

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4903386号

(P4903386)

(45) 発行日 平成24年3月28日(2012.3.28)

(24) 登録日 平成24年1月13日(2012.1.13)

(51) Int.Cl.	F I
GO 6 F 21/22 (2006.01)	GO 6 F 21/22 1 1 6
GO 6 F 12/00 (2006.01)	GO 6 F 12/00 5 1 3 D
GO 6 F 13/00 (2006.01)	GO 6 F 13/00 5 4 O E

請求項の数 39 (全 33 頁)

(21) 出願番号	特願2004-568963 (P2004-568963)	(73) 特許権者	501113353
(86) (22) 出願日	平成15年9月17日(2003.9.17)		シマンテック コーポレーション
(65) 公表番号	特表2005-539334 (P2005-539334A)		Symantec Corporation
(43) 公表日	平成17年12月22日(2005.12.22)		アメリカ合衆国、カリフォルニア州 94
(86) 国際出願番号	PCT/US2003/030178		043, マウンテン ビュー, エリス ス
(87) 国際公開番号	W02004/027653		トリート 350
(87) 国際公開日	平成16年4月1日(2004.4.1)	(74) 代理人	100147485
審査請求日	平成18年7月13日(2006.7.13)		弁理士 杉村 憲司
(31) 優先権主張番号	10/247,002	(72) 発明者	ロウニー ケヴィン ティー
(32) 優先日	平成14年9月18日(2002.9.18)		アメリカ合衆国 カリフォルニア州 94
(33) 優先権主張国	米国 (US)		131 サン フランシスコ ダンカン
(31) 優先権主張番号	10/431,145		ストリート 748
(32) 優先日	平成15年5月7日(2003.5.7)		
(33) 優先権主張国	米国 (US)		

最終頁に続く

(54) 【発明の名称】 事前選択されたデータに関し探索可能な情報コンテンツ

(57) 【特許請求の範囲】

【請求項 1】

メッセージモニタリングシステムが、ネットワーク上で個々の宛先に到達するよう電子的に送信される複数のメッセージを、事前選択されたデータに関してモニターする段階と、

前記メッセージモニタリングシステムにより、前記複数のメッセージの内容探索を実行して、検索された当該複数のメッセージの一又は二以上が当該事前選択されたデータの少なくとも一部分を含んでいるか否かを決定する段階であって、複数の内容フラグメントを含むメッセージの内容探索を実行する段階は、

抽象的データ構造を探索して、前記内容フラグメントのハッシュ又は暗文に対応する前記事前選択されたデータの事前作成されたハッシュ又は暗文を含む抽象的データ構造内のエントリを識別する段階であって、前記抽象的データ構造は、表構造を有する前記事前選択されたデータのインデックスであり、前記抽象的データ構造のエントリは、前記表構造内のセルの内容の事前作成されたハッシュ又は暗文と前記セルの行番号とを含み、前記抽象的データ構造は前記事前選択されたデータのの前記表構造内のセルの内容を表さないものである識別する段階と、

前記内容フラグメントのグループを決定する段階であって、各グループは、同じ行番号を有する前記抽象的データ構造内のエントリの事前作成されたハッシュ又は暗文に対応するハッシュ又は暗文に関連付けられた内容フラグメントを含むものである決定する段階と、

10

20

何れかのグループが少なくとも一つの条件を満たすならば、前記メッセージが前記事前選択されたデータの少なくとも一部を含んでいると決定する段階とを含む決定する段階と、

前記メッセージモニタリングシステムにより、内容探索が実行されたメッセージが前記事前選択されたデータの少なくとも一部を含んでいると決定されたことに応じて、当該メッセージを前記個々のあて先に到達しないようにする段階とを含む方法。

【請求項 2】

前記抽象的データ構造は、前記事前選択されたデータのコピーを含まず、前記少なくとも一つの条件は、何れかのグループが、数量閾値に達する又は超える数量の内容フラグメントを含む場合に満たされる数量閾値であることを特徴とする、請求項 1 に記載の方法。

10

【請求項 3】

前記抽象的データ構造内の各エントリは、前記表構造内のセルの列番号及び前記表構造内のセルの列の型式を更に含むことを特徴とする、請求項 1 に記載の方法。

【請求項 4】

前記内容探索を実行する段階前に、データベースから抽出された前記事前選択されたデータに基づいて前記抽象的データ構造を作成する段階を更に備えていることを特徴とする、請求項 1 に記載の方法。

【請求項 5】

20

前記抽象的データ構造内の各エントリは、セルの列番号を更に含み、前記条件は、内容フラグメントに関する数量閾値であり、前記メッセージの内容探索を実行する段階は更に、

前記メッセージ中のテキストの個々のラインを解析して、前記メッセージを単語又は句である内容フラグメントで解析する段階と、

前記内容フラグメントの少なくとも一部分にハッシュ関数を適用して、前記内容フラグメントのハッシュを作成する段階と、

ハッシュ表衝突リストを各内容フラグメント用に作成する段階であって、前記ハッシュ表衝突リストは、前記抽象的データ構造内のエントリについて事前作成されたハッシュと内容フラグメントのハッシュとの間の一致のリストであり、前記グループのそれぞれが、同じ行番号を有する前記抽象的データ構造内のエントリを識別するハッシュ表衝突リストと関連付けられた内容フラグメントを有する作成する段階と、

30

何れかのグループが、異なる列番号を有する前記抽象的データ構造内のエントリを識別するハッシュ表衝突リストに関連付けられた内容フラグメントの数量閾値を少なくとも含むならば、前記メッセージは前記事前選択されたデータの少なくとも一部分を含んでいると決定する段階と

を含むことを特徴とする、請求項 1 に記載の方法。

【請求項 6】

前記メッセージの内容探索を実行する段階は更に、

前記抽象的データ構造を探索する前に、前記メッセージ内のテキストの個々のラインを解析して、前記メッセージを単語又は句である内容フラグメントで解析し、前記メッセージの中で、前記事前選択されたデータの一部を含んでいるかもしれない内容フラグメントのシーケンスを検出する段階と、

40

前記シーケンス内の内容フラグメントの少なくとも一部分にハッシュ関数又は暗号化関数を適用して、前記内容フラグメントのハッシュ又は暗文を作成する段階と、

前記抽象的データ構造を探索して、前記シーケンス内の内容フラグメントのハッシュ又は暗文に対応する前記事前選択されたデータの事前作成されたハッシュ又は暗文を含む前記抽象的データ構造内のエントリを識別する段階と

を含むことを特徴とする請求項 1 に記載の方法。

【請求項 7】

50

前記事前選択されたデータの一部を含んでいるかもしれない内容フラグメントのシーケンスを検出する前記段階は、

所定のフォーマットを定義付ける探索ルールを用いて、前記所定のフォーマットを有する表現に関する前記メッセージの本文を探索する段階であって、前記所定のフォーマットを有する前記表現は、口座番号、社会保障番号、クレジットカード番号、電話番号、郵便番号、eメールアドレス、現金量又は運転免許証番号の何れかである探索する段階と、

前記表現を取り囲んでいる領域は、前記事前選択されたデータの一部を含んでいるかもしれないと判定する段階であって、前記表現を取り囲んでいる前記領域は、前記表現の前後に隣接するデータフラグメントの量を含む判定する段階とを含むことを特徴とする、請求項6に記載の方法。

10

【請求項8】

ネットワーク上で個々の宛先に到達するよう電子的に送信される複数のメッセージを、埋め込まれている事前選択されたデータに関してモニターする手段と、

前記複数のメッセージの内容探索を実行し、検索された当該複数のメッセージの一又は二以上が当該事前選択されたデータの少なくとも一部分を含んでいるか否かを決定する手段であって、複数の内容フラグメントを含むメッセージの内容探索を実行する段階は、

抽象的データ構造を探索して、前記内容フラグメントのハッシュ又は暗文に対応する前記事前選択されたデータの事前作成されたハッシュ又は暗文を含む抽象的データ構造内のエントリを識別する段階であって、前記抽象的データ構造は、表構造を有する前記事前選択されたデータのインデックスであり、前記抽象的データ構造内のエントリは、前記表構造内のセルの内容の事前作成されたハッシュ又は暗文と前記セルの行番号とを含み、前記抽象的データ構造は前記事前選択されたデータのの前記表構造内のセルの内容を表さないものである識別する段階と、

20

前記内容フラグメントのグループを決定する段階であって、各グループは、同じ行番号を有する前記抽象的データ構造内のエントリの事前作成されたハッシュ又は暗文に一致するハッシュ又は暗文に関連付けられた内容フラグメントを含むものである決定する段階と、

何れかのグループが少なくとも一つの条件を満たすならば、前記メッセージが前記事前選択されたデータの少なくとも一部を含んでいると決定する段階とを含む決定する手段と、

30

内容探索が実行されたメッセージが前記事前選択されたデータの少なくとも一部を含んでいると決定されたことに応じて、当該メッセージを前記個々の宛て先に到達しないようにする手段とを備えていることを特徴とする装置。

【請求項9】

プロセッサ上で実行されると、前記プロセッサに、

ネットワーク上で個々の宛先に到達するよう電子的に送信される複数のメッセージを、事前選択されたデータに関してモニターする段階と、

前記複数のメッセージの内容探索を実行し、検索された当該複数のメッセージの一又は二以上が当該事前選択されたデータの少なくとも一部分を含んでいるか否かを決定する段階であって、複数の内容フラグメントを含むメッセージの内容探索を実行する段階は、

40

抽象的データ構造を探索して、前記内容フラグメントのハッシュ又は暗文に対応する前記事前選択されたデータの事前作成されたハッシュ又は暗文を含む抽象的データ構造内のエントリを識別する段階であって、前記抽象的データ構造は、表構造を有する前記事前選択されたデータのインデックスであり、前記抽象的データ構造のエントリは、前記表構造内のセルの内容の事前作成されたハッシュ又は暗文と前記セルの行番号とを含み、前記抽象的データ構造は前記事前選択されたデータのの前記表構造内のセルの内容を表さないものである識別する段階と、

前記内容フラグメントのグループを決定する段階であって、各グループは、同じ行番号を有する前記抽象的データ構造内のエントリの事前作成されたハッシュ又は暗文に一致

50

するハッシュ又は暗文に関連付けられた内容フラグメントを含むものである決定する段階と、

何れかのグループが少なくとも一つの条件を満たすならば、前記メッセージが前記事前選択されたデータの少なくとも一部を含んでいると決定する段階とを含む決定する段階と、

内容探索が実行されたメッセージが前記事前選択されたデータの少なくとも一部を含んでいると決定されたことに応じて、当該メッセージを前記個々のあて先に到達しないようにする段階と

を含む方法を実行させる指示を記憶するコンピューター読み取り可能記憶媒体。

【請求項 10】

10

表構造を有する事前選択された機密データのデータ要素から導き出された抽象的データ構造であって、当該抽象的データ構造は、表構造を有する前記事前選択されたデータのインデックスであり、前記抽象的データ構造内のエントリは、前記表構造内のセルの内容の事前作成されたハッシュ又は暗文と前記セルの行番号とを含み、当該事前選択された機密データの前記表構造のセルの内容を含んでいない抽象的データ構造を、パーソナルコンピューター装置がサーバーから受信する段階と、

前記サーバーに保存されている前記事前選択された機密データの少なくとも一部がパーソナルコンピューター装置の複数のデータ記憶媒体の内容に含まれていることを示すしを求めて、当該パーソナルコンピューター装置の複数のデータ記憶媒体の内容を探索する段階であって前記探索する段階は、

20

前記抽象的データ構造を探索して、記憶媒体からの内容の内容フラグメントであって、単語又は句を含む前記内容フラグメントのハッシュ又は暗文に一致する前記事前選択されたデータについて事前作成されたハッシュ又は暗文を含む抽象的データ構造内のエントリを識別する段階と、

前記内容フラグメントのグループを決定する段階であって、各グループは、前記抽象的データ構造内の、同じ行番号を有するエントリの事前作成生成されたハッシュ又は暗文に一致するハッシュ又は暗文に関連付けられた内容フラグメントを含むものである決定する段階と

を含む探索する段階と、

前記サーバーに保存されている前記事前選択された機密データの一部が前記パーソナルコンピューター装置に前記パーソナルコンピューター装置のユーザによって保存されていることを検出するよう、当該パーソナルコンピューター装置の複数のデータ記憶媒体の少なくとも一つの内容において、前記グループが少なくとも一つの条件を満たすことに基づいて前記事前選択された機密データの少なくとも一部を検出する段階と、

30

前記事前選択された機密データの一部のデータの検出に関する通知を、前記パーソナルコンピューター装置から前記サーバーに送る段階と、を備えていることを特徴とする方法。

【請求項 11】

前記事前選択された機密データの少なくとも一部が検出された場合、前記検出されたデータへのアクセスを防ぐ段階を更に含んでいることを特徴とする、請求項 10 に記載の方法。

40

【請求項 12】

前記内容は、定期的に探索されることを特徴とする、請求項 10 に記載の方法。

【請求項 13】

前記内容は、前記パーソナルコンピューター装置がネットワークと切断されているときに探索されることを特徴とする、請求項 10 に記載の方法。

【請求項 14】

前記通知を送る段階は、

前記事前選択された機密データを検出すると、前記事前選択された機密データの検出の通知を含んでいるメッセージを作成する段階と、

50

前記メッセージを送信キューに置く段階と、

前記パーソナルコンピュータ装置が前記システムに再接続された後、前記メッセージを前記システムに送信する段階と、を含んでいることを特徴とする、請求項 13 に記載の方法。

【請求項 15】

前記パーソナルコンピュータ装置に関する探索の範囲を定義する指示を前記サーバーから受け取る段階を更に備えていることを特徴とする、請求項 10 に記載の方法。

【請求項 16】

前記パーソナルコンピュータ装置の複数のデータ記憶媒体の内容を探索する前記段階は、前記事前選択された機密データの少なくとも一部の存在に関して、1 つ又は複数の特定のデータオペレーションをモニターする段階を含んでいることを特徴とする、請求項 10 に記載の方法。

