(19)日本国特許庁(JP)　(12)公開特許公報(A)　(11)特許出願公開番号
特開2012-164031
(P2012-164031A)
(43)公開日　平成24年8月30日(2012.8.30)

(51)Int.Cl.		F I			テーマコード（参考）
G06F 17/30	(2006.01)	G06F	17/30	320C	5B017
G09C 1/00	(2006.01)	G06F	17/30	170Z	5B075
G06F 21/24	(2006.01)	G06F	17/30	220C	5J104
		G09C	1/00	660D	
		G06F	12/14	540A	

審査請求 未請求　請求項の数 13　ＯＬ　（全 22 頁）

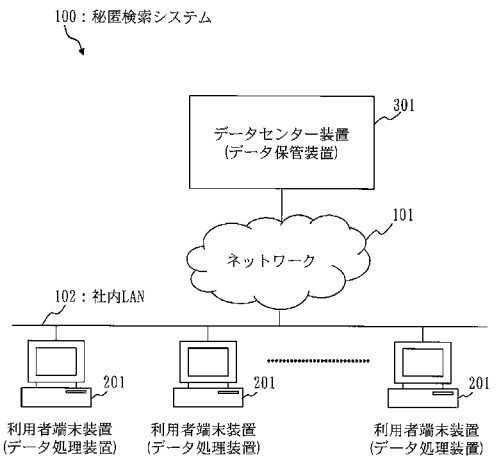
(21)出願番号　特願2011-22147(P2011-22147)
(22)出願日　平成23年2月3日(2011.2.3)

(71)出願人　000006013
三菱電機株式会社
東京都千代田区丸の内二丁目７番３号
(74)代理人　100099461
弁理士　溝井　章司
(74)代理人　100152881
弁理士　山地　博人
(72)発明者　森　拓海
東京都千代田区丸の内二丁目７番３号　三菱電機株式会社内
(72)発明者　松田　規
東京都千代田区丸の内二丁目７番３号　三菱電機株式会社内

最終頁に続く

(54)【発明の名称】データ処理装置及びデータ保管装置及びデータ処理方法及びデータ保管方法及びプログラム

(57)【要約】
【課題】確率的暗号を利用した秘匿検索を安全に高速化する。
【解決手段】利用者端末装置２０１は、保管対象の文書情報に対して指定された保管キーワードから一意に得られる乱数値をエントロピー符号化して、保管対象の文書情報に対応付けられるタグがデータセンター装置３０１で保管される際にタグに付される保管索引値を生成し、保管対象の文書情報とタグと保管索引値をデータセンター装置３０１に送信する。データセンター装置３０１では、保管索引値を用いてタグを分類、索引化することができ、確率的暗号を利用した秘匿検索を安全に高速化することができる。
【選択図】図１



【特許請求の範囲】
【請求項１】
複数の暗号化データと、各暗号化データに対応付けられている、暗号化データの検索の際に照合されるタグデータとを保管するデータ保管装置に接続され、
前記データ保管装置での保管の対象となる保管対象データのキーワードを保管キーワードとして指定するキーワード指定部と、
前記保管キーワードから一意に得られる乱数値をエントロピー符号化して、前記保管対象データの暗号化データに対応付けられるタグデータが前記データ保管装置で保管される際に前記タグデータに付される索引値を生成する索引生成部と、
前記保管対象データの暗号化データと前記タグデータと前記索引値とが含まれる保管要求を、前記データ保管装置に対して送信する通信部とを有することを特徴とするデータ処理装置。
【請求項２】
前記データ処理装置は、
秘匿化された検索キーワードと、前記秘匿化された検索キーワードとの照合の対象となるタグデータの索引値とが含まれる検索要求を受信した際に、前記検索要求に含まれる索引値と一致する索引値と対応付けられているタグデータを抽出し、抽出したタグデータと前記秘匿化された検索キーワードとを照合して暗号化データを検索するデータ保管装置に接続され、
前記索引生成部は、
前記データ保管装置において前記検索要求に含まれる索引値と比較される索引値を生成することを特徴とする請求項１に記載のデータ処理装置。
【請求項３】
前記データ処理装置は、
暗号化データとタグデータと索引値とを対応付けて保管するデータ保管装置に接続され、
前記キーワード指定部は、
前記データ保管装置に暗号化データの検索を行わせる検索キーワードを指定し、
前記データ処理装置は、更に、
前記検索キーワードを秘匿化する検索キーワード秘匿化部を有し、
前記索引生成部は、
前記検索キーワードから一意に得られる乱数値をエントロピー符号化して、前記データ保管装置において、秘匿化された検索キーワードとの照合の対象となるタグデータを選出するために、前記データ保管装置に保管されている索引値と比較される索引値を生成し、
前記通信部は、
前記秘匿化された検索キーワードと前記索引値とが含まれる検索要求を、前記データ保管装置に対して送信することを特徴とする請求項１又は２に記載のデータ処理装置。
【請求項４】
前記索引生成部は、
前記保管キーワードのハッシュ値を算出し、算出したハッシュ値をハフマン符号化して、前記保管対象データの暗号化データに対応付けられるタグデータが前記データ保管装置で保管される際に前記タグデータに付される索引値を生成し、
前記検索キーワードのハッシュ値を算出し、算出したハッシュ値をハフマン符号化して、前記データ保管装置において、前記秘匿化された検索キーワードとの照合の対象となるタグデータを選出するために、前記データ保管装置に保管されている索引値と比較される索引値を生成することを特徴とする請求項３に記載のデータ処理装置。
【請求項５】
データ処理装置に接続され、前記データ処理装置から送信された暗号化データを保管するデータ保管装置であって、
前記データ処理装置から、保管対象の暗号化データと、暗号化データの検索の際に照合

されるタグデータと、前記保管対象の暗号化データに指定された保管キーワードから一意に得られる乱数値をエントロピー符号化して得られた保管索引値とが含まれる保管要求を受信し、受信した前記保管要求に含まれるタグデータのＩＤ（Ｉｄｅｎｔｉｆｉｃａｔｉｏｎ）をタグＩＤとして設定する保管要求受信部と、
前記保管要求受信部により設定されたタグＩＤと前記保管要求に含まれる保管索引値とを対応付けるインデックス情報を生成するインデックス情報生成部と、
前記インデックス情報生成部により生成されたインデックス情報を記憶するインデックス記憶部と、
前記保管要求に含まれる保管対象の暗号化データと、前記保管要求に含まれるタグデータとを対応付けて保管するデータ保管部とを有することを特徴とするデータ保管装置。
【請求項６】
前記インデックス情報生成部は、
前記保管要求に含まれる保管索引値と同じ保管索引値が記述されている既存のインデックス情報が前記インデックス記憶部に記憶されている場合に、既存のインデックス情報において同じ保管検索値と対応付けられている他のタグＩＤとともに、前記保管要求に含まれるタグデータのタグＩＤを既存のインデックス情報において前記保管索引値と対応付けることを特徴とする請求項５に記載のデータ保管装置。
【請求項７】
前記データ保管装置は、
複数のデータ処理装置に接続されており、
前記データ保管装置は、更に、
前記複数のデータ処理装置のうち暗号化データの検索を要求するデータ処理装置から、秘匿化された検索キーワードと、秘匿化前の検索キーワードから一意に得られる乱数値をエントロピー符号化して得られた検索索引値とが含まれる検索要求を受信する検索要求受信部と、
前記インデックス記憶部に記憶されているインデックス情報を参照して、前記検索要求に含まれる検索索引値と一致する保管索引値と対応付けられているタグＩＤを１つ以上選出するタグＩＤ選出部と、
前記タグＩＤ選出部により選出されたタグＩＤに対応するタグデータを前記データ保管部から抽出し、抽出したタグデータごとに、前記検索要求に含まれる秘匿化された検索キーワードとの照合を行い、検索キーワードと一致する保管キーワードから生成されているタグデータを特定し、特定したタグデータと対応付けられている暗号化データを前記データ保管部から抽出するデータ抽出部とを有することを特徴とする請求項５又は６に記載のデータ保管装置。
【請求項８】
前記インデックス情報生成部は、
タグＩＤと保管索引値とを表形式で対応付けるインデックス情報を生成することを特徴とする請求項５～７のいずれかに記載のデータ保管装置。
【請求項９】
前記インデックス情報生成部は、
タグＩＤと保管索引値とを木形式で対応付けるインデックス情報を生成することを特徴とする請求項５～８のいずれかに記載のデータ保管装置。
【請求項１０】
複数の暗号化データと、各暗号化データに対応付けられている、暗号化データの検索の際に照合されるタグデータとを保管するデータ保管装置に接続されているコンピュータが行うデータ処理方法であって、
前記コンピュータが、前記データ保管装置での保管の対象となる保管対象データのキーワードを保管キーワードとして指定するキーワード指定ステップと、
前記コンピュータが、前記保管キーワードから一意に得られる乱数値をエントロピー符号化して、前記保管対象データの暗号化データに対応付けられるタグデータが前記データ

