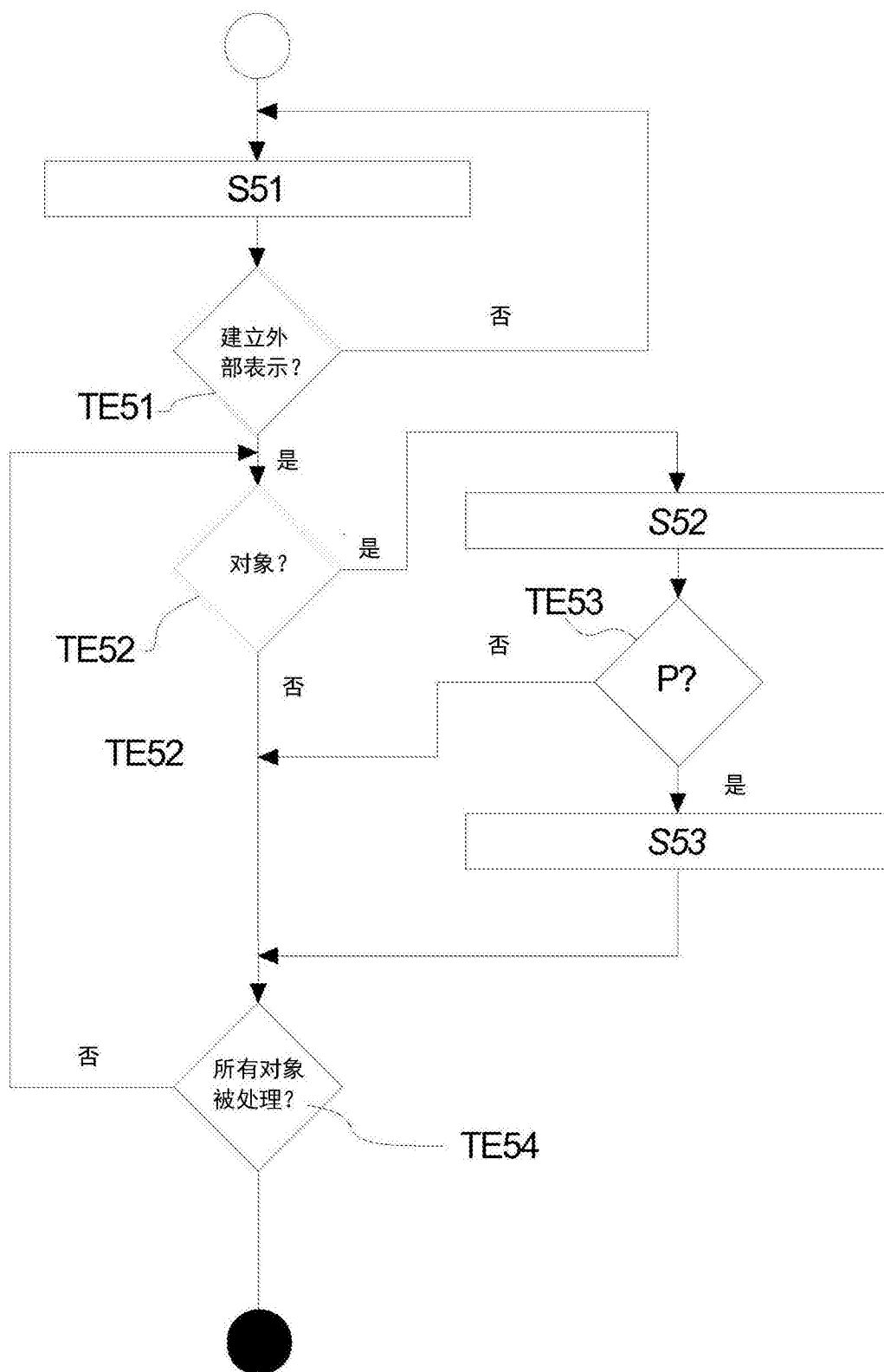


图5

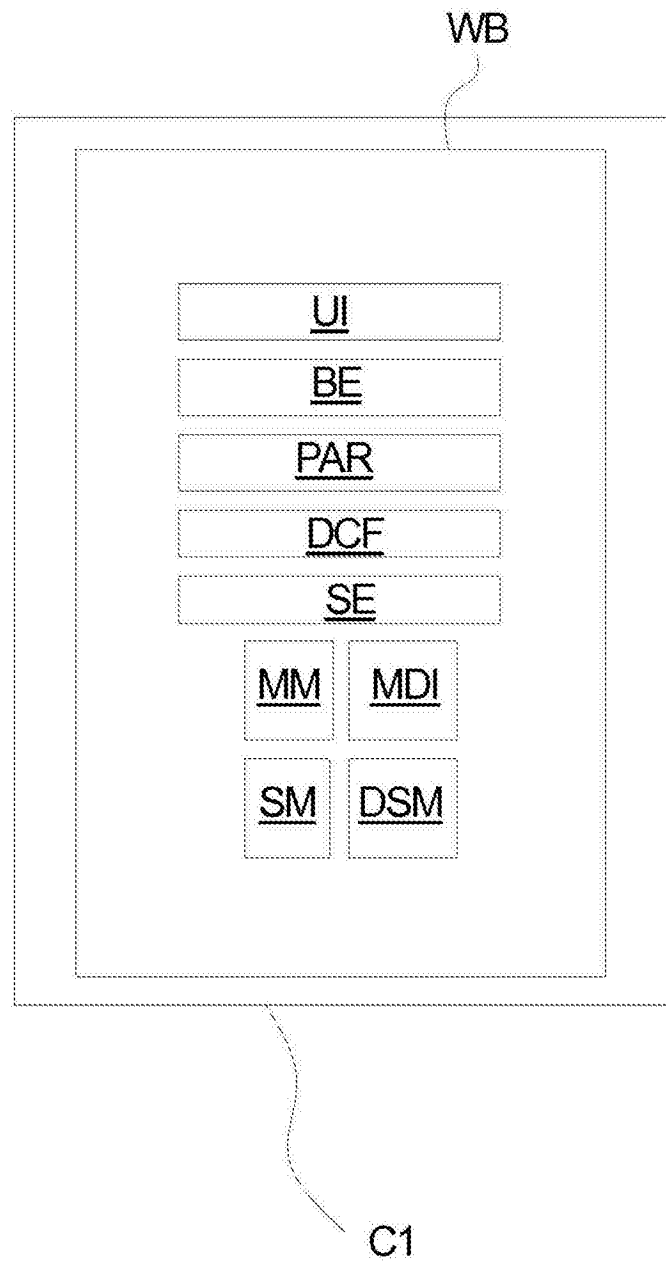


