

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5774513号
(P5774513)

(45) 発行日 平成27年9月9日(2015.9.9)

(24) 登録日 平成27年7月10日(2015.7.10)

(51) Int.Cl. F I
G 0 6 F 12/00 (2006.01) G O 6 F 12/00 5 1 O B
 G O 6 F 12/00 5 2 O P

請求項の数 8 (全 12 頁)

(21) 出願番号 特願2012-24011 (P2012-24011)
 (22) 出願日 平成24年2月7日(2012.2.7)
 (65) 公開番号 特開2013-161342 (P2013-161342A)
 (43) 公開日 平成25年8月19日(2013.8.19)
 審査請求日 平成26年7月25日(2014.7.25)

(73) 特許権者 000233055
 株式会社日立ソリューションズ
 東京都品川区東品川四丁目12番7号
 (74) 代理人 110001678
 特許業務法人藤央特許事務所
 (72) 発明者 西田 晋平
 東京都品川区東品川四丁目12番7号 株
 式会社日立ソリューションズ内
 (72) 発明者 古谷 沙織
 東京都品川区東品川四丁目12番7号 株
 式会社日立ソリューションズ内
 審査官 加内 慎也

最終頁に続く

(54) 【発明の名称】 ファイルリスト生成方法及びシステム並びにプログラム、ファイルリスト生成装置

(57) 【特許請求の範囲】

【請求項1】

検索対象ファイルを管理するファイルサーバから前回の検索インデクス作成処理以降に検索対象ファイルに対して実施されたファイルデータの追加、変更、削除の操作履歴リストを取得する第1のステップと、取得した操作履歴リスト中に同一ファイルについて複数の操作履歴が含まれていた場合には最新の操作履歴のみを取得した後、当該操作履歴を他のファイルの操作履歴と共に集約し、前回の検索インデクス作成処理以降に検索対象ファイルに対して実施された操作履歴リストとの差分リストとして出力する第2のステップとを備え、

前記第2のステップにおいて、前記第1のステップで取得した操作履歴件数が所定件数以上である場合は、取得した操作履歴リストをファイルパス名でソートし、同一のファイルパスに関する操作履歴リストについては同一の分散処理サーバが割り当たるように操作履歴リストを分割し、前記分割した夫々について複数の分散処理サーバで生成した差分リストを集約して出力することを特徴とするファイルリスト生成方法。

【請求項2】

前回の検索インデクス作成処理以降から現在までの期間を複数の期間に分割し、その分割期間毎に各期間の操作履歴リストを前記第1のステップで取得し、その取得した操作履歴リストの処理を複数の分散処理サーバに割当て、前記第2のステップにおいて各分散処理サーバに割当てた操作履歴リスト中に同一ファイルについて複数の操作履歴が含まれていた場合には最新の操作履歴のみを取得した後、当該操作履歴を他の分散処理サーバが分

10

20

散処理した操作履歴と共に集約し、前回の検索インデクス作成処理以降に検索対象ファイルに対して実施された操作履歴リストとの差分リストとして出力することを特徴とする請求項 1 に記載のファイルリスト生成方法。

【請求項 3】

検索対象ファイルを管理するファイルサーバから前回の検索インデクス作成処理以降に検索対象ファイルに対して実施されたファイルデータの追加、変更、削除の操作履歴リストを取得する第 1 の手段と、取得した操作履歴リスト中に同一ファイルについて複数の操作履歴が含まれていた場合には最新の操作履歴のみを取得した後、当該操作履歴を他のファイルの操作履歴と共に集約し、前回の検索インデクス作成処理以降に検索対象ファイルに対して実施された操作履歴リストとの差分リストとして出力する第 2 の手段とを備え、

10

前記第 2 の手段は、前記第 1 の手段が取得した操作履歴件数が所定件数以上である場合は、取得した操作履歴リストをファイルパス名でソートし、同一のファイルパスに関する操作履歴リストについては同一の分散処理サーバが割り当たるように操作履歴リストを分割し、前記分割した夫々について複数の分散処理サーバで生成した差分リストを集約して出力することを特徴とするファイルリスト生成システム。

【請求項 4】

前回の検索インデクス作成処理以降から現在までの期間を複数の期間に分割し、その分割期間毎に各期間の操作履歴リストを前記第 1 の手段で取得し、その取得した操作履歴リストの処理を複数の分散処理サーバに割当て、各分散処理サーバに割当てた操作履歴リスト中に同一ファイルについて複数の操作履歴が含まれていた場合には最新の操作履歴のみを取得した後、当該操作履歴を他の分散処理サーバが分散処理した操作履歴と共に前記第 2 の手段が集約し、前回の検索インデクス作成処理以降に検索対象ファイルに対して実施された操作履歴リストとの差分リストとして出力することを特徴とする請求項 3 に記載のファイルリスト生成システム。

