



(12)发明专利

(10)授权公告号 CN 106980665 B

(45)授权公告日 2020.06.12

(21)申请号 201710170178.4

G06F 16/21(2019.01)

(22)申请日 2017.03.21

(56)对比文件

(65)同一申请的已公布的文献号

申请公布号 CN 106980665 A

CN 103795490 A, 2014.05.14, 全文.

CN 1585553 A, 2005.02.23, 全文.

US 2017024435 A1, 2017.01.26, 全文.

US 8005794 B2, 2011.08.23, 全文.

(43)申请公布日 2017.07.25

(73)专利权人 广东神马搜索科技有限公司

地址 510627 广东省广州市天河区黄埔大

道西平云路163号广电平云广场B塔13

层自编01单元

审查员 王艳坤

(72)发明人 卢申朋 褚建辉 刘东辉 王新栋

(74)专利代理机构 北京展翼知识产权代理事务

所(特殊普通合伙) 11452

代理人 屠长存

(51)Int. Cl.

G06F 16/22(2019.01)

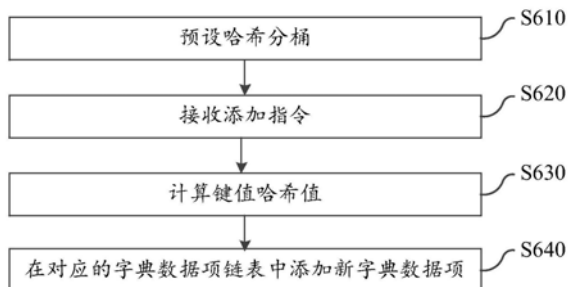
权利要求书3页 说明书11页 附图6页

(54)发明名称

数据字典实现方法、装置及数据字典管理系统

(57)摘要

本发明公开了一种数据字典实现方法、装置及数据字典管理系统,包括:在数据字典文件的指定位置处预留多个哈希分桶信息字段;响应于添加新字典数据项的指令,使用哈希算法计算其键值哈希值;以及在键值哈希值所属的哈希分桶对应的字典数据项链表中添加新字典数据项。由此,提供一种统一的字典,提升查询性能,同时方便后期扩展与兼容,减少周边服务以及管理平台的复杂性。



1. 一种数据字典实现方法,包括:

在数据字典文件的指定位置处预留多个哈希分桶信息字段,其中,哈希算法的值域被映射到所述多个哈希分桶,所述值域中的每个哈希值属于一个哈希分桶,每个哈希分桶包括所述值域中的至少一个哈希值,所述哈希分桶信息字段用于记录对应于所述哈希分桶的字典数据项链表的首地址指针,所述字典数据项链表包括属于所述哈希分桶的至少一个键值哈希值所对应的至少一个字典数据项,所述字典数据项包括键值哈希值、对应的字典值以及地址指针字段,其中地址指针字段用于记录属于同一个字典数据项链表的下一个字典数据项的地址;

响应于添加新字典数据项的指令,使用所述哈希算法计算其键值哈希值;

在所述数据字典文件中写入所述新字典数据项;以及

在原来位于所述字典数据项链表尾部的字典数据项的地址指针字段中写入所述新字典数据项的写入地址,以在所述键值哈希值所属的哈希分桶对应的字典数据项链表中添加所述新字典数据项。

2. 根据权利要求1所述的数据字典实现方法,其中,在所述键值哈希值所属的哈希分桶的哈希分桶信息字段为空的情况下,在所述键值哈希值所属的哈希分桶对应的字典数据项链表中添加所述新字典数据项的步骤包括:

在所述数据字典文件中写入所述新字典数据项;以及

在所述哈希分桶信息字段中写入所述新字典数据项的写入地址。

3. 根据权利要求1或2所述的数据字典实现方法,还包括:

在所述数据字典文件的头部记录空闲空间起始地址指针,其中,所述新字典数据项的写入地址为所述空闲空间起始地址指针所指向的地址;以及

响应于添加所述新字典数据项,更新所述数据字典文件的头部记录的空闲空间起始地址指针。

4. 根据权利要求1所述的数据字典实现方法,还包括:

在所述数据字典文件的头部记录所述数据字典文件大小、分桶数量和字典数据项个数中的至少一项;以及

在添加新字典数据项的过程中实时更新所述字典数据项个数。

5. 根据权利要求1所述的数据字典实现方法,其中,在所述字典数据项链表中已经有包含所述键值哈希值的旧字典数据项的情况下,该方法还包括:

在所述字典数据项链表中,将所述旧字典数据项的前一个字典数据项的地址指针修改为指向所述旧字典数据项的后一个字典数据项的地址。

6. 根据权利要求1所述的数据字典实现方法,其中,所述数据字典文件存储在外部存储器,该方法还包括:

使用mmap方式将所述数据字典文件从外部存储器映射到内部存储器中。

7. 根据权利要求1所述的数据字典实现方法,还包括:

响应于包含查询键值的查询请求,使用所述哈希算法计算查询键值哈希值;

从所述查询键值哈希值所属的哈希分桶的哈希分桶信息字段读取对应的字典数据项链表的首地址;

依次逐项比对所述查询键值哈希值和所述字典数据项链表中各字典数据项的键值哈

希值;以及

在找到与所述查询键值哈希值相同的键值哈希值的情况下,返回对应字典数据项中的字典值。

8. 根据权利要求1所述的数据字典实现方法,其中,所述字典值包括protocolbuffer序列化之后的二进制长度和相应字典数据。

9. 一种数据字典管理系统,包括:

外部存储器,用于存储数据字典文件;以及

处理器,用于执行根据权利要求1-8中任何一项所述的方法,以管理所述数据字典文件。

10. 根据权利要求9所述的数据字典管理系统,还包括:

内部存储器,所述外部存储器上的所述数据字典文件被通过mmap方式映射到所述内部存储器上。

11. 一种数据字典实现装置,包括:

分桶预设装置,用于在数据字典文件的指定位置处预留多个哈希分桶信息字段,其中,哈希算法的值域被映射到所述多个哈希分桶,所述值域中的每个哈希值属于一个哈希分桶,每个哈希分桶包括所述值域中的至少一个哈希值,所述哈希分桶信息字段用于记录对应于所述哈希分桶的字典数据项链表的首地址指针,所述字典数据项链表包括属于所述哈希分桶的至少一个键值哈希值所对应的至少一个字典数据项,所述字典数据项包括键值哈希值、对应的字典值以及地址指针字段,其中,地址指针字段用于记录属于同一个字典数据项链表的下一个字典数据项的地址;

哈希计算装置,用于响应于添加新字典数据项的指令,使用所述哈希算法计算其键值哈希值;

数据写入装置,用于在所述数据字典文件中写入所述新字典数据项;以及

地址写入装置,用于在原来位于所述字典数据项链表尾部的字典数据项的地址指针字段中写入所述新字典数据项的写入地址,以在所述键值哈希值所属的哈希分桶对应的字典数据项链表中添加所述新字典数据项。

12. 根据权利要求11所述的数据字典实现装置,其中,

在所述键值哈希值所属的哈希分桶的哈希分桶信息字段为空的情况下,所述地址写入装置在所述哈希分桶信息字段中写入所述新字典数据项的写入地址。

13. 根据权利要求11或12所述的数据字典实现装置,还包括:

