

(19) 日本国特許庁(JP)

## (12) 特許公報(B2)

(11) 特許番号

特許第6098294号  
(P6098294)

(45) 発行日 平成29年3月22日(2017.3.22)

(24) 登録日 平成29年3月3日(2017.3.3)

(51) Int.Cl.	F 1
G06F 21/62	(2013.01)
G09C 1/00	(2006.01)
H04L 9/32	(2006.01)
G06F 17/30	(2006.01)
GO6F 21/62	GO6F 21/62 3 5 4
GO9C 1/00	GO9C 1/00 6 6 0 D
HO4L 9/32	HO4L 9/00 6 7 3 C
G06F 17/30	GO6F 17/30 1 2 0 A

請求項の数 8 (全 22 頁)

(21) 出願番号	特願2013-70339 (P2013-70339)
(22) 出願日	平成25年3月28日(2013.3.28)
(65) 公開番号	特開2014-194621 (P2014-194621A)
(43) 公開日	平成26年10月9日(2014.10.9)
審査請求日	平成27年12月4日(2015.12.4)

(73) 特許権者	000005223 富士通株式会社 神奈川県川崎市中原区上小田中4丁目1番 1号
(74) 代理人	100107766 弁理士 伊東 忠重
(74) 代理人	100070150 弁理士 伊東 忠彦
(74) 代理人	100146776 弁理士 山口 昭則
(72) 発明者	小櫻 文彦 神奈川県川崎市中原区上小田中4丁目1番 1号 富士通株式会社内

最終頁に続く

(54) 【発明の名称】情報秘匿化装置、情報秘匿化方法

## (57) 【特許請求の範囲】

## 【請求項 1】

識別情報を含むデータが所定の格納数の格納場所に入力されて、入力された前記データを、前記格納場所から、入力順序が早い入力データに対して入力順序が遅い入力データを優先的に、ランダムの順序で出力する第1のランダム化装置と、

前記第1のランダム化装置から出力された前記データの前記識別情報をトークンで置換するトークン化部と、を備えた情報秘匿化装置。

## 【請求項 2】

前記第1のランダム化装置は、

前記格納場所を有して、入力された前記データを前記格納場所に逐次格納する第1の格納部と、

前記第1の格納部の格納数に応じた第1の乱数を生成させる第1の乱数生成部と、を備え、

前記第1の乱数によってランダムに特定される前記第1の格納部の格納場所から格納された前記データを取り出して出力する請求項1に記載の情報秘匿化装置。

## 【請求項 3】

前記トークンを発生させるトークン発生器と、

前記トークン発生器で発生された前記トークンが入力されて、入力された前記トークンをランダムの順序で出力する第2のランダム化装置と、をさらに備えた請求項1又は2記載の情報秘匿化装置。

## 【請求項 4】

前記第2のランダム化装置は、  
所定の格納数の格納場所を有して、入力された前記トーケンを前記格納場所に逐次格納する第2の格納部と、  
前記第2の格納部の格納数に応じた第2の乱数を生成させる第2の乱数生成部と、を備え、  
前記第2の乱数によってランダムに特定される前記第2の格納部の格納場所から格納された前記トーケンを取り出して出力する請求項3に記載の情報秘匿化装置。

## 【請求項 5】

前記第2の格納部の格納場所は、入力順序が早い入力データに対して入力順序が遅い入力データを優先的に取り出す請求項4に記載の情報秘匿化装置。 10

## 【請求項 6】

識別情報を含むデータが所定の格納数の格納場所に入力されて、入力された前記データを、前記格納場所から、入力順序が早い入力データに対して入力順序が遅い入力データを優先的に、ランダムの順序で出力する第1のランダム化処理と、

前記第1のランダム化処理で出力された前記データの前記識別情報をトーケンで置換するトーケン化処理と、をコンピュータが実行する識別情報の情報秘匿化方法。

## 【請求項 7】

データに含まれる識別情報をトーケンで置換するトーケン化部と、  
前記トーケン化部で出力されたデータが所定の格納数の格納場所に入力されて、入力された該データを、前記格納場所から、入力順序が早い入力データに対して入力順序が遅い入力データを優先的に、ランダムの順序で出力する第一のランダム化装置と、を備えたことを特徴とする識別情報の情報秘匿化装置。 20

## 【請求項 8】

データに含まれる識別情報をトーケンで置換するトーケン化処理と、  
前記トーケン化処理で出力されたデータが所定の格納数の格納場所に入力されて、入力された該データを、前記格納場所から、入力順序が早い入力データに対して入力順序が遅い入力データを優先的に、ランダムの順序で出力する第一のランダム化処理と、をコンピュータが実行する識別情報の情報秘匿化方法。

