[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
G06F 17/00 (2006.01)
G06F 9/46 (2006.01)

[12] 发明专利申请公布说明书

[21] 申请号 200680029960.4

[43] 公开日 2009 年 7 月 29 日

[11] 公开号 CN 101495998A

[22] 申请日 2006.7.20
[21] 申请号 200680029960.4
[30] 优先权
[32] 2005. 8.19 [33] US [31] 11/207,547
[86] 国际申请 PCT/US2006/028384 2006.7.20
[87] 国际公布 WO2007/024378 英 2007.3.1
[85] 进入国家阶段日期 2008.2.18
[71] 申请人 微软公司
地址 美国华盛顿州
[72] 发明人 G·A·瑟弗 J·M·索德伯格
[74] 专利代理机构 上海专利商标事务所有限公司
代理人 陈 斌

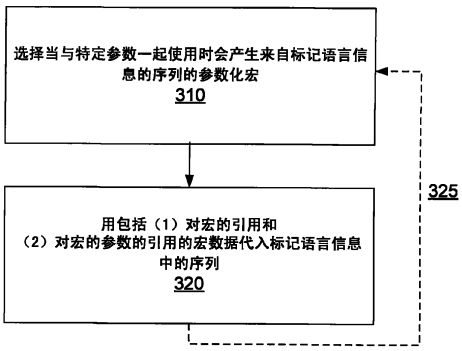
权利要求书 3 页 说明书 11 页 附图 4 页

[54] 发明名称

标记语言数据的编码

[57] 摘要

通过使用多个参数化宏将标记语言文档转换成压紧的标记语言形式。 参数化宏取至少一个参数。当压紧标记语言文档时，用对宏和参数的引用替换元素序列，当该宏被与该参数一起使用时产生该元素序列。 所使用的宏可以是预定的、来自静态字典可以在运行时生成、或两者的混合。 宏的定义可被包括在压紧的标记语言信息中以便允许压紧标记语言信息的接收方将宏引用和参数展开成已经过替换的元素序列。



1．一种用于编码标记语言信息以便传输的方法，其中所述标记语言信息包括序列，所述序列包括至少两个有序元素，所述至少两个有序元素包括至少一个数据元素以及与每一数据元素相关联的至少一个标签元素，所述方法包括：

选择对应于所述序列的宏，其中所述宏取至少一个参数，且对所述宏的所述至少一个参数中的每一个使用所述至少一个数据元素当中的至少一个特定数据元素产生所述序列；以及

用包括对所述宏和对所述标记语言信息中所述特定数据元素的引用的宏数据代入所述标记语言信息中的所述序列，所述代入得到修改后的标记语言信息。

2．如权利要求 1 所述的方法，其特征在于，所述对宏的选择包括：

从至少一个预先存在的宏的集合中选择预先存在的宏。

3．如权利要求 2 所述的方法，其特征在于，所述预先存在的宏的集合是包括宏集合的静态字典，其中所述静态字典是可为所述传输的接收方所用的。

4．如权利要求 1 所述的方法，其特征在于，所述选择预先存在的宏的步骤包括：

生成对应于所述序列的所述宏。

5．如权利要求 1 所述的方法，其特征在于，还包括：

将对所述宏的所述定义添加到所述修改后的标记语言信息中。

6．如权利要求 1 所述的方法，其特征在于，所述为序列选择宏并用对所述宏和所述特定数据元素的引用来代入所述序列的步骤是对所述标记语言信息中的至少两个序列迭代地执行的。

7．如权利要求 1 所述的方法，其特征在于，所述修改后的标记语言信息包括包含至少两个第二序列元素的第二序列，其中所述第二序列元素中的每一个包括对宏、对数据元素或对与数据元素相邻的标签元素的引用，且所述为序列选择宏并用包括对所述宏和所述特定数据元素的引用的宏数据代入所述序列的步骤是对所述标记语言信息中的至少两个序列迭代地执行的。

8．如权利要求 1 所述的方法，其特征在于，还包括：

标记化所述修改后的标记语言信息，所述标记化包括使用标记来代表所述宏数据中的所述宏。

9．一种用于准备标记语言信息以便传输的系统，其中所述标记语言信息包括信息序列，所述序列包括至少两个有序元素，所述至少两个有序元素包括至少一个数据元素以及与每一数据元素相关联的至少一个标签元素，所述系统包括：

宏存储，用于存储宏集合，其中所述宏中的每一个取至少一个元素作为参数；

宏选择器，用于从所述宏集合中选择特定宏，其中所述特定宏当应用于包含一个或多个特定参数的参数集时产生所述信息序列；以及

序列代入器，用于用包括对所述特定宏和参数集的引用的宏数据来代入所述信息序列以产生修改后的标记语言信息。

10．如权利要求 9 所述的系统，其特征在于，所述宏存储包括至少一个预先存在的宏的集合。

11．如权利要求 10 所述的系统，其特征在于，所述预先存在的宏的集合是包括预定义宏的集合的静态字典，其中所述静态字典可为所述传输的接收方所用。

12．如权利要求 9 所述的系统，其特征在于，所述宏选择器还生成对应于所述序列的所述特定宏并将所述特定宏存储在所述宏存储中。

13．如权利要求 9 所述的系统，其特征在于，所述修改后的标记语言信息包括第二序列，其中所述宏选择器从所述宏集合中选择第二特定宏，其中所述第二特定宏当应用于包含一个或多个特定参数的第二参数据集时产生所述第二序列，且其中所述序列代入器用包括对所述第二特定宏和所述第二参数集的引用的第二宏数据代入所述修改后的标记语言信息中的所述第二序列。

