

(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
2010年4月15日(15.04.2010)

PCT

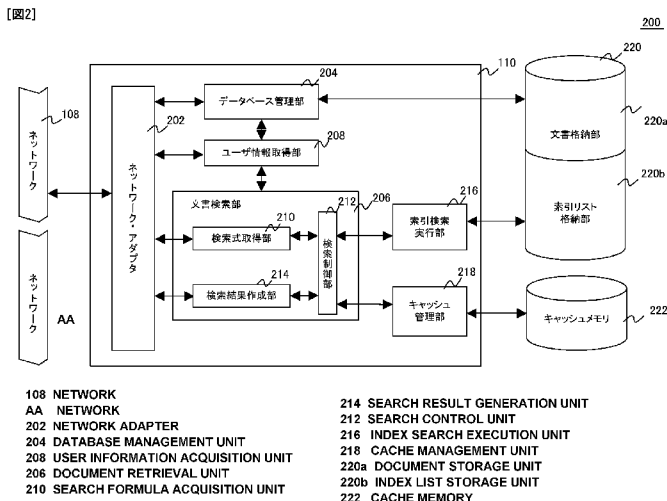
(10) 国際公開番号
WO 2010/041516 A1

- (51) 国際特許分類:
G06F 17/30 (2006.01)
- (21) 国際出願番号: PCT/JP2009/064360
- (22) 国際出願日: 2009年8月14日(14.08.2009)
- (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 2008-261481 2008年10月8日(08.10.2008) JP
- (71) 出願人 (米国を除く全ての指定国について): インターナショナル・ビジネス・マシーンズ・コーポレーション (INTERNATIONAL BUSINESS MACHINES CORPORATION) [US/US]; 10504 ニューヨーク州アーモンク ニュー オーチャード ロード New York (US).
- (72) 発明者; および
- (75) 発明者/出願人 (米国についてのみ): 西川 紘史 (NISHIKAWA hirofumi) [JP/JP]; 〒2428502 神奈川県大和市下鶴間 1 6 2 3 番地 1 4 日本アイ・ビー・エム株式会社大和事業所内 Kanagawa
- (74) 代理人: 上野 剛史, 外 (UENO Takeshi et al.); 〒2428502 神奈川県大和市下鶴間 1 6 2 3 番地 1 4 日本アイ・ビー・エム株式会社大和事業所内 Kanagawa (JP).
- (81) 指定国 (表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, 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,

[続葉有]

(54) Title: INFORMATION PROCESSING APPARATUS, DOCUMENT RETRIEVAL SYSTEM, DOCUMENT RETRIEVAL METHOD, AND PROGRAM

(54) 発明の名称: 情報処理装置、文書検索システム、文書検索方法およびプログラム



(57) Abstract: Provided are an information processing apparatus, a document retrieval system, a document retrieval method, and a program, each for performing document retrieval while ensuring security. An information processing apparatus (110) includes a database management unit (204) for generating an index including an access right to access a document and managing the index together with the document, a user information acquisition unit (208) for acquiring the access right as user information with respect to a user identification value, a document retrieval unit (206) for acquiring the user information and a search request including a search formula, generating a combination search formula by combining an inclusion relational expression including an operator specifying the inclusion relation on the access right of the issuer of the search request with the search formula, and acquiring a shared-level search result which can be shared under a plurality of user identification values in a domain including the user identification values from the combination search formula, and a cache management unit (218) for entering the shared-level search result as a cache item discriminated to be a shared-level one.

(57) 要約:

[続葉有]



WO 2010/041516 A1



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, MK, MT, NL,
NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ,

CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN,
TD, TG).

添付公開書類:

— 国際調査報告 (条約第 21 条(3))

【課題】セキュリティを確保しつつ文書検索を実行する情報処理装置、文書検索システム、文書検索方法およびプログラムを提供する。 【解決手段】情報処理装置 110 は、文書に対するアクセス権限を含ませた索引を生成して、前記文書とともに管理するデータベース管理部 204 と、ユーザ識別値についてアクセス権限をユーザ情報として取得するユーザ情報取得部 208 と、ユーザ情報および検索式を含む検索要求を取得し、検索要求の発行元のアクセス権限についての包含関係を指定する演算子を含む包含関係式を検索式に結合して結合検索式を生成し、結合検索式から、複数のユーザ識別値を含むドメイン内で複数のユーザ識別値の下に共有できる共用レベルの検索結果を取得する文書検索部 206 と共用レベルであることを識別させたキャッシュ項目として前記共用レベルの検索結果を登録するキャッシュ管理部 218 とを含んでいる。

明 細 書

発明の名称：

情報処理装置、文書検索システム、文書検索方法およびプログラム

技術分野

[0001] 本発明は、文書検索に関し、より詳細には、全文索引を使用する場合のセキュリティを確保しつつ、検索効率を向上させる、情報処理装置、文書検索システム、文書検索方法およびプログラムに関する。

背景技術

[0002] ネットワーク技術および情報処理装置の性能向上に伴い、文書の全体の語句を索引として検索する、いわゆる全文索引検索が可能とされている。また、文書検索については、オープン環境で全ユーザに対してすべての文書を検索させるシステムもあるし、また、文書検索システムにアクセスしてきたユーザのアクセス権限をユーザ情報として使用し、アクセス権限に関連してアクセスできる文書を制限することにより、データベースなどに蓄積された文書のセキュリティを確保する、いわゆるセキュリティ検索システムも存在する。

[0003] 文書セキュリティを継承する機能(以下、セキュア検索として参照する。)を有する文書検索システムは、多くの場合、検索対象文書の全文索引内に文書に対するアクセス権情報を保持させ、アクセス索引として文書検索時にアクセス権限の判断を実行する。従来の検索システムでは、多くの場合、ユーザの検索時に取得されるユーザ名やグループ名、またはそれらの階層包含関係(以下、これらを総じてユーザ情報として参照する。)が、閲覧可能なアクセス権にマッチする文書のみを検索結果として抽出することで、セキュア検索を実現する。

[0004] このような文書検索システムは、検索実行時にユーザ情報を含む検索式を作成し、全文検索の基盤を大きく変更することなくセキュア検索を実現可能とする。しかしながら、ユーザ情報を含む検索式は、アクセスしてきたユー

ザのアクセス権限の他、ユーザが属するすべてのグループ名を羅列したOR検索式となる。この結果、従来のセキュア検索は、文書数およびグループ階層が増加するにつれ、グループが含む文書数の増加およびグループ数の増加により、検索対象が非線型的に増加し、検索性能に大きな影響を及ぼすという問題がある。

[0005] 上述した従来技術の問題点を改善するため、検索結果をキャッシュしておき、2回目以降の検索を、高速化する既存技術も知られている。検索結果をキャッシングする処理を採用することで、同一の検索式での検索処理を高速化することができる。しかしながら、セキュア検索についてみれば、ユーザ相互、グループ間相互のアクセス権限を保存しなければならないという特有の問題が生じる。すなわち、キャッシュ結果は、アクセス権限に基づいて登録されなければならないので、ユーザを識別するためにユーザ固有のアクセス権限を含むキャッシュ索引が作成されることになる。この結果、キャッシュ項目は、「同一のユーザ」が「同一の検索式」を使用して実行する場合にのみ、キャッシュヒットを生成させ、それ以外の場合には、キャッシュミスヒットを生じさせてしまい、キャッシュ載りよう効率が高められないという問題があった。

[0006] 上述した問題は、キャッシュヒットの発生が、従来の文書検索システムでは同一ユーザが同一検索式を複数回入力する場合にのみ発生することに起因する。このような検索は、ユーザまたは文書検索システム内で何らかの不都合があった場合を除き多くは生じず、非セキュア検索の場合と比較してキャッシュのヒット率が著しく低下することによる。

[0007] これまで、セキュア検索を可能とする文書検索システムは多数知られている。例えば、特開2005-284608号公報（特許文献1）では、セキュア検索を高速に実現する処理を開示する。特許文献1では、事前にアクセス権限の属性値をデータベースに設定することによりセキュア検索を可能とする。

[0008] 特許文献1では、検索処理以前に、事前に索引に対してアクセス権情報の

組合わせにラベルを付する処理が必要となり、また検索は、ユーザのアクセス権を網羅的にOR検索し、検索結果の和集合を生成する処理が必要とされ、グループの増加およびグループ単位で蓄積される文書数の増加に対して非線型的に検索処理のオーバーヘッドが増加するので、検索効率は、非線型的に非効率化する。この結果、セキュア検索において全文索引検索を対象とするシステムにそのまま適用するには、検索効率の点で現実性が低いものである。

- [0009] また、特開2004-164555号公報（特許文献2）には、検索装置および方法ならびにそのインデックス構築装置が開示されている。特許文献2は、セキュア検索を実現する手法に関する発明であり、特許文献2では、セキュリティ・ドメインを規定し、当該セキュリティ・ドメインに対してアクセス権を設定するための複数のインデックスを登録させることにより、ドメインごとのセキュリティ性を確保する。この際、セキュリティ・ドメインごとのインデックスの生成は、ドメインごとに設定された管理者が、セキュリティ・ドメイン内の文書を収集する収集プログラムを実行させ、セキュリティ・ドメインでアクセス可能文書を収集し、インデックスの作成および編集などにより行われている。
- [0010] 特許文献2の処理によっても、セキュア検索は可能となるが、管理者によるインデックス管理などの負担が増加する。また、収集プログラムによる文書の収集は、結局のところ、文書を検索し抽出するという文書検索処理の実行を伴い、実質的に文書検索プログラムを、管理者用の収集プログラムと、一般ユーザ用の文書検索プログラムというコンテキストで実装しなければならないという、ソフトウェアモジュールの無駄を発生させてしまう。
- [0011] さらに、特許文献2の処理は、文書に対して階層構造を割当て、アクセス権のチェックをインデックス単位で枝狩りすることにより検索対象となる文書数を削減することで、セキュア検索における検索効率またはレスポンス性を向上させるものである。文書に対して階層構造を割当て、枝狩りする処理を使用することによっても、検索対象空間を削減することは可能となる。し

かしながら、文書に階層構造を割当てても、枝の最下層にある文書にまで到達しなければならない場合、結局の所ユーザのアクセス権の網羅的に含むOR検索を実行しなければならない。この結果、検索式によって検索速度やレスポンスは大きく異なることになり、検索処理のスケーラビリティが低下する。

- [0012] また、文書は、そのセキュリティ・レベルに対応して一義的な階層構造を割当てることができない場合もあり、また、セキュリティ・レベルの変更を伴う場合、ツリー構造を再構築しなければならないという不都合がある。さらに、文書に対する階層構造割当てに依存して、本来検索可能な文書であっても、検索対象から排除されてしまう文書が発生することも想定でき、必ずしも汎用的な文書検索に適用できるわけではない。加えて、例えば、ユーザがプライベートなアクセス権限でアクセスする場合には、階層構造に依存せずにアクセスさせなければならないので、複数の索引構造を用意しなければならず、全文検索を前提としたセキュア検索においては、特許文献1と同様に実現性が低いものということができる。

先行技術文献

特許文献

- [0013] 特許文献1：特開2005-284608号公報
特許文献2：特開2004-164555号公報

発明の概要

発明が解決しようとする課題

- [0014] 以上のように、従来のセキュア検索処理は、セキュア検索を実現するために管理者によるセキュリティ管理の負担が発生すること、セキュア検索および文書収集のための類似するソフトウェアの並列実装が必要となる。また、セキュリティ・レベルの変更に対応して索引構造を大幅に変更する必要が生じ、また異なる索引構造を並列的に実装することが必要となるなど、全文索引検索のためのセキュア検索にそのまま適用することが困難であった。加え