## 【発明の詳細な説明】 30

## 【技術分野】

## 【0001】

本発明は、情報秘匿化装置、及び情報秘匿化方法に関する。

## 【背景技術】

## 【0002】

クラウドコンピューティングの発達に伴い、ネットワークを経由して、データベースに蓄積された識別情報を含むデータを外部の情報処理サービスへ送信して分析等を依頼することが増えている。識別情報は、利用者の個人情報やIDなど、個人を特定可能な情報を含んでいる場合があるため、情報漏洩の防止のために、外部にデータを送信する際には識別情報を秘匿化していた。 40

## 【0003】

一方、データベースシステムの仕様によっては、蓄積されたデータを、例えば記録した日時順に出力する等、データベースの仕様に基づく特定の出現パターンを有したものが多く、単に識別情報を秘匿化するのみでは、そのデータの並び順から利用者の特定が可能になってしまう場合があった。

## 【0004】

例えば、特定の疾患有する患者のデータを月別で解析する場合、喻え患者の識別情報を秘匿化したとしても、データベースへの登録順序でデータを出力した場合、データの並び順である程度の患者の特定ができてしまう場合があった。

## 【0005】 50

そこで、従来のデータベースシステムにおいては、例えば、(SELECT \* FROM table ORDER BY RAND())等のコマンドを使用して、蓄積データをランダムにソート処理して出力させる方法が用いられていた。

【0006】

また、利用者の識別情報を管理するトークンにより管理する方法があった。さらに、データベースにおける記憶領域内のデータ格納率を向上させる目的として、同時に扱うデータの集合を1つのまとまりで扱う方法があった（例えば、特許文献1-3等参照）。

【先行技術文献】

【特許文献】

【0007】

10

【特許文献1】特開2005-284679号公報

【特許文献2】特開2011-211593号公報

【特許文献3】特開2008-276336号公報

【発明の概要】

【発明が解決しようとする課題】

【0008】

しかし、上記先行技術において、ソートを使用したランダム化処理は、データのレコード数が増えると累積的に処理時間が遅くなるという問題があった。

【0009】

そこで、一側面では、識別情報を安全にかつ高速に秘匿化する情報秘匿化装置を提供することを目的とする。

20

【課題を解決するための手段】

【0010】

一つの案では、情報秘匿化装置は、識別情報を含むデータが所定の格納数の格納場所に入力されて、入力された前記データを、前記格納場所から、入力順序が早い入力データに対して入力順序が遅い入力データを優先的に、ランダムの順序で出力する第1のランダム化装置と、前記第1のランダム化装置から出力された前記データの前記識別情報をトークンで置換するトークン化部と、を備える。

【発明の効果】

【0011】

30

一態様によれば、識別情報を安全にかつ高速に秘匿化する情報秘匿化装置を提供することができる。

【図面の簡単な説明】

【0012】

【図1】システム全体構成図

【図2】トークン化部の構成図

【図3】トークン化部の動作フローチャート

【図4】順序ランダム化装置のデータ入力フローチャート

【図5】順序ランダム化装置のデータ出力フローチャート

【図6】順序ランダム化装置の処理フローチャート

40

【図7】識別子発番対応表

【図8】第1の実施形態における順序ランダム化装置の初期処理フローチャート

【図9】第1の実施形態における順序ランダム化装置のプール保存フローチャート

【図10】第1の実施形態における順序ランダム化装置のプール取出しフローチャート

【図11】第1の実施形態における仮想配列イメージ図

【図12】第2の実施形態における概要を説明する図

【図13】第2の実施形態におけるプール配列を説明する図

【図14】第2の実施形態における順序ランダム化装置の初期処理フローチャート

【図15】第2の実施形態における順序ランダム化装置の保存動作フローチャート

【図16】第2の実施形態における順序ランダム化装置のプール取出し動作フローチャー

50

ト

【図17】第3の実施形態における概要を説明する図

【図18】第3の実施形態における順序ランダム化装置の初期処理フローチャート

【図19】第3の実施形態における順序ランダム化装置の動作フローチャート

【図20】第4の実施形態におけるシステム全体構成図

【図21】トークン管理装置のハードウェア構成図

【発明を実施するための形態】

【0013】

以下、図面に基づいて本発明の実施の形態を説明する。

【0014】

図1は、情報分析を行うためのシステムの全体構成の一例を説明する全体構成図である。

【0015】

図1において、分析者は、分析者端末20を操作する。分析者端末20は、データベースシステム21に接続されており、データベースシステム21から分析対象のデータを収集する。分析者端末20は、トークン化装置1を介して分析システム22に接続されており、分析対象の情報を分析システム22に送信して、分析結果を受信する。

【0016】

データベースシステム21は、データベース管理システム211と記憶部212を備えている。データベース管理システム211は、データベース管理ソフトウェアによってデータベースの管理、運用を行うシステムである。データモデルとしてはリレーションナルデータベース(RDB)やオブジェクト関係データベースなどが利用され、データベース管理ソフトウェアに応じたデータ構造のデータベースが構築される。本実施形態においては、複数の情報項目(フィールド)が1件のレコードを構成するデータ構造であるRDBについて主に説明する。

