(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
2024 年 3 月 14 日 (14.03.2024)

WIPO | PCT

(10) 国际公布号
WO 2024/051654 A1

(51) 国际专利分类号:
G06F 16/21 (2019.01)

(21) 国际申请号: PCT/CN2023/116818

(22) 国际申请日: 2023 年 9 月 4 日 (04.09.2023)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202211095247.7 2022年9月5日 (05.09.2022) CN

(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:覃鸿巍**(QIN, Hongwei)**; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 裘杰**(QIU, Jie)**; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 钱璟**(QIAN, Jing)**; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 姜俊超**(JIANG, Junchao)**; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 汪坷**(WANG, Ke)**; 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, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ,

(54) **Title:** FILE PROCESSING METHOD AND ELECTRONIC DEVICE

(54) 发明名称: 文件处理方法及电子设备

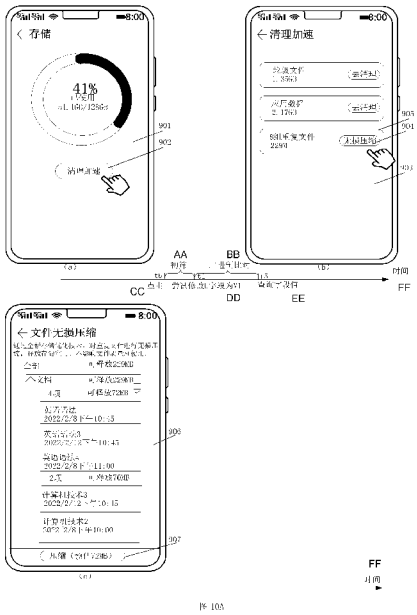


图 10A

AA Preliminary screening
BB Binary alignment
CC Tap
DD Attempt to modify a field U to V1
EE Query a field value
FF Time

(57) **Abstract:** A file processing method and an electronic device, which relate to the technical field of terminals and can reduce the occupation of storage space. The method is applied to an electronic device, the electronic device stores an indication field of a first file, an initial value of the indication field is a first value, and the indication field is modified into a second value by the electronic device when the first file is modified. The method comprises: at a first moment, attempting to modify the value of the indication field of the first file to a third value, the third value being different from the first value, and the third value being different from the second value; if the value of the indication field of the first file is successfully modified to the third value, acquiring the value of the indication field at a second moment; and according to the value of the indication field of the first file at the second moment, determining whether the first file is modified or not in a time period from the first moment to the second moment.

(57) 摘要: 一种文件处理方法及电子设备，涉及终端技术领域，可以降低对存储空间的占用，该方法应用于电子设备，所述电子设备存储有第一文件的指示字段，所述指示字段的初始值为第一值，所述指示字段在所述第一文件被修改时被所述电子设备修改为第二值；所述方法包括：在第一时刻，尝试将所述第一文件的指示字段的值修改为第三值；所述第三值与所述第一值不同；所述第三值与所述第二值不同；若成功将所述第一文件的指示字段的值修改为所述第三值，则在第二时刻，获取所述指示字段的值；根据所述第一文件的指示字段在所述第二时刻的值，判断在所述第一时刻至所述第二时刻的时段内，所述第一文件是否被修改。

WO 2024/051654 A1

[见续页]

LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, 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, 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, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, 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, ME, 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))。

文件处理方法及电子设备

本申请要求于 2022 年 09 月 05 日提交国家知识产权局、申请号为 202211095247.7、发明名称为“文件处理方法及电子设备”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请涉及终端技术领域，尤其涉及文件处理方法及电子设备。

背景技术

目前，很多电子设备利用操作系统中的文件系统管理文件。随着电子设备中存储的文件的数量剧增，文件系统出现了文件管理瓶颈。在一些场景中，电子设备中的文件可占用较多的存储空间，导致电子设备的可用存储空间较低。

发明内容

本申请提供文件处理方法及电子设备，可以高效的管理文件。

为了实现上述目的，本申请实施例提供了以下技术方案：

第一方面，提供一种文件处理方法，应用于电子设备或支持电子设备功能的组件（比如芯片系统），电子设备存储有第一文件的指示字段，指示字段的初始值为第一值，指示字段在第一文件被修改时被电子设备修改为第二值；方法包括：

在第一时刻，尝试将第一文件的指示字段的值修改为第三值；第三值与第一值不同；第三值与第二值不同；

若成功将第一文件的指示字段的值修改为第三值，则在第二时刻，获取指示字段的值；

根据第一文件的指示字段在第二时刻的值，判断在第一时刻至第二时刻的时段内，第一文件是否被修改。

示例性的，如图 13，在 t1 时刻（第一时刻的一个示例），手机将英语语法 3（第一文件的一个示例）的 U 字段（指示字段的一个示例）的值成功修改为 V1（第三值的一个示例），并在 t5 时刻（第二时刻的一个示例），查询 U 字段的值。手机可以根据英语语法 3 的 U 字段在 t5 时刻的值，判断在 t1-t5 时段内，英语语法 3 是否被修改。

本申请实施例中，可以通过第一文件的指示字段的值判断第一文件的修改状态，进而后续可以根据第一文件的修改状态，对第一文件进行相应管理。其中，电子设备无需在第一时刻至第二时刻期间持续监听第一文件的修改状态，而是可以第一时刻、第二时刻之外的其他时刻，中止监听，以降低电子设备用于监听的资源消耗。比如，在第一时刻，电子设备可以通过第一应用尝试将第一文件的指示字段的值修改为 V1，在第一时刻之后，电子设备可以退出第一应用。第二时刻，电子设备再次打开第一应用，获取第一文件的指示字段的值，并根据该指示字段的值，判断第一文件在第一时刻至第二时刻期间是否被修改。可见，本申请实施例中，第一应用无需常驻的情况下，电子设备也可以判断第一文件在某个时段内是否被修改，能够降低对存储空间的占用。

并且，本方案与系统时间无关，因此，即便用户修改系统时间，也不会导致误判。

在一种可能的设计中，尝试将第一文件的指示字段的值修改为第三值，包括：

若满足第一条件，则将指示字段的值修改为第三值；在第一条件下，电子设备的文件系统能够检测到在第一时刻至第二时刻的时段内，针对第一文件的修改操作。

如此，一旦文件在第一时刻至第二时刻期间内被修改，则文件系统能够确保感知到该修改，并将该文件的指示字段的值修改为非 V1 值（比如第二值 V2）。后续，电子设备能够查询该文件的指示字段的值，如果指示字段的值被修改为非 V1 的值，则电子设备确定该文件在第一时刻至第二时刻期间被修改，反之，如果指示字段的值仍为 V1，则确定该文件在第一时刻至第二时刻期间未被修改。

在一种可能的设计中，在第一时刻之后，方法还包括：

检测到第一文件被修改，将第一文件的指示字段的值修改为第二值。

根据第一文件的指示字段在第二时刻的值，判断在第一时刻至第二时刻的时段内，第一文件是否被修改，包括：

若第一文件的指示字段在第二时刻的值为第二值，则确定在第一时刻至第二时刻的时段内，第一文件被修改，或者，若第一文件的指示字段在第二时刻的值不是第三值，则确定在第一时刻至第

二时刻的时段内，第一文件被修改。

示例性的，如图 12，t1 时刻（第一时刻的一个示例）之后，在 t3 时刻，检测到英语语法 2 被修改（该修改为 t1 时刻之后的首次写内存），手机将英语语法 2 的 U 字段的值修改为 V2（第二值的一个示例）。在 t5 时刻（第二时刻的一个示例），手机查询英语语法 2 的 U 字段的值，获知英语语法 2 的 U 字段在 t5 时刻的值为 V2（不再是 V1），则确定在 t1-t5 时段内，英语语法 2 被修改。

在一种可能的设计中，根据第一文件的指示字段在第二时刻的值，判断在第一时刻至第二时刻的时段内，第一文件是否被修改，包括：

若第一文件的指示字段在第二时刻的值为第三值，则确定在第一时刻至第二时刻的时段内，第一文件未被修改，或者，若第一文件的指示字段在第二时刻的值不是第二值，则确定在第一时刻至第二时刻的时段内，第一文件未被修改。

比如，如图 13，手机管家应用扫描完重复文件（比如二进制比对）之后，向用户显示扫描结果之前，在 t5 时刻，手机管家应用查询英语语法 3 这一文件的 U 字段的字段值，获知字段值为 V1（第三值的一个示例）。由于英语语法 3 的 U 字段在 t5 时刻的字段值与在 t1 时刻的字段值相同，均为 V1，则手机管家应用可以确定在 t1 时刻至 t5 时刻期间内，英语语法 3 这一文件未被修改过。

在一种可能的设计中，尝试将第一文件的指示字段的值修改为第三值，包括：若不满足第一条件，则不将指示字段修改为第三值；

方法还包括：在不能成功将第一文件的指示字段的值修改为第三值时，确定在第一时刻至第二时刻的时段内，第一文件被修改。

示例性的，如图 15，假设 t1 时刻之前，手机针对候选重复文件中的计算机技术 1 执行过写内存操作，则重复文件的初筛完成后，在 t1 时刻，由于不满足第一条件，手机管家应用无法成功将计算机技术 1 的 U 字段的值修改为 V1（U 字段的值仍为 V0）。此种情况下，手机管家应用将计算机技术 1 视为已被修改过的文件（不再是重复文件），不再对计算机技术 1 执行去重，以免执行误去重。

在一种可能的设计中，第一条件包括如下一项或多项条件：在第一时刻之前，第一文件未被映射到内存；在第一时刻之前，第一文件被映射到内存，但电子设备未对内存执行过写入操作。

对于某个文件来说，如果第一时刻前未将该文件映射到内存，则后续手机一旦在第一时刻至第二时刻期间内对该文件进行一次或多次写内存操作，其中的首次写内存操作必定会使得手机执行文件系统的相关代码。也就是说，第一时刻至第二时刻期间内的至少首次写内存操作会被文件系统检测到。如此，文件系统可以至少根据第一时刻至第二时刻期间内的至少首次写内存操作，确定该文件被修改，并据此将该文件的 U 字段的值修改为 V2。后续，应用程序可以查询该文件的 U 字段的值，并根据 U 字段的值，判断该文件在第一时刻至第二时刻期间是否被修改。可见，本申请实施例中，文件系统无需检测每次写内存操作，即便文件系统存在漏检写内存操作的现象，文件系统也能根据检测到的一段时间内的至少首次写内存操作，确定文件被修改过。如此，能够提升电子设备的检测精度。

同理，对于某个文件来说，若在第一时刻之前，文件被映射到内存，但电子设备未对内存执行过写入操作，则后续手机一旦在第一时刻至第二时刻期间内对该文件进行一次或多次写内存操作，其中的首次写内存操作必定会使得手机执行文件系统的相关代码。即，一旦文件在第一时刻至第二时刻期间被修改，该修改必定会被文件系统感知到。

基于此原理，若在第一时刻之前，文件被映射到内存，但电子设备未对内存执行过写入操作，则电子设备可以在第一时刻，将该文件的 U 字段的值修改为 V1。后续，一旦该文件在第一时刻至第二时刻期间被修改，则文件的 U 字段值会被文件系统修改为非 V1 值（比如 V2），电子设备可以根据文件的 U 字段值，判断文件在第一时刻至第二时刻期间内是否被修改。

可见，本方案，能够降低 mmap 场景的误判。

在一种可能的设计中，方法还包括：

扫描得到电子设备中的重复文件，重复文件包括第一文件和第二文件；第一文件存储在第一存储空间，第二文件存储在第二存储空间；

确定第一时刻至第二时刻的时段内，第一文件、第二文件未被修改；

显示第一界面；第一界面包括第一文件的信息以及第二文件的信息。

比如，扫描得到英语语法 1-4 为一组重复文件，这四个重复文件分别存储在存储空间 1-4 中，若确定英语语法 2 在第一时刻至第二时刻期间内被修改，英语语法 1、3、4 在第一时刻至第二时刻

期间内未被修改，则意味着英语语法 1、3、4 仍构成重复文件。如图 10A 的（c），手机可显示第一界面 906，第一界面 906 包括英语语法 1、3、4 这组重复文件的信息。

在一种可能的设计中，检测到用户在第一界面中输入的第一操作；

将第一文件以及第二文件压缩至同一存储空间。

比如，如图 10A 的（c），检测到用户点击压缩按钮 907 的操作（第一操作的一个示例），手机对用户选择的一组重复文件（英语语法 1、3、4）执行压缩，压缩后的英语语法 1、3、4 指向同一存储空间。如此，可以释放存储空间。

在一种可能的设计中，在扫描电子设备中的重复文件之后，方法还包括：

存储第一文件的标识和第二文件的标识。比如，图 11 中，电子设备将重复文件的标识存储在重复文件信息库中。

在一种可能的设计中，第一文件的标识包括第一文件的第一索引节点编号；第二文件的标识包括第二文件的第二索引节点编号。

如此，后续，用户打开手机管家，并点击“清理加速”按钮时，手机可以根据存储的重复文件的标识，获知第一文件和第二文件历史上是重复文件。手机可以查询第一文件和第二文件的 U 字段的值，以便确定第一文件、第二文件是否仍构成重复文件。手机无需重新扫描第一文件、第二文件的内容，有助于降低手机的运行消耗。

在一种可能的设计中，根据第一文件的指示字段在第二时刻的值，判断在第一时刻至第二时刻的时段内，第一文件是否被修改，包括：

若第一文件的指示字段在第二时刻的值为第三值，且第一文件的索引节点编号为第一索引节点编号，则确定在第一时刻至第二时刻的时段内，第一文件未被修改。

第二方面，提供一种文件处理方法，应用于电子设备或支持电子设备功能的组件（比如芯片系统），方法包括：

显示第二界面，第二界面包括扫描得到的电子设备中的多个重复文件；多个重复文件分别存储在电子设备的多个存储空间；

响应于用户在第二界面上输入的操作，在第三时刻，获取多个重复文件中每个重复文件的指示字段，重复文件的指示字段的值用于指示重复文件是否被修改；

根据多个重复文件的指示字段在第三时刻的值，将多个重复文件中的部分重复文件压缩至同一存储空间。

比如，如图 10A 的（c），手机显示界面 906（第二界面的一个示例），界面 906 包括扫描得到的多个重复文件（英语语法 1、3、4 这组重复文件，以及计算机 2、3 这组重复文件）。

响应于用户在界面 906 上点击压缩按钮 907，在第三时刻，手机获取用户选择的英语语法 1、3、4 这组重复文件的指示字段，并根据英语语法 1、3、4 各自的指示字段在第三时刻的值，将英语语法 1、3、4 中的部分文件压缩至同一存储空间。

该方案中，能够根据用户的操作，获取第二界面中重复文件的指示字段，并根据各重复文件的指示字段，对部分重复文件进行压缩，一方面，能够释放电子设备的存储空间，另一方面，并非对用户选中的全部重复文件进行压缩，而是根据用户选中的重复文件的指示字段的值，校验是否确实需要对其中的重复文件进行压缩，以降低误压缩的概率，以免误压缩影响电子设备正常运行。可见，该方案能够实现精准的文件压缩，进而能够精准地释放电子设备的存储空间。并且，能够有效避免在执行去重之前，因为其他进程修改了重复文件，去重程序不感知重复文件被修改，而造成的误去重问题。

在一种可能的设计中，每个重复文件的指示字段的初始值为第一值。

在一种可能的设计中，显示第二界面之前，方法还包括：

扫描得到多个候选重复文件；多个候选重复文件包括多个重复文件；

在第一时刻，尝试将多个候选重复文件中每个候选重复文件的指示字段的值修改为第三值；

对多个候选重复文件进行比对，以从多个候选重复文件中确定多个重复文件。

比如，如图 10B,用户点击“清理加速”按钮之后，手机经初筛，得到多个候选重复文件。其中，英语语法 1-5 构成一组候选重复文件，计算机技术 1-4 构成一组候选重复文件。在 t1 时刻（第一时刻的一个示例），手机尝试将多个候选重复文件中每个候选重复文件的指示字段的值修改为 V1（第三值的一个示例）。

在一种可能的设计中，扫描得到多个候选重复文件，包括：

显示第三界面，第三界面包括第一控件；

检测到用户对第一控件的操作，扫描得到多个候选重复文件。

示例性的，如图 10A 的（a），手机显示界面 901（第三界面的一个示例），界面 901 包括清理加速按钮 902（第一控件）。检测到用户点击清理加速按钮 902，手机扫描得到多个候选重复文件（比如图 10B 所示的英语语法 1-5），并根据多个候选重复文件得到多个重复文件（比如图 10B 所示的英语语法 1-4）。

在一种可能的设计中，扫描得到多个候选重复文件，包括：

当电子设备空闲时，扫描得到多个候选重复文件。比如，电子设备夜间执行扫描，或者，锁屏时长达到一定时长后执行扫描。

通过在手机空闲时执行预扫描，可节约后续执行文件去重所需的时间，提高文件去重的效率，提升用户与电子设备之间的人机交互效率。

在一种可能的设计中，在第一时刻之后，方法还包括：

在第二时刻，获取至少一个重复文件的指示字段的值在第二时刻，获取至少一个重复文件的指示字段的值以及至少一个重复文件的索引节点编号；

当满足第三条件时，显示第二界面；第三条件包括：至少一个重复文件的指示字段在第二时刻的值为第三值，且，至少一个重复文件的索引节点编号与数据库中存储的索引节点编号一致。

示例性的，如图 10B，在 t5 时刻（第二时刻的一个示例），手机查询英语语法 1-4 的 U 字段（指示字段）的值，其中，英语语法 2 的 U 字段的值为 V0（不是 V1），英语语法 1、3、4 的 U 字段的值均为 V1，则手机可确定英语语法 1、3、4 在 t1-t5 时段内未被修改。手机可以在如图 10A 的（c）所示界面 906（第二界面）上显示英语语法 1、3、4 的信息。

在一种可能的设计中，尝试将多个候选重复文件中每个候选重复文件的指示字段的值修改为第三值，包括：

若多个候选重复文件中的至少一个候选重复文件满足第一条件，则将至少一个候选重复文件的指示字段的值修改为第三值；在第一条件下，电子设备的文件系统能够检测到在第一时刻至第二时刻的时段内，针对至少一个候选重复文件的修改操作。