图6

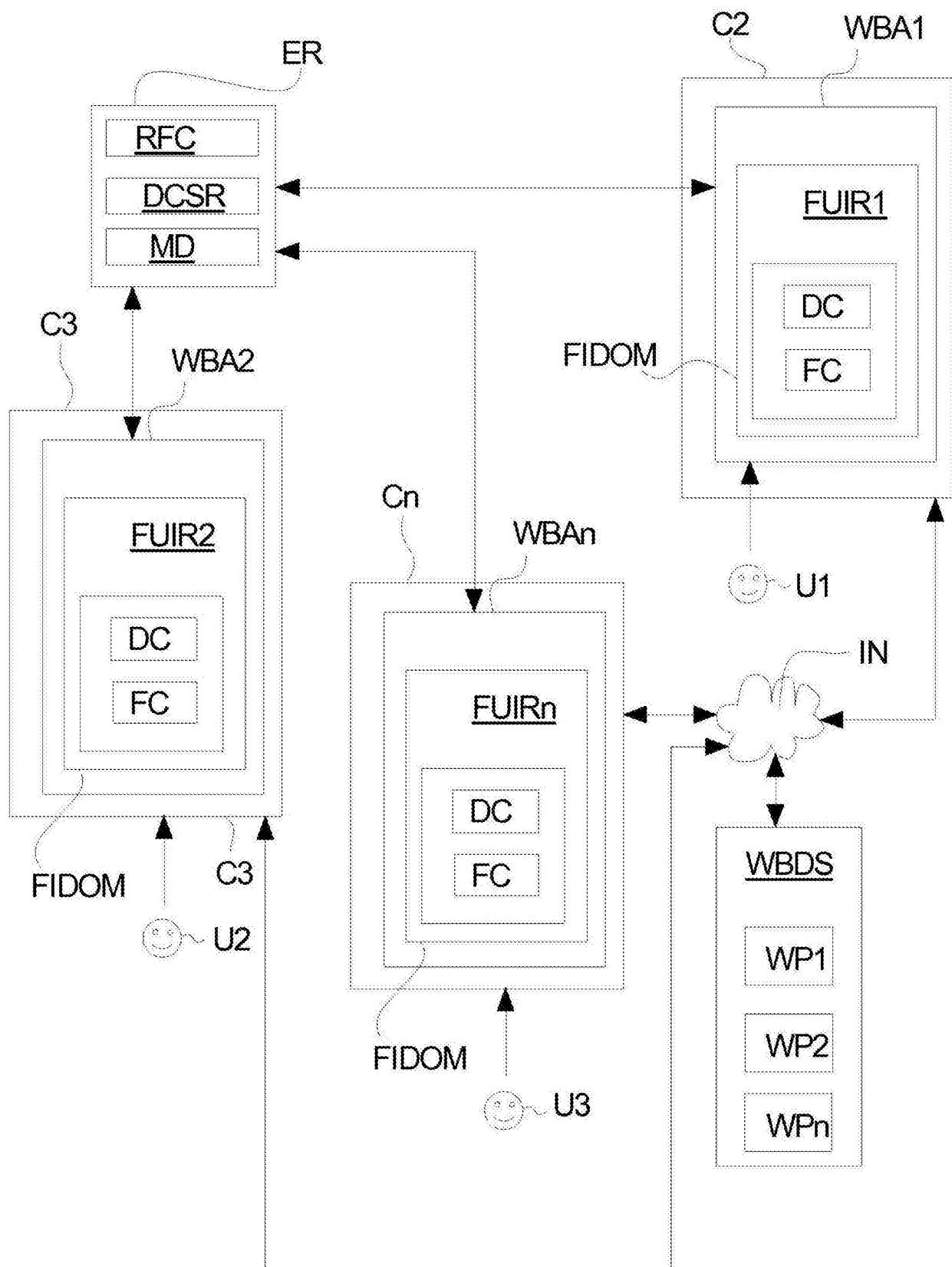


图7