【0017】

トークン管理装置1は、トークン化装置10と復元装置11とを備える。ここで、「トークン」とは、識別情報を秘匿化した情報であり、識別情報と1対1の対応が可能な情報のことをいう。識別情報は、データベースシステム21で管理しているレコードを特定するキー情報(フィールド)であり、利用者の個人情報やIDなど、個人を特定可能な情報を含んでいる場合がある。このため、外部の分析システムで情報を分析する場合、識別情報をトークンに置換して分析システム22に送信する。また、トークン化とは、識別情報をトークンに置き換える処理のことをいう。トークン化は、例えば、IDなどの文字列を特定の文字列(トークン)に置換する処理である。トークン化装置10は、「情報秘匿化装置」として、分析システム22に送信されるデータ中の識別情報をトークン化するとともに、それぞれのトークンが一意の識別情報と対応可能なように対応を記憶して管理する。復元装置11は、分析システム22によって分析された結果に含まれるトークンを元の識別情報に復元する。トークンは、トークン化装置10によって管理されており、トークンと識別情報の一意の対応によって復元される。

【0018】

トークン管理装置1は、例えば分析者端末とLAN接続された装置によって実施がされる。トークン管理装置1は、独立の装置として構成することもできるが、例えば、トークン化装置と復元装置をそれぞれ独立させても良い。また、複数の装置によってシステムとして構成しても良い。また、ネットワーク上のWebサービスのような仮想的な装置として構成しても良い。さらには、分析者端末20上で実行されるソフトウェアとして実装しても良い。

【0019】

分析システム22は、情報の分析を行う、例えば分析者の社外の業務委託先のシステムである。分析依頼者が、自前で分析システムを構築した場合、システムの構築や維持に多大のコストが掛かってしまう。このため、分析システムの利用頻度が低く、稼働率が低い

10

20

30

40

50

ことが予想される場合には、計算機資源の有効利用の観点から、外部の分析システムを利用した方が良い。このときに、分析対象の情報をそのまま外部のシステムに預けると、情報の漏洩の危険が生じるため、分析依頼者は、データベースシステム21から収集したデータの中から、情報分析に必要最低限の情報を選択するとともに、万一情報漏洩が発生したとしても被害を少なくするために、識別情報を秘匿化したデータを分析システム22に送信する。

#### 【0020】

次に、本システムにおけるデータ分析の外部委託処理の概要を、分析データ収集と、分析依頼の二つの手順として説明する。文中の( )書きの数字は、図1のステップと対応している。

10

(分析データの収集)

#### 【0021】

データベースシステム21のデータベース管理システム211は、分析者端末20からの、SQL ( Structured Query Language ) やXQueryなどのクエリ ( 問い合わせ ) に応じて ( 1 ) 、記憶部212に記憶されたデータの検索などのデータベース操作を行う。データベース管理システム211はクエリへのレスポンス ( 応答 ) として検索結果等を分析者端末20に返す ( 2 ) 。この一連の動作によって、分析者端末20は、分析対象となるデータの収集を行う。

(分析依頼)

#### 【0022】

分析者端末20は、データベースシステム21から収集したデータを基に、分析システム22に送信する分析対象のデータをトークン管理装置1に送信する ( 3 ) 。分析者端末20は、収集したデータの中で、分析に不要なレコードやフィールドを削除したり、年齢情報を年代情報 ( 20代、30代等 ) に変更したりする情報の選定をしても良い。この情報の選定は、トークン化装置10で行っても良い。

20

#### 【0023】

トークン管理装置1は、送信された分析対象データに含まれる識別情報をトークン化して秘匿化し、分析システム22に送信する ( 4 ) 。分析システム22は、送信されたデータに含まれる情報を分析して、分析結果をトークン管理装置1に送信する ( 5 ) 。トークン管理装置1は、分析システム22から送信された分析結果の中で、トークンの部分を元の識別情報に復元して、分析結果を分析者端末20に送信する ( 6 ) 。以上の手順にて情報分析の外部委託が実施される。

30

#### 【0024】

次に、図2を用いて、トークン管理装置1のトークン化装置10の構成を説明する。図2は、トークン化装置の構成の一例を説明する構成図である。

#### 【0025】

図2において、トークン化装置10は、順序ランダム化装置101 ( 入力データのシャッフリング ) 、トークン管理部102、識別子発番対応表103、順序ランダム化部104 ( トークンのランダム化 ) 、及びトークン発生器105を備えている。

40

#### 【0026】

順序ランダム化装置101は、入力データのシャッフリングを行う。この実施例では、入力データの中で秘匿したい情報は、識別情報としての「ID」、及び個人情報としての「年齢」であるものとする。IDと年齢は対応付けられて、一つのレコード ( データセット ) となっている。