て、ユーザが帰属されるグループの階層構造が増加するにつれて、非線型的に検索処理のオーバーヘッドが増大し、検索効率が低下するものであり、文書数の増加およびグループの階層構造の増加に柔軟に対応しながら検索効率を改善するものではない。

- [0015] すなわち、本発明は、全文索引を使用するセキュア検索処理の検索効率を向上することが可能な、情報処理装置、文書検索システム、文書検索方法およびプログラムを提供することを目的とする。

課題を解決するための手段

- [0016] 本発明は、上記従来技術の問題点を、セキュア検索のための全文索引構造を修正することなく、キャッシュヒット確率を向上させ、さらに、文書検索式の集合演算回数を削減させることによりセキュア検索における検索効率を改善する。
- [0017] ユーザ単位でのセキュア性の確保は、文書に対してではなく、ユーザに対して階層構造を導入することにより行われ、ユーザの階層構造の識別処理は、ユーザログイン時の情報を利用して検索実行時に検索サーバがユーザ情報として取得する。検索サーバは、ユーザ入力された検索式と、ユーザのアクセス権限から生成される包含関係式とを結合させて結合検索式を生成する。文書検索システムは、結合検索式からユーザのプライベート・レベルの検索を実行する検索式および共用レベルの検索を実行する検索式を生成し、独立して検索を実行させる。
- [0018] 各検索結果は、プライベート・レベルの検索結果および共用レベルの検索結果を識別することが可能なキャッシュ索引にそれぞれ対応付けられて、キャッシュメモリに格納される。
- [0019] 特定の実施形態では、プライベート・レベルの検索結果を索引付けするキャッシュ索引は、プライベート・レベルの検索結果を与える検索式を使用したhash計算により生成され、また共用レベルの検索結果を索引付けするキャッシュ索引は、共用レベルの検索結果を与える検索式を使用したhash計算により生成される。この結果、共用レベルの検索結果についてのキャッシュ索

引の検索は、同一の共用レベルに関連した検索については、検索するユーザが相違してもキャッシュヒットを発生させるが、プライベート・レベルを含むキャッシュ項目に関してキャッシュヒットを発生させず、共用レベルでのキャッシュヒット確率を向上させつつ、プライベート・レベルのセキュリティを保證することができる。

[0020] プライベート・レベルでの検索結果は、同様にしてキャッシュ項目として、プライベート・レベルであることを示すキャッシュ索引とともに登録され、プライベート・レベルでの検索要求の場合には、キャッシュ索引の検索が実行され、キャッシュヒットの場合には、キャッシュ内容が検索結果とされる。

[0021] 一方、キャッシュミスヒットの場合には、共用レベルでも、プライベート・レベルでも全文索引を使用した全文索引検索が実行され、新たな検索結果が生成される。生成された新たな検索結果は、新たな検索結果を生成するために使用した検索式についてのhash計算により生成されるキャッシュ索引に対応付けて登録され、以後、同一の検索式に帰属される検索要求を受領した場合、キャッシュヒットに対応して全文索引検索を実行させることなくキャッシュ内容を返すことを可能とする。

[0022] この結果、本発明では、共用レベルでは、キャッシュ内容は、ユーザが過去に検索したグループ階層単位でアクセス可能な文書を網羅したものとして登録される。この結果、共用レベルでキャッシュヒットが発生した場合、ユーザのプライベート・レベルでのキャッシュヒットまたは索引検索の検索結果と、たかだか1回のOR演算を実行するだけで検索結果集合を生成することができ、集合演算のために消費されるプロセッサ能力を著しく削減することが可能となる。

[0023] また、本発明では、グループに含まれるユーザ数が増加すれば増加するだけ、キャッシュヒット確率が向上しさらに、ユーザのグループ異動や、グループ階層が増加した場合にでも、グループ階層分の集合演算を伴う全文索引検索が実行される確率を最小限に抑制することが可能となり、セキュア検索

効率を向上することができるとともに、グループ階層の増加およびそれに伴う文書の増加に柔軟に対応できるセキュア検索が可能となる。

図面の簡単な説明

- [0024] [図1]本発明の文書検索システム100の概略図。
- [図2]本実施形態の検索サーバ110の機能ブロック図。
- [図3]本実施形態のユーザのグループ階層構造300の実施形態を示した図。
- [図4]本実施形態の検索サーバ110が管理する全文索引400の実施形態を示した図。
- [図5]検索結果をキャッシュする場合の、キャッシュ索引生成処理500を比較した図。
- [図6]本実施形態で、キャッシュメモリ222内に生成されるデータ構造600の実施形態を示した図。
- [図7]本実施形態の検索サーバ110が実行する文書検索処理のフローチャート。
- [図8]図7で説明した、プライベート・レベルでのみアクセス可能な文書に関する検索結果を含む検索結果を生成する文書検索処理のフローチャート。
- [図9]本実施形態のセキュア文書検索によるOR処理の削減およびキャッシュ項目の利用性の改善を、プライベート・レベルでの検索結果を含ませる図8の処理についての結合検索式900を使用して説明する図。
- [図10]本実施形態の文書検索システム100による検索効率改善の機能を説明した図。

発明を実施するための形態

- [0025] 以下、本発明につき、実施形態をもって説明するが、本発明は、後述する実施例に限定されるものではない。図1は、本発明の文書検索システム100の概略図である。文書検索システム100は、インターネット、LANなどのネットワーク108と、ネットワーク108に接続され、ユーザにより操作されて、ネットワーク108を介して検索サーバ110にアクセスする複数の端末装置（以下、単に端末装置102として参照する。）102、1

04、106を含んで構成されている。

[0026] 検索サーバ110は、本実施形態にしたがい、検索サーバ110が管理する文書データベース（図示せず）に格納された文書に対するセキュア検索を実行する。本実施形態では、文書は、電子化された電子データ全般を意味し、テキスト、イメージ、オーディオ、またはこれらのデータが複合的に組み合わせられたマルチメディアコンテンツを含むことができる。ネットワーク108には、検索サーバ110にアクセスするユーザの、アクセス権限を管理するユーザ管理サーバ120が接続されている。ユーザ管理サーバ120は、ユーザが検索サーバ110にログオンする場合、検索サーバ110からの問い合わせを受領し、ログオン時に取得したユーザ識別値（以下、ユーザIDとして参照する。）などを使用してユーザのアクセス権限を検索する。ユーザ管理サーバ120は、ユーザ管理データベース（図示せず）を管理しており、ユーザごとのアクセス権限をリレーショナル・データベースなどを使用して管理している。

[0027] ユーザ管理サーバ120は、ユーザのアクセス権限が見出された場合、ユーザ情報として検索結果を検索サーバ110に返し、検索サーバ110によるユーザのアクセス権限を使用した包含関係式の作成を可能としている。検索サーバ110は、文書データベース（図示せず）を管理しており、文書データベースに登録された文書の全文索引検索を実行するため、検索サーバ110が管理する文書に関して全文索引を、転置索引として生成して文書を索引付けしている。

[0028] なお、図1に示した検索サーバ110とユーザ管理サーバ120は、文書検索システム100が提供するサーバ機能を説明するために分離して示したものであり、特定の実施形態では、例えば、IBM社製のBladeCenter（BladeCenterは、インターナショナル・ビジネス・マシーンズ・コーポレーション社の登録商標である。）などのラックマウント型のサーバ・ユニットとして一体となったサーバとして構成することもできる。

- [0029] 上述した検索サーバ110、ユーザ管理サーバ120は、情報処理装置として実装され、PENTIUM（登録商標）、XEON（登録商標）、CELERON（登録商標）、ATHRON（登録商標）などのPENTIUM（登録商標）互換チップといったCISCアーキテクチャのプロセッサ、または、POWER PC（登録商標）などのRISCアーキテクチャのプロセッサを実装することができ、シングルコアでもマルチコアでもよい。また、検索サーバ110、ユーザ管理サーバ120は、WINDOWS（登録商標）200X、UNIX（登録商標）、LINUX（登録商標）などのオペレーティング・システム（以下、単にOSとして参照する。）により制御されている。
- [0030] また検索サーバ110、ユーザ管理サーバ120は、上述したOSの管理下、C、C++、JAVA（登録商標）、JAVA（登録商標）BEANS、PERL、RUBY、PYTHONなどのプログラミング言語を使用して実装される、CGI、サーバレット、APACHEなどのサーバ・プログラムを実行し、各種サービスを提供している。
- [0031] さらに、検索サーバ110とユーザ管理サーバ120との間のネットワーク108を介したトランザクションのため、これまで知られたいかなるトランザクション・プロトコルでも使用することができる。例えば、検索サーバ110が検索サーバ110にアクセスしたユーザの認証およびアクセス権限検索のため、DAP (Directory Access Protocol) やLDAP (Lightweight Directory Access Protocol) などを使用してユーザ情報に関するトランザクションを生成させることができる。また、ファイル転送などのため、HTTP、FTPなどのファイル転送プロトコルなどを使用することもできるし、RMI (Remote Method Invocation) やRPC (Remote Procedure Call) などの分散コンピューティング環境を使用することもできる。
- [0032] 一方、端末装置102は、パーソナル・コンピュータ、ワークステーションとして実装され、端末装置104は、PDA (Personal Data Assistant) として構成され、端末装置106は、携帯電話などの情報処理装置として実装

されている。また、端末装置 102 は、プロセッサ (MPU) として、これまで知られたいかなるシングルコア・プロセッサまたはデュアルコア・プロセッサを含んで実装されていてもよい。また、端末装置 102、104、106 は、WINDOWS (登録商標)、WINDOES (登録商標) CE、UNIX (登録商標)、LINUX (登録商標)、MAC OS、または携帯電話専用 OS など、これまで知られたいかなる OS により制御されてもよい。

[0033] 端末装置 102、104、106 と検索サーバ 110 との間は、例えば TCP/IP などのトランザクション・プロトコルに基づき、HTTP プロトコルなどのファイル転送プロトコルを使用するトランザクションを使用したネットワークで接続されている。端末装置 102 は、この目的のため、Internet Explorer (商標)、Mozilla、Opera、Netscape Navigator (商標) などのブラウザ・ソフトウェアを実装していて、検索サーバ 110 にアクセスし、ログオン認証の後、文書のセキュア検索を行う。

[0034] なお、検索サーバ 110 と端末装置 102、104、106 との間の特ランザクションは、有線またはワイアレス通信を使用した通信基盤の下で、HTTP プロトコルに限定されず、例えば、RMI (Remote Method Invocation)、RPC (Remote Procedure Call) などの遠隔呼出プロトコルを使用し、CORBA (Common Object Resource Broker Architecture) などの下に実装される、いわゆる分散コンピューティングシステム (DCE) として参照される端末装置-サーバシステムとして実装することもできる。以下、ユーザ管理サーバ 120 は、本実施形態の要旨ではないので、より詳細な説明は省略する。

[0035] 図 2 は、本実施形態の検索サーバ 110 の機能ブロック 200 を示した図である。図 2 に示した各機能ブロックは、プロセッサが、RAM などの実行空間にプログラムを読み込んで、プロセッサがプログラムを実行させることによって、検索サーバ 110 上の機能手段として実現されている。検索サーバ 110 は、ネットワーク・アダプタ 202 と、データベース管理部 204

と、文書検索部 206 とを含んで構成されている。

[0036] ネットワーク・アダプタ 202 は、ネットワーク 108 を介して端末装置 102 からの文書検索要求や文書登録要求を受領し、受領した各要求に対応する処理を実行する機能部に渡し、また検索結果集合から適切なフォーマットの検索結果ファイルを作成し、クライアント 102、104、106 に送付する。検索サーバ 110 は、端末装置 102 から文書登録要求を受領すると、データベース管理部 204 を呼出して、文書データベース 220 に文書および全文索引の登録を実行させる。

