

# (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  
2019年8月8日 (08.08.2019)



(10) 国际公布号  
**WO 2019/148497 A1**

- (51) 国际专利分类号:  
**G06F 17/30** (2006.01)
- (21) 国际申请号: PCT/CN2018/075300
- (22) 国际申请日: 2018年2月5日 (05.02.2018)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (71) 申请人: 华为技术有限公司 (HUAWEI TECHNOLOGIES CO., LTD.) [CN/CN]; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。
- (72) 发明人: 高翔 (GAO, Xiang); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 杜维 (DU, Wei); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 陈俊彦 (CHEN, Chun Yen); 中国

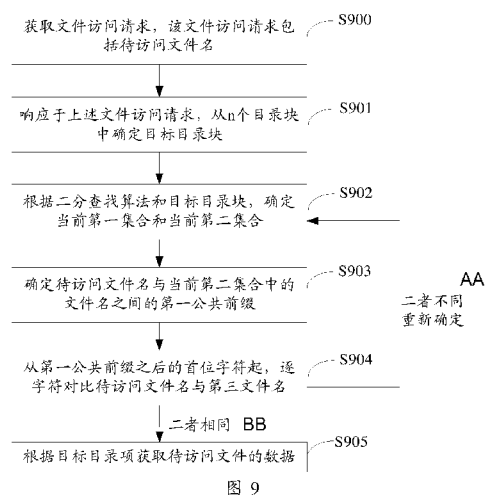
广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 汪宁 (WANG, Ning); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。

(74) 代理人: 北京中博世达专利商标代理有限公司 (BEIJING ZBSD PATENT & TRADEMARK AGENT LTD.); 中国北京市海淀区交大东路31号11号楼8层, Beijing 100044 (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, 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, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,

(54) Title: DATE QUERY METHOD AND DEVICE

(54) 发明名称: 一种数据查询方法及装置



- S900 Acquire a file access request, the file access request comprising a filename to be accessed
- S901 Respond to the file access request and determine a target directory block from among n directory blocks
- S902 Determine a current first set and current second set according to a binary search algorithm and the target directory block
- S903 Determine a first common prefix between the filename to be accessed and filenames in the current second set
- S904 Compare the filename to be accessed with a third filename character by character starting from the first character after the first common prefix
- S905 Acquire data of the file to be accessed according to a target directory entry
- AA The two are not the same, determine again
- BB The two are the same

图 9

(57) Abstract: A data query method and device which relate to the technical field of computers and solve the problem of low efficiency when querying a filename to be accessed. The method comprises: determining a target directory block comprising m directory entries and m filenames, wherein the m directory entries and the m filenames are in one-to-one correspondence and are sorted according to a preset rule; determining a current first set and current second set according to a binary search algorithm and the target directory block, wherein the current first set comprises x continuous filenames in the m filenames, and the current second set comprises x filenames, a first filename and a second filename; and  $m \geq x \geq 1$ ; determining a first common prefix between a filename to be accessed and the filenames in the current second set; comparing the filename to be accessed with a third filename character by character starting from the first character after the first common prefix; if the filename to be accessed is the same as the third filename, then acquiring data of the file to be accessed according to the directory entry corresponding to the third filename.



WO 2019/148497 A1

SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG,  
US, UZ, VC, VN, 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))。

---

(57) 摘要: 一种数据查询方法及装置, 涉及计算机技术领域, 解决了查询待访问文件名效率较低的问题。该方法包括: 确定包括 $m$ 个目录项和 $m$ 个文件名的目标目录块,  $m$ 个目录项和 $m$ 个文件名一一对应且按照预设规则顺序排列; 根据二分查找算法和目标目录块, 确定当前第一集合和当前第二集合, 当前第一集合包括 $m$ 个文件名中连续的 $x$ 个文件名, 当前第二集合包括 $x$ 个文件名、第一文件名以及第二文件名,  $m \geq x \geq 1$ ; 确定待访问文件名与当前第二集合中的文件名之间的第一公共前缀; 从第一公共前缀之后的首位字符起, 逐字符对比待访问文件名与第三文件名; 若待访问文件名与第三文件名相同, 则根据与第三文件名对应的目录项获取待访问文件的数据。

## 一种数据查询方法及装置

### 技术领域

5 本申请实施例涉及计算机技术领域，尤其涉及一种数据查询方法及装置。

### 背景技术

10 文件系统通常采用树形结构组织文件的拓扑关系，这里，文件仅涉及目录文件和普通文件。一般的，树形结构中的叶节点表示普通文件，除叶节点之外的其他节点表示目录文件。目录文件包括多个目录项，每个目录项均包括文件名、文件类型以及索引（inode）号，计算机根据索引号所标识的索引能够获取到文件的数据。计算机在获取某一文件（以文件 A 为例）时，需要在目录文件中查询文件名，以获取与文件 A 的文件名对应的目录项，进而根据获取到的目录项获取文件 A 的数据。

15 目前，一些文件系统（如闪存友好文件系统（Flash Friendly File System, F2FS）、第四代扩展文件系统（Fourth Extended File System, EXT4 等）中目录的结构为包括多级哈希表的哈希（hash）树，每一级哈希表均包括多个哈希值、与每个哈希值对应的文件名以及文件的索引号；另一些文件系统（如新技术文件系统（New technology file system, Ntfs）、B 树文件系统（B-tree file system, Btrfs）中目录的结构为  $n$  ( $n \geq 1$ ) 阶 B+树。

20 在文件系统只读的应用场景中：若该文件系统中目录的结构为包括多级哈希表的哈希树，计算机在获取待访问文件的数据时，需逐级查询哈希表，且在查询的每一哈希表中均先遍历哈希值再匹配文件名，在哈希表的级别较高的情况下，在哈希表中查询待访问文件名的效率较低；此外，哈希树包括有大量的哈希值，导致存储空间的有效利用率较低，而且在哈希表的级别较高的情况下，可能存在高级别哈希表未写满的情况，进一步降低了存储空间的有效利用率；若该文件系统中目录的结构为  $n$  阶 B+树，计算机在获取待访问文件的数据时，该计算机从最小关键字开始顺序查询，或者从根节点开始随机查询，查询待访问文件名的效率比较低；此外， $n$  阶 B+树的叶子节点中的关键字出现在中间节点中，降低了存储空间的有效利用率。

25 综上，在文件系统只读的应用场景中，基于现有的目录结构，计算机查询待访问文件名的效率较低，且计算机的存储空间有效利用率较低。

### 30 发明内容

本申请实施例提供一种数据查询方法及装置，能够解决计算机查询待访问文件名的效率较低，且计算机的存储空间有效利用率较低的问题。

为达到上述目的，本申请实施例采用如下技术方案：

35 第一方面，提供一种数据查询方法，该数据查询方法应用于包括  $n$  ( $n \geq 1$ ) 个目录块的只读文件系统中，每个目录块包括目录项区域和文件名区域。具体的，该数据查询方法为：数据查询装置从上述  $n$  个目录块中确定目标目录块，该目标目录块的目录项区域包括  $m$  个目录项，目标目录块的文件名区域包括  $m$  个文件名， $m$  个目录项与  $m$  个文件名一一对应， $m$  个目录项和  $m$  个文件名均按照预设规则顺序排列，待访问文

文件名位于文件名范围中，文件名范围是由目标目录块的首个文件名与目标目录块的最后一个文件名组成的范围， $m \geq 1$ ；在确定出目标目录块后，数据查询装置依次执行步骤 A1、步骤 B1 以及步骤 C1，这里的步骤 A1 为：根据二分查找算法和目标目录块，确定当前第一集合和当前第二集合，当前第一集合包括  $m$  个文件名中连续的  $x$  个文件名，当前第二集合包括  $x$  ( $m \geq x \geq 1$ ) 个文件名、第一文件名以及第二文件名，第一文件名为排列于  $x$  个文件名中的首个文件名之前且与  $x$  个文件名中的首个文件名相邻的文件名，第二文件名为排列于  $x$  个文件名中的最后一个文件名之后且与  $x$  个文件名中的最后一个文件名相邻的文件名；步骤 B1 为：确定待访问文件名与当前第二集合中的文件名之间的第一公共前缀；步骤 C1 为：从第一公共前缀之后的首位字符起，逐字符对比待访问文件名与第三文件名，该第三文件名为当前第一集合中第一预设位置的文件名；在执行完步骤 C1 之后，若待访问文件名与第三文件名相同，则数据查询装置根据与第三文件名对应的目录项获取待访问文件的数据。

本申请中的数据查询装置确定了待访问文件名与当前第二集合中的文件名之间的第一公共前缀，由于当前第二集合覆盖第一集合，因此，当前第一集合中的所有文件名与待访问文件名之间也存在第一公共前缀，这样，数据查询装置直接从第一公共前缀之后的首位字符起，逐字符对比待访问文件名与第三文件名即可，有效的提高了查询待访问文件的速率。

此外，与现有的目录树结构相比，本申请实施例中的目录结构仅存储有目录项与文件名，并未存储与文件名或目录项相关的其他信息，有效的提高了存储空间利用率。

可选的，在本申请的一种可能的实现方式中，若待访问文件名与第三文件名不同，则数据查询装置根据二分查找算法、当前第一集合和当前第二集合，重新确定第一集合和第二集合，并根据重新确定的第一集合和重新确定的第二集合，执行上述步骤 B1 和上述步骤 C1，直到获取到待访问文件的数据或确定目标目录块未包括待访问文件名。

结合二分查找算法的定义可知，若待访问文件名与第三文件名不同，则数据查询装置进一步缩小查询范围（即重新确定第一集合），并重新确定第一公共前缀。

可选的，在本申请的另一种可能的实现方式中，上述预设规则为字典序顺序，这样，上述“若待访问文件名与第三文件名不同，则根据二分查找算法、当前第一集合和当前第二集合，重新确定第一集合和第二集合”的方法为：若待访问文件名的特征值小于第三文件名的特征值，确定重新确定的第一集合包括当前第一集合中位于第三文件名之前的所有文件名，重新确定的第二集合包括第一文件名、当前第一集合中位于第三文件名之前的所有文件名以及第三文件名；或者，若待访问文件名的特征值大于第三文件名的特征值，确定重新确定的第一集合包括当前第一集合中位于第三文件名之后的所有文件名，重新确定的第二集合包括第三文件名、当前第一集合中位于第三文件名之后的所有文件名以及第二文件名。

可选的，在本申请的另一种可能的实现方式中，上述“确定待访问文件名与当前第二集合中的文件名之间的第一公共前缀”的方法为：确定待访问文件名与第一文件名之间共有的第一前缀；确定待访问文件名与第二文件名之间共有的第二前缀；将第一前缀与第二前缀中长度最小的一个确定为第一公共前缀。

本申请实施例中当前第二集合中的所有文件名组成的文件名范围大于当前第一集合中的所有文件名组成的文件名范围，由于每个目录块中的所有文件名均按照预设规则顺序排列，因此，数据查询装置通过确定待访问文件名与第一文件名之间共有的第一前缀，以及确定待访问文件名与第二文件名之间共有的第二前缀，可将第一前缀与第二前缀中长度最小的一个确定为第一公共前缀。

可选的，在本申请的另一种可能的实现方式中，上述“数据查询装置从  $n$  个目录块中确定目标目录块”的方法为：数据查询装置依次执行步骤 A2、B2、C2；其中，步骤 A2 为：根据二分查找算法和  $n$  个目录块，确定当前第三集合和当前第四集合；其中，当前第三集合包括  $p$  个文件名， $p$  个文件名包括  $p$  个目录块中每个目录块中第二预设位置的文件名，且当前第三集合中的文件名按照预设规则顺序排列， $p$  个目录块为  $n$  个目录块中连续的目录块，当前第四集合包括  $p$  个文件名、第四文件名以及第五文件名，第四文件名为排列于  $p$  个文件名中的首个文件名之前且与  $p$  个文件名中的首个文件名相邻的文件名，第五文件名为排列于  $p$  个文件名中的最后一个文件名之后且与  $p$  个文件名中的最后一个文件名相邻的文件名， $1 \leq p \leq n$ ；步骤 B2 为：确定待访问文件名与当前第四集合中的文件名之间的第二公共前缀；步骤 C2 为：从第二公共前缀之后的首位字符起，逐字符对比待访问文件名与第六文件名；其中，第六文件名为当前第三集合中第三预设位置的文件名；这样，在执行步骤 C2 之后，若待访问文件名与第六文件名相同，则数据查询装置确定第六文件名归属的目录块为目标目录块。

当  $n$  个目录块的存储方式为块间顺序存储时，数据查询装置确定目标目录块的方法与该数据查询装置查询待访问文件名的方法类似。

可选的，在本申请的另一种可能的实现方式中，在待访问文件名与第六文件名不同的情况下，当  $2 \leq p \leq n$  时，数据查询装置根据当前第三集合、当前第四集合和二分查找算法，重新确定第三集合和第四集合，并根据重新确定的第三集合和重新确定的第四集合，执行上述步骤 B2 和上述步骤 C2；当  $p=1$  时，数据查询装置根据当前第三集合包括的文件名确定目标目录块。

可选的，在本申请的另一种可能的实现方式中，在上述预设规则为字典序顺序，对于每个目录块，第二预设位置的文件名为该目录块的首个文件名的情况下，上述“当  $p=1$  时，数据查询装置根据当前第三集合包括的文件名确定目标目录块”的方法为：若待访问文件名的特征值大于当前第三集合中的文件名的特征值，则数据查询装置将当前第三集合中的文件名归属的目录块确定为目标目录块；或者，若待访问文件名的特征值小于当前第三集合中的文件名的特征值，则数据查询装置将位于当前第三集合中的文件名之前且与当前第三集合中的文件名相邻的文件名归属的目录块确定为目标目录块。

可选的，在本申请的另一种可能的实现方式中，在预设规则为字典序顺序的情况下，上述“根据当前第三集合、当前第四集合和二分查找算法，重新确定第三集合和第四集合”的方法为：若待访问文件名的特征值小于第六文件名的特征值，确定重新确定的第三集合包括当前第三集合中位于第六文件名之前的所有文件名，重新确定的第四集合包括第四文件名、当前第三集合中位于第六文件名之前的所有文件名以及第六文件名；或者，若待访问文件名的特征值大于第六文件名的特征值，确定重新确定

的第三集合包括当前第三集合中位于第六文件名之后的所有文件名，重新确定的第四集合包括第六文件名、当前第三集合中位于第六文件名之后的所有文件名以及第五文件名。

5 可选的，在本申请的另一种可能的实现方式中，在上述预设规则为字典序顺序，对于每个目录块，第二预设位置的文件名为该目录块的首个文件名的情况下，若第六文件名为当前第三集合中的首个文件名，待访问文件名的特征值小于第六文件名的特征值，则将位于第六文件名之前且与第六文件名相邻的文件名归属的目录块确定为目标目录块；若第六文件名为当前第三集合中的最后一个文件名，待访问文件名的特征值大于第六文件名的特征值，则将第六文件名归属的目录块确定为目标目录块。

10 可选的，在本申请的另一种可能的实现方式中，上述“确定待访问文件名与当前第四集合中的文件名之间的第二公共前缀”的方法为：确定待访问文件名与第四文件名之间共有的第三前缀；确定待访问文件名与第五文件名之间共有的第四前缀；将第三前缀与第四前缀中长度最小的一个确定为第二公共前缀。

15 可选的，在本申请的另一种可能的实现方式中，在上述  $n$  个目录块按照预设规则顺序排列，并采用完全二叉树方式存储的场景中，上述“数据查询装置从  $n$  个目录块中确定目标目录块”的方法为：数据查询装置依次执行步骤 A3 和步骤 B3，其中步骤 A3 为：确定当前候选目录块和当前第三公共前缀；步骤 B3 为：从当前第三公共前缀之后的首位字符起，逐字符对比待访问文件名与第  $i$  个文件名，第  $i$  个文件名为  $n$  个目录块中第  $i$  个目录块中第四预设位置的文件名， $0 \leq i < n$ ；这样，在执行步骤 B3 后，若待访问文件名与第  $i$  个文件名相同，则数据查询装置重新确定候选目录块为第  $i$  个文件名归属的目录块，并将重新确定的候选目录块确定为目标目录块。

20 可选的，在本申请的另一种可能的实现方式中，若待访问文件名与第  $i$  个文件名不同，则数据查询装置重新确定第三公共前缀、候选目录块以及第  $i$  个文件名，重新确定的第  $i$  个文件名为  $n$  个目录块中第  $j$  个目录块中第四预设位置的文件名；其中，若待访问文件名的特征值大于第  $i$  个文件名的特征值， $j=2i+2$ ；若待访问文件名的特征值小于第  $i$  个文件名的特征值， $j=2i+1$ ， $0 \leq i < j < n$ ；数据查询装置根据重新确定的第三公共前缀、重新确定的候选目录块以及重新确定的第  $i$  个文件名，重新执行步骤 B3，直到确定出目标目录块。

25 可选的，在本申请的另一种可能的实现方式中，在上述预设规则顺序为字典序，第四预设位置的文件名为对应目录块中的首个文件名的情况下，上述“若待访问文件名与第  $i$  个文件名不同，则重新确定候选目录块”的方法为：若待访问文件名的特征值大于第  $i$  个文件名的特征值，则确定重新确定的候选目录块为第  $i$  个文件名归属的目录块；若待访问文件名的特征值小于第  $i$  个文件名的特征值，则确定重新确定的候选目录块为当前候选目录块。

30 可选的，在本申请的另一种可能的实现方式中，在上述预设规则顺序为字典序，第四预设位置的文件名为对应目录块中的首个文件名的情况下，上述“若待访问文件名与第  $i$  个文件名不同，则重新确定第三公共前缀”的方法为：当待访问文件名的特征值大于第  $i$  个文件名的特征值时，将当前第一目标前缀更新为待访问文件名与第  $i$  个文件名之间共有的前缀；将更新后的第一目标前缀与当前第二目标前缀中长度最小

的一个确定为重新确定的第三公共前缀；或者，当待访问文件名的特征值小于第  $i$  个文件名的特征值时，将当前第二目标前缀更新为待访问文件名与第  $i$  个文件名之间共有的前缀；将当前第一目标前缀与更新后的第二目标前缀中长度最小的一个确定为重新确定的第三公共前缀；其中，第一目标前缀的长度和第二目标前缀的长度的初始值均为零，且第一目标前缀的长度和第二目标前缀的长度随着待访问文件名的特征值与第  $i$  个文件名的特征值的大小关系发生变化。

第二方面，提供一种只读文件系统，该只读文件系统的对象包括目录文件，目录文件由  $n$  个目录块组成，每个目录块均包括目录项区域和文件名区域，目录项区域包括至少一个目录项，文件名区域包括至少一个文件名。对应同一目录块而言，该目录块中目录项的数量与文件名的数量相同，且该目录块中的所有目录项以及所有文件名均按照预设规则顺序排列。

可选的，在本申请的一种可能的实现方式中，上述至少一个目录项中的每个目录项均包括索引号、文件类型和与该目录项对应的文件名在所归属的目录块的偏移量；上述文件名区域与目录项区域相邻，且文件名区域位于目录项区域之后。