#### 【0027】

順序ランダム化装置101は、格納部1011と、乱数発生部1012と、を備える。格納部1011は、内部の記憶領域である「プール」を有し、入力データをプールに格納する。格納部1011は、所定の格納数 ( レコード数 ) の入力データをプールの格納場所に格納する。プールは、レコードを記憶する仮想の配列 ( テーブル ) である。

#### 【0028】

50

乱数発生部 1012 は、プールの格納数に応じた乱数を発生させる。例えば、格納数が  $n$  個の場合に対応して、乱数発生部 1012 は、 $0 \sim (n - 1)$  の  $n$  個の範囲で乱数を発生させる。

#### 【0029】

順序ランダム化装置 101 は、乱数発生部 1012 で発生させた乱数により、格納部 1011 のプールに逐次格納されたレコードを任意の順番（ランダム）に取り出す。これにより、データベースシステム 21 の仕様に依存しているデータの出力順序の規則性を破壊することができるため、データベースシステム 21 でのデータ出力時のランダム化処理が不要となり、データベースシステム 21 から高速にデータを出力させることができる。図 2 では、ID が 950001、950002、950003、…と ID の昇順でデータが入力されているが、入力データ順序ランダム化部 101 から出力されるデータは、ID が、950003、950073、950029 と、プールされたデータの中からランダムな順序で出力される。

#### 【0030】

トークン管理部 102 は、トークン化部 1021 と、トークン登録部 1022 を備える。トークン登録部 1022 は、入力されたデータの ID について、その ID が既に登録されているかを識別子発番対応表 103 に問い合わせる。識別子発番対応表 103 には、ID と既に発番されているトークンとが対応づけられて登録されている。識別子発番対応表 103 は、問い合わせされた ID が既に登録済みである場合は、応答としてその ID のトークン値をトークン登録部 1022 に返却する（1）。一方、問い合わせされた ID の登録が無い場合には、ID 未登録の応答を返却する。トークン登録部 1022 は、ID 未登録の応答を受けると、順序ランダム化装置 104 に対してトークンの発番依頼をする（2）。

#### 【0031】

順序ランダム化装置 104 は、格納部 1041 と、乱数発生部 1042 と、を備える。格納部 1041 及び乱数発生部 1042 は、順序ランダム化装置 101 の格納部 1011 及び乱数発生部 1012 と同様の構成であり、説明は省略する。

#### 【0032】

順序ランダム化装置 104 は、トークン管理部 102 から発番依頼を受けると、トークン発生器 105 にて発番されたトークンを、順序ランダム化装置 101 と同様に、格納部 1041 のプールに入れて、乱数発生部 1042 によって発生された乱数により、ランダムな順序でトークンを取り出すシャッフリングを行う。

#### 【0033】

トークン発生器 105 は、例えばトークンを、00001、00002、00003、のようなシリアル番号として発番する。トークン発生器 105 は、シリアル番号を発生することにより、発番されたトークンのユニーク（重複）チェックを行う必要がなくなり、処理が高速化される。また、トークン発生器 105 は、発番されたトークンの番号を記録しておく必要がないため、トークン発生器 105 の処理負荷が軽減できる。なお、トークン発生器 105 で発番されるトークンに、例えばシリアル番号に特定の関数を適用させて、それをトークンとして使用しても良い。

#### 【0034】

発番されたトークンは、順序ランダム化装置 104 によって、00395、00153、00216 のようにランダムな順序にされて、トークン管理部 102 に送信される（3）。順序ランダム化装置 104 により発番されたトークンがランダム化されることにより、トークン発生器 105 で発生させるトークンを発番するときに乱数を発生させる必要がなくなるため、発番の高速化が可能となる。

#### 【0035】

トークン管理部 102 のトークン化部 1021 は、ID をトークンに置換する。また、トークン登録部 1022 は、送信されたトークン値を、ID と対応付けて、識別子発番対応表 103 に登録する（4）。トークン管理部 102 は、トークンと年齢をデータセット

10

20

30

40

50

にして出力データとして出力する。

【0036】

次に、識別子発番対応表103の詳細を、図7を用いて説明する。図7は、識別子発番対応表の一例である。

【0037】

図7において、識別子発番対応表103は、識別子(ID)とトークン値とを対応づけて保存している。ここで識別子(ID)とトークン値とは、一意に対応づけられている。従って、トークン値から識別子に復元することが可能になる。なお、識別子発番対応表103は、トークン値から識別子への復元をすることが目的であるため、例えば、復元の必要がなくなった時点でトークン発生器105と識別子発番対応表103をクリアすることにより、クリア前のトークンとクリア後のトークンが同じ番号で登録されても構わない。