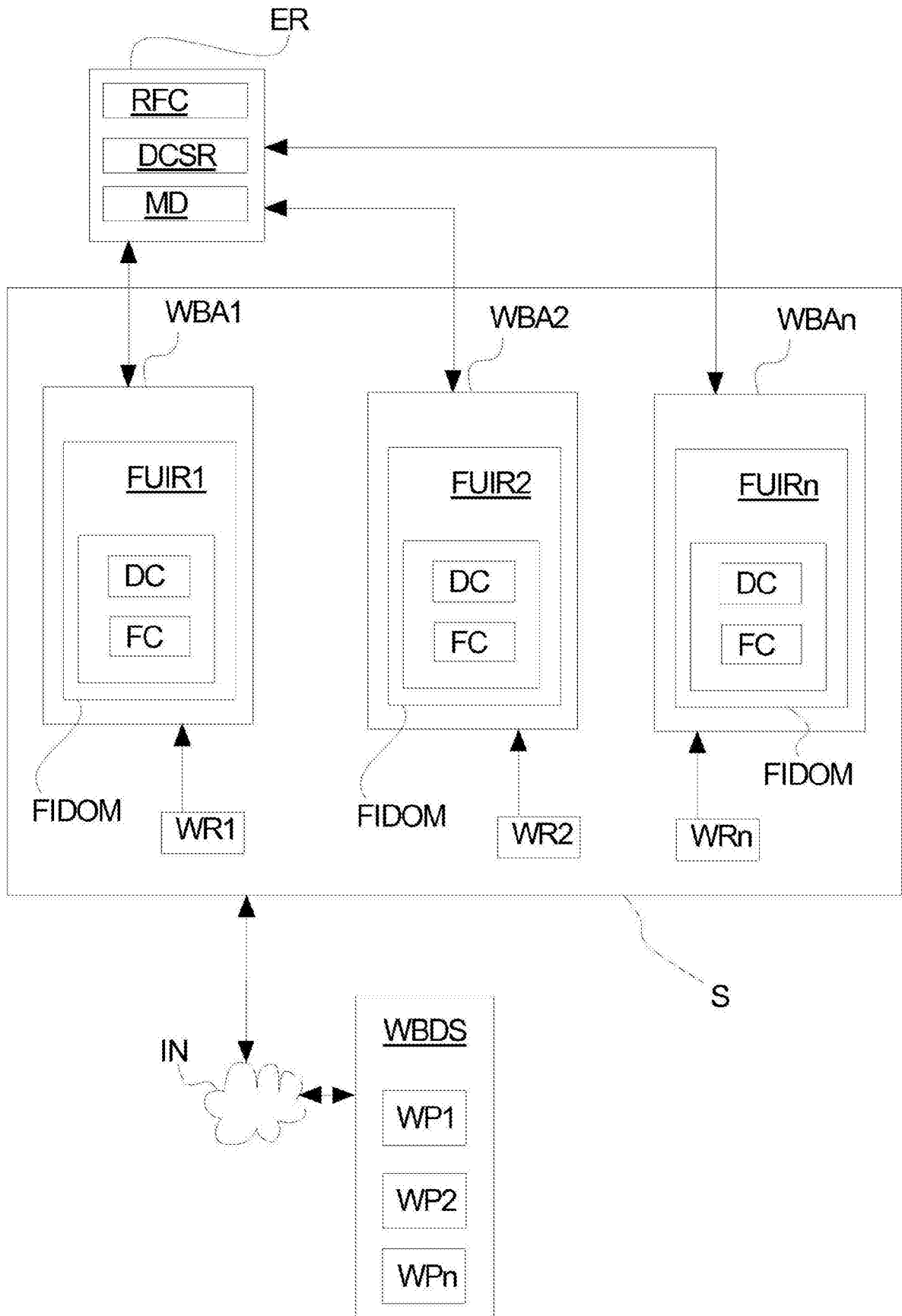


图8

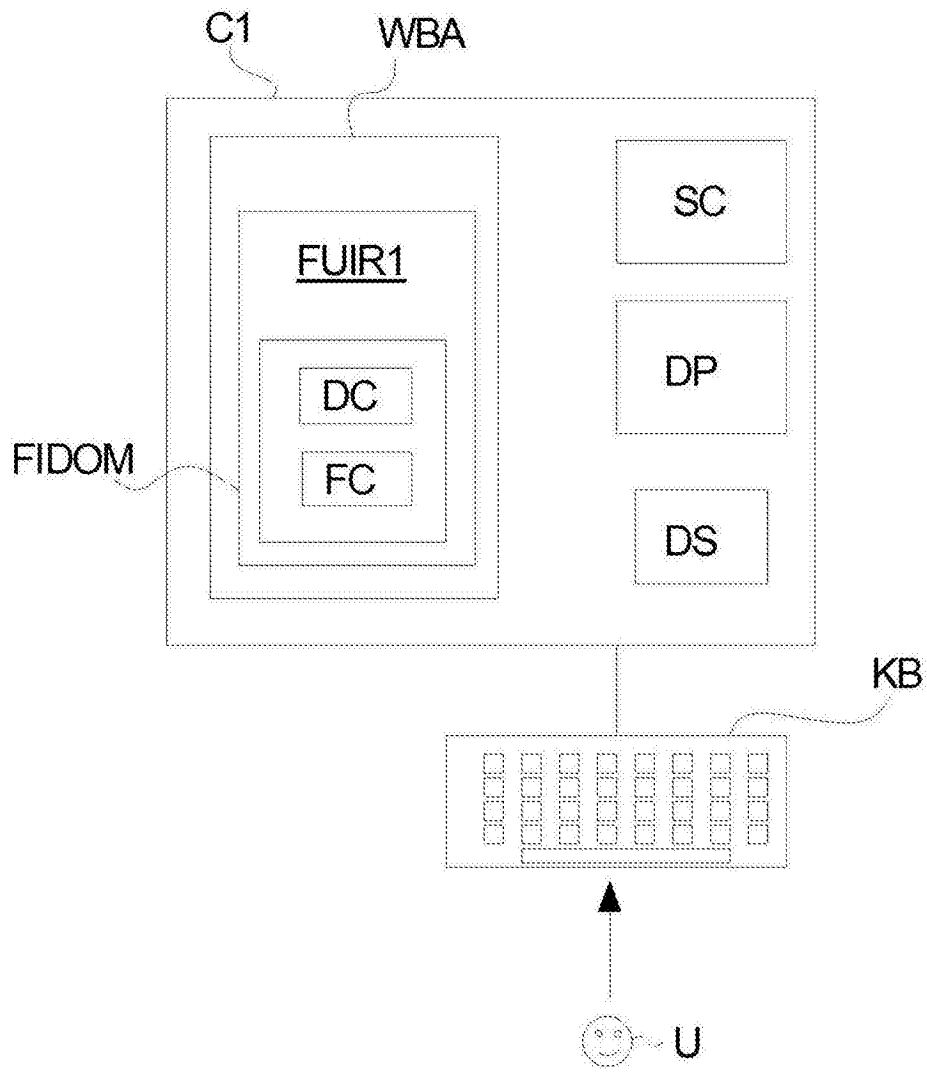