10

【請求項 17】

前記 1 つ又は複数の特定のデータオペレーションの少なくとも 1 つは、ファイル読み取り、ファイル書き込み、ファイル更新、取り外し可能な媒体装置からの読み取り、取り外し可能な媒体装置への書き込み、及び、前記パーソナルコンピュータ装置上を走っているプログラムによる前記複数のデータ記憶媒体の何れかに記憶されているデータへのアクセス、から成るグループから選択されることを特徴とする、請求項 16 に記載の方法。

【請求項 18】

前記事前選択された機密データは、スプレッドシート、フラットファイル、及びデータベースの内の少なくとも 1 つで、組織によって維持されていることを特徴とする、請求項 10 に記載の方法。

20

【請求項 19】

前記抽象的データ構造のエントリは、それぞれ、前記セルの列番号と、随意的に列型式を更に含み、前記探索する段階は更に、

記憶媒体の内容を前記内容フラグメントで解析する段階と、

前記内容フラグメントの少なくとも一部分にハッシュ関数を適用し、前記内容フラグメントのハッシュを作成する段階と、

前記内容フラグメントのそれぞれについてハッシュ表衝突リストを作成する段階であって、前記ハッシュ表衝突リストは、前記抽象的データ構造内のエントリについて事前作成されたハッシュと内容フラグメントのハッシュとの間の一致のリストであり、各グループは、同じ行番号を有する前記抽象的データ構造内のエントリを識別するハッシュ表衝突リストと関連付けられた内容フラグメントを有するハッシュ表衝突リストを作成する段階とを含むことを特徴とする、請求項 18 に記載の方法。

30

【請求項 20】

前記複数のデータ記憶媒体は、メインメモリ、スタティックメモリ、及び大容量記憶メモリから成るグループから選択されることを特徴とする、請求項 10 に記載の方法。

【請求項 21】

前記複数のデータ記憶媒体の内容を探索する段階は、

前記複数のデータ記憶媒体内の各揮発性記憶装置の内容を探索する段階と、

前記複数のデータ記憶媒体内の各永久記憶装置の内容を探索する段階と、を含んでいることを特徴とする、請求項 10 に記載の方法。

40

【請求項 22】

前記パーソナルコンピュータ装置上で走っているアプリケーションによって、前記事前選択されたデータの使用を検出する段階を更に含んでいることを特徴とする、請求項 21 に記載の方法。

【請求項 23】

前記事前選択されたデータを使っている前記アプリケーションを識別する段階と、

前記識別されたアプリケーションを報告する段階と、を更に含んでいることを特徴とする、請求項 21 に記載の方法。

50

【請求項 24】

表構造を有する事前選択された機密データのデータ要素から導き出された抽象的データ構造をパーソナルコンピュータ装置がサーバーから受信する手段であって、前記抽象的データ構造は、表構造を有する前記事前選択されたデータのインデックスであり、前記抽象的データ構造のエントリは、前記表構造内のセルの内容の事前作成されたハッシュ又は暗文と前記セルの行番号とを含み、前記抽象的データ構造は、当該事前選択された機密データの前記表構造内のセルの内容を含んでいない受信する手段と、

前記サーバーに保存されている前記事前選択された機密データの少なくとも一部がパーソナルコンピュータ装置の複数のデータ記憶媒体の内容に含まれていることを示すしるしを求めて、当該パーソナルコンピュータ装置の複数のデータ記憶媒体の内容を探索する手段であって、前記探索する手段は、

10

前記抽象的データ構造を探索して、記憶媒体からの内容の内容フラグメントであって、単語又は句を含む前記内容フラグメントのハッシュ又は暗文に一致する前記事前選択されたデータについて事前作成されたハッシュ又は暗文を含む抽象的データ構造内のエントリを識別する段階と、

前記内容フラグメントのグループを決定する段階であって、各グループは、前記抽象的データ構造内の、同じ行番号を有するエントリの事前作成生成されたハッシュ又は暗文に一致するハッシュ又は暗文に関連付けられた内容フラグメントを含むものである決定する段階と

を含む探索する手段と、

20

前記サーバーに保存されている前記事前選択された機密データの一部が前記パーソナルコンピュータ装置に前記パーソナルコンピュータ装置のユーザによって保存されていることを検出するよう、当該パーソナルコンピュータ装置の複数のデータ記憶媒体の少なくとも一つの内容において、前記グループが少なくとも一つの条件を満たすことに基づいて前記事前選択された機密データの少なくとも一部を検出する手段と、

前記事前選択された機密データの一部のデータの検出に関する通知を、前記パーソナルコンピュータ装置から前記サーバーに送る手段と、
を備えていることを特徴とする装置。

【請求項 25】

前記内容は、定期的に探索されることを特徴とする、請求項 24 に記載の装置。

30

【請求項 26】

前記内容は、前記パーソナルコンピュータ装置がネットワークから切断されているときに探索されることを特徴とする、請求項 24 に記載の装置。

【請求項 27】

前記通知を送るための手段は、

前記事前選択された機密データを検出すると、前記事前選択された機密データの検出の通知を含むメッセージを作成するための手段と、

前記メッセージを送信キューに置くための手段と、

前記パーソナルコンピュータ装置が前記システムに再接続された後、前記メッセージを前記システムに送信するための手段と、を含んでいることを特徴とする、請求項 24 に記載の装置。

40

【請求項 28】

前記パーソナルコンピュータ装置に関する探索の範囲を定義する指示を前記サーバーから受け取るための手段を更に備えていることを特徴とする、請求項 24 に記載の装置。

【請求項 29】

前記パーソナルコンピュータ装置の複数のデータ記憶媒体の内容を探索するための前記手段は、前記事前選択された機密データの少なくとも一部の存在に関して、1つ又は複数の特定のデータオペレーションをモニターするための手段を含んでいることを特徴とする、請求項 24 に記載の装置。

【請求項 30】

50

前記１つ又は複数の特定のデータオペレーションの少なくとも１つは、ファイル読み取り、ファイル書き込み、ファイル更新、取り外し可能な媒体装置からの読み取り、取り外し可能な媒体装置への書き込み、及び、前記パーソナルコンピュータ装置上を走っているプログラムによる前記複数のデータ記憶媒体の何れかに記憶されているデータへのアクセス、から成るグループから選択されることを特徴とする、請求項２９に記載の装置。

【請求項３１】

前記複数のデータ記憶媒体は、メインメモリ、スタティックメモリ、及び、大容量記憶メモリから成るグループから選択されることを特徴とする、請求項２４に記載の装置。

【請求項３２】

前記複数のデータ記憶媒体の内容を探索するための前記手段は、
前記複数のデータ記憶媒体内の各揮発性記憶装置の内容を探索するための手段と、
前記複数のデータ記憶媒体内の各永久記憶装置の内容を探索するための手段と、を含んでいることを特徴とする、請求項２４に記載の装置。

【請求項３３】

前記パーソナルコンピュータ装置上で走っているアプリケーションによって、前記事前選択されたデータの使用を検出するための手段を更に含んでいることを特徴とする、請求項３２に記載の装置。

【請求項３４】

前記事前選択されたデータを使っている前記アプリケーションを識別するための手段と、
前記識別されたアプリケーションを報告するための手段と、を更に含んでいることを特徴とする、請求項３２に記載の装置。

【請求項３５】

プロセッサ上で実行されると、前記プロセッサに、
表構造を有する事前選択された機密データのデータ要素から導き出された抽象的データ構造をパーソナルコンピュータ装置がサーバーから受信する段階であって、前記抽象的データ構造は、表構造を有する前記事前選択されたデータのインデックスであり、前記抽象的データ構造のエントリは、前記表構造内のセルの内容の事前作成されたハッシュ又は暗文と前記セルの行番号とを含み、前記抽象的データ構造は、当該事前選択された機密データの前記表構造内のセルの内容を含んでいない受信する段階と、

前記サーバーに保存されている前記事前選択された機密データの少なくとも一部がパーソナルコンピュータ装置の複数のデータ記憶媒体の内容に含まれていることを示すしるしを求めて、当該パーソナルコンピュータ装置の複数のデータ記憶媒体の内容を、当該パーソナルコンピュータ装置によって探索する段階であって、前記探索する段階は、

前記抽象的データ構造を探索して、記憶媒体からの内容の内容フラグメントであって、単語又は句を含む前記内容フラグメントのハッシュ又は暗文に一致する前記事前選択されたデータについて事前作成されたハッシュ又は暗文を含む抽象的データ構造内のエントリを識別する段階と、

前記内容フラグメントのグループを決定する段階であって、各グループは、前記抽象的データ構造内の、同じ行番号を有するエントリの事前作成生成されたハッシュ又は暗文に一致するハッシュ又は暗文に関連付けられた内容フラグメントを含むものである決定する段階と
を含む探索する段階と、

前記サーバーに保存されている前記事前選択された機密データの一部が前記パーソナルコンピュータ装置に前記パーソナルコンピュータ装置のユーザによって保存されていることを検出するよう、当該パーソナルコンピュータ装置の複数のデータ記憶媒体の少なくとも一つの内容において、前記グループが少なくとも一つの条件を満たすことに基づいて前記事前選択された機密データの少なくとも一部を、当該パーソナルコンピュータ装置によって検出する段階と、

前記事前選択された機密データの一部のデータの検出に関する通知を、前記パーソナル

10

20

30

40

50

コンピューター装置から前記サーバーに送る段階と、
を含む方法を実行させる指示を記憶するコンピューター読み取り可能記憶媒体。

【請求項 3 6】

前記少なくとも一つの条件は、何れかのグループが、数量閾値に達する又は超える数量の内容フラグメントを含む場合に満たされる数量閾値であることを特徴とする、請求項 1 に記載の方法。

【請求項 3 7】

前記少なくとも一つの条件は、何れかのグループが、数量閾値に達する又は超える数量の内容フラグメントを含む場合に満たされる数量閾値であることを特徴とする、請求項 9 に記載のコンピューター読み取り可能記憶媒体。

10

【請求項 3 8】

前記少なくとも一つの条件は、何れかのグループが、数量閾値に達する又は超える数量の内容フラグメントを含む場合に満たされる数量閾値であることを特徴とする、請求項 1 0 に記載の方法。

【請求項 3 9】

前記少なくとも一つの条件は、何れかのグループが、数量閾値に達する又は超える数量の内容フラグメントを含む場合に満たされる数量閾値であることを特徴とする、請求項 2 4 に記載の装置。

【発明の詳細な説明】

【技術分野】

20

【0 0 0 1】

本発明は、データを処理する分野に関しており、具体的には、本発明は、情報コンテンツ内で事前選択された（例えば所有権のある）データを検出することに関する。

【背景技術】

【0 0 0 2】

多くの組織は、リレーショナルデータベース内に大量の安全機密情報を保存している。この型式のデータは、通常、物理的な保護、アクセス制御、周辺の保安制約、及び場合によっては暗号化を含む非常に徹底した保安方策に委ねられている。データベースのデータへのアクセスは、企業内の多くの従業員の仕事を機能させるには必要不可欠なので、この情報が盗難にあったり偶然に配布されたりする可能性は大いにある。情報の盗難は、知的財産の価値という点からも、法令遵守に関連する法律上の信頼性の点からも重大な経営的危険性を意味する。

30

【0 0 0 3】

リレーショナルデータベースシステム

リレーショナルデータベースシステムは、膨大な範囲のアプリケーションに有用である。関係する構造体は、データを問い合わせるのに自然な直感的方法を提示し、下層のディスク記憶システムの詳細をユーザーから隠すという付加的利点を有する様式でデータを保持している。データベースシステムの典型的なアプリケーションは、自然に表構造にフォーマットされる大量の小さなデータを記憶し検索することである。殆どの人に関心を持っている問い合わせの型式は、以下に概要を述べるが、周知のインデックス構造を使って最適化することができるので、リレーショナルデータベースは非常に有用である。

40

【0 0 0 4】

リレーショナルデータベースシステムに要求される問い合わせは、ユーザーが自分の探している表データを簡潔に要求できるようにする構造化問い合わせ言語（SQL）と呼ばれる自然な直感的述語論理を使用する。データベース表には、殆ど常に、SQLに基づく問い合わせを更に効率的にするインデックスが備えられている。これらのインデックスは、Bツリーと呼ばれるデータ構造を使ってメモリ内に記憶されている。現下の議論に最も関係のある、Bツリーの顕著な特長は、以下の通りである。

【0 0 0 5】

Bツリーは、バイナリツリーに基づく抽象的データ構造であり；

50

Bツリーには、インデックス付けする複数のコピーを含んでいなければならない；
Bツリーは、以下に概説するする問い合わせ例を使うのが最も効率的である。

【0006】

多数の問い合わせ例があり、
A=v の形態の正確な一致問い合わせで、ここに；
A は、所与のデータベース表の列又は「属性」であり、
V は、特定の属性値であり、
例えば、SELECT*FROM CUSTOMERS WHERE Income=30,000
v1<A<v2 の形態の範囲問い合わせで、ここに；
A は、所与のデータベース表の列又は「属性」であり、
例えば、SELECT*FROM CUSTOMERS WHERE 30<Income<40
A MATCHES* の形態の接頭語の問い合わせで、ここに、
「s」は特定のストリング値であり、
「s*」は正規表現であり、
例えば、Last_Name MATCHES "Smith*"

10

【0007】

データベースシステムの分野における初期の著作に関して沢山の参考文献がある。第1は、E. F. Codd「大きな共有型データバンク用のデータの関係モデル」ACM通信13(6)、377-387、1970年、によるリレーショナルデータベースに関する独創作業である。

20

【0008】

第2の参考文献は、上に概要を述べた型式の効率的な問い合わせを可能にする基礎的なデータ構造である「Bツリー」データ構造に関する最初に出版された著作の内の1つである。Rudolf BayerとEdward M. McCreightによる「大量の順序付けされたインデックスの編集と管理」、データ記述とアクセスに関する1970 ACM SIGFIDEワークショップの記録、1970年11月15-16日、米国、テキサス州ヒューストン、Rice大学（補遺付き第2版）、107-141頁、ACM、1970、を参照されたい。