10

【0038】

次に、図2で説明したトークン化装置10の動作を、図3のフローチャートを用いて説明する。図3は、トークン化装置10の動作の一例を説明するフローチャートである。

【0039】

図3において、トークン化装置10には、分析者端末20から、秘匿化する情報を含むデータが入力される(S100)。順序ランダム化装置101は、情報をプールに入力する(S101)。トークン管理部102は、順序ランダム化装置101から、レコード単位でデータを取り出す(S102)。取り出されたレコードは、順序ランダム化装置101によって、取り出される順序がランダム化(シャッフリング)されている。なお、ステップS101及びS102のランダム化装置101の動作詳細については、ランダム化装置104の動作詳細とともに後述する実施形態1～実施形態3にて説明する。

20

【0040】

次に、格納部1011のプールにデータがあるか否かをチェックして(S103)、もしデータが全て処理されていれば(S103でNO)、このフローチャートでの処理を終了し、データがプールに残っている場合には(S103でYES)、ID部分が識別子発番対応表103に既に存在しているか否かをチェックする(S104)。

【0041】

ID部分が既存の場合(S104でYES)、識別子発番対応表103は、問い合わせされたIDに対してトークン値を返却し、トークン化部1021は、トークン値でID部分を置き換えて(S105)、レコード単位の情報が出力される(S106)。

30

【0042】

一方、ID部分が識別子発番対応表103に無い場合(S104でNO)、トークン登録部1022は、順序ランダム化装置104にトークン値を要求する(S107)。トークン発生器105は予めトークンを発生させておき、順序ランダム化装置に入力してランダム化されたトークンが準備されている(S108)、トークン登録部1022は、IDとトークンを対応づけて識別子発番対応表103に登録する(S109)。そして、トークン化部1021は、トークン値でID部分を置き換えて(S105)、レコード単位で情報が出力する(S106)。以上の処理は、順序ランダム化装置101の格納部1011のデータが無くなるまでループ処理される(S103)。

40

【0043】

次に、図3のステップS101の情報入力処理の詳細を、図4を用いて説明する。図4は、順序ランダム化装置101及び順序ランダム化装置104のデータ入力の一例を説明するフローチャートである。なお、以下の説明で、順序ランダム化装置101及び順序ランダム化装置104に共通するものは、「順序ランダム化装置」のように符号の記載を省略している。

【0044】

図4において、順序ランダム化装置は、情報の入力があると(S1011)、プールに空きがあるか否かをチェックする(S1012)。プールに空きが無い場合には(S1012でNO)、空きが出るまで情報の保存が待ち状態となる。これにより、プールは入力

50

される情報が終了するまで、確保されたプールを常に全て使用することとなる。

【0045】

プールに空きがあった場合 (S1012でYES)、情報をプールに保存する (S1013)。

【0046】

次に、図3のステップS102の情報出力処理の詳細を、図5を用いて説明する。図5は、順序ランダム化装置のデータ出力の一例を説明するフローチャートである。

【0047】

図5において、図2のトークン管理部102は、順序ランダム化装置に対して情報要求を行う (S1021)。次に、順序ランダム化装置は、プールからランダムに一つの情報を取り出して (S1022)、トークン管理部102に情報を出力する (S1023)。

【0048】

次に、図3で説明した、ステップ108の処理の詳細を、図6を用いて説明する。図6は、順序ランダム化装置104の処理の一例を説明するフローチャートである。

【0049】

図6において、順序ランダム化装置104は、トークン値の発生をトークン発生器105に要求して、トークン値を発生させる (S1081)。トークン発生器105はトークン値を発生させて、発生させたトークン値を順序ランダム化装置104に入力する (S1082)。順序ランダム化装置104のプールに空きがある場合は (S1083でYES)、順序ランダム化装置104は、続けてトークン発生器105にトークン値を発生させて、プールに空きが無くなるまで (S1083でNO) 処理を繰り返す。つまり、順序ランダム化装置104は、トークン管理部102からの発番依頼がされる前にプールにトークン値を事前に準備しておく。これにより、発番依頼があつてからトークン値を発番する場合に比べて、トークン値の発番からプールへの保存までのオーバーヘッドが無くなり、処理の高速化が可能になる。

【0050】

次に、図3のステップS101とステップS102で説明したプールの管理アルゴリズムについて、以下、実施形態1～実施形態3にて詳細に説明する。

【0051】

順序ランダム化装置におけるプールの大きさや管理方法は、ランダム化の性能に大きく影響する。実施形態1～実施形態3では、それぞれのプール管理アルゴリズムについてデータの入出力処理の方法が相違している。なお、ランダム化装置101とランダム化装置104は、以下のいずれの実施形態で実施しても良く、例えば、ランダム化装置101とランダム化装置104とで、同じ実施形態で実施しても良いし、それぞれ異なった実施形態で実施しても良い。

[第1の実施形態]