第三方面，提供一种数据查询装置，该数据查询装置具备如上述第二方面及其任意一种可能的实现方式所述的只读文件系统。该数据查询装置包括处理单元和获取单元。

具体的，上述处理单元，用于从只读文件系统的  $n$  个目录块中确定目标目录块，目标目录块的目录项区域包括  $m$  个目录项，目标目录块的文件名区域包括  $m$  个文件名， $m$  个目录项与  $m$  个文件名一一对应， $m$  个目录项和  $m$  个文件名均按照预设规则顺序排列，待访问文件名位于文件名范围中，文件名范围是由目标目录块的首个文件名与目标目录块的最后一个文件名组成的范围， $m \geq 1$ 。上述处理单元，还用于执行步骤 A1、步骤 B1 以及步骤 C1；其中，步骤 A1 为：根据二分查找算法和目标目录块，确定当前第一集合和当前第二集合，当前第一集合包括  $m$  个文件名中连续的  $x$  个文件名，当前第二集合包括  $x$  个文件名、第一文件名以及第二文件名，第一文件名为排列于  $x$  个文件名中的首个文件名之前且与  $x$  个文件名中的首个文件名相邻的文件名，第二文件名为排列于  $x$  个文件名中的最后一个文件名之后且与  $x$  个文件名中的最后一个文件名相邻的文件名， $m \geq x \geq 1$ ；步骤 B1 为：确定待访问文件名与当前第二集合中的文件名之间的第一公共前缀；步骤 C1 为：从第一公共前缀之后的首位字符起，逐字符对比待访问文件名与第三文件名；其中，第三文件名为当前第一集合中第一预设位置的文件名。上述获取单元，用于若上述处理单元判断出待访问文件名与第三文件名相同，则根据与第三文件名对应的目录项获取待访问文件的数据。

可选的，在本申请的一种可能的实现方式中，上述处理单元，还用于若待访问文件名与第三文件名不同，则根据二分查找算法、当前第一集合和当前第二集合，重新确定第一集合和第二集合，并根据重新确定的第一集合和重新确定的第二集合，执行步骤 B1 和步骤 C1，直到上述获取单元获取到待访问文件的数据或处理单元确定目标目录块未包括待访问文件名。

可选的，在本申请的另一种可能的实现方式中，上述预设规则为字典序顺序，上述处理单元具体用于：若待访问文件名的特征值小于第三文件名的特征值，确定重新

确定的第一集合包括当前第一集合中位于第三文件名之前的所有文件名，重新确定的第二集合包括第一文件名、当前第一集合中位于第三文件名之前的所有文件名以及第三文件名；或者，若待访问文件名的特征值大于第三文件名的特征值，确定重新确定的第一集合包括当前第一集合中位于第三文件名之后的所有文件名，重新确定的第二集合包括第三文件名、当前第一集合中位于第三文件名之后的所有文件名以及第二文件名。

可选的，在本申请的另一种可能的实现方式中，上述处理单元具体用于：确定待访问文件名与第一文件名之间共有的第一前缀；确定待访问文件名与第二文件名之间共有的第二前缀；将第一前缀与第二前缀中长度最小的一个确定为第一公共前缀。

可选的，在本申请的另一种可能的实现方式中，上述处理单元还用于执行步骤 A2、步骤 B2 以及步骤 C2；其中，步骤 A2 为：根据二分查找算法和  $n$  个目录块，确定当前第三集合和当前第四集合；其中，当前第三集合包括  $p$  个文件名， $p$  个文件名包括  $p$  个目录块中每个目录块中第二预设位置的文件名，且当前第三集合中的文件名按照预设规则顺序排列， $p$  个目录块为  $n$  个目录块中连续的目录块，当前第四集合包括  $p$  个文件名、第四文件名以及第五文件名，第四文件名为排列于  $p$  个文件名中的首个文件名之前且与  $p$  个文件名中的首个文件名相邻的文件名，第五文件名为排列于  $p$  个文件名中的最后一个文件名之后且与  $p$  个文件名中的最后一个文件名相邻的文件名， $1 \leq p \leq n$ ；步骤 B2 为：确定待访问文件名与当前第四集合中的文件名之间的第二公共前缀；步骤 C2 为：从第二公共前缀之后的首位字符起，逐字符对比待访问文件名与第六文件名；其中，第六文件名为当前第三集合中第三预设位置的文件名。上述处理单元还用于若待访问文件名与第六文件名相同，则确定第六文件名归属的目录块为目标目录块。

可选的，在本申请的另一种可能的实现方式中，上述处理单元具体用于：当  $2 \leq p \leq n$  时，根据当前第三集合、当前第四集合和二分查找算法，重新确定第三集合和第四集合，并根据重新确定的第三集合和重新确定的第四集合，执行步骤 B2 和步骤 C2；当  $p=1$  时，根据当前第三集合包括的文件名确定目标目录块。

可选的，在本申请的另一种可能的实现方式中，上述预设规则为字典序顺序，对于每个目录块，第二预设位置的文件名为该目录块的首个文件名；当  $p=1$  时，上述处理单元具体用于：若待访问文件名的特征值大于当前第三集合中的文件名的特征值，则将当前第三集合中的文件名归属的目录块确定为目标目录块；或者，若待访问文件名的特征值小于当前第三集合中的文件名的特征值，则将位于当前第三集合中的文件名之前且与当前第三集合中的文件名相邻的文件名归属的目录块确定为目标目录块。

可选的，在本申请的另一种可能的实现方式中，上述预设规则为字典序顺序；上述处理单元具体用于：若待访问文件名的特征值小于第六文件名的特征值，确定重新确定的第三集合包括当前第三集合中位于第六文件名之前的所有文件名，重新确定的第四集合包括第四文件名、当前第三集合中位于第六文件名之前的所有文件名以及第六文件名；或者，若待访问文件名的特征值大于第六文件名的特征值，确定重新确定的第三集合包括当前第三集合中位于第六文件名之后的所有文件名，重新确定的第四集合包括第六文件名、当前第三集合中位于第六文件名之后的所有文件名以及第五文

件名。

可选的，在本申请的另一种可能的实现方式中，上述预设规则为字典序顺序，对于每个目录块，第二预设位置的文件名为该目录块的首个文件名，上述处理单元还用于：若第六文件名为当前第三集合中的首个文件名，待访问文件名的特征值小于第六文件名的特征值，则将位于第六文件名之前且与第六文件名相邻的文件名归属的目录块确定为目标目录块；若第六文件名为当前第三集合中的最后一个文件名，待访问文件名的特征值大于第六文件名的特征值，则将第六文件名归属的目录块确定为目标目录块。

可选的，在本申请的另一种可能的实现方式中，上述处理单元具体用于：确定待访问文件名与第四文件名之间共有的第三前缀；确定待访问文件名与第五文件名之间共有的第四前缀；将第三前缀与第四前缀中长度最小的一个确定为第二公共前缀。

可选的，在本申请的另一种可能的实现方式中，上述  $n$  个目录块按照预设规则顺序排列，并采用完全二叉树方式存储。相应的，上述处理单元，还用于执行步骤 A3 和步骤 B3；其中，步骤 A3 为：确定当前候选目录块和当前第三公共前缀；步骤 B3 为：从当前第三公共前缀之后的首位字符起，逐字符对比待访问文件名与第  $i$  个文件名，第  $i$  个文件名为  $n$  个目录块中第  $i$  个目录块中第四预设位置的文件名， $0 \leq i < n$ 。上述处理单元，还用于若待访问文件名与第  $i$  个文件名相同，则重新确定候选目录块为第  $i$  个文件名归属的目录块，并将重新确定的候选目录块确定为目标目录块。

可选的，在本申请的另一种可能的实现方式中，上述处理单元，还用于若待访问文件名与第  $i$  个文件名不同，则重新确定第三公共前缀、候选目录块以及第  $i$  个文件名，重新确定的第  $i$  个文件名为  $n$  个目录块中第  $j$  个目录块中第四预设位置的文件名；其中，若待访问文件名的特征值大于第  $i$  个文件名的特征值， $j=2i+2$ ；若待访问文件名的特征值小于第  $i$  个文件名的特征值， $j=2i+1$ ， $0 \leq i < j < n$ 。上述处理单元，还用于根据重新确定的第三公共前缀、重新确定的候选目录块以及重新确定的第  $i$  个文件名，重新执行步骤 B3，直到确定出目标目录块。

可选的，在本申请的另一种可能的实现方式中，上述预设规则顺序为字典序，第四预设位置的文件名为对应目录块中的首个文件名。相应的，上述处理单元具体用于：若待访问文件名的特征值大于第  $i$  个文件名的特征值，则确定重新确定的候选目录块为第  $i$  个文件名归属的目录块；若待访问文件名的特征值小于第  $i$  个文件名的特征值，则确定重新确定的候选目录块为当前候选目录块。

可选的，在本申请的另一种可能的实现方式中，上述预设规则顺序为字典序，第四预设位置的文件名为对应目录块中的首个文件名。相应的，上述处理单元具体用于：当待访问文件名的特征值大于第  $i$  个文件名的特征值时，将当前第一目标前缀更新为待访问文件名与第  $i$  个文件名之间共有的前缀；将更新后的第一目标前缀与当前第二目标前缀中长度最小的一个确定为重新确定的第三公共前缀；或者，当待访问文件名的特征值小于第  $i$  个文件名的特征值时，将当前第二目标前缀更新为待访问文件名与第  $i$  个文件名之间共有的前缀；将当前第一目标前缀与更新后的第二目标前缀中长度最小的一个确定为重新确定的第三公共前缀；其中，第一目标前缀的长度和第二目标前缀的长度的初始值均为零，且第一目标前缀的长度和第二目标前缀的长度随着待访

问文件名的特征值与第*i*个文件名的特征值的大小关系发生变化。

5 第四方面，提供一种终端，该终端包括：一个或多个处理器、存储器、通信接口。该存储器、通信接口与一个或多个处理器耦合；存储器用于存储计算机程序代码，计算机程序代码包括指令，当一个或多个处理器执行指令时，终端执行如上述第一方面及其任意一种可能的实现方式所述的数据查询方法。

第五方面，提供一种计算机可读存储介质，所述计算机可读存储介质中存储有指令，当所述指令在上述第四方面所述的终端上运行时，使得所述终端执行如上述第一方面及其任意一种可能的实现方式所述的数据查询方法。

10 第六方面，提供一种包含指令的计算机程序产品，当该计算机程序产品在上述第四方面所述的终端上运行时，使得所述终端执行如上述第一方面及其任意一种可能的实现方式所述的数据查询方法。

在本申请中，上述数据查询装置的名字对设备或功能模块本身不构成限定，在实际实现中，这些设备或功能模块可以以其他名称出现。只要各个设备或功能模块的功能和本申请类似，属于本申请权利要求及其等同技术的范围之内。

15 本申请中第二方面到第六方面，及其各种实现方式的具体描述，可以参考第一方面及其各种实现方式中的详细描述；并且，第二方面到第六方面，及其各种实现方式的有益效果，可以参考第一方面及其各种实现方式中的有益效果分析，此处不再赘述。

本申请的这些方面或其他方面在以下的描述中会更加简明易懂。

## 附图说明

- 20 图 1 为查找二分算法的查询流程示意图；  
 图 2 为现有技术中 Linux 操作系统中的分区结构示意图；  
 图 3 为 F2FS 文件系统的目录结构示意图；  
 图 4 为本申请实施例中数据查询装置的硬件结构示意图；  
 图 5 为本申请实施例中数据查询装置的外部存储器 42~2 的分区结构示意图；  
 25 图 6 为本申请实施例中数据查询装置的目录块的结构分布示意图；  
 图 7 为本申请实施例中数据查询装置的目录块的排列结构示意图一；  
 图 8 为本申请实施例中数据查询装置的目录块的排列结构示意图二；  
 图 9 为本申请实施例提供的数据查询方法的流程示意图；  
 图 10 为本申请实施例中数据查询装置确定目标目录块的流程示意图一；  
 30 图 11 为本申请实施例中数据查询装置确定目标目录块的流程示意图二；  
 图 12 为本申请实施例中第二集合的结构示意图；  
 图 13 为本申请实施例中数据查询装置查询待访问文件名的流程示意图一；  
 图 14 为本申请实施例中数据查询装置的结构示意图一；  
 图 15 为本申请实施例中数据查询装置的结构示意图二。

## 35 具体实施方式

本申请的说明书和权利要求书及上述附图中的术语“第一”、“第二”和“第三”等是用于区别不同对象，而不是用于限定特定顺序。

在本申请实施例中，“示例性的”或者“例如”等词用于表示作例子、例证或说明。本申请实施例中被描述为“示例性的”或者“例如”的任何实施例或设计方案不应被解释

为比其它实施例或设计方案更优选或更具优势。确切而言,使用“示例性的”或者“例如”等词旨在以具体方式呈现相关概念。

为了便于理解本申请实施例,这里先解释本申请实施例涉及到的相关要素。

5 二分查找:又称为折半查找,其基本思想是:在  $m$  个升序排列的元素(如  $A_1$ 、 $A_2$ 、...、 $A_m$ )中查询数据  $x$ ,取  $A[\lfloor(1+m)/2\rfloor]$ 与  $x$  比较,如果二者相等,则查询终止;如果  $x$  小于  $A[\lfloor(1+m)/2\rfloor]$ ,则在小于  $A[\lfloor(1+m)/2\rfloor]$ 的所有元素中进一步查询,直至找到与  $x$  相等的元素;如果  $x$  大于  $A[\lfloor(1+m)/2\rfloor]$ ,则在大于  $A[\lfloor(1+m)/2\rfloor]$ 的所有元素中进一步查询,直至找到与  $x$  相等的元素。其中,  $\lfloor\cdot\rfloor$ 用于表示下取整。

10 示例性的,如图 1 所示,7 个升序排列的数值为{1、4、7、8、10、15、20},若待查询的数值为“7”,则查询过程为:

(1)、比较{1、4、7、8、10、15、20}中第  $\lfloor(1+7)/2\rfloor$ 个数值“8”与“7”的大小。

由于“7”小于“8”,因此继续在小于“8”的所有数值{1、4、7}中查询。

15 (2)、比较{1、4、7}中的第  $\lfloor(1+3)/2\rfloor$ 个数值“4”与“7”的大小。

由于“7”大于“4”,进一步继续在大于“4”的数值{7}中查询。

(3)、{7}中的数值“7”与待查询的“7”相等,查询终止。

本申请实施例用于查询文件名,因此,在本申请实施例中上述  $m$  个元素为  $m$  个文件名。不同文件名之间的大小可以通过文件名的特征值(如在按照字典序排列文件名后,某一文件名的排列顺序等)确定。

20 文件系统提供了一个结构化的数据存储和组织形式,其采用树形结构组织文件的拓扑关系,为用户访问和查询文件提供了方便。为了管理目录下的子文件和子目录,目录文件需要保存该目录下所有子文件的名称和索引(inode)号。其中,文件名是用户可见的,用户通过文件名管理和访问文件;inode 包括文件的基础信息(如文件大小、文件创建时间、文件修改时间等)和多个指向存储有该文件的数据的各个数据块的指针信息,计算机根据 inode 号能够获取到相应的 inode。当应用程序需要读取某一文件时,根据该文件的 inode 所包括的指针信息,即可确定出存储有该文件数据的各个数据块,进而从确定出的数据块中获取该文件的数据。由于文件系统会记录文件的文件名和索引号,因此,在文件系统的应用场景中,用户无需关心文件的数据存储在哪些数据块上,只需要记住文件所属的目录和文件名即可完成文件数据的访问。

30 一般的,存储介质所提供的存储空间包括多个分区(partition),所述多个分区均被挂载在一个或多个文件系统。每个文件的数据均被存入到存储介质的某个分区中。存储介质的每个分区被划分为多个块(block)。对于同一文件系统,每个 block 的大小相同。典型的,每个 block 的大小是 1024 字节(byte)或者 4096 字节。为了便于描述,本申请实施例将存储有目录文件的数据的 block 称为目录块,将存储有其他类型文件的数据的 block 称为数据块。

35 如图 2 所示,Linux 操作系统中,存储介质的每个分区包含有超级区(Super block)、索引区和数据区。超级区存储有文件系统的相关信息,例如:文件系统的类型,block 的数目、block 的大小等信息。索引区包括  $k$  ( $k \geq 1$ ) 个 inode,每个 inode 均包括文件的基础信息(如文件大小、文件创建时间、文件修改时间等)和多个指向存储有该文

件的数据的各个数据块的指针信息。数据区包括普通文件和目录文件的数据，其中，目录文件由  $n$  ( $n \geq 1$ ) 个目录块组成，每个目录块均包括多个目录项，一个目录项与一个文件对应，每个目录项均包括一个文件的索引 (inode) 号、该文件的文件名以及该文件的文件类型等。

5 目前，一些文件系统（如 F2FS 文件系统、EXT4 文件系统等）中目录的结构为包括多级哈希表的哈希树，每一级哈希表均包括多个哈希值、与每个哈希值对应的文件名以及文件的索引号。

10 以 F2FS 文件系统为例，F2FS 文件系统中目录的结构为包括多级哈希表的哈希树，每一级都有一个使用专用数字的哈希桶 (bucket) 的哈希表。每个哈希 bucket 均为目录项数组，哈希 bucket 中的每个目录项包括一个哈希值、与该哈希值对应的文件名以及文件的索引号。表 1 示出了 F2FS 文件系统中目录的结构，该结构为包括 N 级哈希表的哈希树。表 1 中 A 表示 bucket，B 表示目录块，N 表示哈希最大级，A (2B) 表示一个哈希 bucket 包括两个目录块，A (4B) 表示一个哈希 bucket 包括四个目录块。

表 1

第 0 级	A (2B)
第 1 级	A (2B) - A (2B)
第 2 级	A (2B) - A (2B) - A (2B) - A (2B)
... ..	... ..
第 N/2 级	A (2B) - A (2B) - A (2B) - ... .. A (2B)
第 N/2+1 级	A (4B) - A (4B) - A (4B) - ... .. A (4B)
... ..	... ..
第 N 级	A (4B) - A (4B) - A (4B) - A (4B) - ... .. A (4B)

15 结合表 1 可以看出：F2FS 文件系统中目录的结构中，从第 0 级哈希表开始到第 N/2 级哈希表，每一级哈希表中的每个哈希 bucket 均包括两个目录块；从第 N/2+1 级哈希表开始到第 N 级哈希表，每一级哈希表中的每个哈希 bucket 均包括四个目录块。第  $i$  ( $i < N/2$ ) 级哈希表包括  $2^i$  个哈希 bucket，第  $j$  ( $j \geq N/2$ ) 级哈希表包括  $2^{(N/2-1)}$  个哈希 bucket。

20 图 3 示出了上述表 1 所示的哈希树。第 0 级哈希表包括一个哈希 bucket: Bucket 0，Bucket 0 包括目录块 0 和目录块 1 两个目录块。其他级哈希表的内容与第 0 级哈希表、第 1 级哈希表均类似，这里不再进行详细赘述。

