

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2012年10月4日(04.10.2012)



(10) 国際公開番号
WO 2012/132046 A1

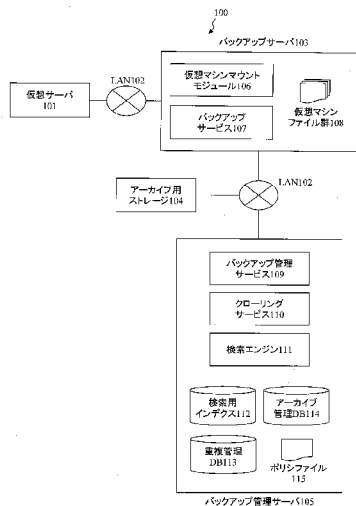
- (51) 国際特許分類:
G06F 12/00 (2006.01) G06F 9/46 (2006.01)
- (21) 国際出願番号: PCT/JP2011/069130
- (22) 国際出願日: 2011年8月25日(25.08.2011)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:
特願 2011-078523 2011年3月31日(31.03.2011) JP
- (71) 出願人(米国を除く全ての指定国について): 株式会社日立ソリューションズ(Hitachi Solutions, Ltd.) [JP/JP]; 〒1400002 東京都品川区東品川四丁目12番7号 Tokyo (JP).
- (72) 発明者; および
- (75) 発明者/出願人(米国についてのみ): 桐畑 康裕 (KIRIHATA Yasuhiro) [JP/JP]; 〒1400002 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 Tokyo (JP).
- (74) 代理人: 平木 祐輔, 外(HIRAKI Yusuke et al.); 〒1056232 東京都港区愛宕2丁目5番1号 愛宕グリーンヒルズMORIタワー32階 Tokyo (JP).
- (81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) 指定国(表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR,

[続葉有]

(54) Title: INFORMATION PROCESSING SYSTEM, BACKUP MANAGEMENT METHOD AND PROGRAM

(54) 発明の名称: 情報処理システム、バックアップ管理方法、及びプログラム

[図1]



(57) Abstract: Provided is technology for enabling searches for specific data included in virtual machine images, without starting up the backed-up virtual machine images. The present invention makes it possible to mount at least one virtual machine image from a repository of backed-up virtual machine images and search files included in that virtual machine image, without starting up the virtual machine image. Also, the present invention realizes functionality that performs appropriate backup management and archiving by detecting files that include specific information such as individual information and email data for archiving, and automatically performing processing such as archiving, transferring and deletion of data from the virtual machine images in accordance with a policy.

(57) 要約: バックアップされた仮想マシンイメージを起動せずに、当該仮想マシンイメージに含まれる特定のデータを検索可能にするための技術を提供する。本発明では、バックアップされた仮想マシンイメージのリポジトリから、少なくとも1つの仮想マシンイメージをマウントし、当該仮想マシンイメージに含まれるファイルを、仮想マシンイメージを起動することなく、検索することを可能にしている。また、本発明では、アーカイブ用のメールデータや個人情報等の特定情報を含むファイルを検出し、ポリシーに応じて仮想マシンイメージからデータをアーカイブ・移動・削除等の処理を自動的に行うことにより、適切にバックアップ管理・アーカイブを行う機能を実現する。

- 101 VIRTUAL SERVER
- 103 BACKUP SERVER
- 104 ARCHIVE STORAGE
- 105 BACKUP MANAGEMENT SERVER
- 106 VIRTUAL MACHINE MOUNT SERVICE
- 107 BACKUP SERVICE
- 108 VIRTUAL MACHINE FILE SET
- 109 BACKUP MANAGEMENT SERVICE
- 110 CRAWLING SERVICE
- 111 SEARCH ENGINE
- 112 INDEX FOR SEARCHING
- 113 DUPLICATE MANAGEMENT DATABASE
- 114 ARCHIVE MANAGEMENT DATABASE
- 115 POLICY FILE

WO 2012/132046 A1

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, ML, MR,
NE, SN, TD, TG).

添付公開書類:

- 国際調査報告 (条約第 21 条(3))
- 補正された請求の範囲 (条約第 19 条(1))

明 細 書

発明の名称：

情報処理システム、バックアップ管理方法、及びプログラム

技術分野

[0001] 本発明は、情報処理システム、バックアップ管理方法、及びプログラムに関し、例えば、仮想マシンのバックアップデータを管理するための技術に関するものである。

背景技術

[0002] クラウドコンピューティングの到来により、Amazon Web ServicesやRackspaseのようなインターネット経由で仮想環境を提供するIaaS事業を行う企業が出現している。それに伴い、オンプレミスにクラウド環境を構築して、社内サービスとして提供するプライベートクラウドのニーズも増えている。

[0003] プライベートクラウドを構築する場合、データセンタにおいてバックエンドで稼働するサーバ環境として、VMwareやXen、Hyper-Vといった仮想化環境を採用するのが一般的である。しかし、この仮想化環境におけるデータのバックアップとバックアップされたデータの管理が重要な課題になっている。

[0004] この課題に対して、通常、各仮想マシンにバックアップエージェントを導入して、このエージェントが指定されたデータのバックアップを取得し、ネットワーク経由でバックアップメディアサーバに送り、メディアサーバからテープライブラリやオフラインストレージ等のバックアップ用デバイスにデータをコピーする方式が主流である。また、ストレージのスナップショット機能を利用して、仮想サーバを格納しているボリュームのスナップショットを取得し、その後、仮想サーバの仮想マシンイメージをそのままバックアップする方式もある。前者に対する後者の方式のメリットは、各仮想サーバにバックアップエージェントをインストールすることなく、バックアップを取得できることである。今後、クラウド環境においては、後者の仮想マシンイメージそのものをバックアップする方式も増えていくと予想される。

[0005] 仮想マシンイメージのバックアップに関し、例えば、特許文献1では、高可用性システムにおいて、サーバシステムが稼働系サーバと待機系サーバで構成されている場合、待機系サーバと同期処理を行う仮想サーバを設置し、稼働系サーバと待機系サーバの同期処理と連動して、待機系サーバと仮想サーバが同期処理を行う。そして、待機系サーバと稼働系サーバの同期処理と待機系サーバと仮想サーバの同期処理を停止した状態で、仮想サーバの仮想マシンイメージを丸ごとバックアップする方式を実現している。

先行技術文献

特許文献

[0006] 特許文献1：特開2010-231257号公報

発明の概要

発明が解決しようとする課題

[0007] 上述のように、特許文献1は、仮想マシンイメージを丸ごとバックアップする仕組みを実現しているが、そのバックアップデータを管理する機能として、イメージデータ内部に含まれているファイル群の検索を実現するものではない。例えば、バックアップデータから特定のファイルや仮想サーバ自体をリカバリしたい場合、仮想マシン内部のファイル内容やシステム構成を検索する必要があるが、特許文献1を含め、現状ではこうした機能は実現できていない。現状では、バックアップデータを起動しなければ検索することができない。

[0008] また、バックアップされた仮想マシンイメージ内には、例えばメールアドレスなど、別途アーカイブ対象となるデータや、個人情報や他社営業機密情報など、バックアップとして複数コピーをとってはいけないデータも含まれている可能性がある。このようなデータについては、データの検出とポリシーに応じたアーカイブ用ストレージへの移動や、バックアップデータからの削除といった機能も必要である。しかし、現状ではこのような機能は実現できていない。

[0009] 本発明はこのような状況に鑑みてなされたものであり、バックアップされた仮想マシンイメージを起動せずに、当該仮想マシンイメージに含まれる特定のデータを検索可能にするための技術を提供するものである。

課題を解決するための手段

[0010] 上記課題を解決するために、本発明では、バックアップされた仮想マシンイメージのリポジトリから、少なくとも1つの仮想マシンイメージをマウントし、当該仮想マシンイメージに含まれるファイルを、仮想マシンイメージを起動することなく、検索することを可能にしている。

[0011] また、本発明では、アーカイブ用のメタデータや個人情報等の特定情報を含むファイルを検出し、ポリシーに応じて仮想マシンイメージからデータをアーカイブ・移動・削除等の処理を自動的に行うことにより、適切にバックアップ管理・アーカイブを行う機能を実現する。

[0012] さらに、本発明では、バックアップデータの重複性や仮想マシンイメージ内に含まれる、OSのシステムファイル等のデータ冗長性を考慮して、検索用インデックスの作成を効率的に行うクローリング方式も実現する。

[0013] 即ち、本発明は、仮想マシンイメージのバックアップを管理する情報処理システムを提供する。当該情報処理システムは、バックアップされた仮想マシンイメージを格納する仮想マシンリポジトリと、仮想サーバの仮想マシンイメージをバックアップし、仮想マシンリポジトリにコピーすると共に、マウント要求に応答して、仮想マシンリポジトリに格納された少なくとも1つの仮想マシンイメージをマウント処理する仮想マシンマウント処理部と、マウントされた仮想マシンイメージにアクセスし、ファイル検索用インデックス作成のためのクローリング処理を実行するクローリング処理部と、を有する。

[0014] 本発明に関連する更なる特徴は、本明細書の記述、添付図面から明らかになるものである。また、本発明の態様は、要素及び多様な要素の組み合わせ及び以降の詳細な記述と添付される特許請求の範囲の様態により達成され実現される。