【0052】

第1の実施形態は、格納数nのプールを用意して、入力データを格納数分入力して、格納数までを発番可能な乱数を用いることにより、格納場所からランダムの順序にデータを出力する。データを出力した格納場所は空きとなる。空きとなった格納場所は、例えば次に入力されるデータによって上書きされながら補充される。また、空きとなった格納場所を後ろのデータにて順次前詰めしても良い。乱数の発生は格納個数の範囲で行い、データベースシステムにおけるランダム化処理はしないため、処理が高速化できる。

【0053】

図8～図11を用いて、第1の実施形態を説明する。図8は、第1の実施形態における順序ランダム化装置の初期処理の一例を説明するフローチャートである。図9は、第1の実施形態における順序ランダム化装置のプール保存の一例を説明するフローチャートである。図10は、第1の実施形態における順序ランダム化装置のプールからの取出しの一例を説明するフローチャートである。さらに、図11は、第1の実施形態における仮想配列の一例を説明するイメージ図である。

10

20

30

40

50

## 【0054】

図8において、順序ランダム化装置の初期処理は、図4におけるステップS1011における初期処理である。初期処理は、データベースシステム21から新たな入力データが入力されたときに1回行われる。順序ランダム化装置は、データサイズに応じたプール(仮想配列)を確保する。本実施例では、 $0 \sim (n - 1)$ のn個のプールを確保している(S10111)。次に、確保したプールの「格納個数」を0にセットして(S10112)、初期化処理を終了する。格納個数は、実際に情報が入力されたプールの数を計数するカウンタであり、この値を格納個数分、フルに保つように入力データを逐次入力する。

## 【0055】

図9において、順序ランダム化装置の入力データのプール保存は、図4におけるステップ1013の処理の詳細を説明したものである。順序ランダム化装置は、プールの空いている格納場所に入力データを格納する(S10131)。次に、格納個数に1を追加してカウンタをインクリメントして処理を終了する(S10132)。

## 【0056】

図10において、順序ランダム化装置のプール取出しは、図5におけるステップS1022の詳細を説明したものである。図2のランダム発生器105は、 $0 \sim (\text{格納個数} - 1)$ の間の整数rを発生する乱数にて、取り出し対象の格納場所を特定する仮想配列の番号である整数rをランダムに生成する(S10221)。次に、プールのr番目に格納されているデータを取り出して、取り出された格納場所は番号が詰められて(S10222)、仮想配列の番号に抜けが無いようにする。次に格納個数に-1を追加して、カウンタをデクリメントして終了する(S10223)。

## 【0057】

図11は、プールで使用される格納場所の仮想配列を説明するイメージ図である。プールには、仮想配列の番号 $0 \sim n$ に対応してデータが格納されている。仮想配列は同一のデータ型がメモリ上に連続して接しているとみなされる仮想的な配列であり、プログラミング上の特定のオペレータによって配列の長さを指定できる。図11では、nが指定されており、 $n + 1$ 個のデータ配列の場合を表している。配列の長さは生成される乱数と対応している。

## 【0058】

以上説明した第1の実施形態においては、格納場所に空きができると次の入力データがすぐに格納されるため、入力されるデータはプールに載せられるだけ載せられることになる。入力データのランダム化は、プールの格納個数が大きい(仮想配列が長い)程入力データの順番をランダムに並び替えるシャッフル効果が高い。しかし、大きなプールを使うには大きなメモリリソースが必要になる。このため、初期化で確保するプールの格納数は、入力されるデータサイズや情報秘匿の必要性によって適宜調整される。

## [第2の実施形態]

## 【0059】

第1の実施形態においては、一度プールに格納されたデータは、同確率でプールから取り出されるため、先に格納されたデータほど先に取り出される確立が高くなる。つまり、取り出されるデータの順番が入力順序に依存することになる。そこで、第2の実施形態においては、プールの構造によって、第1の実施形態の課題を改善している。

## 【0060】

第2の実施形態の概要を図12及び図13を用いて説明する。図12は、第2の実施形態におけるプール管理の概要の一例を説明する図である。図13は、2の実施形態におけるプール配列の一例を説明する図である。

## 【0061】

図12において、順序ランダム化装置は、格納部にプール管理情報を備えている。プール管理情報には、プール配列の要素数(L)、乱数終了値、情報追加する要素番号(0~L-1)、情報追加するカウンタ、1要素に情報追加する数、及びプールに格納する数(L以下)が記録される。要素数L個に設定されたプール配列の1要素の詳細は、重み、乱

10

20

30

40

50

数アクセス用開始ポイント、乱数アクセス用終了ポイント、レコードを格納する仮想配列である。

【0062】

プール管理情報の初期設定において、プール配列の要素数（L）、1配列に情報追加する数、プールに格納する数、各要素の重みが設定される、他の項目の初期値は0又はNULLである。

【0063】

図13において、この実施例ではプール配列Lは4である。すなわち、0～4の5つの要素を備え総データ数よりここでは、上記「プールに格納する数」を4として、最後の1要素は空きとしている。例えば、配列番号0は、重み：1、格納個数：48、乱数アクセス用開始ポイント：0、…、データを格納する仮想配列（48要素）と設定される。また、配列番号1～3も図13に図示すように設定されている。