在一种可能的设计中，在第二时刻，获取至少一个重复文件的指示字段的值，包括：在第二时刻，获取到至少一个重复文件中目标重复文件的指示字段的值为第三值；

方法还包括：在目标重复文件的指示字段在第二时刻的值为第三值的情况下，确定在第一时刻至第二时刻的时段内，目标重复文件未被修改。

在一种可能的设计中，多个候选重复文件包括第三文件；

在第一时刻之后，方法还包括：检测到第三文件被修改，将第三文件的指示字段的值修改为第二值；

在第二时刻，获取到第三文件的指示字段的值不是第三值；

确定在第一时刻至第二时刻的时段内，第三文件被修改。

在一种可能的设计中，多个候选重复文件包括第四文件；

在第一时刻，尝试将多个候选重复文件中每个候选重复文件的指示字段的值修改为第三值，包括：若第四文件不满足第一条件，则不将第四文件的指示字段修改为第三值；

方法还包括：在不能成功将第四文件的指示字段的值修改为第三值时，确定在第一时刻至第二时刻的时段内，第四文件被修改。

在一种可能的设计中，根据多个重复文件的指示字段在第三时刻的值，将多个重复文件中的部分重复文件压缩至同一存储空间，包括：

若多个重复文件中第一文件的指示字段在第三时刻的值为第三值，则对第一文件执行压缩；

或者，若多个重复文件中第一文件的指示字段在第三时刻的值不是第三值，则不对第一文件执行压缩。

示例性的，若多个重复文件中第一文件的指示字段在第三时刻的值为第三值，则对第一文件执行压缩；

或者，若多个重复文件中第一文件的指示字段在第三时刻的值不是第三值，则不对第一文件执行压缩。

示例性的，如图 19，用户点击“压缩”按钮之后，手机执行去重之前，在 t9 时刻（第三时刻），手机管家应用再次查询各重复文件（英语语法 1、3、4、计算机技术 1、2）的 U 字段的值。经判断，英语语法 3 的 U 字段的值在 t1 时刻被成功设置为 V1，但在 t9 时刻 U 字段的值为 V0，则手机管家应用可以确定 t1 时刻至 t9 时刻这段时间内，英语语法 3 被修改过。同理，手机管家应用判断其他重复文件在 t1-t9 时刻是否被修改过。

如图 19，由于英语语法 3 在 t1-t9 时段内被修改过，因此，其不再与英语语法 1、4 构成重复文件，手机管家应用不对英语语法 3 执行去重。反之，t1-t9 时段内，英语语法 1、4 未被修改，则英语语法 1、4 仍构成重复文件，手机管家应用可以在 t10 时刻对英语语法 1、4 进行去重。

如此，通过在实际压缩去重之前，对文件是否为重复文件进行校验，能够降低误去重的概率，提升去重的精准度。

在一种可能的设计中，方法还包括：

将多个重复文件的信息存储在数据库中，重复文件的信息包括如下任一项或多项信息：重复文件的索引节点编号、重复文件的路径、重复文件的名称。

在一种可能的设计中，在目标重复文件的指示字段在第二时刻的值为第三值的情况下，确定在第一时刻至第二时刻的时段内，目标重复文件未被修改，包括：

在目标重复文件的指示字段的值为第三值的情况下，若目标重复文件的索引节点编号未发生改变，则确定在第一时刻至第二时刻的时段内，目标重复文件未被修改。

第三方面，提供一种文件处理装置，应用于电子设备或芯片系统，以应用于电子设备为例，电子设备存储有第一文件的指示字段，指示字段的初始值为第一值，指示字段在第一文件被修改时被电子设备修改为第二值；装置包括：

处理模块，用于：

在第一时刻，尝试将第一文件的指示字段的值修改为第三值；第三值与第一值不同；第三值与第二值不同；

若成功将第一文件的指示字段的值修改为第三值，则在第二时刻，获取指示字段的值；

根据第一文件的指示字段在第二时刻的值，判断在第一时刻至第二时刻的时段内，第一文件是否被修改。

在一种可能的设计中，尝试将第一文件的指示字段的值修改为第三值，包括：

若满足第一条件，则将指示字段的值修改为第三值；在第一条件下，电子设备的文件系统能够检测到在第一时刻至第二时刻的时段内，针对第一文件的修改操作。

在一种可能的设计中，处理模块，还用于在第一时刻之后，检测到第一文件被修改，将第一文件的指示字段的值修改为第二值。

根据第一文件的指示字段在第二时刻的值，判断在第一时刻至第二时刻的时段内，第一文件是否被修改，包括：

若第一文件的指示字段在第二时刻的值为第二值，则确定在第一时刻至第二时刻的时段内，第一文件被修改。

在一种可能的设计中，根据第一文件的指示字段在第二时刻的值，判断在第一时刻至第二时刻的时段内，第一文件是否被修改，包括：

若第一文件的指示字段在第二时刻的值为第三值，则确定在第一时刻至第二时刻的时段内，第一文件未被修改。

在一种可能的设计中，尝试将第一文件的指示字段的值修改为第三值，包括：若不满足第一条件，则不将指示字段修改为第三值；

处理模块，还用于在不能成功将第一文件的指示字段的值修改为第三值时，确定在第一时刻至第二时刻的时段内，第一文件被修改。

在一种可能的设计中，第一条件包括如下一项或多项条件：在第一时刻之前，第一文件未被映射到内存；在第一时刻之前，第一文件被映射到内存，但电子设备未对内存执行过写入操作。

在一种可能的设计中，处理模块，还用于：

扫描得到电子设备中的重复文件，重复文件包括第一文件和第二文件；第一文件存储在第一存储空间，第二文件存储在第二存储空间；

确定第一时刻至第二时刻的时段内，第一文件、第二文件未被修改；

该装置还包括显示模块，显示模块，用于显示第一界面；第一界面包括第一文件的信息以及第二文件的信息。

在一种可能的设计中，该装置还包括输入模块，输入模块，用于检测到用户在第一界面中输入的第一操作；

处理模块，用于根据输入模块检测到的第一操作，将第一文件以及第二文件压缩至同一存储空间。

在一种可能的设计中，该装置还包括存储模块。存储模块，用于在扫描电子设备中的重复文件之后，存储第一文件的标识和第二文件的标识。

在一种可能的设计中，第一文件的标识包括第一文件的第一索引节点编号；第二文件的标识包括第二文件的第二索引节点编号。

在一种可能的设计中，根据第一文件的指示字段在第二时刻的值，判断在第一时刻至第二时刻的时段内，第一文件是否被修改，包括：

若第一文件的指示字段在第二时刻的值为第三值，且第一文件的索引节点编号为第一索引节点编号，则确定在第一时刻至第二时刻的时段内，第一文件未被修改。

第四方面，提供一种文件处理装置，应用于电子设备或芯片系统，该装置包括：

显示模块，用于显示第二界面，第二界面包括扫描得到的电子设备中的多个重复文件；多个重复文件分别存储在电子设备的多个存储空间；

处理模块，用于响应于用户在第二界面上输入的操作，在第三时刻，获取多个重复文件中每个重复文件的指示字段，重复文件的指示字段的值用于指示重复文件是否被修改；

处理模块，还用于根据多个重复文件的指示字段在第三时刻的值，将多个重复文件中的部分重复文件压缩至同一存储空间。

在一种可能的设计中，每个重复文件的指示字段的初始值为第一值。

在一种可能的设计中，处理模块，还用于显示第二界面之前，扫描得到多个候选重复文件；多个候选重复文件包括多个重复文件；

处理模块，还用于在第一时刻，尝试将多个候选重复文件中每个候选重复文件的指示字段的值修改为第三值。

在一种可能的设计中，扫描得到多个候选重复文件，包括：

控制显示模块显示第三界面，第三界面包括第一控件；

检测到用户对第一控件的操作，扫描得到多个候选重复文件。

在一种可能的设计中，扫描得到多个候选重复文件，包括：

当电子设备空闲时，扫描得到多个候选重复文件。

在一种可能的设计中，处理模块，还用于在第一时刻之后，在第二时刻，获取至少一个重复文件的指示字段的值；

显示模块，还用于根据至少一个重复文件的指示字段在第二时刻的值，显示第二界面；第二界面中的多个重复文件为在第一时刻至第二时刻的时段内，修改状态是未被修改的文件。

在一种可能的设计中，尝试将多个候选重复文件中每个候选重复文件的指示字段的值修改为第三值，包括：

若多个候选重复文件中的至少一个候选重复文件满足第一条件，则将至少一个候选重复文件的指示字段的值修改为第三值；在第一条件下，电子设备的文件系统能够检测到在第一时刻至第二时刻的时段内，针对至少一个候选重复文件的修改操作。

在一种可能的设计中，在第二时刻，获取至少一个重复文件的指示字段的值，包括：在第二时刻，获取到至少一个重复文件中目标重复文件的指示字段的值为第三值；

处理模块，还用于在目标重复文件的指示字段在第二时刻的值为第三值的情况下，确定在第一时刻至第二时刻的时段内，目标重复文件未被修改。

在一种可能的设计中，多个候选重复文件包括第三文件；

处理模块，还用于：

在第一时刻之后，检测到第三文件被修改，将第三文件的指示字段的值修改为第二值；

在第二时刻，获取到第三文件的指示字段的值不是第三值；

确定在第一时刻至第二时刻的时段内，第三文件被修改。

在一种可能的设计中，多个候选重复文件包括第四文件；

在第一时刻，尝试将多个候选重复文件中每个候选重复文件的指示字段的值修改为第三值，包括：若第四文件不满足第一条件，则不将第四文件的指示字段修改为第三值；

处理模块，还用于在不能成功将第四文件的指示字段的值修改为第三值时，确定在第一时刻至第二时刻的时段内，第四文件被修改。

在一种可能的设计中，根据多个重复文件的指示字段在第三时刻的值，将多个重复文件中的部分重复文件压缩至同一存储空间，包括：

若多个重复文件中第一文件的指示字段在第三时刻的值为第三值，则对第一文件执行压缩；

或者，若多个重复文件中第一文件的指示字段在第三时刻的值不是第三值，则不对第一文件执行压缩。

在一种可能的设计中，存储模块，用于存储多个重复文件的标识，重复文件的标识包括重复文件的索引节点编号。

在一种可能的设计中，在目标重复文件的指示字段在第二时刻的值为第三值的情况下，确定在第一时刻至第二时刻的时段内，目标重复文件未被修改，包括：

在目标重复文件的指示字段的值为第三值的情况下，若目标重复文件的索引节点编号未发生改变，则确定在第一时刻至第二时刻的时段内，目标重复文件未被修改。

在上述任一方面的任一种可能的设计中，指示字段存储在电子设备的元数据区中。

在上述任一方面的任一种可能的设计中，第一文件的标识包括第一文件的第一索引节点编号；第二文件的标识包括第二文件的第二索引节点编号。

在上述任一方面的任一种可能的设计中，第一条件包括如下一项或多项条件：在第一时刻之前，重复文件未被映射到内存；在第一时刻之前，重复文件被映射到内存，但电子设备未对内存执行过写入操作。

附图说明

图 1A 为本申请实施例提供的重复文件扫描流程的一种示例图；

图 1B 为本申请实施例提供的 i-mtime 字段的一种示例图；

图 2 为本申请实施例提供的写接口方式的写操作的示意图；

图 3 为本申请实施例提供的写内存方式的写操作的示意图；

图 4、图 5 为本申请实施例提供的写操作的另外的示意图；

图 6 为本申请实施例提供的电子设备的硬件结构示意图；

图 7 为本申请实施例提供的电子设备的软件结构示意图；

图 8 为本申请实施例提供的数据区、元数据区的示意图；

图 9 为本申请实施例提供的文件处理方法的示意图；

图 10A 为本申请实施例提供的界面示意图；

图 10B、图 10C 为本申请实施例提供的场景示意图；

图 11 为本申请实施例提供的数据库的示意图；

图 12 为本申请实施例提供的文件被修改的场景示意图；

图 13 为本申请实施例提供的文件未被修改的场景示意图；

图 14 为本申请实施例提供的界面示意图；

图 15 为本申请实施例提供的 V1 设置失败的场景示意图；

图 16 为本申请实施例提供的文件被修改的场景示意图；

图 17 为本申请实施例提供的文件未被修改的场景示意图；

图 18 为本申请实施例提供的 V1 设置失败的场景示意图；

图 19 为本申请实施例提供的场景示意图；

图 20 为本申请实施例提供的界面示意图；

图 21、图 22 为本申请实施例提供的装置的结构示意图。

具体实施方式

本申请的描述中所提到的术语“包括”和“具有”以及它们的任何变形，意图在于覆盖不排他的包含。例如包含了一系列步骤或单元的过程、方法、系统、产品或设备没有限定于已列出的步骤或单元，而是可选地还包括其他没有列出的步骤或单元，或可选地还包括对于这些过程、方法、产

品或设备固有的其它步骤或单元。

需要说明的是，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。

在本申请的描述中，除非另有说明，“多个”的含义是指两个或两个以上。本文中的“和/或”仅仅是一种描述关联对象的关联关系，表示可以存在三种关系，例如，A 和/或 B，可以表示：单独存在 A，同时存在 A 和 B，单独存在 B 这三种情况。

为便于理解，下面先对本申请实施例可能涉及到的相关术语和概念进行介绍。

1、文件去重

一些场景中，文件系统中存在多个重复文件，该多个重复文件原本占用多份存储空间，经过去重后，多个重复文件只占用一份存储空间。重复文件，即内容完全相同文件。示例性的，如图 1A，文件 1 和文件 2 构成重复文件，去重之前，文件 1 和文件 2 分别占用存储空间 1、存储空间 2。电子设备对文件 1 和文件 2 执行去重之后，文件 1 和文件指向同一块存储空间 1，极大释放了电子设备的存储空间。一些方案中，电子设备可以扫描存储的重复文件。重复文件的扫描可包括初筛和二进制比对。其中，在初筛时，电子设备截选文件中的部分内容，通过读取该部分内容，得到一个值。如果两个文件的截选内容的值相同，电子设备初步认为这两个文件有可能是重复文件。之后，电子设备对这两个文件进行二进制比对，逐个字节比对这两个文件的全部内容，当二进制比对的结果一致时，说明两个文件是重复文件。

在另一些场景中，二进制比对还可以替换为其他用于确定重复文件的步骤，比如哈希值比对等，本申请实施例对重复文件的扫描流程的具体实现，不做限制。

2、i-mtime 字段

根据文件在一段时间内是否被修改，可以将文件的修改状态分为两种，一种是在一段时间被被修改，一种是在一段时间内未被修改。

在一些方案中，电子设备可以设置 i-mtime 字段，i-mtime 字段可用于表示文件最后一次被修改的时间。可选的，i-mtime 字段通常可以存储在文件系统的元数据区中。具体的，i-mtime 字段可以存储的元数据区的索引节点（inode）区。示例性的，文件 1 历史上被修改过三次，修改时间分别为 2022-8-22（8：00）、2022-8-23（10：00）、2022-8-23（18：00），则如表 1 所示，文件 1 的 i-mtime 字段的字段值可以为 2022-8-23（18：00）。类似的，文件 2 历史上被修改过两次，修改时间分别为 2022-8-22（18：00）、2022-8-23（10：00），则如表 1 所示，文件 2 的 i-mtime 字段的字段值可以为 2022-8-23（10：00）。

表 1

i-mtime	文件
2022-8-23（18：00）	文件 1
2022-8-23（10：00）	文件 2

在一些方案中，电子设备中的应用程序可以通过查询文件的 i-mtime 字段，获知该文件在一段时间内是否被修改过。作为一种可能的实现方式，应用程序分别在时间点 1、时间点 2 查询文件的 i-mtime 字段，若两次查询到 i-mtime 字段的字段值一致，说明时间点 1 至时间点 2 期间内文件未被修改。反之，若两次查询到 i-mtime 字段的字段值不一致，说明时间点 1 至时间点 2 期间内文件被修改过。

示例性的，如图 1B 的（a），应用程序在 10:00 查询文件 1 的 i-mtime 字段，获知文件 1 最后一次被修改的时间为 8:00，应用程序在 12:00 再次查询文件 1 的 i-mtime 字段，获知文件 1 最后一次被修改的时间为 8:00。此种情况下，两次查询得到的字段值（用于表示文件最后一次被修改的时间）相同，应用程序可确定文件 1 在 10:00-12:00 期间内未被修改过。

再示例性的，如图 1B 的（b），应用程序在 10:00 查询文件 1 的 i-mtime 字段，获知文件 1 最后一次被修改的时间为 8:00，应用程序在 12:00 再次查询文件 1 的 i-mtime 字段，获知文件 1 最后一次被修改的时间为 11:00。此种情况下，两次查询得到的字段值不同，应用程序可确定文件 1 在 10:00-12:00 期间内被修改过。

3、写（write）接口

作为一种写文件的方式，电子设备可以通过写（write）接口执行写文件操作。此种写文件的方式，每次写操作，均会执行文件系统中的一段代码，文件系统可以修改文件的 i-mtime 字段，以便记录文件最后一次被修改的时间。示例性的，如图 2，在 t1 时刻（10:00）调用写接口对文件 1 进行修改时，电子设备会执行文件系统中的代码，文件系统可以修改文件 1 的 i-mtime 字段，以标识文件 1 最后一次被修改的时间为 10:00。类似的，在 t2 时刻（12:00）调用写接口对文件 1 进行修改时，电子设备会执行文件系统中的代码，文件系统可以修改文件 1 的 i-mtime 字段，以标识文件 1 最后一次被修改的时间为 12:00。

4、mmap

作为写文件的另一方式，电子设备可以通过 mmap 接口，将文件映射到一定内存。后续，写文件时，可以通过写内存的方式对文件进行写操作。写内存，可以理解为向文件映射的内存中写入内容，相当于间接地向相应文件进行写入。此种写文件的方式，在对某个文件映射到相应内存之后，电子设备首次在该内存中写入时，会执行文件系统中的相关代码。在一些示例中，后续，在同一内存中写入时，有可能不再执行文件系统中的代码，以免陷入文件系统产生额外开销。

示例性的，如图 3 的（a），电子设备将文件 2 映射到内存 2。如图 3 的（b），在 t1 时刻（10:00），电子设备通过 mmap 接口首次对内存 2 进行写操作，此时，电子设备执行文件系统中的相关代码对内存 2 进行写操作。因此，文件系统能够感知内存 2 关联的文件 2 被修改，并据此修改文件 2 的 i-mtime 字段，以标识文件 2 最后一次被修改的时间为 10:00。

