(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
2008 年12 月18 日 (18.12.2008)

PCT

(10) 国際公開番号
WO 2008/152884 A1

(51) 国際特許分類:
G06F 17/30 (2006.01) *G06F 12/00* (2006.01)

(21) 国際出願番号: PCT/JP2008/059128

(22) 国際出願日: 2008 年5 月19 日 (19.05.2008)

(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2007-154467 2007 年6 月11 日 (11.06.2007) JP

(71) 出願人 および
(72) 発明者: 進藤 達也 (SHINDO, Tatsuya) [JP/JP]; 〒1500013 東京都渋谷区恵比寿一丁目１９番１９号 アクセラテクノロジ株式会社内 Tokyo (JP).

(74) 代理人: 名越 秀夫 (NAKOSHI, Hideo); 〒1060032 東京都港区六本木一丁目９番９号 六本木ファーストビル７階 生田・名越法律特許事務所 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, 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, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

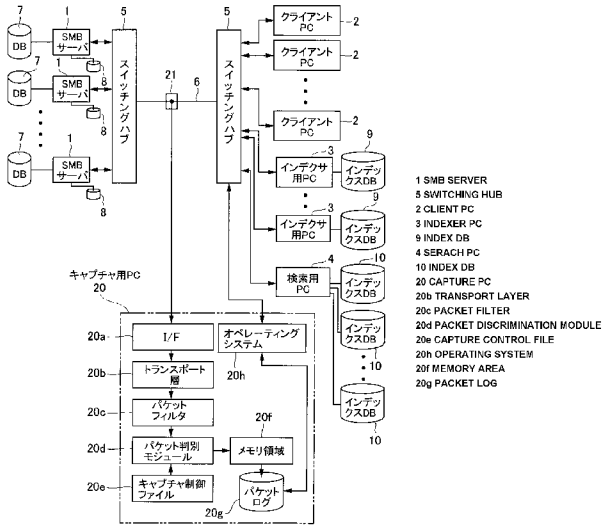
(84) 指定国 (表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE,

[続葉有]

(54) **Title:** FULL-TEXT SEARCH SYSTEM

(54) **発明の名称:** 全文検索システム

[図1]



(57) **Abstract:** Crawl time is shortened. A system comprises an SMB server(1) in which electronic documents are stored, a PC (2) for inputting the new creation, deletion, and partial change of the documents and also inputting the index of a document to be searched, a PC (3) for creating an index for indexing and storing the created index, a PC (4) for storing the index created by the PC (3), a communication path (6) for connection therebetween, and a PC (20) for extracting the update candidate information related to the new creation, deletion, and partial change of the electronic documents from the information accessed from the PC (2) to the server(1) to store it. The PC (3) scans the entire server(1) only when creating the index first. For the update involved in the new creation, partial change, and deletion of the index, the PC (3) checks the server(1) on the basis of the update candidate information stored in the PC (20), extracts the update execution information in which the new creation, deletion, and partial change have actually been executed, and performs the above update of the index on the basis of the update execution information.

[続葉有]

WO 2008/152884 A1

SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

添付公開書類:
— 国際調査報告書
— 補正書・説明書

(57) 要約: クロール時間の短縮。システムは、電子文書が蓄積されたＳＭＢサーバ１と、文書の新規作成，削除，部分的な変更の入力、検索対象文書のインデクスを入力するＰＣ２と、索引用インデクスを作成し、作成したインデクスを格納するＰＣ３と、ＰＣ３で作成されたインデクスを格納するＰＣ４と、これらの間を接続する通信路６とを備え、ＰＣ２からサーバ１へアクセスされる情報のうちから、電子文書の新規作成，削除，部分的な変更に関連する更新候補情報を抽出して、格納するＰＣ２０を備えている。ＰＣ３は、インデクスを最初に作成する際にのみ、サーバ１を全体走査し、インデクスの新規作成，部分的な変更および削除に伴う更新は、ＰＣ２０に格納されている更新候補情報に基づいて、サーバ１を照査して、新規作成，削除，部分的な変更が実際に行われた更新実行情報を抽出し、これに基づいてインデクスの更新を行う。

明 細 書

全文検索システム

技術分野

[0001] この発明は、全文検索システムに関し、特に、全文検索におけるインデクスの更新に関するものであり、大量のテキストデータ群を対象に任意のキーワードを含むデータを探し出す際に用いられ、特に、企業や官公庁など組織内でファイルサーバに蓄積される電子文書(プレーンテキスト、HTML、Word、Excel、Powerpoint、PDFなどオフィス系のフォーマット)を組織のメンバが、必要な電子文書をいつでも探し出せるようにする検索エンジンで使われる。本発明は、特に、検索対象となるデータ数が数T(テラ)クラス以上の大規模ストレッジで顕著な効果を発揮する。大量のテキストデータ群を高速に検索する手法として「転置ファイル法(Inverted file indexing)」が知られており、本発明は、転置ファイル法に代表される「インデクス」を作成するタイプの検索技術にかかわるものである。

背景技術

[0002] 大量のテキストデータ群を高速に検索する方法として、転置ファイル法が知られている。この転置ファイル法は、技術文献1にその詳細が説明されているように、各索引語が出現する文書や文書内での出現位置情報を転置ファイルと呼ばれる表構造の索引に格納し、検索時には、この転置ファイルのみにアクセスして、単語を検索する方法である。図6は、転置ファイル法を用いた企業内検索向け検索システムの代表的な構成例を示している。

[0003] 同図に示した検索システムでは、検索対象となるのがストレッジを持ったSMBサーバ群であり、ここに各種の電子文章がプレーンテキスト、HTML、Word、Excel、Powerpoint、PDFなどオフィス系のフォーマットに格納される。

[0004] ここで、SMBサーバとは、マイクロソフト社がファイルアクセスのためのプロトコルとして採用しているSMB(あるいはCIFS)プロトコルによりアクセス可能なファイルサーバである。クライアントPCは、組織内の各自が使用するPC、あるいは、特定アプリケーションを実行させるPCである。

[0005] インデクサ用PCは、転置ファイル法を実現する上で検索対象ファイルのデータからインバーテッドインデクス(以下インデクス)を生成したり、検索対象ファイルのデータの更新(新規作成, 部分的な変更, 削除)に合わせてインデクスを更新する。

[0006] 対象となるファイルが大規模な場合には、インデクスの生成や更新を分担して行うために複数のインデクサ用PCが用意される。検索用PCは、インデクサ用PCが生成したインデクスを用いて全文検索サービスを提供する。クライアントPCから発行された検索条件に対して検索した結果を返す。

[0007] インデクサ用PCでは、図7に示す2種類のプログラムが動作する。ファイルサーバ用クローラとインデクサである。

[0008] ファイルサーバ用クローラが、検索対象ファイルを格納するSMBサーバ群の全ファイルを最初全て収集し、インデクサの入力としインデクスを作成する(インデクサ全更新機能)。図7の左側に、1つのSMBサーバのファイル構造の詳細が示されている。

[0009] その際に、収集したファイルの一覧をファイルの更新時間(タイムスタンプ)と合わせてタイムスタンプリストとして保存する。図7では、2007年3月20日23時30分00秒に作成したタイムスタンプリストを同図(b)に一覧表として示している。

[0010] 2回目以降のクローリングでは、ファイルサーバ用クローラは、SMBサーバ内の全ファイルについてタイムスタンプを確認し、新しいタイムスタンプリストを作成する。図7では、前回から1日経過した2007年3月21日23時30分00秒に作成したタイムスタンプリストを同図(c)に一覧表として示している。

[0011] 次に、今回の収集により作成したリスト(c)と、前回収集したリスト(b)と照らし合わせる。照らし合わせた結果、新たに追加されたファイル、更新されたファイル、削除されたファイルを見つけ出し、差分リスト(d)を作成する。

[0012] 差分リスト(d)には、変化のあったファイルについて、それが新たに追加されたファイル(New)なのか、更新されたファイル(Update)なのか、削除されたファイル(Delete)なのかを示す印とともに、フアイルのパスとタイムスタンプが列挙される。

[0013] その結果「新たに追加されたファイル」と「更新されたファイル」は、クローラによりファイルの本体を収集する。収集したファイルと差分リストを用いて、それまでに使われてきたインデクスを更新する。

[0014] しかしながら、このような従来の全文検索システムには、特に、インデクスを更新する際に、以下に説明する技術的な課題があった。

[0015] 特許文献1:北研二 他著、「情報検索技術」2005年11月20日北立出版発行 pp. 160－179 第6章6．2 「転置ファイルを用いた全文検索」

発明の開示

発明が解決しようとする課題

[0016] すなわち、図6に示した従来の全文検索システムでは、検索対象となるストレッジの大規模化に対応が困雄になるという問題と、クローリングに時間がかかるという問題があった。

[0017] その理由は、上述したシステムでは、インデクスを更新するために毎回全ての検索対象ファイル群をクローリングしてタイムスタンプを確認しなくてはならないため、この処理がSMBサーバの持つストレッジのサイズに比例して時間がかかり、非就業中の夜中のうちに昨日の変化分を処理できないことも起き得る。 ここで、クローリング時間を短縮するためには、複数のインデクサ用PCから異なるストレッジに分散して、複数スレッドにてクローリングすることで対処できるが、この方法では、本来のアプリケーションの処理に使われるべきストレッジの処理能力を、クローリングのために使ってしまうことになり、システム全体のパフォーマンスが低下することになる。