[0015] 本明細書の記述は典型的な例示に過ぎず、本発明の特許請求の範囲又は適用例を如何なる意味に於いても限定するものではないことを理解する必要がある。

発明の効果

[0016] 本発明によれば、バックアップされた仮想マシンイメージを起動せずに、当該仮想マシンイメージに含まれる特定のデータを検索することができるようになる。

図面の簡単な説明

- [0017] [図1]本発明の実施形態による情報処理システムの概略構成を示す図である。
[図2]アーカイブ管理DBの構成例を示す図である。
[図3]重複管理DBの構成例を示す図である。
[図4]検索用インデックス内のドキュメント管理テーブルの構成例を示す図である。
[図5]検索用インデックス内の転置インデックステーブルの構成例を示す図である。
[図6]クローリング時の仮想マシンマウントモジュールの処理内容を説明するためのフローチャートである。
[図7]クローリング時のクローリングサービスの処理内容を説明するためのフローチャートである。
[図8]データ取得・検索エンジン登録処理の内容を説明するためのフローチャートである。
[図9]バックアップ管理サービスによるポリシーに応じたファイル処理の内容を説明するためのフローチャートである。

発明を実施するための形態

[0018] 本発明は、バックアップされた仮想マシンイメージを起動せずに、バックアップされた仮想マシンイメージ群から、抽出したいファイルを含む仮想イメージファイルをキーワード検索し、当該ファイルを保持する仮想イメージファイルを検出する機能を実現するための技術に関する。

[0019] 以下、添付図面を参照して本発明の実施形態について説明する。添付図面では、機能的に同じ要素は同じ番号で表示される場合もある。なお、添付図面は本発明の原理に則った具体的な実施形態と実装例を示しているが、これらは本発明の理解のためのものであり、決して本発明を限定的に解釈するために用いられるものではない。

[0020] 本実施形態では、当業者が本発明を実施するのに十分詳細にその説明がなされているが、他の実装・形態も可能で、本発明の技術的思想の範囲と精神を逸脱することなく構成・構造の変更や多様な要素の置き換えが可能であることを理解する必要がある。従って、以降の記述をこれに限定して解釈してはならない。

[0021] 更に、本発明の実施形態は、後述されるように、汎用コンピュータ上で稼動するソフトウェアで実装しても良いし専用ハードウェア又はソフトウェアとハードウェアの組み合わせで実装しても良い。

[0022] なお、以後の説明では「テーブル」形式によって本発明の各情報について説明するが、これら情報は必ずしもテーブルによるデータ構造で表現されていなくても良く、リスト、DB、キュー等のデータ構造やそれ以外で表現されていても良い。そのため、データ構造に依存しないことを示すために「テーブル」、「リスト」、「DB」、「キュー」等について単に「情報」と呼ぶことがある。

[0023] また、各情報の内容を説明する際に、「識別情報」、「識別子」、「名」、「名前」、「ID」という表現を用いることが可能であり、これらについてはお互いに置換が可能である。

[0024] 以下では、プログラムとしての「サービス」や「エンジン」等を主語（動作主体）として本発明の実施形態における各処理について説明を行うが、プログラムはプロセッサによって実行されることで定められた処理をメモリ及び通信ポート（通信制御装置）を用いながら行うため、プロセッサを主語とした説明としてもよい。また、プログラムを主語として開示された処理は管理サーバ等の計算機、情報処理装置が行う処理としてもよい。プログラムの

一部または全ては専用ハードウェアで実現してもよく、また、モジュール化されていても良い。各種プログラムはプログラム配布サーバや記憶メディアによって各計算機にインストールされてもよい。また、「モジュール」として示されている処理部は、プログラムとして実現することも可能である。つまり、プログラムとモジュールは相互に置換可能である。

[0025] <システム構成>

図1は、本発明の実施形態によるバックアップ管理システム（計算機システムや情報処理システムともいう）の概略構成を示す図である。バックアップ管理システム100は、仮想サーバ101と、バックアップサーバ103と、アーカイブ用ストレージ104と、バックアップ管理サーバ105と、を有している。仮想サーバ101とバックアップサーバ103は、LAN102で接続され、バックアップサーバ103とバックアップ管理サーバ105、及びアーカイブ用ストレージ104は別のLAN102で接続された構成になっている。図1では示されていないが、仮想サーバ101、バックアップサーバ103、及びバックアップ管理サーバ105は、CPU（プロセッサ）、メモリ、表示装置等を有する通常のコンピュータで構成されており、各モジュールやプログラムはメモリ上に格納され、CPUによって実行されるようになっている。

[0026] このネットワーク構成により、バックアップサーバ103によるバックアップ処理で実行されるデータ通信を干渉することなく、バックアップ管理サーバ105がバックアップサーバ103に対して、検索用のクローリングやデータアーカイブ等の処理を行うことができる。

[0027] (i) 仮想サーバ

仮想サーバ101は、図示しないCPU（プロセッサ）、メモリ、表示装置等を有する通常のコンピュータで構成され、VMwareやHyper-V、Xen等の仮想環境が構築されている。そして、その仮想環境上で複数台のファイルサーバやメールサーバ、個人使用のシンクライアント接続対象となる仮想クライアント等を稼働させることが可能となっている。

[0028] (ii) バックアップサーバ

バックアップサーバ103は、仮想サーバ101上で稼働中の仮想マシンをバックアップ・保管するためのサーバであり、バックアップされた仮想マシンファイル群（仮想マシンイメージ群、仮想マシンイメージのレポジトリともいう）108を有している。また、バックアップサーバ103には、バックアップサービス107及び仮想マシンマウントモジュール106がインストールされ、稼働している。

[0029] バックアップサービス107は、仮想サーバ101にアクセスし、指定された稼働中の仮想マシンイメージをバックアップサーバ103内の仮想マシンファイル群108にコピーするサービス（プログラム）である。

[0030] また、仮想マシンマウントモジュール106は、バックアップ管理サーバ105内のクローリングサービス110からの要求に応じて、仮想マシンイメージをマウントして当該仮想マシンイメージを起動せずにその内容を参照できるようにする機能を持つとともに、指定ファイルのハッシュ値を計算してクローリングサービス110に通知する機能を有している。この通知されたハッシュ値により、クローリングサービス110は重複管理DB113をチェックして、検索エンジン111に既に登録済みかどうかを判定し、登録済みであればデータダウンロード処理は行わずに、メタ情報のみ取得して、検索エンジン111内の同一コンテンツのエントリを利用して複製処理を行う。

[0031] (iii) バックアップ管理サーバ

バックアップ管理サーバ105は、バックアップ管理サービス109と、クローリングサービス110と、検索エンジン111と、検索用インデックス112と、重複管理DB113と、アーカイブ管理DB114と、ポリシーファイル115と、を有している。

[0032] バックアップ管理サービス109は、利用者がバックアップデータの管理機能を利用するためのGUIを提供し、このサービスを用いてバックアップデータ内のファイル検索やデータ管理のためのファイルに関するポリシー設定

等が可能となる。ファイルに関するポリシーとして、例えば、個人情報や他社営業機密情報といった機微データ（機密情報とも言う）の管理ポリシーがある。このポリシーが設定されている場合、バックアップ管理サービス109は、機微データを含むファイル群をクローリング時に検出して、検索用インデックス112に登録しておき、任意のタイミングで機微データをバックアップデータ（仮想マシンファイル群108）から削除、または特定のセキュアなストレージシステムに移行する等の処理を行う。また、別のポリシーとして、例えばファイルサーバやメールサーバに格納された、会計データや業務上の帳票データ、メールデータ等のデータを別途アーカイブするといったアーカイブポリシーがある。このポリシーが設定されている場合、バックアップ管理サービス109は、バックアップデータから会計データや、帳票データ、メールデータを検出して、重複を排除した形でアーカイブ用ストレージ104にコピーを行う。このようなコンテンツの識別は、クローリングの際に実行され、識別結果は検索用インデックス112に記録される。上記のような機微データやアーカイブ用の特殊フォーマットデータ等を識別する機能はクローリングサービス110が保持するが、本発明ではこの識別技術については特に限定しない。公知の識別技術として、例えば、住所や電話番号等の個人情報関連のキーワードによる検出や、ドキュメントのテンプレートとの画像マッチングによる識別、ファイルに付与されたコンテンツ内容の属性を表すメタ情報を利用した検出、ファイル拡張子やヘッダー情報によるファイルタイプ識別等がある。本発明では、これら既存技術を組み合わせて利用するようにしても良い。

[0033] クローリングサービス110は、このコンテンツ識別機能の他に、仮想マシンマウントモジュール106と連携して、バックアップサーバ103上の仮想マシンファイル群108のうちマウントされたファイルにアクセスし、内部のファイル群に関する検索用インデックス作成のためのクローリング処理機能を有する。

[0034] 検索エンジン111は、クローリングサービス110によって登録された

データに対して、利用者が検索できる機能と、検索用インデックス 112 を作成する機能を提供している。検索機能に関しては、例えば、オープンソースの検索エンジンとして Lucene や Senna 等が良く知られている。検索エンジン 111 は、検索用インデックス 112 内を検索することにより、所望のキーワード等を含むファイルを取得することができ、必要に応じて検索結果を表示装置の表示画面上に表示する。そして、例えば、ユーザは、検索結果から所望のファイルを選択して閲覧したり、バックアップ管理サービス 109 を介して当該ファイルをアーカイブ用ストレージ 104 に移動することが可能となる。

[0035] 検索用インデックス 112 は、検索エンジン 111 が利用するデータベースであり、登録されたコンテンツを管理するドキュメント管理テーブル（図 4 参照）と、キーワード検索に利用される転置インデックステーブル（図 5 参照）を保持する。

[0036] 重複管理 DB 113 は、クローリングサービス 110 による検索エンジン 111 の登録の際に、登録対象ファイルが既に検索エンジン 111 に登録されているかどうかを判定するためのデータベースである。