25 基于上述哈希树结构，计算机在 F2FS 文件系统的目录中查询待访问文件名时，首先计算该待访问文件名的哈希值，然后在第 0 级哈希表中扫描哈希值查询包括该待访问文件名和该待访问文件的索引号的目录项。如果没有找到，计算机在第 1 级哈希表中扫描。也就是说，若计算机在上一级哈希表中未找到包括待访问文件名的目录项，则计算机以递增的方式扫描下一级哈希表。在每一级哈希表中，计算机仅需要扫描一个哈希 bucket，该哈希 bucket 的编号是由待访问文件名的哈希值与该级别中的哈希 buckets 数量的相除取余得到的。

30 在 F2FS 文件系统只读的应用场景中，计算机查询待访问文件名时，该计算机需要线性搜索目录项。在哈希表的级别较高、目录项的数量较大的情况下，计算机需要

搜索很多个目录块，查询效率较低。

此外，上述哈希树包括有大量的哈希值，导致存储空间的有效利用率较低。随着级别的增加，高级别哈希表中哈希 bucket 的数量也在适应性的增加。这样，极有可能出现高级别的哈希表未写满的情况，进一步降低了存储空间的有效利用率。

除了哈希树结构的目录之外，还有一些文件系统（如 Ntfs 文件系统、Btrfs 文件系统）中目录的结构为  $n$  ( $n \geq 1$ ) 阶 B+树。

$n$  阶 B+树是一个  $n$  叉排序树。一个 B+树包括根节点、内部节点和叶子节点。其中，根节点可能是一个叶子节点，也可能是一个包括至少两个子节点的节点。 $n$  阶 B+树的所有节点包括  $n$  个关键字，每个关键字不保存数据，只用于索引，所有数据保存在叶子节点。 $n$  阶 B+树的所有叶子节点包括了全部关键字的信息，以及指向这些关键字的指针信息，且叶子节点本身按照关键字的大小自小而大顺序链接。

基于上述  $n$  阶 B+树，计算机在目录中查询待访问文件名时，该计算机从最小关键字开始顺序查询，或者从根节点开始随机查询，查询效率比较低。

此外，B+树的叶子节点中的关键字会在中间节点出现，这种结构导致存储空间的有效利用率较低。

基于上述查询效率低以及计算机存储空间的有效利用率低的问题，本申请实施例提供一种数据查询方法，该方法应用于包括  $n$  ( $n \geq 1$ ) 个目录块的只读文件系统中，每个目录块包括目录项区域和文件名区域，目录项区域包括至少一个目录项，文件名区域包括至少一个文件名，同一目录块中目录项的数量与文件名的数量相同，且同一目录块中的所有目录项以及所有文件名均按照预设规则顺序排列，不同目录块中的目录项的数量可能相同，也可能不同。示例性的， $n$  个目录块中目标目录块的文件名区域包括  $m$  ( $m \geq 1$ ) 个文件名，目标目录块的目录项区域包括  $m$  个目录项， $m$  个文件名与  $m$  个目录项一一对应，所述  $m$  个目录项和所述  $m$  个文件名均按照预设规则顺序排列。

本申请实施例中的数据查询装置在确定出待访问文件名位于目标目录块中的  $m$  个文件名所组成的文件名范围后，即数据查询装置在确定出目标目录块后，根据二分查找算法和该目标目录块，确定包括所述  $m$  个文件名中连续的  $x$  个文件名的当前第一集合以及包括所述  $x$  个文件名、第一文件名和第二文件名的当前第二集合，并确定待访问文件名与当前第二集合中的文件名之间的第一公共前缀，这样，该数据查询装置从第一公共前缀之后的首位字符起，逐字符对比待访问文件名与当前第一集合中第一预设位置的文件名（即第三文件名），当待访问文件名与第三文件名相同时，该数据查询装置根据与第三文件名对应的目录项获取待访问文件的数据。数据查询装置确定了待访问文件名与当前第二集合中的文件名之间的第一公共前缀，由于当前第二集合覆盖第一集合，因此，当前第一集合中的所有文件名与待访问文件名之间也存在第一公共前缀，这样，数据查询装置直接从第一公共前缀之后的首位字符起，逐字符对比待访问文件名与第三文件名即可，有效的提高了查询待访问文件的速率。此外，与现有的目录树结构相比，本申请实施例中的目录结构仅存储有目录项与文件名，并未存储与文件名或目录项相关的其他信息，有效的提高了存储空间的利用率。

本申请实施例中的数据查询装置可以为电脑、手机、平板电脑等终端。图 4 是本

申请实施例提供的一种数据查询装置的结构示意图。参见图 4, 该数据查询装置包括: 通信接口 40、处理器 41 和存储介质 42。其中, 通信接口 40、处理器 41 和存储介质 42 之间通过系统总线 44 连接, 并完成相互间通信。

通信接口 40 用于与其他设备通信, 例如向其他设备共享某一文件的数据。

5 存储介质 42 可用于存储目录文件的数据, 也可以用于存储普通文件的数据, 还可以用于存储软件程序以及应用模块, 处理器 41 通过运行存储在存储介质 42 的软件程序以及应用模块, 从而执行数据查询装置的各种功能应用。

10 存储介质 42 包括内存 42~1 和外部存储器 42~2。内存 42~1 用于暂时存储处理器 41 的运算数据, 以及与外部存储器 42~2 交换的数据等。外部存储器 42~2 用于存储应用程序、目录文件以及普通文件的数据。本申请实施例中, 目录文件由  $n$  ( $n \geq 1$ ) 个目录块组成, 每个目录块包括目录项区域和文件名区域, 目录项区域包括至少一个目录项, 文件名区域包括至少一个文件名, 同一目录块中目录项的数量与文件名的数量相同, 且同一目录块中的所有目录项以及所有文件名均按照预设规则顺序排列, 不同目录块中的目录项的数量可能相同, 也可能不同。目录块的详细解释参考后续描述, 15 这里不对其进行详细描述。在本申请实施方式中, 所述操作系统可以为 Windows 操作系统, 也可以是 Linux 操作系统。

20 在本申请实施例中, 外部存储器 42~2 为非易失性存储器, 例如至少一个磁盘存储器件、电子可擦除可编程只读存储器 (Electrically Erasable Programmable Read-Only Memory, EEPROM)、闪存器件, 例如反或闪存 (NOR flash memory) 或是反及闪存 (NAND flash memory)。非易失存储器存储处理器 41 所执行的操作系统及应用程序。处理器 41 可以从非易失存储器加载运行程序与数据到内存 42~1, 并将数据内容存储于专门用于存储的存储设备中。

25 存储介质 42 可以是独立存在, 通过系统总线 44 与处理器 41 相连接。存储介质 42 也可以和处理器 41 集成在一起。

30 处理器 41 是数据查询装置的控制中心。处理器 41 利用各种接口和线路连接整个数据查询装置的各个部分, 通过运行或执行存储在存储介质 42 内的软件程序和/或应用模块, 以及调用存储在存储介质 42 内的数据, 执行存储设备的各种功能和处理数据, 从而对数据查询装置整体监控。

35 处理器 41 可以仅包括中央处理器 (Central Processing Unit, CPU), 也可以是 CPU、数字信号处理器 (Digital Signal Processor, DSP) 以及通信单元中的控制芯片的组合。在本申请实施方式中, CPU 可以是单运算核心, 也可以包括多运算核心。在具体实现中, 作为一种实施例, 处理器 41 可以包括一个或多个 CPU, 例如图 4 中的处理器 41 包括 CPU 0 和 CPU 1。

系统总线 44 可以是上述元件相互连接并在上述元件之间传递通信的电路, 例如: 40 该系统总线 44 是工业标准体系结构 (Industry Standard Architecture, ISA) 总线、外部设备互连 (Peripheral Component Interconnect, PCI) 总线、扩展工业标准体系结构 (Extended Industry Standard Architecture, EISA) 总线或高级微控制器总线架构 (Advanced Microcontroller Bus Architecture, AMBA) 等。该系统总线 44 可以分为地址总线、数据总线、控制总线等。本申请实施例中为了清楚说明, 在图 4 中将各种总

线都示意为系统总线 44。

需要说明的是，本申请实施例提供的数据库查询方法适用于上述数据库查询装置的文件系统为只读文件系统的应用场景中。

5 为了便于理解本申请实施例提供的数据库查询方法，首先介绍本申请实施例中数据库查询装置的外部存储器 42~2 存储的目录结构。

本申请实施例中数据库查询装置的外部存储器 42~2 中，每个分区包括超级区、索引区和数据区。其中，超级区、索引区和数据区可以参考上述图 2 所示的结构描述，这里不再进行详细赘述。

10 与图 2 所示的结构相同，本申请实施例的数据区中的目录文件也包括 n 个目录块。不同的是，本申请实施例中，n 个目录块中的每个目录块均包括目录项区域和文件名区域，目录项区域包括多个目录项（以 m 个目录项为例进行说明， $m \geq 1$ ），文件名区域包括多个文件名（以 m 个文件名为例进行说明），m 个目录项与 m 个文件名一一对应，且 m 个目录项和 m 个文件名均按照预设规则顺序排列。由于目录项与文件名一一对应，因此，目录项中记录有与其对应的文件名的存储地址。

15 需要说明的是，不同的目录块所包括的目录项的数量可以相同，也可以不同，本申请实施例对此不作具体限定。

本申请实施例中的目录项与现有的目录项有略微的不同。现有的目录项包括文件名、索引号和文件类型等信息。本申请实施例中的目录项包括索引号、文件类型和文件名在目录块的偏移量等信息，并不包括文件名。例如，对于索引号为 S1 的普通文件 20 ABC 而言，现有的与文件名“ABC”对应的目录项包括 S1、ABC、普通文件类型等信息；本申请实施例中将该文件的文件名存储于文件名区域，若该文件在目录块的偏移量为 32，则本申请实施例中与文件名“ABC”对应的目录项包括 S1、32、普通文件类型等信息。

25 为了便于说明，后续内容中的目录项均表示未包括文件名的目录项。

由于文件名的长度是可变的，因此，本申请实施例中目录项和文件名独立存储的结构使得数据库查询装置的存储空间得到有效利用，有效的提高了数据库查询装置的存储空间的利用率。

30 本申请实施例中的文件名区域位于目录项区域之后，且二者相邻，这样，目录项区域中的首个目录项包括的文件名在目录块的偏移量不仅用于指示首个文件名的存储地址，还用于指示目录项区域结束。由于同一目录块中目录项的大小相同，因此，在首个目录项用于指示目录项区域结束的情况下，数据获取装置能够计算出目录项的数量。

35 在一个示例中，图 5 示出了本申请实施例中数据库查询装置的存储介质 42 的分区结构。数据区包括 n 个目录块（目录块 1、目录块 2、... ..、目录块 n），每个目录块均包括目录项区域和文件名区域，目录块 1 的目录项区域包括 m 个目录项（目录项 1、目录项 2、... ..、目录项 m），目录块 1 的文件名区域包括 m 个文件名（文件名 1、文件名 2、... ..、文件名 m），目录项 1 包括文件名 1 的存储位置，目录项 2 包括文件名 2 的存储位置，以此类推，目录项 m 包括文件名 m 的存储位置。数据获取装置根据文件名 1 的存储位置和目录块的起始位置，能够计算出该目录块 1 中目录项的数量

为 m。

由于文件名的长度是变化的，因此，一个目录块包括目录项的数量是不固定的。特别的，若文件 A 的文件名较长，则目录块可能仅包括一个目录项和文件 A 的文件名。图 6 示出了仅包括一个目录项和一个文件名的目录块。

5 可选的，若本申请实施例中的某一目录块的存储空间不足以容纳该文件名，则可以将该文件名存储于该目录块的下一目录块的首个目录项中，也可以将该文件名存储于某一数据块中，并在该目录块的文件名区域中存储用于指向所述某一数据块的偏移信息。

10 本申请实施例中的 n 个目录块以及每个目录块中的所有文件名均按照预设规则顺序排列。

其中，预设规则可以为字典序顺序，也可以为字典序逆序，还可以按照其他有序排列方式，本申请实施例对此不作具体限定。

每个目录块中，目录项与文件名之间一一对应，因此，目录块中的所有目录项的排列顺序与该目录块中的所有文件名的排列顺序相同。

15 为了便于理解，本申请实施例后续涉及到的预设规则均以字典序顺序为例进行说明。

本申请实施例中 n 个目录块的存储方式可以为块间顺序存储或块间按照完全二叉树方式存储。

20 块间顺序存储是指 n 个目录块是根据 n 个文件名按照预设规则顺序排列存储的，这里，n 个文件名中的每个文件名是指与该文件名对应的目录块中第二预设位置的文件名。

其中，目录块中第二预设位置的文件名可以为该目录块的首个文件名，也可以为该目录块的最后一个文件名，还可以为该目录块中除首个文件名和最后一个文件名之外的其他任一文件名，本申请实施例对此不作具体限定。

25 为了便于理解，后续均以目录块中第二预设位置的文件名为该目录块中的首个文件名或该目录块中的最后一个文件名为例进行说明。

30 示例性的，如图 7 所示，数据查询装置包括的目录块中第一个文件名有 ANDY、BABY、CAFE、DASH 以及 EMMA，按照字典序顺序排列，这 5 个文件名的排列顺序依次为：ANDY、BABY、CAFE、DASH、EMMA，因此，数据查询装置中的目录块 1 为第一个文件名为 ANDY 的目录块，目录块 2 为第一个文件名为 BABY 的目录块，目录块 3 为第一个文件名为 CAFE 的目录块，目录块 4 为第一个文件名为 DASH 的目录块，目录块 5 为第一个文件名为 EMMA 的目录块。

35 块间按照完全二叉树方式存储是指 n 个目录块是按照完全二叉树方式存储的。本申请实施例中的 n 个目录块根据每个目录块中第四预设位置的文件名按照所述预设规则排列。

其中，目录块中第四预设位置的文件名可以为该目录块的首个文件名，也可以为该目录块的最后一个文件名，还可以为该目录块中除首个文件名和最后一个文件名之外的其他任一文件名，本申请实施例对此不作具体限定。

示例性的，如图 8 所示，目录块中第四预设位置的文件名为该目录块中的第一个

文件名, 数据查询装置包括的目录块中第一个文件名有 ANDY、BABY、CAFE、DASH 以及 EMMA, 按照字典序顺序排列, 这 5 个文件名的排列顺序依次为: ANDY、BABY、CAFE、DASH、EMMA。这 5 个文件名按照完全二叉树方式存储后, 其排列顺序依次为: DASH、BABY、EMMA、ANDY、CAFE。相应的, 数据查询装置中的目录块 1 为第一个文件名为 DASH 的目录块, 目录块 2 为第一个文件名为 BABY 的目录块, 目录块 3 为第一个文件名为 EMMA 的目录块, 目录块 4 为第一个文件名为 ANDY 的目录块, 目录块 5 为第一个文件名为 CAFE 的目录块。

现以 n 个目录块的存储方式为块间顺序存储为例说明本申请提供的的数据查询方法。具体的, 结合上述图 4 所示的数据查询装置的结构示意图、上述图 5 所示的数据查询装置的存储介质 42 的分区结构、上述图 7 所示的数据查询装置中目录块的一种结构示意图进行详细描述。

具体的, 请参见图 9, 图 9 为本申请实施例提供的一种数据查询方法的流程图。

如图 9 所示, 本申请实施例提供的的数据查询方法具体包括如下步骤。

S900、数据查询装置获取文件访问请求, 该文件访问请求包括待访问文件名。

可选的, 数据查询装置中的某一应用程序在运行过程中触发了获取待访问文件的命令, 本申请实施例中待访问文件的名称统称为待访问文件名。这里, 获取待访问文件的命令也称为文件访问请求, 该文件访问请求包括待访问文件名。

示例性的, 数据查询装置中的某一应用程序在运行过程中触发了命令 “file \*fp; fp= ( “file a” )”, 则数据查询装置获取到了访问文件 a 的请求。

S901、响应于上述文件访问请求, 数据查询装置从 n 个目录块中确定目标目录块。

结合上述描述可知, 本申请实施例中的数据查询装置包括 n 个目录块, 每个目录块均包括目录项区域和文件名区域, 目录项区域包括至少一个目录项, 文件名区域包括至少一个文件名。

本申请实施例中每个目录块中的文件名按照预设规则顺序排序, 因此, 每个目录块的文件名均可组成一个文件名范围。

示例性的, 结合上述图 7, 若目录块 1 包括 3 个文件名: ANDY、APPLE 和 ATTENT, 则该目录块 1 中的文件名组成的文件名范围为: [ANDY, ATTENT]。

这样, 数据查询装置在获取到文件访问请求后, 需要首先确定出待访问文件名具体位于哪一目录块的文件名所组成的文件名范围内, 进而从确定出的目录块中查找待访问文件名。

具体的, 本申请实施例中的数据查询装置从 n 个目录块中确定目标目录块, 即待访问文件名位于目标目录块的文件名所组成的文件名范围内。

容易理解的是, 待访问文件名位于目标目录块的文件名所组成的文件名范围内, 但待访问文件名可以不属于目标目录块包括的文件名。也就是说, 目标目录块可能包括待访问文件名。

结合上一示例, 目录块 1 包括 3 个文件名: ANDY、APPLE 和 ATTENT, 该目录块 1 中的文件名组成的文件名范围为: [ANDY, ATTENT], 待访问文件名 ANGEL 位于文件名范围 [ANDY, ATTENT] 内, 但待访问文件名 ANGEL 不是目录块 1 包括

的文件名。

本申请实施例中的目标目录块的目录项区域包括  $m$  个目录项，目标目录块的文件名区域包括  $m$  个文件名，且  $m$  个目录项与  $m$  个文件名之间一一对应。目录项包括索引号、文件类型以及文件名在目录块的偏移量等信息。

5 具体的，数据查询装置确定目标目录块的方法为：

步骤 A2: 数据查询装置根据二分查找算法和  $n$  个目录块，确定当前第三集合和当前第四集合，当前第三集合包括  $p$  个文件名，该  $p$  个文件名包括  $p$  个目录块中每个目录块中第二预设位置的文件名，且  $p$  个文件名按照预设规则顺序排列，这里， $p$  个目录块为  $n$  个目录块中连续的目录块，当前第四集合包括所述  $p$  个文件名、第四文件名以及第五文件名，第四文件名为排列于  $p$  个文件名中的首个文件名之前且与  $p$  个文件名中的首个文件名相邻的文件名，第五文件名为排列述  $p$  个文件名中的最后一个文件名之后且与  $p$  个文件名中的最后一个文件名相邻的文件名， $1 \leq p \leq n$ ；

步骤 B2: 确定待访问文件名与当前第四集合中的文件名之间的第二公共前缀；

15 步骤 C2: 从第二公共前缀之后的首位字符起，逐字符对比待访问文件名与第六文件名；其中，第六文件名为当前第三集合中第三预设位置的文件名；

若待访问文件名与第六文件名相同，则数据查询装置确定第六文件名归属的目录块为目标目录块。

20 从上述集合的定义可以看出，当前第四集合的文件名组成的文件名范围大于当前第三集合的文件名组成的文件名范围，且当前第四集合包括第四文件名以及第五文件名。特殊的，在  $p=n$  的情况下，第四文件名与第五文件名为空。