保管装置で保管される際に前記タグデータに付される索引値を生成する索引生成ステップと、
前記コンピュータが、前記保管対象データの暗号化データと前記タグデータと前記索引値とが含まれる保管要求を、前記データ保管装置に対して送信する通信ステップとを有することを特徴とするデータ処理方法。
【請求項１１】
データ処理装置に接続され、前記データ処理装置から送信された暗号化データを保管するコンピュータが行うデータ保管方法であって、
前記コンピュータが、前記データ処理装置から、保管対象の暗号化データと、暗号化データの検索の際に照合されるタグデータと、前記保管対象の暗号化データに指定された保管キーワードから一意に得られる乱数値をエントロピー符号化して得られた保管索引値とが含まれる保管要求を受信し、受信した前記保管要求に含まれるタグデータのＩＤ（Ｉｄｅｎｔｉｆｉｃａｔｉｏｎ）をタグＩＤとして設定する保管要求受信ステップと、
前記コンピュータが、前記保管要求受信ステップにより設定されたタグＩＤと前記保管要求に含まれる保管索引値とを対応付けるインデックス情報を生成するインデックス情報生成ステップと、
前記コンピュータが、前記インデックス情報生成ステップにより生成されたインデックス情報を記憶するインデックス記憶ステップと、
前記コンピュータが、前記保管要求に含まれる保管対象の暗号化データと、前記保管要求に含まれるタグデータとを対応付けて保管するデータ保管ステップとを有することを特徴とするデータ保管方法。
【請求項１２】
複数の暗号化データと、各暗号化データに対応付けられている、暗号化データの検索の際に照合されるタグデータとを保管するデータ保管装置に接続されているコンピュータに、
前記データ保管装置での保管の対象となる保管対象データのキーワードを保管キーワードとして指定するキーワード指定ステップと、
前記保管キーワードから一意に得られる乱数値をエントロピー符号化して、前記保管対象データの暗号化データに対応付けられるタグデータが前記データ保管装置で保管される際に前記タグデータに付される索引値を生成する索引生成ステップと、
前記保管対象データの暗号化データと前記タグデータと前記索引値とが含まれる保管要求を、前記データ保管装置に対して送信する通信ステップとを実行させることを特徴とするプログラム。
【請求項１３】
データ処理装置に接続され、前記データ処理装置から送信された暗号化データを保管するコンピュータに、
前記データ処理装置から、保管対象の暗号化データと、暗号化データの検索の際に照合されるタグデータと、前記保管対象の暗号化データに指定された保管キーワードから一意に得られる乱数値をエントロピー符号化して得られた保管索引値とが含まれる保管要求を受信し、受信した前記保管要求に含まれるタグデータのＩＤ（Ｉｄｅｎｔｉｆｉｃａｔｉｏｎ）をタグＩＤとして設定する保管要求受信ステップと、
前記保管要求受信ステップにより設定されたタグＩＤと前記保管要求に含まれる保管索引値とを対応付けるインデックス情報を生成するインデックス情報生成ステップと、
前記インデックス情報生成ステップにより生成されたインデックス情報を記憶するインデックス記憶ステップと、
前記保管要求に含まれる保管対象の暗号化データと、前記保管要求に含まれるタグデータとを対応付けて保管するデータ保管ステップとを実行させることを特徴とするプログラム。
【発明の詳細な説明】
【技術分野】

【０００１】
本発明は、秘匿検索技術に関する。
【背景技術】
【０００２】
秘匿検索とは、暗号化データを暗号化したまま検索する技術である。
近年は、クラウドサービスなどのインターネット上でデータを管理する際に、盗聴などの脅威から守るためのセキュリティ技術として注目されている。
【０００３】
暗号化したデータを検索する手法は、確定的暗号を利用する方法と、確率的暗号を利用する方法の２種類がある。
確定的暗号を利用する方法は、同一のキーワードを暗号化した際に、同一の暗号文が得られるため、暗号化しないデータベース検索システムと同様の高速化手法を利用することができる（例えば、同一キーワードのタグのグルーピング）。
一方、確率的暗号を利用した方式は、同一のキーワードを暗号化する場合でも、異なる暗号文となる。
したがって、単純に（バイナリデータとして）同一の暗号文をグルーピングする方法などを利用することはできない。
特に、確率的暗号を利用した方法では、安全性証明が付加される場合が多く、暗号文から一切の情報を得られないことが、数学的に証明されている。
そのため、キーワードの（部分）情報を利用する索引化は困難である。
【０００４】
次に、確率的暗号を利用したキーワード検索システムを説明する。
【０００５】
確率的暗号を利用した秘匿検索は、ユーザ（検索者、登録者）とデータセンター（データ管理者）の間で、以下のようにしてデータ検索を行うことが一般的である。
データ登録者は、文書情報と、それを検索するためのキーワードを鍵として無意味な平文を暗号化したタグデータ（以下、タグという）を登録する。
データセンターは、タグと文書情報を組にして保存する。
検索時には、検索者は検索したいキーワードで作成した検索クエリをデータセンターに送信する。
データセンターは、検索クエリを用いて、保存しているタグとの一致検査を行う。
一致すれば、組で保存されている文書情報を検索者に返却する。
なお、上記において、文書情報とは、検索システムとは別の暗号で暗号化された文書そのものや、文書名、文書を保管しているデータベースの位置情報などである。
また、上記において、検索クエリとは、検索語に対応するキーワードで暗号化されたタグを復号するための復号鍵である。また、検索クエリから検索キーワードを読み取ることはできない。
【０００６】
なお、単純な秘匿検索方式としては、例えば非特許文献１に記載されている方式がある。
これは、単一のキーワードにおける秘匿検索を実現したものである。
【０００７】
秘匿検索において、データセンターはタグ、検索クエリ、一致検査のいずれにおいても一切の情報も得ることができないため、高いセキュリティが提供される。
しかし、データセンターで行われる一致検査時に、保存されたすべてのタグを検査しなければないため、検索速度が低速であるという課題がある。
この解決法として、特許文献１では、検索履歴を利用して高速化する手法を提案している。また、特許文献２では、ブルームフィルタによって索引木を構成する方法を提案している。
なお、ブルームフィルタとは、フィルタに登録された集合に対して、ある要素が集合に

含まれているかを判定するためのフィルタである。
【先行技術文献】
【特許文献】
【０００８】
【特許文献１】特開２００５－１３４９９０号公報
【特許文献２】特開２００７－５２６９８号公報
【非特許文献】
【０００９】
【非特許文献１】Ｄ Ｂｏｎｅｈ、 Ｇ．Ｄ．Ｃｒｅｓｃｅｎｚｏ、 Ｒ．Ｏｓｔｒｏｖｓｋｙ、Ｐｅｒｓｉａｎｏ Ｇ、 “Ｐｕｂｌｉｃ Ｋｅｙ Ｅｎｃｒｙｐｔｉｏｎ ｗｉｔｈ Ｋｅｙｗｏｒｄ Ｓｅａｒｃｈ”、Ｐｒｏｃｅｅｄｉｎｇｓ ｏｆ ＥＵＲＯＣＲＹＰＴ ’０４、 ｖｏｌ．３０２７ ＬＮＣＳ、 ｐｐ．５０６－５２２ （２００４）
【発明の概要】
【発明が解決しようとする課題】
【００１０】
秘匿検索の最も基本的な方式は非特許文献１である。
この方式はＩＤベース暗号（以下、ＩＢＥ：Ｉｄｅｎｔｉｔｙ－Ｂａｓｅｄ Ｅｎｃｒｙｐｔｉｏｎ）を用いた秘匿検索方式である。
ＩＢＥは、ＩＤを鍵としてデータの暗号化を行う。
ＩＤはＥメールアドレスや住所、名前など、ユーザを一意に識別できるものある。
ＩＤを鍵とするため、誰でもデータの暗号化を行うことができる。
暗号化データは、ＩＤに対応する復号鍵を持つ者（一般的には暗号化に使用したＩＤの所有者）だけが復号できる。
データ登録者はまず、文書情報に設定するキーワードを、ＩＢＥの暗号化機能を利用して、キーワードを鍵（ＩＤ）として無意味な平文を暗号化し、文書情報のタグとして文書情報と共にデータセンターに保存する。
タグは暗号化されているため、キーワードに関する情報は漏洩しない。
次に、検索者は検索したいキーワードと、自身の秘密鍵（ユーザ鍵）から検索クエリを生成し、データセンターに送信する。
第三者は検索クエリからも、キーワード情報を知ることはできない。
検索クエリを受信したデータセンターは、保存されているタグに対し、検索クエリを用いてＩＢＥの復号処理を実行する。
無意味な平文が正しく復号できた場合、暗号化に用いたキーワード（文書情報に設定したキーワード）と検索クエリを生成する際に用いたＩＤ（検索キーワード）が対応していることを意味するため、「一致」という検索結果のみをデータセンターは知ることができる。
【００１１】
特許文献１は、ある検索キーワードで検索した結果をインデックスに保存し、次回以降に同一のキーワードが検索された場合に、高速に検索結果を応答する仕組みである。
特許文献１の方法は、一度検索したキーワードに対しては高速に検索結果を検索者に返却することが可能であるが、一度も検索したことのないキーワードに対しては、データセンターに保存された全てのタグを検索する必要がある。
そのため、膨大なタグを保存するデータベースで、初めて検索されるキーワードに対する検索結果を得るのに、多大な時間を要するという課題がある。
【００１２】
特許文献２の方法は、確定的暗号で暗号化したキーワードからブルームフィルタを作成し、索引木を構築する方法である。
この方法は、データセンターに保存されたブルームフィルタ同士のハミング距離を計算する必要があるため、保存されたブルームフィルタ数が増加すると、索引木生成のための