14．如权利要求 9 所述的系统，其特征在于，还包括：

标记化器，用于标记化所述修改后的标记语言信息，所述标记化器用表示所述宏的标记代入所述修改后的标记语言信息中的所述宏引用。

15．一种含有可由计算机执行的计算机可读指令的计算机可读介质，所述指令用于执行以下步骤，包括：

接收包含宏数据的第一信息，所述宏数据包含对宏以及所述宏的至少一个参数的引用；

获取对所述宏的宏定义；

将所述宏定义应用于所述参数以产生结果序列；以及

用所述结果序列代入所述第一信息中的所述宏数据以产生第二信息。

16．如权利要求 15 所述的计算机可读介质，其特征在于，所述获取宏定义的步骤、将所述宏定义应用于参数以产生结果序列的步骤、以及用所述结果序列代入

所述宏数据的步骤是迭代地执行的。

17. 如权利要求 15 所述的计算机可读介质，其特征在于，获取对所述宏的宏定义包括：

从至少一个预先存在的宏的集合中选择预先存在的宏。

18. 如权利要求 17 所述的计算机可读介质，其特征在于，所述预先存在的宏的集合是包括宏集合的静态字典，其中所述静态字典可为所述传输的接收方所用。

19. 如权利要求 15 所述的计算机可读介质，其特征在于，获取对所述宏的宏定义包括：

从所述第一信息中获取宏定义。

20. 如权利要求 15 所述的计算机可读介质，其特征在于，所述接收包括宏数据的第一信息——所述宏数据包括对宏和所述宏的至少一个参数的引用——的步骤还包括：

展开所述第一信息中的标记。

标记语言数据的编码

背景

当通过网络在发送方和接收方（例如，服务器和服务器）之间传输数据时，在传输进行之前，发送方和接收方都必须知道正传输的数据的格式。例如，如果发送方以用于特定数据库的形式发送数据，则为了使用该数据，接收方必须知道正使用了哪个数据库格式，且必须知道该格式的细节。如果接收方不知道正使用的格式或该格式的细节，则在发送方端正确发送的数据在接收方端将不可识别。

作为示例，数据库格式可包括一连串记录，其中每一记录包含某一大小的记录号，继之以某一大小的姓字段、某一大小的名字段以及某一大小的数据字段。标头可先于这些记录。然而，即使发送方完全遵循该格式来发送数据，但除非接收方知道该格式，否则接收方无法正确理解该数据。

为了确保发送方和接收方两者都具有关于格式的必要信息，通常它们不仅需要运行同样的应用程序，而且需要运行该应用程序的同一版本。例如，如果发送方将来自数据库应用的较新版本的数据发送至运行较老版本的接收方，则接收方的版本可能不能识别该格式，如上所述数据可能丢失或无用。

为了帮助解决这些问题并增加传输时的灵活性，开发了可扩展标记语言（XML），它是基于标准通用标记语言（SGML）的标记语言。标记语言是允许以结构化的方式连同诸如样式、句法和语义信息一起提供内容的语言。XML 被称为可扩展的是因为它不是固定格式标记语言。HTML（超文本标记语言）是定义一种格式的固定格式标记语言。相反，XML 是实际上为元格式的标记语言，是一种允许用户描述其它格式的语言。这允许用户设计标记语言，然后用 XML 来表达它。因此，XML 提供允许格式上的灵活性的灵活标准化数据存储格式，因此可便于发送方与接收方之间的交互，即使没有对严格格式的预先协定。为此，XML 使用类似于（HTML）的基于文字的标签系统以便以结构化的方式描述和存储数据。例如，雇员记录的数据库条目可以如下以 XML 格式表示：

```
<employee>
    <firstname>John</firstname>
    <lastname>Smith</lastname>
```

</employee>

该 XML 数据包括两种类型的元素——标签元素，它以尖括号开始和结束（例如，诸如“<firstname（名）>”的开始标签以及诸如“</firstname>”的结束标签），以及数据元素（例如，“John”）。如图所示，在 XML 文档中，开始和结束标签可嵌套在其它开始和结束标签内。特定元素内出现的所有元素使其开始和结束元素出现在该特定元素的结束标签之前。这定义了树形结构。

以上的示例 XML 包括数据元素“John”和“Smith”，但也包括指示数据元素“John”是 firstname 且连同 lastname（姓）“Smith”也是雇员记录的一部分的信息（在标签元素中）。如果发送方发送该 XML 文件，则识别 XML 的任何应用都能够读取该雇员记录、检索数据并理解其组件。

尽管 XML 不要求接收方知道正在使用哪一文件格式以及该文件格式的细节，但它确有缺点。首先，正在发送的文件由于用于描述数据的大量标签元素而极端庞大。事实上，XML 文件的平均大小比正常数据文件大 2-10 倍。这些较大的文件大小减慢了正在发送的数据的传输时间，且也要求更长的处理时间。从而，传输和消费 XML 可能非常昂贵。

为了平衡灵活性与较快传输和较小文件大小的竞争利益，可使用被称为二进制 XML 的某些技术。尽管不同的二进制 XML 技术可根据所涉及的技术而变化，但在每一二进制 XML 格式中有两个公共特征。

首先，二进制 XML 格式流传送二进制值而非基于字符的值。其次，二进制 XML 格式通过用较短的标记（token）替换标签来对 XML 格式“标记化”。例如，二进制 XML 格式可为以上所示的标签分配以下二进制表示：

1: <employee>

2: </employee>

3: <firstname>

4: </firstname>

5: <lastname>

6: </lastname>

以上所示的记录然后可被呈现为：

1 3 John 4 5 Smith 6 2

（以上示出的数字可用二进制格式呈现；空格没有意义，而仅用于增强显示标记语言文档时的理解。）对基于文字的标签这样的标记表示的代入产生压缩文件，

这可产生大小为原始 XML 文件的四分之一或三分之一的 XML 文件。标签的标记化或者根据某一预定义的标记/标签代入（发送方和接收方双方都已知，被称为“静态字典”）进行，或者根据作为所传输的文件的一部分发送的定义（这样传输的定义被称为“动态字典”）进行。