之后，仍如图 3 的（b），在 t2 时刻，电子设备通过 mmap 接口对内存 2 进行写操作。由于历史上曾对内存 2 进行写操作，为了避免频繁陷入文件系统导致的系统开销，在 t2 时刻电子设备不再陷入文件系统执行对内存 2 的写操作。

再示例性的，如图 4，在 t1 时刻，电子设备将文件映射到某块内存。之后，在 t2 时刻（10:05），电子设备向内存中写入内容，相当于向内存映射的文件中写入相应内容。由于此次写入是向该内存中的首次写入，因此，会执行文件系统中的相关代码，文件系统可以感知文件被修改，并修改该文件的 i-mtime 字段，以标识该文件最后一次被修改的时间为 10:05。之后，在 t3 时刻（10:06），电子设备中的某个应用程序查询文件的 i-mtime 字段。在 t4 时刻（10:07），电子设备第二次向同一块内存中写入内容，且为了频繁陷入文件系统，电子设备本次写内存时并未陷入文件系统中的代码。因此，文件系统不能感知该内存关联的文件被修改，相应的没有修改文件的 i-mtime 字段。t5 时刻，应用程序再次查询文件的 i-mtime 字段。由于 t5、t3 时刻查询到 i-mtime 字段的字段值均为 10:05，则电子设备误认为 t3-t5 时段内（即 10:05-10:07 期间）文件未被修改。

可见，某些场景中，通过 mmap 方式执行写操作时电子设备不会陷入文件系统，导致文件系统无法感知文件被修改，无法跟踪文件的修改，无法及时更新 i-mtime 字段的字段值，导致电子设备对文件是否修改这一事件进行误判。

5.修改系统时间导致文件修改状态的误判

在一些场景中，无论是通过写接口还是 mmap 接口执行写操作，在系统时间被修改的情况下，均可能对文件的修改状态进行误判。示例性的，如图 5，在 t1 时刻（10:00），文件被修改。之后，在 t2 时刻，电子设备中的某个应用程序查询文件的 i-mtime 字段，获知文件最后一次被修改的时间为 10:00。在 t3 时刻，电子设备的系统时间被修改，比如由 10:30 修改为 9:30。之后，t4 时刻（修改系统时间后为 10:00），文件再次被修改。t5 时刻，应用程序再次查询文件的 i-mtime 字段。由于 t5、t3 时刻查询到 i-mtime 字段的字段值均为 10:00，则电子设备误认为 t3-t5 时段内（即 10:05-12:00 期间）文件未被修改。

这里，系统时间可以是用户修改的，也可以是系统自动修改的。

可见，由于系统时间被修改，可能致使两次修改文件时的系统时间相同，进而使得电子设备对文件是否被修改这一事件进行误判，无法满足电子设备精准判断文件是否被修改的业务需求，即，不能可靠的判断文件的修改状态，电子设备的性能较低。

6、inotify

在一些方案中，还可以通过 inotify 机制监听文件是否被修改。应用程序可以通过 inotify 接口通知文件系统需要监听文件的修改状态。后续，文件系统监听到文件被修改的事件，会通知该应用程序文件被修改。此种方式，要求应用程序在需要监听的时段内常驻运行，占用较多系统资源（比如内存资源），电子设备的开销较大。

可见，上述用于判断文件是否被修改的方案，均不能使电子设备获得较好的性能。为解决上述技术问题，本申请实施例提供一种文件处理方法。该方法可适用于电子设备。例如，本申请实施例中的电子设备可通过图 6 中的电子设备来实现。图 6 所示为本申请实施例提供的电子设备的硬件结构示意图。该电子设备 200 包括至少一个处理器 201，存储器 202 以及至少一个收发器 203。

处理器 201 可以是一个通用中央处理器（central processing unit，CPU），微处理器，特定应用集成电路（application-specific integrated circuit，ASIC），或一个或多个用于控制本申请方案程序执行的集成电路。

上述各组件之间可包括通信线路，用于在上述组件之间传送信息。

收发器 203，用于与其他设备通信。在本申请实施例中，收发器可以是模块、电路、总线、接口或者其它能实现通信功能的装置，用于与其他设备通信。可选的，该收发器可以为独立设置的发送器，该发送器可用于向其他设备发送信息，该收发器也可以为独立设置的接收器，用于从其他设备接收信息。该收发器也可以是将发送、接收信息功能集成在一起的部件，本申请实施例对收发器的具体实现不做限制。

存储器 202 可以是只读存储器（read-only memory，ROM）或可存储静态信息和指令的其他类型的静态存储设备，随机存取存储器（random access memory，RAM）或者可存储信息和指令的其他类型的动态存储设备，也可以是电可擦可编程只读存储器（electrically erasable programmable read-only memory，EEPROM）、只读光盘（compact disc read-only memory，CD-ROM）或其他光盘存储、光碟存储（包括压缩光碟、激光碟、光碟、数字通用光碟、蓝光光碟等）、磁盘存储介质或者其他磁存储设备、或者能够用于携带或存储具有指令或数据结构形式的期望的程序代码并能够由计算机存取的任何其他介质，但不限于此。存储器可以是独立存在，通过通信线路与处理器相连接。存储器也可以和处理器集成在一起。

其中，存储器 202 用于存储用于实现本申请方案的计算机执行指令，并由处理器 201 来控制执行。处理器 201 用于执行存储器 202 中存储的计算机执行指令，从而实现本申请下述实施例提供的各个方法。

可选的，本申请实施例中的计算机执行指令也可以称之为应用程序代码、指令、计算机程序或者其它名称，本申请实施例对此不作具体限定。

在具体实现中，作为一种实施例，处理器 201 可以包括一个或多个 CPU。

在具体实现中，作为一种实施例，电子设备 200 可以包括多个处理器。这些处理器中的每一个可以是一个单核（single-CPU）处理器，也可以是一个多核（multi-CPU）处理器。这里的处理器可以指一个或多个设备、电路、和/或用于处理数据（例如计算机程序指令）的处理核。

如图 6 所示为电子设备的示例性结构图。应该理解的是，图示电子设备仅是一个范例，并且在实际应用中电子设备可以具有比图 6 中所示出的更多的或者更少的部件，可以组合两个或更多的部件，或者可以具有不同的部件配置。

电子设备的软件系统可以采用分层架构，事件驱动架构，微核架构，微服务架构，或云架构。例如，该电子设备的软件系统可以是 Android®系统等。本发明实施例以分层架构的 Android 系统为例，示例性说明电子设备的软件结构。

图 7 是本发明实施例的电子设备的软件结构框图。

分层架构将软件分成若干个层，每一层都有清晰的角色和分工。层与层之间通过软件接口通信。在一些实施例中，将 Android 系统分为五层，从上至下分别为应用程序层（简称应用层），应用程序框架层（简称框架层），安卓运行时(Android runtime)和系统库，硬件抽象层（Hardware Abstraction Layer，HAL）以及内核层。

（1）应用层

应用层可以包括一系列应用程序包。如图 7 所示，应用程序包可以包括相机，图库，日历，通话，地图，导航，无线局域网（wireless local area network，WLAN），蓝牙，音乐，视频，短信息等应用程序。

本申请实施例中，应用程序还可以包括设置、手机管家等应用。以手机管家应用为例，手机管家应用可以调用文件系统提供的接口修改相应文件的 U 字段的字段值。后续，手机管家应用可以根据 U 字段的字段值判断相应文件在一段时间内是否被修改过。

可选的，应用程序可以通过如下至少一种方式判断 U 字段的值。

第一种方式：可以由应用程序调用文件系统提供的第一接口，第一接口返回文件的U字段值。

第二种方式：应用程序调用文件系统提供的第二接口，文件系统在执行过程中判断文件的U字段值，并根据U字段值判断是否要继续执行第二接口的剩余步骤。

可选的，上述应用程序可以是运行在用户态的应用程序。

运行在“用户态”的应用程序可被称为“用户态程序”。比如，畅连通信、智慧生活等应用程序均为用户态程序。在一些情况下，用户态程序需要使用操作系统提供的一些功能（诸如读取文件内容），此时，用户态程序会调用操作系统提供的应用编程接口(application programming interface，API)，从用户态陷入内核态执行操作系统内核提供的函数。

（2）框架层

框架层为应用层的应用程序提供API和编程框架。框架层包括一些预先定义的函数。

如图7所示，框架层可以包括活动管理器，窗口管理器，内容提供器，视图系统，电话管理器，资源管理器，通知管理器等。

其中，活动管理器，用于负责所有APP进程的启动，APP四大组件（如活动activity、服务service、广播接收者broadcast receiver、内容提供者content provider）的启动，APP之间及APP内部界面之间切换和各种组件的生命周期。窗口管理器用于管理窗口程序。窗口管理器可以获取显示屏大小，判断是否有状态栏，锁定屏幕，截取屏幕等。内容提供器用来存放和获取数据，并使这些数据可以被应用程序访问。所述数据可以包括视频，图像，音频，拨打和接听的电话，浏览历史和书签，电话簿等。视图系统包括可视控件，例如显示文字的控件，显示图片的控件等。视图系统可用于构建应用程序。显示界面可以由一个或多个视图组成的。例如，包括短信通知图标的显示界面，可以包括显示文字的视图以及显示图片的视图。电话管理器用于提供电子设备的通信功能。例如通话状态的管理(包括接通，挂断等)。资源管理器为应用程序提供各种资源，比如本地化字符串，图标，图片，布局文件，视频文件等等。通知管理器使应用程序可以在状态栏中显示通知信息，可以用于传达告知类型的消息，可以短暂停留后自动消失，无需用户交互。比如通知管理器被用于告知下载完成，消息提醒等。通知管理器还可以是以图表或者滚动条文本形式出现在系统顶部状态栏的通知，例如后台运行的应用程序的通知，还可以是以对话窗口形式出现在屏幕上的通知。例如在状态栏提示文本信息，发出提示音，电子设备振动，指示灯闪烁等。

3、Android Runtime和系统库

如图7所示，Android Runtime负责安卓系统的调度和管理。

系统库可以包括多个功能模块。例如：表面管理器(surface manager)，媒体库(Media Libraries)，三维图形处理库(例如：OpenGL ES)，2D图形引擎(例如：SGL)等。

其中，表面管理器用于对显示子系统进行管理，并且为多个应用程序提供了2D和3D图层的融合。媒体库支持多种常用的音频，视频格式回放和录制，以及静态图像文件等。媒体库可以支持多种音视频编码格式，例如：MPEG4，H.264，MP3，AAC，AMR，JPG，PNG等。三维图形处理库用于实现三维图形绘图，图像渲染，合成，和图层处理等。2D图形引擎是2D绘图的绘图引擎。

（4）内核层

内核层是硬件和软件之间的层。

在本申请实施例中，内核层可以包括文件系统。文件系统是操作系统内核的核心模块，也即主要组成部分之一。文件系统负责管理和存储文件信息，主要是为用户建立文件，存入、读出、修改、转储文件，控制文件的存取，当用户不再使用时撤销文件等。

在一些实施例中，文件系统可运行在内核态。

运行在“用户态”的应用程序可被称为“用户态程序”。比如，畅连通信、智慧生活等应用程序均为用户态程序。在一些情况下，用户态程序需要使用操作系统提供的一些功能（诸如读取文件内容），此时，用户态程序会调用操作系统提供的应用编程接口(application programming interface，API)，从用户态陷入内核态执行操作系统内核提供的函数。

文件系统中的数据可包括元数据以及数据。其中，元数据用来描述文件的特征、属性。可选的，元数据包括但不限于如下信息：文件系统的控制信息、数据结构、文件的描述信息，例如文件长度、文件类型等。文件类型例如为普通文件(regular inode)、目录文件 (directory inode)、软链接(symbol link inode)、特殊文件(special inode)等。

可选的，文件的元数据和数据可以存储在存储设备（比如硬盘）中。如图8，存储设备的存储

区域可包括元数据区以及数据区。

文件的数据可以存储在存储设备的数据区。可选的，数据区中存储的数据可以是基于无损压缩技术进行文件级别的压缩处理后得到的数据。数据区中的数据在存储介质(例如为磁盘、闪存等)的物理存储空间上按照磁盘块的集合来进行存储。其中，同一文件的数据可以在连续的磁盘块中存储，或者，也可以在不连续的磁盘块中交叉存储。

应理解，本申请中引入磁盘块的概念并不意味着存储介质仅限定为磁盘，磁盘块可以用来表示存储介质的物理存储空间经划分后得到的小的物理存储空间。

如图 8，数据存储位置可以用数据所在块（block）的索引表示。比如图 8 中，文件存储在块 909、912、914 中。元数据区可以存储该文件的块索引 909、912、914。后续，对文件执行诸如读写操作时，电子设备可以从元数据区读取文件所在块的索引，并根据索引，在相应块的位置对文件进行访问（比如读取）。

在一些实施例中，文件系统可以为用户态程序提供用于读写文件的接口。作为第一种可能的实现方式，文件系统提供读写接口（例如：read，write），用户态程序调用读、写接口对文件进行读、写处理。

作为第二种可能的实现方式，文件系统提供内存映射接口（例如：mmap），用户态程序调用内存映射接口，将文件映射到内存，并获取映射后的内存地址。之后，用户态程序可以通过内存访问的方式对文件进行访问，比如，通过写内存的方式对文件进行写操作。

在一些实施例中，对于上述第一种文件访问方法，由于每次文件访问的操作都会调用文件系统提供的接口，因此，文件系统可以感知到每次的访问文件操作。

对于上述第二种文件访问方法，对于建立映射后的第一次文件访问（比如第一次写内存），会产生内存缺页，此时会调用文件系统的函数，故文件系统知道用户态程序对文件进行了访问。而后续的访问（比如后续的写内存）不一定会产生内存缺页，故文件系统可能无法感知用户态程序对文件进行了修改。

本申请实施例中，如图 8 所示，文件系统可以在元数据区中新增 unmodified 字段（可简称 U 字段），U 字段可指示相应文件是否被修改。每个文件具有对应的 U 字段。

可选的，文件系统可以设置 U 字段的初始值。比如，初始值设置为 V0。

文件系统可以向应用程序提供接口，应用程序可调用该接口修改 U 字段的字段值（比如修改为 V1）。后续，应用程序可以根据 U 字段的字段值判断相应文件是否被修改。

如下，以电子设备为手机为例，对本申请实施例的技术方案进行详细介绍。如下，首先以文件去重场景为例进行说明。

实施例一

如图 9，本申请实施例的技术方案可包括如下步骤：

S101、文件系统将第一文件的 U 字段的值设置为 V0。

可选的，U 字段可保存在文件系统的元数据区中。

S102、第一时刻，应用程序调用文件系统的接口，尝试将第一文件的 U 字段的值修改为 V1。

示例性的，如图 10A 的（a）所示，手机显示手机管家应用的界面 901，界面 901 可包括“清理加速”按钮 902。检测到用户点击“清理加速”按钮 902 之后，手机可以扫描手机中存储的待清理的数据。待清理的数据可包括多个重复文件，即内容完全相同的多个文件。

在一些实施例中，手机管家应用扫描多个重复文件的过程中，手机管家应用调用文件系统的接口，以尝试将多个重复文件各自对应的 U 字段的字段值由初始值 V0 修改为 V1。可选的，扫描重复文件的过程可包括初筛和二进制比对。手机管家应用可以在初筛得到初步被识别为重复文件的文件（可简称为候选重复文件）之后，二进制比对之前调用文件系统的接口，尝试将各候选重复文件的 U 字段的值修改为 V1。

示例性的，如图 12，在 t6 时刻，用户点击“清理加速”按钮之后，手机可以执行重复文件的初筛，初筛完成后，初步识别英语语法 1-4 为一组候选重复文件，计算机技术 1、2 为一组候选重复文件。以尝试将候选重复文件中英语语法 2 的 U 字段值修改为 V1 为例，用户点击“清理加速”按钮，手机执行重复文件的初筛之后，在 t1 时刻（第一时刻），手机管家应用调用文件系统的接口，尝试将英语语法 2 对应的 U 字段的字段值由 V0 修改为 V1。在图 12 的示例中，由于 t1 时刻之前，手机未执行过针对英语语法 2 的 mmap，满足第一条件，手机管家应用可以成功将英语语法 2 的 U 字

段的值由 V0 修改为 V1。

同理，手机管家应用可以尝试将其他候选重复文件（即初步被识别为重复文件的文件）的 U 字段的值修改为 V1。

作为一种可能的实现方式，应用程序想要将第一文件的 U 字段置为 V1 时，可以调用文件系统的接口向文件系统发送请求消息，该请求消息用于请求将第一文件的 U 字段的初始值修改为 V1。

可选的，V1 可以由应用程序指定，也可以由文件系统指定。

在一些实施例中，文件系统从应用程序接收请求消息之后，判断是否满足修改该文件的 U 字段的值的第一条件，在满足第一条件时，文件系统可以将第一文件的 U 字段的值修改为 V1。反之，在不满足第一条件时，文件系统不将第一文件的 U 字段的值修改为 V1。

可选的，第一条件包括：该文件是否已被映射到内存（例如，应用程序已使用 mmap()对文件进行映射）。

示例性的，如图 15，t1 时刻之前，手机已将计算机技术 1 这一文件映射到内存，不满足第一条件，则在 t1 时刻，手机管家应用不会成功将计算机技术 1 的 U 字段的字段值由 V0 修改为 V1。

图 10B 示出了初筛得到的部分候选重复文件。经初筛，手机管家应用初步识别出英语语法 1-5 为一组候选重复文件，计算机技术 1-4 为一组候选重复文件。在 t1 时刻，手机管家应用可以调用文件系统的接口，以尝试将各候选重复文件的 U 字段的值修改为 V1。示例性的，表 2 示出了 t1 时刻之后，各候选重复文件的 U 字段的字段值。

表 2

文件	U 字段
英语语法 1	V1
英语语法 2	V1
英语语法 3	V1
英语语法 4	V1
英语语法 5	V1
计算机技术 1	V0
计算机技术 2	V1
计算机技术 3	V1
…	…

如表 2，计算机技术 1 的 U 字段的字段值仍为初始值 V0，一个可能的原因是在 t1 时刻之前，手机已将计算机技术 1 这一文件映射到内存。其他文件，诸如英语语法 1-4、计算机技术 2,的 U 字段的字段值均被手机管家应用修改为 V1，说明 t1 时刻之前，手机未将这几个文件映射到内存。