[0037] アーカイブ管理 DB 114 は、バックアップ管理サービス 109 がアーカイブしたデータを管理するデータベースである。アーカイブ管理 DB 114 は、アーカイブ対象データを仮想マシンマウントモジュール 106 経由で仮想マシンファイル群から抽出し、固めて（一まとめにして）アーカイブ用ストレージ 104 に保存される際に更新される。また、アーカイブ管理 DB 114 は、検索エンジン 111 を用いてバックアップ管理サービス 109 経由で、アーカイブされたデータを検索する際に、そのファイルの所在情報を取得する際にも利用される。

[0038] ポリシファイル 115 は、上記のファイル管理ポリシーに関する設定情報を保持したファイルである。

[0039] なお、本実施形態では、バックアップサーバ 103 とバックアップ管理サーバ 105 は、別々のコンピュータで構成されているが、これらは 1 つのコ

ンピュータ内に実装されていても良い。

[0040] <アーカイブ管理DB>

図2は、アーカイブ管理DBの構成例（テーブル形式）を示す図である。アーカイブ管理DB114は、テーブルを構成する属性値として、セッションID201と、日時202と、格納パス203と、カタログ格納パス204と、を有している。

[0041] セッションID201は、アーカイブセッションごとに割り振られるIDであり、アーカイブセッションの識別に利用する。日時202は、アーカイブセッションの実行された日時である。格納パス203は、アーカイブ用ストレージ先のアーカイブデータ格納パス、カタログ格納パス204はアーカイブデータのディレクトリ構成や各ファイルのメタ情報等のカタログ情報のアーカイブ用ストレージ上の格納先を示している。パスに関しては、例えばArc1¥bk1¥dataの場合、アーカイブ用ストレージ104のArc1上のパス¥bk1¥dataという意味である。

[0042] <重複管理DB>

図3は、重複管理DBの構成例（テーブル形式）を示す図である。重複管理DB113は、テーブルを構成する属性値として、コンテンツID301と、ハッシュ値302と、検索エンジン登録カウント303と、を有している。

[0043] コンテンツID301は、検索エンジン111に登録されたコンテンツに固有のIDである。ハッシュ値302は、コンテンツを含むファイルのハッシュ値であり、コンテンツ内容が異なると違う値になる。検索エンジン登録カウント303は、当該コンテンツが検索エンジンに登録されている数を示している。例えば、バックアップリポジトリ（仮想マシンファイル群108）内にファイル名は異なるが同じ内容を保持するファイルが5つ存在する場合を考える。そして、そのバックアップリポジトリの検索用インデックスを構築したとき、ファイル名は異なるが、5つの同じコンテンツが検索エンジンに登録される。このため、検索エンジン登録カウント303は「5」とな

る。

[0044] <ドキュメント管理テーブル>

図4は、検索用インデックス112に含まれるドキュメント管理テーブルの構成例を示す図である。ドキュメント管理テーブルは、テーブルを構成する属性値として、ドキュメントID401と、取得日時402と、コンテンツ403と、メタ情報ポインタ404と、ファイルパス405と、ACLポインタ406と、更新日時407と、サイズ408と、適用ポリシーID409と、アーカイブセッションID410と、アクセス制御エントリ411と、を有している。

[0045] ドキュメントID401は、検索エンジンに登録されたファイルごとに割り振られるID番号である。取得日時402は、クローリングサービスがデータを取得して検索エンジンに登録した日時である。コンテンツ403は、テキスト抽出されたファイル内容である。メタ情報ポインタ404は、各ファイルのメタ情報を格納したテーブルへのポインタ情報である。

[0046] 各ファイルのメタ情報テーブルにおいて、属性であるファイルパス405は、対象ファイルの格納先を示している。ACLポインタ406は、ファイルに設定されたアクセス制御リストへのポインタ情報である。更新日時407は、ファイルの最終更新日時である。サイズ408は、ファイルサイズである。また、適用ポリシーID409は、クローリング時に識別されたファイルの属性に応じて適用されるポリシーのIDを意味する。更に、アーカイブセッションID410は、適用ポリシーに応じてアーカイブ用ストレージにファイルが格納されている場合、そのファイルを格納したアーカイブ処理のセッションIDを示している。このID情報をキーにアーカイブ管理DB114を検索することで、アーカイブデータやカタログデータの格納先が分かる仕組みになっている。

[0047] アクセス制御リストは、アクセス制御エントリ411のリストデータを格納する。アクセス制御エントリ411はアクセス権限とアクセス内容を示している。例えば、図4にあるEveryone:Rは、すべてのユーザにとってReadの

みのアクセス権限を付与するという意味である。更に、例えば、SIDが000111222333の特定ユーザにのみフルコントロールのアクセス権限の与える場合には、000111222333:Fが追加でエントリされる形で、ファイルに対するアクセス権が定義される。

[0048] <転置インデックステーブル>

図5は、検索用インデックス112に含まれる転置インデックステーブルの構成例を示す図である。転置インデックステーブルは、テーブルを構成する属性値として、キーワード501と、ロケーション情報ポインタ502と、ドキュメントID503と、位置情報504と、を有している。

[0049] キーワード501は、ドキュメントに含まれるキーワードである。検索エンジン111にキーワードが渡されたとき、それをキーにして転置インデックスを検索して、キーワードが含まれるドキュメントを探することができるようになっている。ロケーション情報ポインタ502は、それぞれのキーワードが含まれているドキュメント群と各ドキュメント内のキーワードの位置を格納したテーブルへのポインタ情報である。そのポインタ情報は、属性として、指定キーワードを含むドキュメントを特定するためのドキュメントID503と、対応するドキュメント内のキーワードの位置を始点・終点情報のペアで示す位置情報504と、を有している。

[0050] <仮想マシンマウントモジュールの処理内容>

図6は、クローリング時の仮想マシンマウントモジュール106の処理内容を説明するためのフローチャートである。

[0051] クローリング処理の際、クローリングサービス110からマウント要求が発行され、仮想マシンマウントモジュール106は、対象となる仮想マシンイメージのマウント要求を受信すると（ステップ601）、バックアップサービスに問い合わせ、対象となる仮想マシンイメージをマウントする（ステップ602）。

[0052] 次に、仮想マシンマウントモジュール106は、マウントしたディレクトリを共有フォルダ化し、共有フォルダへのアクセスパスをクローリングに通

知する（ステップ603）。これにより、クローリングサービス110がマウントしたディレクトリにアクセス可能となる。

[0053] その後、クローリングサービス110がCIFS/NFSで共有フォルダにアクセスし、クローリングを行うが、仮想マシンマウントモジュール106は、その際にクローリングサービス110から要求を受け、指定ファイル群のハッシュ値を計算して、逐次クローリングサービス110に通知する（ステップ604）。このハッシュ値をベースに、クローリングサービス110は、検索エンジン111にファイルが既に登録済みかどうかを判定し、登録されていない場合はデータをダウンロードして登録処理を進める。この処理により重複したデータのダウンロードが回避される。

[0054] <クローリングサービスの処理内容>

図7は、クローリング時のクローリングサービス110の処理内容を説明するためのフローチャートである。

[0055] 始めにクローリングサービス110は、仮想マシンマウントモジュール106に対して、検索用インデックス作成のターゲットとして指定した仮想マシンイメージのマウント要求を送る（ステップ701）。前述したように（図6参照）、この要求を受けて、仮想マシンマウントモジュール106は、指定された仮想マシンイメージをマウントし、マウントされた共有フォルダのパスを通知し、クローリングサービス110は、このパスを取得する（ステップ702）。

[0056] クローリングサービス110は、通知されたパスに基づいて共有フォルダ上のファイルに順次アクセスして、クローリング処理を行う（ステップ703）。

[0057] クローリングサービス110は、アクセスしたファイルがクローリング対象外かどうかを判定する（ステップ704）。クローリング対象である場合（ステップ704でNOの場合）には、処理はステップ705に移行し、クローリング対象ではない場合（ステップ704でYESの場合）には、処理はステップ708に移行する。より具体的には、アクセスしたファイルがO

Sのシステムファイルやアプリケーションファイル等の、検索対象とならないファイルの場合があり、それを対象外とする。この判定は、例えばNISTが定義しているOSやアプリのファイル群のリストを用いたdeNIST処理をすることで実現可能である。

[0058] ステップ705では、クローリングサービス110は、当該ファイルが検索エンジン111に登録済みであるかどうかについて、ハッシュ値を基に調べる(ステップ705)。登録済みの場合(ステップ705でYESの場合)処理はステップ706に移行する。登録済でない場合(ステップ705でNOの場合)、処理はステップ707に移行する。

[0059] ステップ706では、クローリングサービス110は、検索用インデックス112に登録された情報をコピーし、メタ情報の中で異なる部分を更新する。

[0060] 一方、ステップ707では、クローリングサービス110は、データ取得・検索エンジン登録処理(図8)を行う。

[0061] そして、クローリングサービス110は、対象ファイルがすべてクローリング済みかどうかをチェックする(ステップ708)。共有フォルダ上のすべてクローリングが終わるまで、順次クローリング処理が繰り返される。

[0062] <データ取得・検索エンジン登録処理の詳細>

図8は、クローリング処理におけるデータ取得・検索エンジン登録処理(ステップ707)の詳細を説明するためのフローチャートである。

[0063] まず、クローリングサービス110は、クローリング対象となるファイルを共有フォルダからダウンロードし(ステップ801)、登録用に当該ファイルのファイルパスや更新日時等のメタ情報を取得する(ステップ802)。