【0009】

情報検索システム

30

情報検索は、文書内に見られるテキストデータの記憶と検索を取り扱う広範な分野である。これらのシステムは、表データではなく主に標準的な文書に焦点を当てており、データベースシステムのものとは異なっている。このシステムの初期の例は、コーネル大学でSMARTシステムの一部として開発された。今日、最も良く知られている情報検索アプリケーションは、Google、Inktomi及びAltaVistaの様なウェブベースの探索エンジンである。これらのシステムを使用する一般的な方法は、もっと大きなデジタル文書セットの一部である文書への参照を発見することである。これらのアプリケーションに関するユーザー経験は、通常は、一連の問い合わせと、結果のブラウジングが交錯して構成されている。問い合わせの結果は、関連性が高い順に示されており、ユーザーは、更にブラウジングした後で、問い合わせを精緻化することができる。リレーショナルデータベースについて、これらのシステムが並外れて人気があるのは、人々が最も有用であると分かっている問い合わせの型式に対し迅速に応答するという、基礎をなすインデックスの能力によるものである。

40

【0010】

これらのシステムの大部分は、インデックスが付いた文書の集まりから構築される、いわゆる「用語索引」から導き出されたインデックスに基づいている。これらの用語索引は、各用語について、各文書内でその用語が発生した各場所を一覧表にしているデータ構造を含んでいる。そのようなデータ構造によって、特定の用語を含んでいる全ての文書を迅速に探索できるようになる。用語の集まりを含んでいる全ての文書に問い合わせるユーザーの問い合わせに対して、インデックスは、高次のユークリッド空間内の多数のベクトル

50

を表すように構成される。次いでユーザーの問い合わせ用語のリストも、この空間内のベクトルとして再翻訳される。問い合わせは、文書空間内のどのベクトルが問い合わせベクトルに最も近いかを発見することによって実行される。この最後の段階には、正確さと速度を求めて様々な最適化が施され、「余弦計量」と呼ばれている。

【0011】

先に述べたように、この種のシステムとの典型的なユーザーの対話は、問い合わせ、ブラウジング、精査、そして再度問い合わせに戻る、繰り返しのサイクルである。問い合わせの結果は、通常は、関連性が高い順にランク付けされた多数の文書であり、間違いの可能性の割合が非常に高いこともある。問い合わせに関する幾つかの標準的な例がある。

【0012】

a) 「データベース」と「インデックス」という用語を含んでいる全ての文書
b) 「データベース」又は「インデックス」という用語を含むが「サイベース」は含んでいない全ての文書、のようなブール問い合わせ。

【0013】

a) 「犬」という用語を含んでいる文書でリンクされている全ての文書
b) 「犬」という用語を含んでいる、最も「人気のある」（即ちリンクされている）文書、のようなリンクベースの問い合わせ。

【0014】

情報検索システムの最初の重要な実行プロジェクトの内の1つは、コーネル大学のSMARTシステムである。このシステムは、今日でも使用されている情報検索システムの多くの基本的な構成要素を含んでいる：C. Buckleyによる「SMART情報検索システムの実行」技術レポートTR85-686、コーネル大学、1985年。

【0015】

WAISプロジェクトは、シンキングマシン社製の大量並列処理スーパーコンピュータの初期のアプリケーションだった。これは、インターネット上で利用可能になった最初の情報検索システムの内の1つである。この仕事に関して最初に言及しているのが、Brewster KahleとArt Medlarの「企業ユーザー用の情報システム：広域情報サーバー」技術レポートTMC-199、シンキングマシン社、1991年4月、3.19版である。

【0016】

多くの現在のインターネット探索サービスの市販者の中に、Googleがある。探索の精度におけるGoogleの真のブレークスルーは、インデックスが付けられた文書のテキストとハイパーリンク構造の両方からデータを取り込める能力である。Sergey Brin、Lawrence Pageの「大型ハイパーテキストのウェブ探索エンジンの構造」<http://dbpubs.stanford.edu:8090/pub/1998-8>を参照されたい。

【0017】

ファイルシングリングシステム

インターネットと、デジタル文書をコピーし配信する手頃な手段の成長は、不法又は不適切な文書のコピーを検出するのを助ける技術における研究の関心を高めた。この仕事に関する主要なアプリケーションは、著作権法の違反を検出することであり、盗用を検出することである。この問題は、無差別eメール（AKA要求しないのに送られてくる宣伝用のeメール）の検出及び自動削除にも関係するので、相当な関心がある。これらの技法の大部分を記述するのに利用する技術用語は、文書フラグメントの隣接するシーケンスが、ハッシュコードによって「シングル」されて減少し、文書内で発見されたのと同じシーケンスでルックアップ表内に記憶される「ファイルシングリング」である。

【0018】

ファイルシングリングは、2つの文書の間の類似性を探す非常に迅速な方法を提供する。特定の文書（例えばテキストファイル）を保護するために、文書は、文書を文章毎にハッシングし、これらのハッシングした文章を、迅速に探索するための表内に記憶させることによってシングルされる。新しい文書が著作権で保護された内容のフラグメントを含ん

10

20

30

40

50

でいるかどうかを試験して調べるために、試験メッセージの各フラグメントに同じハッシュ関数が適用され、そのフラグメントが著作権で保護された内容で現れるのと同様の順序で現れるかどうかを見る。その技法は、個々のフラグメントを探索するのに必要な時間が極めて短いので、迅速である。

【0019】

ファイルシングリングシステムとの典型的なユーザーの対話は、能動的ではなく受動的である。ファイルシングリングシステムは、普通は、文書を自動的に処理し、問い合わせ結果を非同期的にユーザーに配信するように設定されている。典型的なファイルシングリングアプリケーションでは、一式のメッセージを用いて組織が組織自体のeメールシステムに配信したくない制約された内容のインデックスを作る、無差別揭示防止になっている。このシナリオでは、「問い合わせ」は、自動的なeメールメッセージの処理と、適切な自動的経路指定に過ぎない。

10

【0020】

文書等個性の問い合わせについては、各試験文書tで、tと同じ内容を有するインデックス付き文書の集まりの中の全文書dを突き止める。無差別揭示検出の場合、セットdを、全ての既知の積極的な無差別揭示メッセージとし、文書tを、入信eメールメッセージとすればよい。

【0021】

カットアンドペースト検出の問い合わせについては、各試験文書tで、何らかのdのフラグメントがtで発生しているインデックス付き文書の集まりの中の全文書dを突き止める。剽窃を検出する場合、セットdを、特定のクラスに関して先に提出された全エッセイにし、文書tを、剽窃の疑いのある学生が書いた新しい文書にすればよい。

20

【0022】

ファイルシングリングにおける主要な発行済み調査プロジェクトは、K O A L A、C O P S及びS C A Mと呼ばれている。それら全てが、性能及び精度を最適にする変数と共に上に述べた基本的なファイルシングリング法の変数を使用している。K O A L Aの情報に関しては、N. H e i n t z eによる「計測可能な文書諮問押捺制度」（電子商取引における第2回U S E N I Xワークショップの会議録、1996年11月）を参照されたい。
<http://www-2.c.s.cmu.edu/afs/cs/user/nch/wwww/koala/main.html>。C O P Sの情報については、S. B r i n、J. D a v i s及びH. G a r c i a-M o l i n aによる「デジタル文書のためのコピー検出機構」（ACM S I G M O D年次会議の会議録、1995年5月）を参照されたい。S C A Mの情報については、N. S h i v a k u m a r及びH. G a r c i a-M o l i n aによる「S C A M：デジタル文書のためのコピー検出機構」（デジタルライブラリの理論及び実践（D L ' 95）における第2回国際会議の会議録、1995年6月）
<http://www-db.stanford.edu/~shiva/SCAM/scaminfo.html>。と、N. S h i v a k u m a r及びH. G a r c i a-M o l i n aによる「計測可能で正確なコピー検出機構の構築」（デジタルライブラリ（D L ' 95）に関する第1回ACM会議の会議録、1996年3月）
<http://www-db.stanford.edu/pub/paper/performance.ps>を参照されたい。

30

40

【0023】

インターネットの内容ろ過システム

内容ろ過システムと呼ばれる様々な市販のアプリケーションが保護手段を実行する。この範疇には、ウェブサイト制約／モニタリングソフトウェアとeメール内容制御の2つの主要なアプリケーション型式がある。どちらの場合も、現在使用されているメインアルゴリズムは、データの誤使用を示すテキストフラグメントの集まりのセットに関して行う正規表現のセットに対するパターンマッチングである。例えば、テキストフラグメント「X X X」を含むURLでの全てのブラウジングを制限することである。eメールの内容制御カテゴリの代表的な例は、「所有権」及び「秘密」という用語を含んでいるが、「ジョー

50

ク」又は「冗談」という用語を含んでいない全eメールを停止し、阻止することである。

【0024】

【非特許文献1】E. F. Codd「大きな共有型データバンク用のデータの関係モデル」

【非特許文献2】Rudolf BayerとEdward M. McCreightによる「大量の順序付けされたインデックスの編集と管理」

【非特許文献3】C. Buckleyによる「SMART情報検索システムの実行」

【非特許文献4】Brewster KahleとArt Medlarの「企業ユーザー用の情報システム：広域情報サーバー」

【非特許文献5】Sergey Brin、Lawrence Pageの「大型ハイパーテキストのウェブ探索エンジンの構造」 10

【非特許文献6】N. Heintzeによる「計測可能な文書諮問押捺制度」

【非特許文献7】S. Brin、J. Davis及びH. Garcia-Molinaによる「デジタル文書のためのコピー検出機構」

【非特許文献8】N. Shivakumar及びH. Garcia-Molinaによる「SCAM：デジタル文書のためのコピー検出機構」

【非特許文献9】N. Shivakumar及びH. Garcia-Molinaによる「計測可能で正確なコピー検出機構の構築」

【発明の開示】

【課題を解決するための手段】 20

【0025】

パーソナルコンピューター装置に記憶されている事前選択されたデータを検出するための方法と装置について説明している。或る実施形態では、本方法は、埋め込まれている事前選択されたデータに関してネットワーク上を電子的に送信されたメッセージをモニターする段階と、メッセージ上で内容探索を実行し、事前選択されたデータから導き出された抽象的データ構造を使って、埋め込まれている事前選択されたデータの存在を検出する段階と、を含んでいる。

【発明を実施するための最良の形態】

【0026】

本発明は、本発明の様々な実施形態に関する以下の詳細な説明及び添付図面から良く理解頂けるであろうが、これらは、本発明を特定の実施形態に限定するものではなく、説明と理解のためのものに過ぎない。 30

【0027】

パーソナルコンピューター装置上のあらゆる場所における機密情報の使用を追跡及びモニターするシステム及び方法をここに説明している。或る実施形態では、このモニタリングは、デスクトップコンピューター又はポータブルコンピューターのようなパーソナルコンピューター装置のデータ記憶媒体の内容探索を実行することによって実施される。別の実施形態では、モニタリングは、メッセージがパーソナルコンピューター装置で送受信されるときに、メッセージの内容探索を実行することによって実施される。更に別の実施形態では、モニタリングは、パーソナルコンピューター装置上で実行されているあらゆるアプリケーション内で潜在的機密情報が使用される前、使用されている間、及び使用された後に内容探索を実行することによって実施される。或る実施形態では、ここで説明しているシステムは、大量のデータベースのデータを取り扱える安全且つ測定可能な方法で、この情報を検出することができる。データベースのデータは、限定するわけではないが、リレーショナルデータベース、スプレッドシート、フラットファイルなどを含む様々なシステム内に記憶されているあらゆる形態の表様式データを備えている。 40

【0028】

以下の説明では、本発明を徹底的に説明するため膨大な詳細事項を記載している。しかしながら、当業者には自明のように、本発明は、これら特定の詳細事項を備えていなくても実施することができる。別の例では、本発明が分かり難くならないように、周知の構造 50

と装置は、詳しく示さずブロック図の形で示している。

【0029】

以下の詳細な説明の幾つかの部分は、コンピューターメモリ内のデータビットに関するオペレーションのアルゴリズムと記号的表現で表示されている。これらのアルゴリズムの記述及び表現は、データ処理技術分野の当業者が、他の当業者に、彼等の仕事の本質を最も効果的に伝えるために用いる手段である。ここでは、そして一般的に、アルゴリズムは、所望の結果に辿り着く自己一貫したステップのシーケンスであると考えられる。このステップは、物理量の物理的な操作を要するステップである。通常は、必ずというわけではないが、これらの量は、記憶し、伝送し、結合し、比較し、或いは操作することのできる電気又は磁気信号の形態を取っている。時々、主として共通に使用する目的で、これらの信号をビット、数値、要素、記号、文字、用語、数字などで表すのが便利であると分かっている。

10

【0030】

しかしながら、以上の及び同様の用語は、全て適切な物理量と関係しており、これらの量に適用する便宜的ラベルに過ぎないことが頭をよぎる。特記しないかぎり、以下の議論から明らかなように、記述全体を通して、「処理」、「演算」、「計算」、「判断」又は「表示」などの様な用語を使用している議論は、コンピューターシステムのレジスタ及びメモリ内で物理（電子）量で表されているデータを、コンピューターシステムメモリ又はレジスタ或いは他のそのような情報記憶装置か、変換又は表示装置内の物理量として同様に表される他のデータに操作及び変換する、コンピューターシステム、又は同様の電子演算装置の動作と処理を指すものと理解されたい。

20

【0031】

本発明は、更に、ここに述べるオペレーションを実行するための装置に関する。この装置は、必要な目的のために特別に作ってもよいし、コンピューターに記憶されているコンピュータープログラムによって選択的に起動又は再構成される汎用コンピューターを備えていてもよい。そのようなコンピュータープログラムは、限定するわけではないが、フロッピー（登録商標）ディスク、光ディスク、CD-ROM及び磁気光ディスクを含む何らかの型式のディスクか、読み取り専用メモリ（ROM）か、ランダムアクセスメモリ（RAM）か、EPROMか、EEPROMか、磁気又は光カードか、電子的命令を記憶するのに適している何らかの型式の媒体の様なコンピューター読み取り可能記憶媒体内に記憶され、それぞれが、コンピューターシステムのバスに連結されている。