25 由于当前第四集合的文件名组成的文件名范围大于当前第三集合的文件名组成的文件名范围，因此，当前第三集合中的文件名也存在待访问文件名与当前第四集合中的文件名之间的第二公共前缀，这样的话，数据查询装置在确定出第二公共前缀后，可从第二公共前缀后的首位字符起，逐字符对比待访问文件名与第六文件名，提高了查询的效率。

容易理解的是，在第四文件名与第五文件名为空的情况下，第二公共前缀初始化为空。

30 具体的，数据查询装置确定待访问文件名与当前第四集合中的文件名之间的第二公共前缀的方法为：数据查询装置确定待访问文件名与第四文件名之间共有的第三前缀，并确定待访问文件名与第五文件名之间共有的第四前缀；然后，该数据查询装置将第三前缀与第四前缀中长度最小的一个确定为第二公共前缀。

35 在一个示例中，结合上述图 7，令当前第四集合为 {ANDY、BABY、CAFE、DASH、EMMA}，当前第三集合为 {BABY、CAFE、DASH}，第四文件名为 ANDY，第五文件名为 EMMA，待访问文件名为 CORE，则数据查询装置确定第三前缀为空，第四前缀为空，则该数据查询装置确定第二公共前缀为空。

在另一个示例中，令当前第四集合为 {A、AC、ACB、ACD、AD、B、C}，当前第三集合为 {AC、ACB、ACD、AD、B}，若待访问文件名为 ACC，第四文件名为 A，第五文件名为 C，数据查询装置确定第三前缀为 A，第四前缀为空，则该数据查询装置确定第二公共前缀为空。

第六文件名为当前第三集合中第三预设位置的文件名，该第三预设位置的文件名可以为当前第三集合中的第  $\lfloor (1+p)/2 \rfloor$  个文件名，也可以为当前第三集合中其他位置的文件名，本申请实施例对此不作具体限定。

5 数据查询装置逐字符对比待访问文件名与第六文件名后，可确定出待访问文件名与第六文件名是否相同。

若待访问文件名与第六文件名相同，则数据查询装置确定第六文件名归属的目录块为目标目录块。可选的，若待访问文件名与第六文件名相同，数据查询装置可以无需再确定目标目录块，该数据查询装置可直接根据第六文件名确定出与待访问文件名对应的目录项，进而根据与待访问文件名对应的目录项，获取到待访问文件的数据。

10 由于数据查询装置在确定目标目录块的过程中直接找到与待访问文件名相同的文件名的过程较为简单，本申请实施例对这一情况不作详细描述。

若待访问文件名与第六文件名不同，当  $2 \leq p \leq n$  时，数据查询装置根据当前第三集合、当前第四集合和二分查找算法，重新确定第三集合和第四集合，并根据重新确定的第三集合和重新确定的第四集合，执行上述步骤 B2 和上述步骤 C2。

15 结合前面对二分查找的描述可知，重新确定的第三集合包括当前第三集合中排列于第六文件名之前或之后的所有文件名。

具体的，若待访问文件名的特征值大于第六文件名的特征值，则重新确定的第三集合包括当前第三集合中排列于第六文件名之后的所有文件名。若待访问文件名的特征值小于第六文件名的特征值，则重新确定的第三集合包括当前第三集合中排列于第六文件名之前的所有文件名。

20 示例性的，结合上述图 7，令当前第三集合包括的文件名为图 7 中的 ANDY、BABY、CAFE、DASH 和 EMMA，第六文件名为 CAFE。若待访问文件名为 ANDY，则重新确定的第三集合包括的文件名为 ANDY 和 BABY。若待访问文件名为 DASH，则重新确定的第三集合包括的文件名为 DASH 和 EMMA。

25 数据查询装置根据重新确定的第四集合执行上述步骤 B2，重新确定第二公共前缀。

30 在一个示例中，令当前第四集合为 {A、AC、ACB、ACD、AD、B、C}，当前第三集合为 {AC、ACB、ACD、AD、B}，若待访问文件名为 ACC，第六文件名为 ACD，第四文件名为 A，第五文件名为 C，数据查询装置确定当前的第二公共前缀为空（当前的第三前缀为 A，当前的第四前缀为空，因此，当前的第二公共前缀为空），并根据该该当前的第二公共前缀从首位字符起逐字符对比待访问文件名 ACC 与文件名 ACD。由于待访问文件名 ACC 的特征值小于文件名 ACD 的特征值，因此，该数据查询装置重新确定第三集合和第四集合，重新确定的第三集合为 {AC、ACB}，重新确定的第四集合为 {A、AC、ACB、ACD}。数据查询装置根据重新确定的第四集合重新确定第二公共前缀。重新确定的第四集合中的第四文件名为 A，重新确定的第四集合中的第五文件名为 ACD，因此，在重新确定第二公共前缀的过程中，第三前缀为 A，第四前缀为 AC，数据查询装置将 A 与 AC 中长度最小的一个确定为第二公共前缀，即数据查询装置重新确定的第二公共前缀为 A。

35 若待访问文件名与第六文件名不同，当  $p=1$  时，数据查询装置根据当前第三集合包括的文件名确定目标目录块。具体的，在第二预设位置的文件名为目录块的首个文

件名的情况下，若待访问文件名的特征值大于当前第三集合中的文件名的特征值，则数据查询装置将当前第三集合中的文件名归属的目录块确定为目标目录块；若待访问文件名的特征值小于当前第三集合中的文件名的特征值，则数据查询装置将位于当前第三集合中的文件名之前且与当前第三集合中的文件名相邻的文件名归属的目录块确定为目标目录块。

5 在一个示例中，每一文件名为对应目录块中的首个文件名，若当前第三集合包括 A、B、C、D、E、F、G、H 以及 I 这九个文件名，第六文件名为 E，待访问文件名为 EA，则数据查询装置确定目标目录块的过程为：①、第二公共前缀为空，数据查询装置根据第二公共前缀从首位字符起逐字符对比待访问文件名 EA 与第六文件名 E；由于待访问文件名 EA 的特征值大于文件名 E 的特征值，该数据查询装置重新确定第三集合为 {F、G、H、I}；②、步骤①中，数据查询装置重新确定的第三集合为 {F、G、H、I}，因此，这一步骤中，当前第三集合为 {F、G、H、I}，若第六文件名为当前第三集合中的第  $\lfloor (1+q)/2 \rfloor$  个文件名，则第六文件名为 G，这一步骤中第二公共前缀为空，数据查询装置根据第二公共前缀从首位字符起逐字符对比待访问文件名 EA 与文件名 G；由于待访问文件名 EA 的特征值小于文件名 G 的特征值，该数据查询装置重新确定第三集合为 {F}，重新确定的第四集合为 {E、F、G}；③、步骤②中，数据查询装置重新确定的第三集合为 {F}，重新确定的第四集合为 {E、F、G}，因此，这一步骤中，当前第三集合为 {F}，当前第四集合为 {E、F、G}，数据查询装置根据 {E、F、G} 确定的第二公共前缀为空；数据查询装置根据当前的第二公共前缀从首位字符起逐字符对比待访问文件名 EA 与文件名 F；由于当前第三集合仅包括一个文件名 F，且待访问文件名 EA 的特征值小于 F 的特征值，因此，目标目录块为位于 F 之前且与 F 相邻的文件名 E 归属的目录块。

10 15 20 25 30 35 在另一个示例中，结合上述图 7，如图 10 所示，若当前第三集合为 {ANDY、BABY、CAFE、DASH、EMMA}，待访问文件名为 CORE，数据查询装置确定目标目录块的过程为：①、此时，第四文件名与第五文件名为空，第二公共前缀为空，若第六文件名为 CAFE，数据查询装置根据该第三公共前缀从首字符逐字符对比待访问文件名 CORE 与文件名 CAFE；由于待访问文件名 CORE 的特征值大于文件名 CAFE 的特征值，数据查询装置重新确定第三集合为 {DASH、EMMA}，并重新确定第二公共前缀为空（待访问文件名 CORE 与文件名 CAFE 之间共有的第三前缀为“C”；待访问文件名 CORE 与第五文件名之间共有的第四前缀为空；因此，重新确定的第二公共前缀为空）；②、步骤①中数据查询装置重新确定的第三集合为 {DASH、EMMA}，因此在这一步骤中，当前第三集合为 {DASH、EMMA}，若第六文件名为 EMMA，数据查询装置根据当前的第二公共前缀从首字符逐字符对比待访问文件名 CORE 与文件名 EMMA；③、由于待访问文件名 CORE 的特征值小于文件名 EMMA 的特征值，数据查询装置重新确定第三集合为 {DASH}，并重新确定第二公共前缀为空；此时，当前第三集合为 {DASH}，当前的第二公共前缀为空，数据查询装置根据当前的第二公共前缀从首位字符起逐字符对比待访问文件名 CORE 与文件名 DASH；由于当前第三集合仅包括一个文件名 DASH，且待访问文件名 CORE 的特征值小于文件名 DASH 的特征值，因此，数据查询装置确定目标目录块为位于 DASH 之前且与 DASH 相邻

的文件名 CAFE 归属的目录块；④、数据查询装置确定目标目录块为目录块 3。

5 在一个示例中，每一文件名为对应目录块中的首个文件名，若当前第三集合为 { A、AC、ACB、ACD、AD、B、C }，第四文件名与第五文件名为空，第六文件名为 ACD，待访问文件名为 ACC，则数据查询装置确定目标目录块的过程为：①、由于第四文件名与第五文件名为空，因此当前的第二公共前缀为空，数据查询装置根据该第二公共前缀从首位字符起逐字符对比待访问文件名 ACC 与第六文件名 ACD；由于待访问文件名 ACC 的特征值小于文件名 ACD 的特征值，该数据查询装置重新确定第三集合为 { A、AC、ACB }，重新确定第四集合为 { 空、A、AC、ACB、ACD }；②、在步骤①中，数据查询装置重新确定的第三集合为 { A、AC、ACB }，因此，步骤②中当前第三集合为 { A、AC、ACB }，若第六文件名为当前第三集合中的第  $\lfloor (1+q)/2 \rfloor$  个文件名，则第六文件名为 AC；步骤①中重新确定第四集合为 { 空、A、AC、ACB、ACD }，因此，步骤②中当前第四集合为 { 空、A、AC、ACB、ACD }，当前第四集合中的第四文件名为空，当前第四集合中的第五文件名为 ACD，相应的，当前的第三前缀为空，当前的第四前缀为 AC，因此，数据查询装置重新确定的第二公共前缀为空；数据查询装置根据该第二公共前缀从首位字符起逐字符对比待访问文件名 ACC 与文件名 AC；由于待访问文件名 ACC 的特征值大于文件名 AC 的特征值，该数据查询装置重新确定第三集合为 { ACB }，重新确定的第四集合为 { AC、ACB、ACD }；③、在步骤②中，数据查询装置重新确定的第三集合为 { ACB }，重新确定的第四集合为 { AC、ACB、ACD }，因此，在步骤③中，当前第三集合为 { ACB }，当前第四集合为 { AC、ACB、ACD }；当前第四集合中的第四文件名为 AC，因此，第三前缀为 “AC”；当前第四集合中的第五文件名为 ACD，因此，第四前缀为 “AC”；数据查询装置将第三前缀 “AC” 与第四前缀 “AC” 中长度最小的一个作为第二公共前缀，因此，当前的第二公共前缀为 “AC”；数据查询装置该第二公共前缀从 “AC” 之后的首位字符起逐字符对比待访问文件名 ACC 与文件名 ACB；由于当前第三集合仅包括一个文件名 ACB，且待访问文件名 ACC 的特征值大于 ACB 的特征值，因此，目标目录块为 ACB 归属的目录块。

特殊的，在第二预设位置的文件名为目录块中首个文件名，第三预设位置的文件名为当前第三集合中的首个文件名，即第六文件名为当前第三集合中的首个文件名的情况下，若待访问文件名的特征值小于第六文件名的特征值，无论 p 的数值是否大于 1，数据查询装置均将位于第六文件名之前且与第六文件名相邻的文件名归属的目录块确定为目标目录块。

35 在一个示例中，结合上述图 7，若当前第三集合为 { ANDY、BABY、CAFE、DASH、EMMA }，待访问文件名为 CORE，数据查询装置确定目标目录块的过程为：①、此时，第四文件名与第五文件名为空，第二公共前缀为空，若第六文件名为 CAFE，数据查询装置根据该第三公共前缀从首字符逐字符对比待访问文件名 CORE 与文件名 CAFE；由于待访问文件名 CORE 的特征值大于文件名 CAFE 的特征值，数据查询装置重新确定第三集合为 { DASH、EMMA }，并重新确定第二公共前缀为空（待访问文件名 CORE 与文件名 CAFE 之间共有的第三前缀为 “C”；待访问文件名 CORE 与第五文件名之间共有的第四前缀为空；因此，重新确定的第二公共前缀为空）；②、步骤①中数据查询装置重新确定的第三集合为 { DASH、EMMA }，因此在这一步骤中，当前第三集

合为{DASH、EMMA}，若第六文件名为 DASH，数据查询装置根据当前的第二公共前缀从首字符逐字符对比待访问文件名 CORE 与文件名 DASH；③、由于待访问文件名 CORE 的特征值小于文件名 DASH 的特征值，数据查询装置确定目标目录块为位于 DASH 之前且与 DASH 相邻的文件名 CAFE 归属的目录块，即确定目标目录块为目录块 3。

在第二预设位置的文件名为目录块中首个文件名，第三预设位置的文件名为当前第三集合中的最后一个文件名，即第六文件名为当前第三集合中的最后一个文件名的情况下，若待访问文件名的特征值大于第六文件名的特征值，数据查询装置均将该第六文件名归属的目录块确定为目标目录块。

在一个示例中，结合上述图 7，若当前第三集合为{ ANDY、BABY、CAFE、DASH、EMMA}，待访问文件名为 END，第六文件名为当前第三集合中的最后一个文件名 EMMA，由于待访问文件名 END 的特征值大于文件名 EMMA 的特征值，因此，数据查询装置确定文件名 EMMA 归属的目录块为目标目录块。

可选的，在 n 个目录块的存储方式为块间顺序存储的场景中，若上述第六文件名为当前第三集合中的第  $\lfloor (1+p)/2 \rfloor$  个文件名，对于每个目录块，若第二预设位置的文件名为该目录块中首个文件名，本申请实施例中的数据查询装置确定目标目录块的伪代码可以为如下代码：

```

{
  ( headprefix, endprefix ) ← ( 0, 0 )           // headprefix 和 endprefix 的初始值
  均为 0//
  left ← 0
  right ← blocks-1      //目录块数目为 blocks，初始化查询闭区间为[0, blocks-1]//
  while left ≤ right do      // 在查询闭区间内执行下述操作//
    mid ← ( left+right ) /2      //确定查询闭区间的中间位置的文件名//
    25 commonprefix ← min ( headprefix, endprefix )    //公共前缀为 headprefix 和
    endprefix 中的最小一个 //
    ( ret, matched ) ← prefixstrcmp ( fname, dirent0name[mid], commonprefix )
    //从公共前缀后的首个字符起，逐字符对比待访问文件名 fname 与位于查询闭区间中间
    30 位置的文件名 dirent0name[mid]，并确定两个文件名之间共有的前缀//
    if ret=0 then
      return mid
      break      //如果 fname 与 dirent0name[mid]相同，则位于中间位置的
    目录块为目标目录块，返回位于中间位置的目录块的标识，并退出//
    else
    35 if ret>0 then
      left← mid+1
      headprefix ←matched    //如果 fname 的特征值大于 dirent0name[mid]
    的特征值，则查询闭区间更新为[mid+1, blocks-1]，并适应 headprefix //
    else

```

```

        right← mid-1
        endprefix ←matched //如果 fname 的特征值小于 dirent0name[mid]的
特征值，则查询闭区间更新为[0, mid-1]，并适应更新 endprefix //
        end if
5       end if
        end while
        return left-1 //返回 left-1 位置的目录块的标识 //
    }

```

10 该代码中的 headprefix 为第三前缀，endprefix 为第四前缀，上述查询闭区间等效于上述第三集合。

可以看出，在确定目标目录块之前，数据查询装置的查询闭区间在不断缩小，且在每一查询闭区间中，数据查询装置均是从公共前缀后的首个字符起，逐字符对比待访问文件名 fname 与位于该查询闭区间中间位置的文件名 dirent0name[mid]，有效的提高了查询目标目录块的速率。

15 数据查询装置在 S901 中确定出目标目录块后，从该目标目录块中查找与待访问文件名相同的文件名。本申请实施例中将与待访问文件名对应的目录项称为目标目录项。数据查询装置在查找到与待访问文件名相同的文件名后，可确定出目标目录项，进而根据目标目录项获取待访问文件的数据。

具体的，数据查询装置在执行 S901 后，执行 S902，并顺序执行后续步骤。

20 S902、数据查询装置根据二分查找算法和目标目录块，确定当前第一集合和当前第二集合。

25 这里，当前第一集合包括 m 个文件名中连续的 x 个文件名，当前第二集合包括所述 x 个文件名、第一文件名以及第二文件名， $m \geq x \geq 1$ 。第一文件名为排列于 x 个文件名中的首个文件名之前且与 x 个文件名中的首个文件名相邻的文件名，第二文件名为排列于 x 个文件名中的最后一个文件名之后且与 x 个文件名中的最后一个文件名相邻的文件名。

目标目录块包括 m 个文件名，数据查询装置根据二分查找算法在目标目录块中的 m 个文件名中进行查询。

30 具体的，数据查询装置根据二分查找算法和目标目录块，确定包括 x 个文件名的当前第一集合以及包括所述 x 个文件名、第一文件名以及第二文件名的当前第二集合。

容易理解的是，从上述对二分查找的描述以及图 1 示出的流程可知，在数据查询装置的查询过程，查询区间不断缩小，且在每一查询区间内数据查询装置均需比较待访问文件名的特征值与该查询区间中间位置的文件名的特征值的大小。为了便于描述，本申请实施例以当前的查询区间为当前第一集合为例进行说明。

35 特殊的，若  $x=m$ ，则第一文件名与第二文件名为空。

S903、数据查询装置确定待访问文件名与当前第二集合中的文件名之间的第一公共前缀。

从上述当前第一集合和当前第二集合的描述可知，当前第二集合中所有文件名组成的文件名范围大于当前第一集合中所有文件名组成的文件名范围，因此，当前第一

集合中的所有文件名均存在第一公共前缀。

容易理解的是，第一文件名与第二文件名为空的情况下，第一公共前缀初始化为空。

5 具体的，数据查询装置确定第一公共前缀的方法为：数据查询装置确定待访问文件名与第一文件名之间共有的第一前缀，并确定待访问文件名与第二文件名之间共有的第二前缀；然后，该数据查询装置将第一前缀与第二前缀中长度最小的一个确定为第一公共前缀。

