



# (12)发明专利申请

(10)申请公布号 CN 107094021 A

(43)申请公布日 2017.08.25

(21)申请号 201710080667.0

(22)申请日 2007.06.18

(30)优先权数据

11/425,137 2006.06.19 US

(62)分案原申请数据

200780023129.2 2007.06.18

(71)申请人 埃塞克斯PA有限责任公司

地址 美国特拉华州

(72)发明人 D·M·门罗

(74)专利代理机构 北京市金杜律师事务所

11256

代理人 王茂华 黄倩

(51)Int.Cl.

H03M 7/30(2006.01)

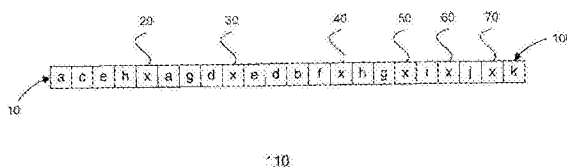
权利要求书2页 说明书8页 附图2页

(54)发明名称

数据压缩

(57)摘要

本发明涉及数据压缩。该数据压缩的方法可以包括：选择数据集内出现的符号串，所述符号串是一个或多个字符；生成与所述符号串在所述数据集内的各个位置相对应的符号串码；对于所述数据集内出现的任何其他符号串，重复所述选择符号串以及所述生成符号串码，以产生任何相应的其他符号串码；以及通过组合各个符号串码来生成压缩数据码。



1. 一种压缩数据集的方法,包括:  
选择所述数据集内出现的符号串,所述符号串是一个或多个字符;  
生成与所述符号串在所述数据集内的各个位置相对应的符号串码,如果符号串包括符号串的重复,则符号串码包括指示所述符号串的一连串出现的存在的数据;  
对于所述数据集内出现的任何其他符号串,重复所述选择符号串以及所述生成符号串码,以产生任何相应的其他符号串码;以及  
通过组合各个符号串码来生成压缩数据码。
2. 根据权利要求1所述的方法,还包括:对于所述各个符号串码中的一个符号串码,使用行程长度编码对所述符号串的各个位置进行编码。
3. 根据权利要求1所述的方法,还包括:以组结尾EOG符号结束所述符号串码。
4. 根据权利要求1所述的方法,其中所述生成符号串码包括:确定所述符号串在所述数据集内的出现之间的距离,以及基于所述距离生成所述符号串码。
5. 根据权利要求4所述的方法,其中所述生成符号串码包括:当先前已编码的符号串的位置被省略时,确定所述符号串在所述数据集内的出现之间的所述距离。
6. 根据权利要求1所述的方法,其中所述选择符号串包括从具有已知长度的数据集中选择符号串,并且  
其中所述生成符号串码包括避免对最后一个符号串进行编码。
7. 根据权利要求1所述的方法,其中所述选择符号串包括从具有标准长度的数据集中选择符号串。
8. 根据权利要求1所述的方法,还包括:将所述数据集的长度作为所述压缩数据码的前缀,或者在编码之前提供所述数据集的所述长度。
9. 根据权利要求1所述的方法,还包括:采用所述符号串的预定次序,或者相对于所述符号串的出现频率对所述符号串进行排序。
10. 根据权利要求1所述的方法,其中所述选择符号串包括选择具有固定的预定义形式的符号串,或者选择形式上具有灵活性的符号串。
11. 根据权利要求1所述的方法,还包括:利用所述符号串作为所述符号串码的前缀或者后缀。
12. 根据权利要求1所述的方法,还包括:  
在第一模式期间以预定次序生成所述符号串码;以及  
在第二模式期间将所述符号串嵌入到所述压缩数据码中。
13. 根据权利要求1所述的方法,还包括:将“不再有组”信号解释为所述数据集的结尾。
14. 根据权利要求1所述的方法,还包括:将关于所述符号串的附加信息与所述符号串组合以形成新的符号串。
15. 根据权利要求1所述的方法,还包括:存储所述压缩数据码,或者通过通信信道发送所述压缩数据码。
16. 一种压缩数据集的设备,包括:  
用于选择所述数据集内出现的符号串的装置,所述符号串是一个或多个字符;  
用于生成与所述符号串在所述数据集内的各个位置相对应的符号串码的装置,如果符号串包括符号串的重复,则符号串码包括指示所述符号串的一连串出现的存在的数据;