空闲地址记录装置,用于在所述数据字典文件的头部记录空闲空间起始地址指针,其中,所述新字典数据项的写入地址为所述空闲空间起始地址指针所指向的地址;以及

空闲地址更新装置,用于响应于添加所述新字典数据项,更新所述数据字典文件的头部记录的空闲空间起始地址指针。

14. 根据权利要求11所述的数据字典实现装置,其中,该装置还包括:

链表修改装置,用于在所述字典数据项链表中已经有包含所述键值哈希值的旧字典数据项的情况下,在所述字典数据项链表中,将所述旧字典数据项的前一个字典数据项的地址指针修改为指向所述旧字典数据项的后一个字典数据项的地址。

15. 根据权利要求11所述的数据字典实现装置,还包括:

查询哈希计算装置,用于响应于包含查询键值的查询请求,使用所述哈希算法计算查询键值哈希值;

分桶查找装置,用于从所述查询键值哈希值所属的哈希分桶的哈希分桶信息字段读取对应的字典数据项链表的首地址;

哈希值比对装置,用于依次逐项比对所述查询键值哈希值和所述字典数据项链表中各字典数据项的键值哈希值;以及

字典值返回装置,用于在找到与所述查询键值哈希值相同的键值哈希值的情况下,返回对应字典数据项中的字典值。

数据字典实现方法、装置及数据字典管理系统

技术领域

[0001] 本发明涉及数据库管理领域,特别涉及一种数据字典实现方法、装置及数据字典管理系统。

背景技术

[0002] 数据字典(Data dictionary),通常指对数据的数据项、数据结构、数据流、数据存储等进行定义和描述的集合,是对数据模型中的数据对象或项目的描述的集合。往往针对一种数据维护一个数据字典,基于数据类型、尺寸等,相应地设定数据存储结构。例如,用于存储电话号码的数据字典和用于存储实体信息的数据字典会分别采用相应设定的不同数据存储结构。

[0003] 为了加快内部系统快速开发,解决数据类型冗杂的局势,迫切需要一种统一的数据字典,通过对统一的数据字典进行管理来方便字典的使用,提升系统的开发效率。

发明内容

[0004] 本发明的目的是提供一种数据字典实现方法、装置及数据字典管理系统,以统一字典,有利于方便构建数据字典,并对分发、校验进行统一的管理,降低管理平台的复杂性。

[0005] 根据本发明的一个方面,提供了一种数据字典实现方法,包括:在数据字典文件的指定位置处预留多个哈希分桶信息字段,其中,哈希算法的值域被映射到多个哈希分桶,值域中的每个哈希值属于一个哈希分桶,每个哈希分桶包括值域中的至少一个哈希值,哈希分桶信息字段用于记录对应于哈希分桶的字典数据项链表的首地址指针,字典数据项链表包括属于哈希分桶的至少一个键值哈希值所对应的至少一个字典数据项,字典数据项包括键值哈希值和对应的字典值;响应于添加新字典数据项的指令,使用哈希算法计算其键值哈希值;以及在键值哈希值所属的哈希分桶对应的字典数据项链表中添加新字典数据项。

[0006] 由此,通过采用键值哈希方式,从多个哈希分桶中快速找到键值哈希值所属的哈希分桶,并在哈希分桶对应的字典数据项链表中添加新字典数据项,以实现任意类型的数据在数据字典的快速存储。另外,由于键值哈希值采用二进制数据,而二进制数据比字符串的效率很高很多,也能相应地提升添加速度和查找速度。

[0007] 优选地,字典数据项还包括地址指针字段,用于记录属于同一个字典数据项链表的下一个字典数据项的地址,在键值哈希值所属的哈希分桶对应的字典数据项链表中添加新字典数据项的步骤包括:在数据字典文件中写入新字典数据项;以及在原来位于字典数据项链表尾部的字典数据项的地址指针字段中写入新字典数据项的写入地址。

[0008] 由此,字典数据链表的首地址指针不为空,即字典数据项链表中已存在至少一个数据项时,在链表的尾部添加新字典数据项,形成数据项单向链表,使得链表在逻辑上是连续的,充分利用链表的可用空间,方便地进行数据存储。

[0009] 优选地,在键值哈希值所属的哈希分桶的哈希分桶信息字段为空的情况下,在键值哈希值所属的哈希分桶对应的字典数据项链表中添加新字典数据项的步骤包括:在数据

字典文件中写入新字典数据项;以及在哈希分桶信息字段中写入新字典数据项的写入地址。

[0010] 由此,字典数据链表的首地址指针为空,即字典数据项链表中尚无字典数据项时,则在键值哈希值所述的哈希分桶的哈希分桶信息字段(链表首地址指针)写入新字典数据项的写入地址,以使新字典数据项成为该字典数据项链表的首项。

[0011] 优选地,该数据字典实现方法还可以包括:在数据字典文件的头部记录空闲空间起始地址指针,其中,新字典数据项的写入地址为空闲空间起始地址指针所指向的地址;以及响应于添加新字典数据项,更新数据字典文件的头部记录的空闲空间起始地址指针。

[0012] 由此,通过空闲空间起始地址指针,能够快速查找到字典文件的空闲地址,相应地在该空闲地址添加新字典数据项,实现数据物理上的连续存储,充分利用字典文件的空闲空间。

[0013] 优选地,该数据字典实现方法还可以包括:在数据字典文件的头部记录数据字典文件大小、分桶数量和字典数据项个数中的至少一项;以及在添加新字典数据项的过程中实时更新字典数据项个数。由此,实现数据的快速加载。

[0014] 优选地,在字典数据项链表中已经有包含键值哈希值的旧字典数据项的情况下,该方法还包括:在字典数据项链表中,将旧字典数据项的前一个字典数据项的地址指针修改为指向旧字典数据项的后一个字典数据项的地址。由此,避免哈希冲突。

[0015] 优选地,数据字典文件存储在外部存储器,该方法还包括:使用mmap方式将数据字典文件从外部存储器映射到内部存储器中。由此,减少用户态和内核态之间的拷贝,提升加载速度。

[0016] 优选地,该数据字典实现方法还可以包括:响应于包含查询键值的查询请求,使用哈希算法计算查询键值哈希值;从查询键值哈希值所属的哈希分桶的哈希分桶信息字段读取对应的字典数据项链表的首地址;依次逐项比对查询键值哈希值和字典数据项链表中各字典数据项的键值哈希值;以及在找到与查询键值哈希值相同的键值哈希值的情况下,返回对应字典数据项中的字典值。由此,加快查询速度,提升查询性能。

[0017] 优选地,字典值包括protocol buffer序列化之后的二进制长度和相应字典数据。字典值采用结构化定义,可以支持多值方式。字典值采用protocol buffer方式,可以方便字典值的扩展,减少对服务的影响。

[0018] 根据本发明的另一方面,提供了一种数据字典管理系统,包括:外部存储器,用于存储数据字典文件;以及处理器,用于执行根据权利要求1-9中任何一项的方法,以管理数据字典文件。

[0019] 优选地,该数据字典管理系统还可以包括:内部存储器,外部存储器上的字典文件被通过mmap方式映射到内部存储器上。

[0020] 根据本发明的又一方面,还提供了一种数据字典实现装置,包括:分桶预设装置,用于在数据字典文件的指定位置处预留多个哈希分桶信息字段,其中,哈希算法的值域被映射到多个哈希分桶,值域中的每个哈希值属于一个哈希分桶,每个哈希分桶包括值域中的至少一个哈希值,哈希分桶信息字段用于记录对应于哈希分桶的字典数据项链表的首地址指针,字典数据项链表包括属于哈希分桶的至少一个键值哈希值所对应的一个和多个字典数据项,字典数据项包括键值哈希值和对应的字典值;哈希计算装置,用于响应于添加新

