



(12) 发明专利申请

(10) 申请公布号 CN 103294749 A

(43) 申请公布日 2013. 09. 11

(21) 申请号 201310024553. 6

(22) 申请日 2013. 01. 23

(30) 优先权数据

2012-024011 2012. 02. 07 JP

(71) 申请人 株式会社日立解决方案

地址 日本东京都

(72) 发明人 西田晋平 古谷沙织

(74) 专利代理机构 北京银龙知识产权代理有限公司

11243

代理人 许静 郭凤麟

(51) Int. Cl.

G06F 17/30(2006. 01)

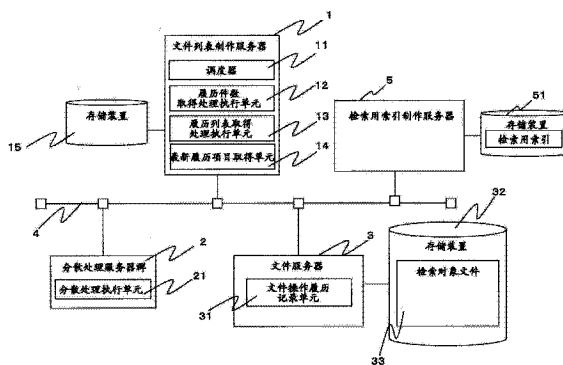
权利要求书2页 说明书7页 附图4页

(54) 发明名称

文件列表生成方法和系统以及程序、文件列表生成装置

(57) 摘要

本发明的文件列表生成方法分析操作履历列表, 高速地制作对同一文件仅包含最新的操作的列表、即在文件服务器内追加、变更、删除的文件数据的最新的一览即差分列表。其特征在于, 具备以下步骤: 第一步骤, 从管理检索对象文件的文件服务器中取得在上一次检索索引制作处理以后对检索对象文件实施的文件数据的追加、变更、删除的操作履历列表; 以及第二步骤, 在取得的操作履历列表中对同一文件包含多个操作履历的情况下取得最新的操作履历之后, 与其它文件的操作履历一起汇总该操作履历, 作为与上一次检索索引制作处理以后对检索对象文件实施的操作履历列表之间的差分列表而输出。



1. 一种文件列表生成方法,其特征在于,包括以下步骤:

第一步骤,从管理检索对象文件的文件服务器中取得在上一次检索索引制作处理以后对检索对象文件实施的文件数据的追加、变更、删除的操作履历列表;以及

第二步骤,在取得的操作履历列表中对同一文件包含多个操作履历的情况下仅取得最新的操作履历之后,与其它文件的操作履历一起汇总该操作履历,作为与上一次检索索引制作处理以后对检索对象文件实施的操作履历列表之间的差分列表而输出。

2. 根据权利要求1所述的文件列表生成方法,其特征在于,

在取得上述操作履历列表的第一步骤中取得的操作履历件数为规定件数以上的情况下,在上述第二步骤中通过多个分散处理服务器按照每个文件路径来分担差分列表的生成而进行处理之后,对每个文件路径的差分列表进行汇总而输出。

3. 根据权利要求1所述的文件列表生成方法,其特征在于,

将从上一次检索索引制作处理以后至当前为止的期间分割为多个期间,在上述第一步骤中按照每个该分割期间取得各期间的操作履历列表,将取得的该操作履历列表的处理分配到多个分散处理服务器,在上述第二步骤中在分配到各分散处理服务器的操作履历列表中对同一文件包含多个操作履历的情况下仅取得最新的操作履历之后,与由其它分散处理服务器进行了分散处理的操作履历一起汇总该操作履历,作为与上一次检索索引制作处理以后对检索对象文件实施的操作履历列表之间的差分列表而输出。

4. 一种文件列表生成系统,其特征在于,具备:

第一单元,其从管理检索对象文件的文件服务器中取得在上一次检索索引制作处理以后对检索对象文件实施的文件数据的追加、变更、删除的操作履历列表;

第二单元,其在取得的操作履历列表中对同一文件包含多个操作履历的情况下仅取得最新的操作履历之后,与其它文件的操作履历一起汇总该操作履历,作为与上一次检索索引制作处理以后对检索对象文件实施的操作履历列表之间的差分列表而输出。

5. 根据权利要求4所述的文件列表生成系统,其特征在于,

在通过上述第一单元取得的操作履历件数为规定件数以上的情况下,在上述第二单元通过多个分散处理服务器按照每个文件路径来分担差分列表的生成而进行处理之后,按照每个文件路径来汇总差分列表而输出。

6. 根据权利要求4所述的文件列表生成系统,其特征在于,

将从上一次检索索引制作处理以后至当前为止的期间分割为多个期间,通过上述第一单元按照每个该分割期间取得各期间的操作履历列表,将取得的该操作履历列表的处理分配到多个分散处理服务器,在分配到各分散处理服务器的操作履历列表中对同一文件包含多个操作履历的情况下仅取得最新的操作履历之后,上述第二单元与由其它分散处理服务器进行了分散处理的操作履历一起汇总该操作履历,作为与上一次检索索引制作处理以后对检索对象文件实施的操作履历列表之间的差分列表而输出。

7. 一种文件列表生成程序,在文件列表制作服务器中使用,其特征在于,使上述文件列表制作服务器作为以下部分而发挥功能:

第一单元,其从管理检索对象文件的文件服务器中取得在上一次检索索引制作处理以后对检索对象文件实施的文件数据的追加、变更、删除的操作履历列表;

第二单元,其在取得的操作履历列表中对同一文件包含多个操作履历的情况下仅取得

最新的操作履历之后,与其它文件的操作履历一起汇总该操作履历,作为与上一次检索索引制作处理以后对检索对象文件实施的操作履历列表之间的差分列表而输出。

8. 根据权利请求 7 所述的文件列表生成程序,其特征在于,

在通过上述第一单元取得的操作履历件数为规定件数以上的情况下,使第二单元作为以下单元而发挥功能:通过多个分散处理服务器按照每个文件路径来分担差分列表的生成而进行处理之后,按照每个文件路径汇总差分列表而输出。