[0018] さらに、上記システムでは、全てのファイルをクローリングして初めて更新を行うことになるために、更新状況を即時にインデクスに反映することが困難になり、タイムリーな更新ができないという問題があった。

[0019] 本発明は、このような従来の問題点に鑑みてなされたものであって、その目的とするところは、検索対象ストレッジの大規模化に容易に対応することができるとともに、ファイルサーバの負荷を増加することなく、クローリング時間の短縮化して、短時間にインデクスに反映することができる全文検索システムを提供することにある。

課題を解決するための手段

[0020] 上記目的を達成するために、本発明は、検索対象となる電子文書が蓄積されたデータベースを有するファイルサーバ群と、前記電子文書の新規作成，削除，部分的な変更を入力するとともに、検索対象文書のインデクスを入力するクライアントPC群と

、前記ファイルサーバ群の前記電子文書を照査して、当該電子文書に含まれている文言の索引用インデクスを作成するとともに、作成したインデクスを格納するインデクスデータベースを有するインデクサ用PC群と、前記インデクサ用PC群で作成されたインデクスを格納する検索用PCとを有し、前記ファイルサーバ群，クライアントPC群，インデクサ用PC群，検索用PCの間のそれぞれがスイッチイングハブを介して接続される通信路とを備え、任意のクライアントPCから入力された検索対象インデクスに基づいて、前記検索用PCから前記検索対象インデクスに該当するインデクスを抽出して、当該クライアントPCに出力する全文検索システムにおいて、前記クライアントPC群から前記ファイルサーバ群へアクセスされる情報のうちから、前記前記電子文書の新規作成，削除，部分的な変更に関連する更新候補情報を抽出して、格納するキャプチャ用PCを備え、前記インデクサ用PC群は、前記インデクスを最初に作成する際にのみ、前記ファイルサーバ群を全体走査し、前記インデクスの新規作成，部分的な変更および削除に伴う更新は、前記キャプチャPCに格納されている前記更新候補情報に基づいて、前記ファイルサーバ群を照査して、新規作成，削除，部分的な変更が実際に行われた更新実行情報を抽出し、当該更新実行情報に基づいて前記インデクスの更新を行うようにした。

[0021] このように構成した全文検索システムでは、ファイルサーバ群（ストレッジ）へのアクセスをイベントとして捕らえる。本発明では、毎回全ファイルをクロール（走査）することはしないで、最初の1回だけ全ファイルをクロールしてインデクスを作成し、2回目からは、ファイルに対するアクセス（新規作成、部分的な更新、削除、アクセス権の変更）をイベントとして捕らえ、各イベントの対象となったファイルのみインデクス更新の対象とする。

[0022] このため、更新の度ごとに全ファイルをクロールする従来の方式に対して、クローリング時間の大幅な短縮化が可能になり、ストレッジの大規模化も簡単に対応することができ、しかも、このような効果は、アプリケーションの処理に使われるべきストレッジの処理能力を、クローリングのために消費することなく得られるので、システム全体のパフォーマンスの低下を回避することができ、さらには、更新状況を即時にインデクスに反映することが可能になり、タイムリーな更新をも可能にする。

[0023] 前記更新候補情報は、前記通信路から取得することができる。

[0024] また、前記更新候補情報は、前記クライアントPC群から前記ファイルサーバ群へアクセスされた情報が、前記ファイルサーバ群に残されているアクセルログから取得することができる。

[0025] 前記キャプチャ用PCは、前記通信路に設置されたタップを介して、前記クライアントPC群から前記ファイルサーバ群へアクセスされる情報を取得して、前記更新候補情報を抽出するパケット判別モジュールと、前記更新候補情報を格納するパケトログと、前記判別モジュールの制御用ファイルとを有することができる。

[0026] 前記インデクスには、前記検索対象となる電子文書へのアクセス権の作成、および、同アクセス権の新規作成，削除，部分的な変更などの更新情報を含ませることができる。

発明の効果

[0027] 本発明に係る全文検索システムによれば、クローリング時間の大幅な短縮化が可能になり、ストレッジの大規模化も簡単に達成することができ、しかも、このような効果は、アプリケーションの処理に使われるべきストレッジの処理能力を、クローリングのために消費することなく得られるので、システム全体のパフォーマンスの低下を回避することができ、さらには、更新状況を即時にインデクスに反映することが可能になり、タイムリーな更新をも可能にする。

図面の簡単な説明

[0028] [図1]本発明にかかる全文検索システムの全体構成を示すブロック図である。

[図2]図1に示したシステムで、更新候補情報を取得・格納する際に手順の一例を示すフローチャート図である。

[図3]図1に示したシステムで、取得した更新候補情報に基づいて、実際に更新されたファイルを確認し、その後インデクス更新する手順を示したフローチャート図である。

[図4]本発明にかかる全文検索システムの他の実施例を示す全体構成を示すブロック図である。

[図5]図4に示したシステムで、更新候補情報を取得・格納する際に手順の一例を示

すフローチャート図である。

[図6]従来の全文検索システムの全体構成を示すブロック図である。

[図7]図4に示したシステムでインデクスを更新する際の説明図である。

符号の説明

[0029]

1	SMBサーバ
2	クライアントPC
3	インデクサPC
4	検索用PC
5	スイッチングハブ
6	通信路
20	キャプチャ用PC
20c	パケットフィルタ
20d	パケット判別モジュール
20e	キャプチャ制御ファイル
20f	メモリ領域
20g	パケットログ
21	タップ

発明を実施するための最良の形態

[0030] 以下に、本発明を実施するための最良の形態を実施例に基づいて説明する。図1から図3は、本発明に係る全文検索システムの一実施例を示している。図1は、システムの全体構成図である。本実施例の全文検索システムは、SMBサーバ1と、クライアントPC2と、インデクサ用PC3と、検索用PC4と、スイッチングハブ5を備えた通信路6を備えている。

[0031] SMBサーバ1は、複数がスイッチングハブ5に並列接続されていて、これらがファイルサーバ群を構成している。各SMBサーバ1には、検索対象となる電子文書が蓄積されたデータベース7と、自サイトにアクセスされた文書の名前や時間，数などを記憶しておくアクセルログ8を有している。具体的には、ストレッジを持ったWindows（登録商標） ServerやSambaサーバ、あるいはNAS（Network Attached Storage）

がこれに相当する。

[0032] クライアントPC2は、複数がスイッチングハブ5に並列接続されていて、これらがクライアントPC群を構成しており、SMBサーバ1に格納する文書の新規作成，削除，部分的な変更を入力し、SMBサーバ1上のデータを操作するアプリケーションプログラムを実行するとともに、検索対象文書のインデクスが入力される。

[0033] インデクサ用PC3は、複数がスイッチングハブ5に並列接続されていて、これらがインデクサ用PC群を構成している。各インデクサ用PC3には、インデクスを格納するインデクスDB9を有している。

[0034] インデクサ用PC3は、ファイルサーバ群（SMBサーバ1群）の電子文書（デレクトリとファイル）を巡回走査して、検索対象となるファイルの属性（更新日時、アクセス権、ファイルの種別）を読み出す。ファイルの属性からインデクスの作成に必要なファイルを読み出す。これらを用いて検索用のインデクスを作成する。

[0035] 検索用PC4は、各インデクサ用PC3で作成されたインデクスのコピーを格納するデータベース10を有しており、これを用いて検索処理機能を提供する。インデクスのコピーは、インデクサ用PC3が新規の更新を終わるごとに作られ、それまで使っていた古いインデクスと交換される。

[0036] スイッチングハブ5を備えた通信路6は、ギガビットイーサーなどのPC間通信路であって、例えば、企業などで採用されているイントラネットがこれに相当する。このような検索システムでは、任意のクライアントPC2から入力された検索対象インデクスに基づいて、検索用PC4から検索対象インデクスに該当するインデクスを抽出して、当該クライアントPC2に出力することになる。

[0037] 以上のような全文検索システムとしての基本的な構成は、前述した従来のシステムと相違はないが、本実施例の全文検索システムは、以下に説明する点に顕著な特徴がある。

[0038] すなわち、本実施例の場合には、クライアントPC2群からSMBサーバ1群（ファイルサーバ群）へアクセスされる情報のうちから、電子文書の新規作成，削除，部分的な変更に関連する更新候補情報を抽出して、格納するキャプチャ用PC20を備えている。

[0039] キャプチャ用PC20は、タップ21を介して通信路6に接続されている。タップ21は、通信路6に影響を与えることなく、通信路6を通過するパケットを取り出すための機器であり、例えば、シスコムethergazer1000(市販商品名)を用いることができる。

[0040] また、キャプチャ用PC20には、ネットワークインターフェイス20a, トランスポート層20b, パケットフィルタ20c, パケット判別モジュール20d, キャプチャ制御用ファイル20e, メモリ領域20f, パケットログ20gを備えている。

[0041] このように構成されたキャプチャ用PC20では、タップ21で取り出した通信パケットを解析し、SMBサーバ1上のファイルに対する操作(その中でも検索用インデクスの更新に関連するもの)履歴をアクセスパケットとしてパケットログ20gに出力する。 タップ21と繋がるポートは、パケットの受信専用であるが、図1の右側のスイッチングハブ5に繋がるポートには、IPアドレスを割り当て、インデクサ用PC3と通信可能とする。より具体的には、スイッチングハブ5とキャプチャ用PC20のオペレーティングシステム20hとを介して、パケットログ20gにアクセス可能になっていて、この機能が実現される。 これにより、アクセスパケットログ20gは、インデクサ用PC3からこの通信経路を介して、参照できるようになっている。キャプチャ用PC20には、キャプチャ(抽出)しないIPアドレスの一覧を予め設定しておくことができる。