可选的，在满足第一条件的情况下，文件系统向应用程序返回设置成功的返回消息，应用程序可据此确定该文件的 U 字段的值已被修改为 V1。

可选的，在不满足第一条件的情况下，文件系统向应用程序返回设置失败的返回消息，应用程序可据此确定该文件的 U 字段的值未被设置为 V1（仍为 V0）。

S103、检测到第一文件被修改，文件系统将第一文件的 U 字段的值修改为 V2。

在一些场景中，用户可以对文件进行修改。比如，如图 12，假设 t1 时刻之前，手机未执行过针对英语语法 2 的写内存操作。t3 时刻，响应于用户对英语语法 2 的修改意图，手机针对英语语法 2 进行首次写内存。具体的，手机对英语语法 2 进行首次写内存操作，必定会执行文件系统中的相关代码。如此，在 t3 时刻，文件系统能够感知英语语法 2 被修改，并将英语语法 2 的 U 字段由 V1 修改为 V2。

可选的，文件系统还可以在元数据区中存储文件的修改时间。如图 12，文件系统可以将 t3 存储到元数据区中。后续，应用程序可以根据元数据区存储的修改时间，确定文件的修改时间。

S104、应用程序判断第一文件的 U 字段的值是否被成功设置为 V1。若第一文件的 U 字段的值被设置为 V1，则执行如下步骤 S106-S110。若第一文件的 U 字段的值未被设置为 V1，则执行步骤 S111。

可选的，应用程序可以从文件系统接收返回消息，该返回消息用于指示第一文件的 U 字段的值是否被成功设置为 V1。

应理解，应用程序将某个重复文件的 U 字段的值成功设置为 V1，意味着，手机在 t1 时刻之前未将该文件映射到内存。此种情况下，后续手机可以将该文件映射到内存，并且，若采用写内存方

式对该文件进行修改，则手机必定会在首次对该文件进行写内存时，执行文件系统中的相关代码。如此，文件系统能够感知该文件被修改，并修改该文件的U字段的值（比如由V1修改为V2）。手机可以根据该文件的U字段的值（V2）确定该文件被修改过。反之，若手机后续未对该文件进行修改，则该文件的U字段的值仍为V1。手机可以根据该文件的U字段的值（V1）确定该文件未被修改。

其中，V2与V1的值不同。在一些实现方式中，V2可以等于V0。

或者，后续若采用write接口方式对该文件进行修改，则手机每次写操作时同样会执行文件系统的相关代码，文件系统仍能感知到该文件被修改，并修改该文件的U字段的值（V2）。反之，若手机后续未对该文件进行修改，则该文件的U字段的值仍为V1。手机可以根据该文件的U字段的值（V1）确定该文件未被修改。

简言之，在应用程序能够在第一时刻成功将文件的U字段的值设置为V1的情况下，无论后续采用写内存方式或write接口方式对文件进行修改，手机均可以根据文件的U字段值后续（比如第二时刻）是否发生变化，精准地判断文件实际上是否被修改。

反之，应用程序未能成功将某个文件的U字段的值设置为V1（仍为V0），意味着，手机此前已将该文件映射到内存。此种情况下，若后续采用写内存（未必是首次写内存）方式对该文件进行修改，为降低开销，手机未必会执行文件系统中的相关代码，文件系统无法跟踪该文件的修改事件，也就无法在文件被修改时更新文件的U字段的值。比如，在文件被修改的情况下，文件系统可能未感知到文件被修改，因此，并未更新文件的U字段的值。再比如，文件未被修改的情况下，文件系统也不会更新文件的U字段的值。这样，无论文件是否被修改，文件的U字段的值均是V0（不会被文件系统更新为V2），手机无法根据U字段的值，判断文件是否被修改。

考虑到应用程序在第一时刻未能成功将某个文件的U字段的值设置为V1的情况下，文件系统未必能感知到文件被修改，进而文件系统未必及时更新文件的U字段的值，导致手机后续在第二时刻无法根据U字段的值判断该文件是否被修改。此种情况下，为了避免数据被误去重，手机可以执行后续步骤S109，确定第一文件不是重复文件，以免对第一文件进行误去重。

在一些实施例中，应用程序缺少相应权限时，也可能导致无法成功将文件的U字段的值由V0修改为V1。本申请实施例中，经过用户授权的系统应用具有修改U字段的权限，以降低误判风险。

或者，在一些实施例中，在缺少存储空间时，应用程序可能无法成功将文件的U字段的值修改为V1。

S105、第二时刻，应用程序查询第一文件的U字段的值。

在一些实施例中，若第一时刻成功将某些候选重复文件（初步被识别为重复文件的文件）的U字段的值修改为V1，则手机可以对这些候选重复文件进行二进制比对，以便确定最终的重复文件。

示例性的，如图10B，t6时刻，用户点击“清理加速”按钮，触发手机执行候选重复文件的初筛，初筛得到候选重复文件包括英语语法1-5这组候选重复文件的，以及计算机技术1-4这组候选重复文件。t1时刻，手机管家应用尝试将各候选重复文件的U字段的值设置为V1。假设，在t1时刻，英语语法1-4，计算机技术2-4的U字段的值均被成功设置为V1，则手机管家应用可以对应用语法1-4这组候选重复文件，以及计算机技术2-4这组候选重复文件进行二进制比对，以确定最终的重复文件。经二进制比对，英语语法1-4构成一组重复文件。计算机技术2-4构成一组重复文件。

在一些实施例中，手机可以将重复文件的信息存储到重复文件信息库中。可选的，重复文件信息库中可以存储重复文件的标识。如图11，重复文件信息库可以包括英语语法1-4这组重复文件中每个重复文件的标识，以及计算机技术2-4这组重复文件中每个重复文件的标识。可选的，重复文件的标识包括但不限于：重复文件的名称、重复文件的索引节点编号（inode number）、重复文件的U字段的值在第一时刻是否被成功设置为V1。其中，文件的索引节点编号用于唯一标识一个文件。内容不同的文件，其索引节点编号不同。

在一些实施例中，扫描得到各重复文件之后，向用户呈现重复文件的扫描结果之前，在第二时刻，应用程序可以查询各重复文件的U字段的值，并根据各重复文件的U字段在第二时刻的值，判断各重复文件在第一时刻至第二时刻期间是否被修改过。

作为一种可能的实现方式，应用程序自己查询文件的U字段的值。或者，应用程序向文件系统发送查询请求，由文件系统查询文件的U字段的值，并向应用程序返回文件的U字段的值。

以判断重复文件中英语语法2的修改状态为例，如图12，初筛完成之后，在t1时刻，初步被识别为重复文件（候选重复文件）的英语语法2的U字段的值被成功设置为V1。之后，手机管家对

各候选重复文件执行二进制比对，最终确定英语语法 1-4 构成一组重复文件，计算机计算 2-4 构成一组重复文件。在扫描得到各重复文件之后，向用户展示重复文件的扫描结果之前，在 t5 时刻，手机管家应用可以查询英语语法 2 的 U 字段的值，并根据 t5 时刻的 U 字段值判断在 t1-t5 这段时间内，英语语法 2 是否被修改。

S106、应用程序判断第一文件的 U 字段的值是否为 V1。若 U 字段的值是 V1，则执行如下步骤 S107。若 U 字段的值不是 V1，则执行如下步骤 S108。

S107、应用程序确定第一文件未被修改。

示例性的，如图 13，手机管家应用扫描完重复文件（二进制比对）之后，向用户显示扫描结果之前，在 t5 时刻，手机管家应用查询英语语法 3 这一重复文件的 U 字段的字段值，获知字段值为 V1。由于英语语法 3 的 U 字段在 t5 时刻的字段值与在 t1 时刻的字段值相同，均为 V1，则手机管家应用可以确定在 t1 时刻（t1 时刻，手机管家应用调用文件系统的接口，尝试将英语语法 3 的 U 字段的字段值由 V0 修改为 V1）至 t5 时刻期间内，英语语法 3 这一文件未被修改过。

同理，手机管家应用可以查询英语语法 1、3、4 的 U 字段的字段值，如果英语语法 1、3、4 的 U 字段值均为 V1，则说明文件系统未检测到英语语法 1、3、4 被修改。那么，手机管家应用可以确定在 t1 时刻至 t5 时刻期间内，英语语法 1、3、4 这三个文件同样未被修改过。因此，英语语法 1、3、4 仍构成重复文件，后续，手机管家应用可以对这组重复文件进行去重，以降低对存储空间的占用。

在一些实施例中，应用程序判断第一文件的索引节点编号以及第一文件的 U 字段在第二时刻的值，若第一文件的索引节点编号不变，且第一文件的 U 字段在第二时刻的值为 V1，则确定在第一时刻至第二时刻的时段内，第一文件未被修改。

S108、应用程序确定第一文件被修改。

示例性的，如图 12，手机管家应用执行完二进制比对，得到重复文件的扫描结果之后，向用户显示扫描结果之前，在 t5 时刻，手机管家应用可以从元数据区中读取英语语法 2 的 U 字段的字段值，获知字段值为 V2。由于英语语法 2 的 U 字段在 t5 时刻的字段值与在 t1 时刻的字段值不同，则手机管家应用可以确定在 t1 时刻（t1 时刻，手机管家应用请求文件系统将英语语法 3 的 U 字段的字段值由 V0 修改为 V1）至 t5 时刻期间内，英语语法 2 这一文件可能曾被修改过。手机不再将英语语法 2 视为重复文件。这样一来，后续，手机管家应用不对英语语法 2 执行去重。

可选的，元数据区中还可以存储有英语语法 2 的修改时间 t3，手机管家应用可以从元数据区中读取修改时间，并据此获知英语语法 2 的修改时间为 t3。

与图 4 所示方案中，文件系统漏检诸如第二次的写内存操作，无法及时更新 i-mtime 字段，导致电子设备误认为在 t3-t5 时段内文件未被修改相比，采用本申请实施例的技术方案，对于某个文件来说，如果第一时刻前未将该文件映射到内存，则后续手机一旦在第一时刻至第二时刻期间内对该文件进行一次或多次写内存操作，其中的首次写内存操作必定会使得手机执行文件系统的相关代码。也就是说，第一时刻至第二时刻期间内的至少首次写内存操作会被文件系统检测到。如此，文件系统可以至少根据第一时刻至第二时刻期间内的至少首次写内存操作，确定该文件被修改。可见，本申请实施例中，文件系统无需检测每次写内存操作，即便文件系统存在漏检写内存操作的现象，文件系统也能根据检测到的一段时间内的至少首次写内存操作，确定文件被修改过。如此，能够提升电子设备的检测精度。

S109、应用程序确定第一文件不是重复文件。

上文已提及，在第一时刻，第一文件的 U 字段的值未被成功设置为 V1，意味着，在第一时刻至第二时刻这段时间内，第一文件可能被修改（可能不再是重复文件），但文件系统未必能够检测到第一文件被修改。此种情况下，应用程序将第一文件视为已修改的文件，第一文件不再是重复文件。并且，应用程序不对第一文件执行去重处理，以免对实际上是非重复文件的第一文件进行误去重。

示例性的，如图 15，假设 t1 时刻之前，手机针对候选重复文件中的计算机技术 1 执行过写内存操作，则重复文件的初筛完成后，在 t1 时刻，手机管家应用无法成功将计算机技术 1 的 U 字段的值修改为 V1（U 字段的值仍为 V0）。此种情况下，手机管家应用可以不再对计算机技术 1 执行二进制比对。手机管家应用可以将计算机技术 1 视为已被修改过的文件（不再是重复文件），不再对计算机技术 1 执行去重，以免执行误去重。

在一些实施例中，手机管家应用确定各重复文件在 t1-t5 这段时间内的修改状态之后，可以向

用户呈现未被修改的一组或多组重复文件。如图 10B，在 t1-t5 时段内，未被修改的文件包括：英语语法 1、3、4、计算机技术 2、3。即，英语语法 1、3、4 仍构成重复文件，计算机技术 2、3 仍构成重复文件。

如图 10A 的（b），在 t2 时刻，手机可以显示界面 903。界面 903 可包括手机中重复文件对应的卡片 905。如图 10A 的（b），根据卡片 905，手机中存储有 8 组重复文件（包括英语语法 1、3、4 这组重复文件以及计算机计算 2、3 这组重复文件），重复文件总共占用 229MB 的存储空间。

在一些示例中，检测到用户点击“无损压缩”按钮 904 之后，手机可以显示如图 10A 的（c）所示界面 906。界面 906 可呈现上述 8 组重复文件的详细信息。比如，如图 10A 的（c），界面 906 中呈现了第一组重复文件“英语语法 1、英语语法 3、英语语法 4”的信息。界面 906 还呈现了第二组重复文件“计算机技术 2、计算机技术 3”的信息。响应于用户的诸如上滑操作等，界面 906 还可以呈现其他重复文件的信息。每组重复文件中的多个文件分别占用多个存储空间。比如，英语语法 1-4 各占用 24MB 的存储空间。

用户可以勾选某一组或多组重复文件，并点击压缩按钮 907。手机可以根据用户的压缩/去重意图，将用户选择的一组或多组重复文件进行去重。比如，如图 10A 的（c），用户勾选英语语法这组重复文件，则手机可以根据用户的压缩意图，将英语语法 1、3、4 这一组文件进行去重。如图 10C，去重之后，英语语法 1、3、4 仅占用一份存储空间（24MB），极大释放了手机的存储空间。

在一些实施例中，执行去重之后，电子设备在用户界面中仍可显示英语语法 1、3、4 这组文件，用户仍可正常打开、使用这组文件中的任意文件。相比于去重之前，去重之后，英语语法 1、3、4 占用的存储空间有所降低。

利用手机管家应用执行完去重处理之后，用户可以关闭手机管家应用。

可选的，上述提及的文件去重是文件压缩的一种可能的实现方式。文件压缩还可以有其他方式，本申请实施例不做限制。

本申请实施例的技术方案，针对不同类型的文件，有不同的方法来判断文件是否在一段时间内被修改过，以满足判断不同类型文件是否被修改的业务需求。

具体的，对于第一时刻之前未被映射到内存的文件 1，电子设备可以在第一时刻将该文件 1 的 U 字段的值由初始值 V0 修改为 V1。之后，在第二时刻，电子设备可以查询该文件 1 的 U 字段的值，并根据 U 字段的值判断该文件 1 是否被修改过。在一些情况下，第二时刻，若文件 1 的 U 字段的值是 V1 之外的其他值（比如 V2），则电子设备可以确定在第一时刻至第二时刻这一时段内，该文件 1 被修改过。在另一些情况下，第二时刻，若文件 1 的 U 字段的值是 V1，则电子设备可以确定在第一时刻至第二时刻这一时段内，该文件 1 未被修改过。

对于第一时刻之前已被映射到内存的文件 2，电子设备在第一时刻不修改该文件 2 的 U 字段的值（U 字段的值仍为初始值 V0）。意味着，在第一时刻至第二时刻这段时间内，文件 2 可能被修改（可能不再是重复文件），但文件系统未必能够检测到文件 2 被修改。此种情况下，电子设备将文件 2 视为已修改的文件，文件 2 不再是重复文件。并且，应用程序不对文件 2 执行去重处理，以免对实际上是非重复文件的文件 2 进行误去重。

需要说明的是，上述 V2 也可以替换成其他非 V1 的值（例如 V0）。本申请实施例对文件系统感知到文件被修改时，将 U 字段具体修改为何值，不做限制。

上述主要以应用程序为手机管家应用为例，对判断文件的修改状态的方法进行说明，应理解，应用程序还可以为其他任何可能的应用程序。比如，应用程序还可以为文件管理器等。文件管理器可以采用本申请实施例的技术方案判断在一段时间内，其目录下的文件是否被修改。

上述主要以用户主动触发手机扫描重复文件为例，在另一些实施例中，手机可以自行扫描手机中的重复文件，并对部分重复文件进行去重，以减少对存储空间的占用。以手机主动扫描重复文件、执行去重为例，示例性的，在时间点 1，应用程序将文件 A、B 的 U 字段的值均修改为 V1。在时间点 2，应用程序读取文件 A 和 B 的内容，认为文件 A 和 B 是重复文件。在时间点 3，文件 A 被修改，文件 A 的 U 字段的值被文件系统修改为 V2。在时间点 4，应用程序查询文件 A、B 的 U 字段的值，获知文件 A 的 U 字段的值为 V1，文件 B 的 U 字段的值为 V2，意味着，文件 A 已被修改，文件 B 未被修改，文件 A、B 不再是重复文件。因此，应用程序不对文件 A、B 执行去重。

再示例性的，在时间点 1，应用程序将文件 A、B 的 U 字段的值均修改为 V1。在时间点 2，应用程序读取文件 A 和 B 的内容，认为文件 A 和 B 是重复文件。在时间点 3，应用程序查询文件 A、B 的

U字段的值，获知文件A、B的U字段的值为V1，意味着，自时间点1开始，文件A、B均未被修改，文件A、B仍构成重复文件。因此，应用程序可以对文件A、B执行去重。

实施例二

本申请实施例还提供一种文件处理方法，该方法中，文件已被映射到内存，还可以包括如下两种情况：文件已被映射到内存，但电子设备未对文件映射的内存进行写入；文件已被映射到内存，且电子设备已对文件映射的内存进行写入。在上述两种情况下，电子设备可以采用不同方式判断文件是否被修改。如下，结合举例进行说明。

示例性的，如图16，初筛得到候选重复文件后，在t1时刻，应用程序想要将英语语法2这一文件的U字段的值修改为V1，则应用程序可以调用文件系统的接口向文件系统发送请求消息。经判断，该文件此前已被mmap（即已被映射到内存），但没有向该文件映射的内存中写入内容，那么，文件系统确定满足修改该文件的U字段的条件，进而可以将该文件的U字段的值由初始值V0修改为V1。之后，应用程序可以对英语语法3与其他候选重复文件进行二进制比对，以确定最终的重复文件。

之后，在t3时刻，电子设备首次对该文件映射的内存进行写操作，电子设备会调用文件系统的函数，故文件系统能够感知到此次对该文件的写操作。如图16，文件系统感知到对该文件的写操作之后，可以将该文件的U字段的值由V1修改为V2。