【0064】

ここで、「重み」とは、各配列間において格納個数に対して、1レコードに割り当てられる乱数の数である。例えば、配列番号0では重み1であり、一つのレコードに一つの乱数が割り当てられ、一方、配列番号3では重み8であり、一つのレコードに8の乱数が割り当てられる。したがって、配列番号0では $48 \times 1 = 48$ 個の乱数が割り当てられて、配列番号1では $32 \times 2 = 64$ 個、配列2では $21 \times 4 = 84$ 個、配列3では $8 \times 9 = 72$ 個の乱数が割り当てられる。つまり、重みが低い場合には格納するレコードの個数を増やして1レコードに割り当てられている乱数の幅を狭くする一方、重みが高い場合には、逆に格納するレコードの個数を減らして1レコードに割り当てられている乱数の幅を広くしている。

【0065】

本実施形態においては、配列番号が後ろの要素に対して、配列番号が前の要素より重みを高く設定している。このプールの管理方法によって、格納された情報が取り出される確率を調整して、シャッフル効果を高くすることができる。

【0066】

この実施例では、上記の通り、発生させる乱数の個数は、 $48 + 64 + 84 + 72 = 268$ 個であるため、乱数の幅は0～267である。例えば、乱数が153であったとすると、プールから取り出されるレコードは、配列番号2の仮想配列10となる。レコードを一つ取り出した場合、取り出したレコードの後ろのレコードは順次前詰にされていく。したがって、配列番号3の要素数は9～8へと減算されることになり、次のレコードが追加されると再び要素数が9に戻ることになる。

【0067】

次に、第2の実施形態における動作を図14～図16のフローチャートを用いて説明する。図14は、第2の実施形態における順序ランダム化装置の初期処理の一例を説明するフローチャートである。図15は、第2の実施形態における順序ランダム化装置の保存動作の一例を説明するフローチャートである。図16は、第2の実施形態における順序ランダム化装置部のプール取り出し動作の一例を説明するフローチャートである。

【0068】

図14において、先ず、プール（仮想配列）に格納する数を仮決めする（S211）。仮決めされた格納数を基に、プールの分割数Lを決める（S212）。この分割数が図12で説明した「プール配列の要素数（L）」になる。

【0069】

次に、プール（仮想配列）に格納する数を決める（S213）。この数は、分割数Lを決めた上で、実際に使用する数を決めるものであり、L以下で任意に設定可能である。

【0070】

次に、図12で説明したプール管理情報の初期化を行う（S214）。具体的には、乱数終了値、情報追加する要素番号、及び情報追加するカウンタを0にセットする。

【0071】

10

20

30

40

50

次に、プール配列の確保と初期化を行う(S 2 1 5)。具体的には、「乱数アクセス用開始ポイント」、及び「乱数アクセス用終了ポイント」に0をセットする。また、「データを格納する仮想配列」は、1要素に情報追加する数の仮想配列を用意する。さらに、「重み」はプール配列の順番が大きいほど大きい値にする。図13においては1、2、4、8、16・・・を例示している。以上で、初期処理を終了する。

【0072】

図15において、順序ランダム化装置の保存動作は、先ず、情報入力がされると(S 2 0 0)、「情報追加する要素番号」から、データを格納するプール配列の要素を求める(S 2 0 1)。

【0073】

次に、その要素の「要素のデータを格納する仮想配列」の「情報追加するカウンタ」番目にデータを格納して、次の操作を行う(S 2 0 2)。「情報追加するカウンタ」をインクリメントする(++)。「乱数アクセス用終了ポイント」に重みを加算代入する(+=重み)。「乱数終了値」に重みを加算代入する(+=重み)。

【0074】

次に、「情報追加するカウンタ」が「1要素に情報追加する数」になったか否かを判断し(S 2 0 3)、NOである場合は本フローチャートの動作を終了する。

【0075】

ステップS 2 0 3でYESの場合、情報を追加する要素がL番目のプール配列の要素であるか否かを判断し(S 2 0 4)、NOである場合には、「情報追加する要素番号」をインクリメントし、追加するカウンタを0、次のプール配列の要素の「乱数アクセス用開始ポイント」と、「乱数アクセス終了ポイント」の値を求めセットする(S 2 0 5)。一方、ステップS 2 0 4でYESの場合には、プール配列の1番目を0番目にマージして1番目を空にし、1番目が空いた分、1つずつ前詰めして移行させ、L番目が空の状態になるようする。