[0042] キャプチャするのは、クライアントPC2からSMBサーバ1上のファイル操作を行っているパケットに限られる。従って、インデクサ用PC3によるクローリング(走査)のためのパケットなどを抽出することがないようにしておく必要があるので、キャプチャの対象外とすべきIPアドレスを設定できるようにしておく。

[0043] ここで、キャプチャ対象とすべきパケットの情報は、ログとしてファイルに出力される前にまずメモリ(メモリ領域20f)上に格納され、複数まとめてファイル出力するようになっている。

[0044] これは、ファイルアクセスの性能を向上させる目的と、メモリ上で「同一ファイルやディレクトリに対する同一アクセスのログを一つにまとめる」処理を行うためである。この際にメモリ上に格納するログ数の上限を、実際に使うサーバのメモリ量に合せて設定できるようにする。

[0045] ログファイル(パケットログ20f)に出力する間隔の最長時間は、以下のように設定す

る。メモリ上に格納されたログが、ある一定時間たった場合、「ログ数上限」に達しなくても強制的にログファイルに吐き出させるため、出力する間隔の最長時間を設定する。

[0046] パケット判別モジュール20dでは、電子文書の新規作成，削除，部分的な変更に関連する更新候補情報の抽出が、以下の記号とフォーマットによりどのアクセスが行われたかが判別される。

T(パス名)←タッチ:そのファイルやディレクトリに触れた

D(パス名)← 削除:そのファイルやディレクトリを削除した

R(パス名，old)(パス名，new)← ファイル名やディレクトリ名を変更した

M(パス名，old)(パス名，new)← ファイルやデイレクトを移動した

[0047] すなわち、本実施例の場合、電子文書の新規作成，削除，部分的な変更に関連する更新候補情報として、T(タッチ)，D(削除)，R(old，new)，M(old，new)が抽出されることになる。

[0048] インデクサ用PC3は、インデクスを最初に作成する際にのみ、SMBサーバ1群(ファイルサーバ群)を全体走査する。すなわち、一番最初にインデクスを作成するときには、検索対象となる全てのファイルを基点リスト(「ファイル巡回」&「インデクス」制御ファイル内で定義)からたどり、ファイル属性を読み出しタイムスダンプリストをつくる。

またその際に列挙した全ての検索対象ファイルをインデクサ用PCに読み出す。これらを用いてインデクスを作成する。なお、インデクサ用PC3は、負荷分散によるインデクス作成／更新時間の短縮のために複数の存在が可能であり、各インデクサPC3に走査対象となるSMBサーバ1を割り当てればよい。

[0049] また、「ファイル巡回」&「インデクス」制御ファイルには、インデクサ用PC3における処理の設定を可能にするために、以下の内容を含む。

・基点リスト ツリー型のディレクトリ構造を持ったファイル

システム内を巡回する基点となるパス名の列

・検索非対称パス列 基点リストからツリーを巡る際に「その部分だけは検索対象にしない」パス列

・検索対象サフィックス txt，htm，html，pdfなど検索対象とするファイルのサフィク

ス列

・ログ確認周期

キャプチャ用PC20の出力する「アクセスパケットログ」を確認する周期

[0050] 一方、インデクサ用PC3は、インデクスの新規作成，部分的な変更および削除に伴う更新は、キャプチャPC20に格納されている更新候補情報に基づいて、SMBサーバ1群を照査して、新規作成，削除，部分的な変更が実際に行われた更新実行情報に基づいてインデクスの更新を行う。

[0051] 以上のより具体的な手順について、図2および3を参照にして説明する。図2は、キャプチャ用PC20による更新候補情報を取得する手順が示されている。 クライアントPC2において、アプリケーション（ワープロ、スプレッドシート、等々）からネットワークファイルアクセスが発生した場合に、まずファイル共有サービス層（Windows（登録商標） システムの場合にはSMBあるいはCIFSと呼ばれるプロトコルが使われる）を経由し、ネットワークの通信プロトコルをサポートするトランスポート層を経由しTCP／IPの規約に従ったパケットが作られ、ネットワークインターフェイス20aで物理的なネットワークアクセスがなされる。

[0052] SMBサーバ1は、この逆の順序で各層にて処理が行われファイル共有サービスがSMBサーバ1に接続されているディスクに対応する処理を行う。このようなネットワーク経由のファイルアクセスが行われている状況において、キャプチャ用PC20は、タップ21を使い、ネットワーク上を流れるパケットをキャプチャする。

[0053] キャプチャ用PC20の内部においてネットワークインターフェイス層およびトランスポート層は、オペレーティングシステム（Windows（登録商標）サーバOSなど）の機能が使われる。そこからパケットを選別して取組むためのパケットフィルタ20cとしては、1ibcapのような標準品を用いる。パケットフィルタ20cにおいて、SMB／CIFSに関連するパケットに絞って取り込むためにTCPポート445番だけを取り出すように設定する。

[0054] これを図2に基づいて具体的に説明すると、手順がスタートすると、s1で、パケットフィルタ20cを用い、ネットワーク（通信路6）からパケットが抽出される。

[0055] 次いで、s2で、抽出したパケットのIPアドレスが、インデクサ用PC3のものか、否か

判断され、インデクサ用PC3のものでなければ、s3に移行する。s3では、抽出したパケットの処理がSMB／CIFSに関連するものか、否かが判断され、これがSMB／CIFSに関連するものであれば、s4が実行される。

[0056] s4では、パケット判断モジュール20dにより、SMBプロトコルとして、前述したT(タッチ)，D(削除)，R(変更)，M(移動)であれば、これらが抽出され更新候補情報として、処理を表す記号と処理対象(ファイル，フォルダ)をメモリ領域20fに記録する。

[0057] 続くs5では、メモリ領域20fで、同一の対象に同一の処理を施すものは、複数残さず、単一にする処理が行われる。次のs6では、メモリ領域20f上の記録が、メモリ上に格納するログ数の条件を越えたか否かが判断され、これが超えている場合には、s7で、メモリ領域20f上の全記録をパケットログ20gに出力して、ここに格納する。

[0058] s6で、メモリ領域20f上の記録が、メモリ上に格納するログ数の条件を越えていないと判断された場合には、s8で、メモリ領域20f上の記録が1つ以上あって、前回ログ出力して以来「パケットログ20gに出力する間隔時間が最長時間を超えたか否かが判断され、越えている場合には、s7に移行して、メモリ領域20f上の全記録をパケットログ20gに出力して、ここに格納する。

[0059] 一方、s8で、バケットログ20gに出力する間隔時間が最長時間を超えていないと判断されると、s1に戻って処理が続行される。なお、s2で、抽出したパケットのIPアドレスが、インデクサ用PC3のものであると判断された場合と、s3で、抽出したパケットの処理がSMB／CIFSに関連するものでないと判断された場合には、s8に移行する。

[0060] 以上のような手順を順次実行することにより、キャプチャPC20のパケットログ20gには、更新候補情報として、T(タッチ)，D(削除)，R(変更)，M(移動)と処理対象(ファイル，フォルダ)とが蓄積されることになる。 このようにしてパケットログ20gに蓄積された更新候補情報は、インデクサ用PC3から定期的ないしは不定期に読まれて、インデクスの更新に使用される。この際には、インデクスの新規作成，部分的な変更および削除に伴う更新は、更新候補情報に基づいて、SMBサーバ1群を照査して、新規作成，削除，部分的な変更が実際に行われた更新実行情報を抽出し、この更新実行情報に基づいてインデクスの更新が行われる。

[0061] この際に行われる具体的な手順を図3に示している。なお、図3に示した手順では、

初回にインデクサPC3で全SMBサーバ1を走査して、インデクスを作成するステップを含んでいる。

[0062] 同図に示した手順がスタートすると、まず、s10で、検索対象とする全ファイル(全SMBサーバ1群のファイル)を「ファイル巡回」で取得し、インデクスを作成する(最初ないしは初回だけ)。作成したインデクスは、データベース9に格納する。同時にタイムスタンプリストを作成する。

[0063] 続くs11では、ファイルアクセルパケツトログ20gが新規に追加されたか否かが判断され、新規に追加されたと判断された場合には、s12に移行する。s12では、新規に追加されたアクセスログの内容を参照して、対象となるファイル, フォルダおよびフォルダ下の全ファイルを巡回し、変化があった対象を差分リストとして出力する。また、タイムスタンプリストをこの差分リストで更新する。

[0064] 続くs13では、巡回の結果、実際に、新規追加, 削除および変更のあったファイル情報を取得し、差分リストと合わせて実際に新規追加, 削除および変更のあった更新実行情報に基づいて、インデクスを更新する。次のs14では、一定時間待機して、s11に戻ることになる。また、s11で、パケットログに新規追加がないと判断された場合も、s14の一定時間の待機を経てs11に戻ることになる。

[0065] さて、以上のように構成した全文検索システムによれば、クライアントPC2群からSMBサーバ1群(ファイルサーバ群)へアクセスされる情報のうちから、電子文書の新規作成, 削除, 部分的な変更に関連する更新候補情報を抽出して、格納するキャプチャ用PC20を備え、インデクサ用PC3群は、インデクスを最初に作成する際にのみ、SMBサーバ1群(ファイルサーバ群)を全体走査し、インデクスの新規作成, 部分的な変更および削除に伴う更新は、キャプチャPC20に格納されている更新候補情報に基づいて、SMBサーバ1群(ファイルサーバ群)を照査して、新規作成, 削除, 部分的な変更が実際に行われた更新実行情報を抽出し、当該更新実行情報に基づいてインデクスの更新を行うようにした。