用于对于所述数据集内出现的任何其他符号串重复所述选择符号串以及所述生成符号串码以产生任何相应的其他符号串码的装置;以及

用于通过组合各个符号串码来生成压缩数据码的装置。

17. 根据权利要求16所述的设备,还包括:用于对于所述各个符号串码中的一个符号串码使用行程长度编码对所述符号串的各个位置进行编码的装置。

18. 根据权利要求16所述的设备,还包括:用于以组结尾EOG符号结束所述符号串码的装置。

19. 根据权利要求16所述的设备,其中用于生成符号串码的装置包括:用于确定所述符号串在所述数据集内的出现之间的距离以及基于所述距离生成所述符号串码的装置。

20. 根据权利要求19所述的设备,其中用于生成符号串码的装置包括:用于当先前已编码的符号串的位置被省略时确定所述符号串在所述数据集内的出现之间的所述距离的装置。

21. 根据权利要求16所述的设备,其中用于选择符号串的装置包括用于从具有已知长度的数据集中选择符号串的装置,并且

其中用于生成符号串码的装置包括用于避免对最后一个符号串进行编码的装置。

22. 根据权利要求16所述的设备,其中用于选择符号串的装置包括用于从具有标准长度的数据集中选择符号串的装置。

23. 根据权利要求16所述的设备,还包括:用于将所述数据集的长度作为所述压缩数据码的前缀或者在编码之前提供所述数据集的所述长度的装置。

24. 根据权利要求16所述的设备,还包括:用于采用所述符号串的预定次序或者相对于所述符号串的出现频率对所述符号串进行排序的装置。

25. 根据权利要求16所述的设备,其中用于选择符号串的装置包括用于选择具有固定的预定义形式的符号串或者选择形式上具有灵活性的符号串的装置。

26. 根据权利要求16所述的设备,还包括:用于利用所述符号串作为所述符号串码的前缀或者后缀的装置。

27. 根据权利要求16所述的设备,还包括:

用于在第一模式期间以预定次序生成所述符号串码的装置;以及

用于在第二模式期间将所述符号串嵌入到所述压缩数据码中的装置。

28. 根据权利要求16所述的设备,还包括:用于将“不再有组”信号解释为所述数据集的结尾的装置。

29. 根据权利要求16所述的设备,还包括:用于将关于所述符号串的附加信息与所述符号串组合以形成新的符号串的装置。

30. 根据权利要求16所述的设备,还包括:用于存储所述压缩数据码或者通过通信信道发送所述压缩数据码的装置。

## 数据压缩

[0001] 相关申请的交叉引用

[0002] 本申请是优先权日为2006年06月19日、国际申请日为2007年06月18日、国际申请号为PCT/US2007/014138、中国专利申请号为200780023129.2、发明名称为“数据压缩”的专利申请的分案申请。

### 技术领域

[0003] 本专利申请涉及数据压缩。

### 背景技术

[0004] 近年来,人们通过网络共享数据变得越来越流行;然而,数据传输在带宽利用方面具有一定成本。所以,例如,大量数据常常被压缩。压缩也可以用于存储数据到磁的或其它介质上。

### 附图说明

[0005] 主题在说明书的结论部分被特别地指出并清楚地请求保护。然而,所请求保护的主体,包括其组织和操作方法以及目的、特征和优点,可以通过结合附图参考以下的详细说明而被最好地理解,其中:

[0006] 图1是说明根据所请求保护的主体的一个实施例的示意图。

[0007] 图2示出一个示例性系统。

### 具体实施方式

[0008] 在以下的详细说明中,许多特定细节被阐述,以便提供对于所请求保护的主体透彻的理解。然而,本领域技术人员将理解,所请求保护的主体可无需这些特定细节而实施。在其它实例中,对熟知的方法、过程、部件和/或电路没有进行详细描述,以免使所请求保护的主体模糊。