字典数据项的指令,使用哈希算法计算其键值哈希值;以及数据添加装置,用于在键值哈希值所属的哈希分桶对应的字典数据项链表中添加新字典数据项。

[0021] 优选地,字典数据项还包括地址指针字段,用于记录属于同一个字典数据项链表的下一个字典数据项的地址,数据添加装置包括:数据写入装置,用于在数据字典文件中写入新字典数据项;以及地址写入装置,用于在原来位于字典数据项链表尾部的字典数据项的地址指针字段中写入新字典数据项的写入地址。

[0022] 优选地,在键值哈希值所属的哈希分桶的哈希分桶信息字段为空的情况下,地址写入装置在哈希分桶信息字段中写入新字典数据项的写入地址。

[0023] 优选地,该数据字典实现装置还可以包括:空闲地址记录装置,用于在数据字典文件的头部记录空闲空间起始地址指针,其中,新字典数据项的写入地址为空闲空间起始地址指针所指向的地址;以及空闲地址更新装置,用于响应于添加新字典数据项,更新数据字典文件的头部记录的空闲空间起始地址指针。

[0024] 优选地,该装置还可以包括:链表修改装置,用于在字典数据项链表中已经有包含键值哈希值的旧字典数据项的情况下,在字典数据项链表中,将旧字典数据项的前一个字典数据项的地址指针修改为指向旧字典数据项的后一个字典数据项的地址。

[0025] 优选地,数据字典实现装置还可以包括:查询哈希计算装置,用于响应于包含查询键值的查询请求,使用哈希算法计算查询键值哈希值;分桶查找装置,用于从查询键值哈希值所属的哈希分桶的哈希分桶信息字段读取对应的字典数据项链表的首地址;哈希值比对装置,用于依次逐项比对查询键值哈希值和字典数据项链表中各字典数据项的键值哈希值;以及字典值返回装置,用于在找到与查询键值哈希值相同的键值哈希值的情况下,返回对应字典数据项中的字典值。

[0026] 通过本发明的数据字典实现方法、装置及数据字典管理系统,能够提供统一的字典,有利于构建数据字典,并对分发、校验进行统一的管理,降低管理平台的复杂性。

附图说明

[0027] 通过结合附图对本公开示例性实施方式进行更详细的描述,本公开的上述以及其它目的、特征和优势将变得更加明显,其中,在本公开示例性实施方式中,相同的参考标号通常代表相同部件。

[0028] 图1示出了根据本发明的数据字典文件概图示意图。

[0029] 图2示出了根据本发明的字典头部概图示意图。

[0030] 图3示出了根据本发明的哈希分桶部分信息概图示意图。

[0031] 图4示出了根据本发明的字典内容部分概图示意图。

[0032] 图5示出了根据本发明一实施例的数据字典管理系统的示意性框图。

[0033] 图6示出了根据本发明一实施例的数据字典实现方法的流程图。

[0034] 图7示出了根据本发明一实施例的添加新字典数据项的方法的流程图。

[0035] 图8示出了哈希分桶信息字段为空时添加新字典数据项示意图。

[0036] 图9示出了哈希分桶信息字段不为空时添加新字典数据项示意图。

[0037] 图10示出了根据本发明另一实施例的数据字典实现方法的流程图。

[0038] 图11示出了根据本发明又一实施例数据字典实现方法的流程图。

- [0039] 图12示出了字典数据项链表中包含旧字典数据项时添加新字典数据项的示意图。
- [0040] 图13示出了根据本发明一实施例的查询键值方法的流程图。
- [0041] 图14示出了根据本发明一实施例的数据字典实现装置的示意性框图。
- [0042] 图15示出了根据本发明一实施例的数据添加装置的示意性框图。
- [0043] 图16示出了根据本发明另一实施例的数据字典实现装置的示意性框图。
- [0044] 图17示出了根据本发明又一实施例的数据字典实现装置的示意性框图。

具体实施方式

[0045] 下面将参照附图更详细地描述本公开的优选实施方式。虽然附图中显示了本公开的优选实施方式,然而应该理解,可以以各种形式实现本公开而不应被这里阐述的实施方式所限制。相反,提供这些实施方式是为了使本公开更加透彻和完整,并且能够将本公开的范围完整地传达给本领域的技术人员。

[0046] 为了更好地理解本发明,在描述本发明之前,首先对文中出现的名词的含义进行解释。

[0047] 数据字典(Data dictionary)以数据库中的数据项(数据元素,数据的基本单元,是数据不可分割的最小单位)为单位,按照一定顺序排列,对其内容作详细说明的数据集。数据字典中存放着系统中所有数据的定义,即对所有数据库结构的描述,是一个定义数据库中使用的数据库中的所有数据元素和结构的含义、类型、数据大小、格式、度量单位、精度以及允许聚值范围的共享仓库,数据字典在表现形式上为数据库中的表。

[0048] 在数据库中,大多数时,表的“列”称为“字段”,每个字段包含某一专题的信息。例如,“通讯录”数据库中,“姓名”、“联系电话”这些都是表中所有行共有的属性,所以把这些列称为“姓名”字段和“联系电话”字段。

[0049] 键-值对(key-value pair)是编程语言对数学概念中映射的实现,键(key)用作元素的索引,值(value)则表示所存储和读取的数据。本公开中以键值表示键(key),以字典值表示值(value)。

[0050] 哈希算法(Hash),即散列函数,是把任意长度的输入通过散列算法,变换成固定长度的输出。哈希表(hash table)是从一个集合A到另一个集合B的映射(mapping),是一种对应关系,这样的对应关系在现实生活中很常见,例如,A→B;人→身份证号;日期→星座等。在哈希表中,上述对应过程称为hashing。集合A中元素a对应集合B中元素b,b被称为a的哈希值(hash value)。

[0051] 本公开中,哈希桶是盛放不同键-值对的容器(即是哈希表)。键值采用哈希方式,一方面可以加快添加速度和查找速度,另一方面,采用二进制数据的键值哈希值比字符串的效率高很多,也能够提升添加和查找速度。

[0052] mmap方式的功能是将一个文件或者其它对象映射进内存(内部存储器)。文件被映射到多个页上,如果文件的大小不是所有页的大小之和,最后一个页不被使用的空间将会清零,mmap在用户空间映射调用系统中作用很大。

[0053] 本公开中,数据字典采用mmap方式进行加载存储,可以减少系统的调用,对系统负载减少了影响,并且通过慢加载机制,更有利于服务的稳定。同时,由于mmap方式常驻内部存储器,当应用异常重启的时候,不需要重新加载字典,节省恢复时间。

[0054] Protocol Buffers是一种轻便高效的结构化数据存储格式,可以用于结构化数据串行化,或者说序列化。它很适合做数据存储或RPC数据交换格式。可用于通讯协议、数据存储等领域的语言无关、平台无关、可扩展的序列化结构数据格式。

[0055] 本文中,字典值是结构化的protocol buffer定义(字段支持常见数据类型定义),其保存的是protocol buffer序列化之后的二进制长度和对应数据,专注于字典存储,查询接口与扩展性。字典元数据是protocol buffer文件中对数据值的定义,专注于数据校验等附属信息。采用protocol buffer可以将字典值与字典元数据相分离,使数据字典使用起来更为便利,方便后期扩展与兼容。