9. 根据权利请求 7 所述的文件列表生成程序,其特征在于,

将从上一次检索索引制作处理以后至当前为止的期间分割为多个期间,通过上述第一单元按照每个该分割期间取得各期间的操作履历列表,将取得的该操作履历列表的处理分配到多个分散处理服务器,使得作为第二单元而发挥功能:在分配到各分散处理服务器的操作履历列表中对同一文件包含多个操作履历的情况下仅取得最新的操作履历之后,与由其它分散处理服务器进行了分散处理的操作履历一起汇总该操作履历,作为与上一次检索索引制作处理以后对检索对象文件实施的操作履历列表之间的差分列表而输出。

10. 一种文件列表生成装置,其特征在于,具备:

第一单元,其从管理检索对象文件的文件服务器中取得在上一次检索索引制作处理以后对检索对象文件实施的文件数据的追加、变更、删除的操作履历列表;以及

第二单元,其在取得的操作履历列表中对同一文件包含多个操作履历的情况下仅取得最新的操作履历之后,与其它文件的操作履历一起汇总该操作履历,作为与上一次检索索引制作处理以后对检索对象文件实施的操作履历列表之间的差分列表而输出。

11. 根据权利请求 10 所述的文件列表生成装置,其特征在于,

上述第一单元在取得的操作履历件数为规定件数以上的情况下,在上述第二单元通过多个分散处理服务器按照每个文件路径分担差分列表的生成而进行处理之后,对每个文件路径汇总差分列表而输出。

12. 根据权利请求 10 所述的文件列表生成装置,其特征在于,

将从上一次检索索引制作处理以后至当前为止的期间分割为多个期间,通过上述第一单元按照每个该分割期间来取得各期间的操作履历列表,将取得的该操作履历列表的处理分配到多个分散处理服务器,在分配到各分散处理服务器的操作履历列表中对同一文件包含多个操作履历的情况下仅取得最新的操作履历之后,上述第二单元与由其它分散处理服务器进行了分散处理的操作履历一起汇总该操作履历,作为与上一次检索索引制作处理以后对检索对象文件实施的操作履历列表之间的差分列表而输出。

## 文件列表生成方法和系统以及程序、文件列表生成装置

### 技术领域

[0001] 本发明涉及一种针对在检索系统中要更新检索索引的文件数据生成其列表的方法和系统以及程序、装置,特别涉及一种在检索索引制作对象的文件服务器具备取得对文件服务器内数据进行的操作履历的接口的情况下根据使用该接口取得的操作履历来高速地制作文件服务器内数据中的、追加、更新、删除的文件的一览的方法、系统、程序以及装置。

### 背景技术

[0002] 近年来,随着计算机性能的高速化、HDD 的大容量化,需要制作庞大数量的非定型文档。因此,从庞大数量的文档中能够高速且准确地搜索出所需的文档的检索系统的必要性提高。为了使检索结果准确,针对保存了成为检索对象的非定型文档的文件服务器,将在制作检索索引之后进行的文件数据的追加、变更、删除操作及时反映到检索索引变得重要。在将这些反映到检索索引时,当还对没有变更的文件数据的检索索引进行更新时,花费很多时间,因此通常仅缩小为追加、变更、删除的文件数据而更新检索索引。因此,需要制作追加、变更、删除的文件数据的一览。

[0003] 为了满足这种检索系统的需求,还存在以下文件服务器:在文件服务器中存储文件数据的操作履历,具备根据来自外部的请求来提供操作履历的接口。

[0004] 已知在以下专利文献 1 中公开的先行技术。

[0005] 专利文献 1:日本特开 2006-268456 号

### 发明内容

[0006] 发明要解决的问题

[0007] 为了将文件数据的追加、变更、删除操作及时反映到检索索引,考虑使用以下方法:使用多个服务器进行分散处理,由此高速地实施与文件服务器内的检索对象文件有关的检索索引的新制作处理、追加、变更、删除的文件的检索索引更新处理。

[0008] 在制作索引、制作更新对象文件数据的一览时,在文件服务器提供返回更新文件的一览的接口的情况下,利用该接口即可,但是在没有具备这种接口的文件服务器的情况下,通常需要扫描存在于文件服务器内的检索索引制作对象范围的全部文件数据,列举对象文件、判断有无更新操作。

[0009] 特别在进行索引更新时,即使在追加、变更、删除的文件数据为少量的情况下,需要扫描全部文件数据,因此追加、变更、删除的文件数据的一览的制作处理导致索引更新处理长时间化。

[0010] 另一方面,考虑以下方法:在文件服务器具备返回文件操作的履历列表的接口的情况下,使用该接口来取得操作履历列表,按照操作履历列表对检索索引反映追加、变更、删除。但是,有时在操作履历列表中对同一文件包含多个操作履历。在该情况下,如果不沿着操作履历列表的时间序列来更新检索索引则无法正确地更新检索索引。在需要这种逐步

的处理的情况下,例如在分割操作履历列表而使用多个服务器进行分散处理而更新检索索引的情况下,在更新检索索引之前需要按照操作履历列表的时间序列顺序来排列应用分散处理结果,假设即使高速地完成操作履历列表的后半部分的处理,也必须等到操作履历列表的前半部分的处理完成为止。在操作履历列表中对同一文件包含多个操作履历这一情况导致需要逐步处理。