[0009] 随后的具体实施方式的某些部分以算法和/或对存储在计算系统内——诸如在计算机和/或计算系统存储器内——的数据比特和/或二进制数字信号的操作的符号表示的形式而给出。这些算法描述和/或表示是由数据处理领域技术人员用来把他们的工作内容传达给该领域其它技术人员的技术。在这里并且通常地,算法被认为是自相一致的运算序列和/或导致期望结果的类似处理。运算和/或处理可包括对物理量的物理操作。典型地——虽然不一定必须,这些量可以采取能够被存储、传送、组合、比较和/或操纵的电信号和/或磁信号的形式。主要由于普遍的使用,把这些信号称为比特、数据、数值、单元、符号、字符、术语、数目、数字等等有时被证明是方便的。然而,应当理解,所有这些以及类似的术语是与适当的物理量相关联的,并且仅仅是方便的标签。除非专门阐述,否则正如从以下的讨论看到的,将会意识到在本说明书的讨论中,利用诸如“处理”、“计算”、“确定”等等的术语指的是诸如计算机或类似的电子计算设备那样的计算平台的动作和/或处理,所述计算

平台在计算平台的处理器、存储器、寄存器和/或其它信息存储、传输和/或显示设备内操纵和/或变换被表示为电和/或磁的物理量和/或其它物理量的数据。

[0010] 此后描述压缩数据的方法的一个特定实施例,虽然所请求保护的主体在范围上不限于这方面。选择在要被压缩的数据集或数据组内出现的一个符号串。生成符号串码,其指示符号串在数据内的一个或多个位置。同样地,对于数据内的另外的符号串,这个过程可以重复进行。各个符号串码然后可以组合以形成数据码。得到的数据码是该数据集或数据组的压缩形式。该数据码可根据需要被发送或存储。

[0011] 一个特定实施例可以通过接连地识别所选择的符号串出现的位置、确定在所选择的符号串的出现之间的距离并生成包括表示所选择的串在数据内的位置的码而被实行,虽然所请求保护的主体在范围上不限于这方面。同样地,在这里,符号可包括任何类型、形式或格式的数据。例如,这可包括记录、文件、扇区、集群、组和/或它们的部分。而且,符号可包括字、字节、比特、文本、字符等等。符号串可包括单个或多个符号。同样地,它们在长度上可以是固定的或可变的。

[0012] 对于一个特定实施例,符号串可以按特定的或预定的次序被编码,虽然这同样仅仅是一个示例性实施例并且所请求保护的主体在范围上不限于这方面。替换地或另外,不同于按次序编码,符号串可以以任何次序被编码。在这样的实施例中,符号串码例如可以加上指示特定的符号串的某些其它码作为前缀,虽然所请求保护的主体在范围上当然不限于这个示例性实施例。同样地,对于一个特定实施例,所采用的方法可以在模式之间切换,所述模式诸如其中符号串码按预定的次序被发送或存储的模式,和其中这个次序没有被预先确定、但其中表示相应的符号串的某些码在符号串码之前或作为符号串码的一部分而被发送的模式。

[0013] 如果符号串包括符号串的重复,则符号串码可包括指示符号串的一连串出现的存在的数据。从压缩的观点来看,指示一连串出现的数据的存储或传输通常比单独发送这样的出现更有效率。在一个特定实施例中,可以通过行程长度编码(run length coding)而将位置用信号通知(signal),但其它编码方法也可能代替或作为附加,正如下面更详细地描述的那样。同样地,编码可以至少部分地是依赖于上下文的。例如,如果要被编码的数据集或数据组的长度是事先已知的,则它可以作为第一符号串码的前缀。至少部分地依赖于实施例的细节,最后的符号串的位置可以被确定而无需发送或存储相应于该特定的符号串的附加信息,同样如下面更详细地描述的那样。

[0014] 而且,关于所有的或某些符号串的辅助信息或附加信息也可以被包括在数据码中。在一个特定实施例中,与特定的符号串有关的附加信息可以在相应的符号串码的结尾或否则与其相关联地被发送。替换地,附加信息可以在发送或存储符号串码后被发送或存储。更一般地,附加信息可以在任何时间或以任何方式被提供,以使得译码器能够把该信息与相应的符号串相关联。例如,要被使用的符号串的列表或表可以被预先确定或预定义。替换地或另外,例如,它可以至少部分基于在要压缩的数据中出现的符号串而被编制。

[0015] 在特定的数据集或数据组的符号串码被存储或被发送后,可以使用结尾码(end code)。在数据集中存在的符号串——除最后的符号串外——的位置被存储或传送后,在至少某些实施例中,可以确定最后的符号串的位置。因此,编码器可以省略存储或传送这样的最后的符号串的位置,正如此后更详细地描述的那样。