尽管文件大小较小，但二进制 XML 技术仍有缺点。首先，存在使该技术低效的冗余代入。例如，如果一数字被用作未经压缩的 XML 文件中的标签，则当使用二进制 XML 时它可能被编码成一不同的数字且然后必须被解码，这未节省空间却有了进行编码/解码的成本。此外，即使当使用二进制 XML 技术时，因为众多标签被重复，数据未被完全压缩成最小文件大小。这可由在单个 XML 文件中包含使用相同标签的众多数据记录的情况示出。在这样的情况中，即使当编码时如<lastname>的基于文字的标签将由数值替换，但仍有相同标签被重复的多个实例。

因此，需要更高效地编码数据并将其编码成更小的文件大小的技术。

概述

根据本发明的一些实施例，通过使用至少一个参数化宏将标记语言文档转换成压紧（compacted）的标记语言形式。该宏用于以更紧凑的形式替换标记语言文档中找到的元素。参数化宏展开为包括作为宏的参数给出的至少某些参数数据在内的元素（标签和/或数据元素）的有序集。以此方式，标记语言数据被压紧。

参数化宏的定义在静态字典或动态字典中找到。如果定义位于动态字典中，则在某些实施例中，该定义随压紧的标记语言数据传输。在某些情况中，所使用的宏将是动态定义的宏与在静态字典中定义的其它宏的混合。

然后可通过使用宏定义将包含在压紧的标记语言数据中的宏和参数展开成它们所表示的元素系列从而将该压紧的标记语言文档转换成未压紧的标记语言形式。

这些和其它实施例将在以下更全面描述。

附图简述

当结合附图阅读时，前述概述以及以下优选实施例的详细描述将被更好地理解。出于说明本发明的目的，在附图中示出了本发明的示例性构造；然而，本发明不限于所公开的具体方法和手段。在附图中：

图 1 是可在其中实现本发明的各方面的示例性计算环境的框图；

图 2 是可在其中实现本发明的示例性系统的框图；

图 3 是用于编码标记语言信息的方法的流程图；

图 4 示出了根据本发明的一个实施例的用于准备供传输的标记语言信息的系统；以及

图 5 是描述根据本发明的一个实施例的将压缩标记语言文档的解码的流程图。

说明性实施例的详细描述

示例性计算环境

图 1 示出了可在其中实现本发明各方面的示例性计算系统环境。计算系统环境 100 只是合适的计算环境的一个示例，并不旨在对本发明的使用范围或功能提出任何限制。也不应该把计算环境 100 解释为对示例性计算环境 100 中示出的任一组件或其组合有任何依赖性或要求。

本发明可用众多其它通用或专用计算系统环境或配置来操作。适合在本发明中使用的公知的计算系统、环境和/或配置的示例包括，但不限于，个人计算机、服务器计算机、手持或膝上型设备、多处理器系统、基于微处理器的系统、机顶盒、可编程消费者电子产品、网络 PC、小型机、大型机、嵌入式系统、包含上述系统或设备中的任一个的分布式计算机环境等。

本发明可在诸如程序模块等由计算机执行的计算机可执行指令的通用语境中描述。一般而言，程序模块包括执行特定任务或实现特定抽象数据类型的例程、程序、对象、组件、数据结构等。本发明也可以在分布式计算环境中实现，其中任务由通过通信网络或其它数据传输介质链接的远程处理设备执行。在分布式计算环境中，程序模块可以位于包括存储器存储设备在内的本地和远程计算机存储介质中。

参考图 1，用于实现本发明的一个示例性系统包括计算机 110 形式的通用计算设备。计算机 110 的组件可以包括，但不限于，处理单元 120、系统存储器 130 和将包括系统存储器在内的各种系统组件耦合至处理单元 120 的系统总线 121。处理单元 120 可表示诸如在多线程处理器上所支持的多重逻辑处理单元。系统总线 121 可以是若干类型的总线结构中的任一种，包括存储器总线或存储器控制器、外围总线和使用各种总线体系结构中的任一种的局部总线。作为示例，而非限制，这样的体系结构包括工业标准体系结构（ISA）总线、微通道体系结构（MCA）总线、扩展的 ISA（EISA）总线、视频电子技术标准协会（VESA）局部总线和外围部件互连（PCI）总线（也被称为 Mezzanine 总线）。系统总线 121 也可被实现为点对点连接、交换结构等通信设备。

计算机 110 通常包括各种计算机可读介质。计算机可读介质可以是能够被计算机 110 访问的任何可用介质，且包括易失性和非易失性介质、可移动和不可移动介质。作为示例，而非限制，计算机可读介质可以包括计算机存储介质和通信介质。计算机存储介质包括以任何方法或技术实现的用于存储诸如计算机可读指令、数据结构、程序模块或其它数据等信息的易失性和非易失性、可移动和不可移动介质。计算机存储介质包括，但不限于，RAM、ROM、EEPROM、闪存或其它存储器技术；CD-ROM、数字多功能盘（DVD）或其它光盘存储；磁带盒、磁带、磁盘存储或其它磁性存储设备；或能用于存储所需信息且可以由计算机 110 访问的任何其它介质。通信介质通常具体化为诸如载波或其它传输机制等已调制数据信号中的计算机可读指令、数据结构、程序模块或其它数据，且包含任何信息传递介质。术语“已调制数据信号”指的是这样一种信号，其一个或多个特征以在信号中编码信息的方式被设定或更改。作为示例，而非限制，通信介质包括诸如有线网络或直接线连接的有线介质，以及诸如声学、RF、红外线和其它无线介质的无线介质。上述中任一个的组合也应包括在计算机可读介质的范围之内。

系统存储器 130 包括易失性和/或非易失性存储器形式的计算机存储介质，诸如只读存储器（ROM）131 和随机存取存储器（RAM）132。基本输入/输出系统 133（BIOS）包含有助于诸如启动时在计算机 110 中元件之间传递信息的基本例程，它通常被存储在 ROM 131 中。RAM 132 通常包含处理单元 120 可以立即访问和/或目前正在处理单元 120 上操作的数据和/或程序模块。作为示例，而非限制，图 1 示出了操作系统 134、应用程序 135、其它程序模块 136 和程序数据 137。