[0037] 本実施形態の全文索引は、特定の形式に限定するものではないが、本実施形態では、全文検索性の効率の観点から、転置索引方式の索引付けを採用する。データベース管理部 204 は、受領した文書を、文書データベース 220 の文書格納部 220 a に格納し、生成した転置索引を、文書データベース 220 の索引リスト格納部 220 b に、文書と対応付けて登録する。なお、文書には、固有の文書識別値が付されていて、転置索引の検索語が、文書識別値に対応付けされて登録される。また、データベース管理部 204 は、当該文書について設定されるアクセス権限を使用して全文索引中にアクセス権限を登録する。全文索引中のアクセス権限は、セキュア検索実行中に、ユーザのアクセス権限を判断するために参照される。

[0038] 本実施形態では、ユーザのアクセス権限は、本実施形態では、ユーザ管理サーバ 120 により一括管理されている。検索サーバ 110 は、ユーザから文書の登録要求を受領すると、ユーザ情報取得部 208 を呼出して、ユーザがログオン時に入力したユーザ ID およびパスワードなどの情報を参照してユーザ管理サーバ 120 にアクセスし、ユーザのアクセス認証を実行する。そして、アクセス認証に成功すると、ユーザ情報取得部 208 は、当該ユーザのアクセス権限をユーザ情報として取得し、データベース管理部 204 に渡すことで、全文索引中にユーザに許可されたアクセス権限に対応する索引を生成し、転置索引に登録する。

[0039] 文書検索部 206 は、本実施形態にしたがい、セキュア検索を実行し、検

索式取得部 210 と、検索制御部 212 と、検索結果作成部 214 とを含んで構成されている。検索式取得部 210 は、文書検索要求が含む検索語および検索語間を接続する論理演算子を取得する。取得した検索式は、検索制御部 212 に送付されて、結合検索式を生成するためにバッファリングされる。検索結果作成部 214 は、検索制御部 212 から検索結果を受領して、検索結果を HTML、XML などの構造化文書または端末装置・プログラムに対応するフォーマットなどの検索結果ファイルを生成し、ネットワーク・アダプタ 202 を介して検索要求の発行元に返却する。また、検索結果作成部 214 は、複数の検索結果を検索制御部 212 から受領した場合、複数の検索結果の和集合を作成し、検索結果を作成する処理を実行する。

[0040] 検索制御部 212 は、本実施形態にしたがい、セキュア検索処理のプロセスを制御する。検索制御部 212 は、ユーザ情報取得部 208 が、ユーザ・ログオン時に取得したアクセス権限から包含関係式を生成し、検索式取得部 210 から受領した検索式と合成して結合検索式を生成する。例えば、検索式取得部 210 が取得した検索式が、「ibm AND patent」であり、ユーザのアクセス権限の包含関係を示すグループ階層として $G_X \in G_Y \in U_A$ (G_Y が G_X を包含する上位グループであり、ユーザ U_A は、 G_X のメンバである。)、というユーザ情報を受領しているものとする。

[0041] 検索制御部 212 は、取得したユーザ情報から包含関係式を、IN 演算子を使用して、「acl:UA IN acl:GX IN acl:GY」として生成し、これを検索式「ibm AND patent」と合成して、結合検索式 { (ibm AND patent) AND (acl:UA IN acl:GX IN acl:GY) } として生成する。IN 演算子は、ユーザのアクセス権限の包含関係を規定する検索処理を指令する演算子、であり、より詳細に後述する。本実施形態の上述したコンテキストにおいて、検索制御部 212 は、セキュア検索を実行する場合に、索引検索実行部 216 およびキャッシュ管理部 218 の検索を制御する。検索制御部 212 は、結合検索式を生成し、結合検索式をキャッシュ管理部 218 および索引検索実行部 216 に渡す。

- [0042] キャッシュ管理部 218 は、L2 キャッシュまたは L3 キャッシュなどとして構成される、キャッシュメモリ 222 を管理している。キャッシュメモリのキャッシュ項目は、検索語およびユーザのアクセス権限を使用して生成した hash 値をキャッシュ索引とし、過去に検索された検索結果をキャッシュ内容とするレコードとしてキャッシュメモリ 222 の容量が許容するまで登録されている。
- [0043] また、キャッシュ管理部 218 は、検索制御部 212 から結合検索式を受領すると、包含関係式が含むプライベート・レベルのアクセス権限に関連する情報を除き、ドメイン・レベルおよびオープンドメイン・レベルの文書を検索するため、受領した検索式の検索語、演算子、およびアクセス権限に関連する情報から、キャッシュ検索を実行するための hash 値を計算する。キャッシュ管理部 218 は、生成したキャッシュ索引を使用してキャッシュメモリ 222 の検索を実行する。
- [0044] 本実施形態において、プライベート・レベルでのアクセス権限とは、ユーザの帰属する階層グループとは関連せず、ユーザのみがアクセスすることを許可するアクセス権限または特権レベルを意味する。このようなプライベート・レベルは、例えば作成中またはレビュー中であって、ネットワーク上に公開する段階にない文書や、高度に閲覧制限を行う必要がある文書、または検索サーバ 110 が特定の商用検索サービスを行っている場合、ゲストまたはサブスクライバなどに応じて設定されるセキュリティ・レベルである。
- [0045] また、ドメイン・レベルとは、複数のユーザを含むユーザ集団を意味し、特定の集団内の部門、部署などに対応する。他の実施形態では、ドメイン・レベルは、同一のアクセス権限を有するユーザ集団とすることができる。さらに、オープンドメイン・レベルは、複数のドメイン・レベルを含むドメイン集団に対応し、特定の集団全体とすることもできるし、また、商用検索システムなどではゲスト・レベルのアクセス権限を有するものとしてすることができる。なお、ドメイン階層は、特定の実施形態において、適宜設定され、プライベート、ドメイン、オープンドメインの 3 階層に限定されるものではない。

い。

- [0046] キャッシュヒットが発生すると、キャッシュ管理部 218 は、キャッシュ内容を検索制御部 212 に返し、検索結果の作成を依頼する。また、キャッシュミスヒットが発生した場合、検索制御部 212 にキャッシュミスヒットを通知する。キャッシュミスヒットが通知されると、検索制御部 212 は、索引検索実行部 216 に対して結合検索式を使用した検索実行を指令する。
- [0047] なお、当該処理では、上述した結合検索式 { (ibm AND Patent) AND (acl:UA IN acl:GX IN acl:GY) } を、プライベート・レベルの検索式 A = (ibm AND Patent) AND (acl:UA) および共用レベルの検索式 B = (ibm AND Patent) AND (acl:GX IN acl:GY) に分離して独立した検索を実行させる。検索式 A は、ユーザのプライベート・レベルに関連する検索結果を与え、検索式 B は、ユーザが帰属されるグループ階層についての検索語に関連する検索結果を与える。検索式 B で与えられる検索結果を生成するのは、キャッシュヒットの際にプライベート・レベルに関連する検索結果がグループ内の全ユーザに共有されてしまうことを防止するためである。
- [0048] 一方、検索式 A は、ユーザのプライベート・レベルでのアクセス権限を有する文書を検索することを可能とするので、ユーザがプライベート・レベルまで含めた検索を要求する場合には、検索結果作成部 214 は、検索式 A および検索式 B の各検索結果を受領し、検索結果として、(検索式 A の検索結果) \cup (検索式 B の検索結果) を生成することにより、ユーザに返す検索結果集合を作成する。一方、ユーザがプライベート・レベルまで含まない検索を要求する場合やユーザが特権レベルを有していない場合、検索結果 B のみの検索が実行され、検索結果作成部 214 は、(検索式 B の検索結果) のみを受領してユーザに返す検索結果を作成する。
- [0049] なお、ユーザに対してどのレベルの検索結果を返すかについては、特定の集団内でのクロースドの文書検索システム 100 では、ユーザが検索要求を送付する際に、ラジオボタンやチェックボックスなどの設定でユーザオンデマンドに指定することもできる。また、文書検索システム 100 が、商用検

素などを行うオープンなシステムである場合には、ログオン時のユーザ認証に基づき、ユーザが特権ユーザであるか否かの判断を使用して、プライベート・レベルの検索結果を生成するか否かを指定し、これをユーザ情報として検索サーバ110が取得することで、検索結果生成処理を変更することができる。

[0050] 検索制御部212は、キャッシュミスヒットと判断した場合、結合検索式を索引検索実行部216に渡す。この際、本実施形態では、結合検索式から、検索式Aおよび検索式Bを作成した後、検索式が含むIN演算子を、例えばSQL (Structured Query Language) などが解釈可能なOR演算子に変更し、索引検索実行部216による全文索引検索実行を可能とする形式に変換する。IN演算子からOR演算子への変換処理は、検索制御部212が実行することもできるし、索引検索実行部216がIN演算子を含む結合検索式を受領して変換処理を実行することもでき、特に変換処理を実行する処理部に限定はない。

[0051] 索引検索実行部216は、変換処理後の検索式Aおよび検索式Bを、使用して独立して全文索引検索を実行させる。全文検索の実行は、例えばSQLパーザなどにより解析し、検索処理を実行し、AND演算子、OR演算子で結合される検索語およびアクセス権限ごとに索引リスト格納部220bを走査して、転置索引により参照される文書識別値に対応する文書のタイトル、格納場所などの情報を抽出する。抽出したデータは、検索制御部212に送付され、検索制御部212からキャッシュ制御部218および検索結果作成部214に送付され、キャッシュメモリ222への登録処理および検索結果が作成される。

[0052] 検索結果作成部214は、キャッシュ内容または索引検索の結果を、複数の検索結果がある場合、検索要求に対応して和集合を計算し、検索結果集合を作成する。その後、文書検索部206は、作成された検索結果を、ネットワーク・アダプタ202からネットワーク108を介して検索要求の発行元に返却し、本実施形態のセキュア検索の一連のトランザクションが終了する

- 。
- [0053] 以上の処理において、検索サーバ110は、キャッシュ管理部218を介して上述した検索式A、検索式B、または検索式Aおよび検索式Bの過去検索結果を、検索式Aおよび検索式Bのhash値をそれぞれキャッシュ索引として登録することが可能となる。このため、例えば、キャッシュ管理部218は、過去に検索を実行したユーザが帰属される同一グループ内で、同一の検索式を与える検索要求を受領した場合、同一グループからの検索要求に対してキャッシュヒットを生成させることが可能となる。この場合、ユーザがプライベート・レベルでの検索を要しない場合、まったく集合演算を実行することなくセキュア検索が可能となる。また、プライベート・レベルまで含めたセキュア検索を実行する場合でも、集合演算は、たかだか1回まで削減でき、この集合演算は、グループ階層の増加に関わらず、キャッシュヒット確率およびプライベート・レベルの検索結果の要否にのみに依存する。
- [0054] 本実施形態では、キャッシュヒットの確率は、特定のグループに帰属されるユーザが増加すればするほど高まり、またグループ階層の増加に関わらずキャッシュ内容として、包含関係のみに関連した検索結果が登録されている。このため、本実施形態では、ユーザ数の増加やグループ階層の増加に依存することなく、全文検索を使用する文書検索システム100の検索効率を著しく改善でき、かつスケーラビリティを保證することが可能となる。
- [0055] 図3は、本実施形態のユーザのグループ階層構造300の実施形態を示す。図3に示すように本実施形態のグループ階層構造300では、特定のユーザ、例えばUA322は、グループG(GB)320に帰属され、さらにGB320が、グループA(GA)310に帰属されている。一方、ユーザD(UD)332、UE334は、グループC(GC)330に帰属され、GC330がGA310に帰属されている。
- [0056] GB320およびGC330とは、階層構造的には関連性を有しないが、GB320およびGC330は、さらに上位階層のグループであるGA310のメンバを構成する。図3に示したグループ階層構造300の階層構造は