[0016] 同样地,由所请求保护的主体覆盖的实施例包括例如能够实施任何这样的方法的计算机程序,和/或能够实施这样的方法的计算机系统。实施例还包括能够实行如下面描述的方法的硬件和/或软件编码器,以及能够从代表数据的压缩的数据码重建原始的数据集或数据组的硬件和/或软件译码器。这些以及许多其它实施例将在所请求保护的主体范围内。

[0017] 在所描述的至少某些实施例中,压缩是无损的,虽然所请求保护的主体在范围上不限于这方面。例如,在某些实施例中,数据码可包括关于被包含在要被压缩的数据内的那些符号串的位置的信息。

[0018] 在这个特定的情形中,要被压缩的相关数据的任何组、集或部分被视作字符或其它符号的有序序列。如果这样的数据代表例如文本,则各个符号可包括一个或多个文本字符,但当然这不是必须的。许多其它的符号也可以被表示。更一般地,符号可以由字节或其它大小的数据组来表示。也有可能使用较长的或短的数据部分,它可以或不一定适合于在例如字节或数字字长度内。如果数据以二进制形式表示,则符号可以取决于具体的实施例而被表示为单个比特或具有固定或可变长度的多个比特。

[0019] 作为一个例子,诸如被存储为文件、记录或数据的其它单一联系之类的数据可以视作整体,或替代地被划分或分割成合适的长度,该长度对于符号组——在这里被称为符号串——足够长,使之可以以某个压缩效率被编码,但足够短到可以方便地处理。下面将更详细地描述各种方法。

[0020] 初始地,例如在一个特定实施例中,建立可在数据集内出现的符号串列表或表。加到这个列表的可以是至少部分基于关于数据的统计量的现有知识或信息的符号串的初始列表。为提供某些简单的例子,例如对于文本,通常的符号串可包括“ee”、诸如“and”或“or”之类经常出现的字、或后面跟随空白的标点符号。当然,所请求保护的主体在范围上不限于这些例子或这个特定实施例。许多可能的变化也将被包括在所请求保护的主体范围内。

[0021] 作为另一个例子,可以在编码开始之前检查特定的数据集,以识别经常出现的符号串。替换地或另外,如果进行划分,则这些划分例如可以通过使用符号串的初始列表而被分开地或独立地处理。例如,可以至少部分地基于可能已经在以前的数据划分中发现的符号串而确定该初始列表。

[0022] 替换地,作为例子,符号串可能随着编码进行而增加,如同在Lempel-Ziv-Welsh编码中发生的那样。然而,在这个实施例例子中,如在下面描述的,编码符号串与在LZW编码中使用的方法非常不同。在LZW中,符号串通过用另一个较长的符号或串进行替换而被编码。为此,LZW有时不能压缩数据集,并且在某些情形下可产生较长的数据集。相反,根据所请求保护的主体实施例很可能导致压缩,正如此后更详细地描述的那样。

[0023] 为了压缩数据组或数据集,可以从可得到的符号串列表中选择第一符号串。该符号串在数据中的出现被定位。第一符号串在数据内的位置被记住。这个过程然后对于数据的另外的符号串重复进行,以便规定这个数据集或数据组。虽然顺序地从数据集或数据组的开始到结尾进行处理可以是一种方便的方法,但是包括符号串的数据可以以任何次序被处理。

[0024] 典型地,编码可以通过硬件或软件编码器被实行。在一个可能的实施例中,编码器

可被配置成在编码后将数据通过通信信道发送到译码器,译码器可被配置成实时地或以其它方式利用接收的编码数据来重建数据集。

[0025] 再次地,所请求保护的主体在范围上不限于特定实施例。所以,以前或此后描述的实施例将仅仅作为例子用于说明。不同于这些具体例子的许多其它方法和/或实施例将被包括在所请求保护的主体范围内。尽管如此,继续说明这些例子,现在参照图1。图1是根据所请求保护的主体一个可能的实施例的示意图。

[0026] 在图1的例子中,示出一行文本以代表数据集或数据组。同样地,符号串——这里是字母“x”——在本例中将被处理。当然,所请求保护的主体不限于这个示例性实施例或任何一个特定实施例。这个例子仅仅是用于解释性说明。许多其它可能的实施例将被包括在所请求保护的主体范围内。