[0064] 次に、クローリングサービス110は、当該ファイルのテキストデータを抽出して(ステップ803)、抽出データを基に、当該ファイルの識別処理(前述の公知の技術を利用する)を行い、適用するポリシーに対応する適用ポリシーIDを判定する(ステップ804)。

- [0065] 最後に、クローリングサービス110は、検索エンジン登録用のフォーマットのデータを生成して、検索エンジン111に登録する（ステップ805）。
- [0066] <ファイル処理の内容>
- 図9は、バックアップ管理サービス109がポリシーに応じて実行するファイル処理の内容を説明するためのフローチャートである。
- [0067] まず、バックアップ管理サービス109は、アーカイブ用ポリシーや機微データ検出ポリシー等、各ポリシーで設定されたスケジュールに応じて、適用ポリシーIDをベースにバックアップデータを検索する（ステップ901）。
- [0068] そして、バックアップ管理サービス109は、適用ポリシーIDでバックアップデータを検索する際に、適用ポリシーIDがアーカイブに対応するか否かを判定する（ステップ902）。
- [0069] 適用ポリシーIDがアーカイブ用ポリシーのIDの場合（ステップ902でYESの場合）、バックアップ管理サービス109は、検索結果より対応ファイル群をリストアップしてアーカイブフォーマットにし、アーカイブ用ストレージに格納する（ステップ905）。例えば、週次でメールデータのアーカイブを取る場合、メールアーカイブの適用ポリシーIDが付与されたエントリを検索し、アーカイブ用データを作成してアーカイブ用ストレージに格納する場合に、このような処理が行われる。
- [0070] 適用ポリシーIDがアーカイブ用ポリシーのIDでない場合（ステップ902でNOの場合）、バックアップ管理サービス109は、さらに、適用ポリシーIDが機微データ管理用ポリシーのIDに一致するかチェックする（ステップ903）。
- [0071] 適用ポリシーIDが機微データ管理用ポリシーのIDに一致する場合（ステップ903でYESの場合）、バックアップ管理サービス109は、ポリシーファイルに設定された機微ファイルに対する処理ポリシーに応じて、管理者への警告やデータのバックアップデータからの削除、指定されたセキュアなストレージへの移動処理等を実施する（ステップ904）。

[0072] 適用ポリシーIDが機微データ管理用ポリシーのIDに一致しない場合（ステップ903でNOの場合）、処理は終了する。

[0073] 以上の構成を採ることにより、仮想マシンイメージのバックアップシステムにおいて、バックアップされたイメージデータを管理する上で、イメージ内にあるファイル検索や、設定されたアーカイブや機微データ検出等のポリシーに応じたデータ管理を実現することができる。また、バックアップデータ内には重複したファイルが大量に存在するが、検索用インデックスを作成・更新する際に、そうした重複を考慮した効率的な処理を行うことができる。

[0074] <まとめ>

本発明の実施形態では、仮想マウントモジュール（仮想マシンマウント処理部）が、仮想サーバの仮想マシンイメージをバックアップし、仮想マシンファイル群（仮想マシンリポジトリ）にコピーする。また、仮想マウントモジュールは、クローリングサービス（クローリング処理部）のマウント要求に回答して、仮想マシンファイル群に格納された少なくとも1つの仮想マシンイメージをマウント処理する。そして、クローリングサービスが、マウントされた仮想マシンイメージにアクセスし、ファイル検索用インデックス作成のためのクローリング処理を実行する。また、バックアップ管理サービスが、マウントされた仮想マシンイメージに含まれるファイルに対して、ファイル操作を実行する。このように、仮想マシンイメージをマウントして、内部に保存された個々のファイルに対する検索・ファイル操作する仕組みを実現することで、内部に保存されたファイルのコンテンツ内容に応じた仮想マシンイメージデータの検索・管理が可能となる。

[0075] また、仮想マシンマウントモジュールは、仮想マシンイメージに含まれるファイルを送信する前に、そのファイルのハッシュ値をクローリングサービスに通知する。そして、クローリングサービスが、ハッシュ値を用いて、同一内容を有するファイルが既に取得済か否か判断し、取得済でないファイルのみ取得するようにする。また、当該ファイルについては、メタ情報のみ取得し、既に取得済のファイルからデータをコピーして、検索用インデックス

に登録する。このようにすることで、ファイル名が異なっても内容が同一のファイルについては、再度取得する必要がなく、処理を効率化することが可能となる。なお、この取得済か否かの判断については、検索エンジンに登録したデータに関する重複管理DBを設け、これを参照して登録・更新処理を行うようにする。これにより、バックアップデータに適した効率的な検索用インデックスに対する処理を行うことができる。

[0076] さらに、検索インデックスに登録されるファイルには、適用ポリシーが付与される。この適用ポリシーは、ファイルに含まれるデータの内容に従って当該ファイルの取り扱いを規定する情報である。この適用ポリシーは、検索エンジンが、クロールサービスが取得したファイルに対して識別処理を実行してデータの内容を識別し、各ファイルに対して付与される。そして、バックアップ管理サービス（バックアップ管理処理部）は、検索用インデックスに登録されたファイルに付与された適用ポリシーに従って、仮想マシンファイル群に格納された仮想マシンイメージに含まれるファイルに対して所定のファイル操作を実行する。このように、検索用インデックス作成時に、コンテンツ内容を判定してアーカイブ管理や機微データ管理等の適用ポリシーを設定することで、検索エンジンに登録されたバックアップデータのポリシーに応じたデータ管理が可能となる。

[0077] なお、本発明は、実施形態の機能を実現するソフトウェアのプログラムコードによっても実現できる。この場合、プログラムコードを記録した記憶媒体をシステム或は装置に提供し、そのシステム或は装置のコンピュータ（又はCPUやMPU）が記憶媒体に格納されたプログラムコードを読み出す。この場合、記憶媒体から読み出されたプログラムコード自体が前述した実施形態の機能を実現することになり、そのプログラムコード自体、及びそれを記憶した記憶媒体は本発明を構成することになる。このようなプログラムコードを供給するための記憶媒体としては、例えば、フレキシブルディスク、CD-ROM、DVD-ROM、ハードディスク、光ディスク、光磁気ディスク、CD-R、磁気テープ、不揮発性のメモリカード、ROMなどが用い

られる。

[0078] また、プログラムコードの指示に基づき、コンピュータ上で稼動しているOS（オペレーティングシステム）などが実際の処理の一部又は全部を行い、その処理によって前述した実施の形態の機能が実現されるようにしてもよい。さらに、記憶媒体から読み出されたプログラムコードが、コンピュータ上のメモリに書きこまれた後、そのプログラムコードの指示に基づき、コンピュータのCPUなどが実際の処理の一部又は全部を行い、その処理によって前述した実施の形態の機能が実現されるようにしてもよい。

[0079] さらに、実施の形態の機能を実現するソフトウェアのプログラムコードを、ネットワークを介して配信することにより、それをシステム又は装置のハードディスクやメモリ等の記憶手段又はCD-RW、CD-R等の記憶媒体に格納し、使用時にそのシステム又は装置のコンピュータ（又はCPUやMPU）が当該記憶手段や当該記憶媒体に格納されたプログラムコードを読み出して実行するようにしても良い。

[0080] 最後に、ここで述べたプロセス及び技術は本質的に如何なる特定の装置に関連することはなく、コンポーネントの如何なる相応しい組み合わせによってでも実装できることを理解する必要がある。更に、汎用目的の多様なタイプのデバイスがここで記述した教授に従って使用可能である。ここで述べた方法のステップを実行するのに、専用の装置を構築するのが有益であることが判るかもしれない。また、実施形態に開示されている複数の構成要素の適宜な組み合わせにより、種々の発明を形成できる。例えば、実施形態に示される全構成要素から幾つかの構成要素を削除してもよい。さらに、異なる実施形態にわたる構成要素を適宜組み合わせてもよい。本発明は、具体例に関連して記述したが、これらは、すべての観点に於いて限定の為ではなく説明の為である。本分野にスキルのある者には、本発明を実施するのに相応しいハードウェア、ソフトウェア、及びファームウェアの多数の組み合わせがあることが解るであろう。例えば、記述したソフトウェアは、アセンブラ、C/C++、perl、Shell、PHP、Java（登録商標）等の広範

囲のプログラム又はスクリプト言語で実装できる。

[0081] さらに、上述の実施形態において、制御線や情報線は説明上必要と考えられるものを示しており、製品上必ずしも全ての制御線や情報線を示しているとは限らない。全ての構成が相互に接続されていても良い。

[0082] 加えて、本技術分野の通常の知識を有する者には、本発明のその他の実装がここに開示された本発明の明細書及び実施形態の考察から明らかになる。記述された実施形態の多様な態様及び／又はコンポーネントは、データを管理する機能を有するコンピュータ化ストレージシステムに於いて、単独又は如何なる組み合わせでも使用することが出来る。明細書と具体例は典型的なものに過ぎず、本発明の範囲と精神は後続する請求範囲で示される。

符号の説明

- [0083] 1 0 1 仮想サーバ
- 1 0 2 LAN
- 1 0 3 バックアップサーバ
- 1 0 4 アーカイブ用ストレージ
- 1 0 5 バックアップ管理サーバ
- 1 0 6 仮想マシンマウントモジュール
- 1 0 7 バックアップサービス
- 1 0 8 仮想マシンファイル群
- 1 0 9 バックアップ管理サービス
- 1 1 0 クローリングサービス
- 1 1 1 検索エンジン
- 1 1 2 検索用インデックス
- 1 1 3 重複管理DB
- 1 1 4 アーカイブ管理DB
- 1 1 5 ポリシファイル
- 2 0 1 セッションID
- 2 0 2 日時