[0056] 由于哈希(hash)只要求从A到B的对应为一个映射,并没有限定该对应关系为一一映射。因此会有这样的可能:两个或多个不同的键值对应同一个哈希值。这种情况叫做哈希碰撞(hash collision),即键值哈希发生冲突。本文中采用六十四位哈希方法,对键值进行哈希,并假设键值哈希之后是不冲突的。

[0057] 如前所述,为了加快内部系统快速开发,解决数据类型冗杂的局势,迫切需要一种统一的数据字典。因此,本发明提出了一种数据字典管理系统,以通过构建一种数据字典,对分发、校验进行统一管理,降低管理平台的复杂性。下面结合附图及实施例详细说明本发明的技术方案。

[0058] 字典文件的数据结构

[0059] 首先结合附图1-4介绍本发明的数据字典文件的数据结构。图1示出了根据本发明的字典文件概图示意图。图2示出了根据本发明的字典头部概图示意图。图3示出了根据本发明的哈希分桶信息概图示意图。图4示出了根据本发明的字典内容部分概图示意图。

[0060] 如图1所示,该数据字典的字典文件包括字典头部、哈希分桶信息和字典内容部分。

[0061] 字典头部为固定长度,可以记录数据字典文件大小、分桶数量和字典数据项个数中的至少一项。优选地,在添加新字典数据项的过程中还可以实时更新字典数据项的个数。

[0062] 在图2所示的优选实施例中,字典头部可以包含三十二字节,前八个字节表示文件大小,接下来的八个字节表示分桶数量,之后的八个字节表示空闲空间指针地址,最后八个字节直接记录字典数据项个数。其中,以unit64_t表示八个字节。

[0063] 在数据字典文件的指定位置处(例如,哈希分桶信息部分)可以预留多个哈希分桶信息字段,哈希分桶信息部分可以包含分桶数量个地址指针,每个地址指针八个字节,如图3所示。其中,哈希分桶信息字段可以用于记录对应于哈希分桶的字典数据项链表的首地址指针。

[0064] 由此,将哈希算法的值域映射到多个哈希分桶,值域中的每个哈希值属于一个哈希分桶,便于快速添加和查找。

[0065] 字典内容部分可以包含已经分配的字典数据项链表和空闲空间。

[0066] 在图4所示的优选实施例中,字典数据项可以包含三部分,即八个字节的地址指针、八个字节的键值哈希值和实际长度的字典值(包含protocol buffer序列化之后二进制长度和对应数据)。并且,地址指针保存指向同一个哈希分桶的下一个数据项的地址,形成数据项单向链表。其中,链表中的前后关系是根据地址指针的指向关系确定的。

[0067] 字典值采用protocol buffer进行保存,可以包含多个字段,其长度是不固定的,

方便字典值字段扩展,减少对现有服务的影响。

[0068] 值得说明的是,本文中的各个字段,除了字典数据项中的具体字典值,其它都具有固定的长度,例如,大多可以为八个字节,或其它长度。并且其优选顺序可以与图1-4中所示的顺序相同。

[0069] 为了使用简单,本发明的字典数据可以提供一次构建,只读查询的方式,不提供数据的更新与删除操作。

[0070] 添加接口

[0071] 下面将结合附图及实施例详细描述本发明的数据字典实现方案。

[0072] 图5示出了根据本发明一实施例的数据字典管理系统的示意性框图。

[0073] 如图5所示,本发明的数据字典管理系统500可以包括外部存储器510和处理器520。

[0074] 外部存储器510用于存储数据字典文件。

[0075] 这里,外部存储器510是指除计算机内存及CPU缓存以外的存储器,断电后仍能保存数据,包括硬盘、软盘、光盘、U盘、磁带等。

[0076] 处理器520可以执行数据字典实现方法,以管理数据字典文件。其详细执行流程及步骤可参见下文中的数据字典实现方法。

[0077] 数据字典管理系统500还可以包括内部存储器530。其中,外部存储器510上的字典文件可以被通过mmap方式映射到内部存储器530上,减少用户态和内核态之间的拷贝,提升加载速度。并且,mmap方式字典常驻内部存储器,当应用异常重启的时候,不需要重新加载字典,节省恢复时间。

[0078] 处理器520可以用以执行根据本发明的数据字典实现方法。下面将参考图6-12对根据本发明的数据字典实现方法进行详细描述。

[0079] 图6示出了根据本发明一实施例的数据字典实现方法的流程图。

[0080] 如图6所示,在步骤S610预设哈希分桶。

[0081] 具体地说,在数据字典文件的指定位置(例如哈希分桶信息部分)处预留多个哈希分桶信息字段。其中,哈希算法的值域被映射到多个哈希分桶,值域中的每个哈希值属于一个哈希分桶,每个哈希分桶包括值域中的至少一个哈希值。

[0082] 哈希分桶信息字段可以用于记录对应于哈希分桶的字典数据项链表的首地址指针,字典数据项链表包括属于哈希分桶的至少一个键值哈希值所对应的至少一个字典数据项,字典数据项包括键值哈希值和对应的字典值。

[0083] 在步骤S620中,接收添加新字典数据项的指令。

[0084] 在步骤S630中,响应于添加新字典数据项的指令,使用哈希算法计算其键值哈希值。

[0085] 数据项包括键值和对应的字典值,其可以是输入的任意数据,可以是字母、数值或者两者的结合。采用哈希算法对键值进行哈希,得到固定的哈希结果,即键值哈希值。新字典数据项包括键值哈希值和实际长度的字典值。

[0086] 在步骤S640中,在键值哈希值所属的哈希分桶对应的字典数据项链表中添加新字典数据项。

[0087] 这里,键值采用哈希方式,通过计算得到的键值哈希值,快速从多个哈希分桶中查

找键值哈希值所属的分桶,并在该分桶对应的字典数据项链表中添加新字典数据项。由此,方便快速查找,提升查询性能。

[0088] 此外,字典数据项还可以包括地址指针字段,用于记录属于同一个字典数据项链表的下一个字典数据项的地址。本公开中,可以通过判断键值哈希值所属的哈希分桶的字典数据项链表中是否包含字典数据项,从而确定新字典数据项的添加(写入)接口。

[0089] 图7示出了根据本发明一实施例的添加新字典数据项步骤的方法流程图。

[0090] 如图7所示,在步骤S710中,在数据字典文件中写入新字典数据项。

[0091] 这里,新字典数据项的写入位置可以是数据字典文件中的任意空闲地址,例如,可以是键值哈希值所属的字典数据项链表的尾部,也可以是数据字典文件中的其它空闲空间。

[0092] 在步骤S720中,判断键值哈希值所属的哈希分桶的哈希分桶信息字段是否为空,在判断结果为空的情况下,进入步骤S730;在判断结果不为空的情况下,进入步骤S740。

[0093] 判断结果为空,即字典数据项链表尚无字典数据项,因此,在步骤S730中,在哈希分桶信息字段写入新字典数据项的写入地址,从而该新字典数据项作为对应的字典数据项链表的首项。其中,哈希分桶信息字段用于记录对应于哈希分桶的字典数据项链表的首地址指针。

[0094] 图8示出了哈希分桶信息字段为空时添加新字典数据项示意图,如图8所示,分桶2的值为空,即分桶2对应的字典数据项链表为空,则在数据字典文件中添加新字典数据项,并在哈希分桶信息字段写入新字典数据项的写入地址,从而该新字典数据项作为对应字典数据项链表的首项。