[0011] 本发明的目的在于,分析操作履历列表,制作对同一文件仅包含最新的操作的列表、即在文件服务器内追加、变更、删除的文件数据的最新一览(以下称为差分列表),并且,利用分散处理服务器群对从该履历列表向差分列表转换的转换处理进行分散执行,由此将大容量存储器返回的大的操作履历列表高速地转换为差分列表,由此能够容易地执行检索索引的新制作、更新处理的分散执行,使检索索引高速地成为最新的状态。

#### [0012] 用于解决问题的方案

[0013] 为了解决上述问题,本发明所涉及的文件列表生成方法的特征在于,具备以下步骤:第一步骤,从管理检索对象文件的文件服务器中取得在上一次检索索引制作处理以后对检索对象文件实施的文件数据的追加、变更、删除的操作履历列表;以及第二步骤,在取得的操作履历列表中对同一文件包含多个操作履历的情况下取得最新的操作履历之后,与其它文件的操作履历一起汇总该操作履历,作为与上一次检索索引制作处理以后对检索对象文件实施的操作履历列表之间的差分列表而输出。

[0014] 另外,其特征不在于,在取得上述操作履历列表的第一步骤中取得的操作履历件数为规定件数以上的情况下,在上述第二步骤中通过多个分散处理服务器按照每个文件路径来分担差分列表的生成而进行处理之后,对每个文件路径的差分列表进行汇总而输出。

[0015] 另外,其特征不在于,将从上一次检索索引制作处理以后至当前为止的期间分割为多个期间,在上述第一步骤中按照每个该分割期间取得各期间的操作履历列表,将取得的该操作履历列表的处理分配到多个分散处理服务器,在上述第二步骤中在分配到各分散处理服务器的操作履历列表中对同一文件包含多个操作履历的情况下仅取得最新的操作履历之后,与由其它分散处理服务器进行了分散处理的操作履历一起汇总该操作履历,作为与上一次检索索引制作处理以后对检索对象文件实施的操作履历列表之间的差分列表而输出。

[0016] 另外,本发明所涉及的文件列表生成系统的特征在于,具备:第一单元,其从管理检索对象文件的文件服务器中取得在上一次检索索引制作处理以后对检索对象文件实施的文件数据的追加、变更、删除的操作履历列表;以及第二单元,其在取得的操作履历列表中对同一文件包含多个操作履历的情况下仅取得最新的操作履历之后,与其它文件的操作履历一起汇总该操作履历,作为与上一次检索索引制作处理以后对检索对象文件实施的操作履历列表之间的差分列表而输出。

[0017] 另外,其特征不在于,在上述第一单元中取得的操作履历件数为规定件数以上的情况下,在上述第二单元通过多个分散处理服务器按照每个文件路径来分担差分列表的生成而进行处理之后,按照每个文件路径来汇总差分列表而输出。

[0018] 另外,其特征不在于,将从上一次检索索引制作处理以后至当前为止的期间分割为多个期间,通过上述第一单元按照每个该分割期间取得各期间的操作履历列表,将取得的该操作履历列表的处理分配到多个分散处理服务器,在分配到各分散处理服务器的操作履

历列表中对同一文件包含多个操作履历的情况下在仅取得最新的操作履历之后,上述第二单元与由其它分散处理服务器进行了分散处理的操作履历一起汇总该操作履历,作为与上一次检索索引制作处理以后对检索对象文件实施的操作履历列表之间的差分列表而输出。

[0019] 本发明所涉及的文件列表生成程序使用于文件列表制作服务器,其特征在于,使上述文件列表制作服务器作为以下部分而发挥功能:第一单元,其从管理检索对象文件的文件服务器中取得在上一次检索索引制作处理以后对检索对象文件实施的文件数据的追加、变更、删除的操作履历列表;以及第二单元,其在取得的操作履历列表中对同一文件包含多个操作履历的情况下在仅取得最新的操作履历之后,与其它文件的操作履历一起汇总该操作履历,作为与上一次检索索引制作处理以后对检索对象文件实施的操作履历列表之间的差分列表而输出。

[0020] 另外,其特征在于,在上述第一单元中取得的操作履历件数为规定件数以上的情况下,使第二单元作为以下单元而发挥功能:通过多个分散处理服务器按照每个文件路径来分担差分列表的生成而进行处理之后,按照每个文件路径汇总差分列表而输出。

[0021] 另外,其特征在于,使作为第二单元而发挥功能:将从上一次检索索引制作处理以后至当前为止的期间分割为多个期间,通过上述第一单元按照每个该分割期间取得各期间的操作履历列表,将取得的该操作履历列表的处理分配到多个分散处理服务器,在分配到各分散处理服务器的操作履历列表中对同一文件包含多个操作履历的情况下在仅取得最新的操作履历之后,与由其它分散处理服务器进行了分散处理的操作履历一起汇总该操作履历,作为与上一次检索索引制作处理以后对检索对象文件实施的操作履历列表之间的差分列表而输出。

[0022] 本发明所涉及的文件列表生成装置的特征在于,具备:第一单元,其从管理检索对象文件的文件服务器中取得在上一次检索索引制作处理以后对检索对象文件实施的文件数据的追加、变更、删除的操作履历列表;以及第二单元,其在取得的操作履历列表中对同一文件包含多个操作履历的情况下仅取得最新的操作履历之后,与其它文件的操作履历一起汇总该操作履历,作为与上一次检索索引制作处理以后对检索对象文件实施的操作履历列表之间的差分列表而输出。

[0023] 另外,其特征在于,在上述第一单元在取得的操作履历件数为规定件数以上的情况下,使第二单元作为以下单元而发挥功能:在上述第二单元通过多个分散处理服务器按照每个文件路径分担差分列表的生成而进行处理之后,对每个文件路径汇总差分列表而输出。