之后，如图16，在扫描得到多个重复文件之后（比如二进制比对之后），向用户展示重复文件的扫描结果之前，在t5时刻，应用程序查询该文件的U字段的值（V2），由于t5时刻查询得到的U字段的值与t1时刻U字段的值不同，因此，应用程序可确定在t1-t5时段内该文件被修改过。反之，如图17，在t5时刻，应用程序查询该文件的U字段的值（V1），由于t5时刻查询得到的U字段的值与t1时刻U字段的值相同，因此，应用程序可确定在t1-t5时段内该文件未被修改过。手机可以将t1-t5时段内未被修改过的重复文件展示给用户。比如，在t1-t5时段内，若重复文件中的英语语法1、3、4未被修改过，则手机可以在图10A的（c）所示界面中展示英语语法1、3、4的信息。

再示例性的，如图18，重复文件的初筛完成之后，t1时刻，应用程序想要将计算机技术1这一文件的U字段的值修改为V1，则应用程序可以调用文件系统的接口向文件系统发送请求消息。经判断，该文件此前已被mmap（即已被映射到内存），且已向该文件映射的内存中写入内容，那么，文件系统确定不满足修改该文件的U字段的条件，因此，文件系统不对该文件的U字段的值进行修改，该文件的U字段的值仍为V0。

当文件的U字段的值在t1时刻未被成功设置为V1时，意味着，在t1时刻至展示重构文件扫描结果的这段时间内，文件可能被修改（不再是重复文件），但文件系统未必能够检测到文件被修改。此种情况下，应用程序将文件视为已修改的文件，文件不再是重复文件。并且，应用程序不对该文件执行去重处理，以免对实际上是非重复文件的文件进行误去重。

在一些实施例中，可选的，在用户点击如图10A的（c）所示压缩按钮907之后，手机执行去重之前，手机管家应用向文件系统发送去重请求，文件系统可以再次检查各重复文件的U字段的值，以便进一步确定各重复文件自第一时刻开始是否被修改过。示例性的，如图19，用户点击“压缩”按钮之后，手机执行去重之前，在t9时刻，文件系统再次查询各重复文件（英语语法1、3、4、计算机技术1、2）的U字段的值。经判断，英语语法3的U字段的值在t1时刻被成功设置为V1，但在t9时刻U字段的值为V0，则手机可以确定t1时刻至t9时刻这段时间内，英语语法3被修改过。同理，手机判断其他重复文件在t1-t9时刻是否被修改过。

如图19，由于英语语法3在t1-t9时段内被修改过，因此，其不再与英语语法1、4构成重复文件，手机管家应用不对英语语法3执行去重。反之，t1-t9时段内，英语语法1、4未被修改，则英语语法1、4仍构成重复文件，手机管家应用可以在t10时刻对英语语法1、4进行去重。

可选的，在用户点击压缩按钮907之后，由文件系统检查文件的U字段的值，可确保检查U字段值和去重操作的原子性，即，检查U字段值，确定满足去重条件，到执行去重的这一过程中，文件不会被修改。

本申请实施例的技术方案，应用程序无需常驻运行。如此，能够在确保精准判断文件的修改状态的同时，降低系统开销。以上述去重场景为例，在需要清理手机中的重复文件时，用户打开手机管家应用，并点击“清理加速”按钮，触发手机执行重复文件扫描流程。之后，用户可以关闭手机

管家应用。后续，当手机剩余的存储空间不足，出现诸如卡顿现象时，用户可以再次打开手机管家应用，并点击“清理加速”按钮，手机可以根据历史上的重复文件扫描结果，以及部分重复文件的修改状态，重新确定手机中的重复文件，以降低重新扫描重复文件带来的资源消耗。

比如，如图 10A 的（c），用户打开手机管家应用后，手机管家应用扫描得到：英语语法 1-4 为一组重复文件，计算机技术 1、计算机技术 2 为一组重复文件。手机可以将英语语法 2-4 这组重复文件的文件标识，以及计算机技术 1、计算机技术 2 这组重复文件的文件标识存储在如图 11 所示的重复文件信息库中。

后续，手机的存储空间不足时，用户再次打开手机管家应用，并点击“清理加速”按钮 902 之后，手机管家应用可以查询如图 11 所示的重复文件信息库中，获知英语语法 1-4 这组历史上的重复文件，以及计算机技术 2-4 这组历史上的重复文件，并查询英语语法 1-4、计算机技术 2-4 这几个重复文件的 U 字段值，以便根据各重复文件的 U 字段值判断各重复文件在一段时段内是否被修改过。

经判断，假设英语语法 1、2 被修改，计算机 4 被修改，英语语法 3、英语语法 4、计算机技术 3、计算机技术 2 均未被修改。那么，手机管家应用可确定英语语法 3、英语语法 4 仍构成重复文件，计算机技术 3、计算机技术 2 仍构成重复文件。如图 14，在 t6 时刻，手机管家应用可在界面 906 中显示上述几组重复文件。

可以看出，再次打开手机管家应用，确定手机中的重复文件的过程中，手机可以减少对部分文件的扫描（比如无需从相应内存中读取历史上重复文件的内容），极大降低了手机的运行消耗。

在一些实施例中，手机管家应用重新确定手机中的重复文件之后，可以更新诸如图 11 所示的重复文件信息库。

上述第一条件仅是示例，应理解，能够保证 t1-t5 时刻的写入操作被文件系统感知的条件，均可作为第一条件，反之，只要无法保证后续的写入被文件系统感知，第一条件就不满足。示例性的，作为另一种可行的实施例，可以判断该文件在第一时刻之前是否被 mmap，且 mmap 是否具有写权限。若文件在第一时刻之前被 mmap，但 mmap 不具有写权限，则确定满足第一条件。若文件在第一时刻之前被 mmap，但 mmap 具有写权限，则确定不满足第一条件。

上述主要以去重场景为例进行说明，应理解，本申请实施例的技术方案还可以应用于其他场景中。

上述以用户点击“清理加速”按钮之后，手机执行重复文件的扫描为例进行说明，在另一些实施例中，手机还可以执行预扫描流程，在空闲时段执行重复文件的扫描，并存储扫描结果。换言之，上述实施例中的重复文件扫描流程，还可以是在手机空闲时（比如夜间）执行的扫描流程。比如，图 12 所示的初筛和二进制比对过程，可以在夜间预先执行。

示例性的，手机可以在夜间执行重复文件的扫描。假设扫描得到英语语法 1-4 为一组重复文件，并可以在扫描过程中，在 t1 时刻，尝试将英语语法 1-4 的 U 字段的值修改为 V1。之后，如图 10A 的（a），用户在白天点击“清理加速”按钮 902，手机可以查询英语语法 1-4 的 U 字段的值。假设英语语法 1-4 的 U 字段的值在 t1 时刻均被成功设置为 V1，英语语法 1、3、4 的 U 字段的值当前仍为 V1，则手机确定英语语法 1、3、4 自 t1 时刻开始未被修改过，这三个文件仍构成重复文件。手机可以在诸如图 10A 的（c）所示界面中展示这三个重复文件的信息。

通过在手机空闲时执行预扫描，可节约后续执行文件去重所需的时间，提高文件去重的效率，提升用户与电子设备之间的人机交互效率。

本申请实施例不限制电子设备执行重复文件预扫描的时机。电子设备可以在适当的场景执行重复文件的预扫描流程。比如，在夜间，在锁屏时长超过阈值时执行预扫描。

或者，在一些实施例中，电子设备也可以提供设置入口，使得用户可以设置执行预扫描的时机。示例性的，如图 20，手机的设置界面 20 中显示开关 201。当开关 201 被开启时，手机可以在空闲时执行预扫描。在另一些实施例中，手机还可以提供设置入口，使得用户可以设置预扫描的时机，比如，用户可以设置夜间或锁屏超过一定阈值时执行预扫描。

上述一个或多个界面均是示例性的，还可以有其他的界面设计方式，本申请对界面的具体设计方式不做限制。

需要说明的是，可以对上述多个实施例进行组合，并实施组合后的方案。可选的，各方法实施例的流程中的一些操作任选地被组合，并且/或者一些操作的顺序任选地被改变。并且，各流程的步骤之间的执行顺序仅是示例性的，并不构成对步骤之间执行顺序的限制，各步骤之间还可以是其

他执行顺序。并非旨在表明所述执行次序是可以执行这些操作的唯一次序。本领域的普通技术人员会想到多种方式来对本文所述的操作进行重新排序。另外，应当指出的是，本文某个实施例涉及的过程细节同样以类似的方式适用于其他实施例，或者，不同实施例之间可以组合使用。

此外，方法实施例中的某些步骤可等效替换成其他可能的步骤。或者，方法实施例中的某些步骤可以是可选的，在某些使用场景中可以删除。或者，可以在方法实施例中增加其他可能的步骤。

并且，上述各方法实施例之间可以单独实施，或结合起来实施。

可选的，电子设备还可以在其他时机、场景中判断 V1 是否设置成功。比如，每次查询文件的字段值之前，判断 V1 是否设置成功。

本申请另一些实施例提供了一种装置，该装置可以是上述电子设备（比如手机等）或服务器。该装置可以包括：显示屏、存储器和一个或多个处理器。该显示屏、存储器和处理器耦合。该存储器用于存储计算机程序代码，该计算机程序代码包括计算机指令。当处理器执行计算机指令时，装置可执行上述方法实施例中手机执行的各个功能或者步骤。该装置的结构可以参图 6 所示的电子设备（比如终端）。

其中，该装置的核心结构可以表示为图 21 所示的结构，装置包括：处理模块 1301、输入模块 1302、存储模块 1303 和显示模块 1304。

处理模块 1301，可包括中央处理器（CPU）、应用处理器（Application Processor，AP）或通信处理器（Communication Processor，CP）中的至少一个。处理模块 1301 可执行与用户电子设备的其他元件中的至少一个的控制和/或通信相关的操作或数据处理。具体地，处理模块 1301 可用于根据一定的触发条件，控制主屏上显示的内容。处理模块 1301 还用于将输入的指令或数据进行处理，并根据处理后的数据确定显示样式。

可选的，处理模块 1301 可以实现为诸如图 6 所示的处理器 201。

输入模块 1302，用于获取用户输入的指令或数据，并将获取到的指令或数据传输到电子设备的其他模块。具体地说，输入模块 1302 的输入方式可以包括触摸、手势、接近屏幕等，也可以是语音输入。例如，输入模块可以是电子设备的屏幕，获取用户的输入操作并根据获取到的输入操作生成输入信号，将输入信号传输至处理模块 1301。

存储模块 1303，可包括易失性存储器和/或非易失性存储器。存储模块用于存储用户终端设备的其他模块中的至少一个相关的指令或数据。

可选的，存储模块 1303 可以实现为诸如图 6 所示的存储器 202。

显示模块 1304，可包括例如液晶显示器(LCD)、发光二极管(LED)显示器、有机发光二极 管(OLED)显示器、微机电系统(MEMS)显示器或电子纸显示器。用于显示用户可观看的内容（例如，文本、图像、视频、图标、符号等）。

可选的，还包括通信模块 1305，用于支持个人终端（通过通信网络）与其他个人终端通信。例如，通信模块可经由无线通信或有线通 信连接到网络，以与其他个人终端或网络服务器进行通信。无线通信可采用蜂窝通信协议中的至少一个，诸如，长期演进(LTE)、高级长期演进(LTE-A)、码分多址(CDMA)、宽带码分多址(WCDMA)、通用移动通信系统(UMTS)、无线宽带 (WiBro)或全球移动通信系统(GSM)。无线通信可包括例如短距通信。短距通信可包括无线保真(Wi-Fi)、蓝牙、近场通信(NFC)、磁条传输(MST)或 GNSS 中的至少一个。

可选的，通信模块 1305 可以实现为诸如图 6 所示的收发器 203。

需要说明的是，装置的各功能模块可执行上述方法实施例中的一个或多个步骤。

本申请实施例还提供一种芯片系统，如图 22 所示，该芯片系统包括至少一个处理器 1401 和至少一个接口电路 1402。处理器 1401 和接口电路 1402 可通过线路互联。例如，接口电路 1402 可用于从其它装置（例如电子设备的存储器）接收信号。又例如，接口电路 1402 可用于向其它装置（例如处理器 1401）发送信号。示例性的，接口电路 1402 可读取存储器中存储的指令，并将该指令发送给处理器 1401。当所述指令被处理器 1401 执行时，可使得电子设备执行上述实施例中的各个步骤。当然，该芯片系统还可以包含其他分立器件，本申请实施例对此不作具体限定。

本申请实施例还提供一种计算机存储介质，该计算机存储介质包括计算机指令，当所述计算机指令在上述电子设备上运行时，使得该电子设备执行上述方法实施例中手机执行的各个功能或者步骤。

本申请实施例还提供一种计算机程序产品，当所述计算机程序产品在计算机上运行时，使得所

述计算机执行上述方法实施例中手机执行的各个功能或者步骤。

通过以上实施方式的描述，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。

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是通过一些接口，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。

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。

以上内容，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。

权 利 要 求 书

1. 一种文件处理方法，其特征在于，应用于电子设备，所述电子设备存储有第一文件的指示字段，所述指示字段的初始值为第一值，所述指示字段在所述第一文件被修改时被所述电子设备修改为第二值；所述方法包括：

在第一时刻，尝试将所述第一文件的指示字段的值修改为第三值；所述第三值与所述第一值不同；所述第三值与所述第二值不同；

若成功将所述第一文件的指示字段的值修改为所述第三值，则在第二时刻，获取所述指示字段的值；

根据所述第一文件的指示字段在所述第二时刻的值，判断在所述第一时刻至所述第二时刻的时段内，所述第一文件是否被修改。

2. 根据权利要求 1 所述的方法，其特征在于，所述指示字段存储在所述电子设备的元数据区中。

3. 根据权利要求 1 或 2 所述的方法，其特征在于，所述尝试将所述第一文件的指示字段的值修改为第三值，包括：

若满足第一条件，则将所述指示字段的值修改为所述第三值；在所述第一条件下，所述电子设备的文件系统能够检测到在所述第一时刻至所述第二时刻的时段内，针对所述第一文件的修改操作。

4. 根据权利要求 1-3 任一项所述的方法，其特征在于，在所述第一时刻之后，所述方法还包括：

检测到所述第一文件被修改，将所述第一文件的指示字段的值修改为所述第二值；

根据所述第一文件的指示字段在所述第二时刻的值，判断在所述第一时刻至所述第二时刻的时段内，所述第一文件是否被修改，包括：

若所述第一文件的指示字段在所述第二时刻的值为所述第二值，则确定在所述第一时刻至所述第二时刻的时段内，所述第一文件被修改，或者，若所述第一文件的指示字段在所述第二时刻的值不是所述第三值，则确定在所述第一时刻至所述第二时刻的时段内，所述第一文件被修改。

5. 根据权利要求 1-3 任一项所述的方法，其特征在于，根据所述第一文件的指示字段在所述第二时刻的值，判断在所述第一时刻至所述第二时刻的时段内，所述第一文件是否被修改，包括：

若所述第一文件的指示字段在所述第二时刻的值为所述第三值，则确定在所述第一时刻至所述第二时刻的时段内，所述第一文件未被修改，或者，若所述第一文件的指示字段在所述第二时刻的值不是所述第二值，则确定在所述第一时刻至所述第二时刻的时段内，所述第一文件未被修改。

6. 根据权利要求 1 或 2 所述的方法，其特征在于，所述尝试将所述第一文件的指示字段的值修改为第三值，包括：若不满足第一条件，则不将所述指示字段修改为第三值；

所述方法还包括：在不能成功将所述第一文件的指示字段的值修改为所述第三值时，确定在所述第一时刻至所述第二时刻的时段内，所述第一文件被修改。

7. 根据权利要求 3 或 6 所述的方法，其特征在于，所述第一条件包括如下一项或多项条件：在所述第一时刻之前，所述第一文件未被映射到内存；在所述第一时刻之前，所述第一文件被映射到内存，但所述电子设备未对所述内存执行过写入操作。

8. 根据权利要求 1-7 任一项所述的方法，其特征在于，所述方法还包括：

扫描得到所述电子设备中的重复文件，所述重复文件包括所述第一文件和第二文件；所述第一文件存储在第一存储空间，所述第二文件存储在第二存储空间；

确定所述第一时刻至所述第二时刻的时段内，所述第一文件、所述第二文件未被修改；

显示第一界面；所述第一界面包括所述第一文件的信息以及所述第二文件的信息。

9. 根据权利要求 8 所述的方法，其特征在于，检测到用户在所述第一界面中输入的第一操作；

将所述第一文件以及所述第二文件压缩至同一存储空间。

10. 根据权利要求 8 或 9 所述的方法，其特征在于，在扫描所述电子设备中的重复文件之后，所述方法还包括：

存储所述第一文件的标识和所述第二文件的标识。

11. 根据权利要求 10 所述的方法，其特征在于，所述第一文件的标识包括所述第一文件的第一索引节点编号；所述第二文件的标识包括所述第二文件的第二索引节点编号。

12. 根据权利要求 11 所述的方法，其特征在于，根据所述第一文件的指示字段在所述第二时刻的值，判断在所述第一时刻至所述第二时刻的时段内，所述第一文件是否被修改，包括：

若所述第一文件的指示字段在所述第二时刻的值为所述第三值，且所述第一文件的索引节点编

号为所述第一索引节点编号，则确定在所述第一时刻至所述第二时刻的时段内，所述第一文件未被修改。

13. 一种文件处理方法，其特征在于，应用于电子设备，所述方法包括：

显示第二界面，所述第二界面包括扫描得到的所述电子设备中的多个重复文件；所述多个重复文件分别存储在所述电子设备的多个存储空间；

响应于用户在所述第二界面上输入的操作，在第三时刻，获取所述多个重复文件中每个重复文件的指示字段，所述重复文件的指示字段的值用于指示所述重复文件是否被修改；

根据所述多个重复文件的指示字段在所述第三时刻的值，将所述多个重复文件中的部分重复文件压缩至同一存储空间。

14. 根据权利要求 13 所述的方法，其特征在于，根据所述多个重复文件的指示字段在所述第三时刻的值，将所述多个重复文件中的部分重复文件压缩至同一存储空间，包括：

