

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2022 年 1 月 27 日 (27.01.2022)



(10) 国际公布号
WO 2022/016327 A1

(51) 国际专利分类号:
G06F 16/182 (2019.01)

(21) 国际申请号: PCT/CN2020/103065

(22) 国际申请日: 2020 年 7 月 20 日 (20.07.2020)

(25) 申请语言: 中文

(26) 公布语言: 中文

(71) 申请人: 中山大学 (SUN YAT-SEN UNIVERSITY)
[CN/CN]; 中国广东省广州市海珠区新港西路 135 号, Guangdong 510275 (CN)。

(72) 发明人: 廖婕 (LIAO, Jie); 中国广东省广州市海珠区新港西路 135 号, Guangdong 510275 (CN)。
解静仪 (XIE, Jingyi); 中国广东省广州市海珠区新港西路 135 号, Guangdong 510275 (CN)。 农

革 (NONG, Ge); 中国广东省广州市海珠区新港西路 135 号, Guangdong 510275 (CN)。

(74) 代理人: 深圳中一联合知识产权代理有限公司 (SHENZHEN ZHONGYI UNION INTELLECTUAL PROPERTY AGENCY CO., LTD.); 中国广东省深圳市福田区园岭街道深南中路 1014 号报春大厦 9 楼 (5 号信箱), Guangdong 518028 (CN)。

(81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL,

(54) Title: SAFE SUFFIX INDEX OUTSOURCING CALCULATION METHOD AND APPARATUS

(54) 发明名称: 一种安全的后缀索引外包计算方法及装置

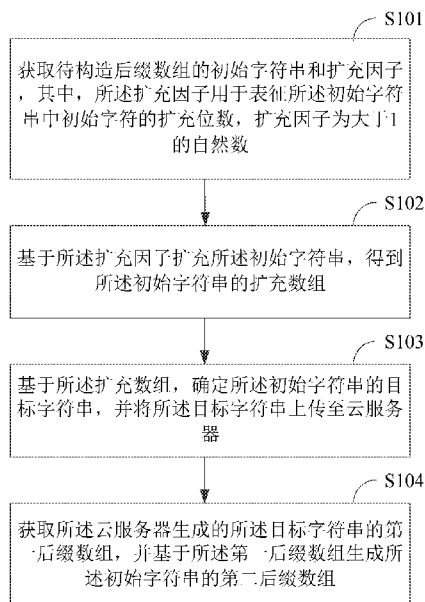


图 2

S101 Obtain an initial character string and an expansion factor of a suffix array to be constructed, the expansion factor being used to characterize the number of expansion digits of the initial characters in the initial character string, and the expansion factor being a natural number greater than 1
S102 Expand the initial character string on the basis of the expansion factor, and obtain an expanded array of the initial character string
S103 On the basis of the expanded array, determine a target character string of the initial character string, and upload the target character string to a cloud server
S104 Obtain a first suffix array of the target character string generated by the cloud server, and generate a second suffix array of the initial character string on the basis of the first suffix array

(57) Abstract: A safe suffix index outsourcing calculation method and apparatus, which are applicable to the technical field of data processing. The method comprises: obtaining an initial character string and an expansion factor of a suffix array to be constructed, the expansion factor being used to characterize the number of expansion digits of the initial characters in the initial character string, and the expansion factor being a natural number greater than 1 (S101); expanding the initial character string on the basis of the expansion factor, and obtaining an expanded array of the initial character string (S102); on the basis of the expanded array, determining a target

PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,
ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US,
UZ, VC, VN, WS, ZA, ZM, ZW。

- (84) 指定国(除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

- 包括国际检索报告(条约第21条(3))。

character string of the initial character string, and uploading the target character string to a cloud server (S103); and obtaining a first suffix array of the target character string generated by the cloud server, and generating a second suffix array of the initial character string on the basis of the first suffix array (S104). In the described method, not only can a suffix array of an initial character string be calculated, but it can also be ensured that the initial character string will not be leaked when calculating the suffix array, thus ensuring the safety of the initial character string.

(57) 摘要: 一种安全的后缀索引外包计算方法及装置, 适用于数据处理技术领域, 该方法包括: 获取待构造后缀数组的初始字符串和扩充因子, 其中, 所述扩充因子用于表征所述初始字符串中初始字符的扩充位数, 扩充因子为大于1的自然数 (S101); 基于所述扩充因子扩充所述初始字符串, 得到所述初始字符串的扩充数组 (S102); 基于所述扩充数组, 确定所述初始字符串的目标字符串, 并将所述目标字符串上传至云服务器 (S103); 获取所述云服务器生成的所述目标字符串的第一后缀数组, 并基于所述第一后缀数组生成所述初始字符串的第二后缀数组 (S104); 该方法既可以计算了初始字符串的后缀数组, 又可以保证在计算后缀数组时不会泄露初始字符串, 保证了初始字符串的安全。

一种安全的后缀索引外包计算方法及装置

技术领域

[0001] 本申请属于数据处理技术领域，尤其涉及一种安全的后缀索引外包计算方法及装置。

背景技术

[0002] 在数据处理中，后缀数组（或后缀索引）是一个重要且被广泛应用的数据结构，可以用于数据压缩、基因组比对、全文检索等多种领域。随着互联网技术的发展，需要存储的数据越来越多，为了有效管理及检索这些数据，亟需可高效构造和检索的全文索引，而后缀数组就是可用于进行全文索引的数组。

[0003] 目前一种可行的外包构造字符串的后缀数组的方法是将需要构造后缀数组的字符串传输至云服务器中，在云服务器中构造字符串的后缀数组，但是目前的云服务器的安全机制并不完全可靠，容易造成数据泄露，因此直接将原始字符串传输至云服务器上无法保证原始字符串的安全，造成构造后缀数组时原始字符串的安全性差。

发明概述

技术问题

[0004] 本申请实施例提供了一种安全的后缀索引外包计算方法及装置，可以解决目前外包构造字符串的后缀数组时字符串的安全性差的问题。

问题的解决方案

技术解决方案

[0005] 第一方面，本申请实施例提供了一种安全的后缀索引外包计算方法，包括：

[0006] 获取待构造后缀数组的初始字符串和扩充因子，其中，所述扩充因子用于表征所述初始字符串中初始字符的扩充位数，扩充因子为大于1的自然数；

[0007] 基于所述扩充因子扩充所述初始字符串，得到所述初始字符串的扩充数组；

[0008] 基于所述扩充数组，确定所述初始字符串的目标字符串，并将所述目标字符串上传至云服务器；

- [0009] 获取所述云服务器生成的所述目标字符串的第一后缀数组，并基于所述第一后缀数组生成所述初始字符串的第二后缀数组。
- [0010] 第二方面，本申请实施例提供了一种安全的后缀索引外包计算装置，包括：
- [0011] 数据获取模块，用于获取待构造后缀数组的初始字符串和扩充因子，其中，所述扩充因子用于表征所述初始字符串中初始字符的扩充位数，扩充因子为大于1的自然数；
- [0012] 数据扩充模块，用于基于所述扩充因子扩充所述初始字符串，得到所述初始字符串的扩充数组；
- [0013] 字符串生成模块，用于基于所述扩充数组，确定所述初始字符串的目标字符串，并将所述目标字符串上传至云服务器；
- [0014] 后缀数组生成模块，用于获取所述云服务器生成的所述目标字符串的第一后缀数组，并基于所述第一后缀数组生成所述初始字符串的第二后缀数组。
- [0015] 第三方面，本申请实施例提供了一种终端设备，包括：存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机程序，其特征在于，所述处理器执行所述计算机程序时实现上述第一方面中任一项所述的安全的后缀索引外包计算方法。
- [0016] 第四方面，本申请实施例提供了一种计算机可读存储介质，所述计算机可读存储介质存储有计算机程序，其特征在于，所述计算机程序被处理器执行时实现上述第一方面中任一项所述的安全的后缀索引外包计算方法。
- [0017] 第五方面，本申请实施例提供了一种计算机程序产品，当计算机程序产品在终端设备上运行时，使得终端设备执行上述第一方面中任一项所述的安全的后缀索引外包计算方法。
- [0018] 可以理解的是，上述第二方面至第五方面的有益效果可以参见上述第一方面中的相关描述，在此不再赘述。

发明的有益效果

有益效果

- [0019] 本申请实施例与现有技术相比存在的有益效果是：本申请根据扩充因子扩充初始字符串，得到扩充数组，基于扩充数据得到初始字符串的目标字符串，最后

获取云服务器生成的目标字符串的第一后缀数组，基于第一后缀数组得到初始字符串的第二后缀数组；本申请通过将初始字符串转换成目标字符串，通过目标字符串计算初始字符串的后缀数组，不直接使用初始字符串计算后缀数组，保证了初始字符串的安全性，本申请得到的第二后缀数组与直接使用初始字符串生成的后缀数组相同，因此本申请既能保证计算的后缀数组是初始字符串的后缀数组，又可以保证在计算后缀数组时不会泄露初始字符串，保证了初始字符串的安全。

对附图的简要说明

附图说明

[0020] 为了更清楚地说明本申请实施例中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本申请的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动性的前提下，还可以根据这些附图获得其他的附图。

[0021] 图1是本申请一实施例提供的安全的后缀索引外包计算方法的应用场景示意图；

[0022] 图2是本申请一实施例提供的安全的后缀索引外包计算方法的流程示意图；

[0023] 图3是本申请一实施例提供的扩充字符串的确定方法的流程示意图；

[0024] 图4是本申请一实施例提供的目标字符串的确定方法的流程示意图；

[0025] 图5是本申请一实施例提供的第二数组的确定方法的具体实现过程的流程示意图；

[0026] 图6是本申请一实施例提供的安全的后缀索引外包计算装置的结构示意图；

[0027] 图7是本申请一实施例提供的终端设备的结构示意图；

[0028] 图8是本申请一实施例提供的计算机的部分结构的框图。

发明实施例

本发明的实施方式

[0029] 以下描述中，为了说明而不是为了限定，提出了诸如特定系统结构、技术之类的具体细节，以便透彻理解本申请实施例。然而，本领域的技术人员应当清楚，在没有这些具体细节的其它实施例中也可以实现本申请。在其它情况中，省

略对众所周知的系统、装置、电路以及方法的详细说明，以免不必要的细节妨碍本申请的描述。

[0030] 应当理解，当在本申请说明书和所附权利要求书中使用时，术语“包括”指示所描述特征、整体、步骤、操作、元素和/或组件的存在，但并不排除一个或多个其它特征、整体、步骤、操作、元素、组件和/或其集合的存在或添加。

[0031] 另外，在本申请说明书和所附权利要求书的描述中，术语“第一”、“第二”、“第三”等仅用于区分描述，而不能理解为指示或暗示相对重要性。

[0032] 目前计算字符串的后缀数组的方法较多，但是在数据规模较大时，普通的计算机依然难以提供构建后缀数组所需要的计算能力，例如：对于长度为 n 的字符串 S ，使用目前整体性能最好的归纳排序方法构建其后缀数组需要 $O(n)$ 的时间开销和 $O(n \log n)$ 位的空间开销；当 n 较大时，后缀排序对计算机的配置要求很高，而通过互联网使用数据中心的云服务器能够以较低成本获得较高的运算性能，性价比更优。

[0033] 目前使用云服务器计算字符串的后缀数组是将原始的字符串不加保密的直接上传至云服务器，而目前的云服务器的安全机制并不完全可靠，容易造成数据泄露，因此直接将原始的字符串上传至云服务器计算后缀数组的方法安全性不高。

[0034] 本申请通过将原始的字符串转换成目标字符串，将目标字符串上传至云服务器，通过目标字符串计算原始的字符串的后缀数组，既能得到准确的原始的字符串的后缀数组，又能保证原始字符串的安全。

[0035] 图1为本申请实施例提供的安全的后缀索引外包计算方法的应用场景示意图，上述安全的后缀索引外包计算方法可以用于计算字符串的后缀数组。其中，存储设备10用于存储初始字符串和扩充因子，终端设备20用于从存储设备10中获取初始字符串和扩充因子，并基于扩充因子和初始字符串生成初始字符串的目标字符串，将目标字符串发送至云服务器30，云服务器30生成目标字符串的第一后缀数组后将第一后缀数组发送至终端设备20，终端设备20通过第一后缀数组计算初始字符串的第二后缀数组。