、U A 3 2 2 ~ U C 3 2 6、U D 3 3 2、U E 3 3 4の階層がアクセス権限の最も狭いプライベート・レベルの階層であり、G B 3 2 0、G C 3 3 0がご同一のグループに帰属される複数のユーザにアクセス権限を認めるドメイン・レベルの階層である。

[0057] さらに、G A 3 1 0は、最もアクセス権限を広く認めるオープンドメイン・レベルの階層である。オープンドメイン・レベルであるG A 3 1 0に割当てられる文書は、G B 3 2 0、G C 3 3 0のいずれのドメインからでも、そのアクセス権限に基づいてアクセスすることが可能とされている。一方、ドメイン・レベルの文書は、各ドメインに帰属されるユーザしか、アクセスできないアクセス権限として割当てられ、プライベート・レベルの文書は、原則的には、当該ユーザのみがアクセスすることができるものとしてセキュア検索が可能とされている。

[0058] なお、図3に示したグループ階層構造300は、本実施形態のコンテキストでは、検索サーバ110ではなく、ユーザ管理サーバ120が一括して管理し、検索サーバ110が、検索要求を受領してオンザフライにアクセス権限をユーザ情報としてユーザ管理サーバ120から受領する。このため、セキュリティ階層設定と、検索処理とは完全に分離することができ、文書検索システム100の設置語のユーザのアクセス権限変更や階層構造変更は、完全に検索サーバとは分離して設定可能となり、検索サーバ110の索引付処理や検索処理には影響を与えることはない。

[0059] 図4は、本実施形態の検索サーバ110が管理する全文索引400の実施形態を示す。全文索引400は、いわゆる転置索引として構成されており、検索サーバ110が文書から抽出した単語と、当該単語を含む文書の文書識別値とが、特定の単語ごとにレコードとして構成されている。その他、図示した全文索引400では、単語と同レベルでアクセス権限を処理するため、アクセス権限索引も、単語を登録するフィールド410に登録されている。例えば、単語「ibm」は、フィールド420に示されるように、文書1、文書2、文書4、…に出現し、単語「patent」は、文書2、文書3、文書4、…

に出現しているのが示されている。

- [0060] また、フィールド410には、文書に対して設定されたアクセス権限が、転置索引として追加されており、例えば、acl:UAとしてユーザUAについてプライベート・レベルでアクセスできる文書が、文書1、文書4、…であることが示されている。また、フィールド410には、acl:GXとして、グループXに帰属されるユーザがアクセスできる文書が、フィールド420において文書3、文書4…であることが索引付けされている。フィールド410に含まれる単語は、検索要求に含まれる検索語として使用され、図示した実施形態では、アクセス権限も検索語と同レベルで検索要求に含まれる。なお、図4に示した転置索引には、文書識別値の他、単語の出現位置を対応付けて登録しておくことができ、少なくとも単語と、文書識別値とが対応付けられている限り、その実装形式には限定はない。
- [0061] ボックス430には、本実施形態で、索引検索実行部216が実行する場合に使用する結合検索式、検索結果、および検索結果をキャッシュする場合のキャッシュ索引生成処理の実施形態を示す。ボックス430内に示された検索要求は、グループGXに属するユーザUAが、自己のプライベート・レベルでの文書検索も含むセキュア検索を要求する場合の実施形態である。ユーザUAは、端末装置102から、自己のユーザID、パスワードなどを入力して検索サーバ110にログオンし、検索要求を、例えば、検索語=ibmとして検索サーバ110に送付する。その後、検索制御部212は、受領した検索式と、包含関係式とを合成する。
- [0062] 説明する実施形態では、検索制御部212がキャッシュ管理部218からキャッシュミスヒットの通知を受け、結合検索式索引検索実行部216の索引検索を説明する。検索サーバ110は、キャッシュミスヒットであると判断した場合、検索制御部212が生成した、「query= ibm AND (acl:UA IN acl:GX)」を修正し、IN演算子を、SQLなどの標準演算子であるOR演算子に修正し、検索式「query=ibm AND (acl:UA OR acl:GX)」を生成する。ボックス430に示した実施形態では、さらに、索引検索実行部216は、q

queryA=ibm AND (acl:UA) および queryB=ibm AND (acl:GX) を生成し、独立した全文索引検索を実行する。なおこれらの検索は、順序的に実行されてもよく、また並列検索が行われてもよい。上述したqueryAは、プライベート・レベルの検索式であり、queryBが、共用レベルの検索式である。

[0063] 図4に示した実施形態では、queryAで、検索語=ibmを含む文書が文書1、文書2、文書4であり、ユーザUAのアクセス権限が文書1および文書4を許容しているので、検索結果Aとして、{文書1、文書4}が生成される。また、グループXのアクセス権限は、文書3、文書4を許容しているので、queryBの検索結果Bは、{文書4}が返される。検索制御部212は、各検索結果A、Bを、特定の目的に応じて、それぞれキャッシュメモリ222に格納する。その際、使用したqueryA、queryBの情報を使用してhash値を計算し、キャッシュ索引とする。さらに検索制御部212は、検索結果A、Bを受領して和集合を生成させ、重複を排除して検索結果={文書1、文書4}を生成する。

[0064] キャッシュ索引を生成するためのhash計算は、hash(ibm+AND+acl:UA)として検索式全体を対象とすることもできるし、hash(ibm)+hash(AND)+hash(acl:UA)として生成することもでき、検索式とキャッシュ内容とを対応付けすることができる限り、いかなるhash計算アルゴリズムでも使用することができる。なお、ハッシュ関数としては、SHA-1、SHA-2、MD-2、MD-4、MD-5どのこれまで知られたいかなるハッシュ関数でも使用することができる。

[0065] 上述したキャッシュ索引生成処理により、少なくともグループXに帰属する全ユーザが同一の検索語を使用して検索要求を発行する場合、プライベート・レベルを除く検索結果のキャッシュ索引は、hash(ibm+AND+acl:GX)となるので、グループXのユーザは、キャッシュ索引hash(ibm+AND+acl:GX)に対応付けられるキャッシュ内容(検索結果)にキャッシュヒットを発生させることができる。一方、プライベート・レベルについてみれば、グループXのユーザMが同一の検索語を発行したとしても、登録されたキャッシュ索引は

、hash(ibm+AND+acl:UA)であり、キャッシュ管理部218が生成するhash値は、hash(ibm+AND+acl:UM)なので、キャッシュミスヒットが発生し、プライベート・レベルのセキュリティが確保できることが理解される。

[0066] 図5は、検索結果をキャッシュする場合の、キャッシュ索引生成処理500を比較した図である。従来のキャッシュ索引生成処理は、ボックス510に示されるように、検索式に含まれる全情報を使用し、全情報をhashする。なお、図5では、hash計算について、hash(ibm, acl:GX, in, acl:GY)として記述するが、図5では、hash計算に使用するデータを例示する目的であり、hash値は、hash計算に使用するべき情報またはhash値をどのように使用して計算することができる。図5に示すように、同一のユーザが同一の検索語および演算子を使用した検索要求でなければキャッシュヒットは発生しない。一方、本実施形態のキャッシュ索引生成処理は、ボックス520で示されるように、ドメイン・レベルでのキャッシュ項目の共有化を行うため、検索式に含まれるプライベート・レベルのアクセス権限に関連する情報を除いた情報を使用してhash値を生成してキャッシュ索引とする。

[0067] なお、本実施形態で、ユーザがプライベート・レベルでの検索を要求する場合、プライベート・レベルに関連する情報のみを含む検索式が生成され、hash値が生成されて、キャッシュメモリ222にキャッシュ索引とともにその検索結果が、キャッシュされる。

[0068] 図6は、本実施形態で、キャッシュメモリ222内に生成されるデータ構造600の実施形態を示す。データ構造600は、キャッシュ索引を登録するフィールド610と、キャッシュ索引により索引付けされるキャッシュ内容（検索結果）とが対応づけられたレコードを含んで生成されている。キャッシュ管理部218は、キャッシュ索引を検索し、ハッシュ値が一致するレコードのキャッシュ内容を、検索結果として検索制御部212に返すことで検索処理の高速化を達成している。

[0069] キャッシュ索引を登録するフィールド610は、検索制御部212が生成した結合検索式が含む検索語（アクセス権限を含む）の中から、プライベー

ト・レベルに関連する情報を除いた検索式のハッシュ値が登録されていて、グループ単位でのキャッシュヒットを達成させている。また、キャッシュ内容は、すでに検索された共用レベルでの検索結果またはプライベート・レベルの検索結果が、フィールド620内に、例えば「DOC(ibm) in GX U DOC(ibm) in GY」などとして登録されている。共用レベルの検索結果およびプライベート・レベルの検索結果は、フィールド610に示されるように、キャッシュ索引で識別されている。なお、キャッシュ内容としては、検索結果の実データを登録しておくこともできるし、キャッシュ容量を効率的に使用することができるように、検索結果の実データを参照させる実アドレス、仮想アドレス、絶対パス名、URI、URLなどを登録しておくことができる。

[0070] 図7は、本実施形態の検索サーバ110が実行する、共用レベルについて検索を実行する場合の文書検索処理のフローチャートを示す。図7の処理は、ステップS700から開始し、ステップS701で、ログオンしたユーザが入力した検索語を含む検索式を取得する。説明する実施形態では、query=ibmが取得される。また、ユーザが検索式としてibm AND patentを入力した場合、query=ibm AND patentが取得される。ステップS702では、ユーザのアクセス権限をユーザ情報としてユーザ管理サーバ120から取得する。ユーザ情報は、ログオンしたユーザのユーザIDなどにより照会されたユーザのプライベート・レベル、ドメイン・レベル、オープンドメイン・レベルなどのアクセス権限を含む情報としてユーザ管理データベース120から取得することができる。

[0071] ステップS703では、取得したユーザの階層レベルごとの包含関係を参照して、例えば、包含関係式を、「acl:UA IN GX IN GY」などとして設定し、包含関係式と、ユーザ入力された検索式、例えば「ibm」や、「ibm AND patent」とをAAND演算子で結合し、結合検索式として、query= {ibm AND (acl:UA IN GX IN GY)} またはquery= {(ibm AND Patent) AND (acl:UA IN GX IN GY)} を生成する。

- [0072] さらに、ステップS703では、上述した結合検索式をキャッシュ管理部218に送付し、包含関係式から、ユーザのプライベート・レベルでのアクセス権限に対応する検索語を除いた情報から共用レベルについての検索式について、hash値、hash(ibm, and, in, GX, in, GY)を計算して、キャッシュ索引を生成する。ステップS704では、キャッシュ管理部218が、キャッシュメモリ222のキャッシュ索引を検索し、キャッシュヒットか否かを判断する。
- [0073] ステップS704でキャッシュヒットと判断された場合(yes)、処理をステップS705に進め、キャッシュメモリからhash値に対応するキャッシュ内容を読み込んで検索結果を取得し、ステップS706で取得した内容を検索結果作成部214に送付して検索結果を作成し、検索要求の発行元にネットワーク108を介して送付し、ステップS712で、一連のトランザクションを終了する。
- [0074] 一方、ステップS704でキャッシュミスヒットと判断された場合(no)、処理をステップS707に分岐させ、索引検索実行部216において共用レベルの検索式を作成する。この処理は、例えば、結合検索式中の包含関係式内に含まれる、プライベート・レベルのアクセス権限および直後のIN演算子までを分離することにより行うことができる。その後、結合検索式中のIN演算子を、OR演算子に変換し、通常のSQLパーザが処理可能な結合検索式に修正する。その後、ステップS708でSQLなどを使用して文書データベース220の索引リスト格納部222bを走査して文書の全文検索をOR演算子単位で実行する。
- [0075] ステップS709では、キャッシュメモリ222の容量を検査し、キャッシュメモリ222の容量が超過する場合(yes)、ステップS710でキャッシュ項目のうち不要なレコードを削除する。キャッシュ項目の削除は、種々の方法および基準を使用して実行することができ、アクセス頻度が一定のしきい値よりも低いことを基準として採用することもできるし、タイムスタンプなど、最新にアクセスされた時刻が古いものを削除するなど、これま