20

【請求項 5】

ファイルリスト作成サーバで使用するプログラムであって、
前記ファイルリスト作成サーバを、

検索対象ファイルを管理するファイルサーバから前回の検索インデクス作成処理以降に検索対象ファイルに対して実施されたファイルデータの追加、変更、削除の操作履歴リストを取得する第 1 の手段と、取得した操作履歴リスト中に同一ファイルについて複数の操作履歴が含まれていた場合には最新の操作履歴のみを取得した後、当該操作履歴を他のファイルの操作履歴と共に集約し、前回の検索インデクス作成処理以降に検索対象ファイルに対して実施された操作履歴リストとの差分リストとして出力する第 2 の手段として機能させ、

30

前記第 2 の手段は、前記第 1 の手段が取得した操作履歴件数が所定件数以上である場合は、取得した操作履歴リストをファイルパス名でソートし、同一のファイルパスに関する操作履歴リストについては同一の分散処理サーバが割り当たるように操作履歴リストを分割し、前記分割した夫々について複数の分散処理サーバで生成した差分リストを集約して出力することを特徴とするファイルリスト生成プログラム。

【請求項 6】

40

前回の検索インデクス作成処理以降から現在までの期間を複数の期間に分割し、その分割期間毎に各期間の操作履歴リストを前記第 1 の手段で取得し、その取得した操作履歴リストの処理を複数の分散処理サーバに割当て、各分散処理サーバに割当てた操作履歴リスト中に同一ファイルについて複数の操作履歴が含まれていた場合には最新の操作履歴のみを取得した後、当該操作履歴を他の分散処理サーバが分散処理した操作履歴と共に集約し、前回の検索インデクス作成処理以降に検索対象ファイルに対して実施された操作履歴リストとの差分リストとして出力する第 2 の手段として機能させることを特徴とする請求項 5 に記載のファイルリスト生成プログラム。

【請求項 7】

検索対象ファイルを管理するファイルサーバから前回の検索インデクス作成処理以降に

50

検索対象ファイルに対して実施されたファイルデータの追加、変更、削除の操作履歴リストを取得する第1の手段と、取得した操作履歴リスト中に同一ファイルについて複数の操作履歴が含まれていた場合には最新の操作履歴のみを取得した後、当該操作履歴を他のファイルの操作履歴と共に集約し、前回の検索インデクス作成処理以降に検索対象ファイルに対して実施された操作履歴リストとの差分リストとして出力する第2の手段とを備え、前記第2の手段は、前記第1の手段が取得した操作履歴件数が所定件数以上である場合は、取得した操作履歴リストをファイルパス名でソートし、同一のファイルパスに関する操作履歴リストについては同一の分散処理サーバが割り当たるように操作履歴リストを分割し、前記分割した夫々について複数の分散処理サーバで生成した差分リストを集約して出力することを特徴とするファイルリスト生成装置。

10

【請求項8】

前回の検索インデクス作成処理以降から現在までの期間を複数の期間に分割し、その分割期間毎に各期間の操作履歴リストを前記第1の手段で取得し、その取得した操作履歴リストの処理を複数の分散処理サーバに割当て、各分散処理サーバに割当てた操作履歴リスト中に同一ファイルについて複数の操作履歴が含まれていた場合には最新の操作履歴のみを取得した後、当該操作履歴を他の分散処理サーバが分散処理した操作履歴と共に前記第2の手段が集約し、前回の検索インデクス作成処理以降に検索対象ファイルに対して実施された操作履歴リストとの差分リストとして出力することを特徴とする請求項7に記載のファイルリスト生成装置。

20

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、検索システムにおいて検索インデクスを更新すべきファイルデータについて、そのリストを生成する方法及びシステム並びにプログラム、装置に係るものであり、特に、検索インデクス作成対象のファイルサーバが、ファイルサーバに対して行われたファイルサーバ内データの操作履歴を取得するインタフェースを備えている場合について、そのインタフェースを用いて取得した操作履歴から、ファイルサーバ内データのうち、追加、更新、削除されたファイルの一覧を高速に作成する方法、システム、プログラム及び装置に関するものである。

【背景技術】

30

【0002】

近年、コンピュータ性能の高速化、HDDの大容量化に伴い、膨大な数の非定型文書が作られるようになってきている。このため、膨大な数の文書の中から、必要とする文書を高速かつ的確に探し出すことの可能な検索システムの必要性が高まっている。検索結果を的確なものとするためには、検索対象となる非定型文書が格納されたファイルサーバに対して検索インデクス作成後に行われたファイルデータの追加、変更、削除操作をタイムリーに検索インデクスに反映することが重要となる。これらを検索インデクスへと反映させる際、変更されていないファイルデータについても検索インデクスを更新すると多大な時間がかかってしまうため、一般的には追加、変更、削除のあったファイルデータのみ絞って検索インデクスを更新する。そのためには、追加、変更、削除されたファイルデータの一覧を作成する必要がある。

40

【0003】

このような検索システムのニーズにこたえるため、ファイルサーバの中にはファイルデータの操作履歴を記憶し、外部からの要求に応じて操作履歴を提供するインタフェースを備えるものも存在している。