[0036] 以下结合图1对本申请实施例的安全的后缀索引外包计算方法进行详细说明。

[0037] 图2示出了本申请提供的安全的后缀索引外包计算方法的示意性流程图，参照图2，对该方法的详述如下：

[0038] S101，获取待构造后缀数组的初始字符串和扩充因子，其中，所述扩充因子用于表征所述初始字符串中初始字符的扩充位数，扩充因子为大于1的自然数。

[0039] 在本实施例中，初始字符串指的是最原始的字符串，也就是待构造后缀数组的字符串，例如，初始字符串可以是 $S[0, n) = \text{banananana\$}$ ，需要计算初始字符串的后缀数组。

[0040] 扩充因子指的是需要将初始字符串中的一个字符扩充到的位数，可以根据需要设置，例如，扩充因子是3，则说明需要将一个字符扩充到3位，也就是3个字符。

[0041] 获取初始字符串和扩充因子的方法可以是存储设备发送过来的，也可以是用户输入的初始字符串和扩充因子。

[0042] S102，基于所述扩充因子扩充所述初始字符串，得到所述初始字符串的扩充数组。

[0043] 在本实施例中，由于扩充因子是用于扩充初始字符串中的初始字符的，所以基于扩充因子可以得到扩充后的字符串，并将扩充后的字符串组成扩充数组。

[0044] 在一种可能的实现方式中，步骤S102的实现过程可以包括：

[0045] S1021，基于所述扩充因子，扩充所述初始字符串中的每个初始字符，得到每个初始字符的扩充字符串，并将所述扩充字符串组成所述扩充数组。

[0046] 在本实施例中，对初始字符串中初始字符进行扩充，需要对每个初始字符均进行扩充，且在扩充时是基于初始字符串进行扩充的，需要根据扩充因子，将初始字符与其后面的初始字符组成扩充字符串。

[0047] 扩充数组中扩充字符串在扩充数组中的位置与对应的初始字符在初始字符串中的位置相同，例如，初始字符在初始字符串中的位置是第3个字符，则第3个初始字符的扩充字符串在扩充数组中的位置是第3个字符串。

[0048] 作为举例，如果初始字符串为bana，第1个初始字符b的扩充字符串为ba，第2个初始字符a的扩充字符串为an，第3个初始字符n的扩充字符串为na，第4个初始字符a的扩充字符串为a，因此扩充字符串组成的扩充数组为[ba, an, na, a]。

- [0049] 如图3所示, 具体的, 步骤S1021的实现过程可以包括:
- [0050] S10211, 选取所述初始字符串中的第 i 个初始字符, 其中, $i \geq 1$ 。
- [0051] 在本实施例中, 初始字符串的下标是从0开始编号的, 例如, 初始字符串为banana\$, 第1个字符为b对应的下标为0, 第2个字符为a对应的下标为1, 第7个字符为\$对应的下标为6。
- [0052] 从初始字符串的第1个初始字符开始选取, 直到选取到最后一个初始字符, 对初始字符串中每一个初始字符进行扩充。
- [0053] S10212, 若 $i+k-1 \leq n$, 则将所述初始字符串中的第 i 个初始字符至第 $i+k-1$ 个初始字符作为所述第 i 个初始字符的扩充字符串, 其中, k 为扩充因子, n 为所述初始字符串中初始字符的总数, $n > 1$ 。
- [0054] 在本实施例中, 由于需要扩充初始字符, 且扩充时增加的字符是初始字符串中与需要扩充的初始字符相连且在初始字符之后的字符, 所以需要根据需要进行扩充的初始字符在初始字符串中的位置确定扩充字符串。
- [0055] 作为举例, 初始字符串为banana\$, 扩充因子为3, 对于第1个初始字符b, 由于 $i+k-1=1+3-1 < 7$, 则将第1个初始字符至第 $i+k-1=1+3-1=3$ 个初始字符作为第1个初始字符的扩充字符串, 即第1个初始字符的扩充字符串为ban。
- [0056] 对于第2个初始字符a, 由于 $i+k-1=2+3-1 < 7$, 则将第2个初始字符至第 $i+k-1=2+3-1=4$ 个初始字符作为第2个初始字符的扩充字符串, 即第2个初始字符的扩充字符串为ana。
- [0057] S10213, 若 $i+k-1 > n$, 则将所述初始字符串中的第 i 个初始字符至第 n 个初始字符作为所述第 i 个初始字符的扩充字符串。
- [0058] 在本实施例中, 如果第 i 个初始字符之后的初始字符的个数小于 $k-1$, 则只需要将第 i 个初始字符至第 n 个初始字符作为所述第 i 个初始字符的扩充字符串即可。
- [0059] 作为举例, 初始字符串为banana\$, 扩充因子为3, 对于第6个初始字符a, 由于 $i+k-1=6+3-1 > 7$, 则将第6个初始字符至第7个初始字符作为第6个初始字符的扩充字符串, 即第6个初始字符的扩充字符串为a\$。
- [0060] 对于第7个初始字符\$, 由于 $i+k-1=7+3-1 > 7$, 则将第7个初始字符至第7个初始字符作为第7个初始字符的扩充字符串, 即第7个初始字符的扩充字符串为\$。

- [0061] S103, 基于所述扩充数组, 确定所述初始字符串的目标字符串, 并将所述目标字符串上传至云服务器。
- [0062] 在本实施例中, 目标字符串是构建的需要代替初始字符串上传至云服务器计算后缀数组的字符串, 由于云服务器不能直接从变换后的目标字符串恢复得到初始字符串, 保证了初始字符串的安全。
- [0063] 在本实施例中, 云服务器基于归纳排序方法生成所述目标字符串的第一后缀数组。
- [0064] S104, 获取所述云服务器生成的所述目标字符串的第一后缀数组, 并基于所述第一后缀数组生成所述初始字符串的第二后缀数组。
- [0065] 在本实施例中, 云服务器生成目标字符串的第一后缀数组后, 将第一后缀数组传回终端设备, 终端设备根据第一后缀数组生成初始字符串的第二后缀数组, 完成对初始字符串的后缀数组的构建。
- [0066] 本申请实施例中, 根据扩充因子扩充初始字符串, 得到扩充数组, 基于扩充数组得到初始字符串的目标字符串, 最后获取云服务器生成的目标字符串的第一后缀数组, 基于第一后缀数组得到初始字符串的第二后缀数组; 本申请通过将初始字符串转换成目标字符串, 通过目标字符串计算初始字符串的后缀数组, 不直接使用初始字符串计算后缀数组, 保证了初始字符串的安全性, 本申请得到的第二后缀数组与直接使用初始字符串生成的后缀数组相同, 因此本申请既能保证计算的后缀数组是初始字符串的后缀数组, 又可以保证在计算后缀数组时不会泄露初始字符串, 保证了初始字符串的安全。
- [0067] 如图4所示, 在一种可能的实现方式中, 步骤S103的实现过程可以包括:
- [0068] S1031, 将所述初始字符串分成预设分块数量个字符串块。
- [0069] 在本实施例中, 分块数量表征将初始字符串中的初始字符分成的块的数量, 分块数量可以是用户输入的, 也可以预先存储在终端设备中的, 还可以是从外部设备中获取的。分块数量可以根据需要设置, 例如, 设置为3、4或5。
- [0070] 各个字符串块中的字符的数量可以相同也可以不同, 也就是可以将初始字符平均分配到各个字符串块中, 也可以不是平均分配的。
- [0071] 作为举例, 如果分块数量为4, 初始字符串为S=banananana\$, 则可以将初始字

字符串分块得到 $S=[ba][nan][ana][na\$]$ ，其中 $[ba]$ 、 $[nan]$ 、 $[ana]$ 、 $[na\$]$ 分别为4个字符串块。

[0072] S1032，基于所述字符串块中的字符，将所述扩充数组分成数组块。

[0073] 具体的，将所述字符串块中的各个字符对应的扩充字符串组成一个数组块，得到所述扩充数据的各个数组块。

[0074] 在本实施例中，字符串块是由初始字符串分块得到的，因此字符串块中的每个字符均对应一个扩充字符串，将字符串块中的字符对应的扩充字符串放在一个数组块中，即完成对扩充数组的分块。

[0075] 作为举例，根据初始字符串得到的字符串块分别为 $[ba][nan][ana][na\$]$ ，初始字符串的扩充数组为：

[0076] $S' = (ban) (ana) (nan) (ana) (nan) (ana) (nan) (ana) (na\$)$
 $(a\$) (\$)$ ，则将扩充数组分块后得到： $S' = [(ban) (ana)][(nan) (ana) (nan)][(ana) (nan) (ana)][(na\$) (a\$) (\$)]$ ，其中： $[(ban) (ana)]$ 、 $[(nan) (ana) (nan)]$ 、 $[(ana) (nan) (ana)]$ 、 $[(na\$) (a\$) (\$)]$ 分别为4个数组块。

[0077] S1033，基于所述字符串块和所述数组块，确定所述初始字符串的目标字符串。

[0078] 在本实施例中，通过字符串块和数组块即可得到目标字符串。

[0079] 在一种可能的实现方式中，步骤S1033的实现过程可以包括：

[0080] S10331，基于各个所述字符串块中最后一个字符在所述初始字符串中的第一下标，得到由第一下标组成的第一数组。

[0081] 在本实施例中，每个字符串块中均存在最后一个字符，且最后一个字符在初始字符串中对应一个第一下标，将所有的字符串块中的最后一个字符对应的第一下标组成第一数组。第一下标在第一数组中的排列顺序与字符串块中均存在的最后一个字符在初始字符串中的排列顺序相同。

[0082] 作为举例，初始字符串 $S=banananana\$$ ，分块后得到 $S=[ba][nan][ana][na\$]$ ， $[ba]$ 中最后一个字符 a 在初始字符串中的第一下标为1， $[nan]$ 中最后一个字符 n 在初始字符串中的第一下标为4， $[ana]$ 中最后一个字符 a 在初始字符串中的第一

下标为7，[na\$]中最后一个字符\$在初始字符串中的第一下标为10，则第一数组 $E[4] = \{1, 4, 7, 10\}$ 。

[0083] S10332，基于各个所述字符串块的块末后缀在所述初始字符串的第一后缀中的排名，得到由排名编号组成的第二数组。

[0084] 在本实施例中，字符串块的块末后缀由该字符串块中的最后一个字符至所述初始字符串的最后一个初始字符组成。

[0085] 初始字符串的第一后缀的个数与初始字符串中初始字符的个数相同，初始字符有几个，初始字符串的第一后缀就有几个，例如，初始字符串中有11个初始字符，则初始字符串有11个第一后缀。

[0086] 其中，初始字符串最后一个初始字符为一个第一后缀，所述初始字符串中第 g 个初始字符至最后一个初始字符组成一个第一后缀， $1 \leq g < n$ ， n 为所述初始字符串中初始字符的总数。

[0087] 如图5所示，具体的，步骤S10332的实现过程可以包括：

[0088] S103321，获取所述初始字符串中的第 j 个字符串块的块末后缀，其中，所述第 j 个字符串块的块末后缀由第 j 个字符串块中的最后一个字符至所述初始字符串的最后一个初始字符组成， $1 \leq j \leq m$ ， m 为分块数量。

[0089] 作为举例， $S = [ba][nan][ana][na\$]$ 中第一个字符串块的块末后缀为anananana\$，第二个字符串块的块末后缀为nanana\$，第三个字符串块的块末后缀为ana\$，第四个字符串块的块末后缀为\$。

[0090] S103323，获取所述初始字符串的各个第一后缀，将各个所述第一后缀按照字典顺序排序，其中，所述初始字符串最后一个初始字符为一个第一后缀，所述初始字符串中第 g 个初始字符至最后一个初始字符组成一个第一后缀， $1 \leq g < n$ ， n 为所述初始字符串中初始字符的总数。

[0091] 作为举例，如果初始字符串为 $S = \text{banananana\$}$ ，初始字符串的第一后缀分别为\$、a\$、na\$、ana\$、nana\$、anana\$、nanana\$、ananana\$、anananana\$、banananana\$；

[0092] 将以上第一后缀进行排序得到\$、a\$、ana\$、anana\$、ananana\$、anananana\$、banananana\$、na\$、nana\$、nanana\$、nananana\$。