计算机 110 也可以包括其它可移动/不可移动、易失性/非易失性计算机存储介质。仅作为示例，图 1 示出了从不可移动、非易失性磁介质中读取或向其写入的硬盘驱动器 140，从可移动、非易失性磁盘 152 中读取或向其写入的磁盘驱动器 151，以及从诸如 CD ROM 或其它光学介质等可移动、非易失性光盘 156 中读取或向其写入的光盘驱动器 155。可以在示例性操作环境下使用的其它可移动/不可移动、易失性/非易失性计算机存储介质包括，但不限于，盒式磁带、闪存卡、数字多功能盘、数字录像带、固态 RAM、固态 ROM 等。硬盘驱动器 141 通常由诸如接口 140 的不可移动存储器接口连接至系统总线 121，磁盘驱动器 151 和光盘驱动器 155 通常由诸如接口 150 的可移动存储器接口连接至系统总线 121。

以上讨论和在图 1 中示出的驱动器及其相关联的计算机存储介质为计算机 110 提供了对计算机可读指令、数据结构、程序模块和其它数据的存储。例如，在

图 1 中，硬盘驱动器 141 被示为存储操作系统 144、应用程序 145、其它程序模块 146 和程序数据 147。注意，这些组件可以与操作系统 134、应用程序 135、其它程序模块 136 和程序数据 137 相同或不同。操作系统 144、应用程序 145、其它程序模块 146 和程序数据 147 在这里被标注了不同的标号是为了说明至少它们是不同的副本。用户可以通过输入设备，诸如键盘 162 和定点设备 161（通常指鼠标、跟踪球或触摸垫）向计算机 110 输入命令和信息。其它输入设备（未示出）可以包括麦克风、操纵杆、游戏手柄、圆盘式卫星天线、扫描仪等。这些和其它输入设备通常由耦合至系统总线的用户输入接口 160 连接至处理单元 120，但也可以由其它接口或总线结构，诸如并行端口、游戏端口或通用串行总线（USB）连接。监视器 191 或其它类型的显示设备也经由接口，诸如视频接口 190 连接至系统总线 121。除监视器以外，计算机也可以包括其它外围输出设备，诸如扬声器 197 和打印机 196，它们可以通过输出外围接口 195 连接。

计算机 110 可使用至一个或多个远程计算机，诸如远程计算机 180 的逻辑连接在网络化环境下操作。远程计算机 180 可以是个人计算机、服务器、路由器、网络 PC、对等设备或其它常见网络节点，且通常包括上文相对于计算机 110 描述的许多或所有元件，尽管在图 1 中只示出存储器存储设备 181。图 1 中所示逻辑连接包括局域网（LAN）171 和广域网（WAN）173，但也可以包括其它网络。这样的联网环境在办公室、企业范围计算机网络、内联网和因特网中是常见的。

当在 LAN 联网环境中使用时，计算机 110 通过网络接口或适配器 170 连接至 LAN 171。当在 WAN 联网环境中使用时，计算机 110 通常包括调制解调器 172 或用于在诸如因特网等 WAN 173 上建立通信的其它装置。调制解调器 172 可以是内置或外置的，它可以通过用户输入接口 160 或其它合适的机制连接至系统总线 121。在网络化环境中，相对于计算机 110 描述的程序模块或其部分可以存储在远程存储器存储设备中。作为示例，而非限制，图 1 示出了远程应用程序 185 驻留在存储器设备 181 上。可以理解，所示的网络连接是示例性的，且可以使用在计算机之间建立通信链路的其它手段。

计算环境 100 一般包括至少某种形式的计算机可读介质。计算机可读介质可以是能够由计算环境 100 访问的任何可用介质。作为示例，而非限制，计算机可读介质可以包括计算机存储介质和通信介质。计算机存储介质包括以任何方法或技术实现的用于存储诸如计算机可读指令、数据结构、程序模块或其它数据等信息的易失性和非易失性、可移动和不可移动介质。计算机存储介质包括，但不限于，RAM、

ROM、EEPROM、闪存或其它存储器技术；CD-ROM、数字多功能盘（DVD）或其它光盘存储；磁带盒、磁带、磁盘存储或其它磁性存储设备；或能用于存储所需信息且可以由计算环境 100 访问的任何其它介质。通信介质通常具体化为诸如载波或其它传输机制等已调制数据信号中的计算机可读指令、数据结构、程序模块或其它数据，且包含任何信息传递介质。术语“已调制数据信号”指的是这样一种信号，其一个或多个特征以在信号中编码信息的方式被设定或更改。作为示例，而非限制，通信介质包括诸如有线网络或直接线连接的有线介质，以及诸如声学、RF、红外线和其它无线介质的无线介质。上述中任一个的组合也应包括在计算机可读介质的范围之内。

使用参数化宏编码标记语言文档

参考图 2，可在其中实现本发明的示例性系统包括发送方 210、网络 220 和接收方 230。发送方 210 和接收方 230 经由网络 220 连接，该网络可由私有网络（例如 LAN）和/或公共网络（例如因特网）组成。在本发明的一个实施例中，发送方 210 使用一个或多个参数化宏压缩标记语言文档以便传输。在本发明的一个实施例中，接收方接收压缩的标记语言文档，并使用该压缩文档中所引用的一个或多个参数化宏来对文档解压以便使用。一般，宏包括关于如何使用参数来生成序列的信息。不旨在对宏类型进行限制——在没有限制的情况下，构想了仅包含示出应如何使用参数来生成序列的定义的简单宏。另外，构想了可执行宏，例如被表达为或包括运行时产生该序列的代码的宏。

在某些实施例中，通过标识可使用参数化宏编码的至少一个序列（包括标签元素和数据元素）来编码标记语言信息以便传输。图 3 是用于编码标记语言信息的方法的流程图。如图 3 中所示，在第一步骤 310 中，选择对应于标记语言文档中一序列的参数化宏。所选的参数化宏是取包括来自该序列的数据元素在内的一个或多个参数的宏，且当该宏与这些参数一起使用时产生该序列。在第二步骤 320 中，用对该宏和该参数的引用来代入标记语言信息中的该序列（用对宏和宏的参数的引用来代入标记语言信息中的该序列）。