でキャッシュ項目の更新のために利用されている削除方法を適宜使用して実行することができる。また、ステップS 7 0 9の判断で、キャッシュメモリの容量が超過していないと判断された場合（no）、直ちに処理をステップS 7 1 1に分岐させる。

[0076] ステップS 7 1 1では、IN演算子を含む検索式からキャッシュ索引を生成し、取得した検索結果をキャッシュ索引に対応付けてキャッシュメモリ2 2 2のレコードとして登録し、キャッシュ項目を更新する。その後、各検索結果を、検索結果作成部2 1 4に送付して和集合を生成して検索結果を生成させる。なお、包含関係式から分離された、プライベート・レベルの検索である、{(ibm AND Patent) AND (acl:UA)}に対応する検索を実行する実施形態については、図8の処理で説明する。

[0077] 図8は、図7で説明した、共用レベルの他、プライベート・レベルでアクセス可能な文書を含む検索結果を生成する文書検索処理のフローチャートである。なお、図8の実施形態では、プライベート・レベルはキャッシュヒットしなかったが、共用レベルについては、キャッシュヒットが発生したものとして説明する。

[0078] 図8に示した実施形態の処理は、概ね図7で説明した処理と共通するが、図8で説明する実施形態では、共用レベルの検索結果については、キャッシュヒットが発生し、ステップS 8 0 5でキャッシュ内容を読み込み、ステップS 8 0 6に渡される。一方、プライベート・レベルについてはステップS 8 0 4でキャッシュヒットせず（no）、ステップS 8 0 8で、包含関係式から分離した(acl:UA IN)からプライベート・レベルでの検索式{ibm AND (acl:UA)}を作成し、ステップS 8 0 9で、プライベート・レベルでの検索式{ibm AND (acl:UA)}を使用して独立して検索を実行する。

[0079] ステップS 8 1 0で、キャッシュ容量が超過しているかを判断し、超過していなければ（no）、ステップS 8 1 2で、プライベート・レベルでの検索結果を、そのhash値をキャッシュ索引とし、キャッシュ項目として対応付けてキャッシュメモリ2 2 2に格納する。また、ステップS 8 1 0の判断で

キャッシュ容量釣果と判断された場合（yes）、ステップS810で、キャッシュのうち不要なレコードを削除し、ステップS812でキャッシュする。

[0080] その後、図8の処理では、処理をステップS806に処理を分岐させ、プライベート・レベルでの検索結果を取得し、予めバッファリングされていた共用レベルの検索結果を、検索結果作成部214において両検索結果の和集合として検索結果集合を作成し、ステップS807で、最終的な検索結果を検索要求の発行元に返す処理を実行する。

[0081] 図8に示した文書検索処理では、結合検索式にプライベート・レベルでのアクセス権限が含まれている場合にも効率的にユーザ入力を利用するセキュリティ検索が可能となる。また、プライベート・レベルのアクセス権限の検索結果を、ドメイン・レベルの検索結果と区別してキャッシュ項目とすることで、プライベート・レベルでのセキュア検索要求を実行する場合でも、キャッシュヒットが発生すれば索引検索を全く実行することなく各キャッシュ内容の和集合を生成することが可能となり、プライベート・レベルでのアクセス権限を含めた全検索結果をキャッシュする場合に比較して大きな処理負荷を生じることなく、同一の検索結果を生成させることができる。

[0082] さらに、特定のユーザBのグループ帰属が変更された場合でも、ユーザBは、すでにキャッシュされた変更先のドメイン・レベルでの検索結果をキャッシュメモリ222から検索でき、一方、ユーザBのプライベート・レベルでの検索結果がキャッシュされている場合、全文索引検索を再実行させることなく、ユーザのグループ帰属変更に対応することができる。

[0083] なお、キャッシュメモリ222のキャッシュ項目は、プライベート・レベル、ドメイン・レベル、またはオープンドメイン・レベルでのアクセス権限が修正される場合にはクリアされ、新規なアクセス権限設定の下でのセキュア検索が行われる。また、キャッシュメモリ222のキャッシュ項目のリフレッシュまたは同期化などの処理のため、これまで知られたいかなる方法でも使用することができる。

- [0084] 図9は、本実施形態のセキュア文書検索によるOR処理の削減およびキャッシュ項目の利用性の改善を、プライベート・レベルでの検索結果を含ませる図8の処理についての結合検索式900を使用して説明する図である。結合検索式の検索を実行した段階で、まったくキャッシュヒットしない場合、結合検索式910で示されるように、OR演算は、プライベート・レベル、ドメイン・レベル、オープンドメイン・レベルの3階層について行われる。このOR演算は、階層レベルが増加するに連れて増加し、セキュア検索の検索効率、文書数の他、階層数にも依存して計算オーバーヘッドが増大することが示される。
- [0085] 一方、本実施形態によれば、少なくともプライベート・レベル以上の階層レベルでの検索結果がキャッシュされているので、結合検索式920で示すように、プライベート・レベルでの検索結果と、キャッシュされた検索結果とをOR演算することで、結合検索式910と同一の検索結果が生成でき、検索効率をはるかに改善することが可能となる。
- [0086] さらに他の実施形態でプライベート・レベルでの検索結果も別途キャッシュ索引を生成してキャッシュする実施形態では、結合検索式930で示すようにキャッシュ項目同士のOR演算を実行することで結合検索式910と同一の結果を生成することができるので、包含関係を含めて結合検索式を生成する処理が追加されたとしても、キャッシュ項目が増加すればするほど全文索引を使用する検索処理が大きく削減でき、検索サーバ110の検索処理は大きく改善される。
- [0087] 図10は、本実施形態の文書検索システム100による検索効率改善の機能を説明した図である。図10に示すように、従来の検索結果のキャッシュ方式では、同一のユーザが同一の検索式を入力した場合を除き、キャッシュヒットは生成されず、キャッシュヒットは、領域1010で示す程度の著しく低い確率となる。一方、本実施形態では、図10に示す例示的な実施形態においてもキャッシュヒットは、領域1010に比較して少なくともユーザ数分だけ増加し、この結果、キャッシュヒットが発生する確率は、領域10

20で示すように、少なくとも3倍に増加する。

[0088] さらに、キャッシュヒットは発生する確率は、図10から示されるように、ドメイン・レベルに帰属されるユーザ数が増加するにつれて線形的に増加することになり、ユーザ数の増加が逆に検索効率を向上させることが可能となる。加えて、本実施形態では、階層レベルごとのOR演算を、0または1まで削減することに加え、キャッシュメモリを併用することが可能となり、本実施形態によれば、全文索引を使用するセキュア検索の検索効率を著しく改善することが可能となる。

[0089] これまで説明したように、本発明によれば、全文検索を使用するセキュア検索の検索効率を著しく改善することが可能な文書検索システム、情報処理装置、文書検索方法およびプログラムを提供することが可能となる。

[0090] 本実施形態の上記機能は、C、C++、Java（登録商標）、Java（登録商標）Beans、Java（登録商標）Applet、Java（登録商標）Script、Perl、Ruby、PYTHONなどのオブジェクト指向プログラミング言語などで記述された装置実行可能なプログラムにより実現でき、当該プログラムは、ハードディスク装置、CD-ROM、MO、フレキシブルディスク、EEPROM、EPROMなどの装置可読な記録媒体に格納して頒布することができ、また他装置が可能な形式でネットワークを介して伝送することができる。

[0091] これまで本実施形態につき説明してきたが、本発明は、上述した実施形態に限定されるものではなく、他の実施形態、追加、変更、削除など、当業者が想到することができる範囲内で変更することができ、いずれの態様においても本発明の作用・効果を奏する限り、本発明の範囲に含まれるものである。

符号の説明

[0092] 100…文書検索システム、102、104、106…端末装置、108…ネットワーク、110…検索サーバ、120…ユーザ管理サーバ、200…機能ブロック（検索サーバ）、202…ネットワーク・アダプタ、204…

データベース管理部、206…文書検索部、208…ユーザ情報取得部、210…検索式取得部、212…検索制御部、214…検索結果作成部、216…索引検索実行部、218…キャッシュ管理部、220…文書データベース、222…キャッシュメモリ

請求の範囲

[請求項1]

文書を検索するための情報処理装置であって、前記情報処理装置は、

前記文書に対するアクセス権限を含ませた索引を生成して、前記文書とともに管理するデータベース管理部と、

前記情報処理装置にアクセスする前記アクセス権限をユーザ情報として取得するユーザ情報取得部と、

前記ユーザ情報および検索式を含む検索要求を取得し、前記ユーザ情報から検索要求の発行元のアクセス権限についての包含関係を指定する演算子を含む包含関係式を前記検索式に結合して結合検索式を生成し、前記結合検索式から、複数のユーザ識別値を含むドメイン内で前記複数のユーザ識別値の下に共有できる共用レベルの検索結果を取得する文書検索部と、

前記共用レベルであることを識別させたキャッシュ項目として前記共用レベルの検索結果を登録するキャッシュ管理部と、

前記共用レベルのキャッシュ項目がない場合に前記結合検索式に基づいて索引検索を実行し、前記共用レベルの検索結果を生成する索引検索実行部と

を含む、情報処理装置。

[請求項2]

前記文書検索部は、

前記検索要求から前記検索式を取得する検索式取得部と、

前記検索式取得部が取得した前記検索式と、前記ユーザ情報から生成した前記包含関係式を作成し、前記検索式に結合して前記結合検索式を生成し、前記結合検索式を使用して前記キャッシュ管理部および前記索引検索実行部の検索を制御する検索制御部と、

前記検索要求に対応する検索結果を取得し、前記検索要求の発行元に送付させる、検索結果作成部と

を含む、請求項1に記載の情報処理装置。

- [請求項3] 前記アクセス権限は、セキュア検索のため、ユーザのみがアクセスできるプライベート・レベル、複数の前記ユーザを含むドメイン・レベル、および複数のドメイン・レベルを含むオープンドメイン・レベルごとにそれぞれ設定される、請求項1に記載の情報処理装置。
- [請求項4] 前記索引は、転置索引として生成された全文索引である、請求項3に記載の情報処理装置。
- [請求項5] さらに前記文書および前記転置索引を管理する文書データベースと、前記検索結果をキャッシュするキャッシュメモリとを含み、前記キャッシュメモリは、前記共用レベルであることを示すキャッシュ索引と前記共用レベルの検索結果とを対応付け、前駆プライベート・レベルではキャッシュヒットしない前記キャッシュ項目として登録する、請求項4に記載の情報処理装置。
- [請求項6] ネットワークを介して文書を検索するための文書検索システムであって、前記文書検索システムは、
ユーザ識別値を使用して文書検索のための検索式を含む検索要求を前記ネットワークを介して送付する端末装置と、
前記端末装置からの前記検索要求を受領して文書検索を実行し、検索結果を前記端末装置に返す検索サーバとを含み、前記検索サーバは、
前記文書に対するアクセス権限を含ませた索引を生成して、前記文書とともに管理するデータベース管理部と、
前記情報処理装置にアクセスする前記アクセス権限をユーザ情報として取得するユーザ情報取得部と、
前記ユーザ情報および検索式を含む検索要求を取得し、前記ユーザ情報から検索要求の発行元のアクセス権限についての包含関係を指定する演算子を含む包含関係式を前記検索式に結合して結合検索式を生成し、前記結合検索式から、複数のユーザ識別値を含むドメイン内で前記複数のユーザ識別値の下に共有できる共用レベルの検索結果を取