[0027] 从在这里被称为“start”、在图1上标示为10的位置开始,文本数据110在本例中被顺序地搜索。当然,如以前提到的,所请求保护的主体不限于以上顺序搜索。但是,如果找到所选符号串——这里还是“x”——的出现,则记住关于该出现的信息,在本例中诸如其位置。在所示的例子中,串“x”的第一次出现出现在开始位置10后的第5位置,如用标号20表示的。符号串“x”再次出现在开始后的第9位置,或从前一次出现向后四个位置,如用标号30表示的。

[0028] 继续进行该例子,“x”再次在前一次出现的五个位置后出现,如用标号40示出的。随后的位置间隔是3,2,2,如分别用标号50,60,70示出的。

[0029] 一旦文本数据110被这样地完全分析到“结尾”位置,即这里的100,则通过存储或发送码而生成表示在该部分数据集中的位置的该码。在这个特定实施例中,实际上,不出现“x”的位置被存储或被用信号通知(signal)。在该情形中,这被称为行程长度编码。这里,不包含x的位置的“行程”被编码。

[0030] 在图1的简单的示例性实施例中,数据包括一行文本,符号串包括一个字母。在所请求保护的主体范围内的该例子的变形中,可以寻找更长的和/或更复杂的符号串的出现。正如下面更详细地描述的,这些可包括具有固定的预定义形式的符号串,或替换地,例如可包括诸如在形式、长度和/或组成上具有灵活性的符号串。灵活性的程度可以预定义,或者可以某种程度上至少部分根据数据的某个或某些特征而计算。下面给出某些另外的示例性符号串。

[0031] • 任何字母、符号、或字符a,诸如“x”。原则上这可包括单个符号位置。

[0032] • 任何双字符 $a_1a_2$ ,诸如“st”。

[0033] • 任何组合 $a_1a_2a_3$ ,诸如“str”。

[0034] • 任何更长的组合 $a_1a_2a_3\cdots a_m$ ,诸如“st\_ng”,其中下划线代表单个符号位置。

[0035] 第一符号串的位置被确定后,可以确定在可能的或已知的符号串列表中的第二符号串的位置。通常,数据组可以这种方式被编码。对于某些实施例,随着编码进行,编码器可以把关于已被定位的符号串的信息——诸如一个或多个位置——实时地发送到译码器。替换地,编码的数据可以在本地被存储为压缩表示。

[0036] 示例性实施例可以用伪码表示如下:

[0037] 对于S=符号串

[0038] 指示S在数据集中的位置

[0039] 结束

[0040] 作为另一个例子,考虑8个符号串S1到S8的短序列S。为了说明起见,符号串在这里包括文本片段,但是所请求保护的主体在范围上不限于这方面。这样的片段是短的,而且符号串可能完全不包括文本,但仍旧在所请求保护的主体范围内。在本例中,在两个字之间的空格也包括符号串,如下面所示:

[0041]

S: “测试”

S1	S2	S3	S4	S5	S6	S7	S8
----	----	----	----	----	----	----	----

[0042]

t	h	e	空格	t	e	s	t
---	---	---	----	---	---	---	---

[0043] 在本例中,符号串的位置可以通过任何方便的方法来指示。用来传送或发送位置的方便的方法是表示为简单数字。以某个次序估计关于数据集的符号串组。为了说明起见,作为例子,假设符号串在这里具有次序‘e’‘t’‘s’‘h’‘space(空格)’,但是所请求保护的主体在范围上当然不限于这个例子或任何特定例子。但是,对于这个例子,以下的符号串可被如下地传送或存储:

[0044]

3 6	1 5 8	7	2	4
扫描 e	扫描 t	扫描 s	扫描 h	扫描空格

[0045] 上面,3和6指示符号串‘e’的位置;1、5和8指示符号串‘t’的位置;等等。

[0046] 对于这个特定实施例,期望一种技术,以指示对于特定符号串没有另外的信息被传送或存储。一种实现方法可以是定义一个用于组结尾的专门的符号(EOG)。对于这个示例性实施例,这可以给出:

[0047]