[0024] 另外,其特征在于,将从上一次检索索引制作处理以后至当前为止的期间分割为多个期间,通过上述第一单元按照每个该分割期间来取得各期间的操作履历列表,将取得的该操作履历列表的处理分配到多个分散处理服务器,在分配到各分散处理服务器的操作履历列表中对同一文件包含多个操作履历的情况下仅取得最新的操作履历之后,上述第二单元与由其它分散处理服务器进行了分散处理的操作履历一起汇总该操作履历,作为与上一次检索索引制作处理以后对检索对象文件实施的操作履历列表之间的差分列表而输出。

[0025] 发明的效果

[0026] 根据本发明,在保管对文件服务器内的数据进行操作的操作履历而具备根据需要返回履历列表的文件服务器中,取得在上一次检索索引制作处理以后对检索对象文件实施的文

件数据的追加、变更、删除的操作履历列表,在取得的该操作履历列表中对同一文件包含多个操作履历的情况下仅取得最新的操作履历之后,与其它文件的操作履历一起汇总该操作履历,作为与上一次检索索引制作处理以后对检索对象文件实施的操作履历列表之间的差分列表而输出,因此能够高速地制作在文件服务器上追加、变更、删除的文件的一览。

[0027] 由此,能够容易地执行检索索引的新制作、更新处理的分散执行,能够高速地进行检索索引的新制作、更新处理,结果是,能够使检索系统的检索结果尽可能准确。

#### 附图说明

[0028] 图 1 表示本发明所涉及的第一实施方式中的系统结构图。

[0029] 图 2 是文件服务器内的索引对象文件的数据结构图。

[0030] 图 3 是从文件服务器返回的履历列表的数据结构图。

[0031] 图 4 是使用分散处理服务器群来分散取得履历列表而取得各文件的最新的履历项目并转化为差分列表的处理的流程图。

#### [0032] 附图标记说明

[0033] 1:文件列表制作服务器;2:分散处理服务器群;3:文件服务器;4:网络;11:调度器;12:履历件数取得处理执行单元;13:履历列表取得处理执行单元;14:最新履历项目取得单元;15:与文件列表制作服务器 1 连接的存储装置;21:分散处理执行单元;31:文件操作履历记录单元;32:与文件服务器 3 连接的存储装置;33:检索对象文件;201:检索对象文件的具体例;300:操作履历列表;301:在操作履历列表中表示服务器的数据;302:在操作履历列表中表示共享文件夹的数据;303:在操作履历列表中表示文件的数据;304:在操作履历列表表示操作种类的数据;305:在操作履历列表中表示操作时刻的数据。

#### 具体实施方式

[0034] 下面,根据附图详细说明实施本发明的情况下的第一实施方式。

[0035] 图 1 是本发明所涉及的第一实施方式中的系统结构图。

[0036] 在图 1 中示出的文件列表生成系统是通过 LAN(Local Area Network:局域网)4 等有线或者无线的通信线路能够以对文件列表制作服务器 1、分散处理服务器群 2 以及文件服务器 3(以下称为“服务器等”)相互进行通信的方式连接的系统。

[0037] 在图 1 中,服务器等通过 LAN4 能够以相互通信的方式进行连接,但是连接方法并不限定于 LAN,例如也可以通过 WAN(Wide Area Network:广域网)、因特网进行连接。另外,在图 1 中,服务器等在相同 LAN 分段上进行连接,但是该结构仅是一例,也可以形成任何结构。并且,在图 1 中,文件列表制作服务器 1、分散处理服务器群 2 以及文件服务器 3 分别各存在一台,但是也可以分别各存在两台以上。另外,文件列表制作服务器 1、分散处理服务器群 2 以及文件服务器 3 不需要分别为不同的装置,例如,还能够在一台装置上实现文件列表制作服务器 1、分散处理服务器群 2 以及文件服务器 3 的功能。

[0038] 在文件服务器 3 中设置文件操作履历记录单元 31,记录对保存到与文件服务器 3 连接的存储装置 32 中的检索对象文件 33 进行的追加、变更、删除等操作的履历,根据来自使用了 HTTP(Hypertext Transfer Protocol:超文本传送协议)等的客户端的请求来返回履历列表。在图 3 中详细说明该履历列表。

[0039] 通过上述结构,文件列表制作服务器 1 利用分散处理服务器群 2 进行从文件服务器 3 取得履历列表以及更新上一次索引时以后起的差分列表转换处理。

[0040] 文件列表制作服务器 1 是 PC 等装置,能够以与存储装置 15 进行通信的方式进行连接。存储装置 15 是磁盘等装置,内置于文件列表制作服务器 1 或者进行外部连接。存储装置 15 以及文件列表制作服务器 1 的主存储装置等作为文件列表制作服务器 1 的存储单元而发挥功能。

[0041] 文件列表制作服务器 1 具备调度器 11、履历件数取得处理执行单元 12、履历列表取得处理执行单元 13 以及最新履历项目取得单元 14。另外,文件列表制作服务器 1 具备 CPU、主存储装置等,CPU 将在存储装置 15 中存储的履历件数取得处理执行单元 12 等的程序加载到主存储装置上,通过执行其指令码来执行各种处理。

[0042] 在定期地制作差分列表而更新检索索引时,调度器 11 参照在存储装置 15 中存储的一览制作处理执行间隔来执行履历件数取得处理执行单元 12 之后,执行履历列表取得处理执行单元 13,从文件服务器 3 取得履历列表,之后,执行最新履历项目取得单元 14 来仅取得在履历列表内包含的与同一文件有关的操作履历中的最新的操作履历而转换为差分列表。作为从履历列表向差分列表的转换处理(S401 等)在后文中说明该一系列的动作。