得する文書検索部と、

前記共用レベルであることを識別させたキャッシュ項目として前記共用レベルの検索結果を登録するキャッシュ管理部と、

前記共用レベルのキャッシュ項目がない場合に前記結合検索式に基づいて索引検索を実行し、前記共用レベルの検索結果を生成する索引検索実行部と

を含む、文書検索システム。

[請求項7]

前記文書検索部は、

前記検索要求から前記検索式を取得する検索式取得部と、

前記検索式取得部が取得した前記検索式と、前記ユーザ情報から生成した前記包含関係式を作成し、前記検索式に結合して前記結合検索式を生成し、前記結合検索式を使用して前記キャッシュ管理部および前記索引検索実行部の検索を制御する検索制御部と、

前記検索要求に対応する検索結果を取得し、前記検索要求の発行元に送付させる、検索結果作成部と

を含む、請求項6に記載の文書検索システム。

[請求項8]

前記ネットワークを介して接続され、前記ユーザ識別値に関連する前記アクセス権限を管理するユーザ管理サーバを含み、前記ユーザ管理サーバは、前記アクセス権限を、セキュア検索のため、ユーザのみがアクセスできるプライベート・レベル、複数の前記ユーザを含むドメイン・レベル、および複数のドメイン・レベルを含むオープンドメイン・レベルごとにそれぞれ管理する、請求項7に記載の文書検索システム。

[請求項9]

前記索引は、転置索引として生成された全文索引である、請求項8に記載の文書検索システム。

[請求項10]

さらに前記文書および前記転置索引を管理する文書データベースと、前記検索結果をキャッシュするキャッシュメモリとを含み、前記キャッシュメモリは、前記共用レベルであることを示すキャッシュ索引

と前記共用レベルの検索結果とを対応付け、前記プライベート・レベルではキャッシュヒットしない前記キャッシュ項目として登録する、請求項9に記載の文書検索システム。

[請求項11]

文書を検索するための情報処理装置が実行する文書検索方法であって、前記文書検索方法は、

前記文書に対する前記アクセス権限を含ませた索引を生成して、前記文書とともに管理するステップと、

前記情報処理装置にアクセスするアクセス権限をユーザ情報として取得するステップと、

前記ユーザ情報および検索式を含む検索要求を取得し、前記ユーザ情報から検索要求の発行元のアクセス権限についての包含関係を指定する演算子を含む包含関係式を生成するステップと、

前記包含関係式を前記検索式に結合して結合検索式を生成し、前記結合検索式を使用して複数のユーザ識別値を含むドメイン内で前記複数のユーザ識別値の下で共用できる共用レベルの文書を検索するステップと、

共用レベルの検索結果であることを識別するキャッシュ索引を生成するステップと、

前記共用レベルの検索結果を前記キャッシュ索引と対応付けてキャッシュ項目として登録するステップと、

少なくとも前に共用レベルの検索結果を含む検索結果集合を生成するステップと

を実行する、コンピュータ実行可能な文書検索方法。

[請求項12]

前記検索するステップは、

前記結合検索式から生成される、前記共用レベルの検索結果を生成するための検索式からhash値を生成するステップと、

前記hash値を使用して前記キャッシュ項目を検索するステップと、

前記キャッシュ項目の検索がキャッシュヒットを与える場合、前記

キャッシュ項目に登録された検索済結果を前記共用レベルの検索結果とし、前記キャッシュ項目の検索がキャッシュミスを与える場合に、前記結合検索式に基づいて索引検索を実行し、前記共用レベルの検索結果とするステップと

を含む、請求項 1 1 に記載の文書検索方法。

[請求項13]

前記検索するステップは、さらに

前記結合検索式から生成される、前記複数のユーザ識別値が共用できないプライベート・レベルの検索結果を生成するための検索式を生成するステップと、

前記プライベート・レベルの検索結果を生成するための検索式を使用して前記プライベート・レベルでの検索結果を取得するステップと、

前記プライベート・レベルの検索結果を、前記プライベート・レベルの検索結果を生成するための検索式のhash値をキャッシュ索引として、前記プライベート・レベルの検索結果に対応付けてキャッシュ項目として登録するステップと

を含む、請求項 1 2 に記載の文書検索方法。

[請求項14]

前記検索結果集合を生成するステップは、前記共用レベルおよび前記プライベート・レベルについての各検索結果から和集合を生成するステップを含む、請求項 1 3 に記載の文書検索方法。

[請求項15]

前記アクセス権限は、セキュア検索のため、ユーザのみがアクセスできるプライベート・レベル、複数の前記ユーザを含むドメイン・レベル、および複数のドメイン・レベルを含むオープンドメイン・レベルごとにそれぞれ設定され、さらに、前記索引は、転置索引として生成された全文索引である、請求項 1 4 に記載の文書検索方法。

[請求項16]

コンピュータが実行する文書検索方法のためのコンピュータ実行可能なプログラムであって、前記プログラムは、コンピュータが、

前記文書に対する前記アクセス権限を含ませた索引を生成して、前

記文書とともに管理するステップと、

前記コンピュータにアクセスするアクセス権限をユーザ情報として取得するステップと、

前記ユーザ情報および検索式を含む検索要求を取得し、前記ユーザ情報から検索要求の発行元のアクセス権限についての包含関係を指定する演算子を含む包含関係式を生成するステップと、

前記包含関係式を前記検索式に結合して結合検索式を生成し、前記結合検索式を使用して複数のユーザ識別値を含むドメイン内で前記複数のユーザ識別値の下で共用できる共用レベルの文書を検索するステップと、

共有レベルの検索結果であることを識別するキャッシュ索引を生成するステップと、

前記共用レベルの検索結果を前記キャッシュ索引と対応付けてキャッシュ項目として登録するステップと、

少なくとも前に共用レベルの検索結果を含む検索結果集合を生成するステップと

を実行するための、コンピュータ実行可能なプログラム。

[請求項17]

前記検索するステップは、

前記結合検索式から生成される、前記共用レベルの検索結果を生成するための検索式からhash値を生成するステップと、

前記hash値を使用して前記キャッシュ項目を検索するステップと、

前記キャッシュ項目の検索がキャッシュヒットを与える場合、前記キャッシュ項目に登録された検索済結果を前記共用レベルの検索結果とし、前記キャッシュ項目の検索がキャッシュミスを与える場合に、前記結合検索式に基づいて索引検索を実行し、前記共用レベルの検索結果とするステップと

を含む、請求項18に記載のプログラム。

[請求項18]

前記検索するステップは、さらに

前記結合検索式から生成される、前記複数のユーザ識別値が共用できないプライベート・レベルの検索結果を生成するための検索式を生成するステップと、

前記プライベート・レベルの検索結果を生成するための検索式を使用して前記プライベート・レベルでの検索結果を取得するステップと、

前記プライベート・レベルの検索結果を、前記プライベート・レベルの検索結果を生成するための検索式のhash値をキャッシュ索引として、前記プライベート・レベルの検索結果に対応付けてキャッシュ項目として登録するステップと

を含む、請求項17に記載のプログラム。

[請求項19]

前記検索結果集合を生成するステップは、前記共用レベルおよび前記プライベート・レベルについての各検索結果から和集合を生成するステップを含み、前記アクセス権限は、セキュア検索のため、ユーザのみがアクセスできるプライベート・レベル、複数の前記ユーザを含むドメイン・レベル、および複数のドメイン・レベルを含むオープンドメイン・レベルごとにそれぞれ設定され、さらに、前記索引は、転置索引として生成された全文索引である、請求項18に記載のプログラム。

[請求項20]

文書を検索するための情報処理装置であって、前記情報処理装置は、

前記文書に対するアクセス権限を含ませた索引を生成して、前記文書とともに管理するデータベース管理部と、

前記情報処理装置にアクセスするユーザ識別値について前記アクセス権限をユーザ情報として取得するユーザ情報取得部と、

前記検索要求から検索式を取得する検索式取得部と、

前記検索式取得部が取得した前記検索式と、前記ユーザ情報から生成した検索要求の発行元のアクセス権限についての包含関係を指定す

る演算子を含む前記包含関係式を作成し、前記検索式に結合して前記結合検索式を生成し、前記結合検索式を使用して検索を制御し、複数のユーザ識別値を含むドメイン内で前記複数のユーザ識別値の下で共用できる共用レベルの検索結果を取得する検索制御部と、

前記検索要求に対応する検索結果を取得し、前記検索要求の発行元に送付させる検索結果作成部とを含む文書検索部と、

前記共用レベルのキャッシュ項目として前記共用レベルの検索結果と対応付けて登録するキャッシュ管理部と、

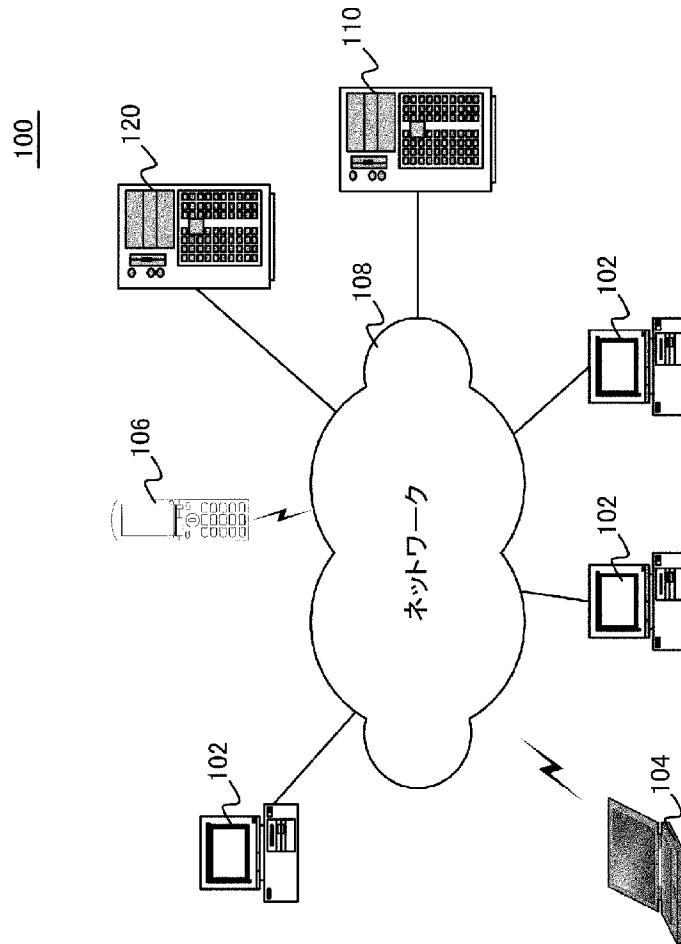
前記共用レベルのキャッシュ項目がない場合に前記結合検索式に基づいて索引検索を実行し、前記共用レベルの検索結果を生成する索引検索実行部と、