計算量が急激に増加するという課題がある。
　また、ブルームフィルタの大きさにより、ブルームフィルタに含まれるキーワードの識別には限度がある。
　特に偽陽性の問題から、ブルームフィルタに登録される可能性のあるキーワード種類数を考慮して、十分に大きなフィルタを用意する必要がある。
　また、ブルームフィルタ１つに対して１つのキーワードが対応している場合（ブルームフィルタに単一のキーワードしか入力しなかった場合）、ブルームフィルタに対応するデータ数を分析することにより、データセンターにキーワード分布に関する情報が漏洩するという課題がある。
【００１３】
　また、他の類似する索引方式でも、単一のキーワードに対して対応するデータ群を管理するものがほとんどである。
　そのため、データセンターはキーワード分布を知ることができ、経験上知られている分布と照合することで、キーワードを推測する「頻度解析」が実施できるという課題がある。
【００１４】
　本発明では、上記のような課題を解決することを主な目的としており、確率的暗号を利用した秘匿検索を安全に高速化することを目的とする。
【課題を解決するための手段】
【００１５】
　本発明に係るデータ処理装置は、
　複数の暗号化データと、各暗号化データに対応付けられている、暗号化データの検索の際に照合されるタグデータとを保管するデータ保管装置に接続され、
　前記データ保管装置での保管の対象となる保管対象データのキーワードを保管キーワードとして指定するキーワード指定部と、
　前記保管キーワードから一意に得られる乱数値をエントロピー符号化して、前記保管対象データの暗号化データに対応付けられるタグデータが前記データ保管装置で保管される際に前記タグデータに付される索引値を生成する索引生成部と、
　前記保管対象データの暗号化データと前記タグデータと前記索引値とが含まれる保管要求を、前記データ保管装置に対して送信する通信部とを有することを特徴とする。
【発明の効果】
【００１６】
　本発明によれば、データ処理装置において、保管キーワードから一意に得られる乱数値をエントロピー符号化して、保管対象データの暗号化データに対応付けられるタグデータがデータ保管装置で保管される際にタグデータに付与される索引値を生成するため、データ保管装置において、索引値を用いてタグデータを分類、索引化することができ、確率的暗号を利用した秘匿検索を安全に高速化することができる。
【図面の簡単な説明】
【００１７】
【図１】実施の形態１に係る秘匿検索システムの構成例を示す図。
【図２】実施の形態１に係る利用者端末装置の構成例を示す図。
【図３】実施の形態１に係るデータセンター装置の構成例を示す図。
【図４】実施の形態１に係る検索要求のデータ構成例を示す図。
【図５】実施の形態１に係る登録データのデータ構成例を示す図。
【図６】実施の形態１に係るインデックス情報のデータ構成例を示す図。
【図７】実施の形態１に係るデータ保管部のデータ保管の例を示す図。
【図８】実施の形態１に係る利用者端末装置における索引値生成処理を示すフローチャート図。
【図９】実施の形態１に係るデータセンター装置におけるデータ登録処理を示すフローチャート図。

【図１０】実施の形態１に係るデータセンター装置における検索処理を示すフローチャート図。
【図１１】実施の形態１に係るハッシュ計算の例とハフマン符号化の例を示す図。
【図１２】実施の形態１に係る利用者端末装置及びデータセンター装置のハードウェア構成例を示す図。
【発明を実施するための形態】
【００１８】
実施の形態１．
本実施の形態では、エントロピー符号を用いてキーワードタグを分類、索引化することで、確率的暗号を利用した秘匿検索を安全に高速化する構成を説明する。
なお、本実施の形態では、エントロピー符号の例として、ハフマン符号を用いて説明を行う。
また、本実施の形態では、ユーザ（検索者、登録者）とデータセンター（データ管理者）の間で、キーワードの鍵付ハッシュ値をハフマン符号化した符号値を索引値とする例を説明する。
【００１９】
本実施の形態における大まかな流れは、次の通りである。
まず、データ登録者は、保存したい文書に関する「文書情報」と文書に関連するキーワードから生成した「タグデータ（以下、タグと表記する）」およびキーワードの鍵付ハッシュ値をハフマン符号で符号化した値（索引値）をデータセンターに保存する。
ここで、文書情報とは、検索システムとは別の暗号で暗号化された文書そのものや、文書名、文書を保管しているデータベースの位置情報などである。
文書情報から文書そのものを閲覧することはできない。
タグは文書情報を検索する際に用いる、キーワードを暗号鍵にして無意味な平文（乱数）を暗号化した値である。
データセンターは、タグと文書情報を組にして保存し、索引値を利用してインデックスを構築する。
検索時には、検索者は検索したいキーワードで作成した検索クエリと、登録時と同様の手順で検索キーワードから生成した索引値をデータセンターに送信する。
データセンターは、受信した索引値から検索対象のタグを限定し、検索クエリを用いて、それらのタグと一致検査を行う。
一致すれば、組で保存されている文書情報を検索者に返却する。
なお、以下では、文書情報をデータセンターに保管する際に、保管対象の文書情報に対して指定するキーワードを保管キーワードとも表記し、検索時に指定するキーワードを検索キーワードとも表記する。
また、文書情報及びタグとともにデータセンターで保管される索引値を保管索引値とも表記し、検索時に検索クエリとともにデータセンターに送信する索引値を検索索引値とも表記する。
【００２０】
次に、本実施の形態に係る秘匿検索システムの構成を説明する。
図１は、本実施の形態に係る秘匿検索システムの構成例を示す図である。
【００２１】
図１において、秘匿検索システム１００は、利用者端末装置２０１、データセンター装置３０１を備える。
利用者端末装置２０１は社内ＬＡＮ（Ｌｏｃａｌ　Ａｒｅａ　Ｎｅｔｗｏｒｋ）１０２に接続されている。
社内ＬＡＮ１０２はネットワーク１０１を介してデータセンター装置３０１と接続されている。
【００２２】
利用者端末装置２０１は、企業のユーザが利用するＰＣ（Ｐｅｒｓｏｎａｌ　Ｃｏｍｐ