[0093] S103324, 对排序后的第一后缀进行编号, 得到各个第一后缀的第一编号。

[0094] 在本实施例中, 每个排序后的第一后缀从0开始编号, 每个第一后缀对应一个第一编号。

[0095] 作为举例, 排序后的第一后缀为\$、a\$、ana\$、anana\$、ananana\$、anananana\$、banananana\$、na\$、nana\$、nanana\$、nananana\$, 则\$的第一编号为0, a\$的第一编号为1, 依次类推, 得到nananana\$的第一编号为10。

[0096] S103325, 将每个字符串块的块末后缀对应的编号作为所述第二数组, 其中, 所述块末后缀对应的编号为与所述块末后缀相同的第一后缀的第一编号。

[0097] 在本实施例中, 每个块末后缀在第一后缀均可以找到与之相同的, 将与之相同的第一后缀的第一编号作为块末后缀对应的编号, 利用块末后缀对应的编号组成第二数组。

[0098] 第二数组中数组元素的顺序与块末后缀对应的数组块的排序相同。

[0099] 作为举例, $S=[ba][nan][ana][na\$]$ 中第一个字符串块的块末后缀为anananana\$, 第二个字符串块的块末后缀为nanana\$, 第三个字符串块的块末后缀为ana\$, 第四个字符串块的块末后缀为\$;

[0100] 排序后的第一后缀为\$、a\$、ana\$、anana\$、ananana\$、anananana\$、banananana\$、na\$、nana\$、nanana\$、nananana\$, 则\$的第一编号为0, a\$的第一编号为1, 依次类推, 得到nananana\$的第一编号为10;

[0101] 则可以确定块末后缀anananana\$的编号为5, 块末后缀nanana\$的编号为9, 块末后缀ana\$的编码为2, 块末后缀\$的编码为0, 第二数组为 $A[4]=\{5, 9, 2, 0\}$ 。

[0102] S10333, 基于所述数组块、所述扩充数组、所述第一数组和所述第二数组, 得到第三数组。

[0103] 在一种可能的实现方式中, 步骤S10333的实现过程可以包括:

[0104] S103331, 基于公式 $V[A[a]]=E[a]$, 将所述第一数组中的数组元素放入所述第三数组中对应的第一位置, 其中, a为所述第一数组中数组元素的下标, $0 \leq a < m$, m为分块数量, $E[a]$ 为第一数组中下标a对应的数组元素, $A[a]$ 为第二数组中的下标a对应的数组元素; $V[A[a]]$ 为第三数组中下标 $A[a]$ 对应的数组元素。

[0105] 在本实施例中, 第一位置指的是第三数组中存放第一数组中的元素的位置, 根

据对应关系可以将第一数组中的元素放入第三数组中，得到第三数组中的部分数组元素。

- [0106] 作为举例，第一数组 $E[4]=\{1, 4, 7, 10\}$ ，第二数组为 $A[4]=\{5, 9, 2, 0\}$ ，则：
- [0107] $a=0$ 时，根据 $V[A[0]]=E[0]$ 得到 $V[5]=1$ ，也就是第三数组中下标5对应整数1；
- [0108] $a=1$ 时，根据 $V[A[1]]=E[1]$ 得到 $V[9]=4$ ，也就是第三数组中下标9对应整数4；
- [0109] $a=2$ 时，根据 $V[A[2]]=E[2]$ 得到 $V[2]=7$ ，也就是第三数组中下标2对应整数7；
- [0110] $a=3$ 时，根据 $V[A[3]]=E[3]$ 得到 $V[0]=10$ ，也就是第三数组中下标0对应整数10。

- [0111] S103332，将目标扩充字符串按照字典顺序进行排序，将排序后的各个目标扩充字符串在扩充数组中对应的下标作为第六数组，其中，所述目标扩充字符串指所述扩充数组中不是块末字符串的字符串，所述块末字符串指数组块中的最后一个字符串。

- [0112] 在本实施例中，由于将扩充数组分块后得到数组块，每个数组块中均有一个块末字符串，因此扩充数组中既包括块末字符串，又包括非块末字符串。块末字符串在扩充数组中的下标已经通过步骤S103331放入第三数组中，所以当前只需将非块末字符串在扩充数组中的下标放入第三数组即可。

- [0113] 作为举例，如果扩充数组为：

- [0114] $S'=[(ban) (ana)][(nan) (ana) (nan)][(ana) (nan) (ana)][(na\$) (a\$) (\$)]$ ，将扩充数组中的非块末字符串进行排序得到的第六数组为： $Q=\{a\$、ana、ana、ban、na\$、nan、nan\}$ 对应的下标分别为9、3、5、0、8、6、2。

- [0115] S103333，基于公式 $V[B_b]=M[b]$ ，将所述第六数组中的元素放入所述第三数组中对应的第二位置，其中， $b \geq 1$ ， $M[b]$ 为所述第六数组中的第b个位置上的数组元素， $V[B_b]$ 为第三数组中第二位置上的数组元素， B_b 为第b个第二位置，第二位置为所述第三数组中除第一位置以外的位置。

- [0116] 在本实施例中，在将第六数组中的字符放入第三数组中时，非块末字符串中存在与块末字符串相同的字符串时，可以基于块末字符串的第二后缀和非块末字符串的第三后缀的大小顺序将非块末字符串的下标放入第三数组中。非块末字

字符串的第三后缀为该非块末字符串对应的初始字符的第一后缀。块末字符串的第二后缀为该块末字符串对应的初始字符所在字符串块的块末后缀。

[0117] 作为举例，第一数组E[4]={1, 4, 7, 10}，第二数组为A[4]={5, 9, 2, 0}，扩充数组为：

[0118] S' = (ban) (ana) (nan) (ana) (nan) (ana) (nan) (ana) (na\$) (a\$) (\$)，第六数组为Q={a\$, ana, ana, ban, na\$, nan, nan}对应的下标分别为9、3、5、0、8、6、2；

[0119] 则得到的第三数组为V[11]={10, 9, 7, 3, 5, 1, 0, 8, 6, 4, 2}。

[0120] 在本实施例中，得到第三数组后，还可以确定第三数组中的桶，将相同扩充字符串对应的下标组成一个桶。作为举例，以上第三数组为V[11]={10, 9, 7, 3, 5, 1, 0, 8, 6, 4, 2}，扩充字符串ana构成的桶为[7、3、5、1]，nan构成的桶为[6、4、2]，还可以确定块末字符串在桶中的位置。

[0121] S10334，基于所述扩充数组，确定第四数组。

[0122] 在一种可能的实现方式中，步骤S10334的实现过程可以包括：

[0123] S103341，按照字典顺序对所述扩充数组中的扩充字符串进行排序，得到排序后的扩充字符串组成的目标数组。

[0124] 在本实施例中，对扩充字符串按照字典顺序进行排序，可以得到排序后的扩充字符串，具体的排序应满足的条件可以参照以下具体实现过程。

[0125] 在一种可能的实现方式中，步骤S103341的实现过程可以包括：

[0126] 按照字典顺序对所述扩充数组中的扩充字符串进行排序，得到扩充字符串序列；在所述扩充字符串序列中存在相同的扩充字符串，且相同的扩充字符串中至少一个是块末字符串时，基于所述初始字符串的后缀和所述块末字符串对应的初始字符所在字符串块的块末后缀，对所述相同的扩充字符串进行排序，得到所述目标数组。

[0127] 在本实施例中，按照字典顺序对扩充字符串进行排序得到的是初次排序后的结果，也就是如果存在相同的扩充字符串，相同的扩充字符串的先后顺序在经过初次排序后是不能确定的，所以，还要对扩充字符串序列进行再次排序。

[0128] 具体的，如果相同的扩充字符串中没有块末字符串，则相同的扩充字符串可以

随意排列。如果相同的扩充字符串中存在块末字符串，可以基于块末字符串的第二后缀和非块末字符串的第三后缀的大小对相同的扩充字符串进行排序。非块末字符串的第三后缀为该非块末字符串对应的初始字符的第一后缀。块末字符串的第二后缀为该块末字符串对应的初始字符所在字符串块的块末后缀。

[0129] 如果非块末字符串的第三后缀小于块末字符串的第二后缀，则将该非块末字符串放在该块末字符串的前面，反之在该块末字符串的后面。如果两个或多个非块末字符串相邻，则按照非块末字符串在扩充数组中的前后位置进行排列。

[0130] 作为举例，如果扩充数组为：

[0131] 初始字符串为S=banananana\$；

[0132] $S' = [(ban) (ana)] [(nan) (ana) (nan)] [(ana) (nan) (ana)] [(na\$) (a\$) (\$)]$ ，对扩充字符串进行排序后得到的目标数组为： $S'' = (\$) (a\$) (ana) (ana) (ana) (ana) (ban) (na\$) (nan) (nan) (nan)$ 。

[0133] (ana) (ana) (ana) (ana) 为四个相同的扩充字符串，分别为第一个数组块中的块末字符串，下标为1；第三个数组块中的块末字符串，下标为7；其他两个为非块末字符串，下标分别为3和5；第一个数组块中的块末字符串的第二后缀为anananana\$；

[0134] 第三个数组块中的块末字符串的第二后缀为ana\$；

[0135] 下标为3的非块末字符串的第三后缀为ananana\$；

[0136] 下标为5的非块末字符串的第三后缀为anana\$；

[0137] 因此排序后的(ana) (ana) (ana) (ana)对应的下标分别为7、3、5、1；

[0138] 依照上述方法排序后的(nan) (nan) (nan)对应的下标分别为6、4、2。

[0139] S103342，确定各个所述字符串块中的块末字符串在所述目标数组中的位置，所述块末字符串指数组块中的最后一个字符串。

[0140] 在本实施例中，确定了扩充字符串的排列顺序，自然可以确定各个块末字符串在排列后的扩充字符串中的位置。

[0141] 作为举例，如果目标数组为：

[0142] $S'' = (\$) (a\$) (ana) (ana) (ana) (ana) (ban) (na\$) (nan) ($

nan) (nan), 其中, (ana) (ana) (ana) (ana) 在扩充数组中对应的下标分别为7、3、5、1, (nan) (nan) (nan) 在扩充数组中对应的下标分别为6、4、2, 则块末字符串在所述目标数组中分别为下标为0、2、5、9的字符串。

[0143] S103343, 对所述目标数组中的扩充字符串进行编号, 得到所述第四数组。

[0144] 在本实施例中, 对所述目标数组中的扩充字符串进行编号可以从0开始, 然后依次加1, 得到第四数组, 第四数组是由编号组成的。具体的编号应该满足的预设规则如下所述。

[0145] 在一种可能的实现方式中, 步骤S103343的实现过程可以包括:

[0146] 依次对所述目标数组中的扩充字符串进行编号, 其中, 相同且相邻的非块末字符串具有相同的编号, 相同且不相邻的非块末字符串具有不同的编号, 块末字符串的编号不等于其他扩充字符串的编号。

[0147] 作为举例, 如果目标数组为:

[0148] S' ' = (\$) (a\$) (ana) (ana) (ana) (ana) (ban) (na\$) (nan) (nan) (nan) (nan), 其中, (ana) (ana) (ana) (ana) 在扩充数组中对应的下标分别为7、3、5、1, (nan) (nan) (nan) 在扩充数组中对应的下标分别为6、4、2, 则块末字符串在所述目标数组中分别为下标为0、2、5、9的字符串, 因此, 第四数组为 $R[11] = \{0, 1, 2, 3, 3, 4, 5, 6, 7, 8, 9\}$ 。

[0149] S10335, 基于所述第三数组和所述第四数组, 确定第五数组。

[0150] 在一种可能的实现方式中, 步骤S10335的实现过程可以包括:

[0151] 基于公式 $Y[V[c]] = R[c]$ 确定第五数组, 其中, $R[c]$ 为第四数组中下标c对应的数组元素, $V[c]$ 为第三数组中下标c对应的数组元素, $Y[V[c]]$ 为第五数组中下标 $V[c]$ 对应的数组元素, $0 \leq c < n$, n 为所述初始字符串中初始字符的总数。

[0152] 在本实施例中, 基于第三数组和第四数组中可以得到第五数组。

[0153] 作为举例, 第三数组为 $V[11] = \{10, 9, 7, 3, 5, 1, 0, 8, 6, 4, 2\}$, 第四数组为 $R[11] = \{0, 1, 2, 3, 3, 4, 5, 6, 7, 8, 9\}$, 则:

[0154] $c=0$ 时, 根据 $Y[V[0]] = R[0]$, $Y[10] = 0$;

[0155] $c=1$ 时, 根据 $Y[V[1]] = R[1]$, $Y[9] = 1$; 依次计算, 可得第五数组为 $Y[11] = \{5, 4, 9, 3, 8, 3, 7, 2, 6, 1, 0\}$ 。

[0156] S10336, 基于各个所述字符串块中的字符数量, 将所述第五数组分成数据块。

[0157] 在本实施例中, 初始字符串中字符串块中的字符数量是多少, 数据块中的字符数量就是多少, 且数据块的排列顺序与字符串块的排列顺序相同。

[0158] 作为举例, 如果分块后的初始字符串为 $S=[ba][nan][ana][na\$]$, 第五数组为 $Y[11]=\{5, 4, 9, 3, 8, 3, 7, 2, 6, 1, 0\}$;

[0159] 则分块后的第五数组为: $Y[11]=\{[5, 4][9, 3, 8][3, 7, 2][6, 1, 0]\}$, $[5, 4]$ 、 $[9, 3, 8]$ 、 $[3, 7, 2]$ 、 $[6, 1, 0]$ 为四个数据块。

[0160] S10337, 将各个所述数据块重新排序, 得到所述目标字符串。

[0161] 在本实施例中, 得到数据块后, 可以将数据块重新排列, 得到排列后的数组记为目标字符串。

[0162] 在一种可能的实现方式中, 步骤S10337的实现过程可以包括:

[0163] 基于第七数组将各个所述数据块重新排序, 得到所述目标字符串, 其中, 第七数组是将所述第五数组中的数据块的下标重新排序后得到的。

[0164] 在本实施例中, 第七数组可以是基于第五数组生成的, 第七数组中的数组元素是由第五数组中的数据块的下标组成的, 数据块的下标随机排列得到第七数组。

[0165] 在本实施例中, 有了第七数组, 基于第七数组中数据块的下标的排列顺序, 将数据块重新排列, 得到目标字符串。

[0166] 作为举例, 如果分块后的第五数组为: $Y[11]=\{[5, 4][9, 3, 8][3, 7, 2][6, 1, 0]\}$, 则数据块 $[5, 4]$ 的下标为0, 数据块 $[9, 3, 8]$ 的下标为1, 数据块 $[3, 7, 2]$ 的下标为2, 数据块 $[6, 1, 0]$ 的下标为3, 因此, 第七数组可以是 $P[4]=\{1, 0, 3, 2\}$, 当然第七数组还可以是数据块的下标其他排列顺序的组合, 在此不做限制;

[0167] 将 $Y[11]=\{[5, 4][9, 3, 8][3, 7, 2][6, 1, 0]\}$ 重新排列, 得到目标字符串为 $X[11]=\{9, 3, 8, 5, 4, 6, 1, 0, 3, 7, 2\}$ 。

[0168] 在一种可能的实现方式中, 步骤S104的实现过程可以包括:

[0169] S1041, 基于所述目标字符串中各个所述数据块中字符的个数, 确定由各个所述数据块中字符的个数组成的第八数组。

[0170] 在本实施例中, 第八数组中的数组元素表征的是数据块中字符的个数, 第八数

组中的数组元素的排列顺序与数据块在目标字符串中的排列顺序相同。

[0171] 作为举例，目标字符串为 $X[11]=\{9, 3, 8, 5, 4, 6, 1, 0, 3, 7, 2\}$ ，数据块分别为 $[9, 3, 8]$ $[5, 4]$ $[6, 1, 0]$ $[3, 7, 2]$ ，则第八数组为 $W[4]=\{3, 2, 3, 3\}$ 。

[0172] S1042，基于所述第八数组，确定第九数组。

[0173] 在一种可能的实现方式中，步骤S1042的实现过程可以包括：

[0174] 将0作为所述第九数组中的第一个数组元素，将所述第八数组中的第一个数组元素作为第九数组中的第二个数组元素；

[0175] 计算所述第八数组中的第一个数组元素至第d个数组元素的和，其中， $1 < d \leq m$ ，m为分块数量；

[0176] 将所述第八数组中的第一个数组元素至第d个数组元素的和作为所述第九数组中的第d+1个数组元素。

[0177] 作为举例，如果第八数组为 $W[4]=\{3, 2, 3, 3\}$ ，第九数组中第三个元素为 $3+2=5$ ，第四个元素为 $3+2+3=8$ ，第五个元素为 $3+2+3+3=11$ ，则第九数组为 $H[5]=\{0, 3, 5, 8, 11\}$ 。

[0178] S1043，基于所述第九数组和所述第一后缀数组，得到所述第二后缀数组。

[0179] 在一种可能的实现方式中，步骤S1043的实现过程可以包括：

[0180] 基于公式 $SA_s[e]=E[P[f]]-W[f]+1+SA_x[e]-H[f]$ ，得到所述第二后缀数组，其中，f基于公式 $H[f] \leq SA_x[e] < H[f+1]$ 确定， $SA_s[e]$ 为第二后缀数组中下标e对应的数组元素， $E[P[f]]$ 为第一数组中下标P[f]对应的数组元素，P[f]为第七数组中下标f对应的数组元素，W[f]为第八数组中下标f对应的数组元素， $SA_x[e]$ 为第一后缀数组中下标e对应的数组元素，H[f]为第九数组中下标f对应的数组元素，H[f+1]为第九数组中下标f+1对应的数组元素。

[0181] 作为举例，如果 $SA_x[11]=\{7, 6, 10, 8, 1, 4, 3, 5, 9, 2, 0\}$ ，第九数组为 $H[5]=\{0, 3, 5, 8, 11\}$ ，第八数组为 $W[4]=\{3, 2, 3, 3\}$ ，第七数组为 $P[4]=\{1, 0, 3, 2\}$ ，第一数组 $E[4]=\{1, 4, 7, 10\}$ ，则：

[0182] $e=0$ ， $SA_x[e]=7$ ， $f=2$ ， $SA_s[0]=E[P[2]]-W[2]+1+SA_x[0]-H[2]=10-3+1+7-5=10$ ；

[0183] $e=1$, $SA_x[e]=6$, $f=2$, $SA_s[1]=E[P[2]]-W[2]+1+SA_x[1]-H[2]=10-3+1+6-5=9$;

[0184] $e=2$, $SA_x[e]=10$, $f=3$, $SA_s[2]=E[P[3]]-W[3]+1+SA_x[2]-H[3]=7-3+1+10-8=7$;

[0185] 依次计算,

[0186] $e=10$, $SA_x[e]=0$, $f=0$, $SA_s[10]=E[P[0]]-W[0]+1+SA_x[10]-H[0]=4-3+1+0-0=2$;

[0187] 因此, 第二后缀数组为 $SA_s[11]=\{10, 9, 7, 5, 3, 1, 0, 8, 6, 4, 2\}$ 。

[0188] 应理解, 上述实施例中各步骤的序号的大小并不意味着执行顺序的先后, 各过程的执行顺序应以其功能和内在逻辑确定, 而不应对本申请实施例的实施过程构成任何限定。

[0189] 对应于上文实施例所述的安全的后缀索引外包计算方法, 图6示出了本申请实施例提供的安全的安全的后缀索引外包计算装置的结构框图, 为了便于说明, 仅示出了与本申请实施例相关的部分。

[0190] 参照图6, 该装置200可以包括: 数据获取模块210、数据扩充模块220、字符串生成模块230和后缀数组生成模块240。

[0191] 其中, 数据获取模块210, 用于获取待构造后缀数组的初始字符串和扩充因子, 其中, 所述扩充因子用于表征所述初始字符串中初始字符的扩充位数, 扩充因子为大于1的自然数;

[0192] 数据扩充模块220, 用于基于所述扩充因子扩充所述初始字符串, 得到所述初始字符串的扩充数组;

[0193] 字符串生成模块230, 用于基于所述扩充数组, 确定所述初始字符串的目标字符串, 并将所述目标字符串上传至云服务器;

[0194] 后缀数组生成模块240, 用于获取所述云服务器生成的所述目标字符串的第一后缀数组, 并基于所述第一后缀数组生成所述初始字符串的第二后缀数组。

[0195] 在一种可能的实现方式中, 数据扩充模块220具体可以包括:

[0196] 数据扩充单元, 用于基于所述扩充因子, 扩充所述初始字符串中的每个初始字符, 得到每个初始字符的扩充字符串, 并将所述扩充字符串组成所述扩充数组

- 。
- [0197] 在一种可能的实现方式中，数据扩充单元具体可以包括：
- [0198] 选取所述初始字符串中的第 i 个初始字符，其中， $i \geq 1$ ；
- [0199] 若 $i+k-1 \leq n$ ，则将所述初始字符串中的第 i 个初始字符至第 $i+k-1$ 个初始字符作为所述第 i 个初始字符的扩充字符串，其中， k 为扩充因子， n 为所述初始字符串中初始字符的总数， $n > 1$ ；
- [0200] 若 $i+k-1 > n$ ，则将所述初始字符串中的第 i 个初始字符至第 n 个初始字符作为所述第 i 个初始字符的扩充字符串。
- [0201] 在一种可能的实现方式中，字符串生成模块230具体可以包括：
- [0202] 第一分块单元，用于将所述初始字符串分成预设分块数量个字符串块；
- [0203] 第二分块单元，用于基于所述字符串块中的字符，将所述扩充数组分成数组块；
- [0204] 字符串生成单元，用于基于所述字符串块和所述数组块，确定所述初始字符串的目标字符串。
- [0205] 在一种可能的实现方式中，第二分块单元具体可以用于：
- [0206] 将所述字符串块中的各个字符对应的扩充字符串组成一个数组块，得到所述扩充数据的各个数组块。
- [0207] 在一种可能的实现方式中，字符串生成单元具体可以包括：
- [0208] 第一数组生成子单元，用于基于各个所述字符串块中最后一个字符在所述初始字符串中的第一下标，得到由第一下标组成的第一数组；
- [0209] 第二数组生成子单元，用于基于各个所述字符串块的块末后缀在所述初始字符串的第一后缀中的排名，得到由排名编号组成的第二数组；
- [0210] 第三数组生成子单元，用于基于所述数组块、所述扩充数组、所述第一数组和所述第二数组，得到第三数组；
- [0211] 第四数组生成子单元，用于基于所述扩充数组，确定第四数组；
- [0212] 第五数组生成子单元，用于基于所述第三数组和所述第四数组，确定第五数组；
- [0213] 数据块生成子单元，用于基于各个所述字符串块中的字符数量，将所述第五数

组分成数据块；

[0214] 目标字符串生成子单元，用于将各个所述数据块重新排序，得到所述目标字符串。

[0215] 在一种可能的实现方式中，第二数组生成子单元具体可以用于：

[0216] 获取所述初始字符串中的第j个字符串块的块末后缀，其中，所述第j个字符串块的块末后缀由第j个字符串块中的最后一个字符至所述初始字符串的最后一个初始字符组成， $1 \leq j \leq m$ ，m为分块数量；

[0217] 获取所述初始字符串的各个第一后缀，将各个所述第一后缀按照字典顺序排序，其中，所述初始字符串最后一个初始字符为一个第一后缀，所述初始字符串中第g个初始字符至最后一个初始字符组成一个第一后缀， $1 \leq g < n$ ，n为所述初始字符串中初始字符的总数；

[0218] 对排序后的第一后缀进行编号，得到各个第一后缀的第一编号；

[0219] 将每个字符串块的块末后缀对应的编号作为所述第二数组，其中，所述块末后缀对应的编号为与所述块末后缀相同的第一后缀的第一编号。

[0220] 在一种可能的实现方式中，第三数组生成子单元，具体可以用于：

[0221] 基于公式 $V[A[a]] = E[a]$ ，将所述第一数组中的数组元素放入所述第三数组中对应的第一位置，其中，a为所述第一数组中数组元素的下标， $0 \leq a < m$ ，m为分块数量，E[a]为第一数组中下标a对应的数组元素，A[a]为第二数组中的下标a对应的数组元素；V[A[a]]为第三数组中下标A[a]对应的数组元素；