10 在一个示例中，如图 11 所示，目标目录块包括 7 个顺序排列的文件名：CAFE、CAGE、CAK、CELL、CORN、DAB、DACE。若当前第一集合为{CAFE、CAGE、CAK、CELL、CORN、DAB、DACE}，第一文件名和第二文件名均为空，第三文件名为 CELL，待访问文件名为 CAK，由于第一文件名和第二文件名均为空，因此，第一前缀和第二前缀均为空，相应的，当前的第一公共前缀为空。

15 若当前第一集合为{CAFE、CAGE、CAK}，则第一文件名为空，第二文件名为 CELL，第一文件名与待访问文件名 CAK 之间共有的第一前缀为空，待访问文件名 CAK 与第二文件名 CELL 之间共有的第二前缀为“C”，第一前缀的长度小于第二前缀，因此，当前的第一公共前缀为空。

S904、数据查询装置从第一公共前缀之后的首位字符起，逐字符对比待访问文件名与第三文件名。

其中，第三文件名为当前第一集合中第一预设位置的文件名。

20 可选的，第一预设位置的文件名可以为当前第一集合中的第  $\lfloor (1+x)/2 \rfloor$  个文件名，也可以为当前第一集合中其他位置的文件名，本申请实施例对此不作具体限定。

由于当前第一集合中的所有文件名均存在第一公共前缀，因此，数据查询装置在确定出第一公共前缀之后，从第一公共前缀之后的首位字符起，逐字符对比待访问文件名与第三文件名，提高了数据查询装置查找与待访问文件名相同的文件名的速率。

25 数据查询装置逐字符对比待访问文件名与第三文件名后，可确定出待访问文件名与第三文件名是否相同。

30 若待访问文件名与第三文件名相同，则说明第三文件名所对应的文件为待访问文件，此时，目标目录项为与第三文件名对应的目录项。数据查询装置根据目标目录项可获取到待访问文件的数据。在这种情况下，数据查询装置在执行 S904 后，继续执行下述步骤 S905。

若待访问文件名与第三文件名不相同，数据查询装置需确定待访问文件名的特征值与第三文件名的特征值之间的大小关系，进而根据该大小关系重新确定第一集合和第二集合，并根据重新确定的第一集合和重新确定的第二集合，执行上述步骤 B1 和上述步骤 C1，直到获取到待访问文件的数据或确定目标目录块未包括待访问文件名。

35 具体的，若待访问文件名的特征值小于第三文件名的特征值，数据查询装置重新确定的第一集合包括当前第一集合中位于第三文件名之前的所有文件名。若待访问文件名的特征值大于第三文件名的特征值，数据查询装置重新确定的第一集合包括当前第一集合中位于第三文件名之后的所有文件名。

示例性的，结合上述图 11，令第一集合为{CAFE、CAGE、CAK、CELL、CORN、

DAB、DACE}, 第三文件名为 CELL。若待访问文件名为 DACE, 则待访问文件名 DACE 的特征值大于第三文件名 CELL 的特征值, 重新确定的第一集合为 {CORN、DAB、DACE}。若待访问文件名为 CAGE, 则待访问文件名 CAGE 的特征值小于第三文件名 CELL 的特征值, 重新确定的第一集合为 {CAFE、CAGE、CAK}。

5 相应的, 若待访问文件名的特征值小于第三文件名的特征值, 数据查询装置重新确定的重新确定的第二集合包括第一文件名、当前第一集合中位于第三文件名之前的所有文件名以及第三文件名。

10 若待访问文件名的特征值大于第三文件名的特征值, 数据查询装置重新确定的第二集合包括第三文件名、当前第一集合中位于第三文件名之后的所有文件名以及第二文件名。

S905、数据查询装置根据目标目录项获取待访问文件的数据。

具体的, 数据查询装置从目标目录项中获取待访问文件的索引号, 并根据待访问文件的索引号获取待访问文件的索引, 进而, 该数据查询装置根据待访问文件的索引即可获取到待访问文件的数据。

15 现结合具体示例说明数据查询装置在目标目录块中查找与待访问文件名相同的文件名的过程。

在一个示例中, 结合图 11, 如图 12 所示, 当前第一集合为 {CAFE、CAGE、CAK、CELL、CORN、DAB、DACE}, 第三文件名为 CELL, 第一文件名和第二文件名均为空, 待访问文件名为 CAK。数据查询装置查询待访问文件名的过程为: ①、此时, 第一公共前缀初始化为空, 数据查询装置根据该第一公共前缀逐字符对比待访问文件名 CAK 与文件名 CELL; 由于待访问文件名 CAK 的特征值小于文件名 CELL 的特征值, 数据查询装置重新确定第一集合为 {CAFE、CAGE、CAK}, 重新确定第二集合中第一文件名为空, 第二文件名为 CELL, 因此, 重新确定的第一公共前缀为空 (待访问文件名 CAK 与文件名 CELL 之间共有的第二前缀为 “C”, 待访问文件名 CAK 与第一文件名之间共有的第一前缀为空, 因此, 重新确定的第一公共前缀为空); ②、在步骤①中重新确定的第一集合为 {CAFE、CAGE、CAK}, 因此, 在步骤②中的当前第一集合为 {CAFE、CAGE、CAK}, 相应的, 当前的第一公共前缀为空, 若第三文件名为 CAGE, 数据查询装置根据当前的第一公共前缀逐字符对比待访问文件名 CAK 与文件名 CAGE, 由于待访问文件名 CAK 的特征值大于文件名 CAGE 的特征值, 数据查询装置重新确定第一集合为 {CAK}, 重新确定第二集合为 {CAGE、CAK、CELL}; ③、在步骤②中重新确定的第一集合为 {CAK}, 重新确定第二集合为 {CAGE、CAK、CELL}, 因此, 在步骤③中的当前第一集合为 {CAK}, 当前第二集合为 {CAGE、CAK、CELL}, 数据查询装置确定当前的第一公共前缀为 “C”, 数据查询装置从 “C” 后的首个字符开始逐字符对比待访问文件名 CAK 与当前第一集合中的文件名 CAK。由于当前第一集合中的文件名 CAK 与待访问文件名 CAK 相同, 因此, 数据查询装置确定目标目录项为与 CAK 对应的目录项。进而, 数据查询装置根据目标目录项获取待访问文件的数据。

可选的, 本申请实施例中数据查询装置在目标目录块中查询待访问文件名的伪代码可以为如下代码:

```

{
  found ← nil //为最终确定出的文件名 found 赋值空指针//
  ( headprefix, endprefix ) ← ( 0, 0 ) // headprefix 和 endprefix 的初始
值均为 0//
5   left ← 0
   right ← ndirents-1 //目录项的数量为 ndirents, 初始化查询闭区间为
[0, ndirents-1] //
   while left ≤ right do // 在查询闭区间内执行下述操作
//
10   mid ← ( left+right ) / 2 //确定查询闭区间的中间位置的文件名//
   commonprefix ← min ( headprefix, endprefix ) //公共前缀为 headprefix 和
endprefix 中的最小一个 //
   ( ret, matched ) ← prefixstrcmp ( fname, direntname[mid], commonprefix ) //
从公共前缀后的首个字符起, 逐字符对比待访问文件名 fname 与位于查询闭区间中间
15 位置的文件名 direntname[mid], 并确定两个文件名之间共有的前缀//
   if ret=0 then
   found ← direntname[mid]
   break //如果 fname 与 direntname[mid]相同, 则最终确定出的文
件名为位于中间位置的文件名, 返回位于中间位置的文件名, 并退出//
20  else
   if ret>0 then
   left ← mid+1
   headprefix ← matched //如果 fname 的特征值大于 direntname[mid]的
特征值, 则查询区间更新为[mid+1, ndirents-1], 并适应更新 headprefix //
25  else
   right ← mid-1
   endprefix ← matched //如果 fname 的特征值小于 direntname[mid]的特
征值, 则查询区间更新为[0, mid-1], 并适应更新 endprefix //
   end if
30  end if
   end while
   return found //返回 found //
}

```

35 该代码中的 headprefix 为第一前缀, endprefix 为第二前缀, fname 为待访问文件名, 上述查询闭区间等效于上述第一集合。

可以看出, 在找到与待访问文件名相同的文件名之前, 数据查询装置的查询区间在不断缩小, 且在每一查询区间中, 数据查询装置均是从公共前缀后的首个字符起, 逐字符对比待访问文件名 fname 与位于该查询区间中间位置的文件名 direntname[mid], 有效的提高了查询速率。

当本申请实施例中  $n$  个目录块的存储方式为块间按照完全二叉树方式存储时，本申请实施例提供的数据查询方法的流程依旧为图 9 示出的流程，数据查询装置也可执行 S900~S905。但是，在  $n$  个目录块的存储方式为块间按照完全二叉树方式存储的场景和  $n$  个目录块的存储方式为块间顺序存储的场景中，数据查询装置从  $n$  个目录块中确定目标目录块的方法不同。

现在对在  $n$  个目录块的存储方式为块间按照完全二叉树方式存储的场景中，数据查询装置从  $n$  个目录块中确定目标目录块的方法进行解释。

具体的，在  $n$  个目录块的存储方式为块间按照完全二叉树方式存储的场景中，数据查询装置确定目标目录块的方法为：

步骤 A3: 数据查询装置确定当前候选目录块和当前第三公共前缀；

步骤 B3: 数据查询装置从当前第三公共前缀之后的首位字符起，逐字符对比待访问文件名与第  $i$  个文件名，该第  $i$  个文件名为  $n$  个目录块中第  $i$  个目录块中第四预设位置的文件名， $0 \leq i < n$ ；

若待访问文件名与第  $i$  个文件名相同，则数据查询装置重新确定候选目录块为第  $i$  个文件名归属的目录块，并将重新确定的候选目录块确定为目标目录块。

第  $i$  个文件名为  $n$  个目录块中第  $i$  个目录块中第四预设位置的文件名。这里，第四预设位置的文件名可以为对应目录块中的首个文件名，也可以为对应目录块中的最后一个文件名，还可以为对应目录块中的其他文件名，本申请实施例对此不作具体限定。

数据查询装置从当前第三公共前缀之后的首位字符起，逐字符对比待访问文件名与第  $i$  个文件名之后，可确定出待访问文件名与第  $i$  个文件名是否相同。

若待访问文件名与第  $i$  个文件名相同，则说明第  $i$  个文件名归属的目录块为目标目录块。此时，数据查询装置可直接根据第  $i$  个文件名确定出与待访问文件名对应的目录项，进而根据与待访问文件名对应的目录项，获取到待访问文件的数据。

若待访问文件名与第  $i$  个文件名不同，则数据查询装置重新确定第三公共前缀、候选目录块以及第  $i$  个文件名，并根据重新确定的第三公共前缀、重新确定的候选目录块以及重新确定的第  $i$  个文件名，重新执行上述步骤 B3，直到确定出目标目录块。

上述重新确定的第  $i$  个文件名为  $n$  个目录块中第  $j$  个目录块中第四预设位置的文件名。其中，若待访问文件名的特征值大于第  $i$  个文件名的特征值， $j=2i+2$ ；若待访问文件名的特征值小于第  $i$  个文件名的特征值， $j=2i+1$ ， $0 \leq i < j < n$ 。

与在  $n$  个目录块的存储方式为块间顺序存储的场景中数据查询装置确定公共前缀的方法类似，在  $n$  个目录块的存储方式为块间按照完全二叉树方式存储的场景中，数据查询装置也确定公共前缀。

在  $n$  个目录块的存储方式为块间按照完全二叉树方式存储的场景中，数据查询装置确定第三公共前缀。

若第四预设位置的文件名为对应目录块中的首个文件名，则数据查询装置重新确定第三公共前缀的方法为：当待访问文件名的特征值大于第  $i$  个文件名的特征值时，数据查询装置将当前第一目标前缀更新为待访问文件名与第  $i$  个文件名之间共有的前缀，并将更新后的第一目标前缀与当前第二目标前缀中长度最小的一个确定为重新确定的第三公共前缀。当待访问文件名的特征值小于第  $i$  个文件名的特征值时，数据查

询装置将当前第二目标前缀更新为待访问文件名与第  $i$  个文件名之间共有的前缀，并将当前第一目标前缀与更新后的第二目标前缀中长度最小的一个确定为重新确定的第三公共前缀。

5 其中，第一目标前缀的长度和第二目标前缀的长度的初始值均为零，且第一目标前缀的长度和第二目标前缀的长度随着待访问文件名的特征值与第  $i$  个文件名的特征值的大小关系发生变化。

此外，在第四预设位置的文件名为对应目录块中的首个文件名的情况下，数据查询装置重新确定候选目录块的方法为：若待访问文件名的特征值大于第  $i$  个文件名的特征值，则数据查询装置确定候选目录块为第  $i$  个文件名归属的目录块；若待访问文件名的特征值小于第  $i$  个文件名的特征值，则数据查询装置确定候选目录块与当前候选目录块相同。

容易理解的是，若  $i=0$ ，当前候选目录块为空，当前第三公共前缀为空。

现结合具体示例说明在  $n$  个目录块的存储方式为块间按照完全二叉树方式存储的场景中，数据查询装置在目标目录块中查找与待访问文件名相同的文件名的过程。

15 示例性的，结合上述图 8，如图 13 所示，若待访问文件名为 CORE，数据查询装置确定目标目录块的过程为：①、候选目录块的初始值为空，第三公共前缀初始值为空，数据查询装置从首位字符起逐字符对比待访问文件名 CORE 与第 0 个文件名 DASH，由于待访问文件名 CORE 的特征值小于 DASH 的特征值，数据查询装置需要进一步比较待访问文件名 CORE 与第 1 ( $2*0+1$ ) 个文件名 BABY。数据查询装置重新确定的第三公共前缀依旧为空，此时，候选目录块保持不变，依旧为空；②、数据查询装置从首位字符起逐字符对比待访问文件名 CORE 与第 1 个文件名 BABY，由于待访问文件名 CORE 的特征值大于 BABY 的特征值，数据查询装置将候选目录块变更为 BABY 归属的目录块（即目录块 2），此外，该查询装置需要进一步比较待访问文件名 CORE 与第 4 ( $2*1+2$ ) 个文件名 CAFE。数据查询装置重新确定的第三公共前缀依旧为空。

25 ③、数据查询装置从首位字符起逐字符对比待访问文件名 CORE 与第 4 个文件名 CAFE。由于待访问文件名 CORE 的特征值大于 CAFE 的特征值，数据查询装置将候选目录块变更为 CAFE 归属的目录块（即目录块 5）；此外，数据查询装置需要进一步比较待访问文件名与第 10 ( $2*4+2$ ) 个文件名；但是，第 10 个文件名不存在；因此，数据查询装置确定 CAFE 归属的目录块为目标目录块，即确定目标目录块为目录块 5。

30 可选的，在  $n$  个目录块的存储方式为块间按照完全二叉树方式存储的场景中，对于每个目录块，若第四预设位置的文件名为该目录块中首个文件名，本申请实施例中的数据查询装置确定目标目录块的伪代码可以为如下代码：

```

{
  candidate ← -1 //候选目录块 candidate 初始值为空//
  35 (headprefix, endprefix) ← (0, 0) // headprefix 和 endprefix 的初始
  值均为 0//
  i ← 0 //从第 i 个文件名开始查询//
  while i < nblocks do //当 i 小于目录块数时执行下述操作//
    commonprefix ← min (headprefix, endprefix) //公共前缀为 headprefix 和

```

```

endprefix 中的最小一个 //
      (ret, matched) ← prefixstrcmp ( fname, dirent0name[i], commonprefix ) //
从公共前缀后的首个字符起，逐字符对比待访问文件名 fname 与第 i 个文件名
dirent0name[i]，并确定两个文件名之间共有的前缀//
5       if ret=0 then
          candidate ← i
          break      //如果 fname 与 dirent0name[i]相同，则第 i 个目录块为目标
          目录块，返回第 i 个目录块的标识，并退出//
        else
10        if ret>0 then
            candidate ← i
            i ← i*2+2
            headprefix ← matched //如果 fname 的特征值大于 dirent0name[i]的特
            征值，则第 i 个文件名更新为第 i*2+2 个文件名，候选目录块 candidate 更新为第 i 个
15        目录块，并适应更新 headprefix //
            else
                i ← i*2+1
                endprefix ← matched //如果 fname 的特征值小于 dirent0name[i]的特
                征值，则第 i 个文件名更新为第 i*2+1 个文件名，并适应更新 endprefix //
20        end if
            end if
        end while
        return candidate //返回 candidate //
    }

```

25 该代码中的 headprefix 相当于第一目标前缀，endprefix 相当于第二目标前缀。

可以看出，数据查询装置均是从公共前缀后的首个字符起，逐字符对比待访问文件名 fname 与第 i 个文件名 dirent0name[i]，有效的提高了查询目标目录块的速率。

30 综上所述，无论 n 个目录块的存储方式是块间顺序存储，还是块间按照完全二叉树方式存储，本申请实施例提供的数据查找方法均可有效的提高查询待访问文件的速率。

本申请实施例提供一种数据查询装置，该数据查询装置用于执行以上数据查询方法中的数据查询装置所执行的步骤。本申请实施例提供的数据查询装置可以包括相应步骤所对应的模块。

35 本申请实施例可以根据上述方法示例对数据查询装置进行功能模块的划分，例如，可以对应各个功能划分各个功能模块，也可以将两个或两个以上的功能集成在一个处理模块中。上述集成的模块既可以采用硬件的形式实现，也可以采用软件功能模块的形式实现。本申请实施例中对模块的划分是示意性的，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。

在采用对应各个功能划分各个功能模块的情况下，图 14 示出上述实施例中所

涉及的数据查询装置的一种可能的结构示意图。如图 14 所示，数据查询装置 140 包括处理单元 1400 和获取单元 1401。

处理单元 1400 用于支持该数据查询装置 10 执行上述实施例中的 S901、S902、S903、S904 等，和/或用于本文所描述的技术的其它过程。

5 获取单元 1401 用于支持该数据查询装置 10 执行上述实施例中的 S900、S905 等，和/或用于本文所描述的技术的其它过程。

其中，上述方法实施例涉及的所有相关内容均可以援引到对应功能模块的功能描述，在此不再赘述。

当然，本申请实施例提供的数据查询装置 140 包括但不限于上述模块，例如：  
10 数据查询装置还可以包括存储单元 1402。

存储单元 1402 可以用于存储该数据查询装置 140 的程序代码和数据。

在采用集成的单元的情况下，本申请实施例提供的数据查询装置的结构示意图如图 15 所示。在图 15 中，数据查询装置 150 包括：处理模块 1500 和通信模块 1501。处理模块 1500 用于对数据查询装置 150 的动作进行控制管理，例如，执行  
15 上述处理单元 1400 执行的步骤，和/或用于执行本文所描述的技术的其它过程。通信模块 1501 用于支持数据查询装置 150 与其他设备之间的交互，例如，执行上述获取单元 1401 执行的步骤。如图 15 所示，数据查询装置 150 还可以包括存储模块 1502，存储模块 1502 用于存储数据查询装置 150 的程序代码和数据，例如存储上述存储单元 1402 所保存的内容。