若所述多个重复文件中第一文件的指示字段在所述第三时刻的值为第三值，则对所述第一文件执行压缩；

或者，若所述多个重复文件中第一文件的指示字段在所述第三时刻的值不是第三值，则不对所述第一文件执行压缩。

15. 根据权利要求 13 或 14 所述的方法，其特征在于，每个重复文件的指示字段的初始值为第一值。

16. 根据权利要求 13-15 任一项所述的方法，其特征在于，显示第二界面之前，所述方法还包括：

扫描得到多个候选重复文件；所述多个候选重复文件包括所述多个重复文件；

在第一时刻，尝试将所述多个候选重复文件中每个候选重复文件的指示字段的值修改为第三值；

对所述多个候选重复文件进行比对，以从所述多个候选重复文件中确定所述多个重复文件。

17. 根据权利要求 16 所述的方法，其特征在于，确定所述多个重复文件之后，所述方法还包括：

将所述多个重复文件的信息存储在数据库中，所述重复文件的信息包括如下任一项或多项信息：所述重复文件的索引节点编号、所述重复文件的路径、所述重复文件的名称。

18. 根据权利要求 16 或 17 所述的方法，其特征在于，扫描得到所述多个候选重复文件，包括：

显示第三界面，所述第三界面包括第一控件；

检测到用户对所述第一控件的操作，扫描得到所述多个候选重复文件。

19. 根据权利要求 16 或 17 所述的方法，其特征在于，扫描得到所述多个候选重复文件，包括：

当所述电子设备空闲时，扫描得到所述多个候选重复文件。

20. 根据权利要求 16-19 任一项所述的方法，其特征在于，尝试将所述多个候选重复文件中每个候选重复文件的指示字段的值修改为第三值，包括：

若所述多个候选重复文件中的至少一个候选重复文件满足第一条件，则将所述至少一个候选重复文件的指示字段的值修改为所述第三值；在所述第一条件下，所述电子设备的文件系统能够检测到在所述第一时刻至所述第二时刻的时段内，针对所述至少一个候选重复文件的修改操作。

21. 根据权利要求 20 所述的方法，其特征在于，在第二时刻，获取所述至少一个重复文件的指示字段的值，包括：在所述第二时刻，获取到所述至少一个重复文件中目标重复文件的指示字段的值为第三值；

所述方法还包括：在所述目标重复文件的指示字段在所述第二时刻的值为第三值的情况下，确定在所述第一时刻至所述第二时刻的时段内，所述目标重复文件未被修改。

22. 根据权利要求 20 所述的方法，其特征在于，所述多个候选重复文件包括第三文件；

在所述第一时刻之后，所述方法还包括：检测到所述第三文件被修改，将所述第三文件的指示字段的值修改为第二值；

在第二时刻，获取到所述第三文件的指示字段的值不是所述第三值；

确定在所述第一时刻至所述第二时刻的时段内，所述第三文件被修改。

23. 根据权利要求 16-19 任一项所述的方法，其特征在于，所述多个候选重复文件包括第四文件；

所述在第一时刻，尝试将所述多个候选重复文件中每个候选重复文件的指示字段的值修改为第三值，包括：若所述第四文件不满足第一条件，则不将所述第四文件的指示字段修改为第三值；

所述方法还包括：在不能成功将所述第四文件的指示字段的值修改为所述第三值时，确定在所述第一时刻至第二时刻的时段内，所述第四文件被修改。

24.根据权利要求 20-23 任一项所述的方法，其特征在于，所述第一条件包括如下一项或多项条件：在所述第一时刻之前，所述重复文件未被映射到内存；在所述第一时刻之前，所述重复文件被映射到内存，但所述电子设备未对所述内存执行过写入操作。

25.根据权利要求 16-24 任一项所述的方法，其特征在于，在第一时刻之后，所述方法还包括：

在第二时刻，获取所述至少一个重复文件的指示字段的值以及所述至少一个重复文件的索引节点编号；

当满足第三条件时，显示所述第二界面；所述第三条件包括：所述至少一个重复文件的指示字段在所述第二时刻的值为第三值，且，所述至少一个重复文件的索引节点编号与数据库中存储的索引节点编号一致。

26.根据权利要求 13-25 任一项所述的方法，其特征在于，所述指示字段存储在所述电子设备的元数据区中。

27. 一种电子设备，其特征在于，包括：处理器以及存储器；

所述存储器用于存储计算机执行指令，所述处理器与所述存储器耦接，当所述电子设备运行时，所述处理器执行所述存储器存储的所述计算机执行指令，以使所述电子设备执行如权利要求 1-12 中任一项所述的方法，或者，执行如权利要求 13-26 中任一项所述的方法。

28.一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，当所述指令在电子设备上运行时，使得所述电子设备执行如权利要求 1-12 中任一项所述的方法，或者执行如权利要求 13-26 中任一项所述的方法。

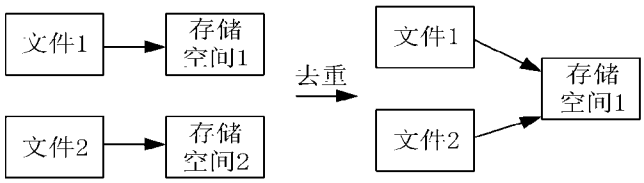

文件1
存储空间1
文件2
存储空间2
去重
文件1
文件2
存储空间1

图 1A

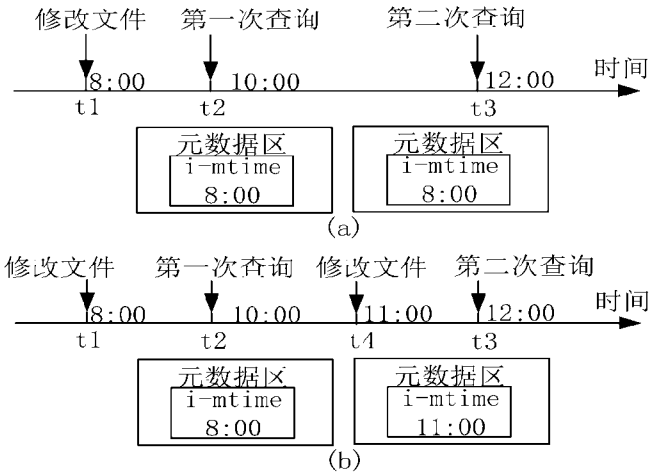

修改文件
第一次查询
第二次查询
8:00
10:00
12:00
时间
t1
t2
t3
元数据区
i-mtime
8:00
元数据区
i-mtime
8:00
(a)
修改文件
第一次查询
修改文件
第二次查询
8:00
10:00
11:00
12:00
时间
t1
t2
t4
t3
元数据区
i-mtime
8:00
元数据区
i-mtime
11:00
(b)

图 1B

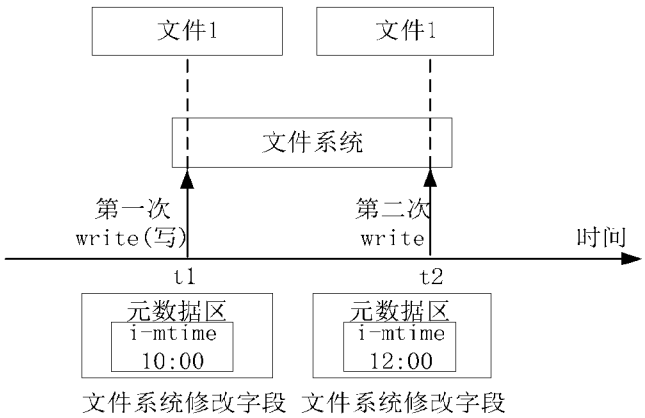

文件1
文件1
文件系统
第一次
write(写)
第二次
write
时间
t1
t2
元数据区
i-mtime
10:00
元数据区
i-mtime
12:00
文件系统修改字段
文件系统修改字段

图 2

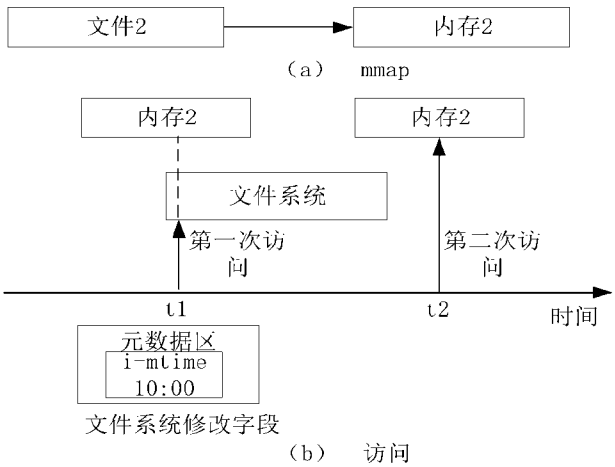

文件2
内存2
(a) mmap
内存2
内存2
文件系统
第一次访问
第二次访问
t1
t2
时间
元数据区
i-mtime
10:00
文件系统修改字段
(b) 访问

图 3

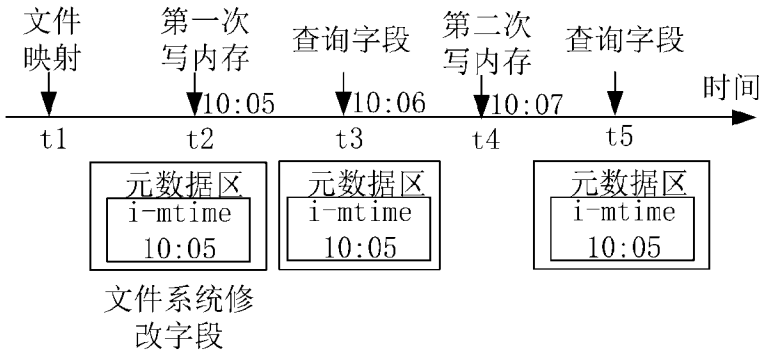

文件映射
第一次写内存
查询字段
第二次写内存
查询字段
时间
10:05
10:06
10:07
t1
t2
t3
t4
t5
元数据区
i-mtime
10:05
元数据区
i-mtime
10:05
元数据区
i-mtime
10:05
文件系统修改字段

图 4

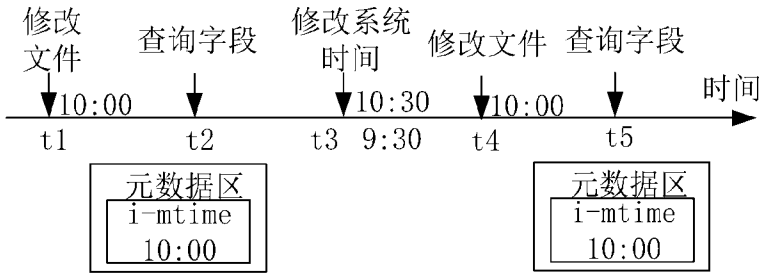

修改文件
查询字段
修改系统时间
修改文件
查询字段
时间
10:00
10:30
10:00
t1
t2
t3
9:30
t4
t5
元数据区
i-mtime
10:00
元数据区
i-mtime
10:00

图 5

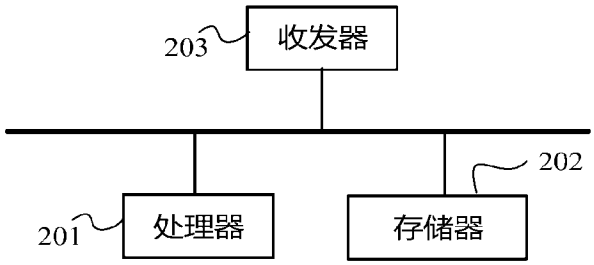

203
收发器
202
201
处理器
存储器

图 6

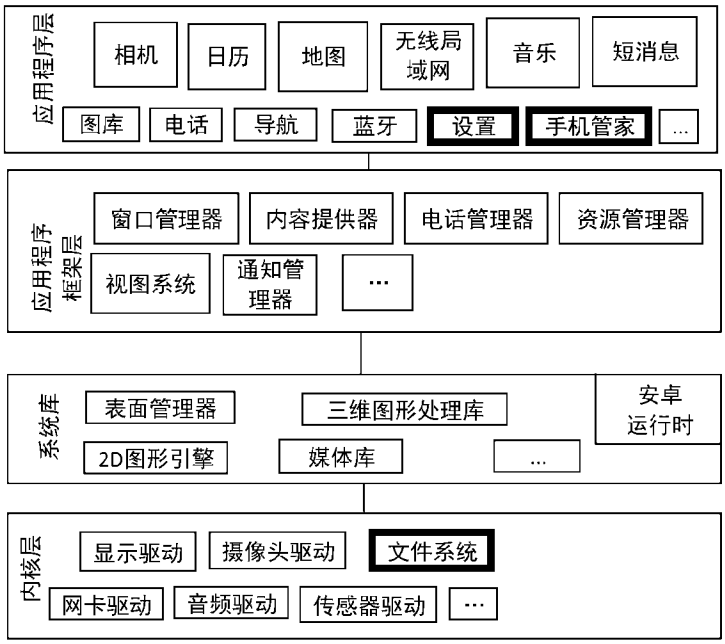

应用程序层
相机
日历
地图
无线局域网
音乐
短消息
图库
电话
导航
蓝牙
设置
手机管家
...
应用程序框架层
窗口管理器
内容提供器
电话管理器
资源管理器
视图系统
通知管理器
...
系统库
表面管理器
三维图形处理库
安卓运行时
2D图形引擎
媒体库
...
内核层
显示驱动
摄像头驱动
文件系统
网卡驱动
音频驱动
传感器驱动
...

图 7

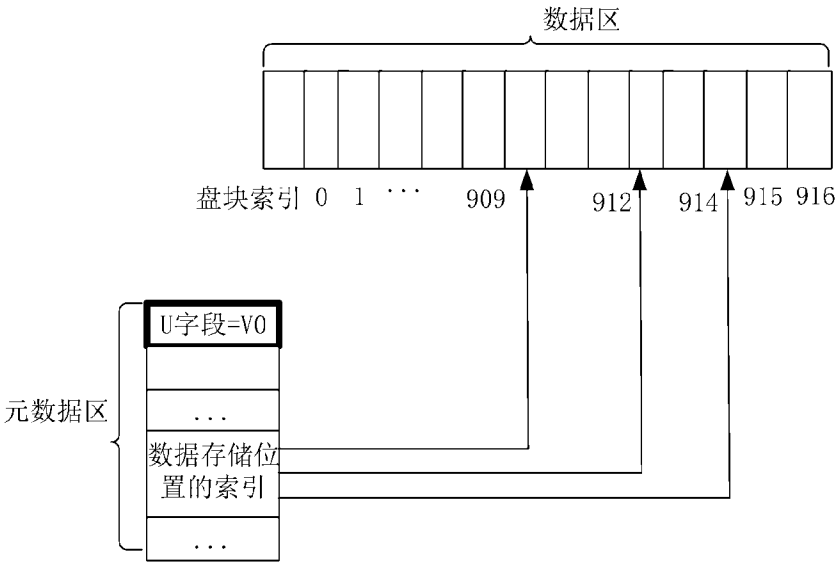

数据区
盘块索引 0 1 ··· 909 912 914 915 916
元数据区
U字段=V0
...
数据存储位置的索引
...

图 8

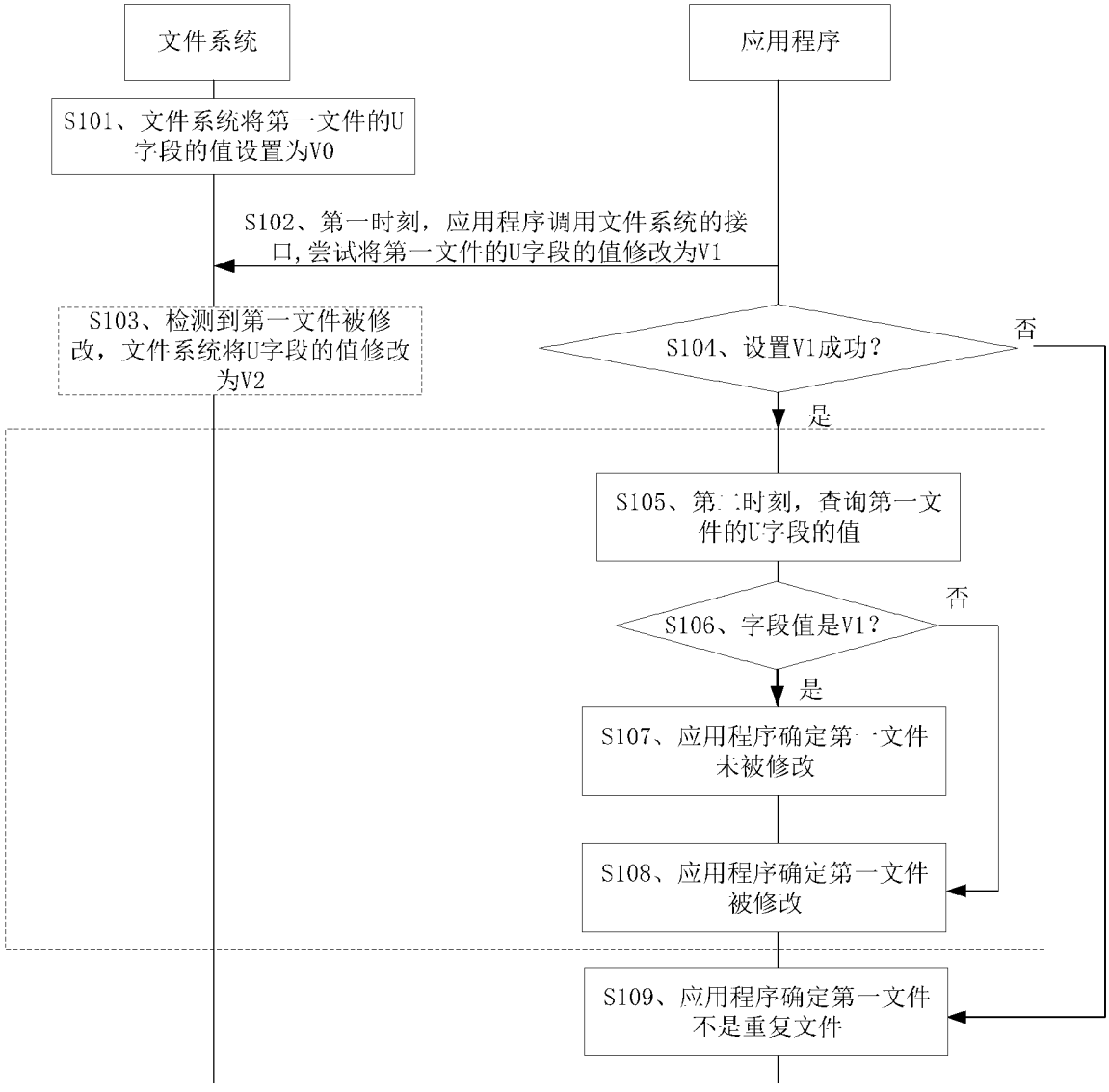
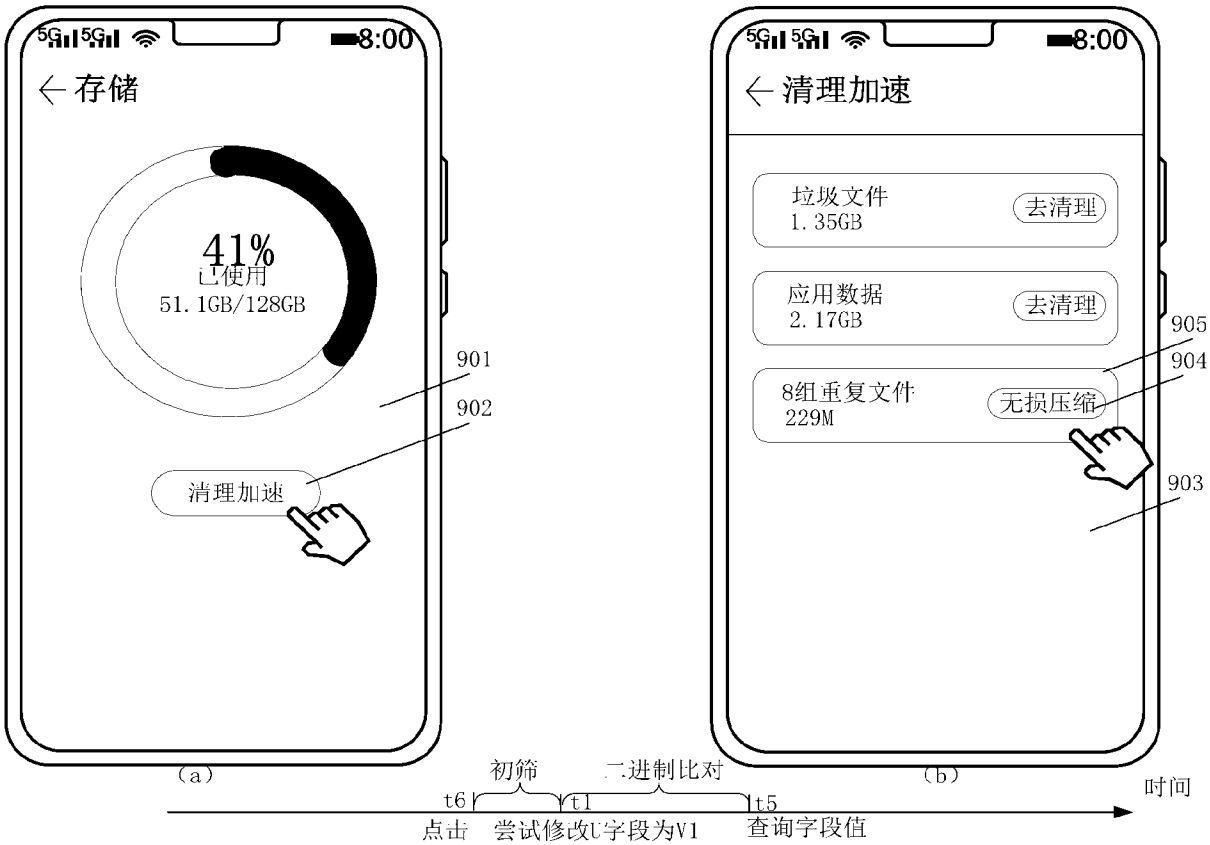