图9

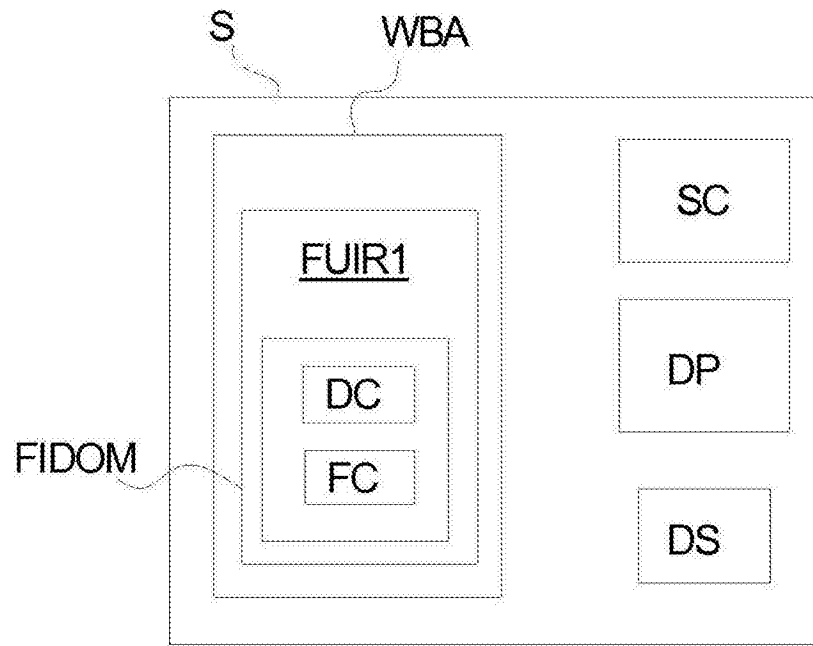


图10