20 其中，处理模块 1500 可以是处理器或控制器，例如可以是 CPU，通用处理器，DSP，专用集成电路（Application-Specific Integrated Circuit，ASIC），现场可编程门阵列（Field Programmable Gate Array，FPGA）或者其他可编程逻辑器件、晶体管逻辑器件、硬件部件或者其任意组合。其可以实现或执行结合本申请公开内容所描述的各种示例性的逻辑方框，模块和电路。所述处理器也可以是实现计算  
25 功能的组合，例如包含一个或多个微处理器组合，DSP 和微处理器的组合等等。通信模块 1501 可以是收发器、RF 电路或通信接口等。存储模块 1502 可以是存储器。

结合图 4，处理模块 1500 可以为图 4 中的处理器 41，通信模块 1501 可以为图 4 中的通信接口 40，存储模块 1502 可以为图 2 中的存储介质 42。

30 其中，上述方法实施例涉及的所有相关内容均可以援引到对应功能模块的功能描述，在此不再赘述。

上述数据查询装置 140 和数据查询装置 150 均可执行上述图 9 所示的数据查询方法，数据查询装置 140 和数据查询装置 150 具体可以是终端。

本申请还提供一种终端，该终端包括：一个或多个处理器、存储器、通信接口。  
35 该存储器、通信接口与一个或多个处理器耦合；存储器用于存储计算机程序代码，计算机程序代码包括指令，当一个或多个处理器执行指令时，终端执行本申请实施例的数据查询方法。

这里的终端可以是视频显示设备，智能手机，便携式电脑以及其它可以处理视频或者播放视频的设备。

本申请另一实施例还提供一种计算机可读存储介质，该计算机可读存储介质包括一个或多个程序代码，该一个或多个程序包括指令，当终端中的处理器在执行该程序代码时，该终端执行如图 9 所示的数据查询方法。

5 在本申请的另一实施例中，还提供一种计算机程序产品，该计算机程序产品包括计算机执行指令，该计算机执行指令存储在计算机可读存储介质中；终端的至少一个处理器可以从计算机可读存储介质读取该计算机执行指令，至少一个处理器执行该计算机执行指令使得终端实施执行图 9 所示的数据查询方法中的数据查询装置的步骤。

10 在上述实施例中，可以全部或部分的通过软件，硬件，固件或者其任意组合来实现。当使用软件程序实现时，可以全部或部分地以计算机程序产品的形式出现。所述计算机程序产品包括一个或多个计算机指令。在计算机上加载和执行所述计算机程序指令时，全部或部分地产生按照本申请实施例所述的流程或功能。

15 所述计算机可以是通用计算机、专用计算机、计算机网络、或者其他可编程装置。所述计算机指令可以存储在计算机可读存储介质中，或者从一个计算机可读存储介质向另一个计算机可读存储介质传输，例如，所述计算机指令可以从一个网站站点、计算机、服务器或数据中心通过有线（例如同轴电缆、光纤、数字用户线（DSL））或无线（例如红外、无线、微波等）方式向另一个网站站点、计算机、服务器或数据中心传输。所述计算机可读存储介质可以是计算机能够存取的任何可用介质或者是包含一个或多个可用介质集成的服务器、数据中心等数  
20 据存储设备。该可用介质可以是磁性介质，（例如，软盘，硬盘、磁带）、光介质（例如，DVD）或者半导体介质（例如固态硬盘 Solid State Disk（SSD））等。

25 通过以上的实施方式的描述，所属领域的技术人员可以清楚地了解到，为描述的方便和简洁，仅以上述各功能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模块完成，即将装置的内部结构划分成不同的功能模块，以完成以上描述的全部或者部分功能。

30 在本申请所提供的几个实施例中，应该理解到，所揭露的装置和方法，可以通过其它的方式实现。例如，以上所描述的装置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装置，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信连接可以通过一些接口，装置或单元的间接耦合或通信连接，可以是电性，机械或其它的形式。

35 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一个物理单元或多个物理单元，即可以位于一个地方，或者也可以分布到多个不同地方。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。

另外，在本申请各个实施例中的各功能单元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。

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，可以存储在一个可读取存储介质中。基于这样的理解，本申请实施例的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的全部或部分可以以软件产品的形式体现出来，该软件产品存储在一个存储介质中，包括若干指令用以使得一个设备（可以是单片机，芯片等）或处理器（processor）执行本申请各个实施例所述方法的全部或部分步骤。而前述的存储介质包括：U盘、移动硬盘、只读存储器（Read-Only Memory, ROM）、随机存取存储器（Random Access Memory, RAM）、磁碟或者光盘等各种可以存储程序代码的介质。

以上所述，仅为本申请的具体实施方式，但本申请的保护范围并不局限于此，任何在本申请揭露的技术范围内的变化或替换，都应涵盖在本申请的保护范围之内。因此，本申请的保护范围应以所述权利要求的保护范围为准。

15

20

## 权 利 要 求 书

1、一种数据查询方法，其特征在于，应用于包括  $n$  个目录块的只读文件系统中，每个目录块包括目录项区域和文件名区域， $n \geq 1$ ，所述数据查询方法包括：

从所述  $n$  个目录块中确定目标目录块，所述目标目录块的目录项区域包括  $m$  个目录项，所述目标目录块的文件名区域包括  $m$  个文件名，所述  $m$  个目录项与所述  $m$  个文件名一一对应，所述  $m$  个目录项和所述  $m$  个文件名均按照预设规则顺序排列，所述待访问文件名位于文件名范围中，所述文件名范围是由所述目标目录块的首个文件名与所述目标目录块的最后一个文件名组成的范围， $m \geq 1$ ；

步骤 A1：根据二分查找算法和目标目录块，确定当前第一集合和当前第二集合，所述当前第一集合包括所述  $m$  个文件名中连续的  $x$  个文件名，所述当前第二集合包括所述  $x$  个文件名、第一文件名以及第二文件名，所述第一文件名为排列于所述  $x$  个文件名中的首个文件名之前且与所述  $x$  个文件名中的首个文件名相邻的文件名，所述第二文件名为排列于所述  $x$  个文件名中的最后一个文件名之后且与所述  $x$  个文件名中的最后一个文件名相邻的文件名， $m \geq x \geq 1$ ；

步骤 B1：确定所述待访问文件名与所述当前第二集合中的文件名之间的第一公共前缀；

步骤 C1：从所述第一公共前缀之后的首位字符起，逐字符对比所述待访问文件名与第三文件名；其中，所述第三文件名为所述当前第一集合中第一预设位置的文件名；

若所述待访问文件名与所述第三文件名相同，则根据与所述第三文件名对应的目录项获取待访问文件的数据。

2、根据权利要求 1 所述的数据查询方法，其特征在于，所述数据查询方法还包括：

若所述待访问文件名与所述第三文件名不同，则根据所述二分查找算法、所述当前第一集合和所述当前第二集合，重新确定第一集合和第二集合，并根据重新确定的第一集合和重新确定的第二集合，执行所述步骤 B1 和所述步骤 C1，直到获取到所述待访问文件的数据或确定所述目标目录块未包括所述待访问文件名。

3、根据权利要求 2 所述的数据查询方法，其特征在于，所述预设规则为字典序顺序；

所述若所述待访问文件名与所述第三文件名不同，则根据所述二分查找算法、所述当前第一集合和所述当前第二集合，重新确定第一集合和第二集合，具体包括：

若所述待访问文件名的特征值小于所述第三文件名的特征值，确定所述重新确定的第一集合包括所述当前第一集合中位于所述第三文件名之前的所有文件名，所述重新确定的第二集合包括所述第一文件名、所述当前第一集合中位于所述第三文件名之前的所有文件名以及所述第三文件名；

或者，

若所述待访问文件名的特征值大于所述第三文件名的特征值，确定所述重新确定的第一集合包括所述当前第一集合中位于所述第三文件名之后的所有文件名，所述重新确定的第二集合包括所述第三文件名、所述当前第一集合中位于所述第三文件名之后的所有文件名以及所述第二文件名。

4、根据权利要求 1-3 中任意一项所述的数据查询方法，其特征在于，所述确定所

述待访问文件名与所述当前第二集合中的文件名之间的第一公共前缀，具体包括：

确定所述待访问文件名与所述第一文件名之间共有的第一前缀；

确定所述待访问文件名与所述第二文件名之间共有的第二前缀；

将所述第一前缀与所述第二前缀中长度最小的一个确定为所述第一公共前缀。

5 5、根据权利要求 1-4 中任意一项所述的数据查询方法，其特征在于，所述从所述 n 个目录块中确定目标目录块，具体包括：

步骤 A2：根据所述二分查找算法和所述 n 个目录块，确定当前第三集合和当前第四集合；其中，所述当前第三集合包括 p 个文件名，所述 p 个文件名包括 p 个目录块中每个目录块中第二预设位置的文件名，且所述当前第三集合中的文件名按照所述预设规则顺序排列，所述 p 个目录块为所述 n 个目录块中连续的目录块，所述当前第四集合包括所述 p 个文件名、第四文件名以及第五文件名，所述第四文件名为排列于所述 p 个文件名中的首个文件名之前且与所述 p 个文件名中的首个文件名相邻的文件名，所述第五文件名为排列于所述 p 个文件名中的最后一个文件名之后且与所述 p 个文件名中的最后一个文件名相邻的文件名， $1 \leq p \leq n$ ；

15 步骤 B2：确定所述待访问文件名与所述当前第四集合中的文件名之间的第二公共前缀；

步骤 C2：从所述第二公共前缀之后的首位字符起，逐字符对比所述待访问文件名与第六文件名；其中，所述第六文件名为所述当前第三集合中第三预设位置的文件名；

20 若所述待访问文件名与所述第六文件名相同，则确定所述第六文件名归属的目录块为所述目标目录块。

6、根据权利要求 5 所述的数据查询方法，其特征在于，若所述待访问文件名与所述第六文件名不同，所述数据查询方法还包括：

25 当  $2 \leq p \leq n$  时，根据所述当前第三集合、所述当前第四集合和所述二分查找算法，重新确定第三集合和第四集合，并根据重新确定的第三集合和重新确定的第四集合，执行所述步骤 B2 和所述步骤 C2；

当  $p=1$  时，根据所述当前第三集合包括的文件名确定所述目标目录块。

7、根据权利要求 6 所述的数据查询方法，其特征在于，所述预设规则为字典序顺序，对于每个目录块，所述第二预设位置的文件名为该目录块的首个文件名；

30 所述当  $p=1$  时，根据所述当前第三集合包括的文件名确定所述目标目录块，具体包括：

若所述待访问文件名的特征值大于所述当前第三集合中的文件名的特征值，则将所述当前第三集合中的文件名归属的目录块确定为所述目标目录块；

或者，

35 若所述待访问文件名的特征值小于所述当前第三集合中的文件名的特征值，则将位于所述当前第三集合中的文件名之前且与所述当前第三集合中的文件名相邻的文件名归属的目录块确定为所述目标目录块。

8、根据权利要求 6 或 7 所述的数据查询方法，其特征在于，所述预设规则为字典序顺序；

所述根据所述当前第三集合、所述当前第四集合和所述二分查找算法，重新确定

第三集合和第四集合，具体包括：

若所述待访问文件名的特征值小于所述第六文件名的特征值，确定所述重新确定的第三集合包括所述当前第三集合中位于所述第六文件名之前的所有文件名，所述重新确定的第四集合包括所述第四文件名、所述当前第三集合中位于所述第六文件名之前的所有文件名以及所述第六文件名；

或者，

若所述待访问文件名的特征值大于所述第六文件名的特征值，确定所述重新确定的第三集合包括所述当前第三集合中位于所述第六文件名之后的所有文件名，所述重新确定的第四集合包括所述第六文件名、所述当前第三集合中位于所述第六文件名之后的所有文件名以及所述第五文件名。

9、根据权利要求 5 所述的数据查询方法，其特征在于，所述预设规则为字典序顺序，对于每个目录块，所述第二预设位置的文件名为该目录块的首个文件名，所述数据查询方法还包括：

若所述第六文件名为所述当前第三集合中的首个文件名，所述待访问文件名的特征值小于所述第六文件名的特征值，则将位于所述第六文件名之前且与所述第六文件名相邻的文件名归属的目录块确定为所述目标目录块；

若所述第六文件名为所述当前第三集合中的最后一个文件名，所述待访问文件名的特征值大于所述第六文件名的特征值，则将所述第六文件名归属的目录块确定为所述目标目录块。

10、根据权利要求 6-9 中任意一项所述的数据查询方法，其特征在于，所述确定所述待访问文件名与所述当前第四集合中的文件名之间的第二公共前缀，具体包括：

确定所述待访问文件名与所述第四文件名之间共有的第三前缀；

确定所述待访问文件名与所述第五文件名之间共有的第四前缀；

将所述第三前缀与所述第四前缀中长度最小的一个确定为所述第二公共前缀。

11、根据权利要求 1-4 中任意一项所述的数据查询方法，其特征在于，所述 n 个目录块按照所述预设规则顺序排列，并采用完全二叉树方式存储，所述从所述 n 个目录块中确定所述目标目录块，具体包括：

步骤 A3：确定当前候选目录块和当前第三公共前缀；

步骤 B3：从所述当前第三公共前缀之后的首位字符起，逐字符对比待访问文件名与第 i 个文件名，所述第 i 个文件名为所述 n 个目录块中第 i 个目录块中第四预设位置的文件名， $0 \leq i < n$ ；

若所述待访问文件名与所述第 i 个文件名相同，则重新确定候选目录块为所述第 i 个文件名归属的目录块，并将重新确定的候选目录块确定为所述目标目录块。

12、根据权利要求 11 所述的数据查询方法，其特征在于，所述数据查询方法还包括：

若所述待访问文件名与所述第 i 个文件名不同，则重新确定第三公共前缀、候选目录块以及所述第 i 个文件名，重新确定的第 i 个文件名为所述 n 个目录块中第 j 个目录块中所述第四预设位置的文件名；其中，若所述待访问文件名的特征值大于所述第 i 个文件名的特征值， $j=2i+2$ ；若所述待访问文件名的特征值小于所述第 i 个文件名的

特征值,  $j=2i+1$ ,  $0 \leq i < j < n$ ;

根据重新确定的第三公共前缀、重新确定的候选目录块以及所述重新确定的第  $i$  个文件名, 重新执行所述步骤 B3, 直到确定出所述目标目录块。

5 13、根据权利要求 12 所述的数据查询方法, 其特征在于, 所述预设规则顺序为字典序, 所述第四预设位置的文件名为对应目录块中的首个文件名, 所述若所述待访问文件名与所述第  $i$  个文件名不同, 则重新确定候选目录块, 具体包括:

若所述待访问文件名的特征值大于所述第  $i$  个文件名的特征值, 则确定所述重新确定的候选目录块为所述第  $i$  个文件名归属的目录块;

10 若所述待访问文件名的特征值小于所述第  $i$  个文件名的特征值, 则确定所述重新确定的候选目录块为所述当前候选目录块。

14、根据权利要求 12 或 13 所述的数据查询方法, 其特征在于, 所述预设规则顺序为字典序, 所述第四预设位置的文件名为对应目录块中的首个文件名, 所述若所述待访问文件名与所述第  $i$  个文件名不同, 则重新确定第三公共前缀, 具体包括:

15 当所述待访问文件名的特征值大于所述第  $i$  个文件名的特征值时, 将当前第一目标前缀更新为所述待访问文件名与所述第  $i$  个文件名之间共有的前缀; 将更新后的第一目标前缀与当前第二目标前缀中长度最小的一个确定为所述重新确定的第三公共前缀;

或者,

20 当所述待访问文件名的特征值小于所述第  $i$  个文件名的特征值时, 将所述当前第二目标前缀更新为所述待访问文件名与所述第  $i$  个文件名之间共有的前缀; 将所述当前第一目标前缀与更新后的第二目标前缀中长度最小的一个确定为所述重新确定的第三公共前缀;

25 其中, 所述第一目标前缀的长度和所述第二目标前缀的长度的初始值均为零, 且所述第一目标前缀的长度和所述第二目标前缀的长度随着所述待访问文件名的特征值与所述第  $i$  个文件名的特征值的大小关系发生变化。

15、一种只读文件系统, 其特征在于, 所述只读文件系统的对象包括目录文件, 所述目录文件由  $n$  个目录块组成, 每个目录块均包括目录项区域和文件名区域, 所述目录项区域包括至少一个目录项, 所述文件名区域包括至少一个文件名;

30 对应同一目录块而言, 该目录块中目录项的数量与文件名的数量相同, 且该目录块中的所有目录项以及所有文件名均按照预设规则顺序排列。

16、根据权利要求 15 所述的只读文件系统, 其特征在于,

所述至少一个目录项中的每个目录项均包括索引号、文件类型和与该目录项对应的文件名在所归属的目录块的偏移量;

35 所述文件名区域与所述目录项区域相邻, 且所述文件名区域位于所述目录项区域之后。

17、一种数据查询装置, 其特征在于, 所述数据查询装置具备如权利要求 15 或 16 所述的只读文件系统, 所述数据查询装置包括:

处理单元, 用于从所述只读文件系统的  $n$  个目录块中确定目标目录块, 所述目标目录块的目录项区域包括  $m$  个目录项, 所述目标目录块的文件名区域包括  $m$  个文件名,

所述 m 个目录项与所述 m 个文件名一一对应，所述 m 个目录项和所述 m 个文件名均按照预设规则顺序排列，所述待访问文件名位于文件名范围中，所述文件名范围是由所述目标目录块的首个文件名与所述目标目录块的最后一个文件名组成的范围， $m \geq 1$ ；

所述处理单元，还用于执行步骤 A1、步骤 B1 以及步骤 C1；其中，

5 所述步骤 A1 为：根据二分查找算法和目标目录块，确定当前第一集合和当前第二集合，所述当前第一集合包括所述 m 个文件名中连续的 x 个文件名，所述当前第二集合包括所述 x 个文件名、第一文件名以及第二文件名，所述第一文件名为排列于所述 x 个文件名中的首个文件名之前且与所述 x 个文件名中的首个文件名相邻的文件名，所述第二文件名为排列于所述 x 个文件名中的最后一个文件名之后且与所述 x 个文件名中的最后一个文件名相邻的文件名， $m \geq x \geq 1$ ；

10 所述步骤 B1 为：确定所述待访问文件名与所述当前第二集合中的文件名之间的第一公共前缀；

15 所述步骤 C1 为：从所述第一公共前缀之后的首位字符起，逐字符对比所述待访问文件名与第三文件名；其中，所述第三文件名为所述当前第一集合中第一预设位置的文件名；

获取单元，用于若所述处理单元判断出所述待访问文件名与所述第三文件名相同，则根据与所述第三文件名对应的目录项获取待访问文件的数据。

18、根据权利要求 17 所述的数据查询装置，其特征在于，

20 所述处理单元，还用于若所述待访问文件名与所述第三文件名不同，则根据所述二分查找算法、所述当前第一集合和所述当前第二集合，重新确定第一集合和第二集合，并根据重新确定的第一集合和重新确定的第二集合，执行所述步骤 B1 和所述步骤 C1，直到所述获取单元获取到所述待访问文件的数据或所述处理单元确定所述目标目录块未包括所述待访问文件名。