先行技術として下記の特許文献1に開示されたものが知られている。

【先行技術文献】

【特許文献】

【0004】

【特許文献1】特開2006-268456号

50

【発明の概要】

【発明が解決しようとする課題】

【0005】

ファイルデータの追加、変更、削除操作をタイムリーに検索インデクスに反映するために、ファイルサーバ内の検索対象ファイルについての検索インデクスの新規作成処理や、追加、変更、削除されたファイルの検索インデクス更新処理を、多数のサーバを用いて分散処理させることで高速に実施する方法が考えられる。

インデクス作成、更新対象ファイルデータの一覧を作成する際、ファイルサーバが更新ファイルの一覧を返すインタフェースを提供している場合は、そのインタフェースを利用すればよいが、そのようなインタフェースを備えていないファイルサーバの場合は、一般的にファイルサーバ内に存在する検索インデクス作成対象範囲のファイルデータ全てを走査して、対象ファイルの列挙や、更新操作の有無を判定する必要がある。

特にインデクス更新の際は、追加、変更、削除されたファイルデータが少量である場合であっても、全てのファイルデータを走査する必要があるため、追加、変更、削除されたファイルデータの一覧の作成処理が、インデクス更新処理の長時間化の要因となっている。

【0006】

一方でファイルサーバがファイル操作の履歴リストを返すインタフェースを備えている場合、そのインタフェースを用いて操作履歴リストを取得し、操作履歴リストに従って検索インデクスに追加、変更、削除を反映する方法が考えられる。しかし、操作履歴リストには同一のファイルに関する複数の操作履歴が含まれている場合がある。この場合、操作履歴リストの時系列に沿って検索インデクスを更新しなければ正しく検索インデクスを更新することができない。このような逐次的な処理が必要な場合、例えば操作履歴リストを分割して多数のサーバを用いて分散処理を行い、検索インデクスを更新する場合、検索インデクス更新前に分散処理結果を操作履歴リストの時系列順に並べて適用する必要があり、仮に操作履歴リストの後半部分の処理が高速に完了したとしても操作履歴リストの前半部分の処理が完了するまで待たなければならない。操作履歴リスト中に同一ファイルに関して複数の操作履歴が含まれていることが逐次処理が必要な原因である。

【0007】

本発明の目的は、操作履歴リストを解析し、同一ファイルについて最新の操作のみが含まれるリスト、すなわちファイルサーバ内で追加、変更、削除されたファイルデータの最新の一覧（以下、差分リストと称する）を作成することであり、かつ、この履歴リストから差分リストへの変換処理を、分散処理サーバ群を利用して分散実行することにより、大容量ストレージが返す大きな操作履歴リストを高速に差分リストに変換することで、検索インデクスの新規作成、更新処理の分散実行を容易に実行できるようにし、検索インデクスを高速に最新の状態にすることである。

【課題を解決するための手段】

【0008】

上記課題を解決するために、本発明に係るファイルリスト生成方法は、検索対象ファイルを管理するファイルサーバから前回の検索インデクス作成処理以降に検索対象ファイルに対して実施されたファイルデータの追加、変更、削除の操作履歴リストを取得する第1のステップと、取得した操作履歴リスト中に同一ファイルについて複数の操作履歴が含まれていた場合には最新の操作履歴のみを取得した後、当該操作履歴を他のファイルの操作履歴と共に集約し、前回の検索インデクス作成処理以降に検索対象ファイルに対して実施された操作履歴リストとの差分リストとして出力する第2のステップとを備え、

前記第2のステップにおいて、前記第1のステップで取得した操作履歴件数が所定件数以上である場合は、取得した操作履歴リストをファイルパス名でソートし、同一のファイルパスに関する操作履歴リストについては同一の分散処理サーバが割り当たるように操作履歴リストを分割し、前記分割した夫々について複数の分散処理サーバで生成した差分リストを集約して出力することを特徴とする。

10

20

30

40

50

また、前回の検索インデクス作成処理以降から現在までの期間を複数の期間に分割し、その分割期間毎に各期間の操作履歴リストを前記第1のステップで取得し、その取得した操作履歴リストの処理を複数の分散処理サーバに割当て、前記第2のステップにおいて各分散処理サーバに割当てた操作履歴リスト中に同一ファイルについて複数の操作履歴が含まれていた場合には最新の操作履歴のみを取得した後、当該操作履歴を他の分散処理サーバが分散処理した操作履歴と共に集約し、前回の検索インデクス作成処理以降に検索対象ファイルに対して実施された操作履歴リストとの差分リストとして出力することを特徴とする。

【0009】

また、本発明に係るファイルリスト生成システムは、検索対象ファイルを管理するファイルサーバから前回の検索インデクス作成処理以降に検索対象ファイルに対して実施されたファイルデータの追加、変更、削除の操作履歴リストを取得する第1の手段と、取得した操作履歴リスト中に同一ファイルについて複数の操作履歴が含まれていた場合には最新の操作履歴のみを取得した後、当該操作履歴を他のファイルの操作履歴と共に集約し、前回の検索インデクス作成処理以降に検索対象ファイルに対して実施された操作履歴リストとの差分リストとして出力する第2の手段とを備え、