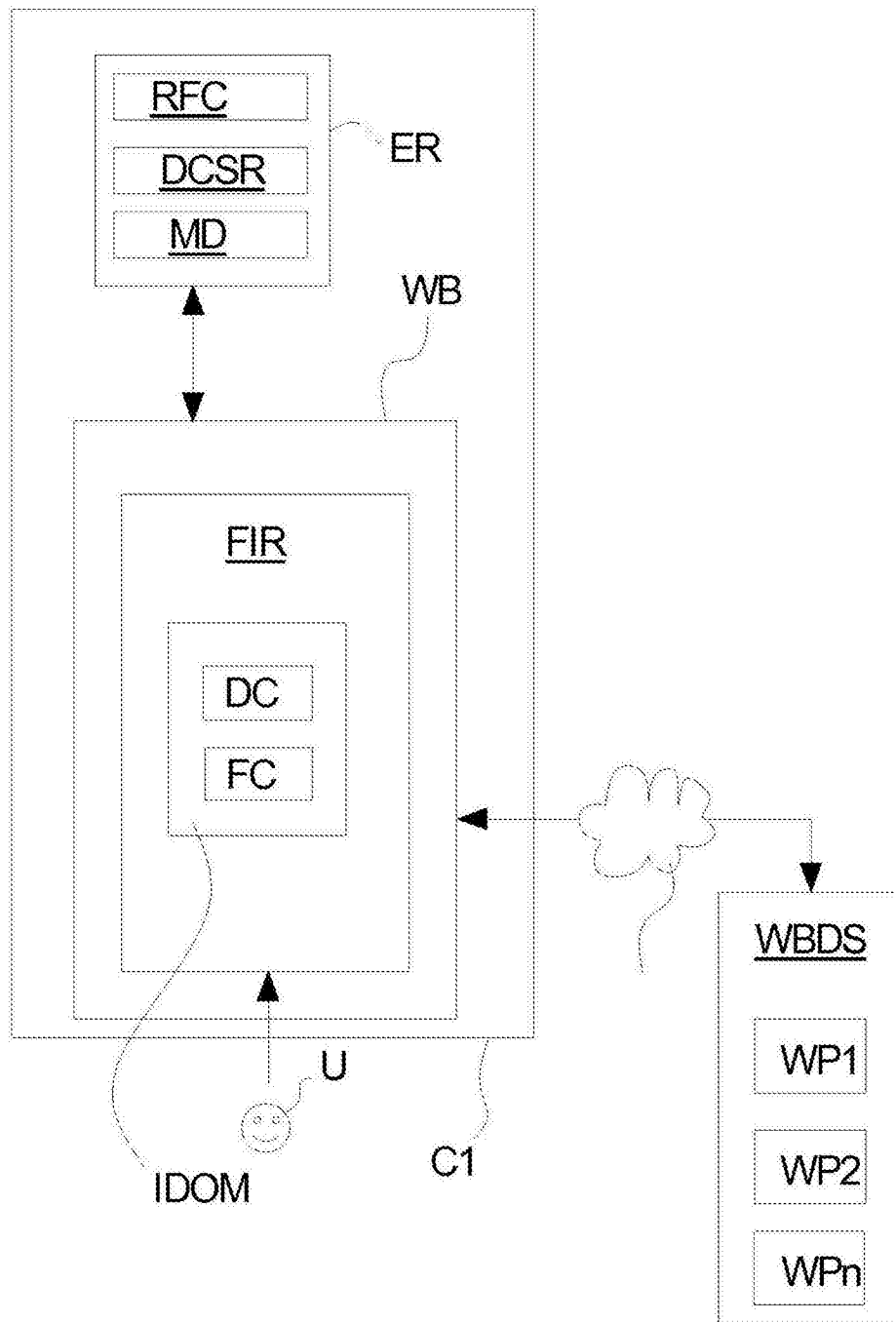


图11

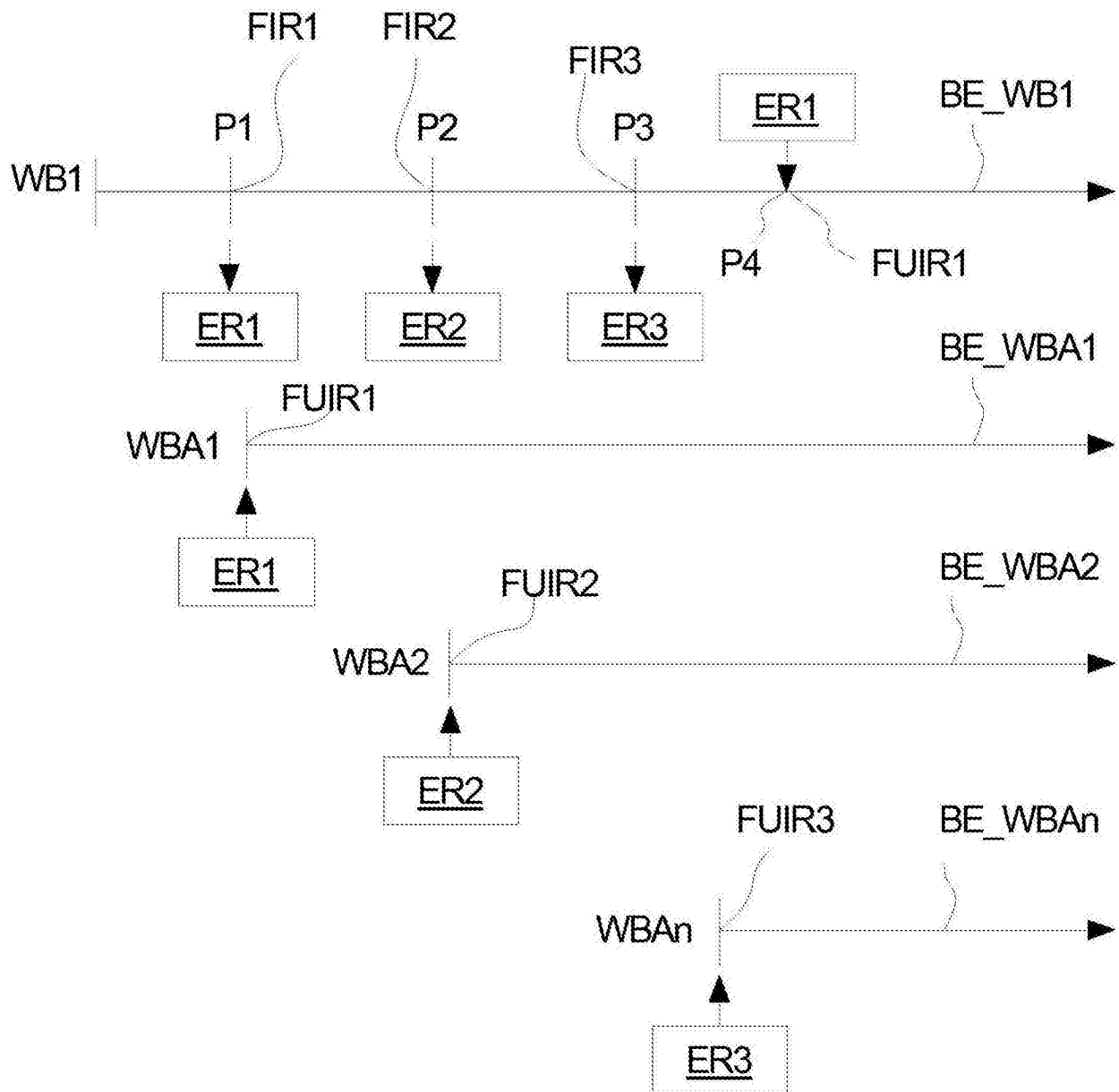