前記文書および前記転置索引を管理する文書データベースと、

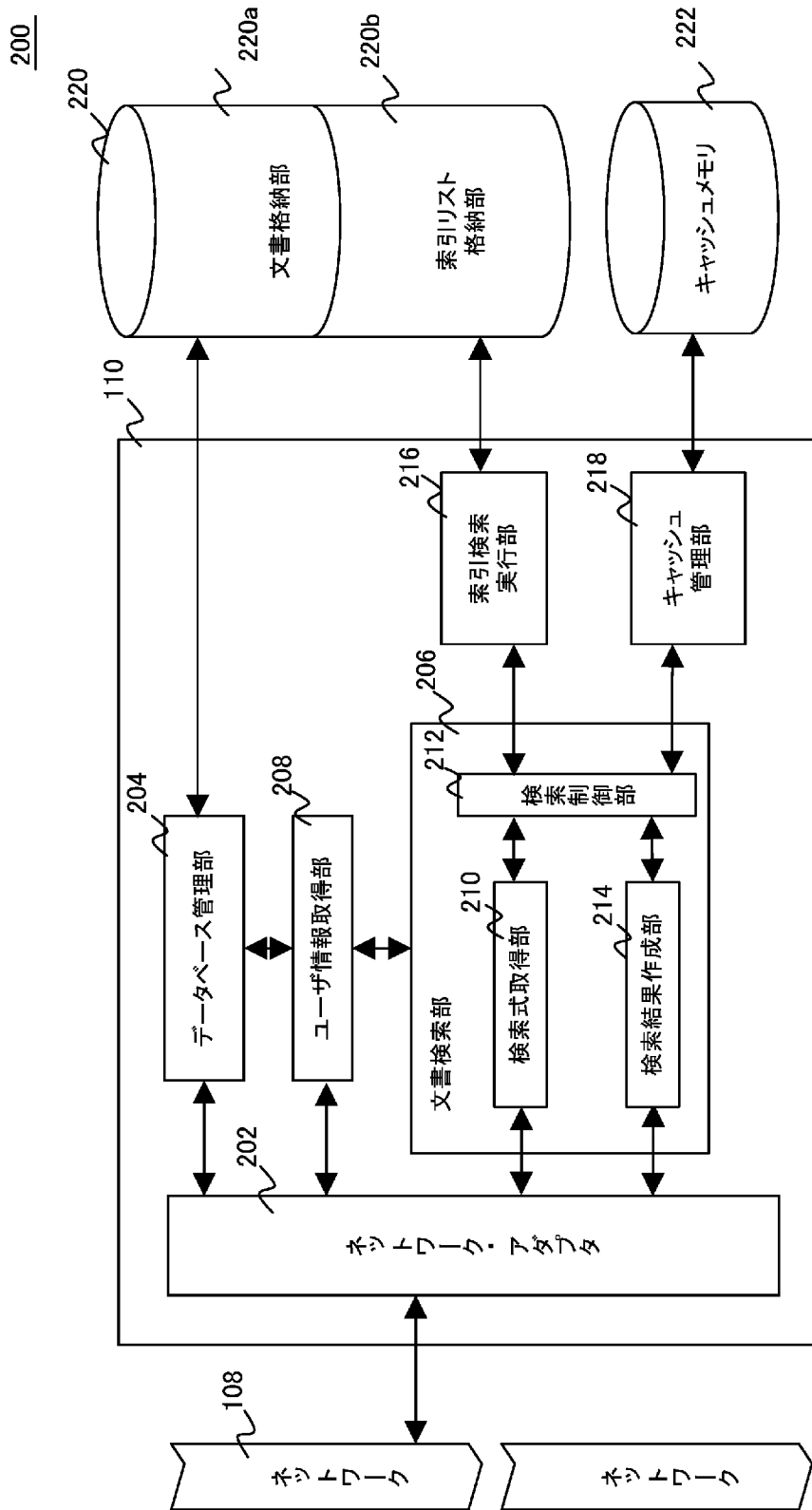
前記共有レベルの検索結果をキャッシュするキャッシュメモリと、
を含み、

前記アクセス権限は、セキュア検索のため、ユーザのみがアクセスできるプライベート・レベル、複数の前記ユーザを含むドメイン・レベル、および複数のドメイン・レベルを含むオープンドメイン・レベルごとにそれぞれ設定され、前記索引は、転置索引として生成された全文索引であり、前記キャッシュメモリは、前記共用レベルであることを示すキャッシュ索引と前記共用レベルの検索結果とを対応付け、前駆プライベート・レベルではキャッシュヒットしない前記キャッシュ項目として登録する、情報処理装置。

[図1]

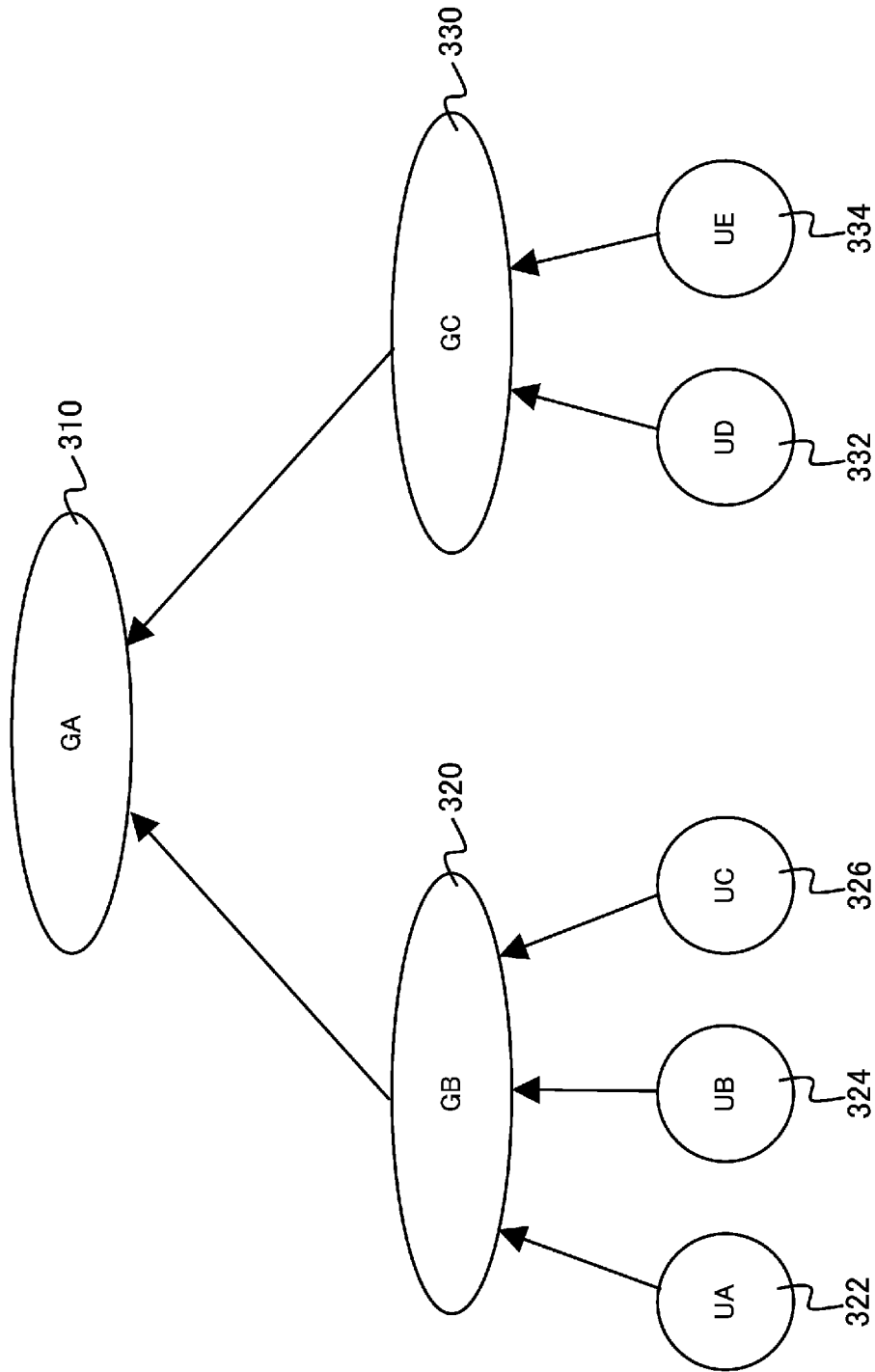


[図2]



[図3]

300



[図4]

400

410 単語	420 出現箇所リスト
ibm	文書 1, 文書 2, 文書 4 ...
patent	文書 2, 文書 3, 文書 4 ...
acl:UA	文書 1, 文書 4 ...
acl:GX	文書 3, 文書 4 ...

430

セキュア検索の検索文:ユーザUAが、グループGXに属する場
合
: query=ibm AND (acl:UA OR acl:GX)
: 検索結果={文書1、文書4}
キャッシュ索引=hash (ibm, and, or, acl:GX)

[図5]

500

520

セキュア検索の検索文:
ユーザUAが、グループGYのグループGXに属する場合:
query=ibm AND (acl:UA IN acl:GX IN acl:GY)
キャッシュ索引=hash (ibm, and acl:GX, in, acl:GY)



510

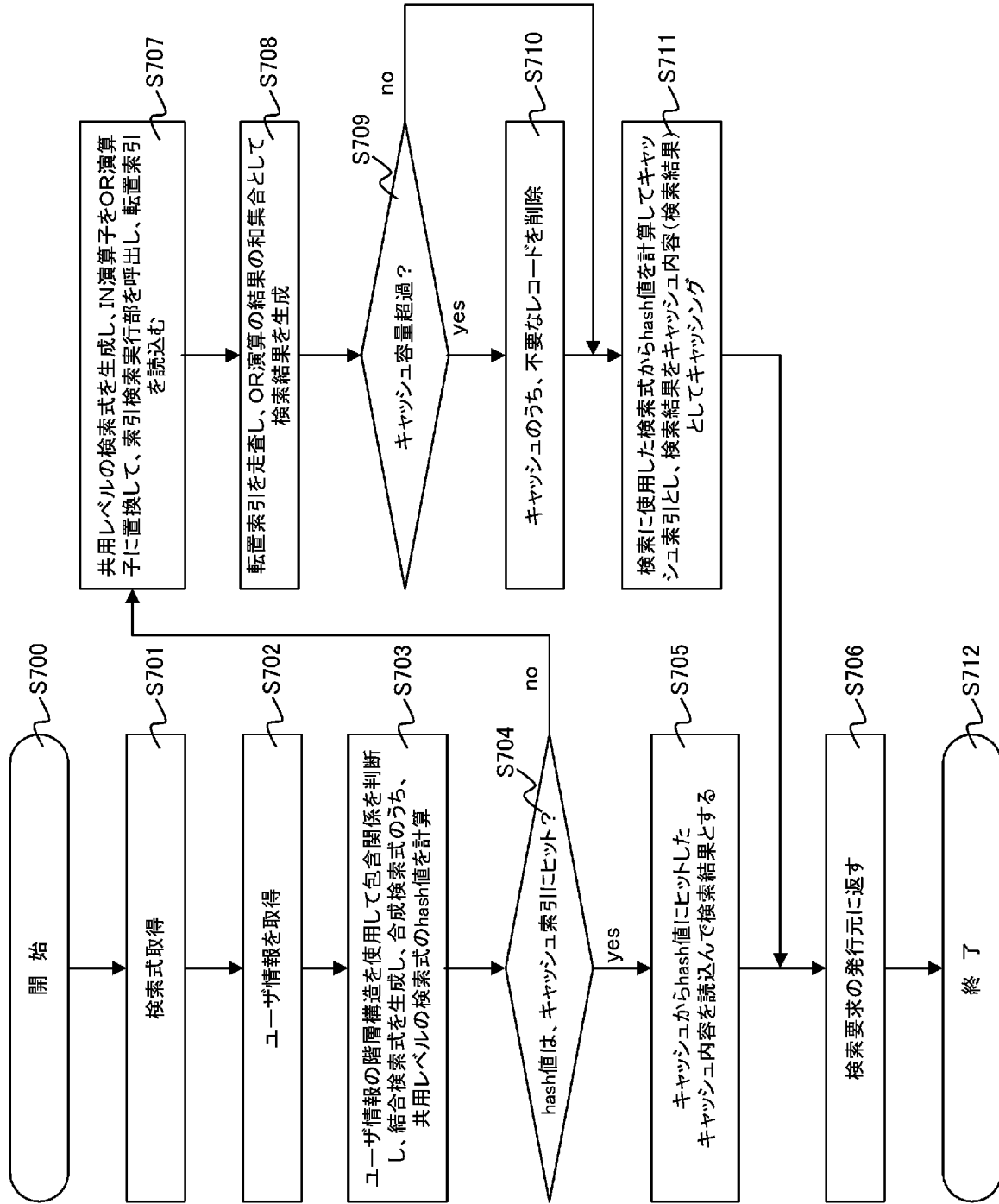
セキュア検索の検索文:
ユーザUAが、グループGYのグループGXに属する場合:
query=ibm AND (acl:UA OR acl:GX OR acl:GY)
キャッシュ索引=hash (ibm, and, acl:UA, or, acl:GX, or, acl:GY)