30

【0032】

ここに呈示されているアルゴリズム及び表示は、本来的に、何れかの特定のコンピューター又は他の装置に関係してはいない。様々な汎用システムを、ここでの教示によるプログラムと共に使用してもよいし、必要な方法ステップを実行するため、更に特別仕様の装置を構築すると便利なことも分かっている。これらの様々なシステムに必要な構造は、以下の説明から明らかになるであろう。更に、本発明は、何れかの具体的なプログラミング言語に関連付けて説明してはいない。ここに記載している本発明の教示を実行するのに、様々なプログラミング言語を使用できるものと理解されたい。

【0033】

機械読み取り可能な媒体は、情報を、機械（例えばコンピューター）によって読み取り可能な形態で記憶又は送信するためのあらゆる機構を含んでいる。例えば、機械読み取り可能な媒体には、読み取り専用メモリ（ROM）、ランダムアクセスメモリ（RAM）、磁気ディスク記憶媒体、光記憶媒体、フラッシュメモリ装置、或いは、電気、光、音響又は他の形態の伝播信号（例えば、搬送波、赤外線信号、デジタル信号など）などが含まれる。

40

【0034】

代表的な実施形態の構成要素

或る実施形態では、ここに記載されている検出方式を実行するためのシステムは、ポリシー管理システム（PMS）とメッセージモニタリングシステム（MMS）の2つの主要

50

な構成要素で構成されている。PMSは、ネットワーク上で送られるメッセージに含まれるか、ポータブルコンピューター、デスクトップコンピューター、パーソナルデジタルアシスタント、携帯電話などの様なパーソナルコンピューター装置のデータ記憶媒体に記憶されているデータ（例えばデータベースのデータ）の使用と送信のための情報保護ポリシーを定めるユーザー入力の受け入れを担当している。従って、このデータは、事前に選択されている。ここで使用する「パーソナルコンピューター装置のデータ記憶媒体」という用語は、パーソナルコンピューター装置内に在るか、一時的又は永久的にパーソナルコンピューター装置用のデータを記憶するパーソナルコンピューター装置がアクセス可能なあらゆる記憶装置を指す。

【0035】

MMSは、ネットワーク上で送られるメッセージ、パーソナルコンピューター装置で処理されるデータ、又はパーソナルコンピューター装置のデータ記憶媒体に記憶されるデータの内容探索の実行と、ユーザーによるPMSに対するポリシーの指定の実施とを担当している。或る実施形態では、これらのシステムは、両方共、情報交換のためあらゆる標準的プロトコルで交信するコンピューターネットワークに連結されている。

【0036】

この実施形態では、通常のオペレーションの途上で、ユーザーは、或る人間によるデータベースのデータの使用又は送信を制限する所与のポリシーを実行するよう決定し、次いでグラフィカルユーザーインターフェースと1つ又は複数のユーザー入力装置（例えば、マウス、キーボードなど）を使って、このポリシーをPMSに手動で入力する。ユーザーインターフェースは、入力を受け取り、PMSを備えたコンピューターシステム又は個々の機械で実行されている。例えば或るポリシーは、顧客サービスにおける所与のグループの個々人が、事前選択されたデータを含むデータファイルを、パーソナルコンピューター装置に取り付けられている取り外し可能な媒体装置に保存するのを停止させる。或る実施形態では、ポリシーは、所望される保護の特性（例えば従業員の或るサブセットだけを制限する）、保護を要するデータの型式（例えばデータベースのデータ）、及び保護を要するデータベースのデータのネットワーク位置（例えば、データベース表の名前、サーバーのIPアドレス、サーバー又はファイル名）を含んでいる。ここでも、この情報は全て、ユーザーに特定の情報を正しいフィールドに入力するよう促す標準的グラフィカルユーザーインターフェースを使って指定される。

【0037】

或る実施形態ではユーザーが調節できるが、デフォルトでは指定された間隔（例えば1日）毎に一回である、或る規則的な間隔で、PMSは、データベースに問い合わせ、保護対象となるデータベースのデータのコピーを抽出し、そのデータから、以下に詳しく説明する抽象的データ構造（以後「インデックス」と呼ぶ）を導き出す。

【0038】

PMSは、次に、このインデックスを、実施すべきポリシーの詳細事項と共にMMSへ送り、MMSがそのポリシーの強制を開始できるようにする。MMSは、インデックスを、強制すべきポリシーの詳細と共にPMSから受け取る。MMSは、インデックスとポリシーの情報を使って、ユーザーが指定したポリシーを強制する。或る実施形態では、MMSは、このインデックスを使って、以下に詳しく論じるように、保護対象のデータベースのデータに関し、出力される各メッセージ（例えばeメール、ウェブメールメッセージなど）を探索する。別の実施形態では、MMSは、このインデックスを使用して、以下に詳しく論じるように、保護対象のデータベースのデータに関し、パーソナルコンピューター装置のデータ記憶媒体の内容、及び／又はユーザーとパーソナルコンピューター装置の間の対話の内容を探索する。

【0039】

代表的なワークフローの概要を図1に示すが、ここでは、最高価値の情報が識別され、ポリシーが生み出され、監視と強制が実行され、訴訟の対象となるビジネス情報に結びつくようになっている。

【0040】

ネットワークベースのオペレーションのモード

或る実施形態では、メッセージモニタリングシステムは、「監視モード」と「強制モード」の2つの方法の内の一方で構成されている。図2は、2つのネットワーク構成を示している。監視モードでは、MMSは、ポリシー違反に関してトラフィック及びレポートを観察できるネットワーク上のどこかに配置されているが、メッセージが出て行くときに阻止するようには構成されていない。これは、PMSが情報にアクセスしている図2Aに示されている。PMSは、スイッチ、タップ及びファイアウォールを介してインターネットに連結されている。MMSは、タップを使ってネットワークメッセージをモニターする。「強制モード」では、MMSは、違反に関しトラフィックとレポートを観察できるが、更に、メッセージを遮って経路を変更し、メッセージの最終的な宛先を変えることができる。これは、PMSが情報にアクセスして、スイッチとファイアウォールを介してインターネットに連結されている図2Aに示されている。この実施形態では、MMSは、一連のサーバーを使ってトラフィックをモニターし、メッセージが事前選択された情報を含んでいるようだと判断した場合は、例えば、特定のサーバーへとトラフィックの経路を変更する。MMSは、様々な層のプロトコル毎に異なるサーバーを使用することができる。

10

【0041】

メッセージの経路変更は、強制ではない。替わりに、MMSは、出て行くメッセージを遮って停めるよう構成することもできる。「強制モード」のポリシーの一つの例では、適切な懲戒行為を行えるように、ポリシーに違反する全てのメッセージを、ポリシーに違反した人のマネージャーに送るようになっている。

20

【0042】

両方のオペレーションのモードでは、多くのMMSをインストールして、それぞれが内容を検出するのに必要なインデックスに関する固有のコピーを備えておくようにすることができる。この平行処理形態は、尺度の問題と、情報の出口について可能性のある複数のポイントの防御とに役立つ。

【0043】

両方の形態において、MMSは、様々なアプリケーション層のプロトコル（例えば、SMTP、HTTP、FTP、AIM、ICQ、SOAPなど）を使って移送されるメッセージを積極的に解析する。

30

【0044】

或る実施形態では、2つのサブシステム（PMSとMMS）は、1つのローカルエリアネットワーク（LAN）上で走る。しかしながら、PMSとMMSは、同じ物理的又は論理的システム内に統合できる。この統合された形態は、システムを作るのに必要な商品経費を制御できるため、より適している。

【0045】

更に別の代替実施形態では、PMSとMMSは、必ずしも同じLAN上にはない。PMSは、データベースの情報と同じLAN上に在るが、MMSは、PMSが在るLANとは異なるLAN上に在る。この構成では、2つの異なるLANは、インターネットによって最終的には1つに連結されるが、ファイアウォール、ルータ及び／又は他のネットワーク装置によって分離される。これは、或る会社が、（法律事務所又は調査代理店の様な）彼等のデータベースのデータを必要としている他の会社を、最初の会社のデータベースのデータのポリシーに違反することから制限したい場合に好都合な構成である。

40

【0046】

図3は、データベースのデータを保護するためのプロセスの或る実施形態のフロー図である。このプロセスは、ハードウェア（回路、専用論理など）、（汎用コンピュータシステム又は専用機上を走っているような）ソフトウェア、又は両者の組み合わせを備えた処理論理によって実行される。

【0047】

図3に示すように、処理論理は、事前選択されたデータに関してメッセージをモニター

50

する。(処理ブロック301)。次に、処理論理は、メッセージが事前選択されたデータを有しているか否かを判断する(処理ブロック302)。有していなければ、処理は、ブロック3011へ移る。有していれば、処理論理は、個人的な送信/受信メッセージが、メッセージ内の情報を送信/受信する承認を受けているか否かを判断する(処理ブロック303)。承認を受けていれば、プロセスは終了し、処理は処理ブロック301へ移る。承認を受けていなければ、処理論理は、メッセージを遮り、メッセージの経路を変更し、メッセージをログするなどの1つ又は複数のアクションを取り(処理ブロック304)、処理は、処理ブロック301へ移る。

【0048】

クライアントベースのオペレーションのモード

クライアントベースのオペレーションのモードは、データーの潜在的な誤使用を含むユーザーのオペレーションを検出するために、パーソナルコンピューター装置のユーザーが取るモニタリング動作に向けられている。これらのユーザーのオペレーションは、例えば、演算システムの何れかの記憶装置上の制限されているデータベースのデータをセーブ又はアクセスする段階と、アプリケーション内で制限されているデータベースのデータを使用する段階と、制限されているデータベースのデータを印刷する段階と、何れかのネットワーク通信プロトコルで制限されているデータベースのデータを使用する段階などを含んでいる。或る実施形態では、ユーザーの動作のモニタリングは、パーソナルコンピューター装置の局所記憶システムにアクセス又はセーブされる内容、或いは様々なアプリケーション層プロトコル(例えば、SMTP、HTTP、FTP、AIM、ICQ、SOAPなど)を使って移送される内容の何れかを解析し探索することによって実行される。別の実施形態では、ユーザーの動作のモニタリングは、ユーザーとパーソナルコンピューター装置の間で交換されるデータを捕らえて解釈することによって実行される。

【0049】

図9は、事前選択された機密データのクライアントベースの保護に関するシステムの1つの実施形態のブロック図である。

【0050】

図9に示すように、サーバー902は、ネットワーク906を介してクライアントのコンピューター(クライアントと呼ぶ)910と通信する。ネットワーク906は、専用ネットワーク(例えば、ローカルエリアネットワーク(LAN))でもよいし、公開ネットワーク(例えば、ワイドエリアネットワーク(WAN))でもよい。クライアント910は、組織内の異なる従業員に属するコンピューターである。各クライアント910は、例えば、デスクトップコンピューター、ポータブルコンピューター(例えばラップトップ)、又は間欠的なネットワーク接続によって作動する他の何らかのコンピューターである。内容モニタリングシステム(ここでは、メッセージモニタリングシステム又はMMSとも呼ぶ)912は、各クライアント912上に在り、事前選択された機密データに関するこのクライアントのデータ記憶媒体の内容を探索することと、ユーザーとクライアント912の間で交換された内容を捕らえて解釈することを担当している。データ記憶媒体は、例えば、メインメモリ、スタティックメモリ、大容量記憶メモリ(例えばハードディスク)、又は、クライアントコンピューター用のファイル又は他の文書を一時的又は永久的に記憶する何らかの他の記憶装置を含んでいる。或る実施形態では、MMS912は、ファイルの読み取り、ファイルの書き込み、ファイルの更新のような特定のデータオペレーションをモニターし、取り外し可能な媒体装置(例えば、フロッピー(登録商標)ドライブ、ユニバーサルシリアルバス(USB)装置、コンパクトディスクレコーダブル(CDR)装置など)の読み取り書き込みを行う。MMS912のオペレーションは、取り外し可能で移動可能な装置による機密データの喪失を防止し易くする。例えば、MMS912のオペレーションは、ユーザーが、クライアント910に記憶されている機密データをフロッピー(登録商標)ディスクにコピーし、機密データを有するファイルをUSBベースの取り外し可能な記憶装置に移し、ラップトップ又はデスクトップコンピューターから機密データを印刷又はeメールし、機密データを承認されていないアプリケーションで使用する

10

20

30

40

50

などの場合に発生する機密データの漏洩を防ぐ。

【0051】

サーバー902は、組織の中で、ここに記載している検出方式の構築を担当している。サーバー902は、PMS904とメッセージコレクタ914を含んでいる。PMS904は、機密データの使用を制御する一式の安全ポリシーを維持している。一式の安全ポリシーは、機密データの潜在的な誤使用についてその人のコンピューターをモニターしなければならない従業員を識別し、探索を実行する機密データを指定し、探索の範囲を定義する（例えば、特定の記憶媒体、データオペレーションなど）。この情報に基づいて、PMS904は、対応するクライアント910を探索するか否かについて各MMSに指示し、探索に用いるインデックスを送る。インデックスは、安全ポリシーに基づいて1つ又は複数のクライアント912に対して事前選択された特定の機密データから導き出される。メッセージコレクタ914は、クライアント910のユーザーによるデータの誤使用を通知する、MMS912から受信するメッセージの収集を担当している。

10

【0052】

或る実施形態では、各MMS912は、サーバー902とのネットワーク接触を維持できないとき（例えば、ラップトップ910が週末に家に持ち帰られたり、他のネットワークに移されたり、盗まれるなど）、スタンドアローンで作動することができる。例えば、ユーザーがラップトップ910をネットワーク906から切断すると、ラップトップ910上で走っているMMS912は、ユーザーが家庭でラップトップで作業している間に、ラップトップ910のデータ記憶媒体の定期的な内容探索を実行する。具体的には、MMS912は、ラップトップ910のローカルファイルシステム、eメールメッセージアーカイブなどを探索する。更に、MMS912は、PMS902に指示された場合は、特定のデータオペレーション（例えば、ファイルの読み取り、ファイルの書き込み、ファイルの更新、フロッピー（登録商標）ディスクの様な取り外し可能な媒体装置に対する読み取り書き込み）をモニターする。或る実施形態では、MMS912は、クライアント910の何れかのデータ記憶媒体上で事前選択されたデータを検出すると、事前選択されたデータの検出の通知を含むメッセージを作成し、このメッセージを送信キューに配置する。後で、ネットワークへの接続が最確立されたときに、送信キューからのメッセージがメッセージコレクタ914へ送信される。或る実施形態では、PMS904が維持しているポリシーは、事前選択されたデータが検出されると、MMS912が事前選択されたデータへのアクセスを防ぐよう要求する。