前記第2の手段は、前記第1の手段が取得した操作履歴件数が所定件数以上である場合は、取得した操作履歴リストをファイルパス名でソートし、同一のファイルパスに関する操作履歴リストについては同一の分散処理サーバが割り当たるように操作履歴リストを分割し、前記分割した夫々について複数の分散処理サーバで生成した差分リストを集約して

出力することを特徴とする。

また、前回の検索インデクス作成処理以降から現在までの期間を複数の期間に分割し、その分割期間毎に各期間の操作履歴リストを前記第1の手段で取得し、その取得した操作履歴リストの処理を複数の分散処理サーバに割当て、各分散処理サーバに割当てた操作履歴リスト中に同一ファイルについて複数の操作履歴が含まれていた場合には最新の操作履歴のみを取得した後、当該操作履歴を他の分散処理サーバが分散処理した操作履歴と共に前記第2の手段が集約し、前回の検索インデクス作成処理以降に検索対象ファイルに対して実施された操作履歴リストとの差分リストとして出力することを特徴とする。

【0010】

本発明に係るファイルリスト生成プログラムは、ファイルリスト作成サーバで使用するプログラムであって、

前記ファイルリスト作成サーバを、

検索対象ファイルを管理するファイルサーバから前回の検索インデクス作成処理以降に検索対象ファイルに対して実施されたファイルデータの追加、変更、削除の操作履歴リストを取得する第1の手段と、取得した操作履歴リスト中に同一ファイルについて複数の操作履歴が含まれていた場合には最新の操作履歴のみを取得した後、当該操作履歴を他のファイルの操作履歴と共に集約し、前回の検索インデクス作成処理以降に検索対象ファイルに対して実施された操作履歴リストとの差分リストとして出力する第2の手段として機能させ、

前記第2の手段は、前記第1の手段が取得した操作履歴件数が所定件数以上である場合は、取得した操作履歴リストをファイルパス名でソートし、同一のファイルパスに関する操作履歴リストについては同一の分散処理サーバが割り当たるように操作履歴リストを分割し、前記分割した夫々について複数の分散処理サーバで生成した差分リストを集約して

出力することを特徴とする。

また、前回の検索インデクス作成処理以降から現在までの期間を複数の期間に分割し、その分割期間毎に各期間の操作履歴リストを前記第1の手段で取得し、その取得した操作履歴リストの処理を複数の分散処理サーバに割当て、各分散処理サーバに割当てた操作履歴リスト中に同一ファイルについて複数の操作履歴が含まれていた場合には最新の操作履歴のみを取得した後、当該操作履歴を他の分散処理サーバが分散処理した操作履歴と共に集約し、前回の検索インデクス作成処理以降に検索対象ファイルに対して実施された操作

10

20

30

40

50

履歴リストとの差分リストとして出力する第2の手段として機能させることを特徴とする。

【0011】

本発明に係るファイルリスト生成装置は、検索対象ファイルを管理するファイルサーバから前回の検索インデクス作成処理以降に検索対象ファイルに対して実施されたファイルデータの追加、変更、削除の操作履歴リストを取得する第1の手段と、取得した操作履歴リスト中に同一ファイルについて複数の操作履歴が含まれていた場合には最新の操作履歴のみを取得した後、当該操作履歴を他のファイルの操作履歴と共に集約し、前回の検索インデクス作成処理以降に検索対象ファイルに対して実施された操作履歴リストとの差分リストとして出力する第2の手段とを備え、

10

前記第2の手段は、前記第1の手段が取得した操作履歴件数が所定件数以上である場合は、取得した操作履歴リストをファイルパス名でソートし、同一のファイルパスに関する操作履歴リストについては同一の分散処理サーバが割り当たるように操作履歴リストを分割し、前記分割した夫々について複数の分散処理サーバで生成した差分リストを集約して出力することを特徴とする。

また、前回の検索インデクス作成処理以降から現在までの期間を複数の期間に分割し、その分割期間毎に各期間の操作履歴リストを前記第1の手段で取得し、その取得した操作履歴リストの処理を複数の分散処理サーバに割当て、各分散処理サーバに割当てた操作履歴リスト中に同一ファイルについて複数の操作履歴が含まれていた場合には最新の操作履歴のみを取得した後、当該操作履歴を他の分散処理サーバが分散処理した操作履歴と共に前記第2の手段が集約し、前回の検索インデクス作成処理以降に検索対象ファイルに対して実施された操作履歴リストとの差分リストとして出力することを特徴とする。