[0222] 将目标扩充字符串按照字典顺序进行排序，将排序后的各个目标扩充字符串在扩充数组中对应的下标作为第六数组，其中，所述目标扩充字符串指所述扩充数组中不是块末字符串的字符串，所述块末字符串指数组块中的最后一个字符串；

[0223] 基于公式 $V[B_b] = M[b]$ ，将所述第六数组中的数组元素放入所述第三数组中对应的第二位置，其中， $b \geq 1$ ，M[b]为所述第六数组中的第b个位置上的数组元素，V[B_b]为第三数组中第二位置上的数组元素，B_b为第b个第二位置，第二位置为所述第三数组中除第一位置以外的位置。

[0224] 在一种可能的实现方式中，第四数组生成子单元具体可以用于：

- [0225] 按照字典顺序对所述扩充数组中的扩充字符串进行排序，得到排序后的扩充字符串组成的目标数组；
- [0226] 确定各个所述字符串块中的块末字符串在所述目标数组中的位置，所述块末字符串指数组块中的最后一个字符串；
- [0227] 对所述目标数组中的扩充字符串进行编号，得到所述第四数组。
- [0228] 在一种可能的实现方式中，第四数组生成子单元具体还可以用于：
- [0229] 按照字典顺序对所述扩充数组中的扩充字符串进行排序，得到扩充字符串序列；
- [0230] 在所述扩充字符串序列中存在相同的扩充字符串，且相同的扩充字符串中至少一个是块末字符串时，基于所述初始字符串的后缀和所述块末字符串对应的初始字符所在字符串块的块末后缀，对所述相同的扩充字符串进行排序，得到所述目标数组。
- [0231] 在一种可能的实现方式中，第四数组生成子单元具体还可以用于：
- [0232] 依次对所述目标数组中的扩充字符串进行编号，其中，相同且相邻的非块末字符串具有相同的编号，相同且不相邻的非块末字符串具有不同的编号，块末字符串的编号不等于其他扩充字符串的编号。
- [0233] 在一种可能的实现方式中，第五数组生成子单元具体可以用于：
- [0234] 基于公式 $Y[V[c]] = R[c]$ 确定第五数组，其中， $R[c]$ 为第四数组中下标 c 对应的数组元素， $V[c]$ 为第三数组中下标 c 对应的数组元素， $Y[V[c]]$ 为第五数组中下标 $V[c]$ 对应的字符， $0 \leq c < n$ ， n 为所述初始字符串中初始字符的总数。
- [0235] 在一种可能的实现方式中，目标字符串生成子单元具体可以用于：
- [0236] 基于第七数组将各个所述数据块重新排序，得到所述目标字符串，其中，第七数组是将所述第五数组中的数据块的下标重新排序后得到的。
- [0237] 在一种可能的实现方式中，所述第一后缀数组的生成方法包括：
- [0238] 基于归纳排序方法生成所述目标字符串的第一后缀数组。
- [0239] 在一种可能的实现方式中，后缀数组生成模块240具体可以包括：
- [0240] 第八数组生成单元，用于基于所述目标字符串中各个所述数据块中字符的个数，确定由各个所述数据块中字符的个数组成的第八数组；

- [0241] 第九数组生成单元，用于基于所述第八数组，确定第九数组；
- [0242] 第二后缀数组生成单元，用于基于所述第九数组和所述第一后缀数组，得到所述第二后缀数组。
- [0243] 在一种可能的实现方式中，第九数组生成单元具体可以用于：
- [0244] 将0作为所述第九数组中的第一个数组元素，将所述第八数组中的第一个数组元素作为第九数组中的第二个数组元素；
- [0245] 计算所述第八数组中的第一个数组元素至第d个数组元素的和，其中， $1 < d \leq m$ ，m为分块数量；
- [0246] 将所述第八数组中的第一个数组元素至第d个数组元素的和作为所述第九数组中的第d+1个数组元素。
- [0247] 在一种可能的实现方式中，第二后缀数组生成单元具体可以用于：
- [0248] 基于公式 $SA_s[e] = E[P[f]] - W[f] + 1 + SA_x[e] - H[f]$ ，得到所述第二后缀数组，其中，f基于公式 $H[f] \leq SA_x[e] < H[f+1]$ 确定， $SA_s[e]$ 为第二后缀数组中下标e对应的数组元素， $E[P[f]]$ 为第一数组中下标P[f]对应的数组元素，P[f]为第七数组中下标f对应的数组元素，W[f]为第八数组中下标f对应的数组元素， $SA_x[e]$ 为第一后缀数组中下标e对应的数组元素，H[f]为第九数组中下标f对应的数组元素，H[f+1]为第九数组中下标f+1对应的数组元素。
- [0249] 需要说明的是，上述装置/单元之间的信息交互、执行过程等内容，由于与本申请方法实施例基于同一构思，其具体功能及带来的技术效果，具体可参见方法实施例部分，此处不再赘述。
- [0250] 所属领域的技术人员可以清楚地了解到，为了描述的方便和简洁，仅以上述各功能单元、模块的划分进行举例说明，实际应用中，可以根据需要而将上述功能分配由不同的功能单元、模块完成，即将所述装置的内部结构划分成不同的功能单元或模块，以完成以上描述的全部或者部分功能。实施例中的各功能单元、模块可以集成在一个处理单元中，也可以是各个单元单独物理存在，也可以两个或两个以上单元集成在一个单元中，上述集成的单元既可以采用硬件的形式实现，也可以采用软件功能单元的形式实现。另外，各功能单元、模块的具体名称也只是为了便于相互区分，并不用于限制本申请的保护范围。上述系

统中单元、模块的具体工作过程，可以参考前述方法实施例中的对应过程，在此不再赘述。

[0251] 本申请实施例还提供了一种终端设备，参见图7，该终端设备400可以包括：至少一个处理器410、存储器420以及存储在所述存储器420中并可在所述至少一个处理器410上运行的计算机程序，所述处理器410执行所述计算机程序时实现上述任意各个方法实施例中的步骤，例如图2所示实施例中的步骤S101至步骤S104。或者，处理器410执行所述计算机程序时实现上述各装置实施例中各模块/单元的功能，例如图6所示模块210至240的功能。

[0252] 示例性的，计算机程序可以被分割成一个或多个模块/单元，一个或者多个模块/单元被存储在存储器420中，并由处理器410执行，以完成本申请。所述一个或多个模块/单元可以是能够完成特定功能的一系列计算机程序段，该程序段用于描述计算机程序在终端设备400中的执行过程。

[0253] 本领域技术人员可以理解，图7仅仅是终端设备的示例，并不构成对终端设备的限定，可以包括比图示更多或更少的部件，或者组合某些部件，或者不同的部件，例如输入输出设备、网络接入设备、总线等。

[0254] 处理器410可以是中央处理单元(Central Processing Unit, CPU)，还可以是其他通用处理器、数字信号处理器(Digital Signal Processor, DSP)、专用集成电路(Application Specific Integrated Circuit, ASIC)、现成可编程门阵列(Field-Programmable Gate Array, FPGA)或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。

[0255] 存储器420可以是终端设备的内部存储单元，也可以是终端设备的外部存储设备，例如插接式硬盘，智能存储卡(Smart Media Card, SMC)，安全数字(Secure Digital, SD)卡，闪存卡(Flash Card)等。所述存储器420用于存储所述计算机程序以及终端设备所需的其他程序和数据。所述存储器420还可以用于暂时地存储已经输出或者将要输出的数据。

[0256] 总线可以是工业标准体系结构(Industry Standard Architecture, ISA)总线、外部设备互连(Peripheral Component, PCI)总线或扩展工业标准体系结

构 (Extended Industry Standard Architecture, EISA) 总线等。总线可以分为地址总线、数据总线、控制总线等。为便于表示, 本申请附图中的总线并不限定仅有一根总线或一种类型的总线。

[0257] 本申请实施例提供的安全的后缀索引外包计算方法可以应用于计算机、平板电脑、笔记本电脑、上网本、个人数字助理 (personal digital assistant, PDA) 等终端设备上, 本申请实施例对终端设备的具体类型不作任何限制。

[0258] 以所述终端设备为计算机为例。图8示出的是与本申请实施例提供的计算机的部分结构的框图。参考图8, 计算机包括: 通信电路510、存储器520、输入单元530、显示单元540、音频电路550、无线保真 (wireless fidelity, WiFi) 模块560、处理器570以及电源580等部件。

[0259] 下面结合图8对计算机的各个构成部件进行具体的介绍:

[0260] 通信电路510可用于收发信息或通话过程中, 信号的接收和发送, 特别地, 将图像采集设备发送的图像样本接收后, 给处理器570处理; 另外, 将图像采集指令发送给图像采集设备。通常, 通信电路包括但不限于天线、至少一个放大器、收发信机、耦合器、低噪声放大器 (Low Noise Amplifier, LNA)、双工器等。此外, 通信电路510还可以通过无线通信与网络和其他设备通信。上述无线通信可以使用任一通信标准或协议, 包括但不限于全球移动通讯系统 (Global System of Mobile communication, GSM)、通用分组无线服务 (General Packet Radio Service, GPRS)、码分多址 (Code Division Multiple Access, CDMA)、宽带码分多址 (Wideband Code Division Multiple Access, WCDMA)、长期演进 (Long Term Evolution, LTE)、电子邮件、短消息服务 (Short Messaging Service, SMS) 等。

[0261] 存储器520可用于存储软件程序以及模块, 处理器570通过运行存储在存储器520的软件程序以及模块, 从而执行计算机的各种功能应用以及数据处理。存储器520可主要包括存储程序区和存储数据区, 其中, 存储程序区可存储操作系统、至少一个功能所需的应用程序 (比如声音播放功能、图像播放功能等) 等; 存储数据区可存储根据计算机的使用所创建的数据 (比如音频数据、电话本等) 等。此外, 存储器520可以包括高速随机存取存储器, 还可以包括非易失性存储

器，例如至少一个磁盘存储器件、闪存器件、或其他易失性固态存储器件。

[0262] 输入单元530可用于接收输入的数字或字符信息，以及产生与计算机的用户设置以及功能控制有关的键信号输入。具体地，输入单元530可包括触控面板531以及其他输入设备532。触控面板531，也称为触摸屏，可收集用户在其上或附近的触摸操作（比如用户使用手指、触笔等任何适合的物体或附件在触控面板531上或在触控面板531附近的操作），并根据预先设定的程式驱动相应的连接装置。可选的，触控面板531可包括触摸检测装置和触摸控制器两个部分。其中，触摸检测装置检测用户的触摸方位，并检测触摸操作带来的信号，将信号传送给触摸控制器；触摸控制器从触摸检测装置上接收触摸信息，并将它转换成触点坐标，再送给处理器570，并能接收处理器570发来的命令并加以执行。此外，可以采用电阻式、电容式、红外线以及表面声波等多种类型实现触控面板531。除了触控面板531，输入单元530还可以包括其他输入设备532。具体地，其他输入设备532可以包括但不限于物理键盘、功能键（比如音量控制按键、开关按键等）、轨迹球、鼠标、操作杆等中的一种或多种。

[0263] 显示单元540可用于显示由用户输入的信息或提供给用户的信息以及计算机的各种菜单。显示单元540可包括显示面板541，可选的，可以采用液晶显示器（Liquid Crystal Display, LCD）、有机发光二极管（Organic Light-Emitting Diode, OLED）等形式来配置显示面板541。进一步的，触控面板531可覆盖显示面板541，当触控面板531检测到在其上或附近的触摸操作后，传送给处理器570以确定触摸事件的类型，随后处理器570根据触摸事件的类型在显示面板541上提供相应的视觉输出。虽然在图8中，触控面板531与显示面板541是作为两个独立的部件来实现计算机的输入和输入功能，但是在某些实施例中，可以将触控面板531与显示面板541集成而实现计算机的输入和输出功能。

[0264] 音频电路550可提供用户与计算机之间的音频接口。音频电路550可将接收到的音频数据转换后的电信号，传输到扬声器由扬声器转换为声音信号输出；另一方面，传声器将收集的声音信号转换为电信号，由音频电路550接收后转换为音频数据，再将音频数据输出处理器570处理后，经通信电路510以发送给比如另一计算机，或者将音频数据输出至存储器520以便进一步处理。