在步骤 310 中，对宏的选择可以是从预先存在宏集合中选择宏。例如，为发送方已知并假定为接收方已知的静态字典可以是所选宏的源。或者，可生成对宏的选择以与该序列相对应。例如，如果标记语言信息包含仅数据元素不同的若干相似的序列，则可创建当与作为参数的数据元素一起使用时生成该正确序列的宏。在这

样的情况中，这样的宏可用于替换每一个这样的序列（使用不同的参数来正确生成序列）。

在某些实施例中，当宏被生成且它不是静态字典的一部分时，宏定义可被包括在传输的压缩标记语言信息中。更一般地，接收方展开标记语言信息所需的任何宏的定义可被包括作为压缩标记语言信息的一部分。或者，这样的定义可用其它方式向接收方公开。例如，可向接收方公开单独的字典文件。例如，如果压缩了多个标记语言文件，且使用了不是接收方可用的静态字典一部分的公共的宏集，则该公共宏集仅需一次传输给发送方。在一单独字典文件中或仅在一个压缩文件中将公共宏集传输给接收方是足够的。以此方式，接收方将具有关于所使用的宏的信息，而无需传输同一宏定义的多个副本。公开宏定义的这些技术不是互斥的，而可适当地一起使用。

如箭头 325 所示，步骤 310 和 320 可在替换若干序列的情况下迭代执行。箭头 325 被示为虚线是因为该迭代并不是存在于所有实施例中。可使用例如 Lempel Ziv Welch（LZW）算法的变型的无损数据压缩算法，通过选择压缩来自数据的哪一序列或如何生成将提供更好的压缩的宏以促进该过程。

尽管所替换的初始序列由标记语言的元素组成，但随着迭代继续，压缩标记语言信息将包括宏引用。这样的宏引用也可能被包括在由宏引用替换的序列中。除如图 3 中所示的使用参数化宏以外，也可使用非参数化宏来压缩标记语言信息。另外，可使用标记化来进一步压缩标记语言信息。标记化可在用宏引用和参数代入序列的过程之前或之后进行。

作为压缩过程的实例，考虑以下标记语言数据：

```
<file>
    <record>
        <firstname>John</firstname>
        <lastname>Adams</lastname>
        <birthyear>1735</birthyear>
         <inauguration_year>1797</inauguration_year>
    </record>
    <record>
        <firstname>John</firstname>
        <lastname>Hancock</lastname>
        <birthyear>1737</birthyear>
    </record>
    <record>
```

```
            <firstname>Thomas</firstname>
            <lastname>Paine</lastname>
            <birthyear>1737</birthyear>
        </record>
        <record>
            <firstname>Patrick</firstname>
            <lastname>Henry</lastname>
            <birthyear>1735</birthyear>
        </record>

    </file>
```

（空格在此示例中没有意义，而仅为便于理解给出。）可以看到，该文件包含四个记录，每一个包含 firstname（名）、lastname（姓）以及 birthyear（出生年份）。另外，一个记录包含 inauguration_year（就职年份）。在一个示例中，在步骤 310，选择对应于序列：“<record> <firstname> John </firstname> <lastname> Adams </lastname> <birthyear> 1735 </birthyear>”的宏。该宏 M1 当使用参数 p1、p2、p3 时产生“<record> <firstname> p1 </firstname> <lastname> p2 </lastname> <birthyear> p3 </birthyear>”。该宏可以是现有宏，或可专门创建。当 p1 为“John”，p2 为“Adams”，p3 为“1735”时，该宏产生正确的序列。在步骤 320 中，在标记语言信息中用宏引用和参数信息代入该序列。因此，修改后的标记语言信息为：

```
    <file>
            M1 John Adams 1735
            <inauguration_year>1797</inauguration_year>
        </record>
        <record>
            <firstname>John</firstname>
            <lastname>Hancock</lastname>
            <birthyear>1737</birthyear>
        </record>
        <record>
            <firstname>Thomas</firstname>
            <lastname>Paine</lastname>
            <birthyear>1737</birthyear>
        </record>
        <record>
            <firstname>Patrick</firstname>
            <lastname>Henry</lastname>
```

```
            <birthyear>1735</birthyear>
        </record>

    </file>
```

也可使用该宏 M1 来代入三个其它位置（以<record>开始并以</birthyear>结束的每一序列）。这三个外加代入之后的结果为：

```
    <file>
            M1 John Adams 1735
            <inauguration_year>1797</inauguration_year>
        </record>
            M1 John Hancock 1737
        </record>
            M1 Thomas Paine 1737
        </record>
            M1 Patrick Henry 1735
        </record>

    </file>
```

可使用另一个宏 M2 来进行另外两次代入，当宏 M2 与参数 p4 和 p5 一起使用时产生“M1 p4 p5 1737 </record>”。这产生如下的压缩标记语言信息：

```
    <file>
            M1 John Adams 1735
            <inauguration_year>1797</inauguration_year>
        </record>
            M2 John Hancock
            M2 Thomas Paine
            M1 Patrick Henry 1735
        </record>

    </file>
```

在此示例中，可以看到，标记语言信息的大小大大减少。一般，标记语言信息中初始存在的重复序列的数目和长度与以宏引用代入序列的益处直接相关。可使用标记化或其它技术以便于进一步压缩经压缩的标记语言信息的大小。

图 4 示出了根据本发明的一个实施例的用于准备标记语言信息以便传输的系统。如图所示，系统 400 包括宏存储 410、宏选择器 420 以及序列代入器 430。宏存储 410 存储在压缩标记语言信息时使用的宏。宏选择器 420 选择在用对宏和参数的引用代入标记语言信息中的序列时使用的宏。序列代入器 430 作出代入。宏选择器 420 也可生成宏，并将其存储在宏存储 410 中，然后选择该宏用于代入序列。宏