ｕｔｅｒ）である。
利用者端末装置２０１は、文書情報とそれを検索するためのタグをデータセンター装置３０１に保管するとともに、データセンター装置３０１に蓄積したタグを検索し、データセンター装置３０１から文書情報を取り出す。
なお、利用者端末装置２０１は、データ処理装置の例である。
【００２３】
データセンター装置３０１は、企業内で作成された文書情報およびタグを保管する大容量の記憶装置を持つサーバ装置である。
また、利用者端末装置２０１から送信されるキーワードの暗号文（またはハッシュ値）をエントロピー符号化した値を利用した索引値を持ち、保存されたタグを効率的に検索する機能を備える。
タグは暗号化された状態で保存されるため、データセンター装置３０１はタグからキーワードを知ることはできない。
なお、データセンター装置３０１は、データ保管装置の例である。
【００２４】
ネットワーク１０１は、社内ＬＡＮ１０２とデータセンター装置３０１を接続する通信路である。
代表的なネットワーク１０１の例はインターネットがある。
【００２５】
社内ＬＡＮ１０２は、企業内に施設された通信路であり、企業内で利用される様々なサーバ装置やＰＣが接続される。
なお、通信路は専用線や無線、ルータなどで構成される複雑な通信路となる。
【００２６】
図２は、利用者端末装置２０１の構成例を示すブロック図である。
利用者端末装置２０１は、文書・鍵格納領域２０２、文書情報管理部２０３、利用者Ｉ／Ｆ（Ｉｎｔｅｒｆａｃｅ）部２０４、検索クエリ生成部２０５、タグ生成部２０６、通信部２０７、索引生成部２０８を備える。
【００２７】
文書・鍵格納領域２０２は、データセンター装置３０１に保存する文書情報を生成するためのオリジナルの文書と、検索クエリの生成のためのユーザ秘密鍵、およびタグ生成のための公開鍵を保存する。
文書情報として、文書を暗号化したデータを利用する場合、文書情報を暗号化するための鍵も保存する。
また、索引生成部２０８において、利用者Ｉ／Ｆ２０４から得られたキーワードを暗号化（またはハッシュ化）するための鍵を保存する。
【００２８】
文書情報管理部２０３は、文書・鍵格納領域２０２に保存されている文書をもとに、文書情報を生成する。
文書情報からは、文書を閲覧することはできない。
文書情報の例としては、検索システムとは別の暗号で暗号化された文書そのものや、文書名、文書を保管しているデータベースの位置情報などがある。
つまり、文書情報管理部２０３は、データセンター装置３０１での保管の対象となる文書（保管対象データ）の暗号化を行って文書情報（保管対象データの暗号化データ）を生成する。
【００２９】
利用者Ｉ／Ｆ部２０４は、秘匿検索システム１００を操作するためのインタフェースである。
検索クエリ生成部２０５や、タグ生成部２０６のためのキーワードをユーザから入力する機能を備える。
つまり、利用者Ｉ／Ｆ部２０４は、ユーザからの指示に従って、保管キーワードや検索

キーワードを指定する。
利用者Ｉ／Ｆ部２０４は、キーワード指定部の例である。
【００３０】
検索クエリ生成部２０５は、ユーザが利用者Ｉ／Ｆ部２０４を介して入力した検索したいキーワードと、文書・鍵格納領域２０２に保存されたユーザ鍵から検索クエリを生成する。
つまり、検索クエリ生成部２０５は、検索クエリとして検索キーワードに対応するタグを復号するための復号鍵を生成する。
検索クエリ生成部２０５は、検索キーワード秘匿化部の例である。
【００３１】
タグ生成部２０６は、ユーザが利用者Ｉ／Ｆ部２０４を介して入力した文書に設定したいキーワードと、文書・鍵格納領域２０２に保存された公開鍵からタグデータを生成する。
【００３２】
通信部２０７は、データセンター装置３０１に対して検索を実行するための検索要求（検索クエリと検索索引値）や、データセンター装置３０１に対して新規に文書情報とタグを保存するための登録データ（文書情報、タグ、保管索引値）を、データセンター装置３０１に送信するために、社内ＬＡＮ１０２に接続するものである。
なお、登録データは、保管要求ともいう。
【００３３】
索引生成部２０８は、利用者Ｉ／Ｆ部２０４から受信したキーワードを、文書・鍵格納領域２０２に保存してある共通鍵（データ登録者と検索者との間の共通鍵）を用いてハッシュ値を計算する（以下、鍵付ハッシュ値という）。
さらに鍵付ハッシュ値に対して、ハフマン符号化を実施して得られた値を索引値として出力する。
つまり、索引生成部２０８は、保管キーワードから一意に得られる乱数値をハフマン符号化して、タグがデータセンター装置３０１で保管される際にタグに付される索引値（保管索引値）を生成する。
また、索引生成部２０８は、検索キーワードから一意に得られる乱数値をハフマン符号化して、データセンター装置３０１において、検索クエリ（秘匿化された検索キーワード）との照合の対象となるタグを選出するために、データセンター装置３０１に保管されている保管索引値と比較される索引値（検索索引値）を生成する。
【００３４】
図３は、データセンター装置３０１の構成例を示すブロック図である。
データセンター装置３０１は、検索要求受信部３０２、検索処理部３０３、インデックス記憶部３０４、データ保管部３０５、索引分類部３０６、検索要求回答部３０７、登録データ受信部３０８を備える。
【００３５】
登録データ受信部３０８は、利用者端末装置２０１から登録データ（保管要求）を受信すると、受信した登録データに含まれるタグにタグＩＤ（Ｉｄｅｎｔｉｆｉｃａｔｉｏｎ）を付与し、索引分類部３０６に保管索引値とタグＩＤを、データ保管部３０５にタグと文書情報をそれぞれ転送する。
タグＩＤは、タグを一意に特定することができる識別子である。
なお、登録データ受信部３０８は、保管要求受信部の例である。
【００３６】
索引分類部３０６は、登録データ受信部３０８から登録データとして保管索引値とタグＩＤを受信し、インデックス記憶部３０４にインデックス情報を追加する機能を持つ。
つまり、索引分類部３０６は、タグＩＤと保管索引値とを対応付けるインデックス情報を生成する。
索引分類部３０６は、インデックス情報生成部の例である。

また、索引分類部３０６は、後述する検索処理部３０３から受信した検索索引値から、対応するタグＩＤ群をインデックス記憶部３０４から取り出す機能を持つ。

【００３７】

インデックス記憶部３０４は、保管索引値とタグＩＤとの対応付けが示されるインデックス情報を保存する機能を持つ。

なお、インデックス記憶部３０４に保存するインデックス情報では、一つの保管索引値に１つ又は複数のタグＩＤが対応する。

【００３８】

データ保管部３０５は、登録データ受信部３０８から受信した文書情報とタグとを対応付けて保存する。

【００３９】

検索要求受信部３０２は、利用者端末装置２０１から送信された検索要求を受信し、検索処理部３０３へ転送する。

【００４０】

検索処理部３０３は、検索要求受信部３０２から検索要求を受信し、検索要求に含まれる検索索引値を索引分類部３０６に送信し、検索索引値と一致する保管索引値がインデックス記憶部３０４に存在する場合は、保管索引値から得られた検索対象タグをデータ保管部３０５から取り出し、検索クエリとの一致検査を行う。

そして、検索処理部３０３は、一致検査の結果より得られた文書情報を検索要求回答部３０７を介して利用者端末装置２０１に送信する。

つまり、検索処理部３０３は、インデックス記憶部３０４に記憶されているインデックス情報を参照して、検索要求に含まれる検索索引値と一致する保管索引値と対応付けられているタグＩＤを１つ以上選出する。

更に、検索処理部３０３は、選出したタグＩＤに対応するタグをデータ保管部３０５から抽出し、抽出したタグごとに、検索要求に含まれる検索クエリ（秘匿化された検索キーワード）との照合を行い、検索キーワードと一致する保管キーワードから生成されているタグを特定し、特定したタグと対応付けられている文書情報をデータ保管部３０５から抽出する。

検索処理部３０３は、タグＩＤ選出部及びデータ抽出部の例に相当する。

【００４１】

検索要求回答部３０７は、検索処理部３０３によって検索された文書情報を検索要求元の利用者端末装置２０１に送信する。

【００４２】

図４は、本実施の形態に係る検索要求２００１のデータ構造の一例を示す。

【００４３】

ここでは、この構成例を検索要求Ａとする。

検索要求Ａは検索クエリ２００２と索引値２００４から構成される。

検索クエリ２００２は、文書・鍵格納領域２０２に保存されたユーザ秘密鍵と検索キーワード２００３を用いて生成する。

生成された検索クエリ２００２からは、検索キーワード２００３を知ることはできない。

データセンター装置３０１は、検索クエリ２００２を用いてタグとの一致検査を行い、検索キーワード２００３と同じ保管キーワードから生成されたタグを抽出する。

索引値２００４は、検索キーワード２００３から図８の手順で得られるエントロピー符号値である。

【００４４】

図５は、本実施の形態に係る登録データ２３０１のデータ構造の一例を示す。

【００４５】

ここでは、この構成例を登録データＡとする。

登録データＡは文書情報２３０２、タグ２３０３、索引値２３０５から構成される。

文書情報２３０２は、例えば、検索システムとは別の暗号で暗号化された文書そのものや、文書名、文書を保管しているデータベースの位置情報などであり、文書情報から文書そのものを閲覧することはできない。
タグ２３０３は、ユーザによって付与された、文書情報２３０２に対する保管キーワード２３０４と文書・鍵格納領域２０２に保存された公開鍵を鍵として無意味な平文を暗号化したデータである。
索引値２３０５は、保管キーワード２３０４から図８の手順で得られるエントロピー符号値である。