- [0265] WiFi属于短距离无线传输技术，计算机通过WiFi模块560可以帮助用户收发电子邮件、浏览网页和访问流式媒体等，它为用户提供了无线的宽带互联网访问。虽然图8示出了WiFi模块560，但是可以理解的是，其并不属于计算机的必须构成，完全可以根据需要在不改变发明的本质的范围内而省略。
- [0266] 处理器570是计算机的控制中心，利用各种接口和线路连接整个计算机的各个部分，通过运行或执行存储在存储器520内的软件程序和/或模块，以及调用存储在存储器520内的数据，执行计算机的各种功能和处理数据，从而对计算机进行整体监控。可选的，处理器570可包括一个或多个处理单元；优选的，处理器570可集成应用处理器和调制解调处理器，其中，应用处理器主要处理操作系统、用户界面和应用程序等，调制解调处理器主要处理无线通信。可以理解的是，上述调制解调处理器也可以不集成到处理器570中。
- [0267] 计算机还包括给各个部件供电的电源580（比如电池），优选的，电源580可以通过电源管理系统与处理器570逻辑相连，从而通过电源管理系统实现管理充电、放电、以及功耗管理等功能。
- [0268] 本申请实施例还提供了一种计算机可读存储介质，所述计算机可读存储介质存储有计算机程序，所述计算机程序被处理器执行时可实现上述安全的后缀索引外包计算方法各个实施例中的步骤。
- [0269] 本申请实施例提供了一种计算机程序产品，当计算机程序产品在移动终端上运行时，使得移动终端执行时可实现上述安全的后缀索引外包计算方法各个实施例中的步骤。
- [0270] 所述集成的单元如果以软件功能单元的形式实现并作为独立的产品销售或使用时，可以存储在一个计算机可读取存储介质中。基于这样的理解，本申请实现上述实施例方法中的全部或部分流程，可以通过计算机程序来指令相关的硬件来完成，所述的计算机程序可存储于一计算机可读存储介质中，该计算机程序在被处理器执行时，可实现上述各个方法实施例的步骤。其中，所述计算机程序包括计算机程序代码，所述计算机程序代码可以为源代码形式、对象代码形式、可执行文件或某些中间形式等。所述计算机可读介质至少可以包括：能够将计算机程序代码携带到拍照装置/终端设备的任何实体或装置、记录介质、计

算机存储器、只读存储器（ROM，Read-Only Memory）、随机存取存储器（RAM，Random Access Memory）、电载波信号、电信信号以及软件分发介质。例如U盘、移动硬盘、磁碟或者光盘等。在某些司法管辖区，根据立法和专利实践，计算机可读介质不可以是电载波信号和电信信号。

[0271] 在上述实施例中，对各个实施例的描述都各有侧重，某个实施例中未详述或记载的部分，可以参见其它实施例的相关描述。

[0272] 本领域普通技术人员可以意识到，结合本文中所公开的实施例描述的各示例的单元及算法步骤，能够以电子硬件、或者计算机软件和电子硬件的结合来实现。这些功能究竟以硬件还是软件方式来执行，取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能，但是这种实现不应认为超出本申请的范围。

[0273] 在本申请所提供的实施例中，应该理解到，所揭露的装置/网络设备和/或方法，可以通过其它的方式实现。例如，以上所描述的装置/网络设备实施例仅仅是示意性的，例如，所述模块或单元的划分，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式，例如多个单元或组件可以结合或者可以集成到另一个系统，或一些特征可以忽略，或不执行。另一点，所显示或讨论的相互之间的耦合或直接耦合或通讯连接可以是通过一些接口，装置或单元的间接耦合或通讯连接，可以是电性，机械或其它的形式。

[0274] 所述作为分离部件说明的单元可以是或者也可以不是物理上分开的，作为单元显示的部件可以是或者也可以不是物理单元，即可以位于一个地方，或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

[0275] 以上所述实施例仅用以说明本申请的技术方案，而非对其限制；尽管参照前述实施例对本申请进行了详细的说明，本领域的普通技术人员应当理解：其依然可以对前述各实施例所记载的技术方案进行修改，或者对其中部分技术特征进行等同替换；而这些修改或者替换，并不使相应技术方案的本质脱离本申请各实施例技术方案的精神和范围，均应包含在本申请的保护范围之内。

权利要求书

- [权利要求 1] 一种安全的后缀索引外包计算方法，其特征在于，包括：
- 获取待构造后缀数组的初始字符串和扩充因子，其中，所述扩充因子用于表征所述初始字符串中初始字符的扩充位数，扩充因子为大于1的自然数；
- 基于所述扩充因子扩充所述初始字符串，得到所述初始字符串的扩充数组；
- 基于所述扩充数组，确定所述初始字符串的目标字符串，并将所述目标字符串上传至云服务器；
- 获取所述云服务器生成的所述目标字符串的第一后缀数组，并基于所述第一后缀数组生成所述初始字符串的第二后缀数组。
- [权利要求 2] 如权利要求1所述的的安全的后缀索引外包计算方法，其特征在于，所述基于所述扩充因子扩充所述初始字符串，得到所述初始字符串的扩充数组，包括：
- 基于所述扩充因子，扩充所述初始字符串中的每个初始字符，得到每个初始字符的扩充字符串，并将所述扩充字符串组成所述扩充数组。
- [权利要求 3] 如权利要求2所述的的安全的后缀索引外包计算方法，其特征在于，所述基于所述扩充因子，扩充所述初始字符串中的每个初始字符，得到每个初始字符的扩充字符串，包括：
- 选取所述初始字符串中的第 i 个初始字符，其中， $i \geq 1$ ；
- 若 $i+k-1 \leq n$ ，则将所述初始字符串中的第 i 个初始字符至第 $i+k-1$ 个初始字符作为所述第 i 个初始字符的扩充字符串，其中， k 为扩充因子， n 为所述初始字符串中初始字符的总数， $n > 1$ ；
- 若 $i+k-1 > n$ ，则将所述初始字符串中的第 i 个初始字符至第 n 个初始字符作为所述第 i 个初始字符的扩充字符串。
- [权利要求 4] 如权利要求1所述的的安全的后缀索引外包计算方法，其特征在于，所述基于所述扩充数组，确定所述初始字符串的目标字符串，包括：
- 将所述初始字符串分成预设分块数量个字符串块；

基于所述字符串块中的字符，将所述扩充数组分成数组块；

基于所述字符串块和所述数组块，确定所述初始字符串的目标字符串

。

[权利要求 5] 如权利要求4所述的安全的后缀索引外包计算方法，其特征在于，所述基于所述字符串块中的字符，将所述扩充数组分成数组块，包括：将所述字符串块中的各个字符对应的扩充字符串组成一个数组块，得到所述扩充数据的各个数组块。

[权利要求 6] 如权利要求4所述的安全的后缀索引外包计算方法，其特征在于，所述基于所述字符串块和所述数组块，确定所述初始字符串的目标字符串，包括：

基于各个所述字符串块中最后一个字符在所述初始字符串中的第一下标，得到由第一下标组成的第一数组；

基于各个所述字符串块的块末后缀在所述初始字符串的第一后缀中的排名，得到由排名编号组成的第二数组；

基于所述数组块、所述扩充数组、所述第一数组和所述第二数组，得到第三数组；

基于所述扩充数组，确定第四数组；

基于所述第三数组和所述第四数组，确定第五数组；

基于各个所述字符串块中的字符数量，将所述第五数组分成数据块；将各个所述数据块重新排序，得到所述目标字符串。

[权利要求 7] 如权利要求6所述的安全的后缀索引外包计算方法，其特征在于，所述基于各个所述字符串块的块末后缀在所述初始字符串的第一后缀中的排名，得到由排名编号组成的第二数组，包括：

获取所述初始字符串中的第j个字符串块的块末后缀，其中，所述第j个字符串块的块末后缀由第j个字符串块中的最后一个字符至所述初始字符串的最后一个初始字符组成， $1 \leq j \leq m$ ，m为分块数量；

获取所述初始字符串的各个第一后缀，将各个所述第一后缀按照字典顺序排序，其中，所述初始字符串最后一个初始字符为一个第一后缀

，所述初始字符串中第 g 个初始字符至最后一个初始字符组成一个第一后缀， $1 \leq g < n$ ， n 为所述初始字符串中初始字符的总数；

对排序后的第一后缀进行编号，得到各个第一后缀的第一编号；

将每个字符串块的块末后缀对应的编号作为所述第二数组，其中，所述块末后缀对应的编号为与所述块末后缀相同的第一后缀的第一编号。

[权利要求 8]

如权利要求6所述的安全的后缀索引外包计算方法，其特征在于，所述基于所述数组块、所述扩充数组、所述第一数组和所述第二数组，得到第三数组，包括：

基于公式 $V[A[a]] = E[a]$ ，将所述第一数组中的数组元素放入所述第三数组中对应的第一位置，其中， a 为所述第一数组中数组元素的下标， $0 \leq a < m$ ， m 为分块数量， $E[a]$ 为第一数组中下标 a 对应的数组元素， $A[a]$ 为第二数组中的下标 a 对应的数组元素； $V[A[a]]$ 为第三数组中下标 $A[a]$ 对应的数组元素；

将目标扩充字符串按照字典顺序进行排序，将排序后的各个目标扩充字符串在扩充数组中对应的下标作为第六数组，其中，所述目标扩充字符串指所述扩充数组中不是块末字符串的字符串，所述块末字符串指数组块中的最后一个字符串；

基于公式 $V[B_b] = M[b]$ ，将所述第六数组中的数组元素放入所述第三数组中对应的第二位置，其中， $b \geq 1$ ， $M[b]$ 为所述第六数组中的第 b 个位置上的数组元素， $V[B_b]$

为第三数组中第二位置上的数组元素， B_b 为第 b 个第二位置，第二位置为所述第三数组中除第一位置以外的位置。

[权利要求 9]

如权利要求6所述的安全的后缀索引外包计算方法，其特征在于，所述基于所述扩充数组，确定第四数组，包括：

按照字典顺序对所述扩充数组中的扩充字符串进行排序，得到排序后的扩充字符串组成的目标数组；

确定各个所述字符串块中的块末字符串在所述目标数组中的位置，所

述块末字符串指数组块中的最后一个字符串；

对所述目标数组中的扩充字符串进行编号，得到所述第四数组。

[权利要求 10] 如权利要求9所述的安全的后缀索引外包计算方法，其特征在于，所述按照字典顺序对所述扩充数组中的扩充字符串进行排序，得到排序后的扩充字符串组成的目标数组，包括：

按照字典顺序对所述扩充数组中的扩充字符串进行排序，得到扩充字符串序列；

在所述扩充字符串序列中存在相同的扩充字符串，且相同的扩充字符串中至少一个是块末字符串时，基于所述初始字符串的后缀和所述块末字符串对应的初始字符所在字符串块的块末后缀，对所述相同的扩充字符串进行排序，得到所述目标数组。

[权利要求 11] 如权利要求9所述的安全的后缀索引外包计算方法，其特征在于，所述对所述目标数组中的扩充字符串进行编号，得到所述第四数组，包括：

依次对所述目标数组中的扩充字符串进行编号，其中，相同且相邻的非块末字符串具有相同的编号，相同且不相邻的非块末字符串具有不同的编号，块末字符串的编号不等于其他扩充字符串的编号。

[权利要求 12] 如权利要求6所述的安全的后缀索引外包计算方法，其特征在于，所述基于所述第三数组和所述第四数组，确定第五数组，包括：
基于公式 $Y[V[c]] = R[c]$ 确定第五数组，其中， $R[c]$ 为第四数组中下标 c 对应的数组元素， $V[c]$ 为第三数组中下标 c 对应的数组元素， $Y[V[c]]$ 为第五数组中下标 $V[c]$ 对应的字符， $0 \leq c < n$ ， n 为所述初始字符串中初始字符的总数。

[权利要求 13] 如权利要求6所述的安全的后缀索引外包计算方法，其特征在于，所述将各个所述数据块重新排序，得到所述目标字符串，包括：
基于第七数组将各个所述数据块重新排序，得到所述目标字符串，其中，第七数组是将所述第五数组中的数据块的下标重新排序后得到的。