20

30

【0053】

図10は、パーソナルコンピューター装置ベースの、事前選択されたデータの保護のためのプロセスの或る実施形態のフロー図である。このプロセスは、ハードウェア（回路、専用論理など）、（汎用コンピューターシステム又は専用機上を走っているような）ソフトウェア、又は両方の組み合わせを備えた処理論理によって実行される。処理論理は、クライアント910のようなパーソナルコンピューター装置上に在る。

【0054】

図10に示すように、処理論理は、パーソナルコンピューター装置で実行される内容探索の範囲を定義する指示を受け取る（処理ブロック1002）。或る実施形態では、指示は、探索しなければならないデータ記憶媒体と探索の周期を指定している。或る実施形態では、指示は、事前選択された機密データの存在に関してモニターすべきデータのオペレーションも指定している。

40

【0055】

次に、処理論理は、事前選択された機密データから導き出された抽象的データ構造即ちインデックスを受け取る（処理ブロック1004）。幾つかの抽象的データ構造の実施形態について、以下に更に詳細に論じる。

【0056】

処理ブロック1006では、処理論理は、抽象的データ構造を使って、事前選択された機密データに関して、パーソナルコンピューター装置のデータ記憶媒体の内容を探索する

50

。内容探索の範囲は、サーバーから受け取った指示で定義されている。探索は、このパーソナルコンピュータ装置のデータのデータ記憶媒体の内容、及び／又は、ユーザーとパーソナルコンピュータ装置の間でやり取りされた内容に対して行われる。或る実施形態では、内容探索は、所定の時間間隔で周期的に実行される。処理論理に用いられる探索技法の幾つかの実施形態について、以下に更に詳しく論じる。「パーソナルコンピュータ装置のデータ記憶媒体」という用語は、例えば、磁気ディスク、揮発性ランダムアクセスメモリ、取り外し可能な媒体、テープバックアップシステム、遠隔ネットワークアドレス指定可能記憶装置などを含むパーソナルコンピュータ装置がアクセス可能なあらゆる形態のデータ記憶装置を指す。或る実施形態では、処理論理は、パーソナルコンピュータ装置上で走っているアプリケーションによって、事前選択されたデータの使用を検出するため、揮発性記憶装置を探索する。使用が検出されれば、処理論理は、事前選択されたデータを使用しているアプリケーションを特定する。

10

【0057】

処理論理が事前選択されたデータ（又はその一部）の存在を検出すれば（処理ボックス1008）、処理論理は、PMSにより維持されているポリシーが、事前選択されたデータへのアクセスの阻止を要求しているか否かを判断する（処理ボックス1009）。或る実施形態では、検出されたデータへのアクセスが、このデータへのアクセスを試みているアプリケーションに対して阻止される。

【0058】

阻止が必要な場合、処理論理は、事前選択されたデータへのアクセスを阻止（処理ブロック1010）し、更に、パーソナルコンピュータ装置が、サーバー又は何れかの他の指定された装置とのネットワーク接触を維持できるか否かを判断する（処理ボックス1011）。この判断が正である場合、処理論理は、検出の通知を含むメッセージをサーバーへ送る（処理ブロック1012）。通知は、パーソナルコンピュータ装置と検出されたデータを特定する。或る実施形態では、通知は、パーソナルコンピュータ装置上で走っているときに事前選択されたデータを使っていたアプリケーションを特定する。

20

【0059】

パーソナルコンピュータ装置とサーバーの間が接続されていない場合、処理論理は、将来ネットワークとの接続が再確立されたときにサーバーへ送信するために、このメッセージをキュー内に置く（処理ブロック1014）。

30

【0060】

先に論じたように、パーソナルコンピュータ装置ベースのモニタリングは、パーソナルコンピュータ装置に保存され処理される内容の監視を考慮している。プロトコルに基づいてフィルタリングを実行する既存のデスクトップベースのファイアウォールとは違って、ここで説明している内容探索は、ファイルシステムやメモリバンク内の事前選択されたデータベースのデータ、又はアプリケーションがアクセスしているプロセス内のデータの追跡に関する具体的な探索の問題に取り組んでいる。

【0061】

要求者に資格証明書（例えばパスワード）を促すことによって、未承認のアクセスを禁止するアクセス制御技法に関して、ここで説明している機密データのクライアントベースの保護は、パーソナルコンピュータ装置内に保存されている内容を、この内容がダウンロードされたか又はアクセス制御システムを介してアクセスされた後でモニターする。

40

【0062】

デスクトップベースの暗号化／暗号解読パッケージシステムは、一般的に、サーバーベースの機構を頼りにデータを暗号化し、デスクトップベースの機構を頼りにデータを解読して閲覧するが、このシステムは、データを解読する暗号キーへのアクセスを制限することによって、データの誤使用を防ぐ働きをしている。ここで説明している機密データのクライアントベースの保護は、暗号作成法の包路線の「全くの」外側に残されており、従って第三者によって盗まれ易いデータを保護するのに使用することができる。

【0063】

50

添付書類内の敵意を隠したコードの存在を検出するのに通常用いられるアンチウイルスの解決法に対し、ここに説明している機密データのクライアントベースの保護は、隠されたコードの存在ではなく、事前選択されたデータベースのデータの存在を検出することに向けられている。

【0064】

そのパーソナルコンピュータ装置に送られてくる全ての内容をモニターする内容フィルターを使ってハードウェアのオペレーションを駆動するよう書かれているソフトウェアの形態をしたドライバフィルターは、事前選択されたデータに関してパーソナルコンピュータ装置のデータ記憶媒体の探索を実行できる能力に欠けている。

【0065】

1つ又は複数のシステム実施形態に関する安全要件

この検出システムの実施形態は情報の安全ポリシーを強制するのに用いられるので、このシステムの安全特性は最高のものである。或る実施形態では、このシステムの主要目的は、データベースのデータに関わる安全ポリシーを強制することである。これは、このシステムがデータベースのデータを扱う方法が非常に安全であることを示唆している。データベースのデータを保護する過程で、システムがデータベースのデータを盗むための新しい道を開けば、最終的な目的が覆される。

【0066】

或る実施形態では、MMSは、ネットワークを流れる膨大な数のメッセージをモニター及び／又は阻止するやり方で展開される。これは、トラフィックが集中しているネットワークの様々なポイント（例えば、ルーター、メールシステム、ファイアウォール、デスクトップコンピューター、eメールアーカイブシステムなど）にMMSをインストールすることを意味している。これは、MMSが、ネットワーク上のこれらの集中ポイントの1つの後ろか前の何れかにインストールされることを意味している。システムのこのような配置は、システムがメッセージの例外的な閲覧をできるようにし、本システムを使っている組織の効用を増大させる。不都合なことに、このような配置によって、MMSは、第三者が未承認のネットワークアクセスを使用してネットワークを取り囲んでいる機密保護ペリメーターを犯し、ネットワーク内に含まれているデータを盗むネットワークベースの攻撃（一般的に「ハッキング」と呼ばれている）を、非常に被り易くもなる。このような配置によって、MMSは、MMSがモニターしている同じ従業員による「ハッキング」攻撃を受け易くなる。

【0067】

別の実施形態では、MMSは、パーソナルコンピュータ装置上に局所的に展開され、局所的な記憶媒体の使用、パーソナルコンピュータ装置上で走っているアプリケーションによる分類されたデータの使用、及び装置に対するネットワーク通信に関する監視の実行を担当している。このようなシステムの配置によって、システムは、コンピューター装置を操作している人間がアクセスし使用する情報を例外的に閲覧できるようになり、そのシステムを使用している組織の効用を高める。しかしながら、このような配置によって、MMSは、MMSがモニターしている同じ従業員による「ハッキング」攻撃を受け易くなる。

【0068】

PMSの機密に関する懸念も、そのソフトウェアがMMSの使用するインデックスを作るために情報源に直接問い合わせる点で、高い。

【0069】

従って、或る実施形態ではネットワーク上のMMSの配置、或いは別の実施形態ではパーソナルコンピュータ装置上のMMSの配置が、MMSを攻撃に曝すことになる。或る実施形態では、これらの攻撃は、ローカルエリアネットワーク（LAN）の内側から、又は組織が維持しているWAN及び／又はインターネットリンクを通してLANの外側から来る。別の実施形態では、攻撃がパーソナルコンピュータ装置のユーザーから来ることもある。ここでの具体的な機密に関する懸念は、MMSが、保護しようとしているリレー

10

20

30

40

50

ショナルデータベースからの貴重なデータベースのデータを含んでいることである。懸念は、ハッカー又はパーソナルコンピューター装置のユーザーが、リレーショナルデータベースが実際に走っている、もっと徹底的にガードされたコンピューターからではなく、MMSからデータを盗もうとすることである。

【0070】

アプリケーションに対する第2のそして関係する機密に関する懸念は、MMSが、PMSが展開されているLANとは異なるLANで展開されている場合に発生する。先に述べたように、これは、データベースのデータを共有する2つの組織に亘って安全ポリシーを実施するのを助ける重要な構成である。MMSに記憶されている情報は、ここでも、情報安全の脅威に曝される。

10

【0071】

様々な実施形態が、これらの安全の脅威を直接取り扱う。ここに説明しているこれらの実施形態の画期的性の1つの態様は、PMS/MMSの対が、保護しようとしているデータのコピーを含んでいないインデックスを交換することである。上に述べたように、PMSは、MMSがポリシーを強制できるように、データベースのデータから導き出した抽象的なデータ構造をMMSへ送る。この保護を実現するために考えられる1つの方法は、単にデータベースをMMSにコピーするだけか、又は（同じく安全の観点から）内容がポリシーと矛盾していないことを確認するためにMMSがデータベースに直接問い合わせできるようにすることである。この方法の問題は、この方法が、それまでは無かった相当な安全の脆弱性を持ち込むことである。この危険な方法では、回復させることは病気より難しい。

20

【0072】

或る実施形態では、PMSは、データベースのデータのコピーを含んでいないか、又はデータベースのデータの暗号化された又はハッシュされたコピーだけを含んでいるデータベースからインデックスを作成する。そのようなインデックスは、データベースのデータのフラグメントに関係する多くのタプルを記憶するためのデータ構造を提供するタプル記憶機構を使って作成される。タプル記憶機構の例には、ハッシュ表、ベクトル、アレイ、ツリー、リスト又はリレーショナルデータベース管理システムの表が含まれる。以下に記載のプロセスでは、インデックスに記憶されているデータは、他の要素に対するデータベース内のその要素の相対位置を保持しているに過ぎない。例えば、ハッシュ表の場合、インデックスは、データベースのデータの各フラグメント（データベースセル内のデータフラグメント）毎に、フラグメントのハッシュコードを、その行番号、列番号及び列の型式と共に記憶している。

30

【0073】

この同じ解決法の他の実施形態は、保護されている知的財産のフラグメントを含んでいるインデックスを使用して、その情報を安全の脅威に曝すことにより解決法の価値を下けている。或る実施形態では、ここで具体的に述べている技法は、MMSを走らせるホストにハッカーが侵入した場合に、盗難に曝されるデータが非論理的となるように、データ自身のあらゆる表現を記憶しないようにしている。

【0074】

以下に述べるプロセスで説明する代替実施形態を実施すれば、機能を強化することができる。この代替実施形態では、システム内のデータの大部分を表しているデータベースからの頻繁に用いられるストリングと数字の少量のコピーだけが、インデックス内に、データベースの表内のデータの相対位置に関する情報の残りと共に、直接記憶される。これは、ハッシュコードの代わりに、これらの一般的なストリング自体のコピーを記憶することによって行われる。この代替方法では、システムは、（これらの一般的な用語に関して）行番号、列番号及びデータベースのデータの型式を記憶するが、ここではハッシュコードを記憶する代わりに、ストリング自体を記憶する。それほど一般的ではないデータベースの残りのセルでは、具体的にはこれらのストリングのコピーを記憶しないで、行番号、列番号及びデータベースのデータの型式だけが記憶される。この方法は、データベース内のストリングと数字のデータの統計的な分布は、最も一般的な用語が記憶されているデータ

40

50

の全体量の非常に大きな割合を占めるようにスキューされることが多いという事実を利用している。少数の一般的な用語が問い合わせの大部分を占めるので、これらの一般的な用語を別々のインデックスに記憶すれば、インデックスの問い合わせが効率的になり、これらの問い合わせは、文献（例えば、ハッシュ表ルックアップ、ビットマップなど）から標準的で迅速な技法を使って走らせることができる。これが安全上の脆弱性ではないという理由は、データベースのデータの量の不均衡な共有部分を占めるこの少数の用語は、最も価値の少ないデータであるからである。「ジョン」及び「スミス」という用語は、名前を含むデータベースの中では非常に一般的であるが、これらの用語の盗難は比較的価値が低い。この実施形態では、システムは、より高い値の、一般的でない用語のデータ（例えばクレジットカード番号、SSN、一般的でない名前など）のコピーの記憶を慎重に回避している。この実施形態では、先に述べた実施形態のように、システムは、データベース内のセルの配置に関する情報のハッシュコードとタプルのみを記憶することによって、機密情報のあらゆるコピーの記憶を回避する。

10

【0075】

事前選択されたデータの検出

或る実施形態では、事前選択されたデータの検出のプロセスは、索引付けと探索の2つの主なオペレーション又は段階を含んでいる。索引付け段階では、システムが、事前選択されたデータからインデックスを構築する。事前選択されたデータは、関係を表フォーマットに構成できるようなデータであればどのようなデータでもよい。つまり、事前選択されたデータは、表フォーマットで記憶される（例えば、リレーショナルデータベース内のデータ、エクセルのスプレッドシート内のデータなど）か、表フォーマットで記憶するのではないが、表フォーマットで記憶できるような関係（例えば、フラットファイル又はパスワードデータベース内にカンマ分離値として記憶されるデータ、オブジェクト志向データベース内のリレーショナルデータなど）を有している。