图12

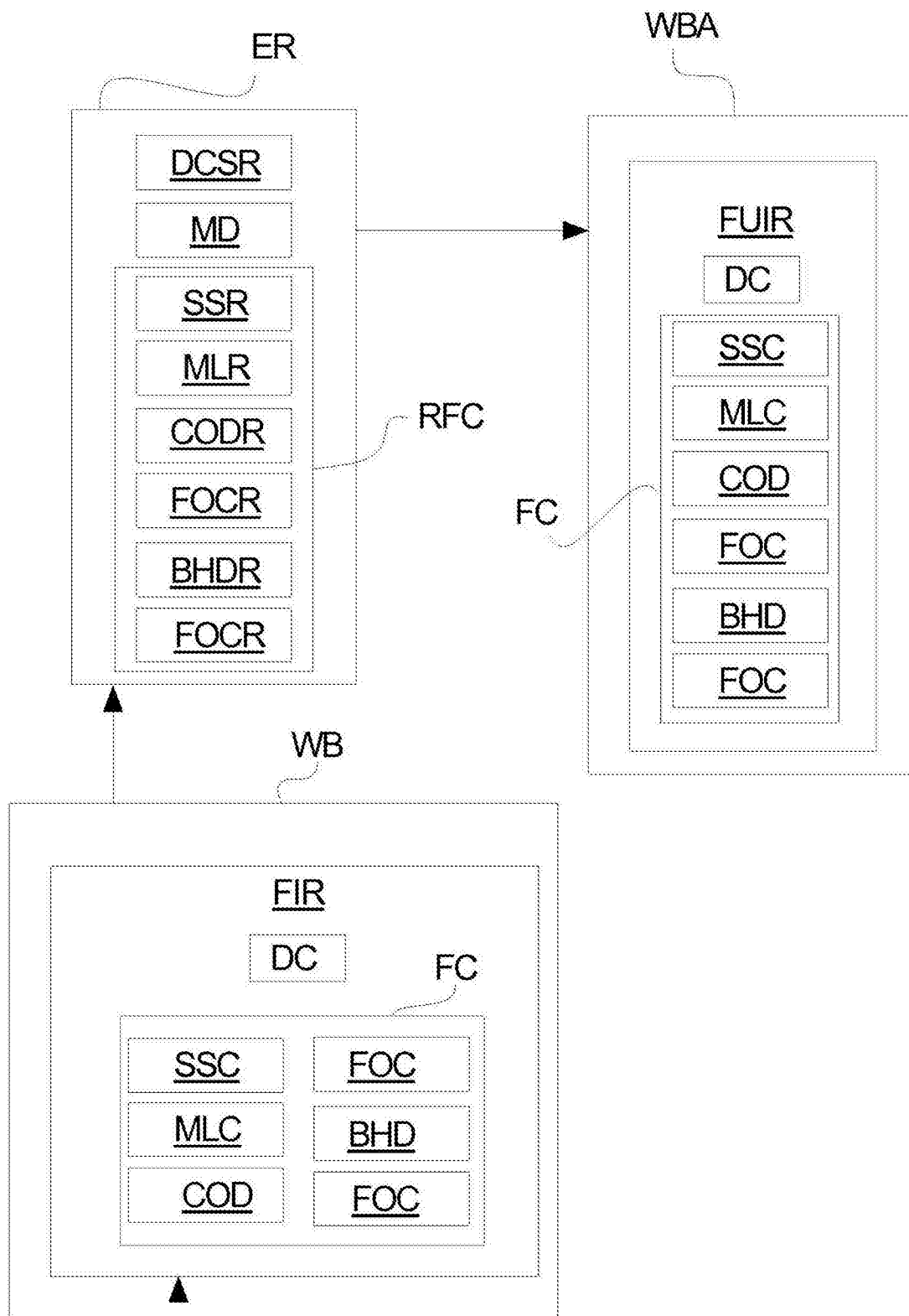


图13