[0066] このため、更新の度ごとに全ファイルをクロールする従来の方式に対して、クローリング時間の大幅な短縮化が可能になり、ストレッジの大規模化も簡単に達成することができ、しかも、このような効果は、アプリケーションの処理に使われるべきストレッジ

の処理能力を、クローリングのために消費することなく得られるので、システム全体のパフォーマンスの低下を回避することができ、さらには、更新状況を即時にインデクスに反映することが困難になり、タイムリーな更新をも可能にする。

[0067] なお、上記実施例で示したインデクスの作成および更新には、電子文書だけでなく、検索対象となる電子文書へのアクセス権の作成、および、同アクセス権の新規作成，削除，部分的な変更などの更新情報を含ませることができる。

[0068] また、上記実施例で示した更新候補情報は、通信路6からタップ21を介して取得すること以外に、SMBサーバ1のログ出力（図1に示したアクセスログ8）から抽出することができる。 図4，5は、SMBサーバ1のログ出力から更新候補情報を抽出する場合の実施例を示しており、この実施例においては、上記実施例と同一もしくは相当する部分に同一符号を付して、その説明を省略するとともに、以下にその特徴点についてのみ説明する。

[0069] 図4は、この実施例に係る全文検索システムの全体構成を示しており、この実施例では、クライアントPC2群からSMBサーバ1群（ファイルサーバ群）へアクセスされる情報のうちから、電子文書の新規作成，削除，部分的な変更に関連する更新候補情報を抽出して、格納するキャプチャ用PC20に替えて、ほぼ同一機能を有するログ収集用PC30を備えている。

[0070] ログ収集用PC30は、SMBサーバ1群が並列接続されたスイッチングハブ5に電気的に接続され、ログ収集ルーチン30aと、アクセルログフィルタ30bと、ログ判定モジュール30cと、メモリ領域30dと、パケットログファイル30eとを備えている。

[0071] このように構成されたログ収集用PC30では、ログ収集ルーチン30aで収集した情報は、アクセルログフィルタ30bとログ判別モジュール30cとにより判別されて、電子文書の新規作成，削除，部分的な変更に関連する更新候補情報の抽出が行われる。

[0072] ここで、キャプチャ対象とすべき情報は、ログとしてファイルに出力される前にまずメモリ（メモリ領域20f）上に格納され、複数まとめてファイル出力するようになっている。

[0073] これは、上記実施例で説明したように、ファイルアクセスの性能を向上させる目的と、メモリ上で「同一ファイルやディレクトリに対する同一アクセスのログを一つにまとめる

」処理を行うためである。この際にメモリ上に格納するログ数の上限を、実際に使うサーバのメモリ量に合せて設定できるようにする。

[0074] ログファイル(パケットログファイル30e)に出力する間隔の最長時間は、以下のように設定する。メモリ上に格納されたログが、ある一定時間たった場合、「ログ数上限」に達しなくても強制的にログファイルに吐き出させるため、出力する間隔の最長時間を設定する。

[0075] 本実施例においても、上記実施例と同様に、インデクサ用PC3は、インデクスを最初に作成する際にのみ、SMBサーバ1群(ファイルサーバ群)を全体走査し、一番最初にインデクスを作成するときには、検索対象となる全てのファイルを基点リスト(「ファイル巡回」&「インデクス」制御ファイル内で定義)からたどり、ファイル属性を読み出しタイムスダンプリストをつくる。

[0076] また、インデクサ用PC3は、上記実施例と同様に、インデクスの新規作成, 部分的な変更および削除に伴う更新は、ログ収集用PC30に格納されている更新候補情報に基づいて、SMBサーバ1群を照査して、新規作成, 削除, 部分的な変更が実際に行われた更新実行情報に基づいてインデクスの更新を行う。

[0077] 図5は、更新候補情報をログ出力から抽出して、蓄積する際のログ収集用PC30の手順を示している。図5の手順がスタートすると、s30で、ログ収集ルーチン30aにて、ネットワークを経由して、SMBサーバ1群のログ出力8が収集される。

[0078] 次いで、s31で、収集したログ情報のうち、SMBサーバ1に対する書込み, 更新, 作成, 属性変更に関するもののみが、アクセルログフィルタ30bにより抽出される。続く、s32では、ログ判別モジュール30cにより、メモリ領域30d上に、SMBサーバ1に対する同一の対象, 同一の処理を一つにまとめて書き出す処理が行われる。

[0079] 次のs33では、メモリ領域30d上の記録が、メモリ上に格納するログ数の条件を越えたか否かが判断され、これが超えている場合には、s34で、メモリ領域30d上の全記録をパケットログファイル30eに出力して、ここに格納して、s36に移行する。

[0080] s33で、メモリ領域30d上の記録が、メモリ上に格納するログ数の条件を越えていないと判断された場合には、s35で、前回ログ出力して以来パケットログファイル30eに出力する間隔時間が設定値を超えたか否かが判断され、越えている場合には、s34

に移行して、メモリ領域30d上の全記録をパケットログファイル30eに出力して、ここに格納して、s36に移行する。

[0081] 一方、s35で、パケットログファイル30eに出力する間隔時間が設定時間を超えていないと判断されると、s36により一定時間待機して、s30に戻って処理が続行される。

[0082] 以上のような手順を順次実行することにより、ログ収集用PC30のパケットログファイル30eには、更新候補情報が蓄積されることになる。このようにしてパケットログファイル30eに蓄積された更新候補情報は、インデクサ用PC3から定期的ないしは不定期に読まれて、インデクスの更新に使用される。

[0083] この際には、インデクスの新規作成，部分的な変更および削除に伴う更新は、更新候補情報に基づいて、SMBサーバ1群を照査して、新規作成，削除，部分的な変更が実際に行われた更新実行情報を抽出し、この更新実行情報に基づいてインデクスの更新が行われる。

[0084] この際に行われる具体的な手順は、上記実施例の図3に示したものと同一である。以上のように構成した実施例でも、上記実施例と同等の作用効果が得られる。

[0085] なお、最近のOSでは、セキュリティ管環や内部統制的な要請から、どのような処理が内部的に行われたかをログファイルとして残す機能を持つものが多い。このログファイルをもとに、その中からファイルのアクセスに関するものだけを選択することで上記実施例と同様にイベントを抽出することが可能となる。

[0086] また、NAS：Network Attached Storageなどの専用ストレッジそれ自身が一種のコンピュータであり、そこへのアクセスをログとして出力したり、内部アクセス状況をプログラムインターフェイスから取り出すことができるものがある。このようなNASを使う場合には、NAS自身の機能からイベントの抽出が可能となる。

[0087] SMBサーバ1がこのようにアクセスログを抽出する機能を持つ場合には、上記実施例で示したパケットをキャプチャする処理は不要となり、このようなアクセスログを用いて上記実施例と同様の処理が可能になる。

産業上の利用可能性

[0088] 本発明にかかる全文検索システムによれば、クローリング時間の大幅な短縮化が可能になり、ストレッジの大規模化も簡単に達成することができ、しかも、このような効果

は、アプリケーションの処理に使われるべきストレッジの処理能力を、クローリングのために消費することなく得られるので、システム全体のパフォーマンスの低下を回避することができ、さらには、更新状況を即時にインデクスに反映することが可能になり、タイムリーな更新をも可能にするので、全文検索の分野において有効に活用することができる。

請求の範囲

[1] 検索対象となる電子文書が蓄積されたデータベースを有するファイルサーバ群と、

前記電子文書の新規作成，削除，部分的な変更を入力するとともに、検索対象文書のインデクスを入力するクライアントPC群と、

前記ファイルサーバ群の前記電子文書を照査して、当該電子文書に含まれている文言の索引用インデクスを作成するとともに、作成したインデクスを格納するインデクスデータベースを有するインデクサ用PC群と、

前記インデクサ用PC群で作成されたインデクスを格納する検索用PCとを有し、

前記ファイルサーバ群，クライアントPC群，インデクサ用PC群，検索用PCの間のそれぞれがスイッチイングハブを介して接続される通信路とを備え、

任意のクライアントPCから入力された検索対象インデクスに基づいて、前記検索用PCから前記検索対象インデクスに該当するインデクスを抽出して、当該クライアントPCに出力する全文検索システムにおいて、

前記クライアントPC群から前記ファイルサーバ群へアクセスされる情報のうちから、前記前記電子文書の新規作成，削除，部分的な変更に関連する更新候補情報を抽出して、格納するキャプチャ用PCを備え、

前記インデクサ用PC群は、前記インデクスを最初に作成する際にのみ、前記ファイルサーバ群を全体走査し、

前記インデクスの新規作成，部分的な変更および削除に伴う更新は、前記キャプチャPCに格納されている前記更新候補情報に基づいて、前記ファイルサーバ群を照査して、新規作成，削除，部分的な変更が実際に行われた更新実行情報を抽出し、当該更新実行情報に基づいて前記インデクスの更新を行うこと特徴とする全文検索システム。