- [权利要求 14] 如权利要求1所述的安全的后缀索引外包计算方法，其特征在于，所述
第一后缀数组的生成方法包括：
基于归纳排序方法生成所述目标字符串的第一后缀数组。
- [权利要求 15] 如权利要求6所述的安全的后缀索引外包计算方法，其特征在于，所
述基于所述第一后缀数组生成所述初始字符串的第二后缀数组，包括
：
基于所述目标字符串中各个所述数据块中字符的个数，确定由各个所
述数据块中字符的个数组成的第八数组；
基于所述第八数组，确定第九数组；
基于所述第九数组和所述第一后缀数组，得到所述第二后缀数组。
- [权利要求 16] 如权利要求15所述的安全的后缀索引外包计算方法，其特征在于，所
述基于所述第八数组，确定第九数组，包括：
将0作为所述第九数组中的第一个数组元素，将所述第八数组中的第
一个数组元素作为第九数组中的第二个数组元素；
计算所述第八数组中的第一个数组元素至第d个数组元素的和，其中
， $1 < d \leq m$ ，m为分块数量；
将所述第八数组中的第一个数组元素至第d个数组元素的和作为所述
第九数组中的第d+1个数组元素。
- [权利要求 17] 如权利要求15所述的安全的后缀索引外包计算方法，其特征在于，所
述基于所述第九数组和所述第一后缀数组，得到所述第二后缀数组，
包括：
基于公式 $SA_s[e] = E[P[f]] - W[f] + 1 + SA_x[e] - H[f]$ ，得到所述第二后
缀数组，其中，f基于公式 $H[f] \leq SA_x[e] < H[f+1]$ 确定， $SA_s[e]$ 为第二
后缀数组中下标e对应的数组元素， $E[P[f]]$ 为第一数组中下标P[f]对
应的数组元素，P[f]为第七数组中下标f对应的数组元素，W[f]为第
八数组中下标f对应的数组元素， $SA_x[e]$ 为第一后缀数组中下标e对
应的数组元素，H[f]为第九数组中下标f对应的数组元素，H[f+1]为
第九数组中下标f+1对应的数组元素。

- [权利要求 18] 一种安全的后缀索引外包计算装置，其特征在于，包括：
- 数据获取模块，用于获取待构造后缀数组的初始字符串和扩充因子，其中，所述扩充因子用于表征所述初始字符串中初始字符的扩充位数，扩充因子为大于1的自然数；
- 数据扩充模块，用于基于所述扩充因子扩充所述初始字符串，得到所述初始字符串的扩充数组；
- 字符串生成模块，用于基于所述扩充数组，确定所述初始字符串的目标字符串，并将所述目标字符串上传至云服务器；
- 后缀数组生成模块，用于获取所述云服务器生成的所述目标字符串的第一后缀数组，并基于所述第一后缀数组生成所述初始字符串的第二后缀数组。
- [权利要求 19] 一种终端设备，包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机程序，其特征在于，所述处理器执行所述计算机程序时实现如权利要求1至17任一项所述的安全的后缀索引外包计算方法。
- [权利要求 20] 一种计算机可读存储介质，所述计算机可读存储介质存储有计算机程序，其特征在于，所述计算机程序被处理器执行时实现如权利要求1至17任一项所述的安全的后缀索引外包计算方法。