[0043] 图 2 是表示文件服务器 3 的检索对象文件 33 的结构的具体例的概念图。

[0044] 在该概念图示出的文件服务器 3 内的检索对象文件 33 的结构中,某一文件服务器 3 的识别名称为“server1”。在“server1”中共享两个目录,分别根据“share1”、“share2”这种共享名称来能够唯一地进行识别。在“server1”的“share1”、“share2”中存在未图示的目录和文件。例如,在“server1”的“share1”中存在“etc”、“doc”这两个目录,在“etc”目录中存在“file1.doc”、“file2.xml”这两个文件,在“doc”目录中存在“file3.doc”这一个文件。同样地,在“server1”的“share2”中存在一个目录“pjt”,在“pjt”目录中存在“pjt1”、“pjt2”、“pjt3”这三个目录,其中,在“pjt1”目录中存在“file4.txt”、“file5.doc”这两个文件。

[0045] 图 3 是使用图 2 的示例来具体地表示文件服务器 3 的文件操作履历记录单元 31 所返回的履历列表的概念图。

[0046] 在履历列表 300 中,首先包括表示是与哪一个服务器有关的履历列表的数据 301。在本例中在 XML 的“objectlogs”这种要素中设置“rootURI”这种名称的属性并加入 `http://server1/` 这种数据,表示图 2 示出的“server1”。

[0047] 接着,包括是与哪一个共享目录有关的履历列表的数据 302。在本例中在 XML 的“container”这种要素中设置“name”这种名称的属性并加入“share1”这种数据,表示图 2 示出的“share1”这种共享文件夹。

[0048] 在各操作履历中包括是与哪一个文件有关的操作还是用于识别文件的数据 303。在本例中,在 XML 的“object”这种要素中设置“uri”这种名称的属性并加入“etc/file1.doc”这种数据,表示处于图 2 示出的“share1”这种共享文件夹属下的“etc”这种文件夹中的“file.doc”这种文件。

[0049] 同时,在操作履历中包括用于表示是进行哪一个操作的数据 304,在本例中,设置“action”这种属性并加入“create”这种数据,表示制作了该文件的情况。

[0050] 作为能够包含于“action”属性的值除了表示对检索对象文件追加了新的文件、目



录的情况的“create”以外,还考虑表示变更了现有的文件、目录的内容或者元数据的情况的“modify”、表示删除了现有的文件、目录的情况的“delete”。

[0051] 并且,在操作履历中包括表示何时进行了该操作的数据 305。在本例中,使用从 1970 年 1 月 1 日 0 时 0 分起的毫秒来表示设置“timestamp”这种属性而进行操作的日期和时间。

[0052] 在操作履历中按照时间序列的顺序来记录了对同一文件进行的操作。数据 306 表示与数据 303 所示的文件相同的文件的其它操作比数据 305 所示的时刻更后进行的情况。

[0053] 这样,按照 XML 的“object”要素的出现顺序来表示对某一服务器 301 的共享文件夹 302 的属下进行的文件或者目录的追加、更新、删除操作的顺序,图 3 的 XML 成为表示对某一服务器 301 的共享文件夹 302 属下进行的操作的履历的履历列表。

[0054] 在图 3 中按照 XML (Extensible Markup Language :可延伸标记语言) 形式描述了履历列表 300,不一定必须是 XML 形式,例如也可以是 JSON (Java Script Object Notation : Java 脚本对象记号) 形式、CSV (Comma Separated Values :逗号分隔值) 形式等其它形式。

[0055] 图 4 是使用分散处理服务器 2 群分散取得履历列表 300 而取得各文件的最新的履历项目并转化为差分列表的处理的流程图。

[0056] 到图 4 的 S401 为止的处理如下。如上所述,在调度器 11 定期地更新检索索引时,参照在存储装置 15 中存储的一览制作处理执行间隔,为了制作差分列表而开始进行处理。

[0057] 此外,检索用索引制作服务器 5 制作检索索引,而保存到存储装置 51。调度器 11 指示检索索引的更新。

[0058] 文件列表制作服务器 1 的履历件数取得处理单元 12 对文件服务器 3 使用 HTTP 等的通信单元,询问并取得从上一次索引更新时起在当前时刻期间的履历列表内包含的履历的件数 (S401)。

[0059] 接着,判断取得的件数是否为最低件数以上 (S402)。该判断是为了避免以下情况而进行的:在列表内的履历件数对通过并列进行从在履历列表内包含的同一文件的履历中取得最新的操作履历项目而转换为差分列表的处理来高速化这种本发明的目的变得太少的情况下,由进行并列处理产生的开销超过并列处理的高速化的优点。此外,最低件数参照在文件列表制作服务器 1 上的设定文件等中存储的值。该值例如是 50000 这种值,期望估计并设定一台的文件列表制作服务器 1 或者分散处理服务器群 2 中的 1 台至少在几分钟以内能够完成从在履历列表内包含的同一文件的履历中取得最新的操作履历项目而转换为差分列表的处理的件数。

[0060] 在 S402 的判断中取得的列表内的履历件数为最低件数以上的情况下,为了在分散处理服务器群 2 中并列地请求履历列表取得请求,而进行分割 (S402)。

[0061] 例如在履历列表件数为 1000000 件的情况下,以每一台的履历列表件数与最低件数相等的方式,分割成第一台的服务器从第一件至第 50000 件,第二台的服务器从第 50001 件至第 100000 件。此外,在该实施方式中使用履历件数取得处理执行单元 12 根据件数来分割履历列表取得请求,但是除了件数以外还考虑通过期间进行分割的方式。即,例如在上一次索引更新为三周前的情况下,考虑以下分割方法:第一台的服务器请求从三周前至两周前为止的一周量的履历列表,第二台的服务器请求从两周前至一周前为止的一周量的履历列表,第三台的服务器从一周前至当前时刻为止的一周量的履历列表。本发明能够具有

以下结构:这样将从上一次索引制作处理日以后至当前为止的期间分割为多个期间,取得每各期间的操作履历列表,将各期间的操作履历列表分配到多个分散处理服务器,汇总输出由各分散处理服务器提取出的最新的操作履历列表。