3 6 EOG	1 5 8 EOG	7 EOG	2 EOG	4 EOG
扫描 e	扫描 t	扫描 s	扫描 h	扫描空格

[0048] 然而,可能期望更有效的编码。例如,对于8个位置,通过传统的编码方法可以采用每个符号串3比特来传送,这与传送EOG非常不同。

[0049] 为了提高压缩效率,我们可以把行程长度编码(RLC)应用到接连的符号串的位置。可以使用任何形式或类型的编码来编码行程长度,并且所请求保护的主体不限于特定的形式或类型。例如,可以采用Golomb编码器或二进制算术编码器来提供几个例子。这些编码器例如可用来编码一连串0和1,指示在数据组中符号串不出现和出现(或反之亦然)的一个或多个位置。当然,这些仅仅是例子,并且所请求保护的主体在范围上不限于这样的例子。编码行程长度的许多其它可能的方法是可以想到的并且被包括在所请求保护的主体范围内。

[0050] 但是,继续进行这个例子,在这样的实施例中,符号串的位置可以被行程长度编码。在这种情形下,例如,代替编码3后跟随6来表示符号串‘e’的位置,可以编码2来表明在‘e’第一次出现之前跳过两个位置,2表明从位置3又跳过两个位置到达位置6,等等:

[0051]

2 2 EOG	0 3 2 EOG	6 EOG	1 EOG	3 EOG
扫描 e	扫描 t	扫描 s	扫描 h	扫描空格

[0052] 这通过引入6个符号——除了EOG以外还有0、1、2、3和6——而提供压缩,这是比9个符号——1到8加上EOG——更小的编码集。

[0053] 然而,压缩可以通过类似于“重要性切换 (significance switching)”的方法而被进一步改进。更具体地,这里,对于每个被编码的接连的符号串,随着越来越多的符号串变成由位置确定,要被编码的位置变得更少。这个关于位置的附加信息可用来提供附加压缩。

[0054] 在本例中,例如随着符号串‘e’的位置变成已知的,这些位置在作为用于以后的符号串的可能位置时可被省略。实际上,为了编码,是编码多少个要跳过的未知的位置,而不是编码要跳过的全部位置,由此提供附加压缩。在本例中,没有行程长度编码时是:

[0055]

0010 0100	10x0 1x01	x0x0 xx1x	x1xo xxxx	xxx1 xxxx
扫描 e	扫描 t	扫描 s	扫描 h	扫描空格

[0056] 这里的符号x表示已经知道从而不必用信号通知的位置。“0”和“1”的总数目从40减小到20,并且作为行程长度码被进一步减小:

[0057]

2 2 EOG	0 2 1 EOG	2 EOG	0 EOG	0 EOG
扫描 e	扫描 t	扫描 s	扫描 h	扫描空格

[0058] 在这个特定实施例中,所使用的不同的符号的数目被进一步减小到0、1、2和EOG。为了规定‘e’的位置,RLC是相同的,因为事先不知道位置。然而,在编码‘t’时,包含‘e’的位置可被省略,因为它们已经被占用,所以不必计入用于编码和/或译码的RLC中。

[0059] 而且,一旦编码器到达要被编码的最后的符号串,它的位置就可被确定而不用传送或编码任何附加信息。更具体地,在这个示例性实施例中,这个符号串占用先前没有用信号通知或编码的位置。然而,这假设要被编码的数据集的长度是已知的。如果是的话,在遇到针对‘h’的EOG后,在本例中剩余的符号串隐含地是‘space’,过程可以终结而不用附加编码或传送。所以在给定的例子中,假设已知‘space’是最后的符号串,附加压缩可以如下地得到:

[0060]

2 2 EOG	0 2 1 EOG	2 EOG	0 EOG	
扫描 e	扫描 t	扫描 s	扫描 h	扫描空格

[0061] 然而,如果数据集的长度和/或符号串的数目未知,在某些情形下也可能想要采用类似的方法。在例如被编码的数据集的长度是未知的情形下,可能有各种方法,并且所请求保护的专利在范围上不限于一种特定方法。例如,可以通过具有标准长度来处理。替换地,长度信息可以作为被编码的信息的前缀。然而,如果多个数据集被编码并且大多数具有相同的长度,从压缩观点来看,用长度作前缀是效率低的。同样地,继续这个想法,对于长数据通信,如果采用标准长度来再划分数据,则可以通过在小于标准长度的结尾处传送或编码数据集来处理偏离标准长度的变化。