20

【発明の効果】

【0012】

本発明によれば、ファイルサーバ内のデータに対する操作の履歴を保管し、要求に応じて履歴リストを返すインタフェースを備えたファイルサーバにおいて、前回の検索インデクス作成処理以降に検索対象ファイルに対して実施されたファイルデータの追加、変更、削除の操作履歴リストを取得し、その取得した操作履歴リスト中に同一ファイルについて複数の操作履歴が含まれていた場合には最新の操作履歴のみを取得した後、当該操作履歴を他のファイルの操作履歴と共に集約し、前回の検索インデクス作成処理以降に検索対象ファイルに対して実施された操作履歴リストとの差分リストとして出力するようにしているため、ファイルサーバ上で追加、変更、削除されたファイルの一覧を高速に作成することが可能となる。

30

これによって、検索インデクスの新規作成、更新処理の分散実行を容易に実行できるようになり、検索インデクスの新規作成、更新処理を高速に行うことが可能となり、結果として、検索システムによる検索結果を可能な限りの確なものとすることができる。

【図面の簡単な説明】

【0013】

【図1】本発明に係る第一の実施形態におけるシステム構成図である。

【図2】ファイルサーバ内のインデクス対象ファイルのデータ構成図である。

40

【図3】ファイルサーバから返される履歴リストのデータ構成図である。

【図4】分散処理サーバ群を用いて履歴リストを分散して取得し、各ファイルの最新の履歴項目を取得して差分リストに変換する処理のフローチャートである。

【発明を実施するための形態】

【0014】

以下、本発明を実施する場合の第一の実施形態について、図面に基づき詳細に説明する。

図1は、本発明に係る第一の実施形態におけるシステム構成図である。

図1において示したファイルリスト生成システムは、ファイルリスト作成サーバ1、分散処理サーバ群2、およびファイルサーバ3（以下「サーバ等」という）を、LAN（Lo

50

cal Area Network) 4等の有線又は無線の通信回線により互いに通信可能に接続したシステムである。

【0015】

図1においては、サーバ等はLAN4によって互いに通信可能に接続されているが、接続方法はLANに限定されるものではなく、例えばWAN(Wide Area Network)やインターネットによって接続されていてもよい。また、図1においては、サーバ等が同一LANセグメント上で接続されているが、この構成は一例に過ぎず、どのような構成になっていてもよい。さらに、図1においては、ファイルリスト作成サーバ1、分散処理サーバ群2、およびファイルサーバ3はそれぞれ1台ずつ存在しているが、それぞれ2台以上存在しても構わない。また、ファイルリスト作成サーバ1、分散処理サーバ群2、およびファイルサーバ3はそれぞれ異なる装置である必要はなく、例えば、ファイルリスト作成サーバ1、分散処理サーバ群2、およびファイルサーバ3の機能を1台の装置によって実現することも可能である。

10

【0016】

ファイルサーバ3にはファイル操作履歴記録手段31が備わっており、ファイルサーバ3に接続された記憶装置32に格納された検索対象ファイル33に対して行われた追加、変更、削除などの操作の履歴を記録し、HTTP(Hypertext Transfer Protocol)などを用いたクライアントからの要求に応じて履歴リストを返すようになっている。この履歴リストの詳細は図3で説明する。

【0017】

以上のような構成により、ファイルリスト作成サーバ1は分散処理サーバ群2を利用してファイルサーバ3から履歴リストの取得と前回インデクスを更新したとき以降からの差分リスト変換処理を行う。

20

【0018】

ファイルリスト作成サーバ1はPC等の装置であり、記憶装置15と通信可能に接続されている。記憶装置15は磁気ディスク等の装置であり、ファイルリスト作成サーバ1に内蔵され又は外部接続される。記憶装置15と、ファイルリスト作成サーバ1の主記憶装置等は、ファイルリスト作成サーバ1の記憶手段として機能する。

【0019】

ファイルリスト作成サーバ1はスケジューラ11、履歴リスト件数取得処理実行手段12、履歴リスト取得処理実行手段13、最新履歴項目取得手段14を備えている。また、ファイルリスト作成サーバ1はCPU、主記憶装置等を備えており、CPUは記憶装置15に記憶されている履歴リスト件数取得処理実行手段12等のプログラムを主記憶装置上にローディングし、その命令コードを実行することで各種の処理を実行する。

30

【0020】

スケジューラ11は、差分リストを定期的作成して検索インデクスを更新するにあたり、記憶装置15に記憶されている一覧作成処理実行間隔を参照して履歴リスト件数取得処理実行手段12を実行した後、履歴リスト取得処理実行手段13を実行して、ファイルサーバ3から履歴リストを取得し、その後、最新履歴項目取得手段14を実行して履歴リストに含まれる同一ファイルに関する操作履歴のうち最新のもののみを取得して差分リストへと変換する。この一連の動作については履歴リストから差分リストへの変換処理(S401等)として後述する。

40

【0021】

図2は、ファイルサーバ3の検索対象ファイル33の構成の具体例を示した概念図である。

この概念図に示すファイルサーバ3内の検索対象ファイル33の構成においては、あるファイルサーバ3の識別名が“server1”である。“server1”では2つのディレクトリが共有されており、それぞれ“share1”、“share2”という共有名により一意に識別可能である。“server1”の“share1”、“share2”には、図示したようなディレクトリおよびファイルが存在する。例えば、“server1”の“share1”には、“etc”、“doc”の2つ