【００４６】

本実施の形態における索引値は２進数のビット列である。
インデックス情報の構造は索引値から検索対象のタグＩＤが得られればよい。
本実施の形態では、インデックス情報は、最も単純な転置インデックスの形式を例とするが、２分木やＢ木を用いた索引木を構成してもよい。

【００４７】

図６は、インデックス記憶部３０４に記憶されるインデックス情報３００１のデータ構造の一例を示す。
インデックス情報３００１は、索引値３００２とタグＩＤ３００３から構成される。
索引値３００２は、文書情報に設定されたキーワードの鍵付ハッシュ値をエントロピー符号化した値（保管索引値）である。
タグＩＤ３００３は、索引値３００２を生成するために用いた保管キーワードのタグに対応するＩＤである。
検索の際には、検索者が指定した検索索引値と同じ保管索引値と対応付けられているタグＩＤ群から得られるタグを一致検査対象とする。

【００４８】

図７のようにデータ保管部３０５は、タグＩＤ３０５１、タグ３０５２、文書情報３０５３を列に持つような表の形式でデータを保存する。
タグＩＤ３０５１は、登録データ受信部３０８によって付与され、インデックス記憶部３０４に保存されるタグＩＤと一致する。

【００４９】

図８は、本実施の形態における索引値の計算方法を説明するためのフローチャートである。

【００５０】

まず、ステップＳ３０６１において、索引生成部２０８は、利用者Ｉ／Ｆ部２０４から入力されたキーワードを文書・鍵格納領域２０２に保存してある鍵を利用してハッシュ値を計算する（本実施の形態ではハッシュ値だが、共通鍵暗号化した値としてもよい）。
このとき、キーワードのハッシュ値を計算する際に用いるハッシュ関数は、同一のキーワードからは同じ値が計算される必要がある。
また、データセンターがキーワードのハッシュ値を計算できないように、データ登録者と、検索者のみ暗号化の鍵を保持することが望ましい。
例えば、暗号化データベースとは無関係の確定的暗号や鍵付ハッシュ関数を利用する。
本実施の形態では、用語の混乱を避け、説明を簡単にするため、キーワードの鍵付ハッシュ値を利用するものとする。

【００５１】

次に、ステップＳ３０６２において、索引生成部２０８は、ステップＳ３０６１で生成した値に対し、エントロピー符号化を実施する。
本実施の形態では、エントロピー符号としてハフマン符号を用いることとする。

【００５２】

次に、ステップＳ３０６３において、索引生成部２０８は、ステップＳ３０６２で出力された符号値を、索引値として通信部２０７に送信する。
エントロピー符号として、ハフマン符号を利用した例を次に示す。

【００５３】
キーワードとして、「ｏｒｉｇｉｎａｌ」「ｃｏｎｆｉｒｍ」「ｓｈａｒｅ」を持つ文書情報をそれぞれＤ（“ｏｒｉｇｉｎａｌ”）、Ｄ（“ｃｏｎｆｉｒｍ”）、Ｄ（“ｓｈａｒｅ”）とする。
また、それぞれの文書情報に付加されるタグをｔａｇ（“ｏｒｉｇｉｎａｌ”）、ｔａｇ（“ｃｏｎｆｉｒｍ”）、ｔａｇ（“ｓｈａｒｅ”）とし、そのタグに付与されるタグＩＤをそれぞれ１、２、３とする。
（タグ、タグＩＤ）として表すと、（ｔａｇ（“ｏｒｉｇｉｎａｌ”）、１）、（ｔａｇ（“ｃｏｎｆｉｒｍ”）、２）、（ｔａｇ（“ｓｈａｒｅ”）、３）となる。
【００５４】
次に、キーワード「ｏｒｉｇｉｎａｌ」「ｃｏｎｆｉｒｍ」「ｓｈａｒｅ」のハッシュ値を計算する。
例えば、ハッシュアルゴリズムであるＳＨＡ２５６でそれぞれのハッシュ値を取ると、図１１（ａ）に示すような１６進数（０～９、ａ～ｆ）６４桁の数値が得られる。
ＳＨＡ２５６の性質から、このハッシュ値からキーワードを求めることはできない。
【００５５】
次に、得られたハッシュ値をハフマン符号化する。
ハフマン符号化を実施すると、図１１（ｂ）に示すよう２進数の数値を得ることができる。
このキーワードのハッシュ値をハフマン符号化した２進数を索引値とする。
【００５６】
上記ハフマン符号値は、符号化時の変換表がなければハッシュ値を復元することはできない。
本実施の形態では、ハフマン符号の変換表を破棄するため、符号値からハッシュ値を復元することはできない。
また、「ｏｒｉｇｉｎａｌ」と「ｃｏｎｆｉｒｍ」の索引値が同じであるため、ｔａｇ（“ｏｒｉｇｉｎａｌ”）とｔａｇ（“ｃｏｎｆｉｒｍ”）はデータセンター装置３０１のインデックス情報において同一のエントリに保存されることになる。
そのため、単純な索引技術のように、索引値からタグの分布を読み取ることができなくなるため、キーワードの頻度解析を防止することが可能であり、より安全である。
【００５７】
次に、秘匿検索システム１００の動作について説明する。
図９は、本実施の形態のデータ登録処理の例を説明するフローチャートである。
この処理はデータセンター装置３０１で実施される処理である。
【００５８】
まず、ステップＳ４０１において、データセンター装置３０１は、利用者端末装置２０１からネットワーク１０１を経由して送信される登録データＡを、登録データ受信部３０８で受信する。
登録データＡは、図５の登録データ２３０１のようになっている。
登録データ受信部３０８は、受信したタグ２３０３にタグＩＤを付与する。
そして、タグＩＤと文書情報２３０２とタグ２３０３をデータ保管部３０５に送信し、タグＩＤと索引値２３０５を索引分類部３０６に送信する。
【００５９】
次に、ステップＳ４０２において、索引分類部３０６は、登録データ受信部３０８から受信した索引値が、既にインデックス情報に存在するかを検査するためにインデックス記憶部３０４を参照する。
【００６０】
次に、ステップＳ４０３において、索引分類部３０６は、登録データ受信部３０８から受信した索引値がインデックス記憶部３０４に保存されたインデックス情報３００１のエントリに含まれるかどうかを検査する。

既に該当するエントリがある場合（該当あり）、ステップＳ４０４に進む。
該当するエントリがない場合（該当なし）は、ステップＳ４０６に進む。
【００６１】
次に、ステップＳ４０４において、索引分類部３０６は、該当するインデックス情報３００１のエントリのタグＩＤ３００３に、登録データ受信部３０８から受信したタグＩＤを追加する。
【００６２】
次に、ステップＳ４０５おいて、データ保管部３０５が、登録データ受信部３０８から受信したタグＩＤと文書情報２３０２とタグ２３０３を保存する。
【００６３】
次に、ステップＳ４０３で、インデックス記憶部３０４に保存されたインデックス情報３００１に登録データ受信部３０８から受信した索引値が存在しない場合について説明する。
【００６４】
ステップＳ４０６において、索引分類部３０６は、登録データ受信部３０８から受信した索引値と対応するタグＩＤから新たにエントリを作成し、インデックス情報３００１に新しいエントリを追加する。
その後ステップＳ４０５を実行する。
【００６５】
以上がデータ登録処理の動作の説明である。
次にデータ検索処理の説明をする。
図１０は、本実施の形態のデータ検索処理の例を説明するフローチャートである。
【００６６】
まず、ステップＳ５０１において、データセンター装置３０１は、利用者端末装置２０１からネットワーク１０１を経由して送信される検索要求Ａを、検索要求受信部３０２で受信する。
【００６７】
次に、ステップＳ５０２において、検索要求受信部３０２は、検索要求Ａを検索処理部３０３に転送する。
検索処理部３０３はインデックス情報を参照するために、索引分類部３０６に検索要求Ａに含まれる索引値２００４を転送する。
【００６８】
次に、ステップＳ５０３において、索引分類部３０６は、インデックス記憶部３０４に保存されているインデックス情報３００１に、検索処理部３０３から受信した索引値２００４が含まれているかどうか検査する。
【００６９】
次に、Ｓ５０３の検査で検索要求Ａの索引値２００４がインデックス記憶部３０４に存在しなかった場合（該当なし）は、ステップＳ５０４において、索引分類部３０６は検索処理部３０３に対して検索要求Ａに含まれる検索キーワード２００３がデータ保管部３０５に存在しないと通知する。
検索処理部３０３は、データ保管部３０５を検索せずに、検索要求回答部３０７を介して、検索要求のあった利用者端末装置２０１へ該当データが存在しない旨を回答する。
【００７０】
次に、Ｓ５０３の検査の結果、受信した検索要求Ａの索引値２００４がインデックス記憶部３０４に存在した場合は（該当あり）、ステップＳ５０５において、索引分類部３０６によってインデックス情報３００１から索引値２００４に対応するタグＩＤ群が返却される。
検索処理部３０３は、索引値２００４に対応するタグＩＤ群を用いて、該当するタグをデータ保管部３０５から参照し、検索要求Ａに含まれる検索クエリ２００２を用いてキーワード一致検査を行う。