[0062] 一种方法可包括在编码之前提供数据组的总长度。如果有标准划分长度,则编码器和译码器可以确定有多少个划分,以及最后的划分的长度,如果它小于标准长度的话。例如,如果102个符号串长度的数据集被编码,划分的标准长度是5,则存在21个划分,其中最后的一个长度为2。再次地,如前所述,总长度可以以许多方式被包括在数据码中,并且所给出的例子将不是限制性的。所请求保护的主体将包括这些示例性方法以及其它可能的方法。

[0063] 同样地,对于处理符号串,也有几个方法是可能的。例如,如前所述,可以采用符号串的预定的次序。然而,替换地,符号串可以相对于它们的出现频率被排序,如果出现频率已知或能够被确定或近似的话。例如,仅仅作为例子,使用英文时,可以具有以下的次序:‘e’ ‘t’ ‘a’ ‘o’ ‘i’ ‘n’ ‘s’ ‘h’ 等等。“空格”可以以它的统计量在样本中表示的这样的次序被包括。另外,可以有不出现的符号串,它可以形成要用信号通知的空组。通过这样的方法,编码器和译码器都知道符号串的次序。因此,例如,如果译码器接收到EOG,它就能够确定下一个适当的符号串。

[0064] 另一种方法可以用例如符号串明确作为数据集的前缀。同样地,符号串替换地可以作为后缀或否则被嵌入,这样,译码器可以根据编码的数据进行适当的确定。同样有可能,系统可以采用两种模式,一种模式中符号串的预定的次序被传送,另一种模式中符号串作为前缀或被嵌入。这些模式可以以任何次序出现,并且在一个例子中,可以通过用于该目的的专门的符号串来指示模式切换。

[0065] 在再一个可能的实施例中,编码器和译码器可以采用从已接收的符号串构建新的符号串的方法。这种灵活度可允许编码器选择或改变符号串和通信的模式,以改进压缩。例如,可能是这样的情形,其中不是所有的预定的符号串都被使用,以避免通过重复EOG符号而用信号通知大量空组,例如可以有新的符号串,其用信号通知不再有组,这实际上意味着数据集的结尾或数据的结尾。

[0066] 在许多情形下,辅助信息可以伴随数据组。例如,在文本的情形下,字体、大小、深浅、颜色和/或样式可包括这样的辅助信息。这可以以许多方式被传送或存储。例如,辅助信息可以嵌入行程长度,或被插入到编码的数据的任何位置,这样,译码器可以适当地把它与符号串相关联。在一个实施例中,可能希望处理辅助信息使之与符号串组合作为符号串本身,因而形成附加组。例如,作为一个简单的例子,斜体‘e’可以形成与正常的‘e’分开的符号串。同样地,在一个实施例中,如果期望的话,可利用专门的符号来在不同的嵌入样式或方法之间切换。

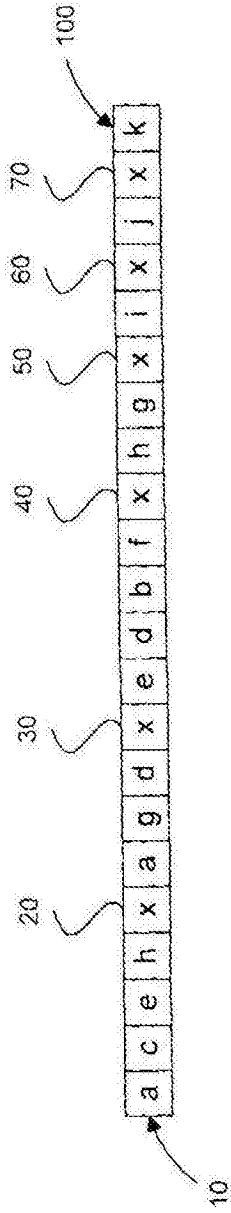
[0067] 根据所请求保护的主体实施例可被应用到所有类型的数据的编码,包括非数字数据,诸如例如在进行编码之前通过任何方便的映射而被转换成数字形式的符号数据。正如指出的,实施例对于行程长度编码执行得很好,但是当然要理解所请求保护的主体不限于这种应用。意图将所请求保护的主体实施例应用到多种不同类型的数据编码中的任一项。所以,所请求保护的主体将不在可用于的数据类型方面被限制。