[0095] 判断结果不为空,即键值哈希值所属的哈希分桶的字典数据项链表中已包含至少一个字典数据项,因此,在步骤S740中,可以在原来位于字典数据项链表尾部的字典数据项的地址指针字段中写入新字典数据项的写入地址,从而将新字典数据项添加到字典数据项链表的尾部。

[0096] 图9示出了哈希分桶信息字段不为空时添加新字典数据项示意图,如图9所示,分桶2的值不为空,对应的字典数据项链表中已包含至少一个字典数据项,并且链表中不包含该键值哈希值对应的字典数据项(哈希不冲突),则直接在字典数据项链表的尾部添加新字典数据项(图9中链表尾部的阴影部分),并将新字典数据项的写入地址写入原字典数据项链表尾部的字典数据项的地址指针字段中,使得原来位于链表尾部的字典数据项的地址指针指向新字典数据项的地址,以形成新的数据项单向链表。

[0097] 由此,将需要添加的新字典数据项添加到对应的接口。应当理解,上述步骤,尤其是步骤S710与步骤S720、S730、S740的顺序是非固定的,在具体实现过程中,也可以先执行步骤S720的判断步骤(以及步骤S730和步骤S740的地址写入步骤),再执行步骤S710的写入步骤,对此不做限制。

[0098] 在上述添加新字典数据项的方法步骤中,新字典数据项的添加接口可以是数据字典文件的空闲空间。鉴于数据字典文件中空闲空间繁多,为了不遗漏并充分利用空闲空间,解决冗余问题,可以在数据字典文件的头部记录空闲空间起始地址指针,并在空闲空间地址指针指向的地址中写入新字典数据项。

[0099] 图10示出了根据本发明另一实施例的数据字典实现方法的流程图。

- [0100] 如图10所示,在步骤S1010中,在数据字典文件的头部记录空闲空间起始地址指针。
- [0101] 在步骤S1020预设哈希分桶,在数据字典文件的指定位置处预留多个哈希分桶信息字段。
- [0102] 在步骤S1030中,接收添加新字典数据项的指令。
- [0103] 在步骤S1040中,响应于添加新字典数据项的指令,使用哈希算法计算键值哈希值。
- [0104] 在步骤S1050中,在空闲空间起始地址指针指向的地址写入新字典数据项。其中,新字典数据项包括键值哈希值和对应的字典值。
- [0105] 在步骤S1060中,响应于添加新字典数据项,更新数据字典文件的头部记录的空闲空间起始地址指针。优选地,可以更新为与写入新字典数据项的空间在物理上连续的下一个空闲空间的地址。
- [0106] 需要说明的是,上述步骤,特别是步骤S1010与步骤S1020可以是不固定的,即,也可以先预设哈希分桶,然后在数据字典文件的头部记录空闲空间起始地址指针。
- [0107] 由此,根据空闲空间起始地址指针顺序地在字典文件中添加新字典数据项,数据字典文件虽然在逻辑上被分为多个链表,但是其在物理上是连续的,从而能够充分利用哈希分桶的空闲空间,更好地发挥mmap的作用。
- [0108] 如前所述,键值哈希之后也可能存在冲突(哈希碰撞)问题,即,在链表中已经包含键值哈希值对应的字典数据项,因此,每次对键值进行哈希之后,还需要检测字典数据项链表中是否已包含该键值哈希值的旧字典数据项,以避免冲突。
- [0109] 图11示出了根据本发明又一实施例数据字典实现方法的流程图。
- [0110] 如图11所示,在步骤S1110中,与前文相同的方法执行预设哈希分桶步骤。
- [0111] 在步骤S1120中,接收添加新字典数据项的指令。
- [0112] 在步骤S1130中,响应于添加新字典数据项的指令,使用哈希算法计算键值哈希值。
- [0113] 在步骤S1140中,例如可以依次逐项对比计算的键值哈希值和字典数据项链表中各字典数据项的键值哈希值,判断计算的键值哈希值所属的哈希分桶的字典数据项链表中是否已包含该键值哈希值的旧字典数据项,若判断结果为是,则进入步骤S1150,若判断结果为否,则进入步骤S1160。
- [0114] 在步骤S1150中,在字典数据项链表中已经包含键值哈希值的旧字典数据项的情况下,在字典数据项链表中,将旧字典数据项的前一个字典数据项的地址指针修改为指向旧字典数据项的后一个字典数据项的地址,同时,在键值哈希值所属的哈希分桶对应的字典数据项链表的尾部写入新字典数据项。其中,修改地址指针可以视为将旧字典数据项从其所在的字典数据项链表中删除。
- [0115] 在步骤S1160中,在字典数据项链表中不包含旧字典数据项的情况下,在键值哈希值所属的哈希分桶对应的字典数据项链表的尾部写入新字典数据项。
- [0116] 并且,在步骤S1160写入新字典数据项时,也可以返回步骤S1140中,再次检测字典数据项链表中是否已包含键值哈希值对应的旧字典数据项。
- [0117] 应当理解,上述步骤,尤其是步骤S1150和步骤S1160,其顺序是非固定的,在具体

过程中也可以先执行步骤S1160的写入过程,再执行步骤S1150的修改地址步骤,对此不做限制。

[0118] 图12示出了字典数据项链表中包含旧字典数据项时添加新字典数据项的示意图,如图12所示,分桶2的首地址的值不为空,并且,链表中已包含键值哈希值对应的旧字典数据项(图中网格部分)。此时,在字典数据项链表的尾部添加新字典数据项(图中链表尾部的阴影部分),同时将旧字典数据项的前一个字典数据项的地址指针修改为指向旧字典数据项的后一个字典数据项的地址(图中删除部分),即,在字典数据项链表中删除旧字典数据项。由此,避免冲突。

[0119] 查询接口

[0120] 本发明的数据字典还可以用于查询键值。

[0121] 图13示出了根据本发明一实施例的查询键值方法的流程图。

[0122] 如图13所示,在步骤S1310中,接收包含查询键值的查询请求。

[0123] 在步骤S1320中,响应于包含查询键值的查询请求,使用哈希算法计算键值哈希值。

[0124] 在步骤S1330中,从查询键值哈希值所属的哈希分桶的哈希分桶信息字段读取对应的字典数据项链表的首地址。

[0125] 在步骤S1340中,依次逐项(单向)比对查询键值哈希值和字典数据项链表中各字典数据项的键值哈希值。

[0126] 在步骤S1350中,判断是否从字典数据项链表中找到与查询键值哈希值相同的键值哈希值,若是,则进入步骤S1360,若否,则进入步骤S1370。

[0127] 在步骤S1360中,在找到与查询键值哈希值相同的键值哈希值的情况下,返回对应字典数据项中的字典值,先提取出二进制长度,之后protocol buffer反序列化对应数据结构。

[0128] 在步骤S1370中,查找至字典数据项链表的尾部,在未找到与查询键值哈希值相同的键值哈希值的情况下,返回查询失败信息,即显示未查询到该键值。

[0129] 由此,通过采用哈希的方式,先从多个分桶中找到查询键值哈希值所属的哈希分桶,再从哈希分桶对应的字典数据项链表中查找与查询键值哈希值相同的键值哈希值,以方便快速查找,提升查询性能。