【００７１】
次に、Ｓ５０５の検索の結果、検索要求Ａに一致するタグがある場合に、検索処理部３０３は、ステップＳ５０４において、該当するタグに対応する文書情報をデータ保管部３０５から読み出し、検索要求回答部３０７を介して、検索要求のあった利用者端末装置２０１へ回答する。
データ保管部３０５に、検索要求に該当するデータが存在しない場合は、検索処理部３０３は、検索要求回答部３０７を介して、検索要求のあった利用者端末装置２０１へ該当データが存在しない旨を回答する。
【００７２】
以上の手順により、暗号化データベースにおいて、暗号化したままデータを検索する方法を安全かつ高速に実施することができる。
【００７３】
以上の実施の形態によれば、検索キーワードのハッシュ値のエントロピー符号を利用することで、検索キーワードに関する情報を漏らすことなく、索引を構成することが可能となり、高速な検索を実施できるという効果がある。
【００７４】
また、本実施の形態で構成する索引は１つ以上のキーワードに対応する検索タグをグルーピングすることが可能であり、索引からキーワードに対応するデータの分布を秘匿するという効果がある。
【００７５】
また、保存されるキーワードの種類の増加に対して、索引を再構成する必要がなく、索引の維持のための計算量を低減するという効果がある。
【００７６】
本実施の形態ではキーワードの鍵付ハッシュ値をハフマン符号化した値を索引値として利用する例を開示したが、キーワードから一意に得られる値であればよく、ハッシュ関数や共通鍵暗号（確定的暗号）を利用してもよい。
【００７７】
また、本実施の形態では、エントロピー符号としてハフマン符号を利用したが、エントロピー符号であればよく、例えば算術符号であっても良い。
【００７８】
また、本実施の形態では、最も単純な転置インデックスの形式を例としたが、索引値から検索対象のタグＩＤが得られればよく、２分木やＢ木を用いた索引木を構成しても良い。
【００７９】
以上、秘匿検索におけるエントロピー符号を利用した索引手法を開示したが、これは、文書情報のキーワード検索に限らず、任意のデータに対応した検索のためのキーデータ（検索キー）であれば、応用可能なことは明らかである。
つまり、本実施の形態で説明したキーワードとは、単語やセンテンスに限らず、あらゆる形式のキーデータを意味する。
このように、本実施の形態によれば、データを暗号化したままで複数の検索キーを用いた検索を高速に実施することができる。
したがって、画像検索、動画検索、音声検索などへの応用が可能である。
【００８０】
以上、本実施の形態では、
検索キーワードおよび、文書情報に設定するキーワードから索引値を生成する索引生成部と、
新規登録対象のキーワードタグと文書情報を受信する登録データ受信部と、
前記登録データ受信部から受信した全てのキーワードタグと文書情報を保存するデータ保管部と、
キーワードから一意に得られる乱数値（暗号文やハッシュ値）をエントロピー符号化し

た値（索引値）と、その索引値に対応するキーワードタグを示すタグＩＤとからなるエントリを複数保持するインデックスを記憶するインデックス記憶部と、
検索要求に含まれる索引値が前記インデックス記憶部に存在する場合に、検索対象となるタグＩＤから該当するタグをデータ保管部から取り出し、キーワードが一致するタグかどうかを検査する検索処理部と、
データ検索時には、索引処理部から受信した索引値から、対応するタグＩＤをインデックス記憶部から取り出し、データ登録時には登録データ受信部から受信した索引値とタグを用いてインデックス記憶部に索引情報を追加する索引分類部とを備える秘匿検索システムを説明した。

【００８１】
また、本実施の形態では、
前記索引生成部は、データ登録時の文書情報に設定するキーワードやデータ検索時の検索キーワードからエントロピー符号値を計算して索引値として出力し、
前記索引分類部は、前記索引生成部によって出力された索引値と、索引値に対応するタグを示すタグＩＤを保存および参照することを説明した。

【００８２】
また、本実施の形態では、
前記索引分類部は、前記索引生成部によって出力された索引値と、索引値に対応するタグを示すタグＩＤを表形式でインデックス記憶部に保存および参照することを説明した。

【００８３】
また、本実施の形態では、
前記索引分類部は、前記索引生成部によって出力された索引値と、索引値に対応するタグを示すタグＩＤを木形式（２分木、Ｂ木など）でインデックス記憶部に保存および参照することを説明した。

【００８４】
最後に、本実施の形態に示した利用者端末装置２０１及びデータセンター装置３０１のハードウェア構成例について説明する。
図１２は、本実施の形態に示す利用者端末装置２０１及びデータセンター装置３０１のハードウェア資源の一例を示す図である。
なお、図１２の構成は、あくまでも利用者端末装置２０１及びデータセンター装置３０１のハードウェア構成の一例を示すものであり、利用者端末装置２０１及びデータセンター装置３０１のハードウェア構成は図１２に記載の構成に限らず、他の構成であってもよい。

【００８５】
図１２において、利用者端末装置２０１及びデータセンター装置３０１は、プログラムを実行するＣＰＵ９１１（Ｃｅｎｔｒａｌ　Ｐｒｏｃｅｓｓｉｎｇ　Ｕｎｉｔ、中央処理装置、処理装置、演算装置、マイクロプロセッサ、マイクロコンピュータ、プロセッサともいう）を備えている。
ＣＰＵ９１１は、バス９１２を介して、例えば、ＲＯＭ（Ｒｅａｄ　Ｏｎｌｙ　Ｍｅｍｏｒｙ）９１３、ＲＡＭ（Ｒａｎｄｏｍ　Ａｃｃｅｓｓ　Ｍｅｍｏｒｙ）９１４、通信ボード９１５、表示装置９０１、キーボード９０２、マウス９０３、磁気ディスク装置９２０と接続され、これらのハードウェアデバイスを制御する。
更に、ＣＰＵ９１１は、ＦＤＤ９０４（Ｆｌｅｘｉｂｌｅ　Ｄｉｓｋ　Ｄｒｉｖｅ）、コンパクトディスク装置９０５（ＣＤＤ）と接続していてもよい。また、磁気ディスク装置９２０の代わりに、ＳＳＤ（Ｓｏｌｉｄ　Ｓｔａｔｅ　Ｄｒｉｖｅ）、光ディスク装置、メモリカード（登録商標）読み書き装置などの記憶装置でもよい。
ＲＡＭ９１４は、揮発性メモリの一例である。ＲＯＭ９１３、ＦＤＤ９０４、ＣＤＤ９０５、磁気ディスク装置９２０の記憶媒体は、不揮発性メモリの一例である。これらは、記憶装置の一例である。
本実施の形態で説明した文書・鍵格納領域２０２、インデックス記憶部３０４及びデー