- 203 格納パス
- 204 カタログ格納パス
- 301 コンテンツID
- 302 ハッシュ値
- 303 検索エンジン登録カウント
- 401 ドキュメントID
- 402 取得日時
- 403 コンテンツ
- 404 メタ情報ポインタ
- 405 ファイルパス
- 406 ACLポインタ
- 407 更新日時
- 408 サイズ
- 409 適用ポリシーID
- 410 アーカイブセッションID
- 411 アクセス制御エントリ
- 501 キーワード
- 502 ロケーション情報ポインタ
- 503 ドキュメントID
- 504 位置情報

請求の範囲

- [請求項1] 仮想マシンイメージのバックアップを管理する情報処理システムであって、
- バックアップされた仮想マシンイメージを格納する仮想マシンリポジトリと、
- 仮想サーバの仮想マシンイメージをバックアップし、前記仮想マシンリポジトリにコピーすると共に、前記仮想マシンリポジトリに格納された少なくとも1つの仮想マシンイメージをマウント処理する仮想マシンマウント処理部と、
- 前記マウントされた仮想マシンイメージにアクセスし、ファイル検索用インデックス作成のためのクローリング処理を実行するクローリング処理部と、を有し、
- 前記仮想マシンマウント処理部は、前記クローリング処理部の要求に応じて前記仮想マシンイメージをマウントすることを特徴とする情報処理システム。
- [請求項2] 請求項1において、
- 前記仮想マシンマウント処理部は、前記仮想マシンイメージに含まれるファイルのハッシュ値を前記クローリング処理部に通知し、
- 前記クローリング処理部は、前記ハッシュ値を用いて、同一内容を有するファイルが既にダウンロードされて登録済か否か判断し、登録済でないファイルのみダウンロードすることを特徴とする情報処理システム。
- [請求項3] 請求項2において、
- さらに、前記仮想マシンイメージに含まれる対象ファイルの検索用インデックスを作成する検索エンジンを有し、
- 前記クローリング処理部は、登録済のファイルについては、前記対象ファイルのメタ情報のみを前記仮想マシンマウント処理部から取得し、

前記検索エンジンは、登録済のファイルのデータをコピーし、前記取得されたメタ情報を更新して前記検索用インデックスを作成することを特徴とする情報処理システム。

[請求項4]

請求項3において、

前記検索エンジンは、前記クローリング処理部によって取得されたファイルに対して識別処理を実行してデータの内容を識別し、ファイルに含まれるデータの内容に従って当該ファイルの取り扱いを規定する適用ポリシーであって、前記データの内容に適合する適用ポリシーを前記検索用インデックスに付与することを特徴とする情報処理システム。

[請求項5]

請求項4において、

さらに、前記検索用インデックスに登録されたファイルに付与された前記適用ポリシーに従って、前記仮想マシンリポジトリに格納された仮想マシンイメージに含まれるファイルに対して所定のファイル操作を実行するバックアップ管理処理部を有することを特徴とする情報処理システム。

[請求項6]

仮想マシンイメージのバックアップを管理する情報処理システムにおけるバックアップ管理方法であって、

前記情報処理システムは、バックアップされた仮想マシンイメージを格納する仮想マシンリポジトリと、仮想マシンマウント処理部と、クローリング処理部と、を有し、

前記バックアップ管理方法は、

前記仮想マシンマウント処理部が、仮想サーバの仮想マシンイメージをバックアップし、前記仮想マシンリポジトリにコピーするステップと、

前記仮想マシンマウント処理部が、前記クローリング処理部の要求に応じて、前記仮想マシンリポジトリに格納された少なくとも1つの仮想マシンイメージをマウントするステップと、

前記クローリング処理部が、前記マウントされた仮想マシンイメージにアクセスし、ファイル検索用インデックス作成のためのクローリング処理を実行するステップと、
を有することを特徴とするバックアップ管理方法。

[請求項7]

請求項6において、

さらに、前記仮想マシンマウント処理部が、前記仮想マシンイメージに含まれるファイルのハッシュ値を前記クローリング処理部に通知するステップと、

前記クローリング処理部が、前記ハッシュ値を用いて、同一内容を有するファイルが既にダウンロードされて登録済か否か判断し、登録済でないファイルのみダウンロードするステップと、
を有することを特徴とするバックアップ管理方法。

[請求項8]

請求項7において、

前記情報処理システムは、さらに、前記仮想マシンイメージに含まれる対象ファイルの検索用インデックスを作成する検索エンジンを有し、

前記方法は、さらに、

前記クローリング処理部が、登録済のファイルについては、前記対象ファイルのメタ情報のみを前記仮想マシンマウント処理部から取得するステップと、

前記検索エンジンが、登録済のファイルのデータをコピーし、前記取得されたメタ情報を更新して前記検索用インデックスを作成するステップと、

を有することを特徴とするバックアップ管理方法。

[請求項9]

請求項8において、

さらに、前記検索エンジンが、前記クローリング処理部によって取得されたファイルに対して識別処理を実行してデータの内容を識別するステップと、

前記検索エンジンが、ファイルに含まれるデータの内容に従って当該ファイルの取り扱いを規定する適用ポリシーであって、前記データの内容に適合する適用ポリシーを前記検索用インデックスに付与するステップと、
を有することを特徴とするバックアップ管理方法。

[請求項10]

請求項9において、

前記情報処理システムは、さらに、バックアップ管理処理部を有し、

前記方法は、さらに、前記バックアップ管理処理部が、前記検索用インデックスに登録されたファイルに付与された前記適用ポリシーに従って、前記仮想マシンリポジトリに格納された仮想マシンイメージに含まれるファイルに対して所定のファイル操作を実行するステップを有することを特徴とするバックアップ管理方法。

[請求項11]

仮想マシンイメージのバックアップを管理する情報処理システムにおけるバックアップ管理を実行するためのプログラムであって、

前記プログラムは、少なくとも1つのコンピュータに、

仮想サーバの仮想マシンイメージをバックアップし、当該仮想マシンイメージを格納するための仮想マシンリポジトリにコピーする処理と、

マウント要求に応答して、前記仮想マシンリポジトリに格納された少なくとも1つの仮想マシンイメージをマウントする処理と、

前記マウントされた仮想マシンイメージにアクセスし、ファイル検索用インデックス作成のためのクローリング処理を実行する処理と、
を実行させるプログラム。

補正された請求の範囲
[2012年7月3日 (03.07.2012) 国際事務局受理]

[請求項 1] (補正後)

仮想マシンイメージのバックアップを管理する情報処理システムであつて、
バックアップされた仮想マシンイメージを格納する仮想マシンリポジトリと、
仮想サーバの仮想マシンイメージをバックアップし、前記仮想マシンリポジトリ
にコピーすると共に、前記仮想マシンリポジトリに格納された少なくとも 1 つの仮
想マシンイメージをマウント処理する仮想マシンマウント処理部と、

前記マウントされた仮想マシンイメージにアクセスし、ファイル検索用インデッ
クス作成のためのクローリング処理を実行するクローリング処理部と、を有し、

前記仮想マシンマウント処理部は、前記クローリング処理部の要求に応じて前記
仮想マシンイメージをマウントし、

前記仮想マシンマウント処理部は、前記仮想マシンイメージに含まれるファイル
のハッシュ値を前記クローリング処理部に通知し、

前記クローリング処理部は、前記ハッシュ値を用いて、同一内容を有するファイ
ルが既にダウンロードされて登録済か否か判断し、登録済でないファイルのみダウ
ンロードすることを特徴とする情報処理システム。

[請求項 2] (削除)

[請求項 3] (補正後)

請求項 1 において、

さらに、前記仮想マシンイメージに含まれる対象ファイルの検索用インデックス
を作成する検索エンジンを有し、

前記クローリング処理部は、登録済のファイルについては、前記対象ファイルの
メタ情報のみを前記仮想マシンマウント処理部から取得し、

前記検索エンジンは、登録済のファイルのデータをコピーし、前記取得されたメ
タ情報を更新して前記検索用インデックスを作成することを特徴とする情報処理シ
ステム。

[請求項 4]

請求項 3 において、

補正された用紙 (条約第19条)

前記検索エンジンは、前記クローリング処理部によって取得されたファイルに対して識別処理を実行してデータの内容を識別し、ファイルに含まれるデータの内容に従って当該ファイルの取り扱いを規定する適用ポリシーであって、前記データの内容に適合する適用ポリシーを前記検索用インデックスに付与することを特徴とする情報処理システム。

[請求項5]

請求項4において、

さらに、前記検索用インデックスに登録されたファイルに付与された前記適用ポリシーに従って、前記仮想マシンリポジトリに格納された仮想マシンイメージに含まれるファイルに対して所定のファイル操作を実行するバックアップ管理処理部を有することを特徴とする情報処理システム。

[請求項6] (補正後)

仮想マシンイメージのバックアップを管理する情報処理システムにおけるバックアップ管理方法であって、

前記情報処理システムは、バックアップされた仮想マシンイメージを格納する仮想マシンリポジトリと、仮想マシンマウント処理部と、クローリング処理部と、を有し、

前記バックアップ管理方法は、

前記仮想マシンマウント処理部が、仮想サーバの仮想マシンイメージをバックアップし、前記仮想マシンリポジトリにコピーするステップと、

前記仮想マシンマウント処理部が、前記クローリング処理部の要求に応じて、前記仮想マシンリポジトリに格納された少なくとも1つの仮想マシンイメージをマウントするステップと、