[0062] 接着,使用 HTTP 等的通信单元对文件服务器 3 请求并取得履历列表 (S404)。此时,在 S402 的判断中列表内的履历件数为最低件数以上的情况下,按照 S403 的分割处理通过分散处理服务器群 2 来并列进行请求。

[0063] 在 S402 的判断中列表内的履历件数小于最低件数的情况下,通过一台的服务器请求并取得履历列表。

[0064] 接着,从取得的履历列表中取得由文件路径、操作种类、操作时刻构成的操作履历 (S405)。能够对表示图 3 的服务器的数据 301、表示共享文件夹的数据 302 以及表示文件的数据 303 进行连接来制作文件路径。

[0065] 接着,统计与同一文件路径有关的操作履历 (S406)。

[0066] 接着,关于各个文件路径,参照操作时刻,仅保留最接近当前时刻的操作的操作履历,丢弃其它履历 (S407)。

[0067] S405、S406、S407 的处理步骤 S402 的判断结果,通过分散处理服务器群 2 来并列进行处理。

[0068] 接着,汇总全部文件路径的操作履历 (S408)。由汇总得到的操作履历构成的列表是通过分散处理服务器群 2 来在并列处理中从同一文件的履历中结束取得最新的操作履历项目的差分列表,在该时间点中完成从履历列表中向差分列表的转换处理。

[0069] 此外,在跨过要分割的件数、期间而产生操作履历的重复的情况下,担心没有完全汇总而产生文件重复。但是,在这种情况下,能够以下那样进行解决。即,在操作履历件数为规定的件数以上的情况下,在多个分散处理服务器中按照每个文件路径通过件数来划分差分列表的生成而分担,但是与同一文件路径有关的操作履历必须通过一个服务器进行差分列表生成。

[0070] 这是由于,首先,按照文件路径的字典顺序来分类取得的履历,通过各分散处理服务器来交换已经分类的履历,以相同的分散处理服务器担当处理的方式决定分担量,由此能够实现相同的文件路径。

[0071] 即使在通过期间分割的情况下,也通过同样的分担方法,即使跨过期间的分割值而存在操作履历的情况下,也不会重复产生排除泄露,能够生成仅保留最新的履历的差分列表。