50

のディレクトリが存在し、“etc”ディレクトリには“file1.doc”、“file2.xml”の2つのファイルが、“doc”ディレクトリには“file3.doc”の1つのファイルが存在する。同様に、“server1”の“share2”には、1つのディレクトリ“pjt”が存在し、“pjt”ディレクトリには“pjt1”、“pjt2”、“pjt3”の3つのディレクトリが存在し、そのうち“pjt1”ディレクトリには“file4.txt”、“file5.doc”の2つのファイルが存在する。

【0022】

図3はファイルサーバ3のファイル操作履歴記録手段31が返す履歴リストについて、図2の例を用いて具体的に示した概念図である。

履歴リスト300には、まず、どのサーバについての履歴リストかを示すデータ301が含まれている。この例ではXMLの“objectlogs”という要素に“rootURI”という名の属性を設けて“http://server1/”というデータを入れて、図2で示した“server1”を表すようにしている。

10

次にどの共有ディレクトリに関する履歴リストかを示すデータ302が含まれている。この例ではXMLの“container”という要素に“name”という名の属性を設けて“share1”というデータを入れて、図2で示した“share1”という共有フォルダを表すようにしている。

【0023】

各操作履歴には、どのファイルについての操作であるか、ファイルを識別するためのデータ303が含まれている。この例ではXMLの“object”という要素に“uri”という名の属性を設けて“etc/file1.doc”というデータを入れて、図2で示した“share1”という共有フォルダ配下にある“etc”というフォルダの中の“file.doc”というファイルを表すようにしている。

20

同時に、操作履歴には、どのような操作が行われたかを示すためのデータ304が含まれており、この例では“action”という属性を設けて“create”というデータを入れて、そのファイルが作成されたことを表すようにしている。

“action”属性に含まれる可能性のある値としては検索対象ファイルに新たなファイルやディレクトリが追加されたことを示す“create”のほか、既存のファイル、ディレクトリの内容、またはメタデータが変更されたことを示す“modify”、既存のファイル、ディレクトリが削除されたことを示す“delete”が考えられる。

30

更に、操作履歴には、その操作がいつ行われたかを示すデータ305が含まれている。この例では“timestamp”という属性を設けて操作が行われた日時を1970年1月1日0時0分からのミリ秒で表すようにしている。

操作履歴の中には同一ファイルに対して行われた操作が時系列順に記録されている。データ306はデータ303が示すファイルと同一のファイルに関して別の操作が、データ305が示す時刻よりも後に行われていることを示す。

【0024】

このように、あるサーバ301の共有フォルダ302の配下に対して行われたファイルまたはディレクトリの追加、更新、削除操作の順序を、XMLの“object”要素の出現順で表すことによって、図3のXMLはあるサーバ301の共有フォルダ302配下に対して行われた操作の履歴を表す履歴リストとなっている。

40

【0025】

図3において履歴リスト300はXML(Extensible Markup Language)形式で記述しているが、必ずしもXML形式である必要はなく、例えばJSON(JavaScript Object Notation:Java = 登録商標)形式やCSV(Comma Separated Values)形式など他の形式でもよい。

図4は、分散処理サーバ2群を用いて履歴リスト300を分散して取得し、各ファイルの最新の履歴項目を取得して差分リストに変換する処理のフローチャートである。

図4のS401に至るまでの処理は、前述のように、スケジューラ11が、検索インデックスを定期的に更新するにあたり、記憶装置15に記憶されている一覧作成処理実行間隔

50

を参照して、差分リストを作成するために処理を開始させる。

【0026】

なお、検索インデクスは検索用インデクス作成サーバ5が作成し、記憶装置51に格納している。検索インデクスの更新はスケジューラ11が指示する。

【0027】

ファイルリスト作成サーバ1の履歴リスト件数取得処理手段12は、ファイルサーバ3に対してHTTP等の通信手段を用いて前回インデクス更新時から、現時刻の間の履歴リストに含まれる履歴の件数を問い合わせ取得する(S401)。

【0028】

次に、取得した件数が最低件数以上かどうかを判断する(S402)。これは履歴リストに含まれる同一ファイルの履歴から最新の操作履歴項目を取得し差分リストへと変換する処理を並列に行うことで高速化するという本発明の目的に対して、履歴リストの件数があまりに少なかった場合に、並列処理を行うことによるオーバーヘッドが並列処理による高速化のメリットを上回ってしまうことを避けるために行うものである。なお、最低件数はファイルリスト作成サーバ1上の設定ファイルなどに記憶している値を参照する。この値は、例えば50000といった値であり、一台のファイルリスト作成サーバ1または分散処理サーバ群2のうちの1台が少なくとも数分以内に履歴リストに含まれる同一ファイルの履歴から最新の操作履歴項目を取得し差分リストへと変換する処理を完了できる件数を推定して設定することが望ましい。