20

【0076】

図4は、事前選択されたデータに索引を付けるためのプロセスの1つの実施形態のフロー図である。このプロセスは、ハードウェア（回路、専用論理など）、（汎用コンピュータシステム又は専用機上を走っているような）ソフトウェア又はそれらの組み合わせを備えた処理論理によって実行される。

【0077】

図4に示すように、処理論理は、事前選択されたデータが標準的表フォーマットに記憶されているか否かを判断する段階（処理ボックス402）で始まる。記憶されていないければ、処理論理は、事前選択されたデータを標準的表フォーマットに変換する（処理ブロック404）。出来た表内の各セルは、事前選択されたデータのフラグメントを記憶している。或る実施形態では、各データフラグメントはトークンである。トークンは、単一の単語でも単語のクラスター（例えば、引用符でくくられた単語）でもよい。例えば、「this」という単語はデータベースのセル内に記憶されているトークンを表すが、「this token」という句は、単一のストリングとしてデータベースのセル内に記憶されている場合、独立したトークンを表す。

30

【0078】

次に、処理論理は、事前選択されたデータから導き出された、タプル記憶構造を作成する（処理ブロック406）。タプル記憶構造は、事前選択されたデータのフラグメントと関連付けられた多くのタプルを記憶するための機構を提供する。タプル記憶構造の例には、ハッシュ表、ベクトル、アレイ、ツリー又はリストが含まれる。タプル記憶構造の各型式は、何れかの所与の内容フラグメントに関して一式のタプル（タプル記憶構造内に一致するものが無ければ、その一式のタプルは空となる）を検索するための方法に関係付けられている。

40

【0079】

更に、処理論理は、対応するタプルのデータベース内の各データフラグメントの位置に関する情報を記憶する（処理ブロック408）。或る実施形態では、データフラグメント

50

の位置に関する情報は、データベース内のデータフラグメントを記憶している行の番号を含んでいる。別の実施形態では、この情報は、データベース内のデータフラグメントを記憶している列の番号と、随意的に列のデータ型式も含んでいる。

【0080】

その後、処理論理は、タプルを所定の順序（例えば、昇順辞書式順序）にソートする（処理ブロック410）。

【0081】

従って、出来上がった抽象的データ構造（即ち、インデックス）は、大きな全体の文脈内におけるデータ記録の相対位置に関する情報を含んでいるだけであり、事前選択されたデータ自体のフラグメントは何も含んでいない。

10

【0082】

或る実施形態では、インデックスの内容は、更にインデックスを盗難から守るために、暗号法的に（例えば、ハッシュ関数で、又は暗号キーを備えた暗号関数を使って）取り扱われている。

【0083】

事前選択されたデータの検出プロセスの「探索」段階について、更に詳しく論じる。図5は、事前選択されたデータに関する情報内容を探索するためのプロセスの1つの実施形態のフロー図である。このプロセスは、ハードウェア（回路、専用論理など）、（汎用コンピュータシステム又は専用機上を走っているような）ソフトウェア、又は両方の組み合わせを備えた処理論理によって実行される。

20

【0084】

図5に示すように、処理論理は、情報内容を受け取る段階で始まる（処理ブロック502）。情報内容は、ファイル（例えば、コンピュータのハードドライブに記憶されているアーカイブ済eメールメッセージ）、又は、ネットワークで送信されたデータのブロック（何れかの型式のネットワークプロトコルを使って、ネットワーク上を送信されたeメールメッセージ）に含まれている。

【0085】

次に、処理論理は、情報内容の中で、事前選択されたデータの一部を含んでいる可能性のある内容フラグメントのシーケンスを検出する（処理ブロック504）。先に述べたように、事前選択されたデータは、保護する必要のある所有権のあるデータベースのデータであるか、固有の表構造を有する何か他の種類のデータである。つまり、事前選択されたデータは、表フォーマットで記憶してもよい（例えば、リレーショナルデータベース内のデータ、エクセルのスプレッドシート内のデータなど）し、表フォーマットで記憶するのではないが、表フォーマットで記憶できるような関係（例えば、フラットファイル又はパスワードデータベース内にカンマ分離値として記憶されているデータ、オブジェクト志向データベース内のリレーショナルデータなど）を有しているものでもよい。

30

【0086】

或る実施形態では、検出された内容フラグメントのシーケンスは、情報内容内の一式の隣接するトークンである。各トークンは、単語又は句に対応する。検出された内容フラグメントのシーケンスは、受け取った情報内容の一部か又は情報内容全体である。

40

【0087】

或る実施形態では、処理論理は、内容フラグメントのシーケンスが列フォーマット済データに似ていると判定する際に、内容フラグメントのシーケンスが、事前選択されたデータの一部を含んでいる可能性があるかと判断する。この判定は、受信した情報内容を解析して分離線を識別し（例えば、タグ< c r >又は< c r >< l f >で表示される）、これらの分離線が同数のトークンと、随意的に同様のトークンのデータ形式を含んでいることを見つけ出すことによって、行われる。

【0088】

別の実施形態では、処理論理は、全情報内容を解析し、事前選択されたデータに関して隣接するトークンのブロックを探索する際に、内容フラグメントのシーケンスが、事前選

50

扱されたデータの一部を含んでいる可能性がある」と判断する。或る実施形態では、隣接するトークンのブロックは、各ブロックのユーザーが指定した幅、及び、情報内容の中の各ブロックのユーザーが指定した位置（例えば、ユーザーは、2つの隣接するブロックを、或る数のトークンで分離することを要求する）の様なユーザーが指定したパラメーターに基づいて定義される。

【0089】

更に別の実施形態では、処理論理は、情報内容の中に事前に定義されたフォーマットの表現を見つけ出す際に、内容フラグメントのシーケンスが、事前選択されたデータの一部を含んでいる可能性がある」と判断する。そのような表現は、例えば、口座番号、社会保障番号、クレジットカードの番号、電話番号、郵便番号、eメールアドレス、金融値又は数値を示すテキストフォーマット（例えば、数字を伴う「\$」印）などである。この表現が見つけれ

10

【0091】

更に別の実施形態では、処理論理は、受信した情報内容に関係付けられた或る特性が、以前の違反の履歴に基づいて、情報内容の中に事前選択されたデータを含んでいる可能性があることを示していると判定する際に、内容フラグメントのシーケンスが、事前選択されたデータの一部を含んでいる」と判断する。これらの特性には、例えば、情報内容の宛先（例えば、電子メッセージの受信者）、情報内容の起点、情報内容に関係付けられた送信の時間、情報内容に関係付けられた送信のサイズ、送信に含まれているファイルの型式（例えば、多目的インターネットメールエクステンション（MIME）型のファイル）などが含まれる。或る実施形態では、以前の違反の履歴は、事前選択されたデータの検出の度に、事前選択されたデータが検出された情報内容の特性を識別し、これらの特性を以前の違反のデータベース内に記録することによって保持される。その後、処理論理は、新しい情報内容の中の内容フラグメントのシーケンスが事前選択されたデータの一部を含んでいるか否かを判定するときに、新しい情報内容の特性を識別し、これらの特性について以前の違反のデータベースを探索する。一致が見つければ、処理論理は、一致特性に関係付けられた以前の違反が、新しい情報内容の中に事前選択されたデータが含まれている可能性を表示しているか否かを判定する。この表示は、一致特性に関係付けられた以前の違反の数、又は一致特性に関係付けられた以前の違反の頻度に基づいていてもよい。例えば、この表示は、特定の送信者が犯した違反の合計数に基づいていてもよいし、それらの違反の所与の期間に亘る頻度に基づいていてもよい。

20

30

【0092】

その後、事前選択されたデータの一部を含んでいる可能性がある内容フラグメントのシーケンスを検出する際に、処理論理は、これらの内容フラグメントの何れかのサブセットが、事前選択されたデータのサブセットと一致するか否かを判定する（処理ブロック506）。この判定は、事前選択されたデータの表構造を定義するインデックス（ここでは抽象的データ構造と呼ぶ）を使って行われる。

【0093】

図6は、事前選択されたデータから導き出された抽象的データ構造内の内容フラグメントのサブセットに対する一致を見つけ出すためのプロセスの1つの実施形態のフロー図である。このプロセスは、ハードウェア（回路、専用論理など）、（汎用コンピューターシステム又は専用機上を走っているような）ソフトウェア又はそれらの組み合わせを備えた論理を処理することによって実行される。

40

【0094】

図6に示すように、処理論理は、図5の処理ブロック504で識別された内容フラグメントのシーケンスを内容フラグメント（例えば、トークン）で解析する段階で始まる。次いで、処理論理は、各内容フラグメント毎に、一式の一致タプルに対して抽象的データ構造を探索する（処理ブロック602）。例えば、情報内容に含まれている「Smith」

50

という単語が、抽象的データ構造内に反映されている事前選択されたデータ内に複数回発生しているかもしれない。具体的には、これらの発生のそれぞれは、抽象的データ構造内に対応するタプルを有している。処理論理は、探索の間に、事前選択されたデータ内の「Smith」という単語の発生に対応する一式のタプルを検索する。各タプルは、事前選択されたデータを記憶しているデータベース又は表の中のこのデータフラグメントの位置に関する情報を記憶する。或る実施形態では、位置情報は、データフラグメントを記憶しているセルの行番号を含んでいる。別の実施形態では、位置情報は、このセルの列番号と、随意的にその列のデータ型式も含んでいる。

【0095】

次に、処理論理は、全内容フラグメントで見つけ出された一致タプルのセットを組み合わせて（処理ブロック604）、次に、組み合わせられた一致タプルのセットを、行番号でグループLに分類する（処理ブロック606）。その結果、各グループL（ここではアキュムレータと呼ぶ）は、全てが同じ列番号を有する一致タプルのセットを含み、即ち、各グループLの一致タプルのセットは、全てがデータベース内の同じ行から出ているように見える事前選択されたデータのフラグメントに対応している。

10

【0096】

更に、処理論理は、各グループに含まれている一致タプルのセットの数によってグループLをソートし（処理ブロック608）、或る実施形態では、際立った列の数を備えたタプルのセットを有するグループを選択する（処理ブロック610）。その後、処理論理は、選択されたグループが十分に大きい数の一致タプルのセットを有しているか否かを判断する（処理ブロック612）。例えば、1つのグループの一致タプルのセットの数が「3」を上回っている場合、情報内容が、データベース内の同じ行の4つ又はそれ以上の列からのデータを含んでいる公算が高い。

20

【0097】

探索プロセスの代表的な実施形態について述べる。図7A-7Cは、事前選択されたデータのハッシュ表インデックスを使って入信してくるメッセージを探索するためのプロセスの代替実施形態のフロー図である。このプロセスは、ハードウェア（回路、専用論理など）、（汎用コンピューターシステム又は専用機上を走っているような）ソフトウェア又はそれらの組み合わせを備えた処理論理によって実行される。

【0098】

図7Aに示すように、処理論理は、入信してくるメッセージを解析する段階で始まる（処理ブロック702）。次に、処理論理は、入信してくるメッセージの解析された部分が列フォーマットされたデータを含んでいるか否か判定する（処理ブロック704）。或る実施形態では、語彙解析を使用して、（例えば、ラインを分離するのに使用されるタグ<cr>又は<cr><lf>を見つけ出すことによって）入信してくるメッセージの解析された部分のラインを識別し、隣接するラインの中で見つけ出されたトークンの数が数と型式において同じであることを検出する。或る実施形態では、処理論理は、各トークンの型式をトークンの総数と共に記憶する。

30

【0099】

処理ボックス704でなされた判断が否であれば、処理は、処理ブロック702に移る。そうでなければ、処理は処理ブロック706へ移り、そこで処理論理がiを、列フォーマットされたデータに似ている最初のラインと等しくなるよう設定する。

40

【0100】

次に、処理論理は、ラインi内の各トークンにハッシュ関数H(k)を適用し（処理ブロック708）、ラインi内の各トークンに関しハッシュ表内のH(k)で一式のタプルを見つけ出し、そのタプルをリストLに加え、リストLを、各アキュムレータのタプルが同じ行数値を有する一式のアキュムレータに再分類する（処理ブロック712）。更に、処理論理は、各Aiの長さによってそのリストLをソートし（処理ブロック714）、ソートされたリストL内の固有の列の出現を確認する（処理ブロック716）。処理ブロック710では、随意的な事前処理論理を実行してリストLに挿入する前にトークンをろ過し

50

、元のトークン k の語彙型式に一致する型式を備えたタプルだけが L に加えられるようにする。実施形態の中には、固有の列の出現を確認する段階が速度又は簡潔さの理由で省略されるものもある。更に別の実施形態では、タプルは、行番号のみを含んでいる（即ち、列番号も型式のインジケータも含んでいない）単なる「単集合」である。

【0101】

その後、入信してくるメッセージが、列フォーマットされたデータに似たもっと多くのラインを含んでいる場合（処理ブロック718）、処理論理は、列フォーマットされたデータと似た次のラインに i を逐増し（処理ブロック722）、処理は処理ブロック706へ移る。そうでない場合、処理論理は、所定のサイズを上回り且つ固有の列番号を有している A_i を備えたテキストのラインを報告する（処理ブロック720）。 10

【0102】

図7Bに示すように、処理論理は、ユーザーが指定した「幅」（ W ）と「ジャンプ」（ J ）のパラメーターを受け取り（処理ブロック732）、入信してくるメッセージを解析する（処理ブロック734）段階で始まる。パラメーター W は、一回繰り返す間に、探索することになる隣接するトークンの各ブロック内の隣接するトークンの数を指定し、パラメーター J は、2つの隣接するブロックの間の必要なトークン数を指定する。

【0103】

次に、処理論理は、位置変数（ S_t ）の値をゼロに設定し（処理ブロック736）、 S_t で始まる W の隣接するメッセージのトークンを集めることによって、探索するブロック（「テキストブロック」）を定義する（処理ブロック738）。 20