[図6]

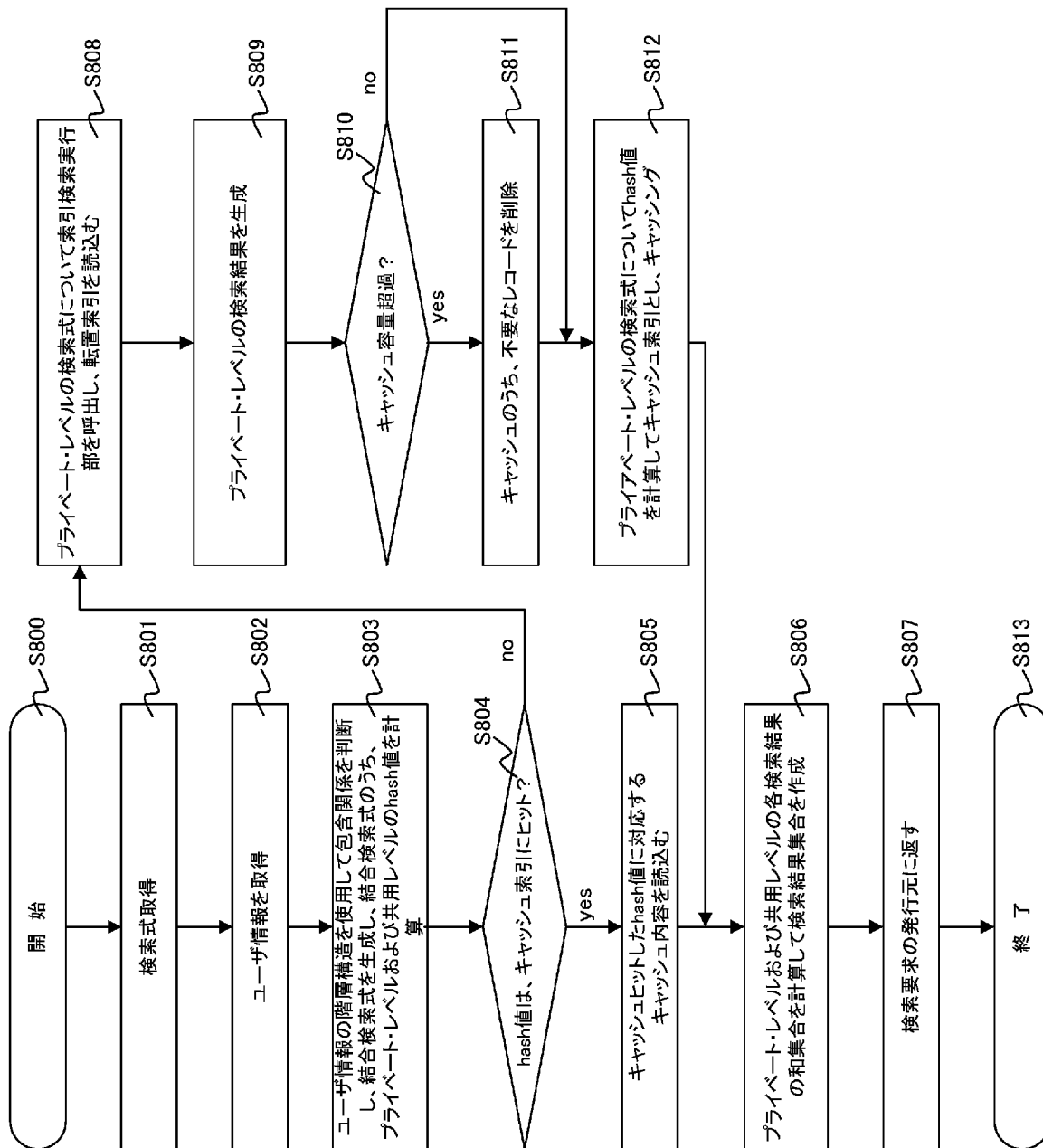
600

610 キャッシュ索引	620 キャッシュ内容(検索結果)
hash (ibm, acl:GX, in, acl:GY)	DOC(ibm) in GX U DOC(ibm) in GY
hash (ibm, acl:GX, in, acl:GY, in acl:GZ)	DOC(ibm) in GX U DOC(ibm) in GY U DOC(ibm) in GZ
hash (ibm, acl:GX)	DOC(ibm) in GX
hash (ibm, acl:UA)	DOC(ibm) in UA
:	:
:	:
:	:

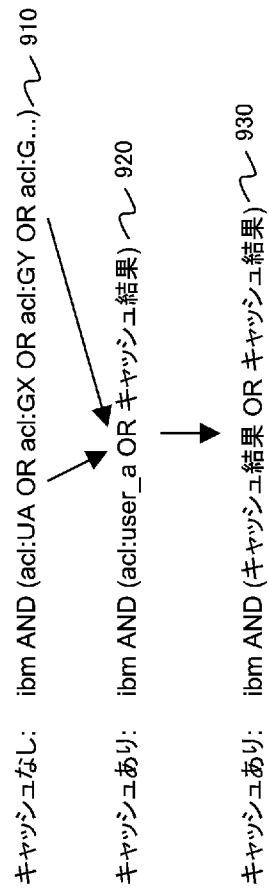
[図7]



[図8]

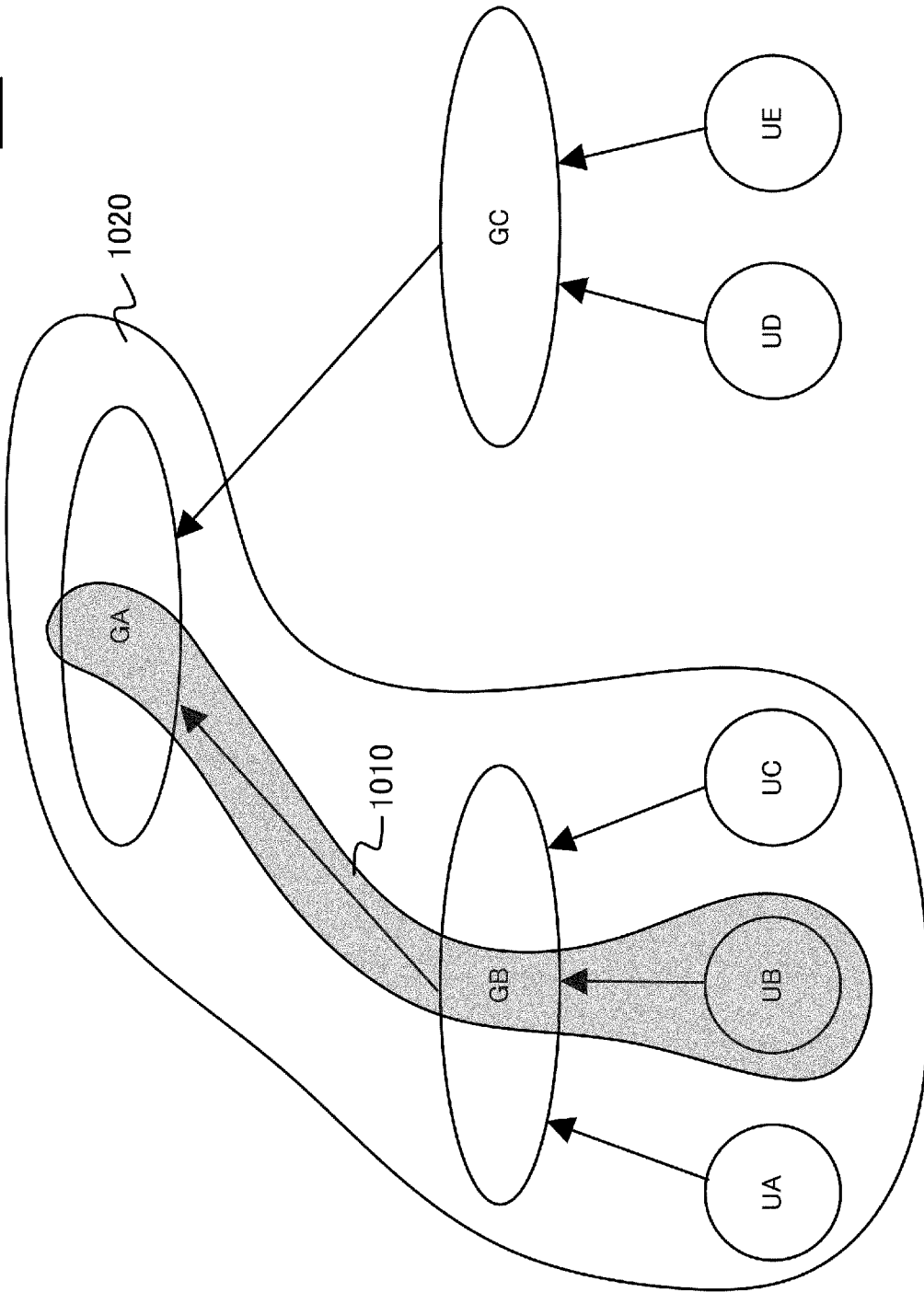


[図9]

900

[] 10

1000



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

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

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	1971-2009年
日本国実用新案登録公報	1996-2009年
日本国登録実用新案公報	1994-2009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)

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
A	JP 2001-344245 A (富士通株式会社) 2001. 12. 14, 全文全図 & US 2001-27451 A1	1-20
A	JP 2003-162441 A (沖電気株式会社) 2003. 06. 06, 全文全図 & US 2003-101200 A1	1-20
A	JP 9-325968 A (マイクロソフト コーポレーション) 1997. 12. 16, 段落【0027】 & US 6199082 B & EP 774722 A	1-20

C欄の続きにも文献が列挙されている。

パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

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

の日の後に公表された文献

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

国際調査を完了した日

28. 09. 2009

国際調査報告の発送日

13. 10. 2009

国際調査機関の名称及びあて先

日本国特許庁 (ISA/J P)
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)

長 由紀子

5M

4233

電話番号 03-3581-1101 内線 3599