[2] 前記更新候補情報は、前記通信路から取得することを特徴とする請求項1記載の全文検索システム。

[3] 前記更新候補情報は、前記クライアントPC群から前記ファイルサーバ群へアクセスされた情報が、前記ファイルサーバ群に残されているアクセルログから取得することを特徴とする請求項1記載の全文検索システム。

[4] 前記キャプチャ用PCは、前記通信路に設置されたタップを介して、前記クライアントPC群から前記ファイルサーバ群へアクセスされる情報を取得して、前記更新候補情報を抽出するパケット判別モジュールと、

前記更新候補情報を格納するパケトログと、

前記判別モジュールの制御用ファイルとを有することを特徴とする請求項2記載の全文検索システム。

[5] 前記インデクスには、前記検索対象となる電子文書へのアクセス権の作成、および、同アクセス権の新規作成，削除，部分的な変更などの更新情報を含ませることを特徴とする請求項1～5のいずれか1項記載の全文検索システム。

補正された請求の範囲

[2008年10月7日（07.10.2008）国際事務局受理]

[1] （補正後）検索対象となる電子文書が蓄積されたデータベースを有するファイルサーバ群と、

前記電子文書の新規作成，削除，部分的な変更を入力するとともに、検索対象文書のインデクスを入力するクライアントPC群と、

前記ファイルサーバ群の前記電子文書を照査して、当該電子文書に含まれている文言の索引用インデクスを作成するとともに、作成したインデクスを格納するインデクスデータベースを有するインデクサ用PC群と、

前記インデクサ用PC群で作成されたインデクスを格納する検索用PCとを有し、

前記ファイルサーバ群，クライアントPC群，インデクサ用PC群，検索用PCの間のそれぞれがスイッチイングハブを介して接続される通信路とを備え、

任意のクライアントPCから入力された検索対象インデクスに基づいて、前記検索用PCから前記検索対象インデクスに該当するインデクスを抽出して、当該クライアントPCに出力する全文検索システムにおいて、

前記クライアントPC群から前記ファイルサーバ群へアクセスされる情報のうちから、前記前記電子文書の新規作成，削除，部分的な変更に関連する更新候補情報を抽出して、格納するキャプチャ用PCを備え、

前記インデクサ用PC群は、前記インデクスを最初に作成する際にのみ、前記ファイルサーバ群を全体走査し、

前記インデクスの新規作成，部分的な変更および削除に伴う更新は、前記キャプチャPCに格納されている前記更新候補情報に基づいて、前記ファイルサーバ群を照査して、新規作成，削除，部分的な変更が実際に行われた更新実行情報を抽出し、当該更新実行情報に基づいて前記インデクスの更新を行う全文検索システムであって、

前記更新候補情報は、前記通信路から取得することを特徴とする全文検索システム。

[2] （補正後） 検索対象となる電子文書が蓄積されたデータベースを有するファイルサーバ群と、

前記電子文書の新規作成，削除，部分的な変更を入力するとともに、検索対象文書のインデクスを入力するクライアントPC群と、

前記ファイルサーバ群の前記電子文書を照査して、当該電子文書に含まれている文言の索引用インデクスを作成するとともに、作成したインデクスを格納するインデクスデータベースを有するインデクサ用PC群と、

前記インデクサ用PC群で作成されたインデクスを格納する検索用PCとを有し、

前記ファイルサーバ群，クライアントPC群，インデクサ用PC群，検索用PCの間のそれぞれがスイッチイングハブを介して接続される通信路とを備え、

任意のクライアントPCから入力された検索対象インデクスに基づいて、前記検索用PCから前記検索対象インデクスに該当するインデクスを抽出して、当該クライアントPCに出力する全文検索システムにおいて、

前記クライアントPC群から前記ファイルサーバ群へアクセスされる情報のうちから、前記前記電子文書の新規作成，削除，部分的な変更に関連する更新候補情報を抽出して、格納するキャプチャ用PCを備え、

前記インデクサ用PC群は、前記インデクスを最初に作成する際にのみ、前記ファイルサーバ群を全体走査し、

前記インデクスの新規作成，部分的な変更および削除に伴う更新は、前記キャプチャPCに格納されている前記更新候補情報に基づいて、前記ファイルサーバ群を照査して、新規作成，削除，部分的な変更が実際に行われた更新実行情報を抽出し、当該更新実行情報に基づいて前記インデクスの更新を行う全文検索システムであって、

前記更新候補情報は、前記クライアントPC群から前記ファイルサーバ群へアクセスされた情報が、前記ファイルサーバ群に残されているアクセルログから取得することを特徴とする全文検索システム。

[3] （補正後） 前記キャプチャ用PCは、前記通信路に設置されたタップを介して、前記クライアントPC群から前記ファイルサーバ群へアクセスされる情報を取得して、前記更新候補情報を抽出するパケット判別モジュールと、

前記更新候補情報を格納するパケトログと、

前記判別モジュールの制御用ファイルとを有することを特徴とする請求項1記載の全文検索システム。

[4] （補正後） 前記インデクスには、前記検索対象となる電子文書へのアクセス権の作成、および、同アクセス権の新規作成，削除，部分的な変更などの更新情報を含ませることを特徴とする請求項1～3のいずれか1項記載の全文検索システム。