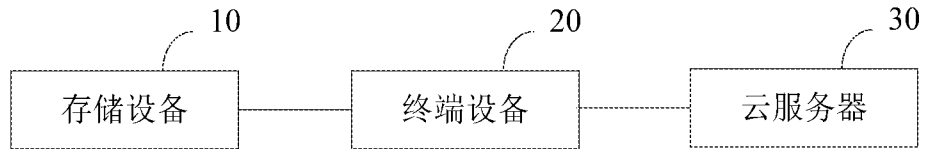


图 1

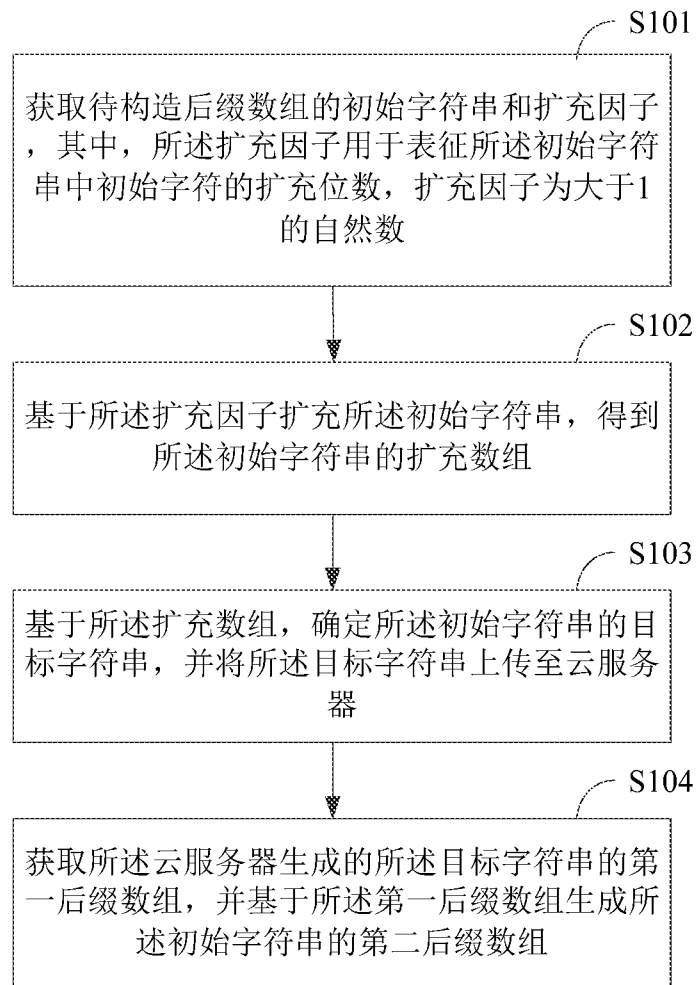


图 2

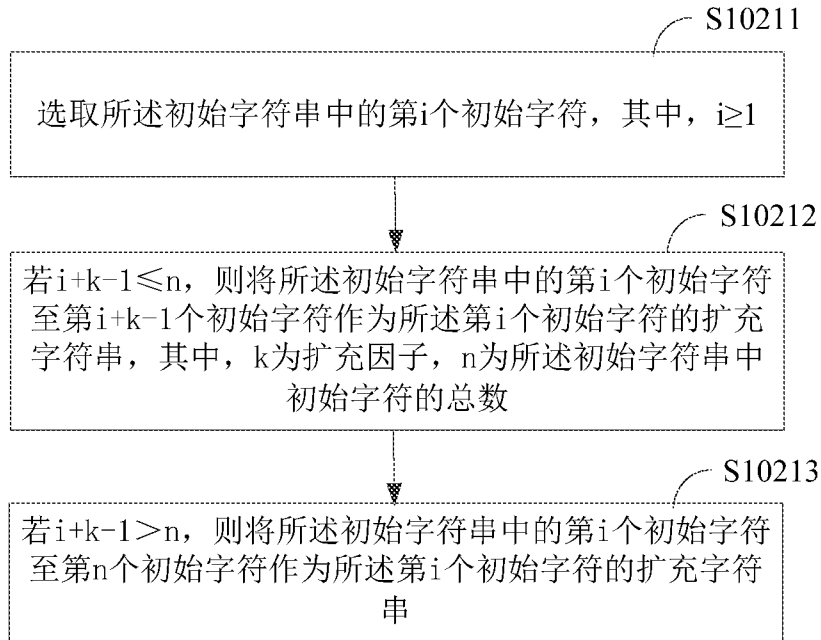


图 3

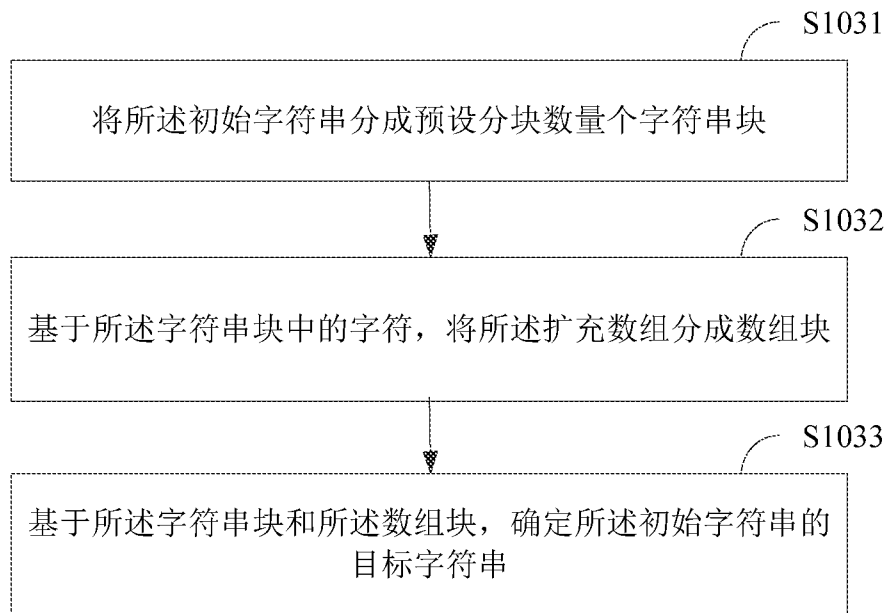


图 4

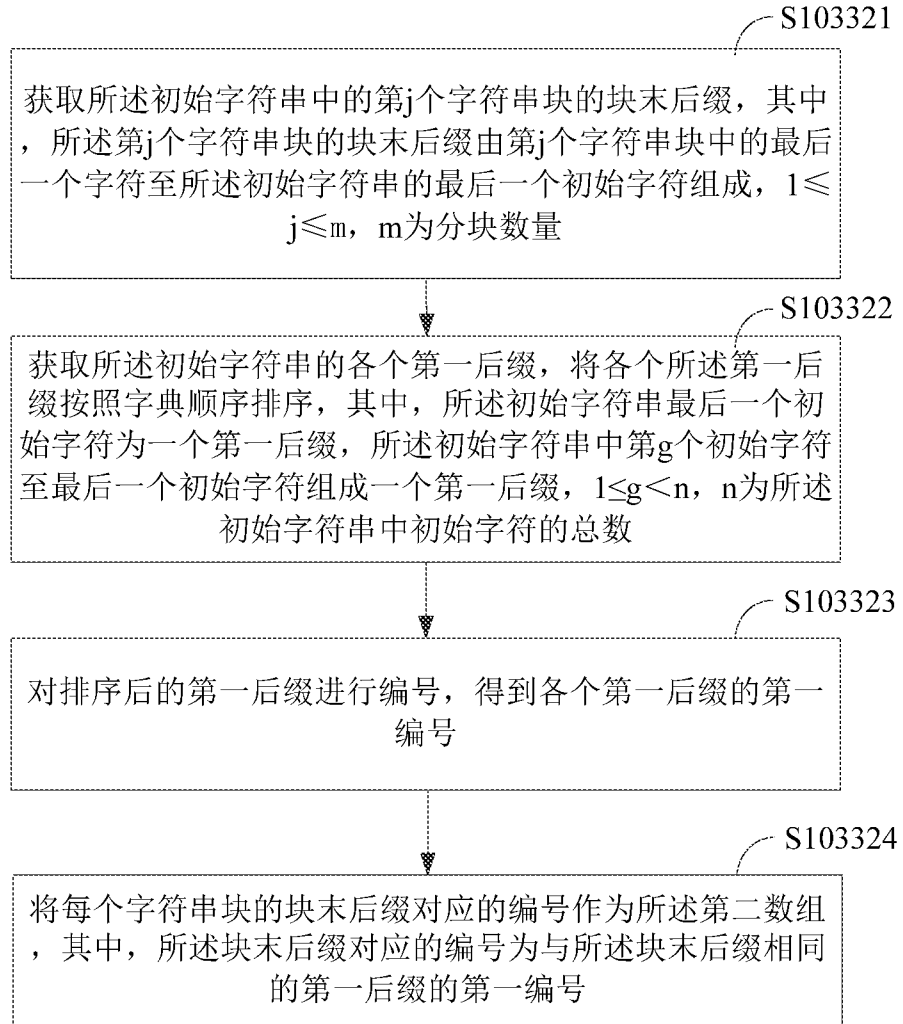


图 5

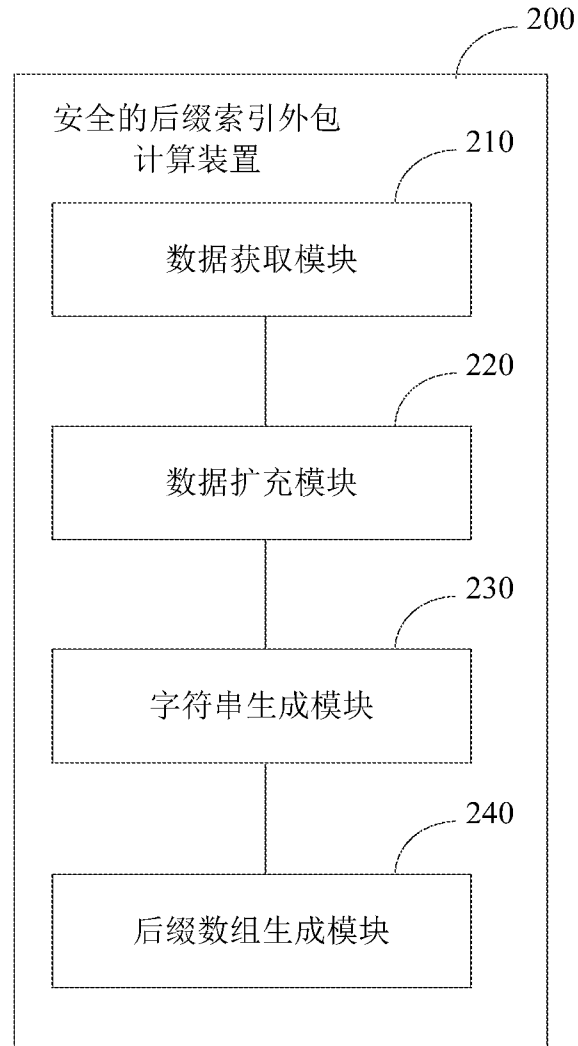


图 6

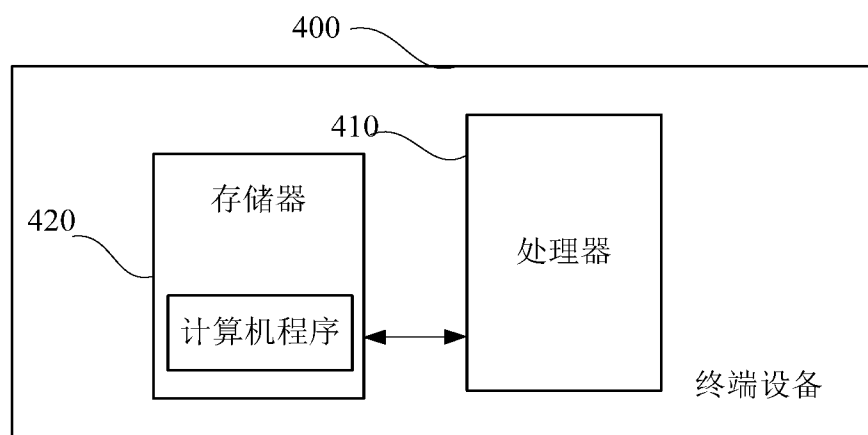


图 7

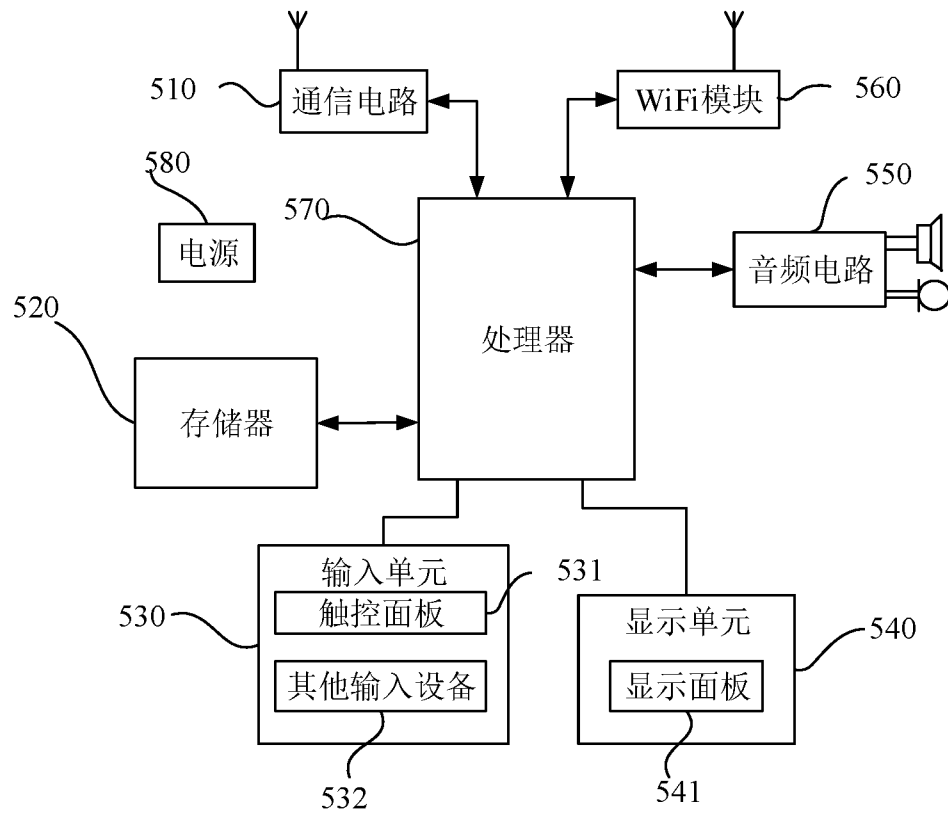


图 8

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2020/103065

A. CLASSIFICATION OF SUBJECT MATTER

G06F 16/182(2019.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

WPI, EPODOC, CNPAT, CNKI, IEEE: 后缀数组, 后缀索引, 数据结构, 全文索引, 构造, 字符串, 安全, 泄露, 扩充, 增加, 位数, 云服务器, 代替, 上传, 传输, 分块, 排序, suffix array, SA, suffix index, data structure, full-text index, construct, string, security, reveal, expand, add, digit, cloud server, replace, upload, block, sort

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 109284273 A (SUN YAT-SEN UNIVERSITY) 29 January 2019 (2019-01-29) claims 1-10, description paragraphs [0051]-[0088], figures 1-4	1-20
A	CN 103810228 A (NVIDIA CORP.) 21 May 2014 (2014-05-21) entire document	1-20
A	CN 109299152 A (SUN YAT-SEN UNIVERSITY) 01 February 2019 (2019-02-01) entire document	1-20
A	US 2014350917 A1 (XEROX CORPORATION) 27 November 2014 (2014-11-27) entire document	1-20
A	GE, Nong et al. "Linear Suffix Array Construction by Almost Pure Induced-Sorting" 2009 Data Compression Conference, 26 May 2009 (2009-05-26), entire document	1-20



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance
 “E” earlier application or patent but published on or after the international filing date
 “L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
 “O” document referring to an oral disclosure, use, exhibition or other means
 “P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
 “X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
 “Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
 “&” document member of the same patent family

Date of the actual completion of the international search

07 April 2021

Date of mailing of the international search report

20 April 2021

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2020/103065

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	109284273	A	29 January 2019	None			
CN	103810228	A	21 May 2014	DE	102013218594	A1	08 May 2014
				US	2014123147	A1	01 May 2014
				TW	201439965	A	16 October 2014
CN	109299152	A	01 February 2019	None			
US	2014350917	A1	27 November 2014	None			

国际检索报告

国际申请号

PCT/CN2020/103065

A. 主题的分类 G06F 16/182(2019.01) i 按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类																				
B. 检索领域 检索的最低限度文献(标明分类系统和分类号) G06F 包含在检索领域中的除最低限度文献以外的检索文献 在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用)) WPI, EPDOC, CNPAT, CNKI, IEEE: 后缀数组, 后缀索引, 数据结构, 全文索引, 构造, 字符串, 安全, 泄露, 扩充, 增加, 位数, 云服务器, 代替, 上传, 传输, 分块, 排序, suffix array, SA, suffix index, data structure, full-text index, construct, string, security, reveal, expand, add, digit, cloud server, replace, upload, block, sort																				
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>CN 109284273 A (中山大学) 2019年 1月 29日 (2019 - 01 - 29) 权利要求1-10, 说明书第[0051]-[0088]段, 图1-4</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 103810228 A (辉达公司) 2014年 5月 21日 (2014 - 05 - 21) 全文</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 109299152 A (中山大学) 2019年 2月 1日 (2019 - 02 - 01) 全文</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>US 2014350917 A1 (XEROX CORPORATION) 2014年 11月 27日 (2014 - 11 - 27) 全文</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>GE, Nong 等. "Linear Suffix Array Construction by Almost Pure Induced-Sorting" 2009 Data Compression Conference, 2009年 5月 26日 (2009 - 05 - 26), 全文</td> <td>1-20</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	A	CN 109284273 A (中山大学) 2019年 1月 29日 (2019 - 01 - 29) 权利要求1-10, 说明书第[0051]-[0088]段, 图1-4	1-20	A	CN 103810228 A (辉达公司) 2014年 5月 21日 (2014 - 05 - 21) 全文	1-20	A	CN 109299152 A (中山大学) 2019年 2月 1日 (2019 - 02 - 01) 全文	1-20	A	US 2014350917 A1 (XEROX CORPORATION) 2014年 11月 27日 (2014 - 11 - 27) 全文	1-20	A	GE, Nong 等. "Linear Suffix Array Construction by Almost Pure Induced-Sorting" 2009 Data Compression Conference, 2009年 5月 26日 (2009 - 05 - 26), 全文	1-20
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求																		
A	CN 109284273 A (中山大学) 2019年 1月 29日 (2019 - 01 - 29) 权利要求1-10, 说明书第[0051]-[0088]段, 图1-4	1-20																		
A	CN 103810228 A (辉达公司) 2014年 5月 21日 (2014 - 05 - 21) 全文	1-20																		
A	CN 109299152 A (中山大学) 2019年 2月 1日 (2019 - 02 - 01) 全文	1-20																		
A	US 2014350917 A1 (XEROX CORPORATION) 2014年 11月 27日 (2014 - 11 - 27) 全文	1-20																		
A	GE, Nong 等. "Linear Suffix Array Construction by Almost Pure Induced-Sorting" 2009 Data Compression Conference, 2009年 5月 26日 (2009 - 05 - 26), 全文	1-20																		
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。																				
<table border="0"> <tr> <td> * 引用文件的具体类型: "A" 认为不特别相关的表示了现有技术一般状态的文件 "E" 在国际申请日的当天或之后公布的在先申请或专利 "L" 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的) "O" 涉及口头公开、使用、展览或其他方式公开的文件 "P" 公布日先于国际申请日但迟于所要求的优先权日的文件 </td> <td> "T" 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 "X" 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 "Y" 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 "&" 同族专利的文件 </td> </tr> </table>			* 引用文件的具体类型: "A" 认为不特别相关的表示了现有技术一般状态的文件 "E" 在国际申请日的当天或之后公布的在先申请或专利 "L" 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的) "O" 涉及口头公开、使用、展览或其他方式公开的文件 "P" 公布日先于国际申请日但迟于所要求的优先权日的文件	"T" 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 "X" 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 "Y" 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 "&" 同族专利的文件																
* 引用文件的具体类型: "A" 认为不特别相关的表示了现有技术一般状态的文件 "E" 在国际申请日的当天或之后公布的在先申请或专利 "L" 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的) "O" 涉及口头公开、使用、展览或其他方式公开的文件 "P" 公布日先于国际申请日但迟于所要求的优先权日的文件	"T" 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 "X" 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 "Y" 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 "&" 同族专利的文件																			
国际检索实际完成的日期 2021年 4月 7日		国际检索报告邮寄日期 2021年 4月 20日																		
ISA/CN的名称和邮寄地址 中国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451		受权官员 王洋 电话号码 (86-10)53961523																		

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2020/103065

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	109284273	A	2019年 1月 29日	无			
CN	103810228	A	2014年 5月 21日	DE	102013218594	A1	2014年 5月 8日
				US	2014123147	A1	2014年 5月 1日
				TW	201439965	A	2014年 10月 16日
CN	109299152	A	2019年 2月 1日	无			
US	2014350917	A1	2014年 11月 27日	无			