タ保管部３０５は、ＲＡＭ９１４、磁気ディスク装置９２０等により実現される。
通信ボード９１５、キーボード９０２、マウス９０３、ＦＤＤ９０４などは、入力装置の一例である。
また、通信ボード９１５、表示装置９０１などは、出力装置の一例である。
【００８６】
通信ボード９１５は、図１に示すように、ネットワークに接続されている。
例えば、通信ボード９１５は、ＬＡＮ、インターネットの他、ＷＡＮ（ワイドエリアネットワーク）、ＳＡＮ（ストレージエリアネットワーク）などに接続されていても構わない。
【００８７】
磁気ディスク装置９２０には、オペレーティングシステム９２１（ＯＳ）、ウィンドウシステム９２２、プログラム群９２３、ファイル群９２４が記憶されている。
プログラム群９２３のプログラムは、ＣＰＵ９１１がオペレーティングシステム９２１、ウィンドウシステム９２２を利用しながら実行する。
【００８８】
また、ＲＡＭ９１４には、ＣＰＵ９１１に実行させるオペレーティングシステム９２１のプログラムやアプリケーションプログラムの少なくとも一部が一時的に格納される。
また、ＲＡＭ９１４には、ＣＰＵ９１１による処理に必要な各種データが格納される。
【００８９】
また、ＲＯＭ９１３には、ＢＩＯＳ（Ｂａｓｉｃ　Ｉｎｐｕｔ　Ｏｕｔｐｕｔ　Ｓｙｓｔｅｍ）プログラムが格納され、磁気ディスク装置９２０にはブートプログラムが格納されている。
利用者端末装置２０１及びデータセンター装置３０１の起動時には、ＲＯＭ９１３のＢＩＯＳプログラム及び磁気ディスク装置９２０のブートプログラムが実行され、ＢＩＯＳプログラム及びブートプログラムによりオペレーティングシステム９２１が起動される。
【００９０】
上記プログラム群９２３には、本実施の形態の説明において「～部」（「インデックス記憶部３０４及びデータ保管部３０５」以外、以下同様）として説明している機能を実行するプログラムが記憶されている。プログラムは、ＣＰＵ９１１により読み出され実行される。
【００９１】
ファイル群９２４には、本実施の形態の説明において、「～の判断」、「～の計算」、「～の暗号化」、「～の符号化」、「～の比較」、「～の照合」、「～の参照」、「～の検索」、「～の抽出」、「～の検査」、「～の生成」、「～の設定」、「～の登録」、「～の選択」、「～の入力」、「～の受信」等として説明している処理の結果を示す情報やデータや信号値や変数値やパラメータが、「～ファイル」や「～データベース」の各項目として記憶されている。
「～ファイル」や「～データベース」は、ディスクやメモリなどの記録媒体に記憶される。
ディスクやメモリなどの記憶媒体に記憶された情報やデータや信号値や変数値やパラメータは、読み書き回路を介してＣＰＵ９１１によりメインメモリやキャッシュメモリに読み出される。
そして、読み出された情報やデータや信号値や変数値やパラメータは、抽出・検索・参照・比較・演算・計算・処理・編集・出力・印刷・表示などのＣＰＵの動作に用いられる。
抽出・検索・参照・比較・演算・計算・処理・編集・出力・印刷・表示のＣＰＵの動作の間、情報やデータや信号値や変数値やパラメータは、メインメモリ、レジスタ、キャッシュメモリ、バッファメモリ等に一時的に記憶される。
また、本実施の形態で説明しているフローチャートの矢印の部分は主としてデータや信号の入出力を示す。

データや信号値は、ＲＡＭ９１４のメモリ、ＦＤＤ９０４のフレキシブルディスク、ＣＤＤ９０５のコンパクトディスク、磁気ディスク装置９２０の磁気ディスク、その他光ディスク、ミニディスク、ＤＶＤ等の記録媒体に記録される。

また、データや信号は、バス９１２や信号線やケーブルその他の伝送媒体によりオンライン伝送される。

【００９２】

また、本実施の形態の説明において「～部」として説明しているものは、「～回路」、「～装置」、「～機器」であってもよく、また、「～ステップ」、「～手順」、「～処理」であってもよい。

すなわち、本実施の形態で説明したフローチャートに示すステップ、手順、処理により、本発明に係るデータ処理方法及びデータ保管方法を実現することができる。

また、「～部」として説明しているものは、ＲＯＭ９１３に記憶されたファームウェアで実現されていても構わない。

或いは、ソフトウェアのみ、或いは、素子・デバイス・基板・配線などのハードウェアのみ、或いは、ソフトウェアとハードウェアとの組み合わせ、さらには、ファームウェアとの組み合わせで実施されても構わない。

ファームウェアとソフトウェアは、プログラムとして、磁気ディスク、フレキシブルディスク、光ディスク、コンパクトディスク、ミニディスク、ＤＶＤ等の記録媒体に記憶される。

プログラムはＣＰＵ９１１により読み出され、ＣＰＵ９１１により実行される。

すなわち、プログラムは、本実施の形態の「～部」としてコンピュータを機能させるものである。あるいは、本実施の形態の「～部」の手順や方法をコンピュータに実行させるものである。

【００９３】

このように、本実施の形態に示す利用者端末装置２０１及びデータセンター装置３０１は、処理装置たるＣＰＵ、記憶装置たるメモリ、磁気ディスク等、入力装置たるキーボード、マウス、通信ボード等、出力装置たる表示装置、通信ボード等を備えるコンピュータである。

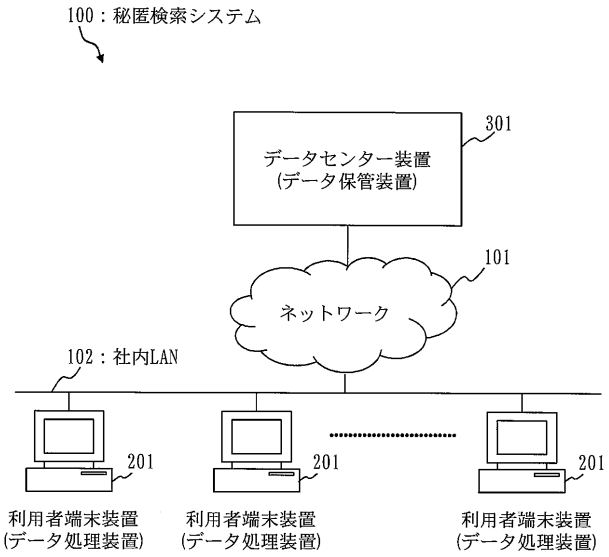
そして、上記したように「～部」として示された機能をこれら処理装置、記憶装置、入力装置、出力装置を用いて実現するものである。

【符号の説明】

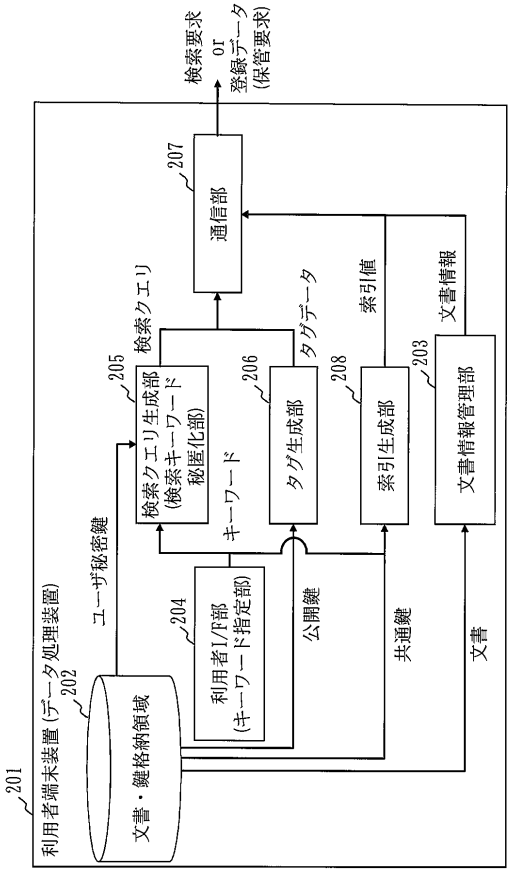
【００９４】

１００　秘匿検索システム、１０１　ネットワーク、１０２　社内ＬＡＮ、２０１　利用者端末装置、２０２　文書・鍵格納領域、２０３　文書情報管理部、２０４　利用者Ｉ／Ｆ部、２０５　検索クエリ生成部、２０６　タグ生成部、２０７　通信部、２０８　索引生成部、３０１　データセンター装置、３０２　検索要求受信部、３０３　検索処理部、３０４　インデックス記憶部、３０５　データ保管部、３０６　索引分類部、３０７　検索要求回答部、３０８　登録データ受信部。

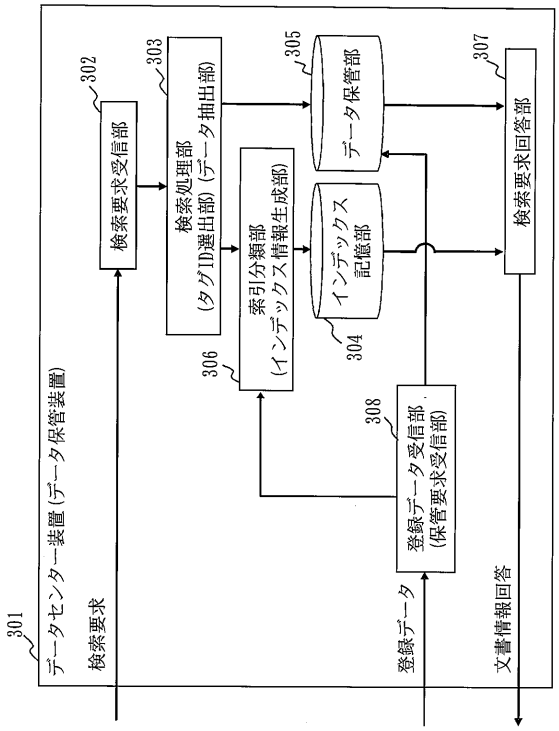
【図１】


100：秘匿検索システム
データセンター装置
(データ保管装置)
301
ネットワーク
101
102：社内LAN
201
利用者端末装置
(データ処理装置)