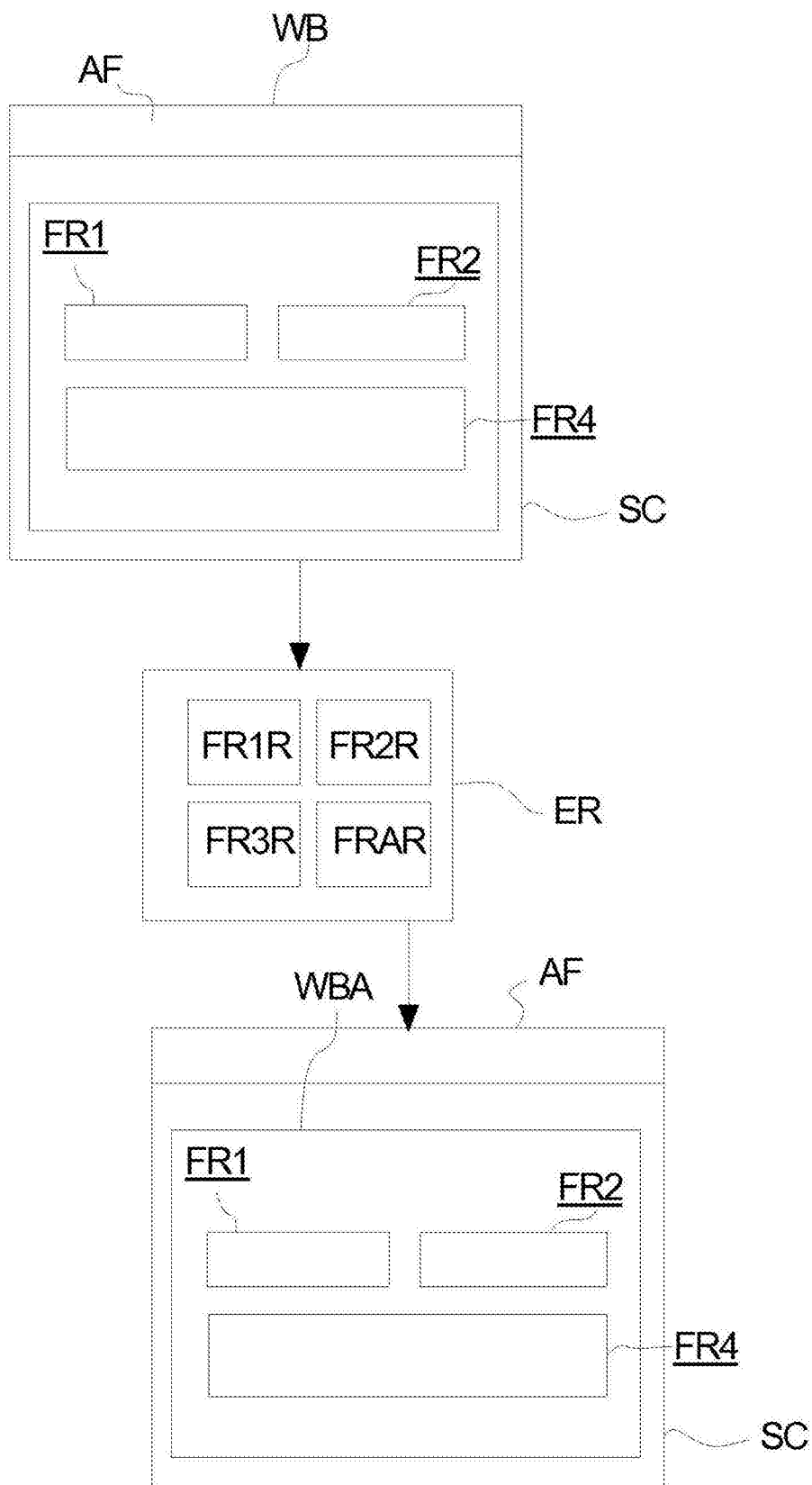


图14

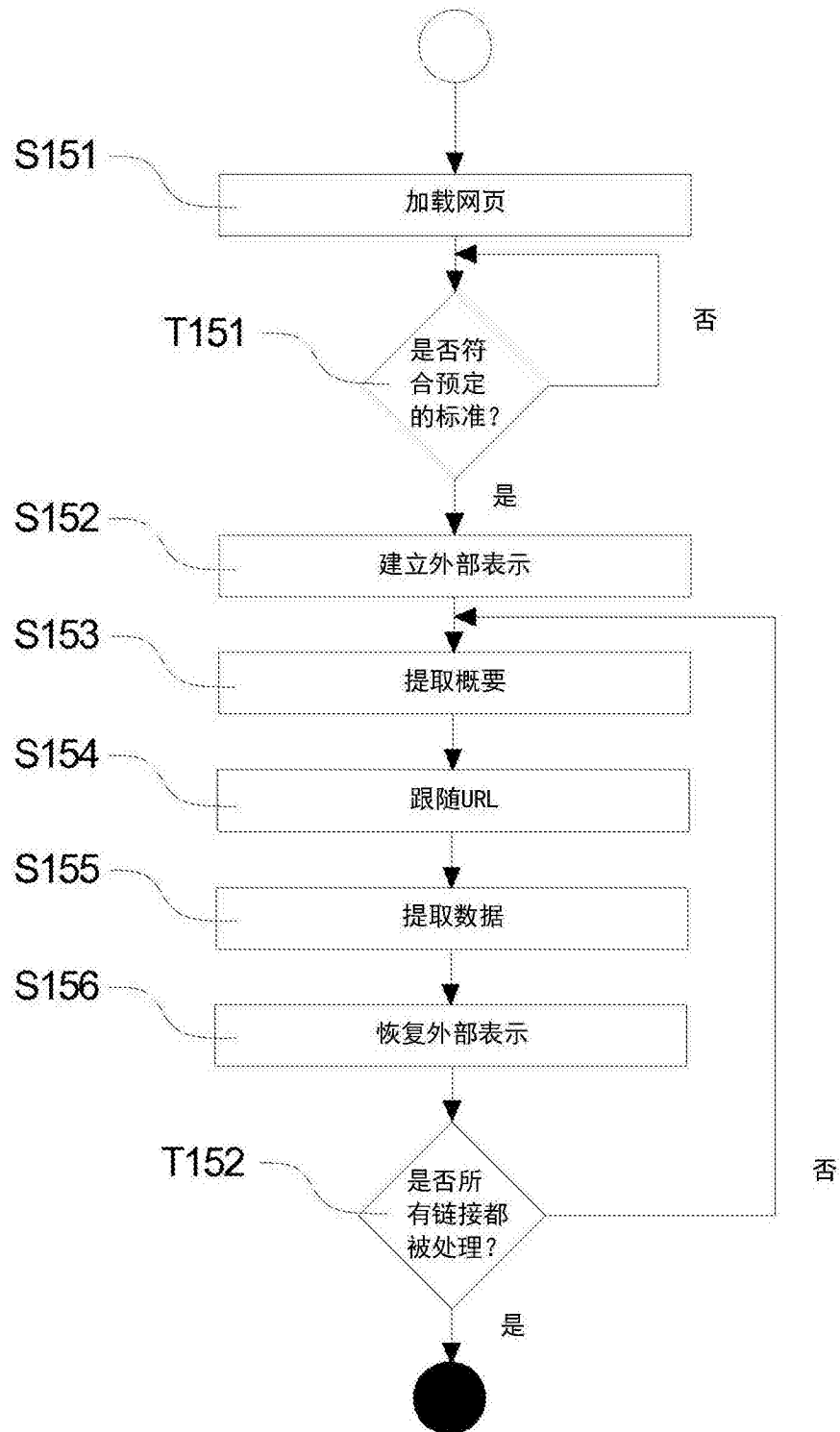