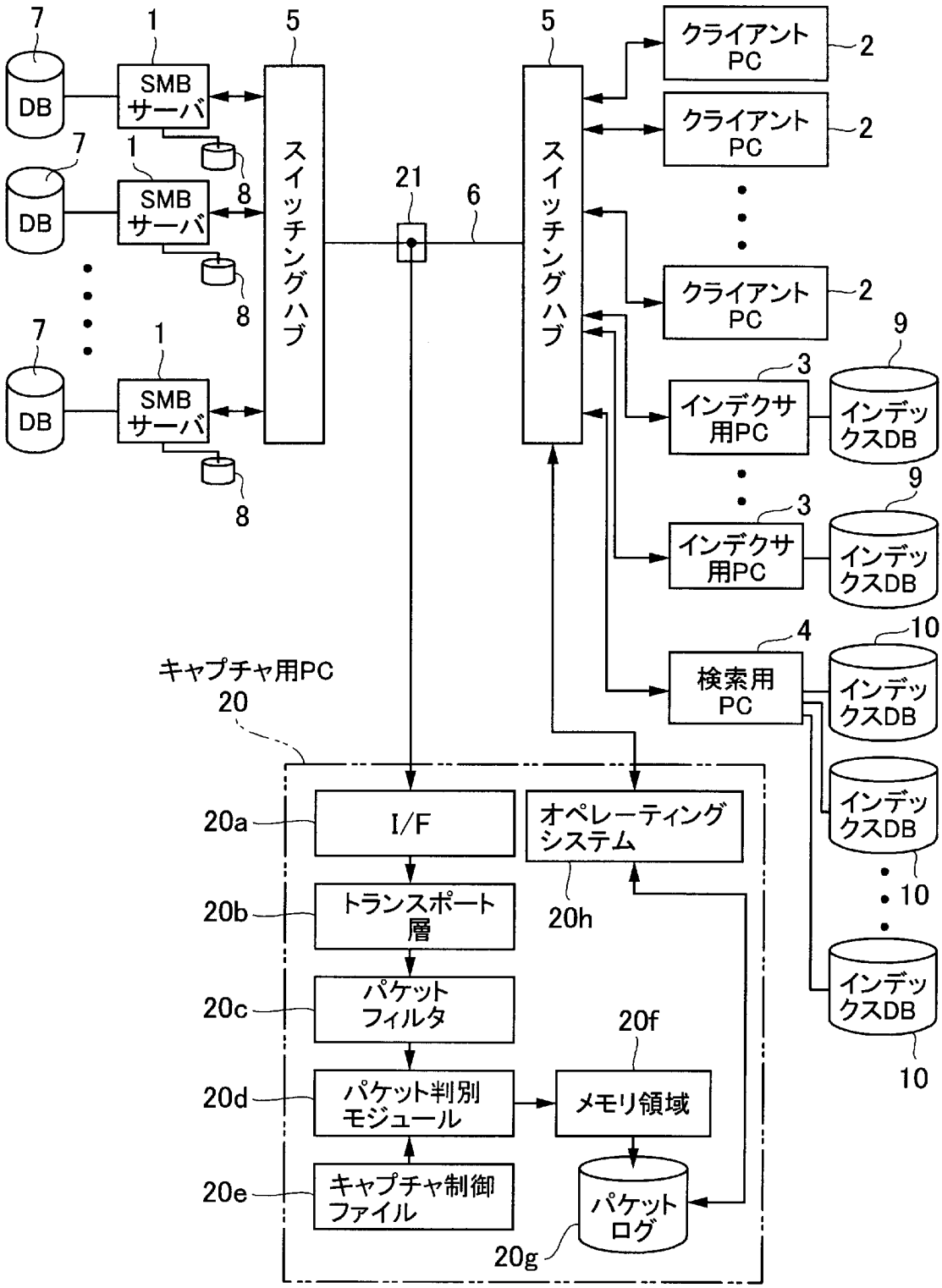
[5] （削除）

条約第19条(1)に基づく説明書

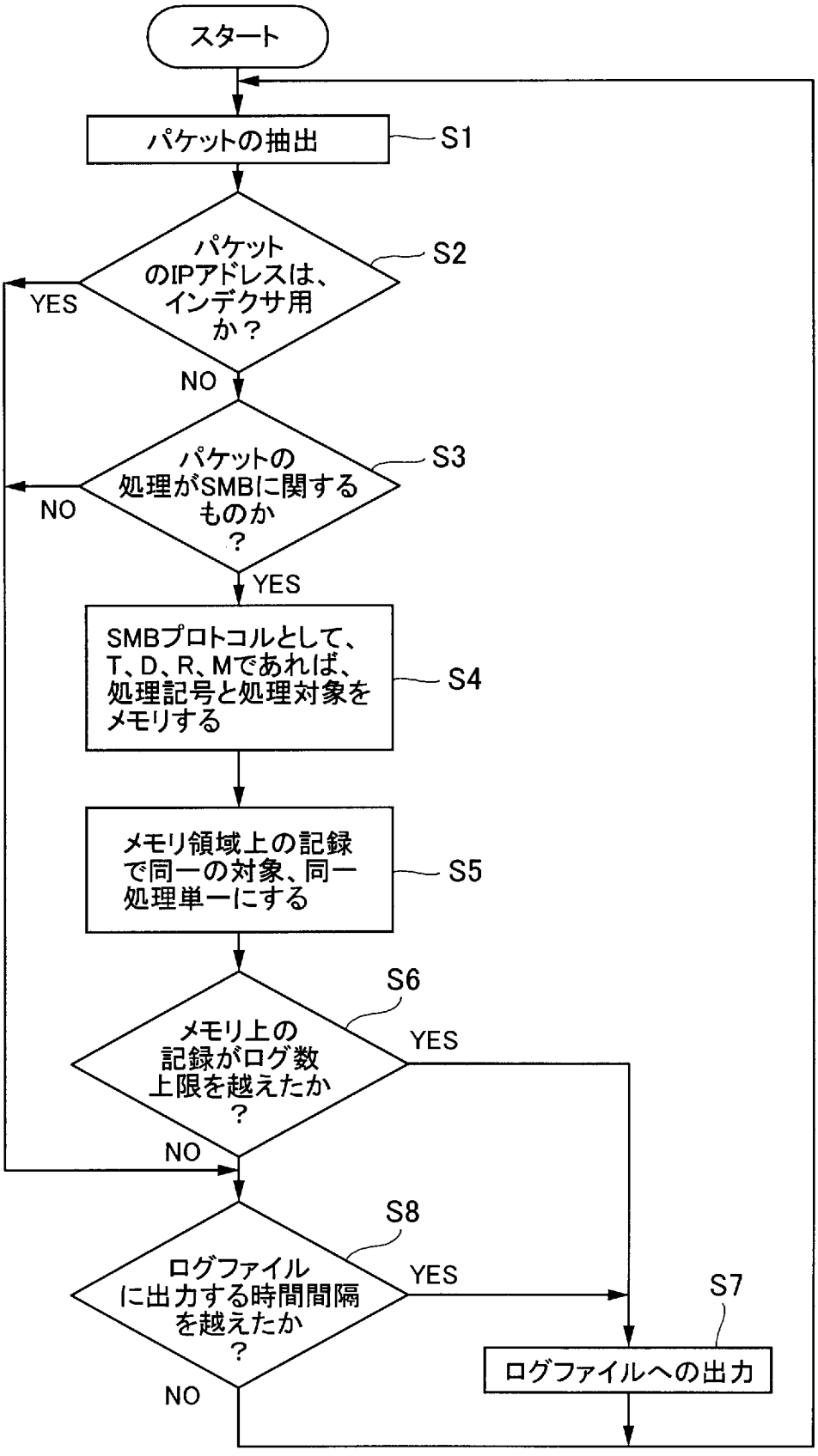
出願人は、請求項1～4を補正した。

請求の範囲第1項において、更新候補情報を通信路から取得することを明確にし、また、第2項において、更新候補情報は、クライアントPC群からファイルサーバ群へアクセスされた情報が、ファイルサーバ群に残されているアクセルログから取得することを明確にするものである。

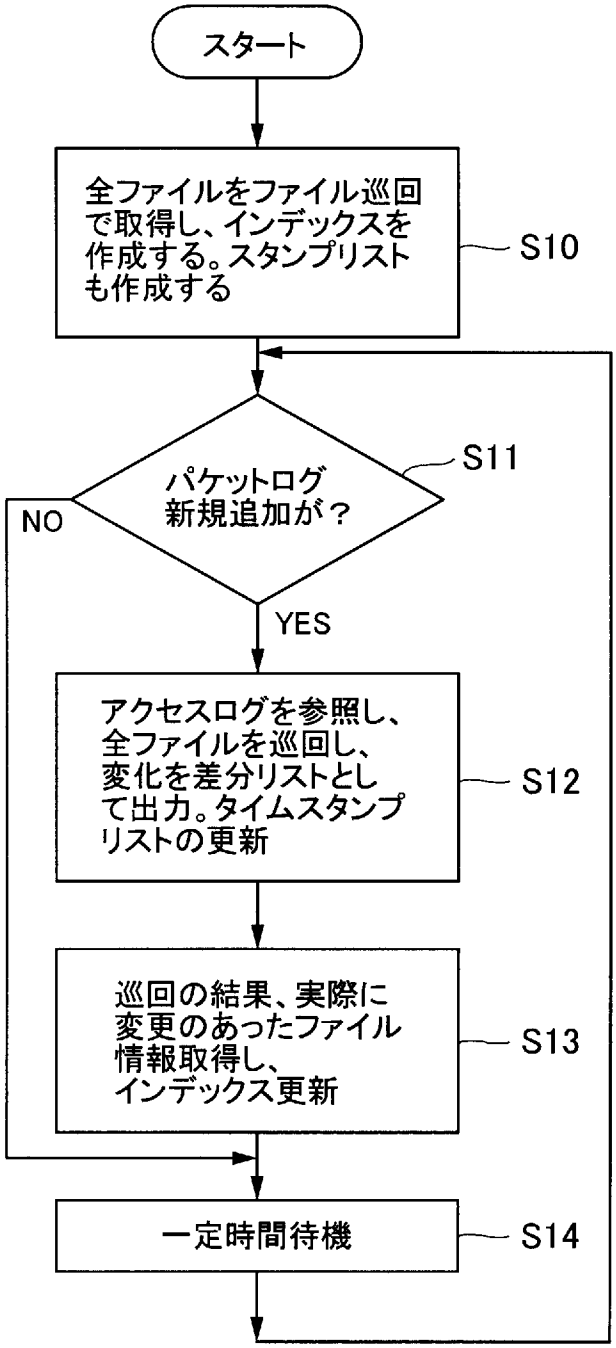
[図1]


7
1
5
DB
SMB
サーバ
8
スイッチングハブ
21
6
5
クライアント
PC
2
9
3
インデクサ
用PC
インデックス
DB
4
10
検索用
PC
キャプチャ用PC
20
20a
I/F
オペレーティング
システム
20h
20b
トランスポート
層
20c
パケット
フィルタ
20f
20d
パケット判別
モジュール
メモリ領域
20e
キャプチャ制御
ファイル
パケット
ログ
20g

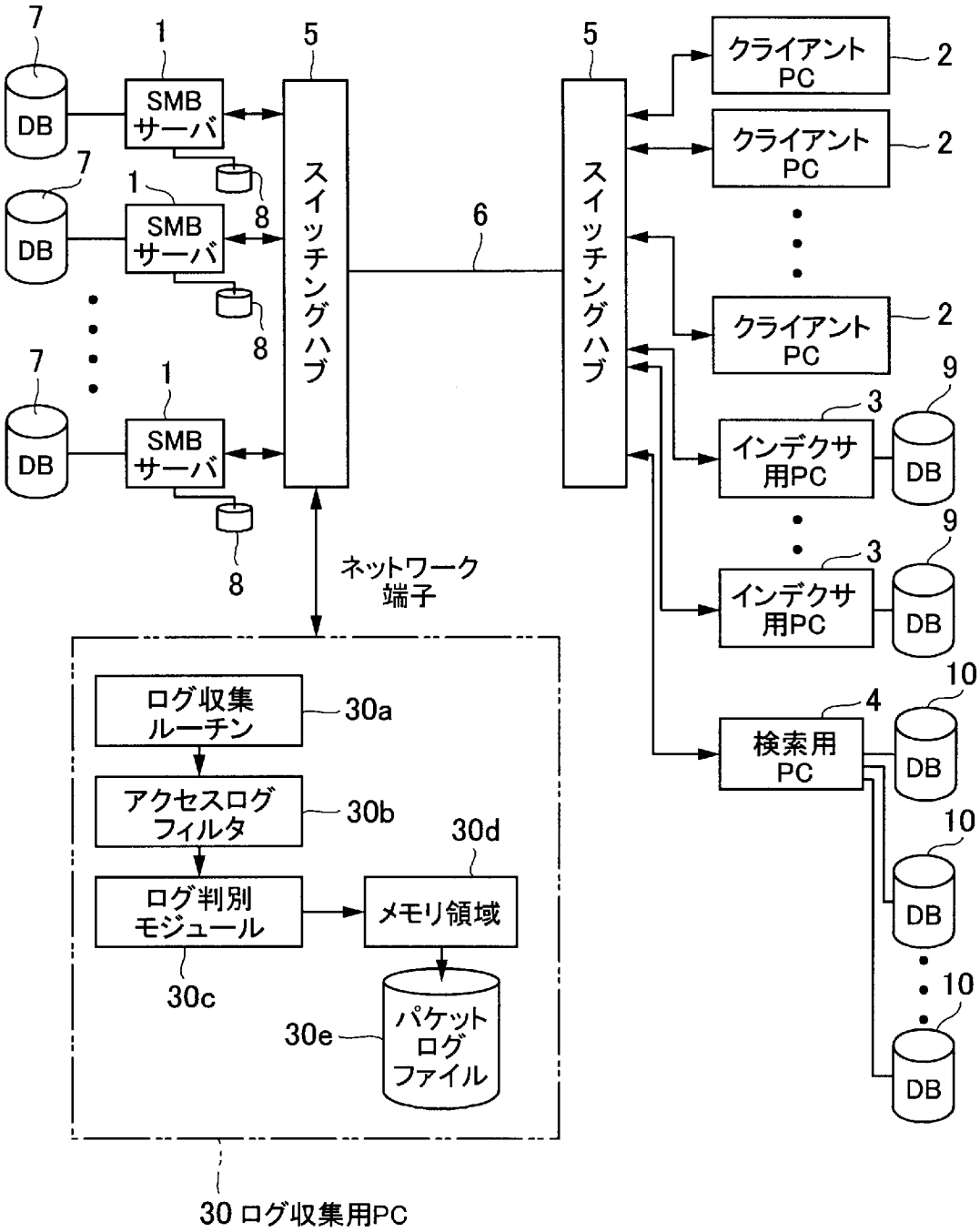
[図2]