[0130] 此外,图5至图13所示的数据字典的实现方法也可以由一种数据字典实现装置实现,下面结合附图14-17进行详细描述。

[0131] 图14示出了根据本发明一实施例的数据字典实现装置的示意性框图。

[0132] 如图14所示,本发明的数据字典实现装置1400可以包括分桶预设装置1410、哈希计算装置1420和数据添加装置1430。

[0133] 分桶预设装置1410,用于在数据字典文件的指定位置处预留多个哈希分桶信息字段,其中,哈希算法的值域被映射到多个哈希分桶,值域中的每个哈希值属于一个哈希分桶,每个哈希分桶包括值域中的至少一个哈希值,哈希分桶信息字段用于记录对应于哈希分桶的字典数据项链表的首地址指针,字典数据项链表包括属于哈希分桶的至少一个键值哈希值所对应的一个和多个字典数据项,字典数据项包括键值哈希值和对应的字典值。

[0134] 哈希计算装置1420,用于响应于添加新字典数据项的指令,使用哈希算法计算其

键值哈希值。

[0135] 数据添加装置1430,用于在键值哈希值所属的哈希分桶对应的字典数据项链表中添加新字典数据项。

[0136] 数据字典实现装置1400所实现的数据字典可以参见图6的相关描述,在此不再赘述。

[0137] 此外,字典数据项还可以包括地址指针字段,用于记录属于同一个字典数据项链表的下一个字典数据项的地址,以便于在添加新字典数据项之后,能够及时在前一个字典数据项的地址指针中写入新字典数据项的写入地址,形成数据项单向链表。

[0138] 图15示出了根据本发明一实施例的数据添加装置的示意性框图。

[0139] 如图15所示,本发明的数据添加装置1530可以包括数据写入装置1531和地址写入装置1533。

[0140] 数据写入装置1531,用于在数据字典文件中写入新字典数据项。其中,可以根据键值哈希值所属的哈希分桶的哈希分桶信息字段是否为空,确定新字典数据项的添加接口,详细的执行步骤可以参见图7的相关描述。

[0141] 地址写入装置1533,用于在相应的地址指针处写入新字典数据项的写入地址。其中,在键值哈希值所属的哈希分桶的哈希分桶信息字段为空的情况下,将所添加的新字典数据项的写入地址写入哈希分桶信息字段(链表首地址);在键值哈希值所属的哈希分桶的哈希分桶信息字段不为空的情况下,将所添加的新字典数据项的写入地址写入原来位于字典数据项链表尾部的字典数据项的地址指针,详细执行步骤可以参见图7至图9的相关描述。

[0142] 图16示出了根据本发明另一实施例的数据字典实现装置的示意性框图。

[0143] 如图16所示,本发明的数据字典实现装置除包括图14所示的装置外,还可以包括空闲地址记录装置1440和空闲地址更新装置1450。

[0144] 空闲地址记录装置1440用于在数据字典文件的头部记录空闲空间起始地址指针,其中,新字典数据项的写入地址为空闲空间起始地址指针所指向的地址。

[0145] 空闲地址更新装置1450用于响应于添加新字典数据项,更新数据字典文件的头部记录的空闲空间起始地址指针。

[0146] 由此,实时更新字典文件头部记录的空闲空间起始地址指针,使得字典文件在物理上连续,以充分使用数据字典文件的空闲空间,提升字典在系统中的加载速度。

[0147] 此外,该数据字典实现装置还可以包括链表修改装置1460,用于在字典数据项链表中已经有包含键值哈希值的旧字典数据项的情况下,在字典数据项链表中,将旧字典数据项的前一个字典数据项的地址指针修改为指向旧字典数据项的后一个字典数据项的地址。由此,跳过旧字典数据项的对应节点,视为在链表中删除该旧字典数据项的节点,避免哈希冲突。

[0148] 图17示出了根据本发明又一实施例的数据字典实现装置的示意性框图。

[0149] 如图17所示,本发明的数据字典实现装置1400除包含图14和图16所示的装置外,还可以包括查询哈希计算装置1470、分桶查找装置1480、哈希值比对装置1490和字典值返回装置14100。

[0150] 查询哈希计算装置1470,用于响应于包含查询键值的查询请求,使用哈希算法计

算查询键值哈希值。

[0151] 分桶查找装置1480,用于从查询键值哈希值所属的哈希分桶的哈希分桶信息字段读取对应的字典数据项链表的首地址。

[0152] 哈希值比对装置1490,用于依次逐项比对查询键值哈希值和字典数据项链表中各字典数据项的键值哈希值。

[0153] 字典值返回装置14100,用于在找到与查询键值哈希值相同的键值哈希值的情况下,返回对应字典数据项中的字典值。

[0154] 由此,通过哈希方式,从多个分桶中找到查询键值哈希值所属的哈希分桶,再将查询键值哈希值逐项与该哈希分桶对应的字典数据项链表中的字典数据项的键值哈希值进行比对,直到找到与该查询键值哈希值相同的键值哈希值,否则返回查询失败信息。

[0155] 至此,已经结合附图详细说明了本发明的数据字典实现方法、装置及数据字典管理系统。通过本发明,能够方便地构建一种统一的数据字典,并对分发、校验进行统一的关联,降低管理平台的复杂性。

[0156] 此外,根据本发明的方法还可以实现为一种计算机程序,该计算机程序包括用于执行本发明的上述方法中限定的上述各步骤的计算机程序代码指令。或者,根据本发明的方法还可以实现为一种计算机程序产品,该计算机程序产品包括计算机可读介质,在该计算机可读介质上存储有用于执行本发明的上述方法中限定的上述功能的计算机程序。本领域技术人员还将明白的是,结合这里的公开所描述的各种示例性逻辑块、模块、电路和算法步骤可以被实现为电子硬件、计算机软件或两者的组合。

[0157] 附图中的流程图和框图显示了根据本发明的多个实施例的系统和方法的可能实现的体系架构、功能和操作。在这点上,流程图或框图中的每个方框可以代表一个模块、程序段或代码的一部分,所述模块、程序段或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。也应当注意,在有些作为替换的实现中,方框中所标记的功能也可以以不同于附图中所标记的顺序发生。例如,两个连续的方框实际上可以基本并行地执行,它们有时也可以按相反的顺序执行,这依所涉及的功能而定。也要注意,框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合,可以用执行规定的功能或操作的专用的基于硬件的系统来实现,或者可以用专用硬件与计算机指令的组合来实现。

[0158] 以上已经描述了本发明的各实施例,上述说明是示例性的,并非穷尽性的,并且也不限于所披露的各实施例。在不偏离所说明的各实施例的范围和精神的情况下,对于本技术领域的普通技术人员来说许多修改和变更都是显而易见的。本文中所用术语的选择,旨在最好地解释各实施例的原理、实际应用或对市场中的技术的改进,或者使本技术领域的其它普通技术人员能理解本文披露的各实施例。