【0029】

S402の判断で取得した履歴リストの件数が最低件数以上だった場合は、履歴リスト取得要求を、分散処理サーバ群2で並列に要求するために、分割する(S402)。

例えば履歴リスト件数が1000000件だった場合は、1台あたりの履歴リスト件数が最低件数と等しくなるように、1台目のサーバは1件目から50000件目、2台目のサーバは50001件目から100000件目というように分割していく。なお、この実施形態では履歴リスト件数取得処理実行手段12を用いて件数を元に履歴リスト取得要求を分割しているが、件数のほかに期間で分割する方式も考えられる。すなわち、例えば前回のインデクス更新が3週間前だった場合は1台目のサーバが3週間前から2週間前までの1週間分の履歴リストを要求し、2台目のサーバが2週間前から1週間前までの1週間分の履歴リストを要求し、3台目のサーバが1週間前から現時刻までの1週間分の履歴リストを要求するような分割手法が考えられる。本発明は、このように前回インデクス作成処理日以降から現在までの期間を複数の期間に分割し、各期間毎の操作履歴リストを取得し、各期間の操作履歴リストを複数の分散処理サーバに割当て、各分散処理サーバが抽出した最新の操作履歴リストを集約して出力するように構成することが可能である。

【0030】

次に、ファイルサーバ3に対してHTTP等の通信手段を用いて履歴リストを要求し、取得する(S404)。このとき、S402の判断で履歴リストの件数が最低件数以上だった場合はS403の分割処理に従って分散処理サーバ群2で並列に要求を行う。

S402の判断で履歴リストの件数が最低件数に満たなかった場合は1台のサーバで履歴リストの要求を行い、取得する。

【0031】

次に、取得した履歴リストからファイルパス、操作種別、操作時刻からなる操作履歴を取得する(S405)。ファイルパスは図3のサーバを表すデータ301、共有フォルダを表すデータ302、ファイルを表すデータ303を繋げることによって作成できる。

次に、同一のファイルパスについての操作履歴を集計する(S406)。

次に、それぞれのファイルパスについて、操作時刻を参照し、最も現時刻に近い操作の操作履歴のみを残して、他の履歴を捨てる(S407)。

S405、S406、S407の処理はS402の判断結果に関わらず分散処理サーバ群2によって、並列に処理する。

【0032】

10

20

30

40

50

次に全てのファイルパスの操作履歴を集約する（S408）。集約された操作履歴からなるリストは、分散処理サーバ群2によって並列処理で同一ファイルの履歴から最新の操作履歴項目を取得し終えた差分リストであり、この時点で履歴リストから差分リストへの変換処理が完了する。

【0033】

なお、分割する件数や期間をまたいで操作履歴の重複が発生した場合に、集約し切れずにファイル重複が発生してしまうことが懸念される。しかし、このような場合には次のようにして解決することが可能である。すなわち、操作履歴件数が所定の件数以上であった場合に複数の分散処理サーバでファイルパスごとに差分リストの生成を件数で区切って分担して行うが、同一ファイルパスに関する操作履歴は必ず一つのサーバで差分リスト生成を行う。

10

これは、まず取得した履歴をファイルパスの辞書順でソートし、ソート済みの履歴を各分散処理サーバで交換し、同一のファイルパスについては同一の分散処理サーバが処理を担当するように分担量を決定することにより実現可能である。

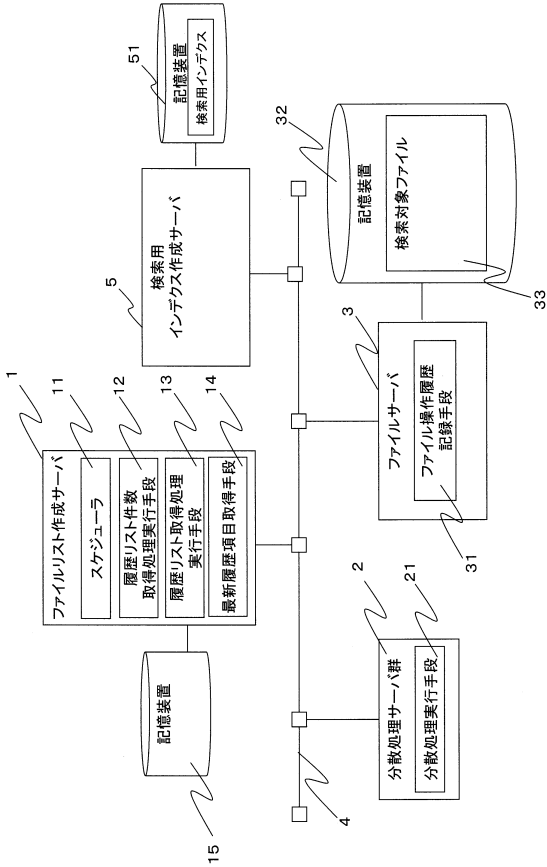
期間によって分割された場合も、同様の分担方法によって、期間の分割値をまたいで操作履歴が存在した場合も重複排除漏れを起こすことなく、最新の履歴のみを残した差分リストを生成することが可能である。