【0104】

更に、処理論理は、テキストブロック内の各トークンにハッシュ関数 $H(k)$ を適用し（処理ブロック740）、テキストブロック内の各トークン毎にハッシュ表内の $H(k)$ で一式のタプルを見つけ出し、テキストブロック内の対応するトークンと同じ型式を有するタプルをリスト L に追加し（処理ブロック742）、リスト L を、一式のアキュムレータに再分類し（処理ブロック744）、各 A_i の長さでそのリスト L をソートし（処理ブロック746）、ソートされたリスト L 内の、固有の列の出現を確認する（処理ブロック748）。

【0105】

その後、処理論理は、トークンの J 数だけ S_t を漸増し（処理ブロック750）、位置 S_t がなおメッセージ内にあるか否か判定する（処理ブロック752）。判断が正であれば、処理は処理ブロック738へ移る。そうでない場合、処理論理は、所定のサイズを上回り且つ固有の列番号を有している A_i を備えているテキストブロックを報告する（処理ブロック758）。 30

【0106】

図7Cに示すように、処理論理は、入信してくるメッセージを解析し（処理ブロック764）、ユーザーの指定したフォーマットを有する最初の表現を探す（処理ブロック766）段階で始まる。そのような表現は、例えば、口座番号、社会保障番号、クレジットカード番号、金融値又は数値を示すテキストフォーマット（例えば、数字を伴う「\$」印）などである。一致表現が見付からなければ、処理は処理ブロック764へ移る。見つけ出されれば、処理は、処理ブロック768へ移り、そこで処理論理は、一致表現の前後の W の隣接するトークンを集めることによって、探索するブロック（「テキストブロック」）を定義する。例えば、テキストブロックは、一致表現の直前の10個のトークンと、一致表現自体と、一致表現の直後の10個のトークンで構成される。 40

【0107】

更に、処理論理は、テキストブロック内の各トークンにハッシュ機能 $H(k)$ を適用し（処理ブロック770）、テキストブロック内の各トークン毎にハッシュ表内の $H(k)$ で一式のタプルを見つけ出し、テキストブロック内の対応するトークンと同じ型式を有するタプルをリスト L に追加し（処理ブロック772）、リスト L を、一式のアキュムレータに再分類し（処理ブロック774）、各 A_i の長さでそのリスト L をソートし（処理ブ 50

ロック 776)、ソートされたリスト L 内の、固有の列の出現を確認する(処理ブロック 778)。

【0108】

その後、処理論理は、メッセージがユーザーの指定したフォーマットの表現をそれ以上有しているか否かを判定する(処理ボックス 780)。判定が正であれば、処理は処理ブロック 768 へ移る。そうでない場合、処理論理は、所定のサイズを上回り且つ固有の列番号を有している A_i を備えているテキストブロックを報告する(処理ブロック 782)。

【0109】

代表的なアプリケーション

或る実施形態では、通常のオペレーションの途上では、(保護が必要な記録が常駐している)組織のデータベースとの安全な通信が行えるように、PMS は会社のネットワーク上に配置されていると想定されている。通常のオペレーションの途上では、更に、MMS が、組織の全ての外部との e メール通信をモニター及び/又は遮断できるように配置されていると想定されている。

【0110】

この例では、組織が、1) 名前、2) 姓、3) クレジットカード番号、4) 残高の 4 つの列が含まれている「カスタマーレコード」と呼ばれるデータベース表を保護しようとしていると想定する。この組織の従業員は、PMS が提供するユーザーインターフェースアプリケーションを使用して、カスタマーレコード表が e メールによる盗難に対する保護を必要としていると指定することになる。すると、PMS は、データベース内のセルのストリング値から導き出されたハッシュ表で構成されているカスタマーレコード表内の記録のインデックスを作る。つまり、セル内の数値を使ってハッシュ表内の数値を調べる。ハッシュ表自体には、各行番号、列番号及びセル自体のデータ型式の記録が含まれている。ハッシュ表にしばしば見られる衝突の場合、「衝突リスト」は、行番号、列番号及び型式に関するそのような記録を多数保持している。データベース表内の全セルがそのような構造にハッシュされると、インデックスが作成され、MMS への送信の準備が整う。インデックスにはデータベースのデータ自体の記録は含まれていないことに注目されたい。これは、このシステムが満たす重要な安全上の制約である。

【0111】

MMS は、インデックスを受け取った後で、メッセージを解析し、メモリ内のハッシュ表を、PMS で作成されたのと同じ様式で再度作成する。

【0112】

MMS は、外部との e メールメッセージをピックアップしてそれを解析する際には、以下に説明するやり方でこのインデックスを使用し、これらの eメールの何れかに、データベースからのデータが含まれているか否かを検出する。これは、eメールメッセージからのテキストの個々のラインを解析することによって行われる。これには、周囲のファイル型式をデコードする段階と、全てのものを生のテキストに変換する段階(例えば、マイクロソフトのワードファイルから全てのフォーマット情報を剥ぎ取り、テキスト自体のみを残す)が含まれている。テキストのこの一連のラインは、「スペース」符号又は他の形態の句読点の様な分離マークを探すことによって個々の単語で解析される。これらの単語はテキストのトークンである。このシステムは、テキストトークンの各ラインに対して、ハッシュ関数を各トークンに適用することによって、インデックスを調べる。このオペレーションの結果が、そのライン上の各トークンに関するハッシュ表の衝突リストとなる。先に説明したように、各衝突リストは、それ自体が、可能性のある行番号、列番号及び型式のトリプレットを記憶する一式のデータ要素である。全トリプレットのユニオンが全ての衝突リストから取られていれば、そして一式のトリプレットが、同じ行番号だが異なる列番号を備えていることが分かれば、高い確率で、eメールメッセージからのテキストのこのラインにはデータベースからの記録が含まれている。なお、ここで使用する「タプル」という用語は、行番号、列番号及び型式のトリプレットの特定の場合限定されず、これら 3 つのパラメーターの全ては含まれていないデータ構造を指す。例えば、或る実施形態

では、或るタプルには、行番号は含まれているが、列番号とデータベースのデータの型式は含まれていない。

【0113】

先行技術との比較

データベース問い合わせ機構は、ここで説明している教示と大幅に異なっている。1つの相違点は、Bツリーには、実際に、Bツリーがインデックスを付けるデータベース表のフラグメントが含まれていることである。上記の方法では、インデックスの内側にはデータベースのデータのコピーは記憶されていない。これが重要なのは、上記の通り、MMSは、データを漏洩から守るためにインデックスのコピーを持たなければならないが、同時に、相当な脅威に曝されるネットワーク内の位置に最も良く展開されるからである。MMSが使用するインデックスを、データベースのデータの何れの構成要素からも自由に保つことが、重要な要件である。

10

【0114】

標準的なデータベース問い合わせ機構とここに概説する本発明との間のもう1つの相違点は、必要な問い合わせの型式と関係がある。リレーショナルデータベースに用いられる標準的な問い合わせのセットは、AND又はORのような連結語を使用する述語論理に基づいている。この基本システムは、通常、eメール及びウェブメールのメッセージへとカットアンドペーストされるデータベースのデータを検出するには上手く働かない。eメールメッセージへとカットアンドペーストされるデータベースのデータは、通常、レポートからのもので、各ラインに、異質で、データベース表の内側には見られないデータが含まれていることが多い。一つの例は、例えば、一群の消費者に対する会計情報が含まれているeメールメッセージである。そのようなメッセージには、例えば、名前、姓、社会保障番号など、保護を必要とするコアデータベースからの多量の記録が含まれているが、コアデータベース表に無い情報も含まれている。代表的例は、他のデータベースから「連結された」情報である。もう1つの例は、データベースのデータのフィールドを分離する単純なラインフォーマットトークンである。これらの各ライン上で見られるこの余分なデータの可能性の故に、出て行くメッセージのライン上の各トークンに適用されるAND及びORのような標準的な述語論理の接続語が、(ORの場合)過剰なヒットか、(ANDの場合)ゼロヒットを作り出す。ここでの説明では、本システムは、nがライン内のトークンの総数より大幅に少なくとも、全てデータベース表の同じ行からのn個又はそれ以上のトークンの存在を検出できる。これは、本発明と、データベース及び文書問い合わせ機構に関する上記先行技術との間のもう1つの重要な相違点である。

20

30

【0115】

上記技法と情報検索技術の間には、幾つかの重大な相違点がある。第1に、これらのシステムのインデックスは、保護対象のデータベース内に記憶されている同じ用語を(用語索引内に)含んでいる。ここでも、システムはハッカーの脅威を受ける可能性のあるネットワーク上の位置にこのインデックスを展開するので、明らかな欠点となる。第2に、これらの問い合わせシステムは、AND及びORのような述語論理の形態を使って、ブール問い合わせを実行する。先に述べたように、この方法は、他の表からの異質なデータと「連結されている」可能性のあるデータベースの記録を検出するには、明らかに不利である。

40

【0116】

ファイルシングリングの技法は、ここに述べる技法と似ているが、実質的に異なっている。ファイルシングリングでは、関心事の主体はテキストデータ(散文、ソフトウェア、概要など)である。ここに述べる技法では、データベースのデータを保護することに焦点が絞られている。1つの相違点は、所与のデータベース表からのデータベースのデータは、試験メッセージで任意に並べ換えられる行順又は列順で現われることである。これらの並べ換えは、通常、データベースのデータを抽出するために適用される問い合わせ機構の単純な結果である。データベースの問い合わせは、任意の列順、及び任意の行順で出てくるデータベースのデータのブロックになる。このため、ファイルシングリングの基本的な

50

技法をデータベースのデータに適用しても、働かない。ファイルシングリングは、保護されている文書と試験文書との間に同じ線形シーケンスが続くことを前提としている。

【0117】

インターネットの内容ろ過システムとここに述べる教示との間には、多くの重要な相違点がある。先に述べたように、インターネットの内容ろ過システムは、キーワード探索に基づいている。上に述べた最新の技法は、保護したいデータベースから、抽象的データ構造を構築する。この抽象的データ構造は、保護しようとしているテキストのフラグメントを含んでいない。キーワードろ過システムは、探索しているテキストの幾つかの表現を、そのテキストに関する問い合わせを実行するために、含んでいなければならない。第2の重要な相違点は、これらのインターネットの内容ろ過システムには、データベースのデータを保護する意図がないことである。データベースのデータに関する組織の機密ポリシーに対する違反を検出するために通常の表現一致を使用すると、検出の方法が非常に不正確になる。これらのシステムは、主に、インターネットがポルノ又は虐待的な内容及び言語に関係している際に、従業員がインターネットを悪用するのを止めるのに利用される。そのようなシステムは、データベースのデータの保護に適用すると、データベースの記録と整合を取るのに、通常の表現を使用する。このことも、データベースのデータのフラグメントを、安全の危険性が最大であるネットワーク上のコンピューターに伝送することになる。

10

【0118】

代表的なコンピューターシステム

20

図8は、ここで述べた1つ又は複数のオペレーションを実行する代表的なコンピューターシステムのブロック図である。図8に示すように、コンピューターシステム800は、代表的なクライアント850又はサーバー800のコンピューターシステムを備えている。コンピューターシステム800は、情報を伝達するための通信機構又はバス811と、情報を処理するためにバス811に連結されているプロセッサ812とを備えている。プロセッサ812は、限定するわけではないが、例えばPentiumTM、PowerPCTM、AlphaTMなどのようなマイクロプロセッサを含んでいる。

【0119】

システム800は、ランダムアクセスメモリ(RAM)、又は、プロセッサ812によって実行される情報及び指示を記憶するためのバス811に連結されている他のダイナミック記憶装置804(メインメモリと呼ばれている)を更に備えている。メインメモリ804は、プロセッサ812が指示を実行している間に、一時的な変数又は中間情報を記憶するのに用いられる。

30

【0120】

コンピューターシステム800は、読み取り専用メモリ(ROM)、及び／又はプロセッサ812用のスタティック情報及び指示を記憶するためにバス811に連結されている他のスタティック記憶装置806と、磁気ディスク又は光ディスクのようなデータ記憶装置807及びその対応するディスクドライブを更に備えている。データ記憶装置807は、情報及び指示を記憶するためバス811に連結されている。

【0121】

40

コンピューターシステム800は、更に、情報をコンピューターのユーザーに表示するためバス811に連結されている陰極線管(CRT)又は液晶表示装置(LCD)の様な表示装置821に連結されている。英数字及び他のキーを含んでいる英数字入力装置822も、プロセッサ812に情報及びコマンド選択を伝達するためバス811に連結されている。追加のユーザー入力装置は、指示情報及びコマンド選択をプロセッサ812に伝達し、ディスプレイ上のカーソルの動きを制御するためバス811に連結されている、マウス、トラックボール、トラックパッド、スタイラス又はカーソル方向キーの様なカーソルコントロール823である。

【0122】

バス811に連結されているもう1つの装置は、紙、フィルム又は同様の型式の媒体の

50

様な媒体上に指示、データ又は他の情報を印刷するために用いられるハードコピー装置 824 である。更に、スピーカー及び／又はマイクロホンの様な音声記録及び再生装置も、コンピューターシステム 800 と音響的インターフェースを取るために、随意的にバス 811 に連結される。この他、バス 811 に連結される装置には、電話又は携帯装置と通信するための有線／無線通信機器 825 がある。

【0123】

なお、システム 800 及び関連ハードウェアの構成要素の何れか又は全てを、本発明で用いることができる。しかしながら、他の構成のコンピューターシステムが、本装置の一部又は前部を含んでいてもよい。

【0124】

以上の説明を読んだ後では、当業者には、本発明に関する多くの変更及び修正が疑いもなく明白になったことであろうが、分かり易くするために図示し説明している具体的な実施形態は、本発明を限定するものではない。従って、様々な実施形態の詳細についての言及は、本発明に必須であると見なされる特徴のみを列挙している請求項の範囲を限定する意図はない。

【図面の簡単な説明】

【0125】

【図1】ワークフローの1つの実施形態を示している。

【図2A】代表的なオペレーションのモードを示している。

【図2B】代表的なオペレーションのモードを示している。

【図3】データベースのデータを保護するためのプロセスの1つの実施形態のフロー図である。

【図4】データベースのデータに索引を付けるためのプロセスの1つの実施形態のフロー図である。

【図5】事前選択されたデータの情報内容を探索するためのプロセスの1つの実施形態のフロー図である。

【図6】事前選択されたデータから導き出された抽象的データ構造内の内容フラグメントのサブセットに一致するものを見つけ出すためのプロセスの1つの実施形態のフロー図である。