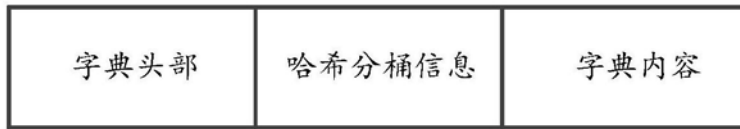


图1

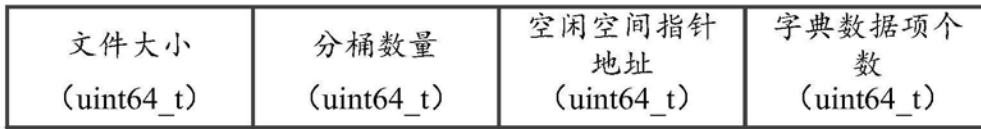


图2



图3

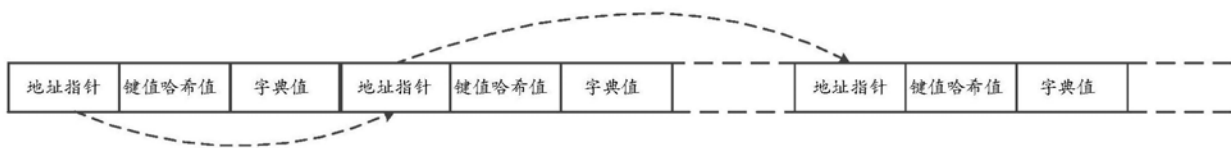


图4



图5

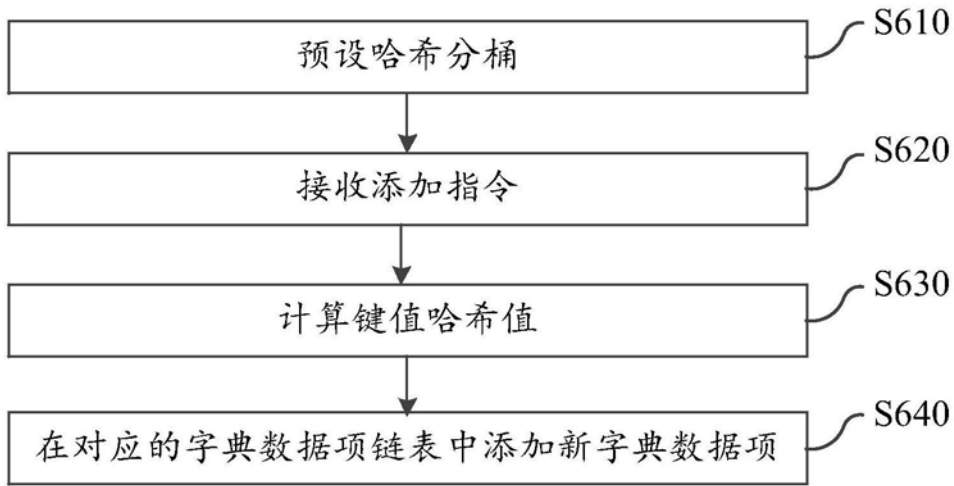


图6

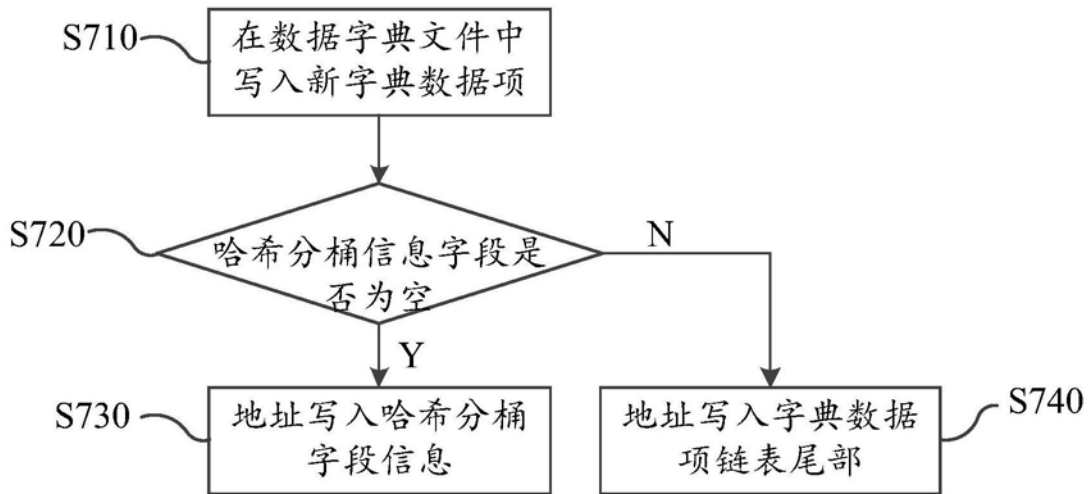


图7

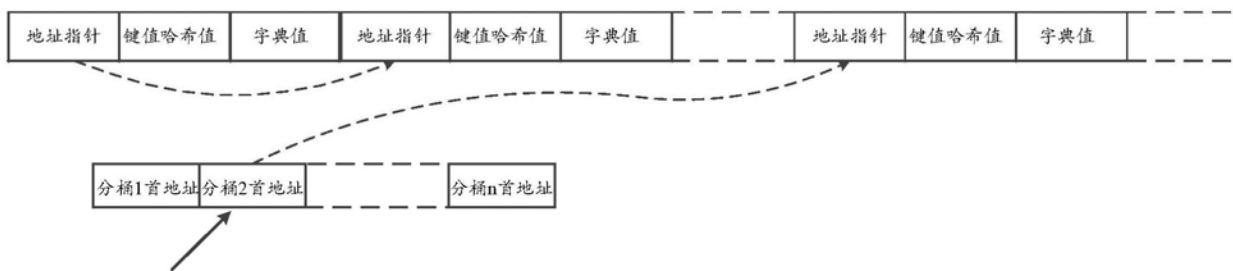


图8