【符号の説明】

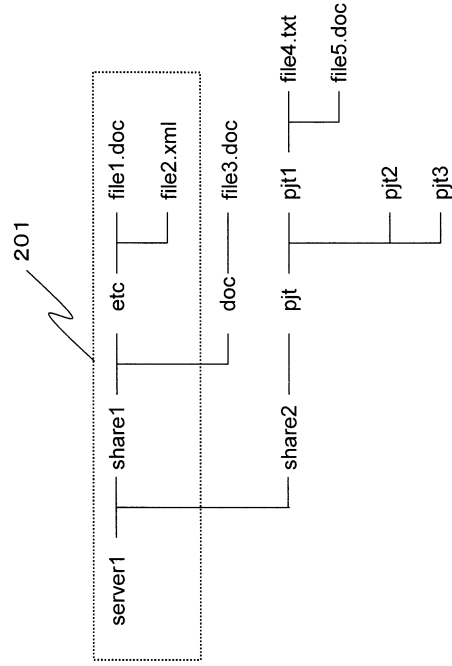
【0034】

- | | | |
|-----|--------------------------|----|
| 1 | ファイルリスト作成サーバ | 20 |
| 2 | 分散処理サーバ群 | |
| 3 | ファイルサーバ | |
| 4 | ネットワーク | |
| 11 | スケジューラ | |
| 12 | 履歴リスト件数取得処理実行手段 | |
| 13 | 履歴リスト取得処理実行手段 | |
| 14 | 最新履歴項目取得手段 | |
| 15 | ファイルリスト作成サーバ1に接続された記憶装置 | |
| 21 | 分散処理実行手段 | |
| 31 | ファイル操作履歴記録手段 | 30 |
| 32 | ファイルサーバ3に接続された記憶装置 | |
| 33 | 検索対象ファイル | |
| 201 | 検索対象ファイルの具体例 | |
| 300 | 操作履歴リスト | |
| 301 | 操作履歴リストにおいてサーバを表すデータ | |
| 302 | 操作履歴リストにおいて共有フォルダを表すデータ | |
| 303 | 操作履歴リストにおいてファイルを表すデータ | |
| 304 | 操作履歴リストにおいて操作種別を表すデータ | |
| 305 | ...操作履歴リストにおいて操作時刻を表すデータ | |

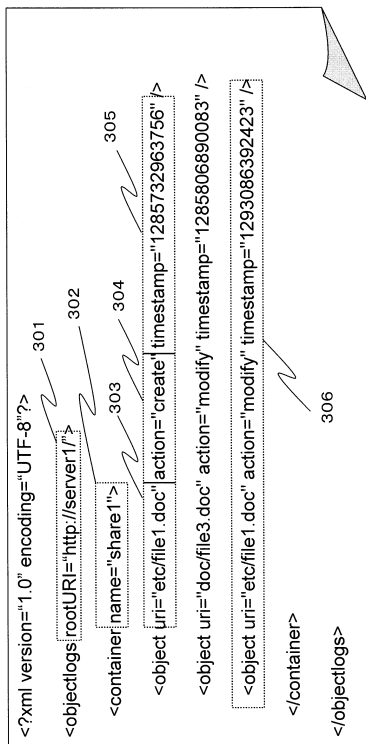
【図1】



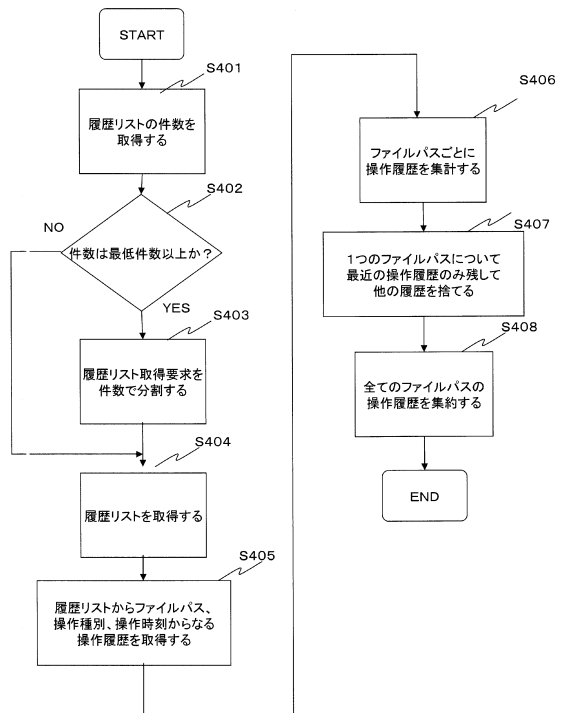
【図2】



【図3】



【図4】



フロントページの続き

- (56)参考文献 特開平08-328924(JP,A)
特開2007-193660(JP,A)
特開昭61-075925(JP,A)
特開2006-018632(JP,A)

- (58)調査した分野(Int.Cl., DB名)
G06F 12/00