图15

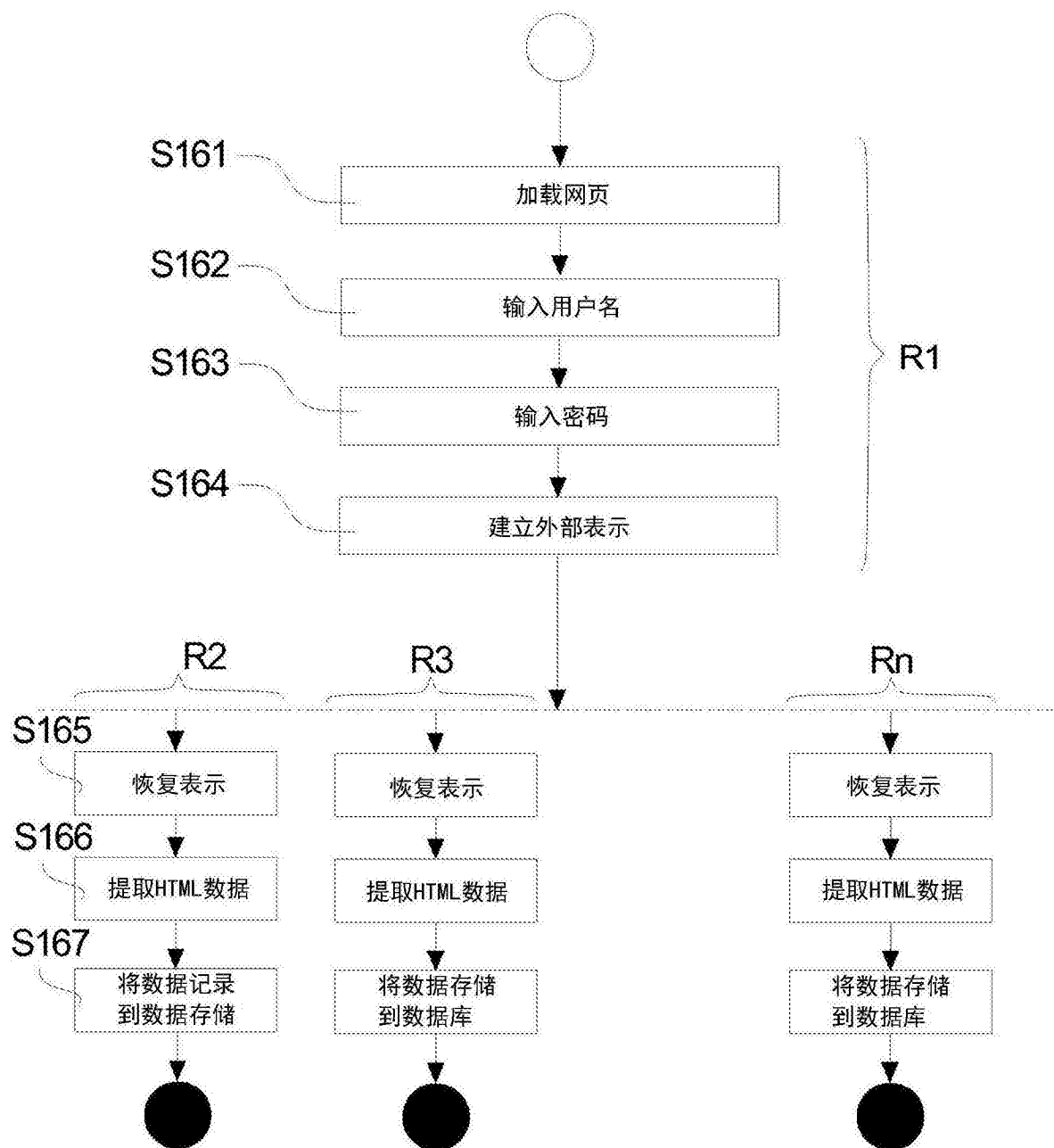


图16