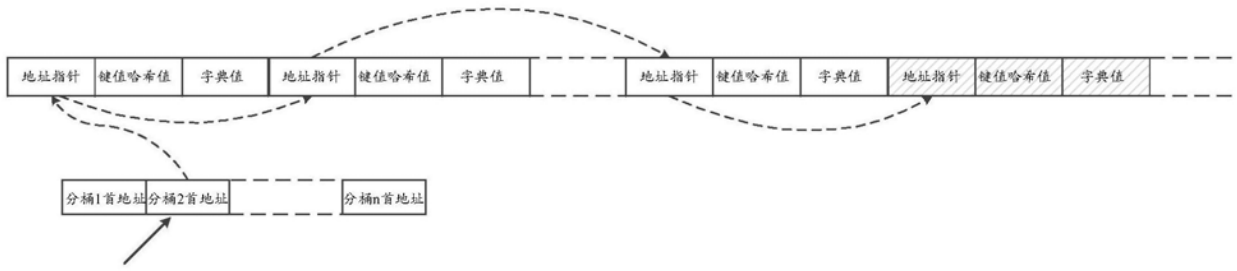


图9

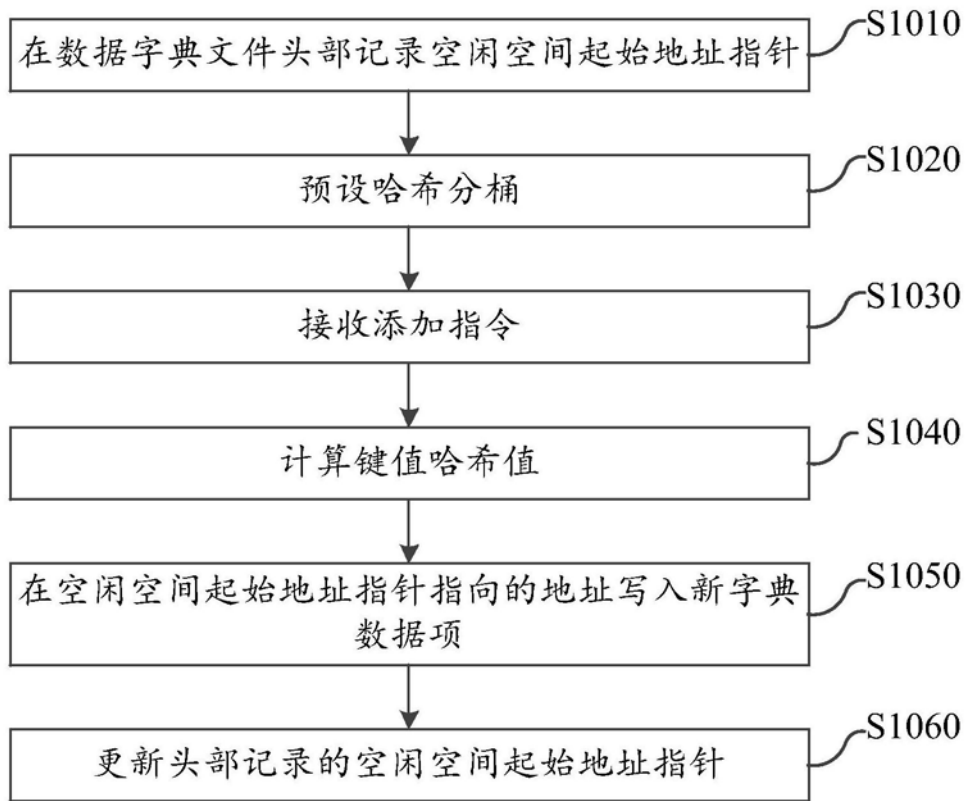


图10

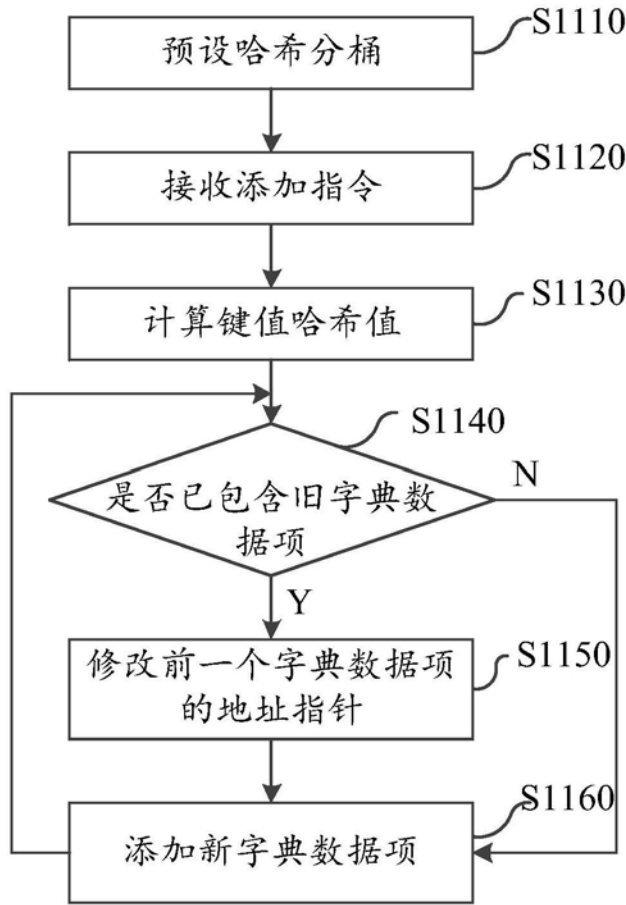


图11

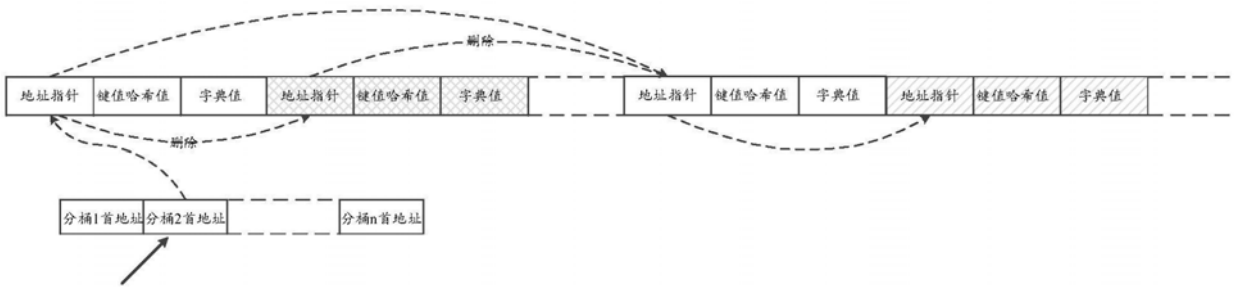


图12

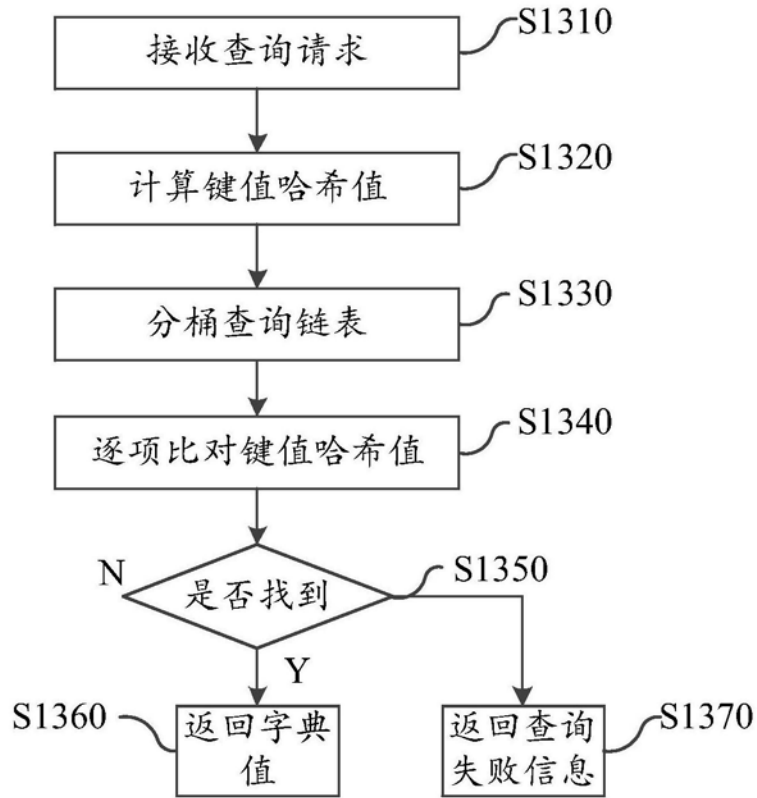


图13



图14



图15



图16

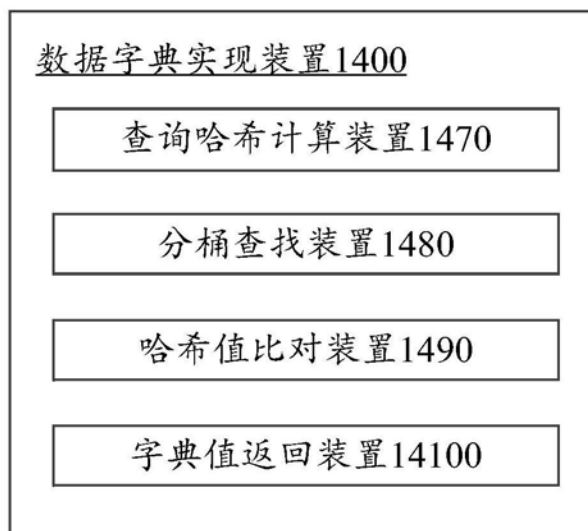


图17