前記クローリング処理部が、前記マウントされた仮想マシンイメージにアクセスし、ファイル検索用インデックス作成のためのクローリング処理を実行するステップと、

前記仮想マシンマウント処理部が、前記仮想マシンイメージに含まれるファイルのハッシュ値を前記クローリング処理部に通知するステップと、

前記クローリング処理部が、前記ハッシュ値を用いて、同一内容を有するファイルが既にダウンロードされて登録済か否か判断し、登録済でないファイルのみダウンロードするステップと、
を有することを特徴とするバックアップ管理方法。

[請求項7] (削除)

[請求項8] (補正後)

請求項6において、

前記情報処理システムは、さらに、前記仮想マシンイメージに含まれる対象ファイルの検索用インデックスを作成する検索エンジンを有し、

前記方法は、さらに、

前記クローリング処理部が、登録済のファイルについては、前記対象ファイルのメタ情報のみを前記仮想マシンマウント処理部から取得するステップと、

前記検索エンジンが、登録済のファイルのデータをコピーし、前記取得されたメタ情報を更新して前記検索用インデックスを作成するステップと、
を有することを特徴とするバックアップ管理方法。

[請求項9]

請求項8において、

さらに、前記検索エンジンが、前記クローリング処理部によって取得されたファイルに対して識別処理を実行してデータの内容を識別するステップと、

前記検索エンジンが、ファイルに含まれるデータの内容に従って当該ファイルの取り扱いを規定する適用ポリシーであって、前記データの内容に適合する適用ポリシーを前記検索用インデックスに付与するステップと、
を有することを特徴とするバックアップ管理方法。

[請求項10]

請求項9において、

前記情報処理システムは、さらに、バックアップ管理処理部を有し、

前記方法は、さらに、前記バックアップ管理処理部が、前記検索用インデックスに登録されたファイルに付与された前記適用ポリシーに従って、前記仮想マシンリポ

ジトリに格納された仮想マシンイメージに含まれるファイルに対して所定のファイル操作を実行するステップを有することを特徴とするバックアップ管理方法。

[請求項 1 1] (補正後)

仮想マシンイメージのバックアップを管理する情報処理システムにおけるバックアップ管理を実行するためのプログラムであって、

前記プログラムは、少なくとも1つのコンピュータに、

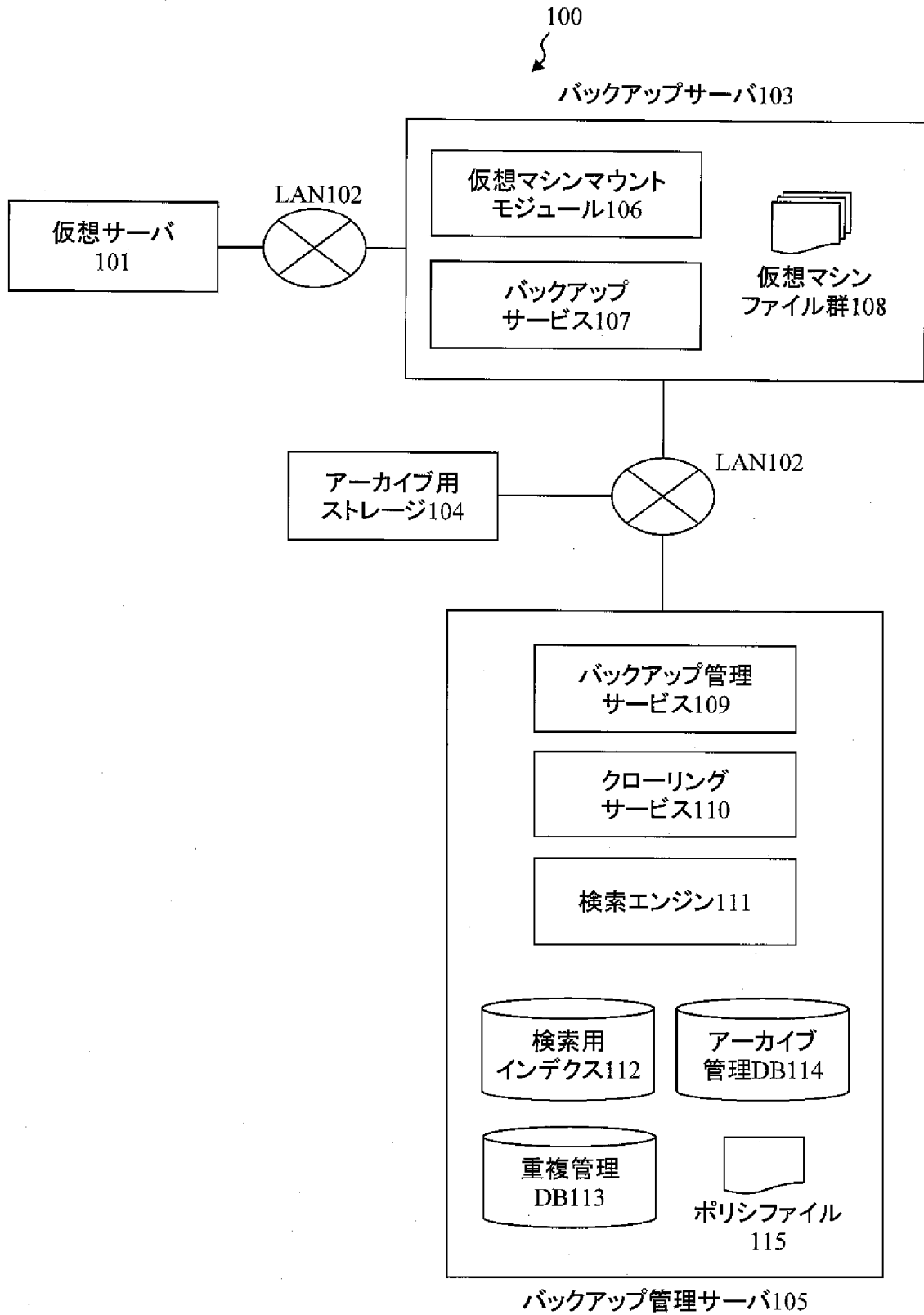
仮想サーバの仮想マシンイメージをバックアップし、当該仮想マシンイメージを格納するための仮想マシンリポジトリにコピーする処理と、

マウント要求に応答して、前記仮想マシンリポジトリに格納された少なくとも1つの仮想マシンイメージをマウントする処理と、

前記マウントされた仮想マシンイメージにアクセスし、ファイル検索用インデックス作成のためのクローリング処理を実行する処理と、

前記仮想マシンイメージに含まれるファイルのハッシュ値を取得し、当該ハッシュ値を用いて、同一内容を有するファイルが既にダウンロードされて登録済か否か判断し、登録済でないファイルのみダウンロードする処理と、
を実行させるプログラム。

[図1]



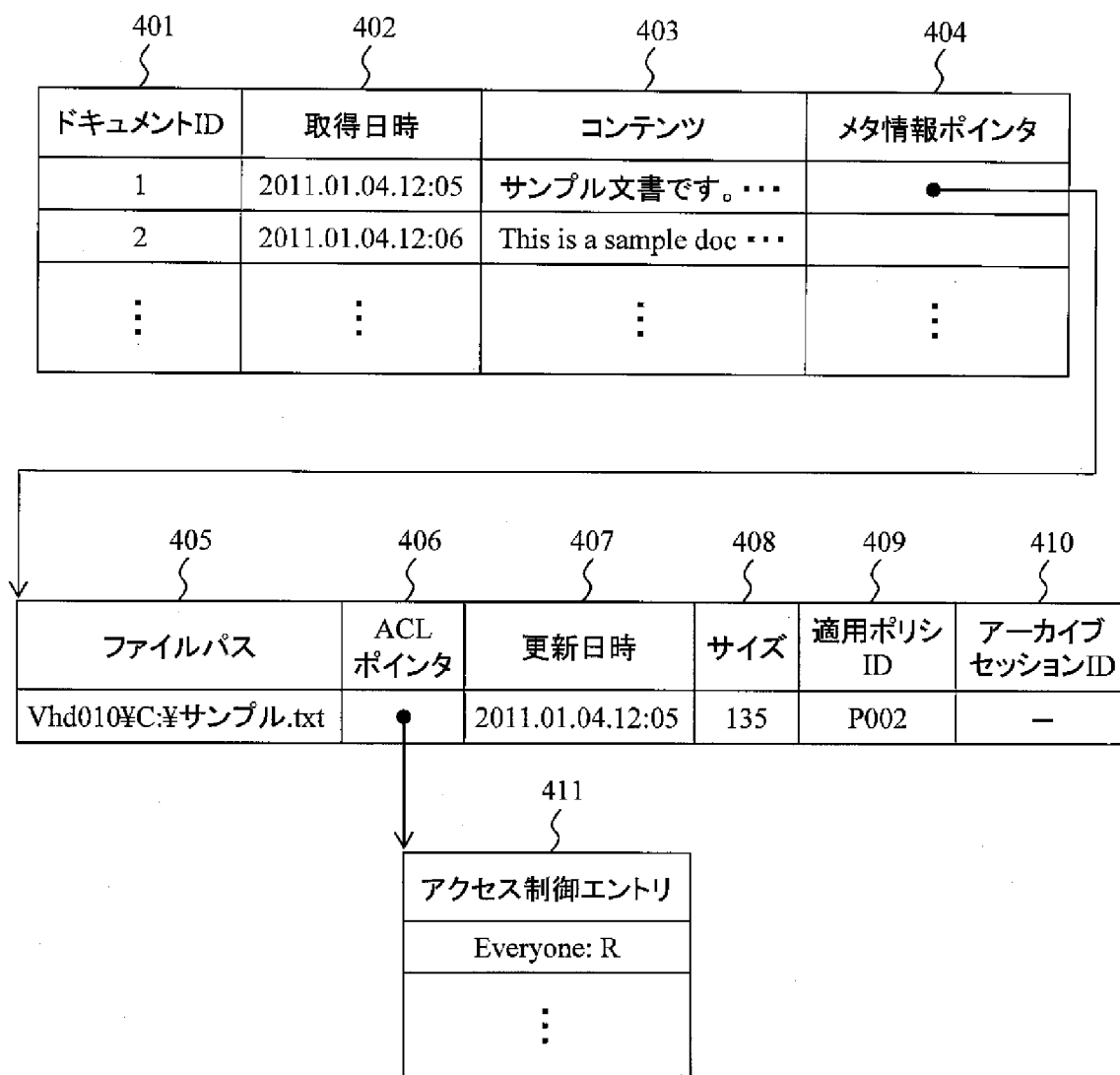
[図2]