文件系统
应用程序
S101、文件系统将第一文件的U字段的值设置为V0
S102、第一时刻，应用程序调用文件系统的接口，尝试将第一文件的U字段的值修改为V1
S103、检测到第一文件被修改，文件系统将U字段的值修改为V2
S104、设置V1成功？
否
是
S105、第二时刻，查询第一文件的U字段的值
S106、字段值是V1？
否
是
S107、应用程序确定第一文件未被修改
S108、应用程序确定第一文件被修改
S109、应用程序确定第一文件不是重复文件

图 9


8:00
存储
41%
已使用
51.1GB/128GB
901
902
清理加速
(a)
8:00
清理加速
垃圾文件
1.35GB
去清理
应用数据
2.17GB
去清理
8组重复文件
229M
无损压缩
905
904
903
(b)
初筛
二进制比对
t6
t1
t5
时间
点击
尝试修改U字段为V1
查询字段值

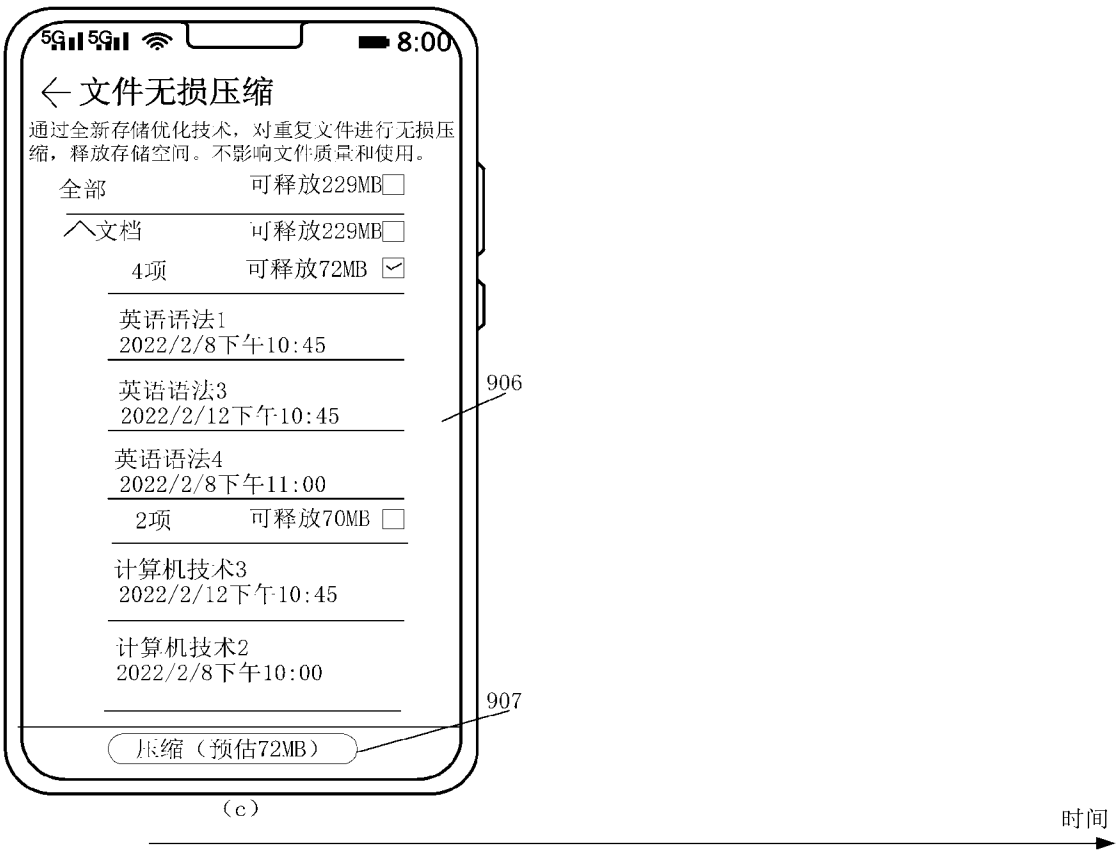

8:00
文件无损压缩
通过全新存储优化技术，对重复文件进行无损压缩，释放存储空间。不影响文件质量和使用。
全部
可释放229MB
文档
可释放229MB
4项
可释放72MB
英语语法1
2022/2/8下午10:45
英语语法3
2022/2/12下午10:45
英语语法4
2022/2/8下午11:00
2项
可释放70MB
计算机技术3
2022/2/12下午10:45
计算机技术2
2022/2/8下午10:00
906
907
压缩（预估72MB）
(c)
时间

图 10A

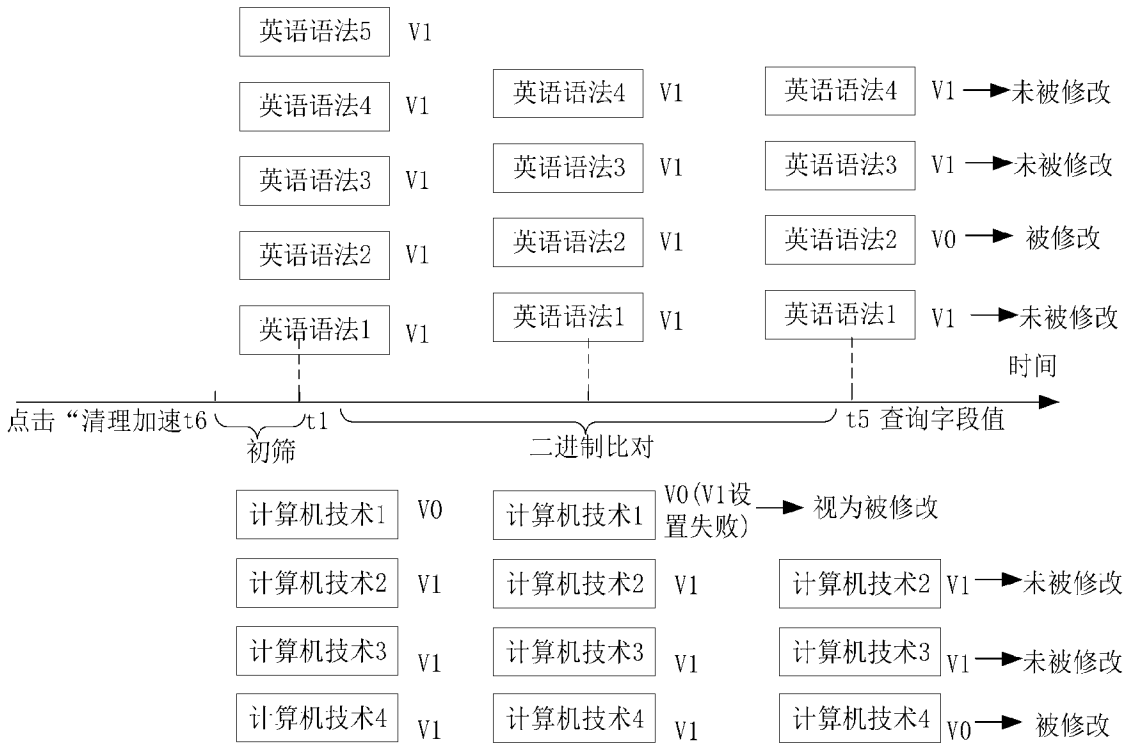


图 10B

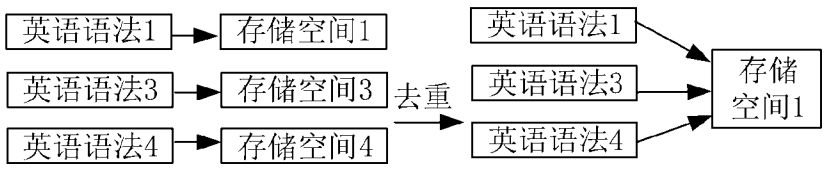


图 10C

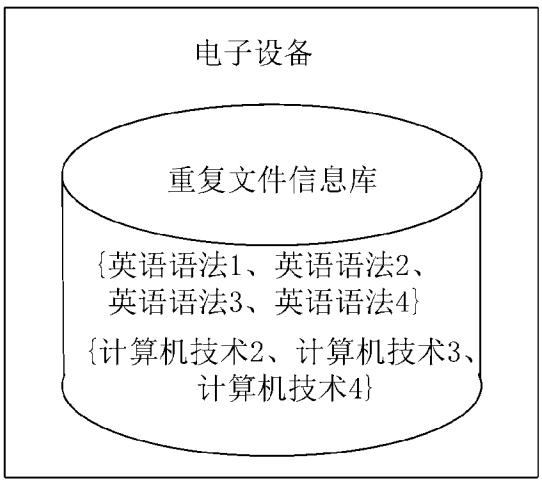


图 11

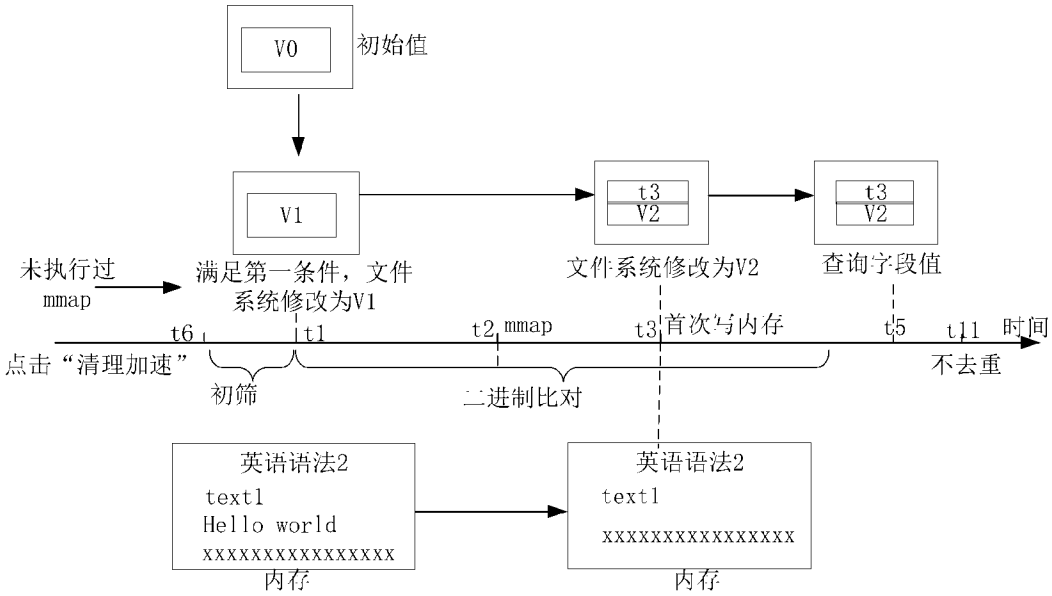

V0
初始值
V1
t3
V2
t3
V2
未执行过
mmap
满足第一条件，文件
系统修改为V1
文件系统修改为V2
查询字段值
t6
t1
t2 mmap
t3 首次写内存
t5
t11
时间
点击“清理加速”
初筛
二进制比对
不去重
英语语法2
text1
Hello world
xxxxxxxxxxxxxxxx
内存
英语语法2
text1
xxxxxxxxxxxxxxxx
内存

图 12

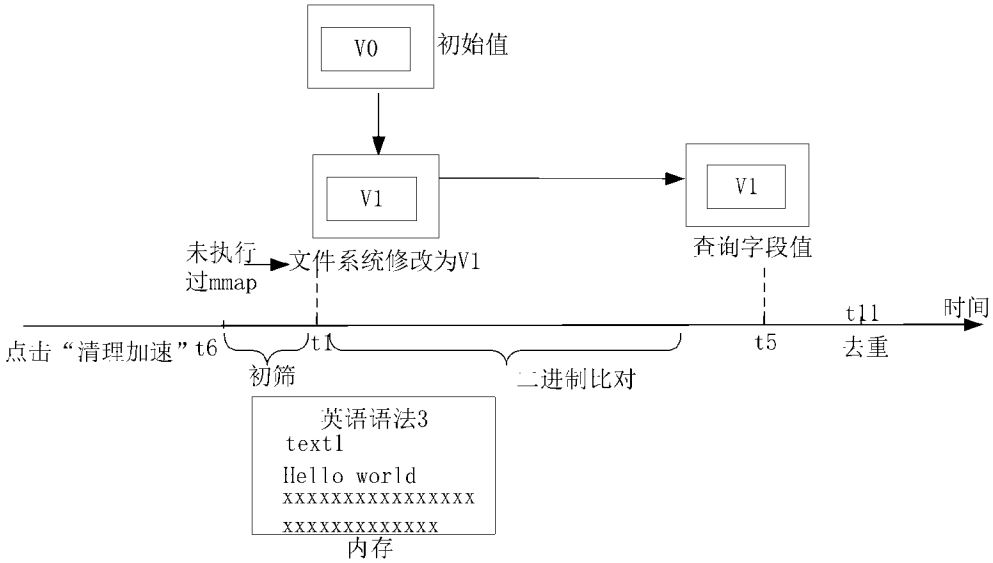

V0
初始值
V1
V1
未执行
过mmap
文件系统修改为V1
查询字段值
t11
时间
点击“清理加速”t6
t1
t5
去重
初筛
二进制比对
英语语法3
text1
Hello world
xxxxxxxxxxxxxxxx
xxxxxxxxxxxxx
内存

图 13

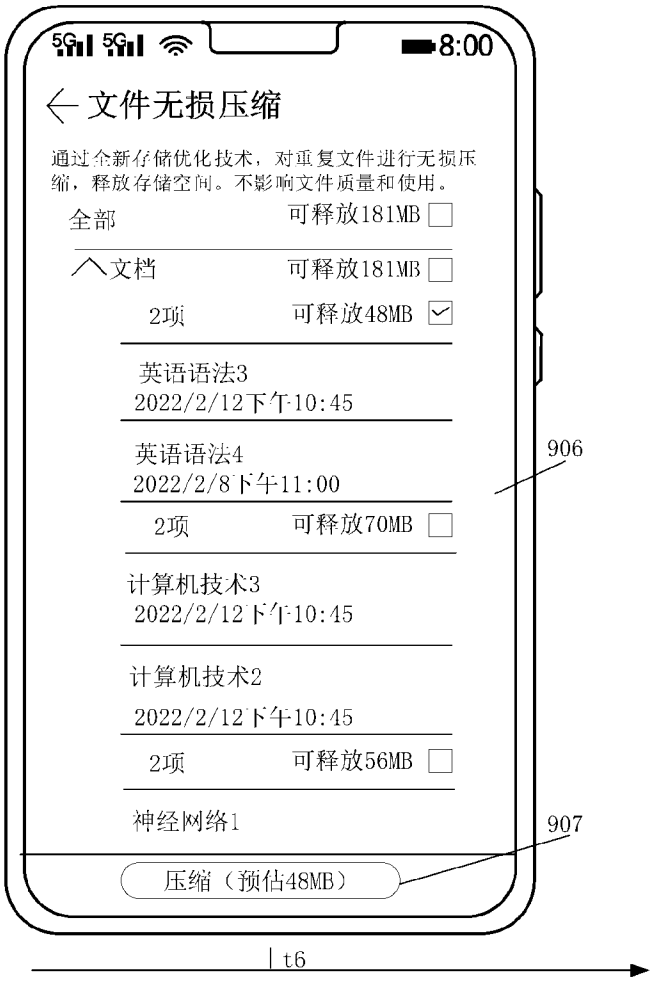

8:00
文件无损压缩
通过全新存储优化技术，对重复文件进行无损压缩，释放存储空间。不影响文件质量和使用。
全部
可释放181MB
文档
可释放181MB
2项
可释放48MB
英语语法3
2022/2/12下午10:45
英语语法4
2022/2/8下午11:00
906
2项
可释放70MB
计算机技术3
2022/2/12下午10:45
计算机技术2
2022/2/12下午10:45
2项
可释放56MB
神经网络1
907
压缩（预估48MB）
t6