存储410可加载有一个或多个静态字典的内容。可由序列代入器430替换若干序列，包括含有之前由序列代入器430放在那里的宏的序列。另外，在一个实施例中，使用标记化器来标记化修改后的标记语言信息。

使用参数化宏解码标记语言文档

为了解码压缩标记语言信息，使用该压缩标记语言信息中引用的宏的定义来产生代入该压缩标记语言信息中的宏数据（宏引用和参数引用）的序列。图5是示出了根据本发明的一个实施例的用于压缩标记语言文档的解码的流程图。如图5中所示，在步骤510中，接收包含宏数据（宏引用和参数应用）的信息。

在步骤520中，获取所引用宏的宏定义。宏可从静态字典获取。或者，宏可从包括在压缩标记语言文档中的定义、从相关数据文件、或从某个其它来源获取。

在步骤530中，应用该宏定义，产生结果序列，在步骤540中，该结果序列被代入到该信息中。可迭代地执行这些步骤以便通过代入来“解压”多个宏引用。

当该压缩标记语言文档已经过标记化时，接收信息的步骤510还包括展开这些标记——实际上是反转标记化过程——以便于产生标记化之前的文档。

可迭代执行步骤510、520和530，替换若干序列。当在压缩标记语言文档中存在对非参数化宏的引用时，也根据该宏定义通过代入来展开对该非参数化宏的引用。

可使用包括解码器的系统以将已编码文档解码。因为，必须应用宏来产生期望形式的标记语言文档，因此宏可以说是驱动了解码器。

结论

注意到，仅出于解释的目的而提供了前述示例，它们决不被解释为对本发明的限制。尽管参考各个实施例描述了本发明，但可以理解，此处所使用的词语是描述和说明的词语，而非限制的词语。此外，尽管此处参考特定手段、材料和实施例描述了本发明，但本发明不旨在限于此处所公开的细节；相反，本发明延及所有功能上等效的结构、方法和使用，诸如落入所附权利要求书范围内的那些。受益于本说明书的教导的本领域技术人员可实现对此的各种修改，且改变可在不背离本发明在其各方面中的范围和精神的情况下作出。

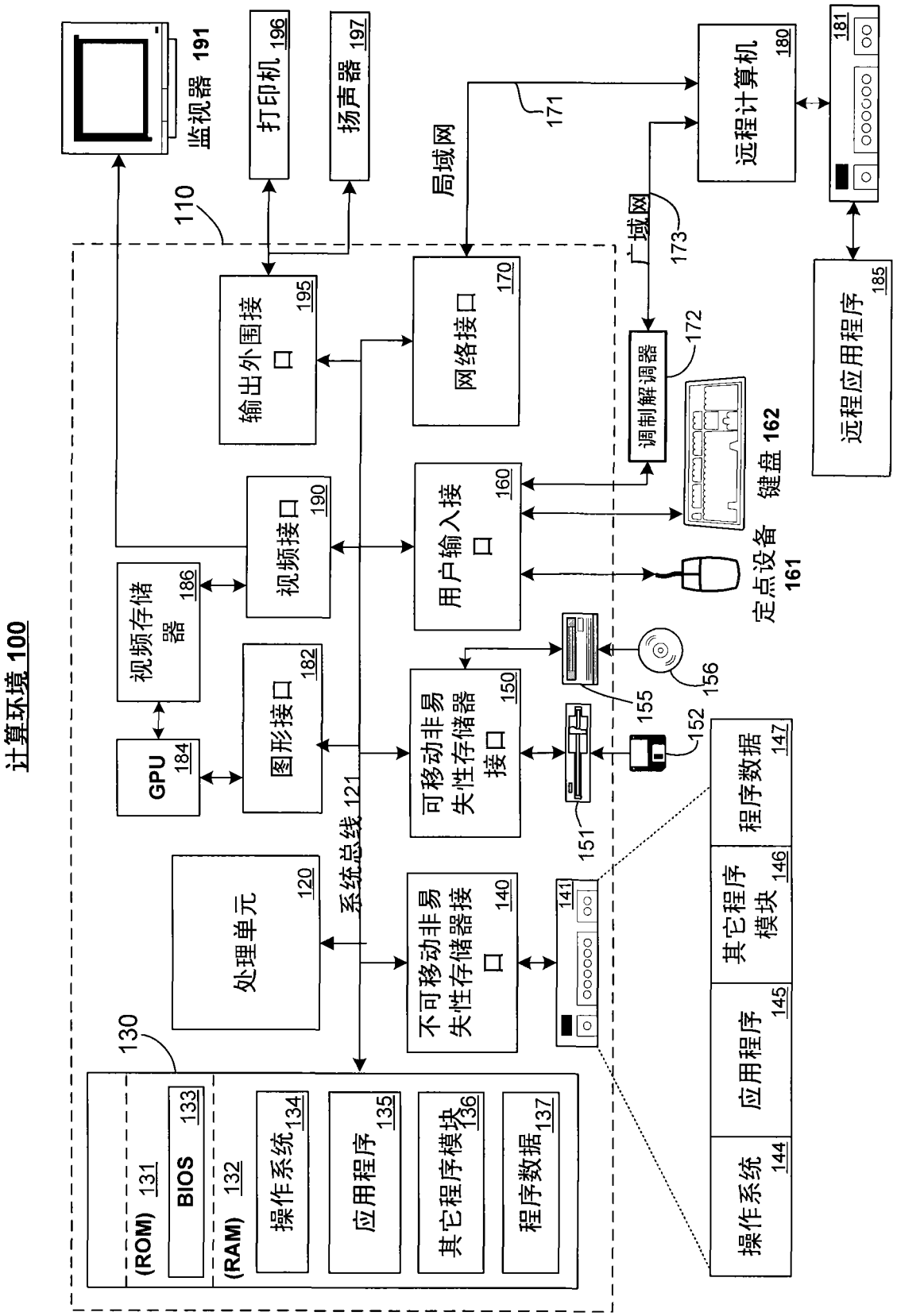

计算环境 100
110
监视器 191
打印机 196
扬声器 197
局域网
171
远程计算机 180
181
广域网
173
输出外围接口 195
网络接口 170
远程应用程序 185
调制解调器 172
键盘 162
定点设备 161
用户输入接口 160
视频接口 190
视频存储器 186
图形接口 182
GPU 184
可移动非易失性存储器接口 150
155
156
152
151
处理单元 120
系统总线 121
不可移动非易失性存储器接口 140
141
130
(ROM) 131
BIOS 133
(RAM) 132
操作系统 134
应用程序 135
其它程序模块 136
程序数据 137
操作系统 144
应用程序 145
其它程序模块 146
程序数据 147