[0068] 当然,将会看到,虽然已描述了特定的实施例,但所请求保护的主体在范围上不限于特定的实施例或实施方案。例如,一个实施例可以是硬件形式,诸如被实现为在设备上或在多个设备的组合上运行,而另一个实施例可以是软件形式。同样地,实施例可以以固件实现,或例如作为硬件、软件和/或固件的组合而实现。同样地,虽然所请求保护的主体在范围

上不限于这方面,但一个实施例可包括一个或多个物品,诸如存储介质或存储媒体。该存储介质——诸如一个或多个CD-ROM和/或盘——可以在其上存储指令,它如果被诸如计算机系统、计算平台或其它系统那样的系统执行,可以导致根据所请求保护的主题的方法的实施例——诸如之前描述的实施例之一——被执行。作为一个可能的例子,计算平台可包括一个或多个处理单元或处理器、一个或多个诸如显示器、键盘和/或鼠标之类的输入/输出设备,和/或一个或多个存储器,诸如静态随机存取存储器、动态随机存取存储器、快闪存储器和/或硬盘驱动器。

[0069] 图2示出系统200。系统200包括连接到数据集204的分析器202。分析器202从数据集204接收符号串206,并由此产生符号串码208。如上所述的,分析器202可以是硬件、软件和/或固件的组合。

[0070] 在前述的说明中,描述了所请求保护的主题的各方面。出于解释的目的,特定的数目、系统和/或配置被阐述以提供对所请求保护的主题的透彻了解。然而,从本公开内容获益的本领域技术人员应当看到,所请求保护的主题可以无需特定的细节而被实践。在其它实例中,公知的特征被省略和/或被简化,以免使所请求保护的主题模糊。虽然这里已经说明和/或描述了某些特征,现在对本领域技术人员而言可以有許多修改、替换、改变和/或等价物。所以,应当理解,所附权利要求意图覆盖在所请求保护的主题的实质精神内的所有的这样的修改和/或改变。



110

图1

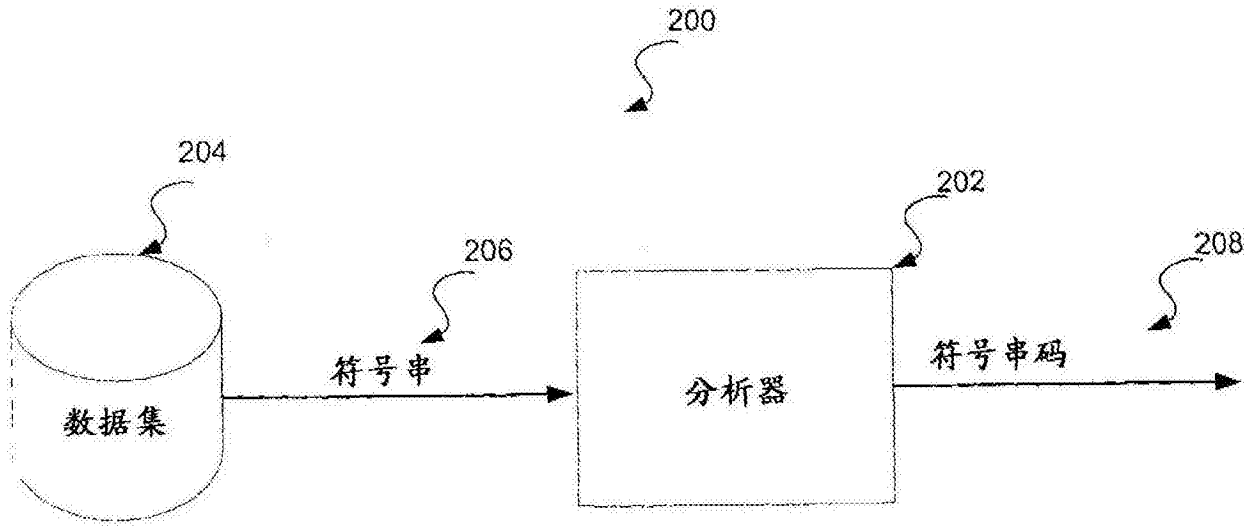


图2