【図7A】事前選択されたデータのハッシュ表インデックスを使って、入信してくるメッセージを探索するためのプロセスの代替実施形態のフロー図である。

【図7B】事前選択されたデータのハッシュ表インデックスを使って、入信してくるメッセージを探索するためのプロセスの代替実施形態のフロー図である。

【図7C】事前選択されたデータのハッシュ表インデックスを使って、入信してくるメッセージを探索するためのプロセスの代替実施形態のフロー図である。

【図8】本明細書で説明している1つ又は複数のオペレーションを実行する代表的なコンピューターシステムのブロック図である。

【図9】事前選択された機密データをクライアントベースで保護するためのシステムの1つの実施形態のブロック図である。

【図10】事前選択された機密データをクライアントベースで保護するためのプロセスの1つの実施形態のフロー図である。

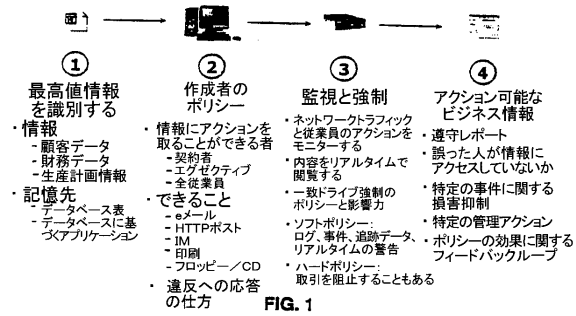
10

20

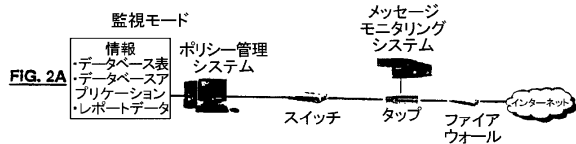
30

40

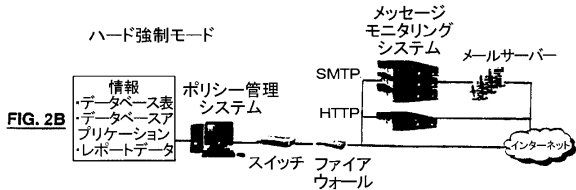
【図 1】



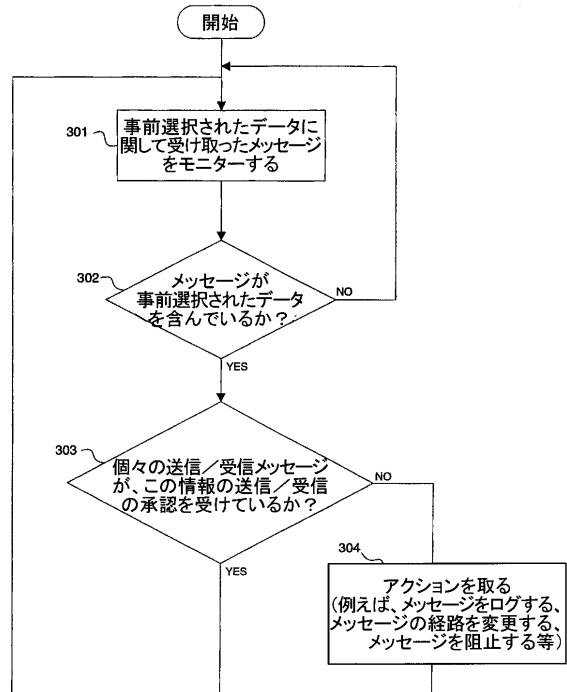
【図 2 A】



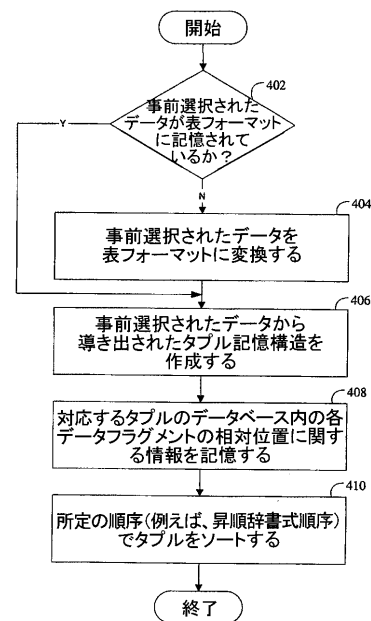
【図 2 B】



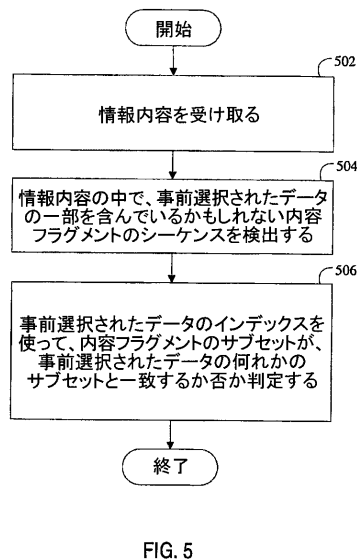
【図 3】



【図 4】



【図 5】



【図 6】

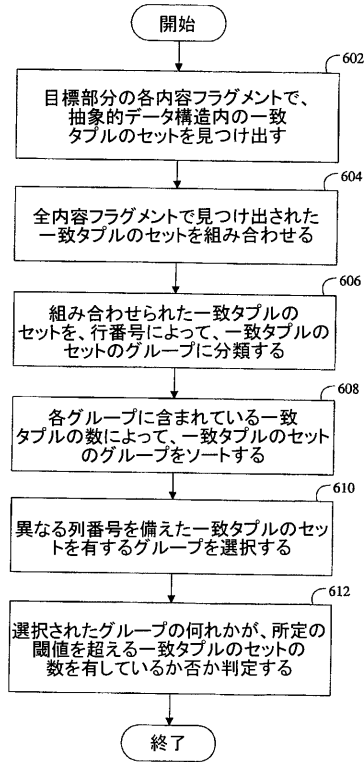


FIG. 6

【図 7 A】

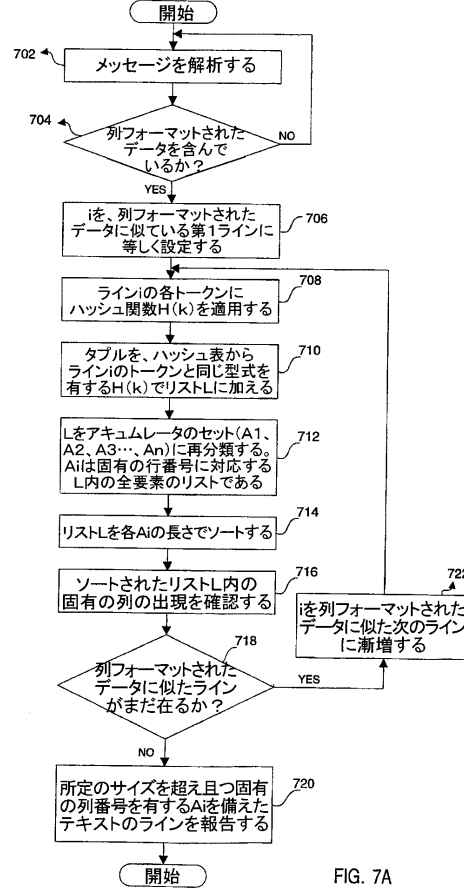


FIG. 7A

【図 7 B】

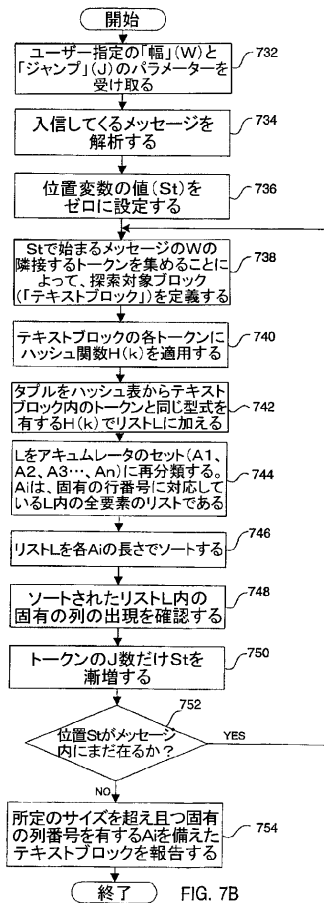


FIG. 7B

【図 7 C】

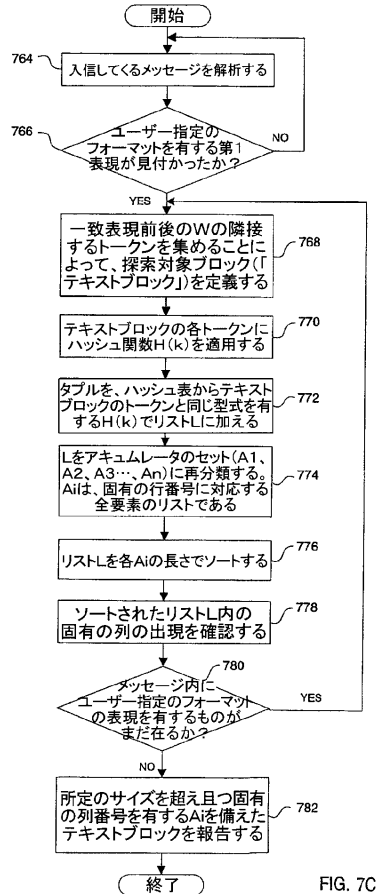


FIG. 7C

【図 8】

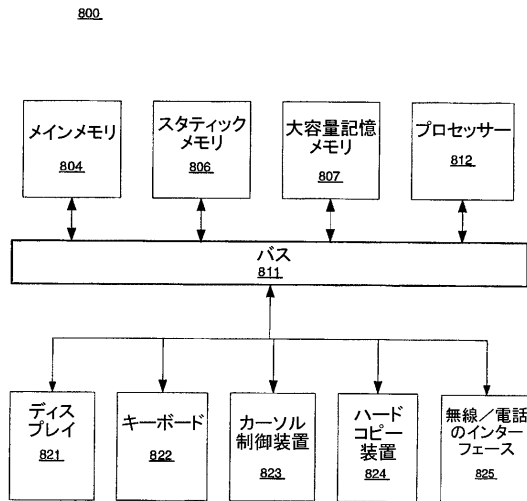


FIG. 8

【図 9】

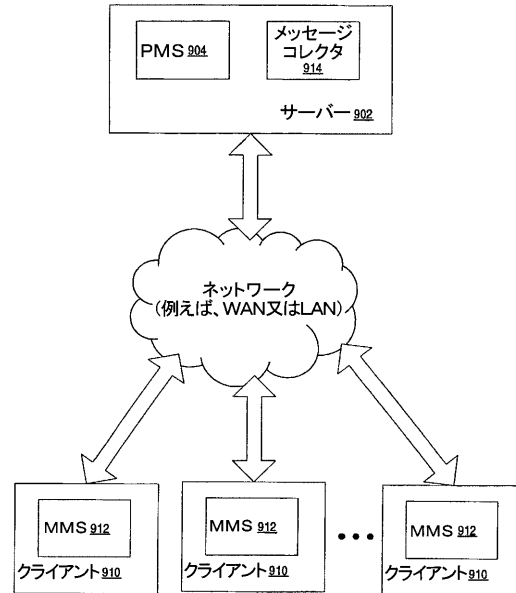


FIG. 9

【図 10】

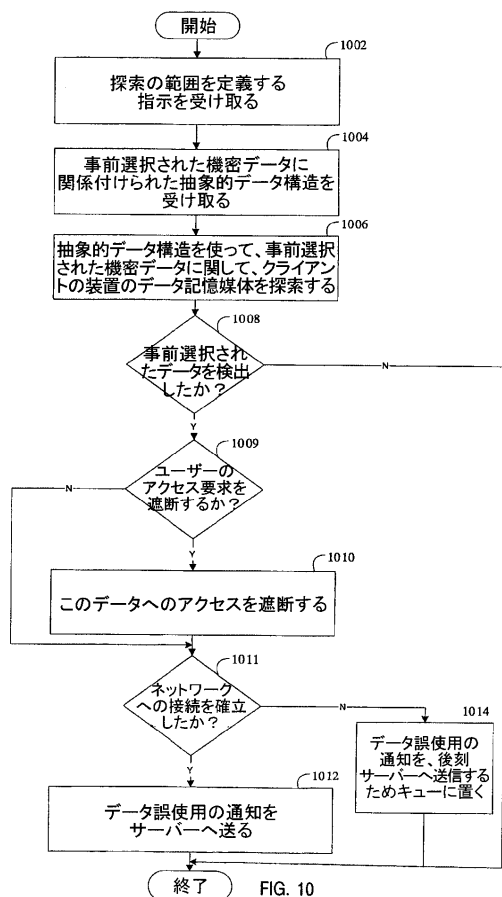


FIG. 10

フロントページの続き

(31)優先権主張番号 10/607,718

(32)優先日 平成15年6月27日(2003.6.27)

(33)優先権主張国 米国(US)

(72)発明者 ウルフ マイケル アール

アメリカ合衆国 カリフォルニア州 94123 サン フランシスコ オクタヴィア ストリート 2511

(72)発明者 ゴバラクリシュナン ミシリー

アメリカ合衆国 カリフォルニア州 94131 サン フランシスコ サーティース ストリート 313 #3

(72)発明者 フリードマン ヴィタリー

アメリカ合衆国 カリフォルニア州 94002 ベルモント サウス ロード 510

(72)発明者 アンサネリ ジョセフ

アメリカ合衆国 カリフォルニア州 94127 サン フランシスコ モントレー ブールヴァード 1700

審査官 宮司 卓佳

(56)参考文献 特開2002-189643 (JP, A)

米国特許第06442607 (US, B1)

米国特許第05701464 (US, A)

(58)調査した分野(Int.Cl., DB名)

G06F 21/22

G06F 12/00

G06F 13/00