图 1

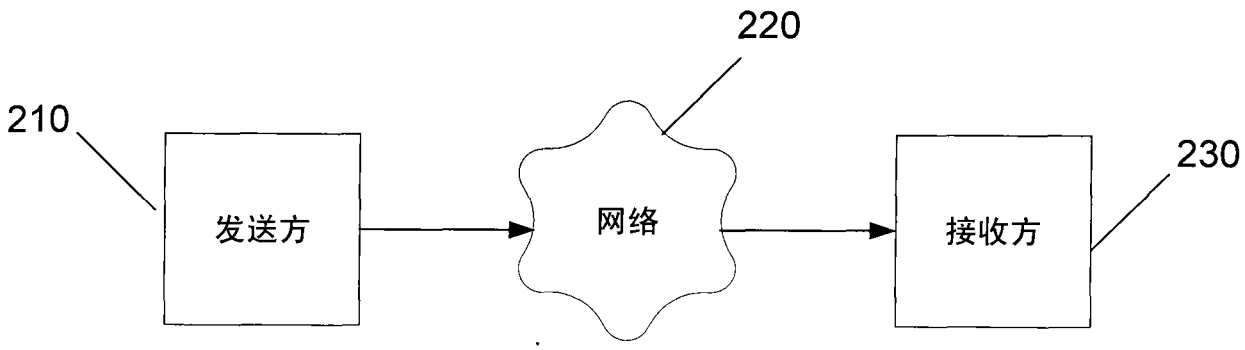

220
210
发送方
网络
接收方
230

图 2

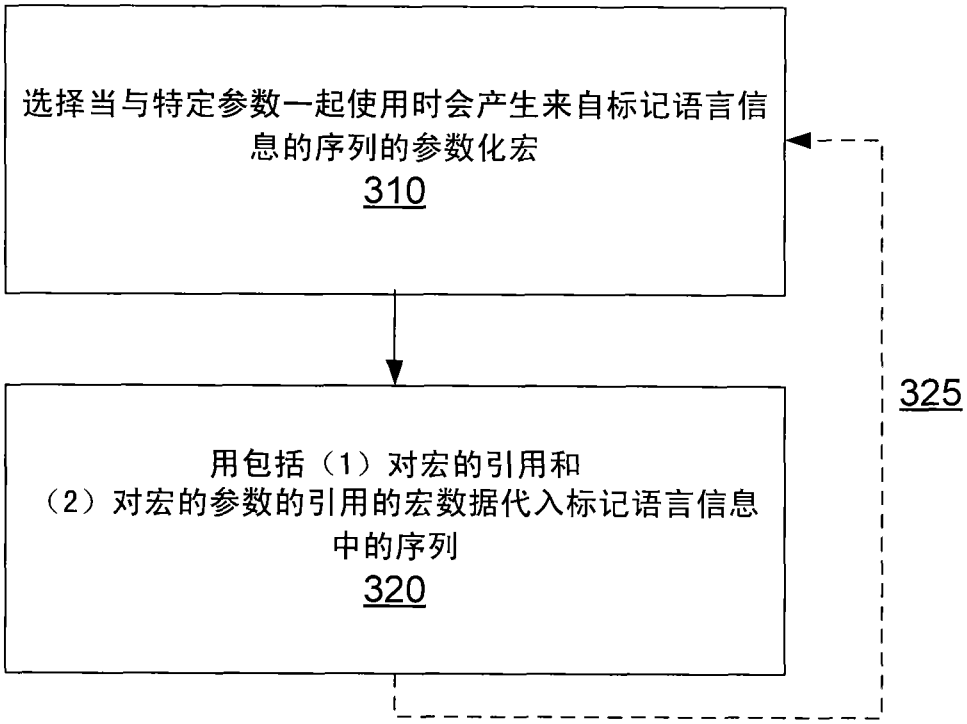

选择当与特定参数一起使用时会产生来自标记语言信息的序列的参数化宏
310
用包括（1）对宏的引用和
（2）对宏的参数的引用的宏数据代入标记语言信息中的序列
320
325