25 19、根据权利要求 18 所述的数据查询装置，其特征在于，所述预设规则为字典序顺序，所述处理单元具体用于：

若所述待访问文件名的特征值小于所述第三文件名的特征值，确定所述重新确定的第一集合包括所述当前第一集合中位于所述第三文件名之前的所有文件名，所述重新确定的第二集合包括所述第一文件名、所述当前第一集合中位于所述第三文件名之前的所有文件名以及所述第三文件名；

30 或者，

若所述待访问文件名的特征值大于所述第三文件名的特征值，确定所述重新确定的第一集合包括所述当前第一集合中位于所述第三文件名之后的所有文件名，所述重新确定的第二集合包括所述第三文件名、所述当前第一集合中位于所述第三文件名之后的所有文件名以及所述第二文件名。

35 20、根据权利要求 17-19 中任意一项所述的数据查询装置，其特征在于，所述处理单元具体用于：

确定所述待访问文件名与所述第一文件名之间共有的第一前缀；

确定所述待访问文件名与所述第二文件名之间共有的第二前缀；

将所述第一前缀与所述第二前缀中长度最小的一个确定为所述第一公共前缀。

21、根据权利要求 17-20 中任意一项所述的数据查询装置，其特征在于，  
所述处理单元，还用于执行步骤 A2、步骤 B2 以及步骤 C2；其中，

所述步骤 A2 为：根据所述二分查找算法和所述  $n$  个目录块，确定当前第三集合  
和当前第四集合；其中，所述当前第三集合包括  $p$  个文件名，所述  $p$  个文件名包括  $p$   
个目录块中每个目录块中第二预设位置的文件名，且所述当前第三集合中的文件名按  
照所述预设规则顺序排列，所述  $p$  个目录块为所述  $n$  个目录块中连续的目录块，所述  
当前第四集合包括所述  $p$  个文件名、第四文件名以及第五文件名，所述第四文件名为  
排列于所述  $p$  个文件名中的首个文件名之前且与所述  $p$  个文件名中的首个文件名相邻  
的文件名，所述第五文件名为排列于所述  $p$  个文件名中的最后一个文件名之后且与所  
述  $p$  个文件名中的最后一个文件名相邻的文件名， $1 \leq p \leq n$ ；

所述步骤 B2 为：确定所述待访问文件名与所述当前第四集合中的文件名之间的  
第二公共前缀；

所述步骤 C2 为：从所述第二公共前缀之后的首位字符起，逐字符对比所述待访  
问文件名与第六文件名；其中，所述第六文件名为所述当前第三集合中第三预设位置  
的文件名；

所述处理单元，还用于若所述待访问文件名与所述第六文件名相同，则确定所述  
第六文件名归属的目录块为所述目标目录块。

22、根据权利要求 21 所述的数据查询装置，其特征在于，所述处理单元具体用于：

当  $2 \leq p \leq n$  时，根据所述当前第三集合、所述当前第四集合和所述二分查找算法，  
重新确定第三集合和第四集合，并根据重新确定的第三集合和重新确定的第四集合，  
执行所述步骤 B2 和所述步骤 C2；

当  $p=1$  时，根据所述当前第三集合包括的文件名确定所述目标目录块。

23、根据权利要求 22 所述的数据查询装置，其特征在于，所述预设规则为字典序  
顺序，对于每个目录块，所述第二预设位置的文件名为该目录块的首个文件名；当  $p=1$   
时，所述处理单元具体用于：

若所述待访问文件名的特征值大于所述当前第三集合中的文件名的特征值，则将  
所述当前第三集合中的文件名归属的目录块确定为所述目标目录块；

或者，

若所述待访问文件名的特征值小于所述当前第三集合中的文件名的特征值，则将  
位于所述当前第三集合中的文件名之前且与所述当前第三集合中的文件名相邻的文件  
名归属的目录块确定为所述目标目录块。

24、根据权利要求 22 或 23 所述的数据查询装置，其特征在于，所述预设规则为  
字典序顺序；所述处理单元具体用于：

若所述待访问文件名的特征值小于所述第六文件名的特征值，确定所述重新确定  
的第三集合包括所述当前第三集合中位于所述第六文件名之前的所有文件名，所述重  
新确定的第四集合包括所述第四文件名、所述当前第三集合中位于所述第六文件名之  
前的所有文件名以及所述第六文件名；

或者，

若所述待访问文件名的特征值大于所述第六文件名的特征值，确定所述重新确定

的第三集合包括所述当前第三集合中位于所述第六文件名之后的所有文件名，所述重新确定的第四集合包括所述第六文件名、所述当前第三集合中位于所述第六文件名之后的所有文件名以及所述第五文件名。

5 25、根据权利要求 21 所述的数据查询装置，其特征在于，所述预设规则为字典序顺序，对于每个目录块，所述第二预设位置的文件名为该目录块的首个文件名，所述处理单元还用于：

若所述第六文件名为所述当前第三集合中的首个文件名，所述待访问文件名的特征值小于所述第六文件名的特征值，则将位于所述第六文件名之前且与所述第六文件名相邻的文件名归属的目录块确定为所述目标目录块；

10 若所述第六文件名为所述当前第三集合中的最后一个文件名，所述待访问文件名的特征值大于所述第六文件名的特征值，则将所述第六文件名归属的目录块确定为所述目标目录块。

26、根据权利要求 22-25 中任意一项所述的数据查询装置，其特征在于，所述处理单元具体用于：

15 确定所述待访问文件名与所述第四文件名之间共有的第三前缀；

确定所述待访问文件名与所述第五文件名之间共有的第四前缀；

将所述第三前缀与所述第四前缀中长度最小的一个确定为所述第二公共前缀。

27、根据权利要求 17-20 中任意一项所述的数据查询装置，其特征在于，所述 n 个目录块按照所述预设规则顺序排列，并采用完全二叉树方式存储；

20 所述处理单元，还用于执行步骤 A3 和步骤 B3；其中，

所述步骤 A3 为：确定当前候选目录块和当前第三公共前缀；

所述步骤 B3 为：从所述当前第三公共前缀之后的首位字符起，逐字符对比待访问文件名与第 i 个文件名，所述第 i 个文件名为所述 n 个目录块中第 i 个目录块中第四预设位置的文件名， $0 \leq i < n$ ；

25 所述处理单元，还用于若所述待访问文件名与所述第 i 个文件名相同，则重新确定候选目录块为所述第 i 个文件名归属的目录块，并将重新确定的候选目录块确定为所述目标目录块。

28、根据权利要求 27 所述的数据查询装置，其特征在于，

30 所述处理单元，还用于若所述待访问文件名与所述第 i 个文件名不同，则重新确定第三公共前缀、候选目录块以及所述第 i 个文件名，重新确定的第 i 个文件名为所述 n 个目录块中第 j 个目录块中所述第四预设位置的文件名；其中，若所述待访问文件名的特征值大于所述第 i 个文件名的特征值， $j=2i+2$ ；若所述待访问文件名的特征值小于所述第 i 个文件名的特征值， $j=2i+1$ ， $0 \leq i < j < n$ ；

35 所述处理单元，还用于根据重新确定的第三公共前缀、重新确定的候选目录块以及所述重新确定的第 i 个文件名，重新执行所述步骤 B3，直到确定出所述目标目录块。

29、根据权利要求 28 所述的数据查询装置，其特征在于，所述预设规则顺序为字典序，所述第四预设位置的文件名为对应目录块中的首个文件名；所述处理单元具体用于：

若所述待访问文件名的特征值大于所述第 i 个文件名的特征值，则确定所述重新

确定的候选目录块为所述第 i 个文件名归属的目录块;

若所述待访问文件名的特征值小于所述第 i 个文件名的特征值, 则确定所述重新确定的候选目录块为所述当前候选目录块。

5 30、根据权利要求 28 或 29 所述的数据查询装置, 其特征在于, 所述预设规则顺序为字典序, 所述第四预设位置的文件名为对应目录块中的首个文件名; 所述处理单元具体用于:

10 当所述待访问文件名的特征值大于所述第 i 个文件名的特征值时, 将当前第一目标前缀更新为所述待访问文件名与所述第 i 个文件名之间共有的前缀; 将更新后的第一目标前缀与当前第二目标前缀中长度最小的一个确定为所述重新确定的第三公共前缀;

或者,

15 当所述待访问文件名的特征值小于所述第 i 个文件名的特征值时, 将所述当前第二目标前缀更新为所述待访问文件名与所述第 i 个文件名之间共有的前缀; 将所述当前第一目标前缀与更新后的第二目标前缀中长度最小的一个确定为所述重新确定的第三公共前缀;

其中, 所述第一目标前缀的长度和所述第二目标前缀的长度的初始值均为零, 且所述第一目标前缀的长度和所述第二目标前缀的长度随着所述待访问文件名的特征值与所述第 i 个文件名的特征值的大小关系发生变化。

20 31、一种终端, 其特征在于, 所述终端包括: 一个或多个处理器、存储器、通信接口;

所述存储器、所述通信接口与所述一个或多个处理器耦合; 所述存储器用于存储计算机程序代码, 所述计算机程序代码包括指令, 当所述一个或多个处理器执行所述指令时, 所述终端执行如权利要求 1-14 中任意一项所述的数据查询方法。

25 32、一种计算机可读存储介质, 所述计算机可读存储介质中存储有指令, 其特征在于, 当所述指令在终端上运行时, 使得所述终端执行如权利要求 1-14 中任意一项所述的数据查询方法。

30 33、一种包含指令的计算机程序产品, 其特征在于, 当所述计算机程序产品在终端上运行时, 使得所述终端执行如权利要求 1-14 中任意一项所述的数据查询方法。