图 14

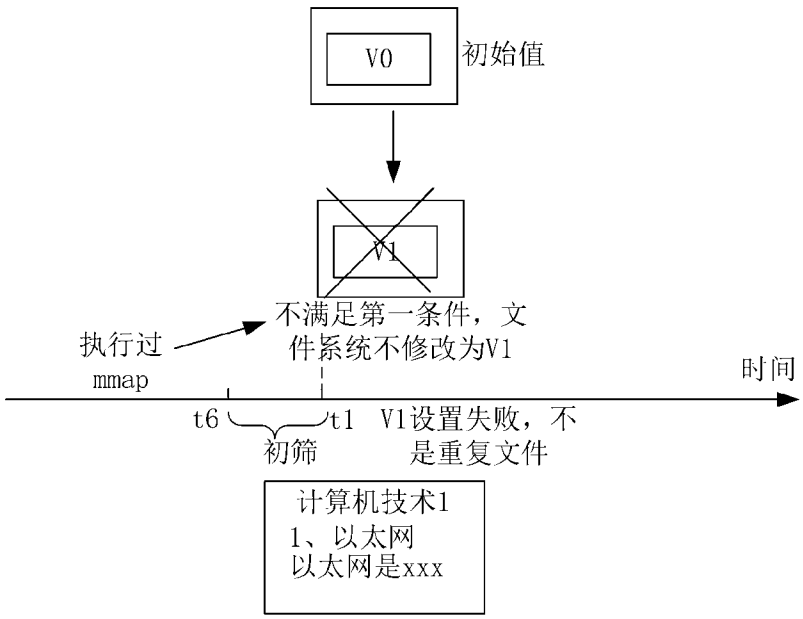

V0
初始值
V1
不满足第一条件，文件系统不修改为V1
执行过mmap
时间
t6
t1
初筛
V1设置失败，不是重复文件
计算机技术1
1、以太网
以太网是xxx

图 15

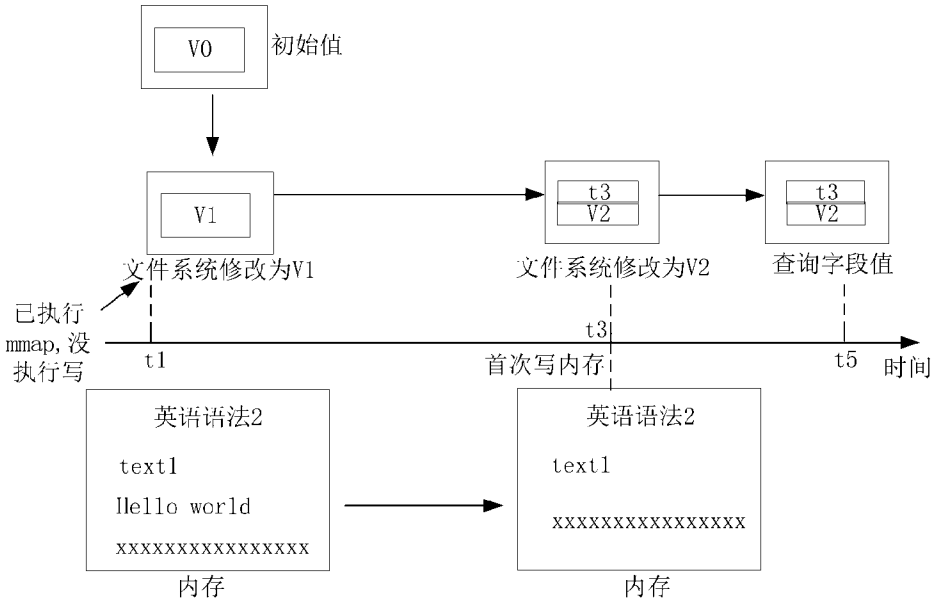

V0
初始值
V1
文件系统修改为V1
t3
V2
文件系统修改为V2
t3
V2
查询字段值
已执行
mmap, 没
执行写
t1
t3
首次写内存
t5
时间
英语语法2
text1
Hello world
xxxxxxxxxxxxxxxx
内存
英语语法2
text1
xxxxxxxxxxxxxxxx
内存

图 16

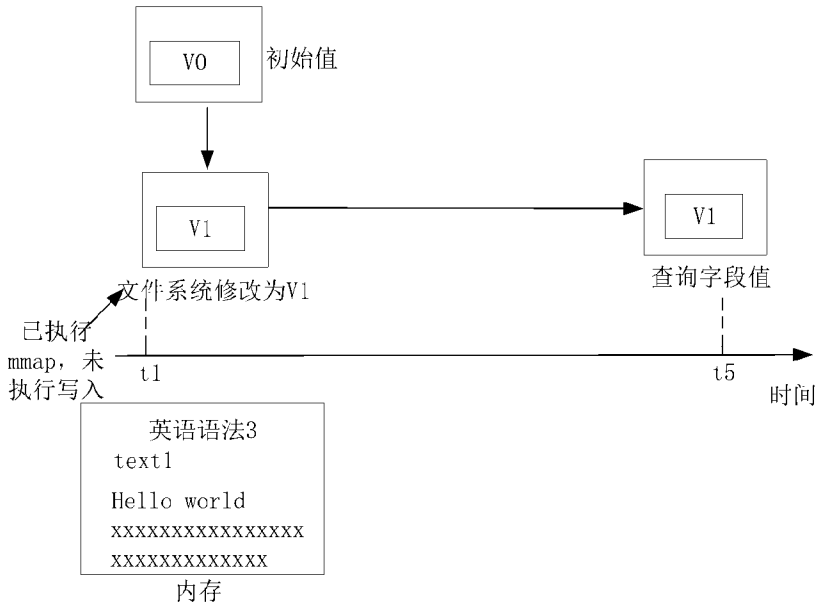

V0
初始值
V1
文件系统修改为V1
V1
查询字段值
已执行
mmap，未
执行写入
t1
t5
时间
英语语法3
text1
Hello world
xxxxxxxxxxxxxxxx
xxxxxxxxxxxxx
内存

图 17

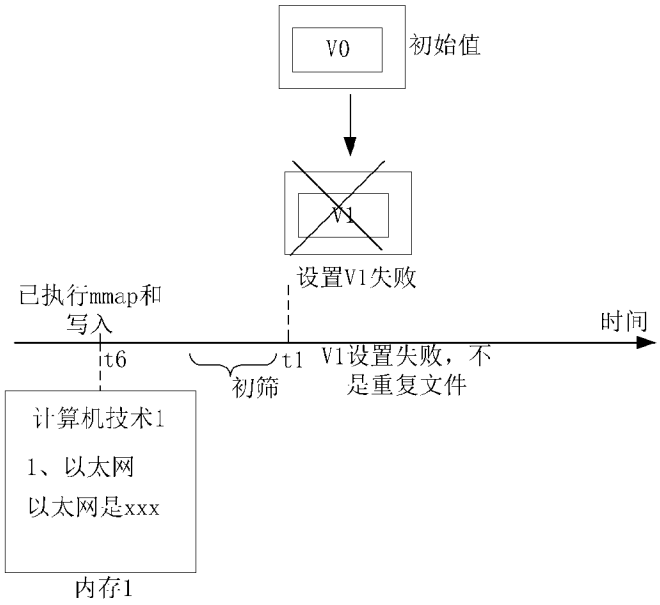

V0
初始值
V1
设置V1失败
已执行mmap和
写入
时间
t6
t1
初筛
V1设置失败，不
是重复文件
计算机技术1
1、以太网
以太网是xxx
内存1

图 18

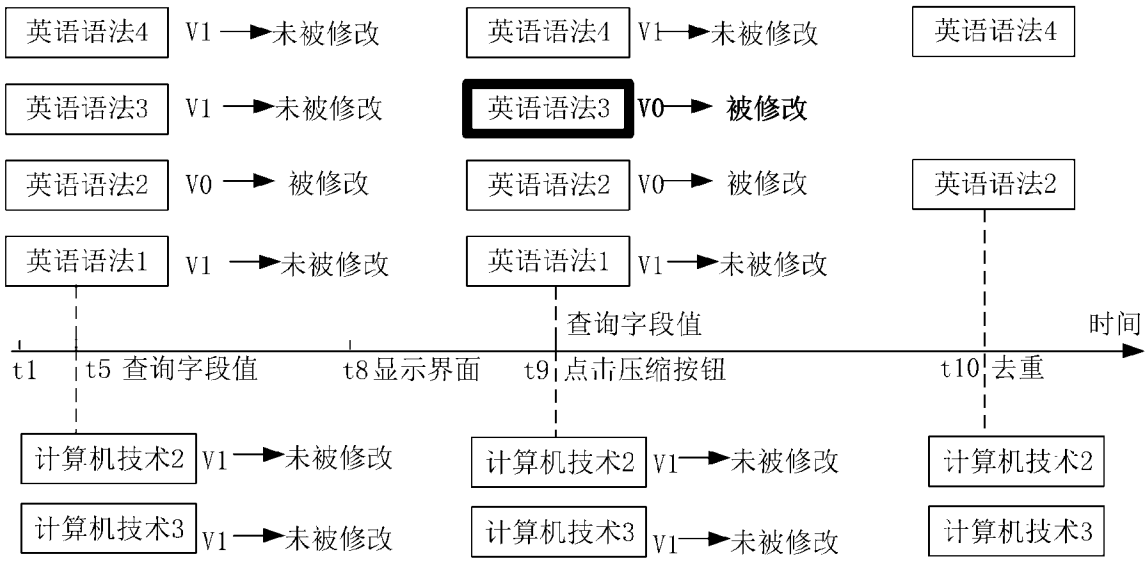

英语语法4
V1
未被修改
英语语法3
V1
未被修改
英语语法2
V0
被修改
英语语法1
V1
未被修改
英语语法4
V1
未被修改
英语语法3
V0
被修改
英语语法2
V0
被修改
英语语法1
V1
未被修改
英语语法4
英语语法2
查询字段值
时间
t1
t5 查询字段值
t8显示界面
t9
点击压缩按钮
t10
去重
计算机技术2
V1
未被修改
计算机技术3
V1
未被修改
计算机技术2
V1
未被修改
计算机技术3
V1
未被修改
计算机技术2
计算机技术3

图 19

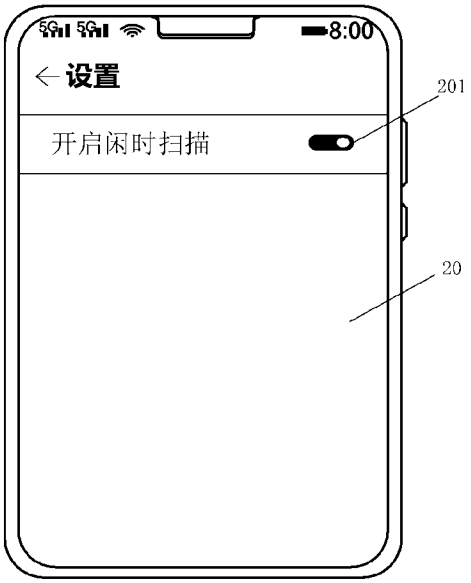

8:00
设置
开启闲时扫描
201
20

图 20

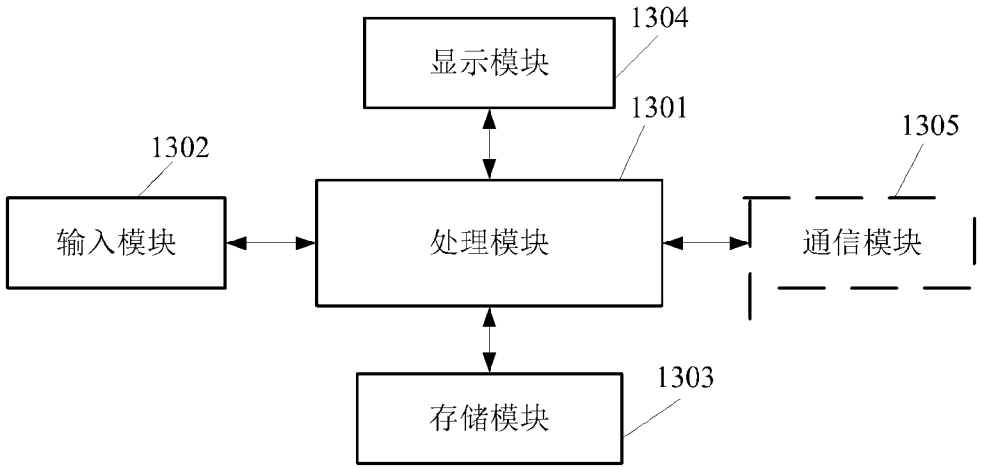

显示模块
1304
1301
1302
1305
输入模块
处理模块
通信模块
存储模块
1303

图 21

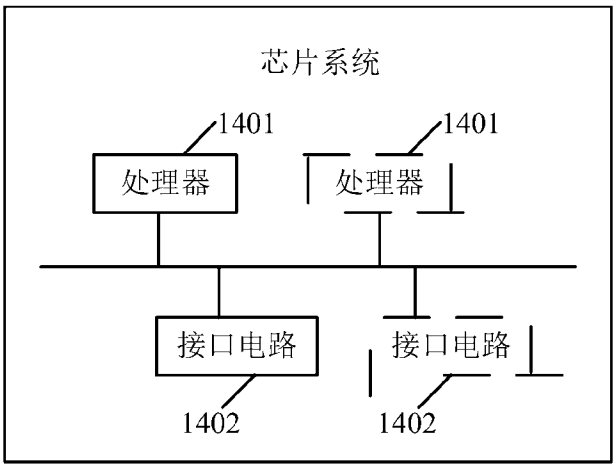

芯片系统
1401
1401
处理器
处理器
接口电路
接口电路
1402
1402

图 22

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2023/116818

A. CLASSIFICATION OF SUBJECT MATTER

G06F16/21(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, H04L

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)

VEN, CNABS, CNTXT, WOTXT, EPTXT, USTXT, CNKI, IEEE: 文件, 修改, 改变, 编辑, 更新, 时间, 时刻, 时段, 重复, 去重, 同步, 同名, 聚类, 写内存, 写接口, 压缩, file, modify, change, edit, update, time, period, duplication, synchronization, simulation, clustering, write, memory, interface, compression

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 112783848 A (HANGZHOU DTDREAM TECHNOLOGY CO., LTD.) 11 May 2021 (2021-05-11) description, paragraphs 0048-0133, and figures 1-4	1-12, 27-28
X	CN 114880299 A (BEIJING BYTEDANCE NETWORK TECHNOLOGY CO., LTD.) 09 August 2022 (2022-08-09) description, paragraphs 0042-0079 and 0289-0300, and figures 1-10	13-15, 26-28
A	CN 114116611 A (HUAWEI TECHNOLOGIES CO., LTD.) 01 March 2022 (2022-03-01) entire document	1-28
A	CN 108647317 A (NEUSOFT CORP.) 12 October 2018 (2018-10-12) entire document	1-28
A	EP 0774715 A1 (STAC ELECTRONICS) 21 May 1997 (1997-05-21) entire document	1-28

☐ 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
"D" document cited by the applicant in the international application
"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	Date of mailing of the international search report
14 November 2023	**17 November 2023**

Name and mailing address of the ISA/CN	Authorized officer
China National Intellectual Property Administration (ISA/CN) **China No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088**	Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2023/116818

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN 112783848 A	11 May 2021	None	
CN 114880299 A	09 August 2022	None	
CN 114116611 A	01 March 2022	None	
CN 108647317 A	12 October 2018	None	
EP 0774715 A1	21 May 1997	JPH 1049416 A	20 February 1998
		US 2002107877 A1	08 August 2002

国际检索报告

国际申请号
PCT/CN2023/116818

A. 主题的分类

G06F16/21(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，H04L

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））

VEN，CNABS，CNTXT，WOTXT，EPTXT，USTXT，CNKI，IEEE:文件，修改，改变，编辑，更新，时间，时刻，时段，重复，去重，同步，同名，聚类，写内存，写接口，压缩，file，modify，change，edit，update，time，period，duplication，synchronization，simulation，clustering，write，memory，interface，compression

C. 相关文件

类 型*	引用文件，必要时，指明相关段落	相关的权利要求
A	CN 112783848 A (杭州数梦工场科技有限公司) 2021年5月11日 (2021 - 05 - 11) 说明书第0048-0133段、附图1-4	1-12，27-28
X	CN 114880299 A (北京字跳网络技术有限公司) 2022年8月9日 (2022 - 08 - 09) 说明书第0042-0079，0289-0300段、附图1-10	13-15，26-28
A	CN 114116611 A (华为技术有限公司) 2022年3月1日 (2022 - 03 - 01) 全文	1-28
A	CN 108647317 A (东软集团股份有限公司) 2018年10月12日 (2018 - 10 - 12) 全文	1-28
A	EP 0774715 A1 (STAC ELECTRONICS) 1997年5月21日 (1997 - 05 - 21) 全文	1-28

☐ 其余文件在C栏的续页中列出。

☑ 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

“D” 申请人在国际申请中引证的文件

“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
2023年11月14日	2023年11月17日
ISA/CN的名称和邮寄地址 中国国家知识产权局 中国北京市海淀区蓟门桥西土城路6号 100088	受权官员 张雯 电话号码 (+86) 010-53961314

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2023/116818

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	112783848	A	2021年5月11日	无			
CN	114880299	A	2022年8月9日	无			
CN	114116611	A	2022年3月1日	无			
CN	108647317	A	2018年10月12日	无			
EP	0774715	A1	1997年5月21日	JPH	1049416	A	1998年2月20日
				US	2002107877	A1	2002年8月8日