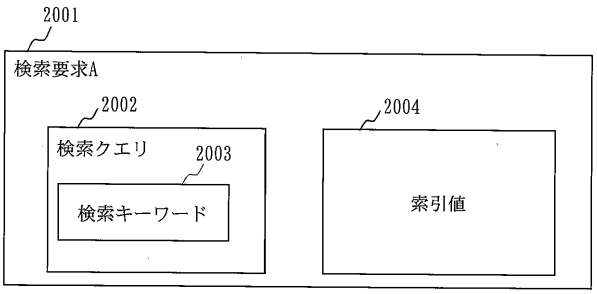
【図２】


検索要求
or
登録データ
(保管要求)
通信部
207
検索クエリ
タグデータ
索引値
文書情報
205
206
208
203
検索クエリ生成部
(検索キーワード
秘匿化部)
タグ生成部
索引生成部
文書情報管理部
キーワード
ユーザ秘密鍵
利用者I/F部
(キーワード指定部)
204
公開鍵
共通鍵
文書
利用者端末装置(データ処理装置)
201
文書・鍵格納領域
202

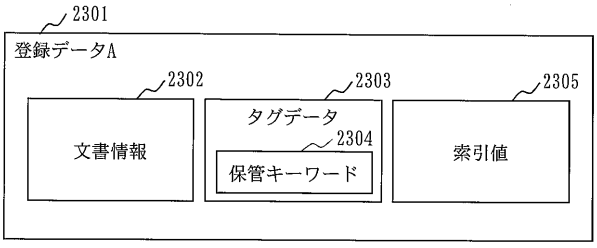
【図３】


データセンター装置(データ保管装置)
301
検索要求
登録データ
文書情報回答
検索要求受信部
302
検索処理部
(タグID選出部)(データ抽出部)
303
索引分類部
(インデックス情報生成部)
306
データ保管部
305
インデックス
記憶部
304
検索要求回答部
307
登録データ受信部
(保管要求受信部)
308

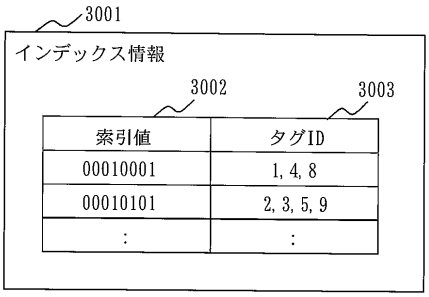
【図４】


2001
検索要求A
2002
検索クエリ
2003
検索キーワード
2004
索引値

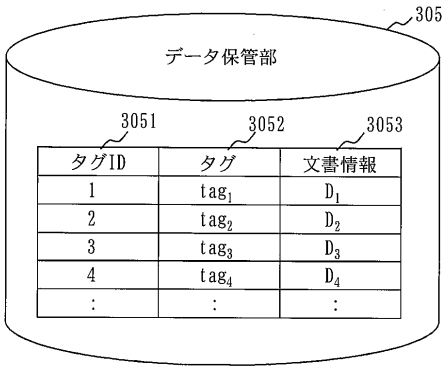
【図５】


2301
登録データA
2302
文書情報
2303
タグデータ
2304
保管キーワード
2305
索引値

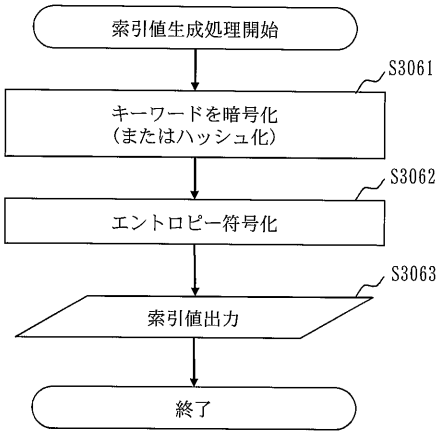
【図６】



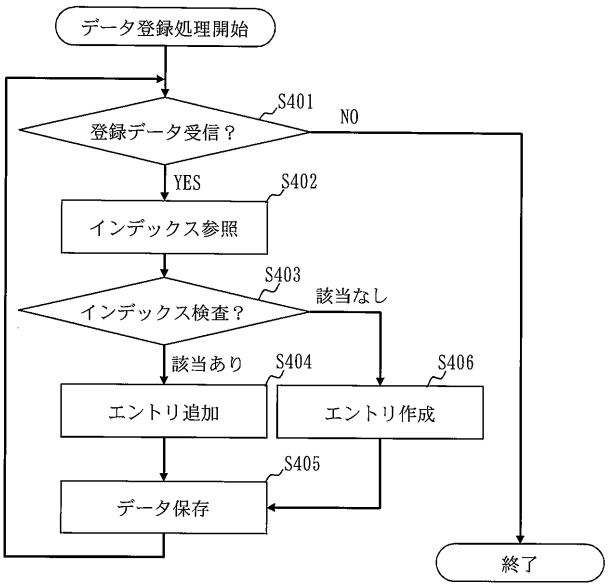
【図７】



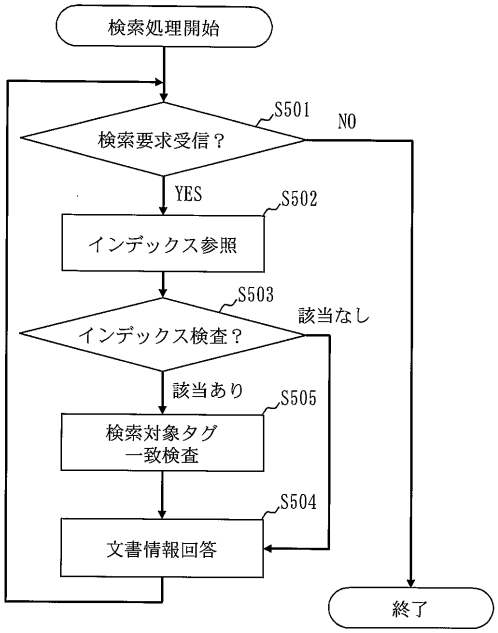
【図８】



【図９】



【図１０】



【図１１】

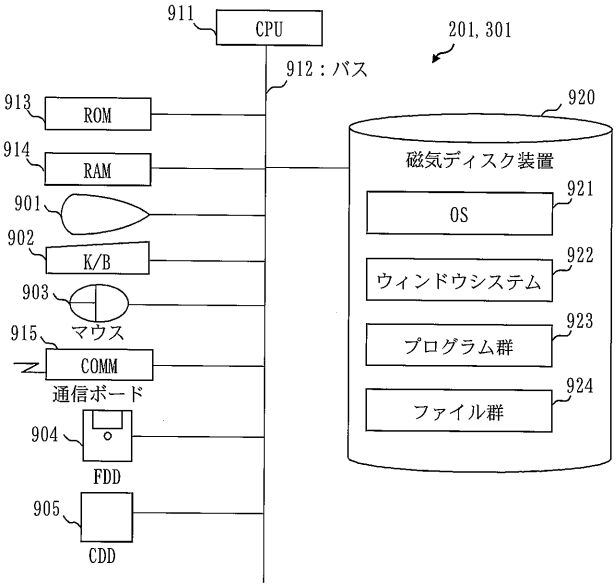
(a)

「original」 →76f0404d9cb6dc90824f4c51edc98ebf7b1b0f8977700f4c19eb33032087 0d6
「confirm」 →a322ab4557014b320a766ab6477a78b7a6eb17ef0d0addf56ad654f1c3081ade
「share」 →c09ac5b4c87530fcfb070d216ceb7253dff4019218d32bf08684bed5103593e9

(b)

「original」 →0000000000100010001101000101011100010011011101111011011
「confirm」 →0000000001000100011010001010111000100110111011110111101
「share」 →0000000001000100011010011100011100100110111010111101111

【図１２】



フロントページの続き

(72)発明者　伊藤　隆
東京都千代田区丸の内二丁目７番３号　三菱電機株式会社内
(72)発明者　服部　充洋
東京都千代田区丸の内二丁目７番３号　三菱電機株式会社内
(72)発明者　平野　貴人
東京都千代田区丸の内二丁目７番３号　三菱電機株式会社内

Ｆターム(参考) 5B017 AA08 BA05 BA07 CA16
5B075 ND20 NK10 NK49 PP30
5J104 AA12 AA16 EA26 JA03 JA21 NA02 NA12 NA27 NA37 PA14