セッションID	日時	格納パス	カタログ格納パス
B001	2010.12.07.12:00	Arc1¥bk1¥data	Arc1¥bk1¥catalog
B002	2010.12.08.12:00	Arc1¥bk2¥data	Arc1¥bk2¥catalog
⋮	⋮	⋮	⋮

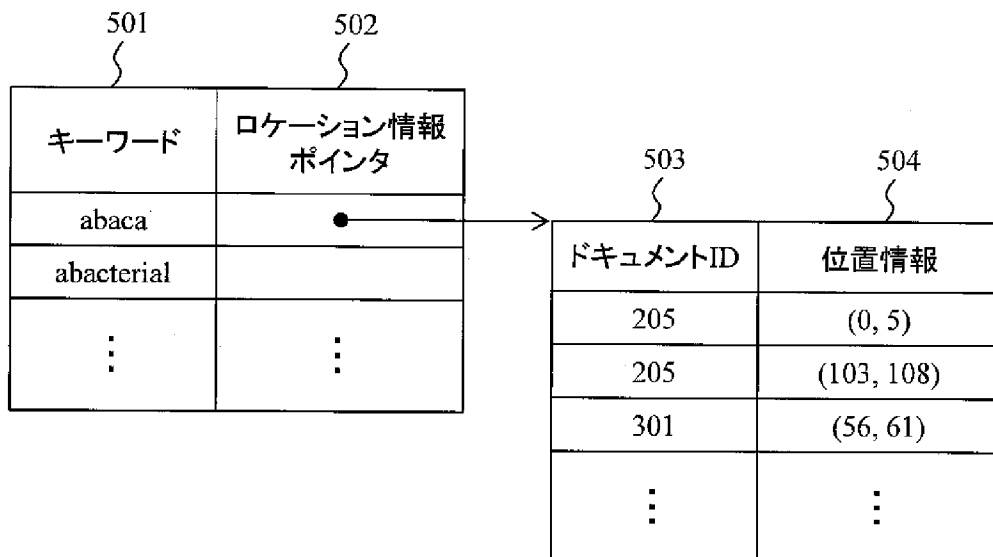
[図3]

コンテンツID	ハッシュ値	検索エンジン登録カウント
F001	23EG43A1TYAW	1
F002	153RQT66HUI	5
⋮	⋮	⋮

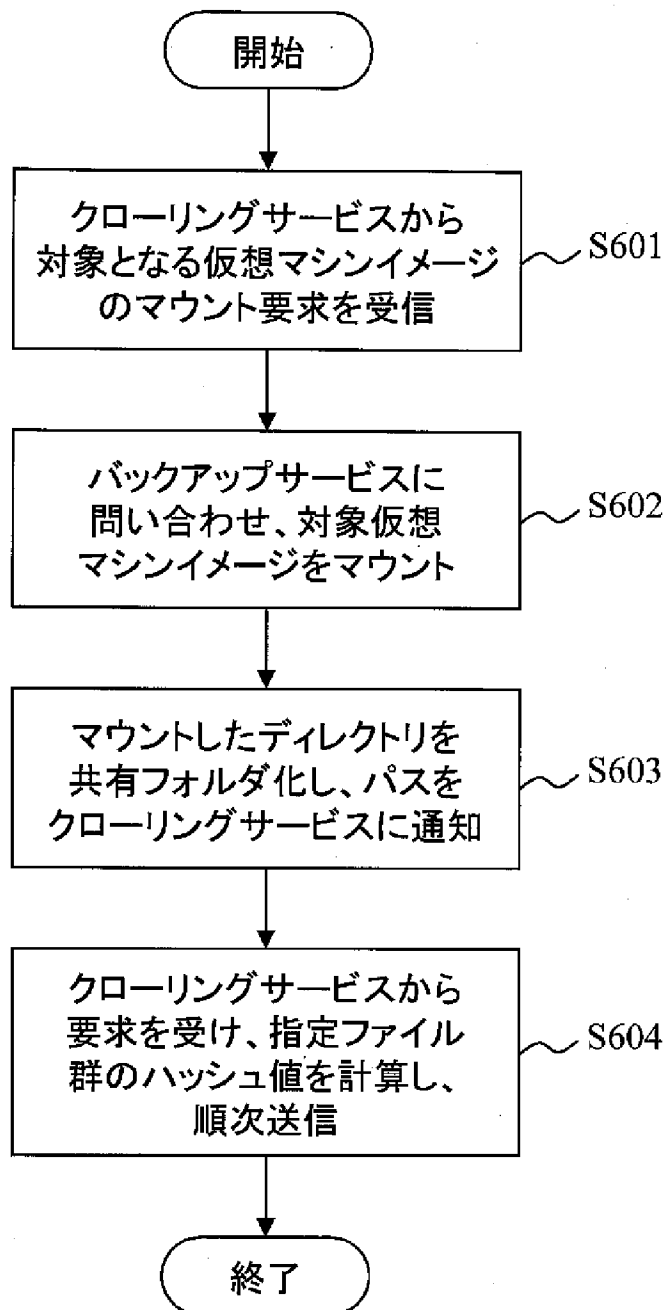
[図4]



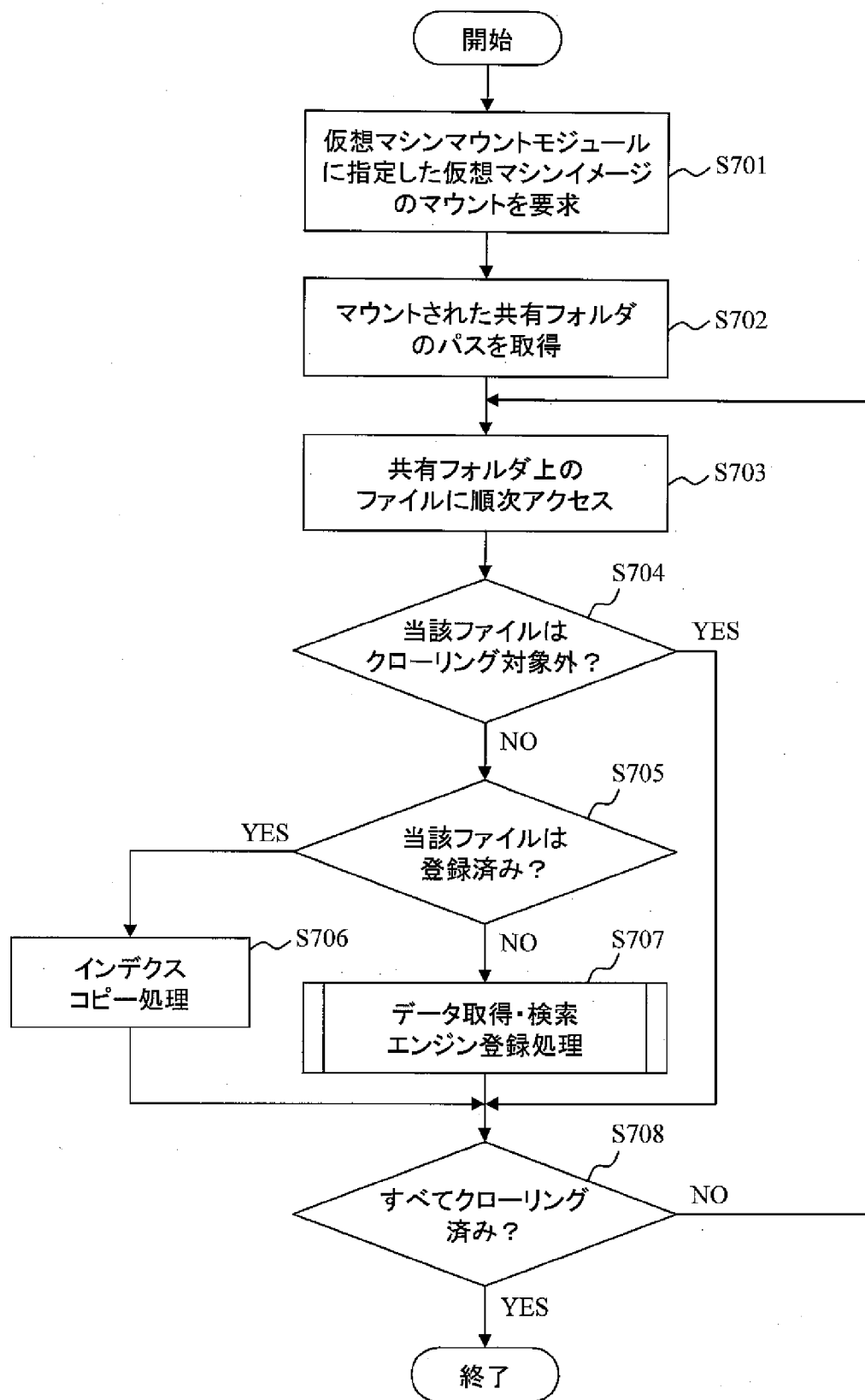
[図5]