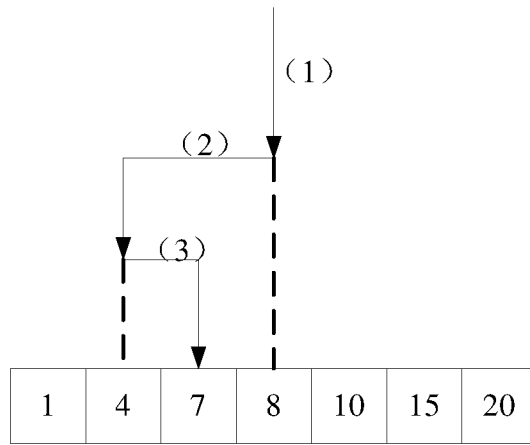


图 1

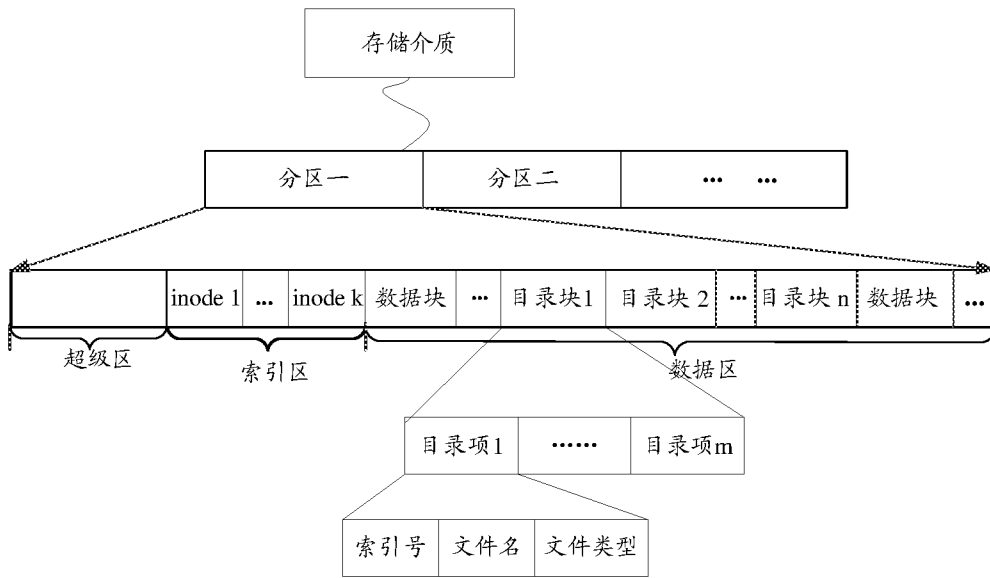


图 2

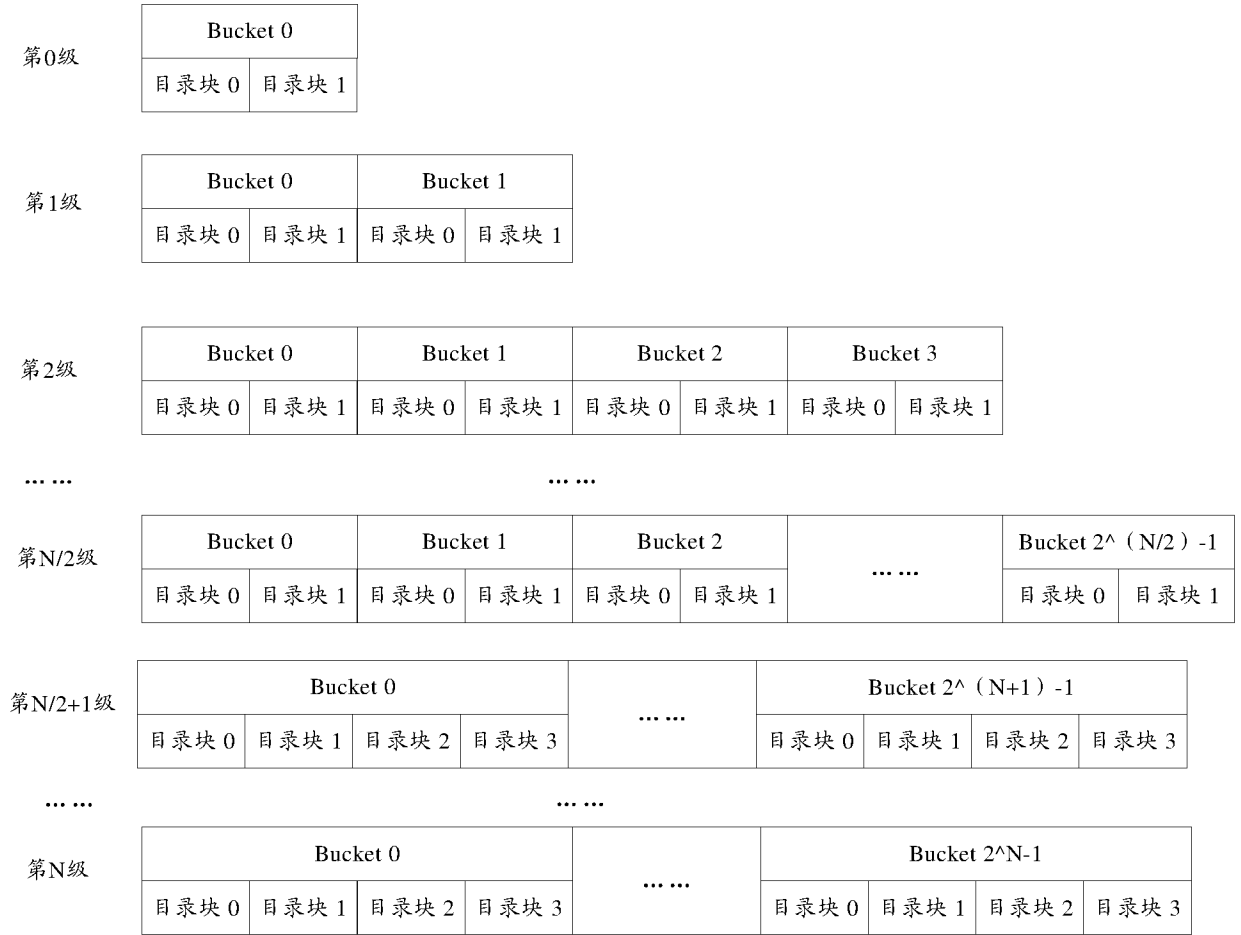


图 3

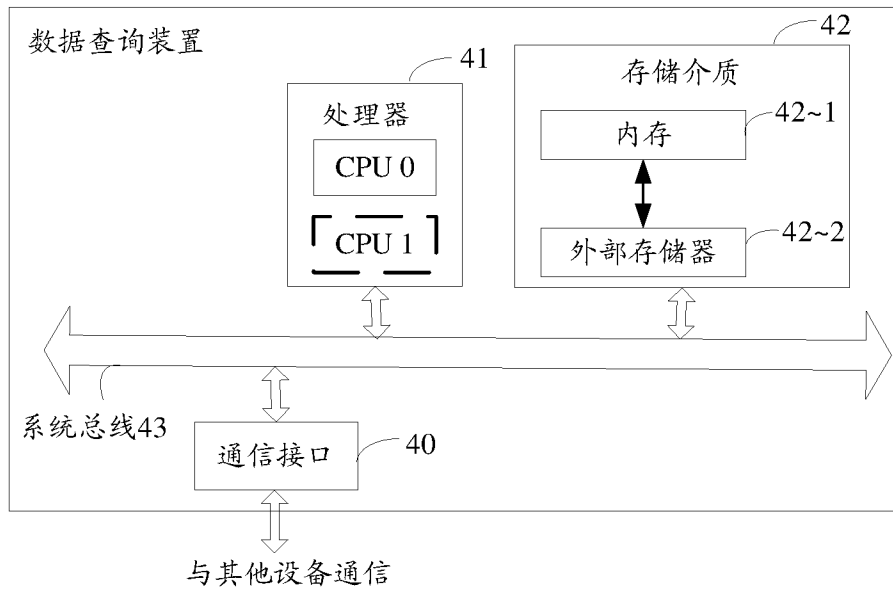


图 4

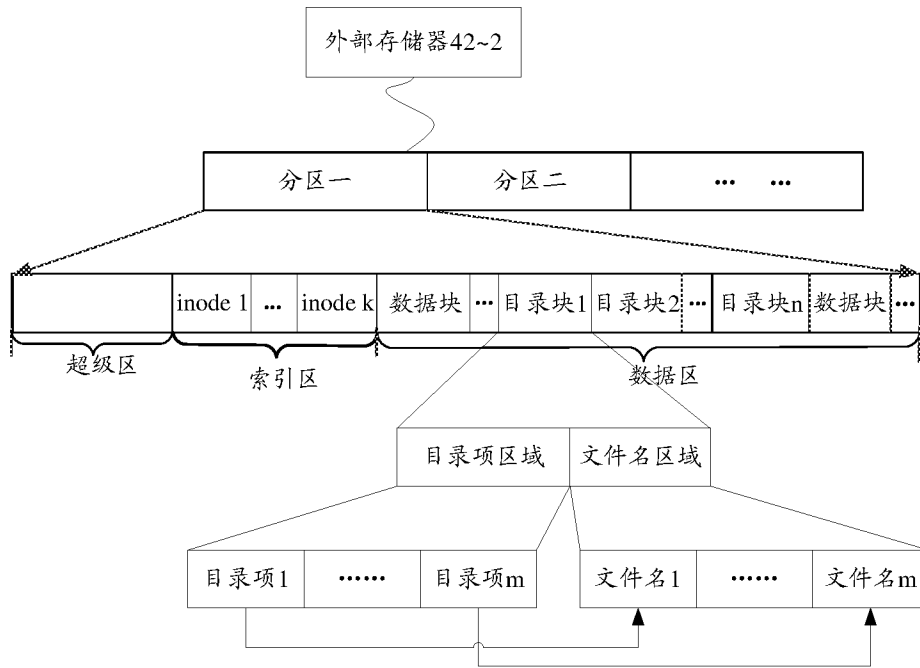


图 5

目录块



图 6

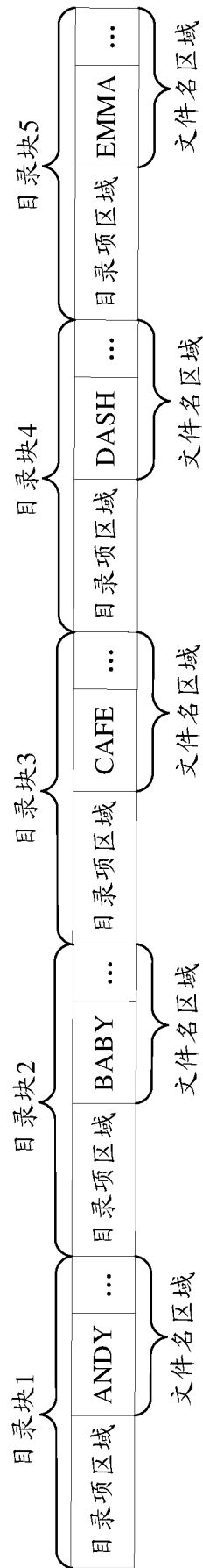


图7

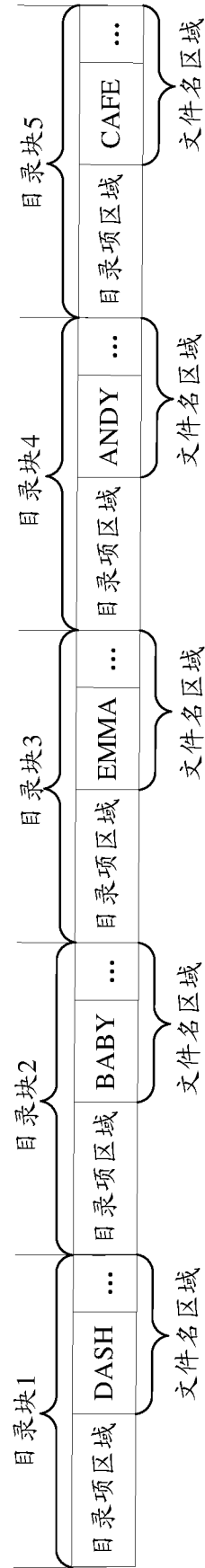


图 8

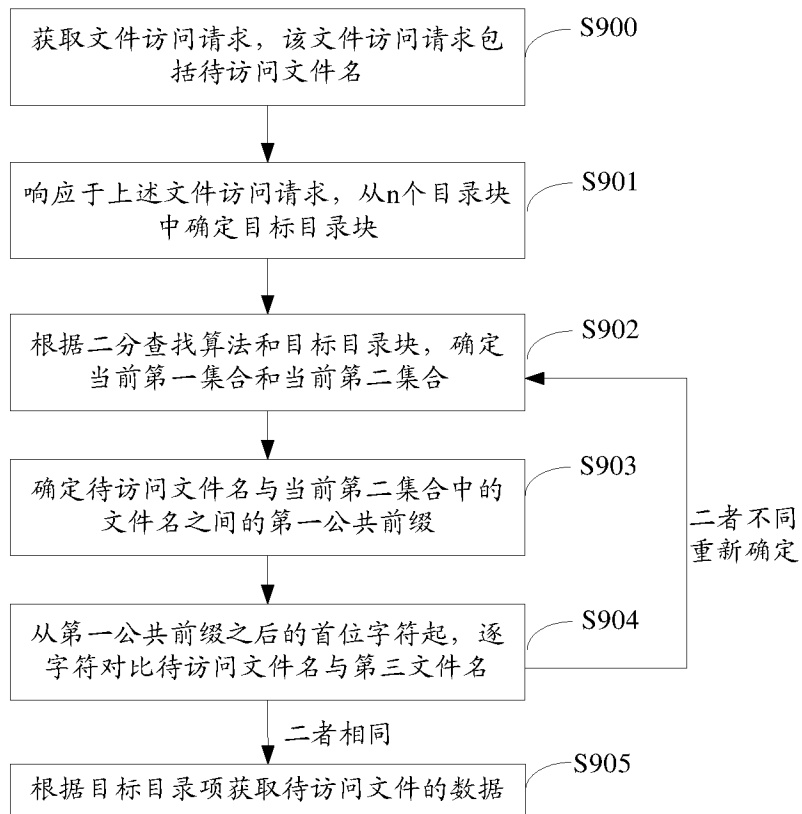


图 9

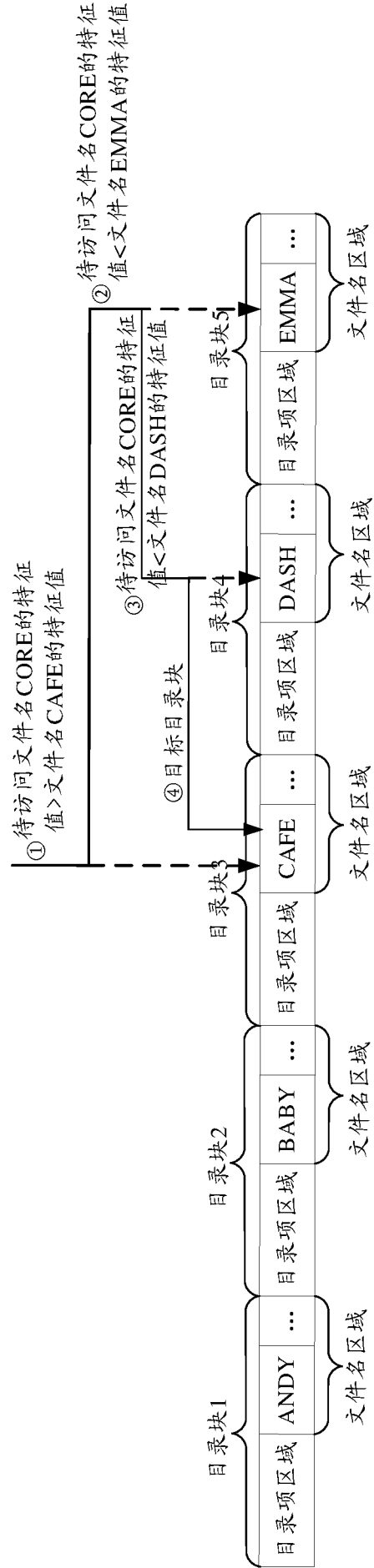


图10

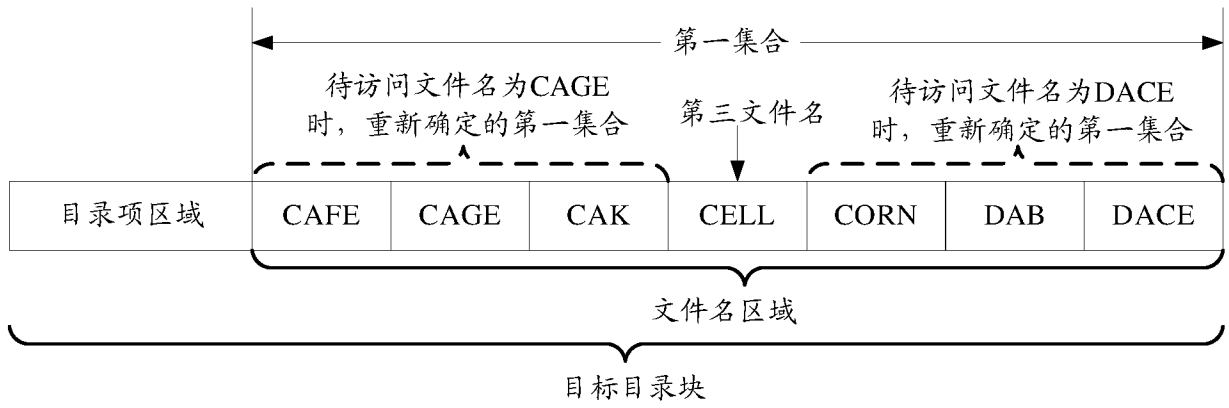


图 11

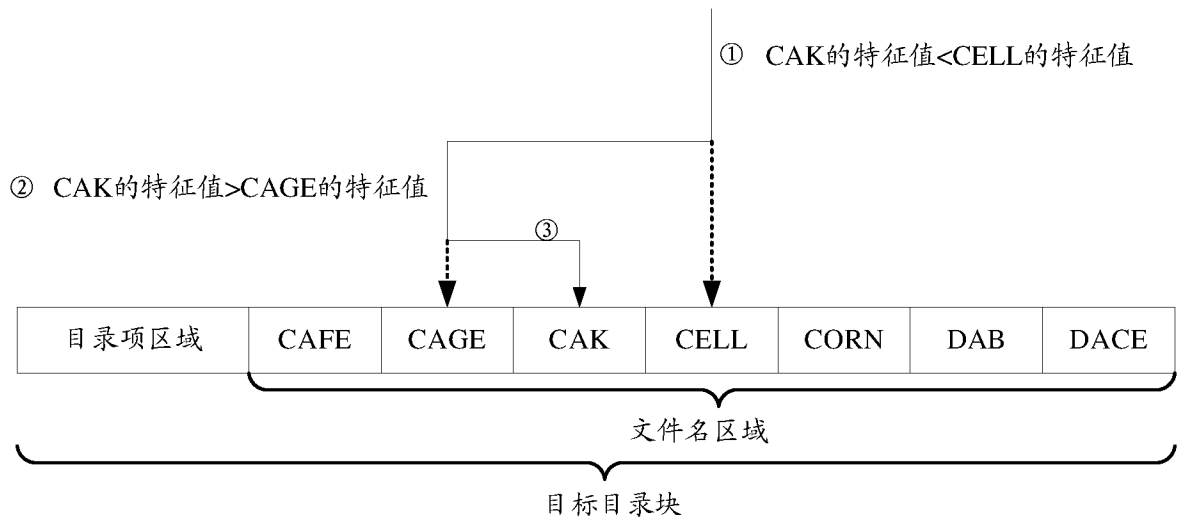


图 12

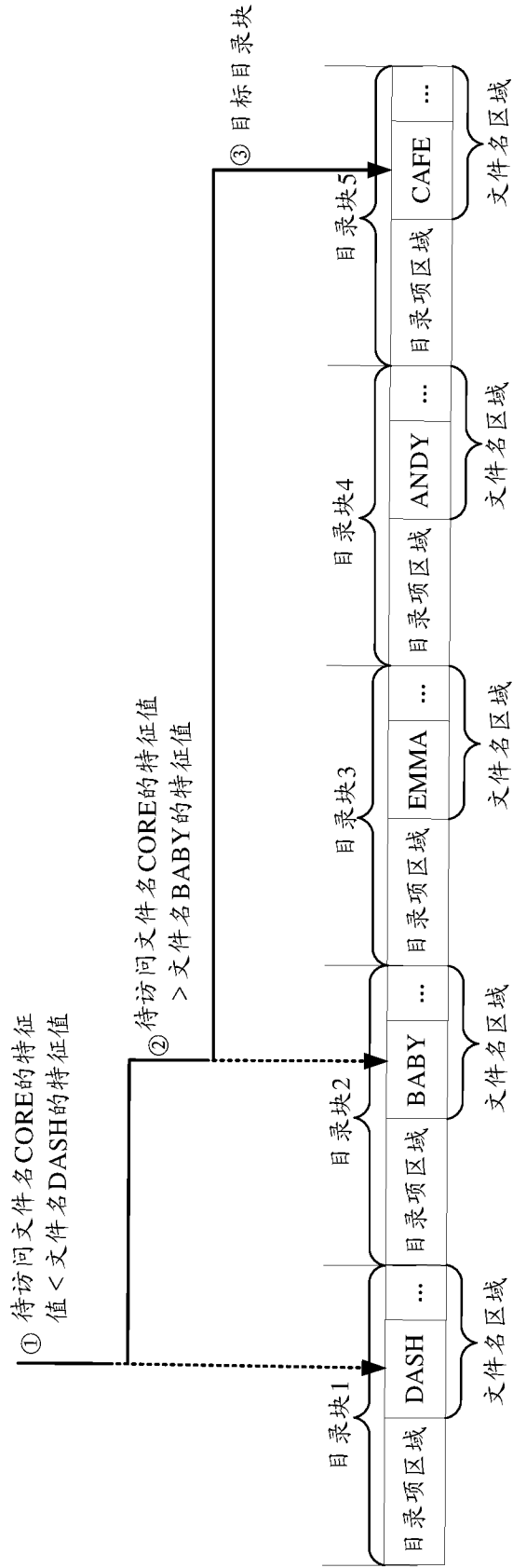


图13

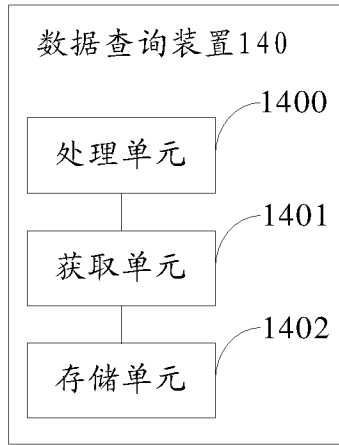


图 14

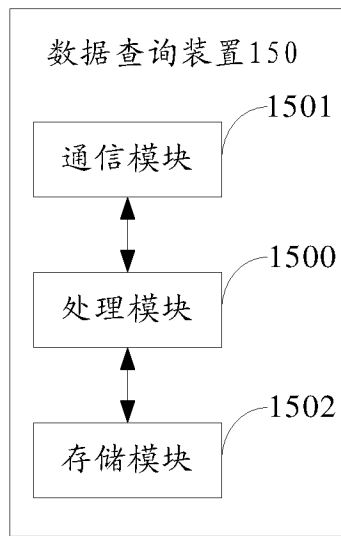


图 15

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2018/075300

<b>A. CLASSIFICATION OF SUBJECT MATTER</b> G06F 17/30(2006.01)i  According to International Patent Classification (IPC) or to both national classification and IPC		
<b>B. FIELDS SEARCHED</b> 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) CNPAT, CNKI, WPI, EPODOC: 文件, 目录, 块, 文件名, 查询, 查找, 检索, 效率, 范围, 连续, 相邻, 哈希, 字典, 相等, 相同, file, directory, block, file name, retrieve, efficiency, scope, continuous, adjacent, hash, dictionary, same		
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 104050251 A (IPANEL. TV INC.) 17 September 2014 (2014-09-17) claim 1, and description, paragraphs [0023]-[0047]	15, 16
A	CN 104050251 A (IPANEL. TV INC.) 17 September 2014 (2014-09-17) claim 1, and description, paragraphs [0023]-[0047]	1-14, 17-33
A	CN 104537016 A (HUAWEI TECHNOLOGIES CO., LTD.) 22 April 2015 (2015-04-22) entire document	1-33
A	CN 103942205 A (SHENZHEN TENCENT COMPUTER SYSTEMS CO., LTD.) 23 July 2014 (2014-07-23) entire document	1-33
A	US 2013159250 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 20 June 2013 (2013-06-20) entire document	1-33
A	庞海飞等 (PANG, Haifei et al.). "基于Linux文件检索方法的改进研究 (Research on Improvement of File Retrieval Method Based on Linux)" <i>火力与指挥控制 (Fire Control &amp; Command Control)</i> , Vol. 42, No. 2, 28 February 2017 (2017-02-28), pp. 145-148	1-33
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> 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 <b>22 October 2018</b>		Date of mailing of the international search report <b>09 November 2018</b>
Name and mailing address of the ISA/CN <b>State Intellectual Property Office of the P. R. China (ISA/ CN) No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088 China</b> Facsimile No. <b>(86-10)62019451</b>		Authorized officer   Telephone No.

**INTERNATIONAL SEARCH REPORT**  
**Information on patent family members**

International application No.

**PCT/CN2018/075300**

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	104050251	A	17 September 2014	None			
CN	104537016	A	22 April 2015	None			
CN	103942205	A	23 July 2014	US	2014244606	A1	28 August 2014
				WO	2014110940	A1	24 July 2014
				HK	1199527	A0	03 July 2015
US	2013159250	A1	20 June 2013	US	2010174690	A1	08 July 2010
				CN	101819577	A	01 September 2010

<b>A. 主题的分类</b> G06F 17/30 (2006.01)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>B. 检索领域</b> 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)) CNPAT, CNKI, WPI, EPODOC: 文件, 目录, 块, 文件名, 查询, 查找, 检索, 效率, 范围, 连续, 相邻, 哈希, 字典, 相等, 相同, file, directory, block, file name, retrieve, efficiency, scope, continuous, adjacent, hash, dictionary, same		
<b>C. 相关文件</b>		
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求
X	CN 104050251 A (深圳市茁壮网络股份有限公司) 2014年 9月 17日 (2014 - 09 - 17) 权利要求1, 说明书第[0023]-[0047]段	15-16
A	CN 104050251 A (深圳市茁壮网络股份有限公司) 2014年 9月 17日 (2014 - 09 - 17) 权利要求1, 说明书第[0023]-[0047]段	1-14, 17-33
A	CN 104537016 A (华为技术有限公司) 2015年 4月 22日 (2015 - 04 - 22) 全文	1-33
A	CN 103942205 A (深圳市腾讯计算机系统有限公司) 2014年 7月 23日 (2014 - 07 - 23) 全文	1-33
A	US 2013159250 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 2013年 6月 20日 (2013 - 06 - 20) 全文	1-33
A	庞海飞 等. "基于Linux文件检索方法的改进研究" 火力与指挥控制, 第42卷, 第2期, 2017年 2月 28日 (2017 - 02 - 28), 第145-148页	1-33
<input type="checkbox"/> 其余文件在C栏的续页中列出。		
<input checked="" type="checkbox"/> 见同族专利附件。		
* 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	国际检索报告邮寄日期	
2018年 10月 22日	2018年 11月 9日	
ISA/CN的名称和邮寄地址	受权官员	
中华人民共和国国家知识产权局(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	武文琛  电话号码 86-(10)-53961312	

国际检索报告  
关于同族专利的信息

国际申请号  
PCT/CN2018/075300

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	104050251	A	2014年 9月 17日	无			
CN	104537016	A	2015年 4月 22日	无			
CN	103942205	A	2014年 7月 23日	US	2014244606	A1	2014年 8月 28日
				WO	2014110940	A1	2014年 7月 24日
				HK	1199527	A0	2015年 7月 3日
US	2013159250	A1	2013年 6月 20日	US	2010174690	A1	2010年 7月 8日
				CN	101819577	A	2010年 9月 1日

表 PCT/ISA/210 (同族专利附件) (2015年1月)