スタート
パケットの抽出
S1
パケット
のIPアドレスは、
インデクサ用
か？
S2
YES
NO
パケットの
処理がSMBに関する
ものか
？
S3
NO
YES
SMBプロトコルとして、
T、D、R、Mであれば、
処理記号と処理対象を
メモリする
S4
メモリ領域上の記録
で同一の対象、同一
処理単一にする
S5
S6
メモリ上の
記録がログ数
上限を越えたか
？
YES
NO
S8
ログファイル
に出力する時間間隔
を越えたか
？
YES
S7
ログファイルへの出力
NO

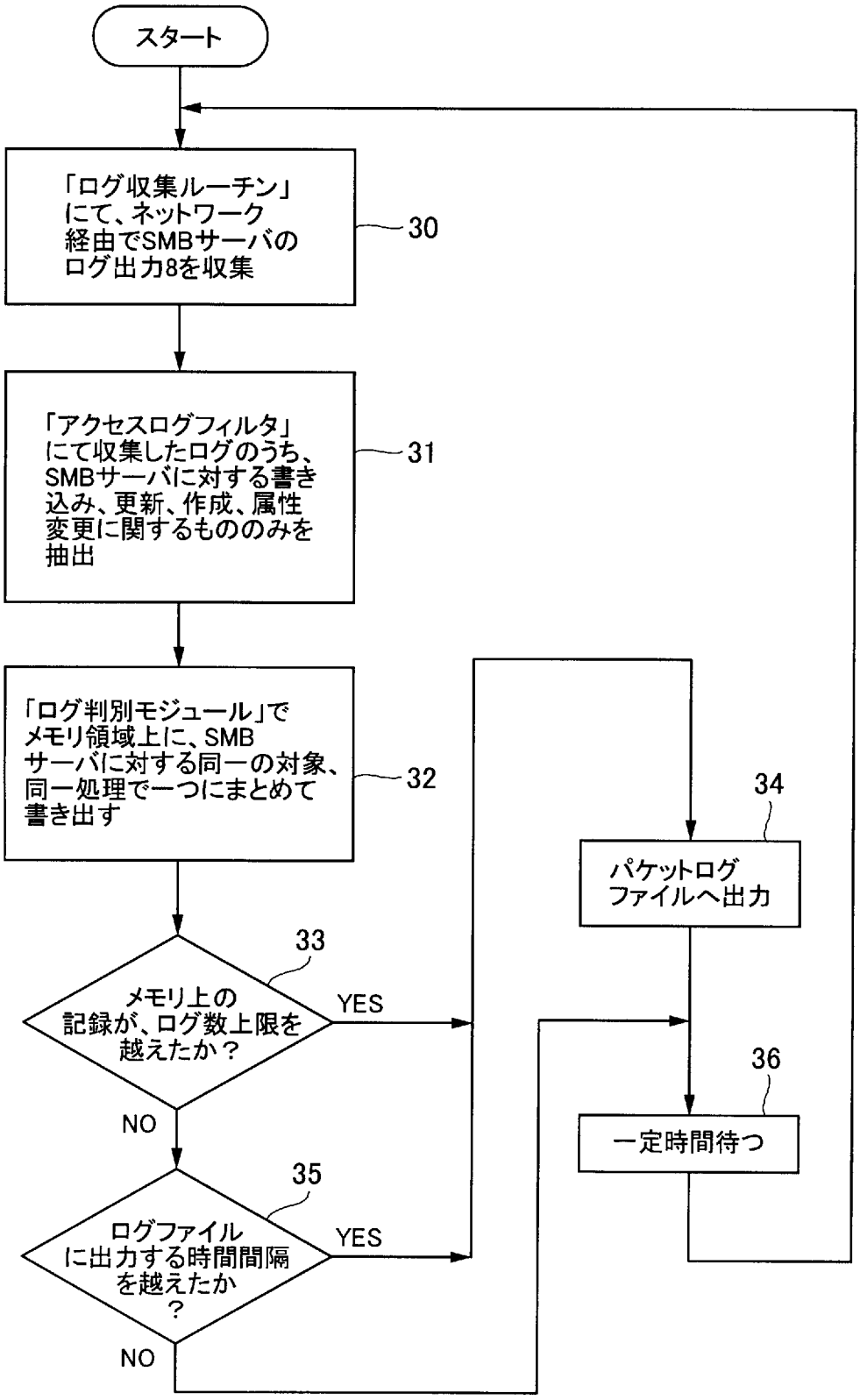
[図3]


スタート
全ファイルをファイル巡回で取得し、インデックスを作成する。スタンプリストも作成する
S10
パケットログ新規追加が？
S11
NO
YES
アクセスログを参照し、全ファイルを巡回し、変化を差分リストとして出力。タイムスタンプリストの更新
S12
巡回の結果、実際に変更のあったファイル情報取得し、インデックス更新
S13
一定時間待機
S14

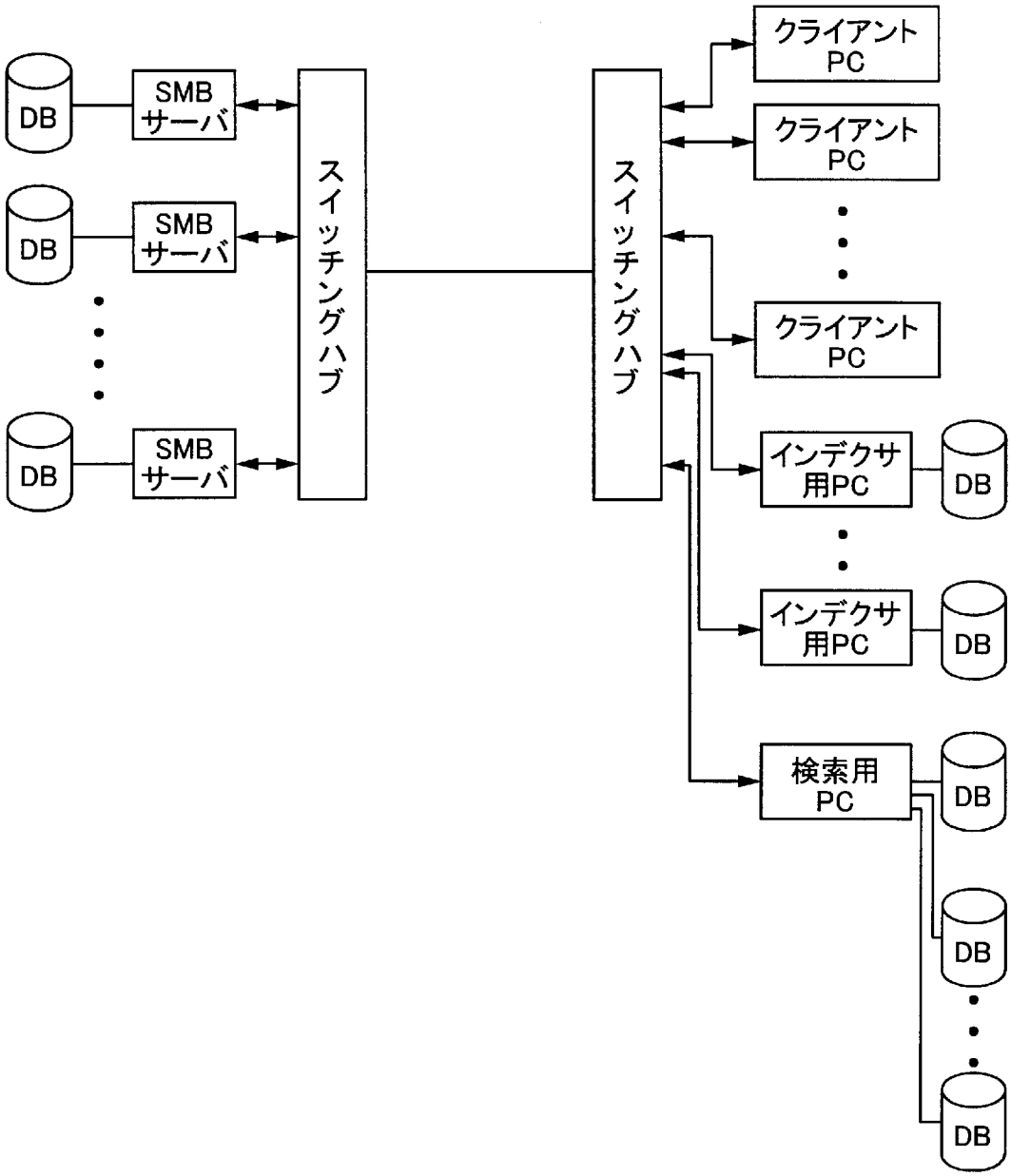
[図4]