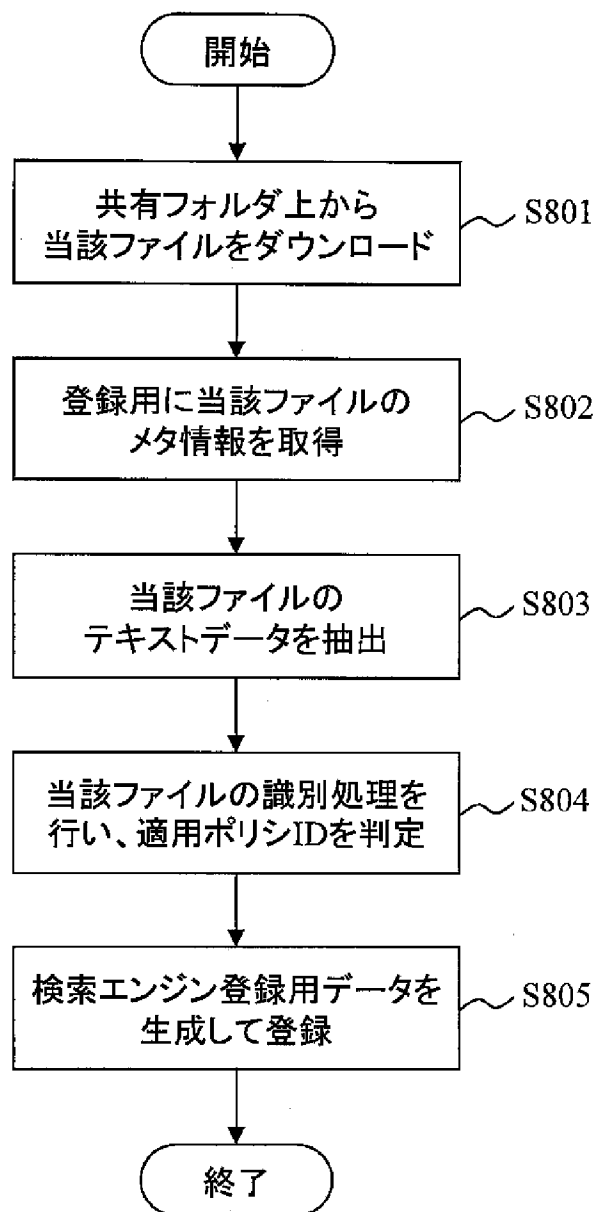
[図6]



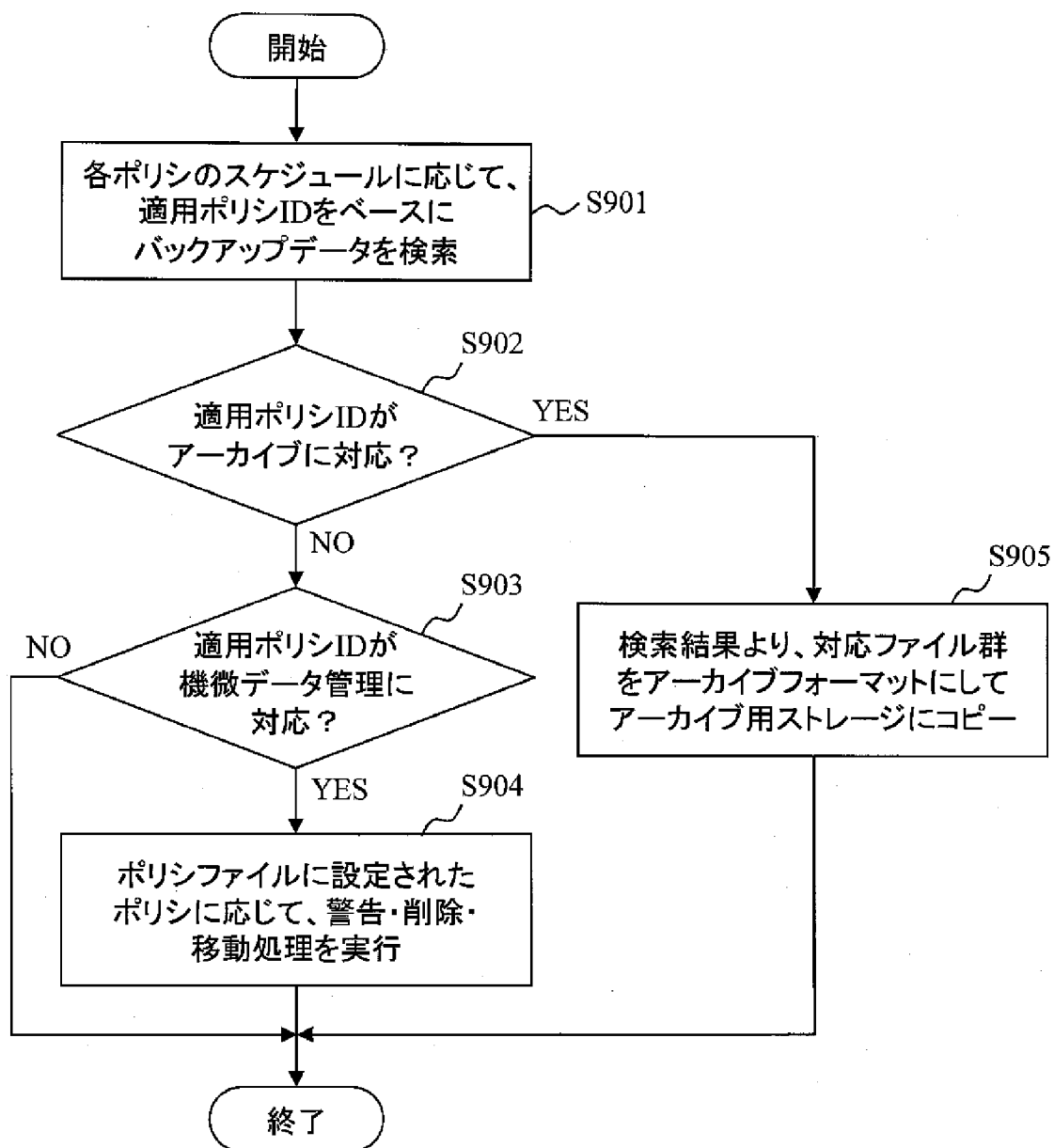
[図7]



[図8]



[図9]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2011/069130

A. CLASSIFICATION OF SUBJECT MATTER

G06F12/00(2006.01) i, G06F9/46(2006.01) 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)

G06F12/00, G06F9/46

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

Jitsuyo Shinan Koho	1922-1996	Jitsuyo Shinan Toroku Koho	1996-2011
Kokai Jitsuyo Shinan Koho	1971-2011	Toroku Jitsuyo Shinan Koho	1994-2011

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y A	JP 2010-33374 A (Hitachi, Ltd.), 12 February 2010 (12.02.2010), paragraphs [0002] to [0005], [0016] to [0071]; fig. 1 to 6 & US 2010/0030825 A1	1, 6, 11 2-5, 7-10
Y A	JP 2010-231257 A (NEC Corp.), 14 October 2010 (14.10.2010), paragraphs [0033] to [0035]; fig. 1, 6 (Family: none)	1, 6, 11 2-5, 7-10

Further documents are listed in the continuation of Box C.

See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search
08 September, 2011 (08.09.11)

Date of mailing of the international search report
20 September, 2011 (20.09.11)

Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

A. 発明の属する分野の分類 (国際特許分類 (IPC))

Int.Cl. G06F12/00(2006.01)i, G06F9/46(2006.01)i

B. 調査を行った分野

調査を行った最小限資料 (国際特許分類 (IPC))

Int.Cl. G06F12/00, G06F9/46

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	1922-1996年
日本国公開実用新案公報	1971-2011年
日本国実用新案登録公報	1996-2011年
日本国登録実用新案公報	1994-2011年

国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
Y A	JP 2010-33374 A (株式会社日立製作所) 2010.02.12, [0002]-[0005]段落, [0016]-[0071]段落, 第1-6図 & US 2010/0030825 A1	1, 6, 11 2-5, 7-10
Y A	JP 2010-231257 A (日本電気株式会社) 2010.10.14, [0033]-[0035]段落, 第1, 6図 (ファミリーなし)	1, 6, 11 2-5, 7-10

C欄の続きにも文献が列挙されている。

パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

「A」特に関連のある文献ではなく、一般的な技術水準を示すもの
 「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
 「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)
 「O」口頭による開示、使用、展示等に言及する文献
 「P」国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献
 「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
 「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
 「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
 「&」同一パテントファミリー文献

国際調査を完了した日

08.09.2011

国際調査報告の発送日

20.09.2011

国際調査機関の名称及びあて先

日本国特許庁 (ISA/JP)
 郵便番号100-8915
 東京都千代田区霞が関三丁目4番3号

特許庁審査官 (権限のある職員)

桜井 茂行

電話番号 03-3581-1101 内線 3599

5M

2945