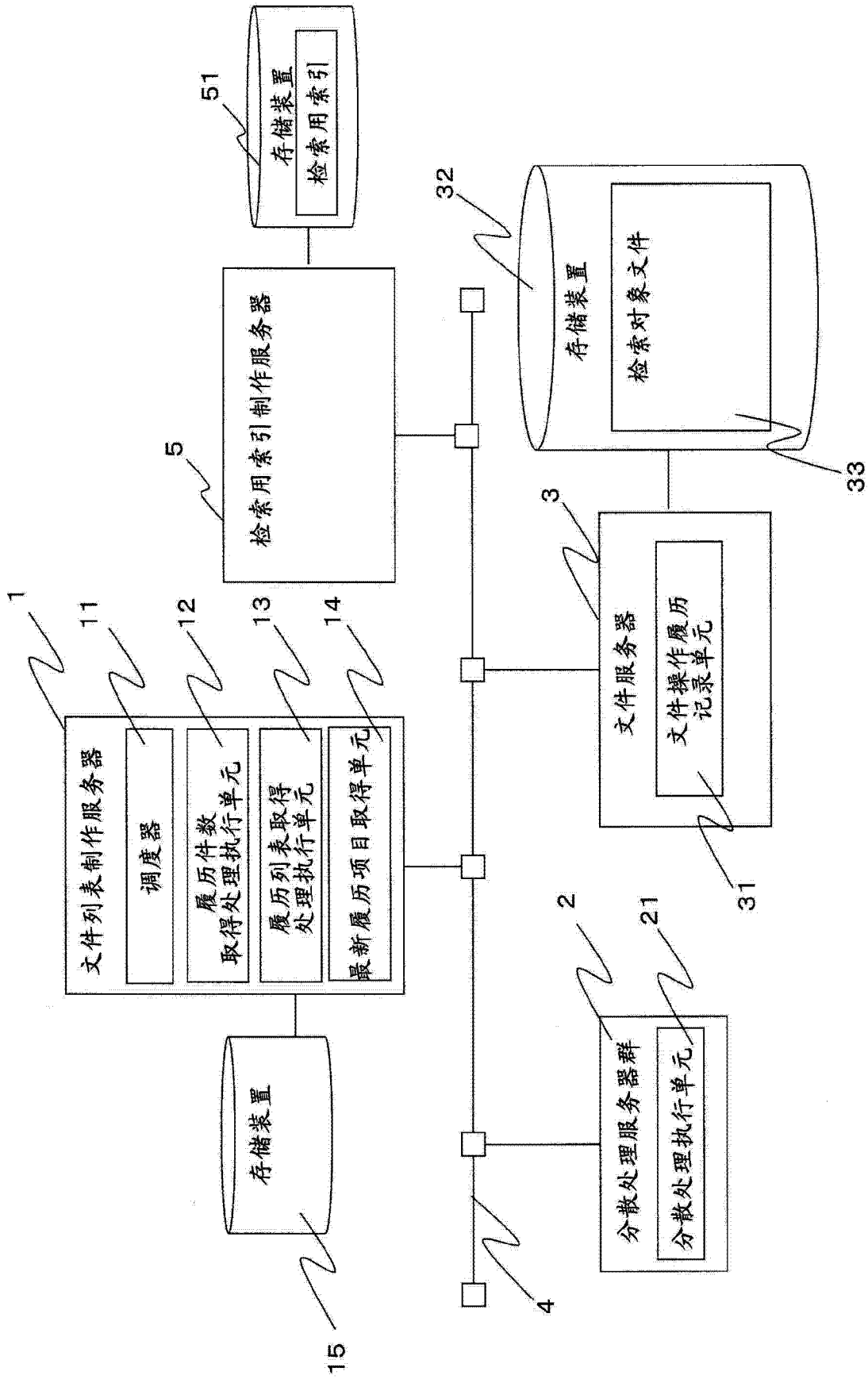


图 1

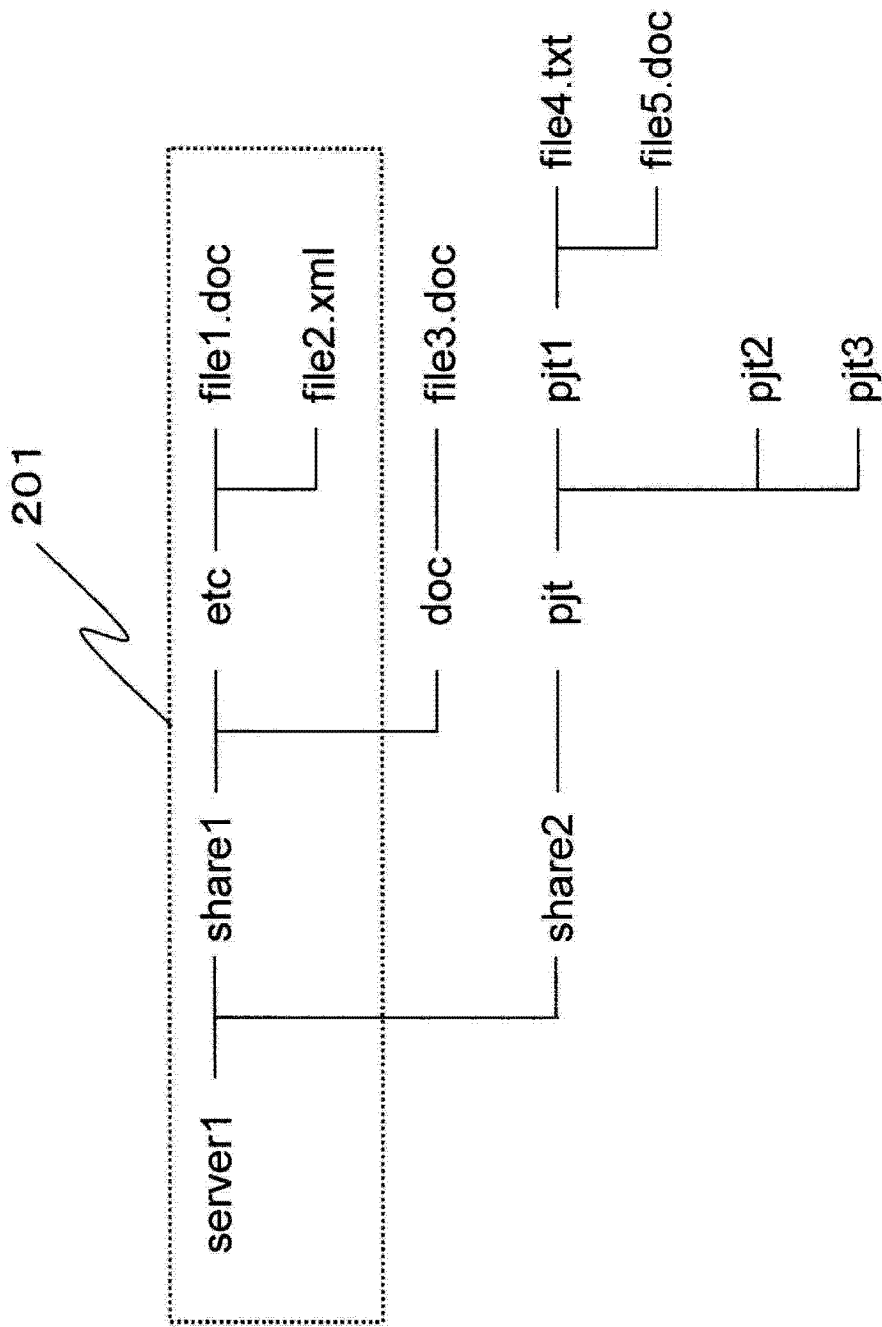


图 2