7
DB
1
SMB
サーバ
8
5
スイッチング
ハブ
6
5
スイッチング
ハブ
クライアント
PC
2
3
インデクサ
用PC
9
4
検索用
PC
10
ネットワーク
端子
ログ収集
ルーチン
30a
アクセスログ
フィルタ
30b
ログ判別
モジュール
30c
30d
メモリ領域
30e
パケット
ログ
ファイル
30 ログ収集用PC

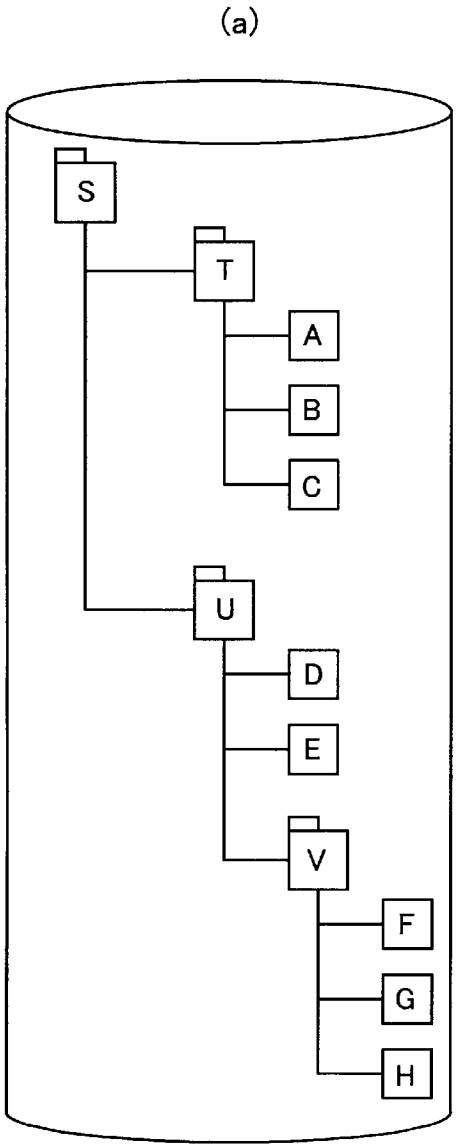
[図5]


スタート
「ログ収集ルーチン」にて、ネットワーク経由でSMBサーバのログ出力8を収集
30
「アクセスログフィルタ」にて収集したログのうち、SMBサーバに対する書き込み、更新、作成、属性変更に関するもののみを抽出
31
「ログ判別モジュール」でメモリ領域上に、SMBサーバに対する同一の対象、同一処理で一つにまとめて書き出す
32
33
メモリ上の記録が、ログ数上限を越えたか？
YES
NO
35
ログファイルに出力する時間間隔を越えたか？
YES
NO
34
パケットログファイルへ出力
36
一定時間待つ

[図6]


DB
SMB
サーバ
DB
SMB
サーバ
DB
SMB
サーバ
スイッチングハブ
スイッチングハブ
クライアント
PC
クライアント
PC
クライアント
PC
インデクサ
用PC
DB
インデクサ
用PC
DB
検索用
PC
DB
DB
DB

[図7]



前回インデクス時のタイムスタンプリスト

2007.3.20. 23:30:10

(b)

パス	更新日時
¥S¥T¥A	2007.3.20. 18:16:23
¥S¥T¥B	2007.3. 1. 10: 4:11
¥S¥T¥C	2007.1.10. 7:13:21
¥S¥U¥D	2007.3.20. 11:22:34
¥S¥U¥E	2007.3.20. 20:17:21
¥S¥U¥V¥F	2007.2. 3. 8:56:21
¥S¥U¥V¥G	2007.3.20. 17:41:26

今回のタイムスタンプリスト

(c)

¥S¥T¥A	2007.3.20. 18:16:23
¥S¥T¥B	2007.3.21. 12:24:11
¥S¥T¥C	2007.1.10. 7:13:21
¥S¥U¥D	2007.3.20. 11:22:34
¥S¥U¥E	2007.3.20. 20:17:21
¥S¥U¥V¥F	2007.3.21. 3:19:34
¥S¥U¥V¥G	2007.3.20. 17:41:26
¥S¥U¥V¥H	2007.3.21. 11:23:15

差分リスト

(d)

Update	¥S¥T¥B	2007.3.21. 12:24:11
Update	¥S¥U¥V¥F	2007.3.21. 3:17:34
NEW	¥S¥U¥V¥H	2007.3.21. 11:23:15

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2008/059128

A. CLASSIFICATION OF SUBJECT MATTER
G06F17/30(2006.01)i, *G06F12/00*(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)
G06F17/30, G06F12/00

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	1996-2008
Kokai Jitsuyo Shinan Koho	1971-2008	Toroku Jitsuyo Shinan Koho	1994-2008

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
JSTPlus(JDreamII), NRI Cyber Patent

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X Y A	Shankar Pasupathy et al., Making enterprise storage more search-friendly, Proc. of the 20th ACM symposium on Operating systems principles, ACM, 2005, [online] [from ACM-DL] [retrieval date 04 June, 2008 (04.06.08)]	1,3 5 2,4
Y	Tatsuya SHINDO, "Kigyo o Atsuku suru Saishin Technology Kigyonai Kensaku System Kakushu Server kara Data o Shushu Access-ken to Rendo shite Kensaku", Nikkei Communications, 01 July, 2006 (01.07.06), No.465, pages 146 to 150	5
Y	JP 2001-344245 A (Fujitsu Ltd.), 14 December, 2001 (14.12.01), All pages; Figs. 1 to 43 & US 2001/0027451 A1	5

☒ 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	Date of mailing of the international search report
04 June, 2008 (04.06.08)	17 June, 2008 (17.06.08)
Name and mailing address of the ISA/ Japanese Patent Office Facsimile No.	Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2008/059128

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	Yu TAKITA et al., "Zenbun Kensaku Engine o Riyo shita File System no Namae Kukan Kakucho", Transactions of Information Processing Society of Japan, 15 March, 2006 (15.03.06), Vol.47, No.SIG3(ACS13), pages 16 to 26	1-5
A	Mikio HIRABAYASHI, "Sakusha ga Kataru Hyper Estraier no Subete", Software Design, 18 March, 2007 (18.03.07), Vol.197, pages 88 to 98	1-5
A	Joji OTSUKA et al., "Center Kanrigata Fusei Access Kenchi System no Teian", IEICE Technical Report(ISEC2002-13 to 33), 18 July, 2002 (18.07.02), Vol.102, No.211, pages 39 to 44	1-5
A	Tatsuya SHINDO, "Zenbun Kensaku Gijutsu to Gyomu System eno Oyo", JAVA PRESS, 15 April, 2004 (15.04.04), Vol.35 , pages 125 to 133	1-5

国際調査報告

国際出願番号　ＰＣＴ／ＪＰ２００８／０５９１２８

Ａ．　発明の属する分野の分類（国際特許分類（ＩＰＣ））

Int.Cl.　G06F17/30(2006.01)i, G06F12/00(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.　G06F17/30, G06F12/00

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	１９７１－２００８年
日本国実用新案登録公報	１９９６－２００８年
日本国登録実用新案公報	１９９４－２００８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）

JSTPlus(JDreamII), NRI サイバーパテント

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	関連する請求の範囲の番号
X	Shankar Pasupathy、外２名, Making enterprise storage more search-friendly, Proc. of the 20th ACM symposium on Operating systems principles, ACM, 2005, [online][from ACM-DL][検索日 2008/6/4]	1,3
Y		5
A		2,4
Y	進藤達也, 企業を熱くする最新テクノロジ　企業内検索システム　各種サーバーからデータを収集　アクセス権と連動して検索, 日経コミュニケーション, 2006.07.01, 第465号, p.146-150	5

☑　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	国際調査報告の発送日
０４．０６．２００８	１７．０６．２００８

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）　５Ｍ　３２５２
鈴木　和樹
電話番号　０３－３５８１－１１０１　内線　３５９９

Ｃ（続き）．　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	関連する請求の範囲の番号
Y	JP 2001-344245 A (富士通株式会社) 2001.12.14, 全頁, 第1-43図 & US 2001/0027451 A1	5
A	滝田裕、外1名, 全文検索エンジンを利用したファイルシステムの名前空間拡張, 情報処理学会論文誌, 2006.03.15, 第47巻, 第SIG3(ACS13)号, p.16-26	1-5
A	平林幹雄, 作者が語るＨｙｐｅｒ　Ｅｓｔｒａｉｅｒのすべて, ＳｏｆｔｗａｒｅＤｅｓｉｇｎ, 2007.03.18, 第197号, p.88-98	1-5
A	大塚丈司、外2名, センター管理型不正アクセス検知システムの提案, 電子情報通信学会技術研究報告(ISEC2002-13～33), 2002.07.18, 第102巻, 第211号, p.39-44	1-5
A	進藤達也, 全文検索技術と業務システムへの応用, ＪＡＶＡ　ＰＲＥＳＳ, 2004.04.15, 第35巻, p.125-133	1-5