图 3

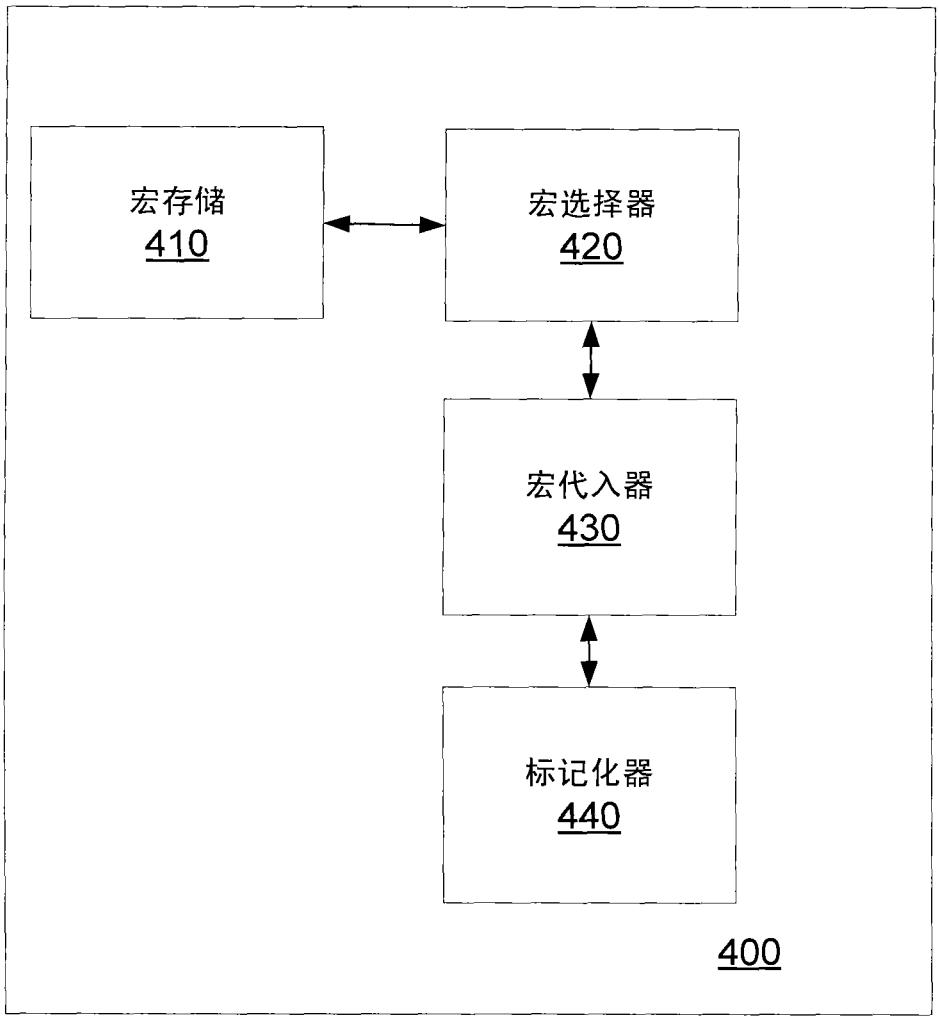


图 4

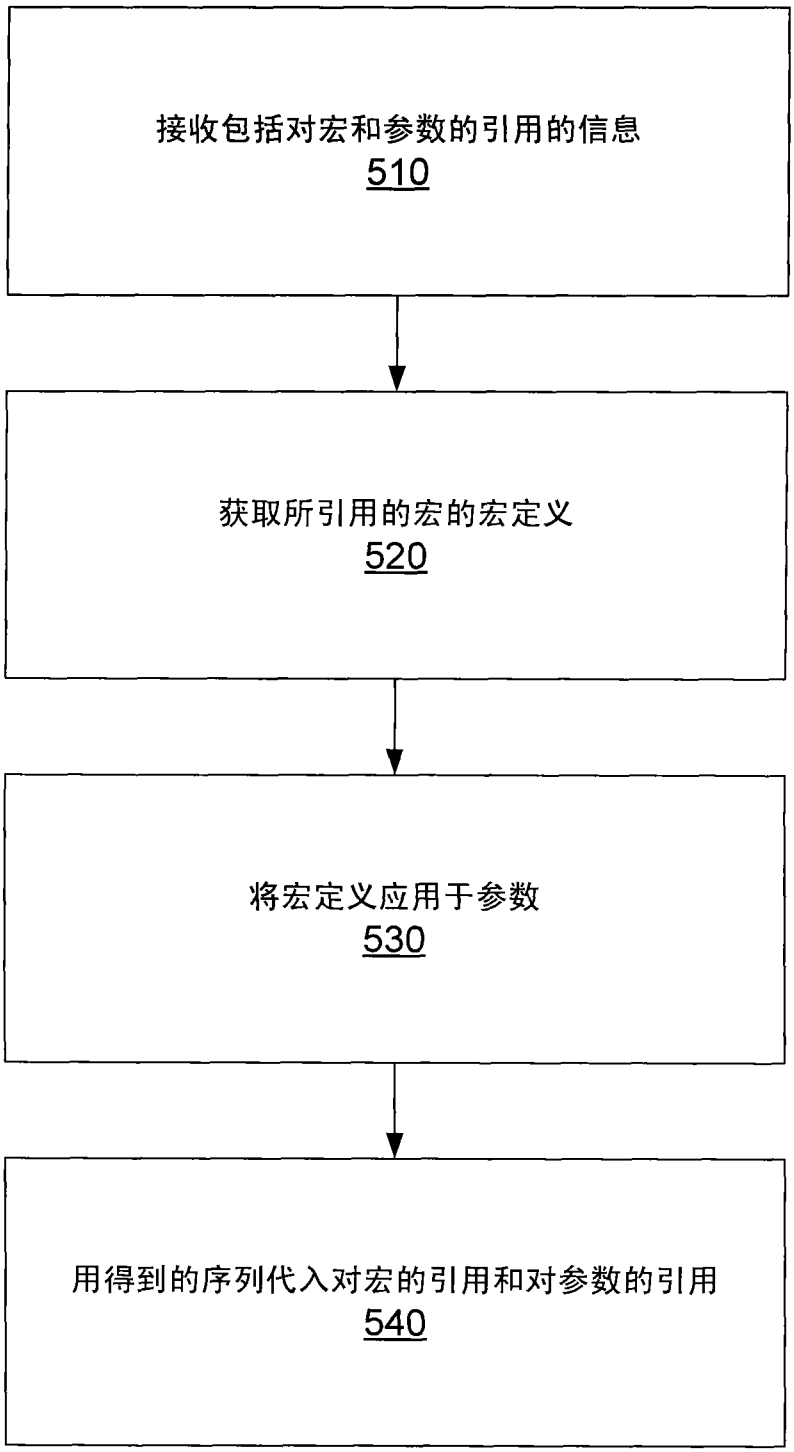

接收包括对宏和参数的引用的信息
510
获取所引用的宏的宏定义
520
将宏定义应用于参数
530
用得到的序列代入对宏的引用和对参数的引用
540

图 5