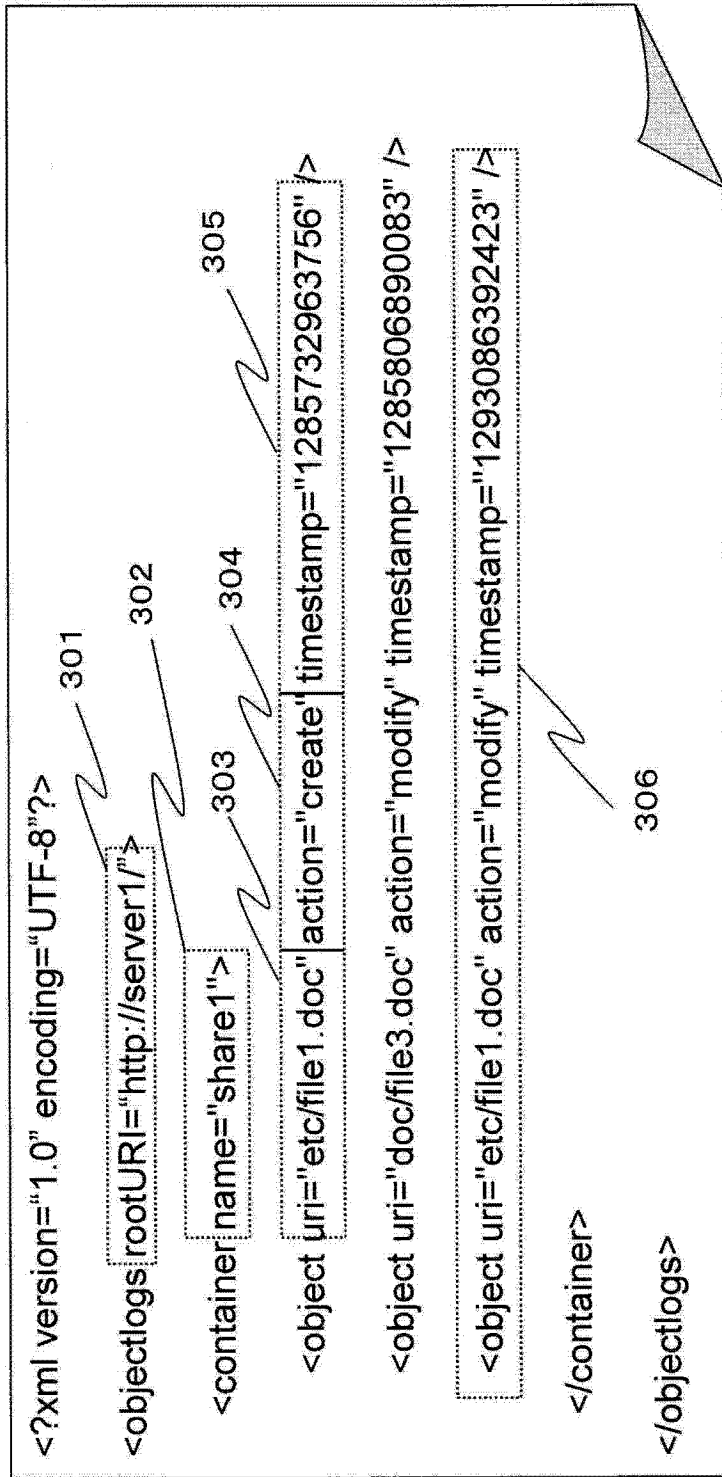


图 3

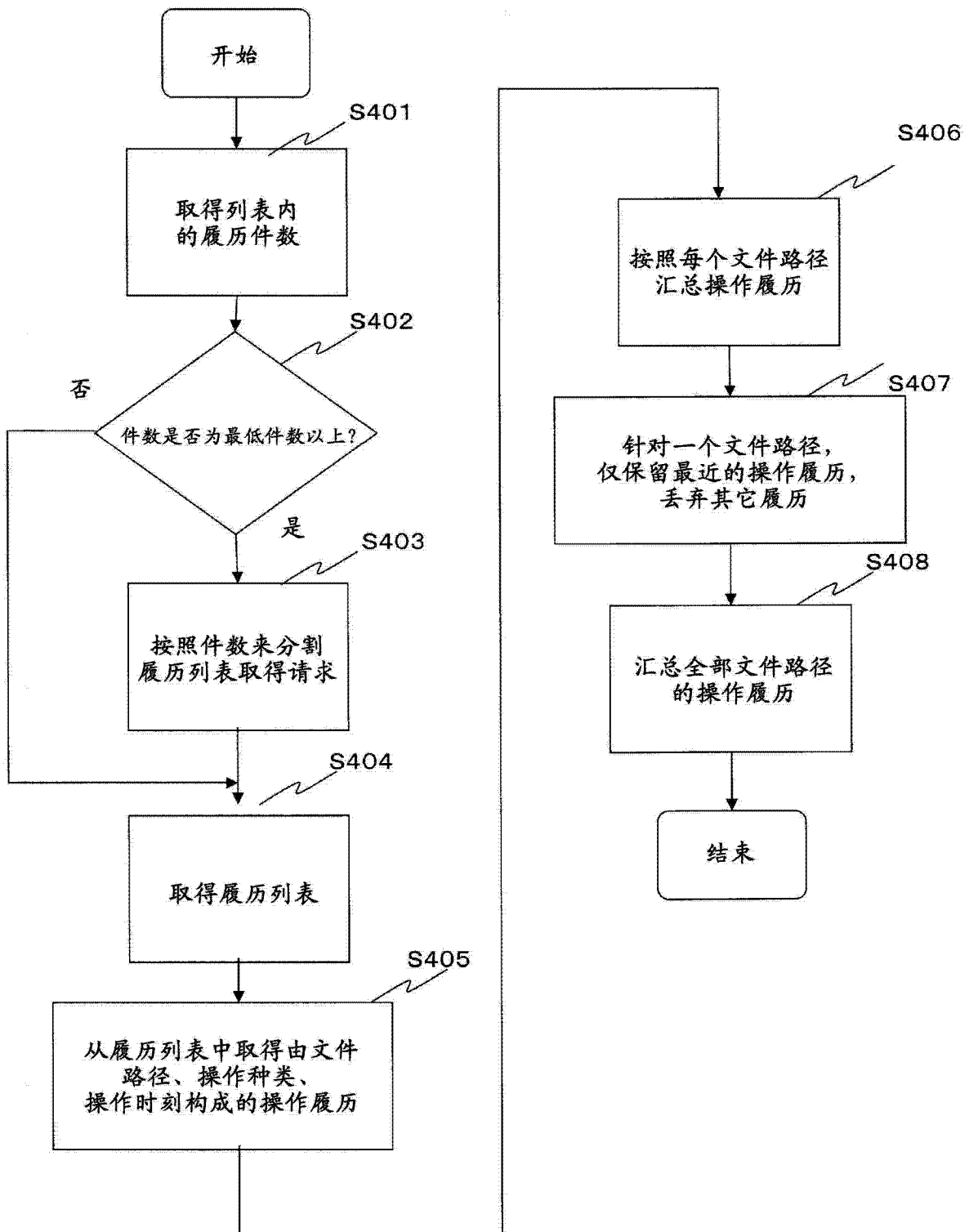


图 4