【0076】

図16において、順序ランダム化装置部のプール取り出し動作は、先ず、0～乱数終了値の間の整数rをランダムに生成する(S 2 2 1)。整数rは、例えば使用するプログラミング言語の最大値を指定した整数乱数生成関数によって得ることができる。整数rを決定する乱数は、取り出しのタイミングで逐次生成する。しかし、例えば取り出しのタイミングとは非同期に乱数を用意しておき、取り出しのタイミングで整数rに適用しても良い。特に乱数の計算に時間がかかる場合には、事前準備によって処理を高速化することができる。

【0077】

次に、プール配列の「乱数アクセス用開始ポイント」と「乱数アクセス用終了ポイント」から、整数rに対応するプール配列の要素を求める(S 2 2 2)。図12及び図13で説明したとおり、整数rによって対応する配列が一意に決まることになる。

【0078】

次に、プール配列の要素からrに該当するレコードのデータを格納する仮想配列の格納場所を求め取り出し、取り出した格納場所は後ろのレコードによって前詰めされる(S 2 2 3)。前詰めのタイミングは、レコードを取り出して直ぐに行う。但し、例えば、所定数のレコードを読み出してから一度に所定数分前詰めしても良い。その場合、読み出しに使用される整数rについて、空のレコードが指定された場合の整数rの再指定について考慮しておく。これにより、保存と読み出しの処理を非同期に行うことも可能となる。

【0079】

次に、プール配列の要素以降の「乱数アクセス用開始ポイント」、「乱数アクセス用終了ポイント」、及び「乱数終了値」のメンテナンスを行う(S 2 2 4)。レコードが前詰めされた場合、上記各設定値も前詰めに合わせて変更される。しかし、ステップS 2 2 3で説明したとおり、前詰めのタイミングによっては上記メンテナンスのタイミングも変わってくる。以上で順序ランダム化装置部のプール取り出し動作を終了する。

[第3の実施形態]

10

20

30

40

50

## 【0080】

第1の実施形態、および第2の実施形態においては、プール配列によって入力データのシャッフルを行ったが、例えば、データベースのレコード数が多く、レコードの登録日や更新日順に入力されるデータに対しては、実施形態1又は2における局所的な並び替えでは全体的なシャッフルが不十分となる場合がある。そこで、第3の実施形態においては、データ全体をランダム化することにより、より高いシャッフル効果を得ることができる。

## 【0081】

第3の実施形態を、図17～図19によって説明する。図17は、第3の実施形態における概要を説明する図である。図18は、第3の実施形態における順序ランダム化装置の初期処理の一例を説明するフローチャートである。図19は、第3の実施形態における順序ランダム化装置の動作の一例を説明するフローチャートである。

10

## 【0082】

図17において、順序ランダム化装置のプールは、n個のテーブルを有する2次記憶装置に接続されている。プールに入力された全入力データは、n個のテーブルにランダムに全て分配されて記憶される。それぞれのテーブルに記憶されたデータは、各テーブル毎にランダムに指定されて読み出されて出力データとなる。

## 【0083】

図18において、順序ランダム化装置の初期処理は、先ず、プールの内部を複数に分割して、テーブル用キャッシュを作れる数をnとする(S310)。プールは2次記憶とのデータのキャッシュで使用するだけなので、容量についてはキャッシュとして十分であれば良い。それぞれのキャッシュは2次記憶の入出力キャッシュとして使用される。

20

## 【0084】

次に、n個の集合を管理するテーブルを2次記憶上に作成する(S311)。

## 【0085】

次に、入力データを1つずつ読み込みながら、1～nの乱数を生成して、生成された乱数に対応するプールに作成されたテーブル用キャッシュに格納し、一杯になったら2次記憶上のテーブルに退避させる(S312)。

## 【0086】

次に、各テーブル毎に、ランダムに最初の読み出し位置と読み出し順(前から後、後ろから前)を決めセットする(S313)。読み出しテーブルを1とする。

30

## 【0087】

図19において、順序ランダム化装置のプールからの取出しは、先ず、読み出しテーブルがないか否かを判断する(S300)。読み出しテーブルが無い場合には、このフローチャートの動作を終了する(S300でYES)。読み出しテーブルがある場合(S300でNO)、読み出しテーブルからデータを取出す(S301)。

## 【0088】

次に、読み出しテーブルのスキャンが一巡したか否かを判断する(S302)。スキャンが一巡した場合(S302でYES)、読み出しテーブルから除外する(S303)。

## 【0089】

次に、読み出しテーブルがないか否かを判断する(S304)。読み出しテーブルがある場合(S304でNO)、又は読み出しテーブルのスキャンが一巡していない場合(S302でNO)、読み出しテーブルを次に勧める(S305)。

40

## 【0090】

全てのテーブルの読み出しが終了した場合(S304でYES)、順序ランダム化装置のプールからの取出し処理を終了する。

## [第4の実施形態]

## 【0091】

第1の実施形態～第3の実施形態は、トークン管理装置1の順序ランダム化装置についての実施形態であったが、同様の順序ランダム化装置は、データベースシステムの中に構

50

築することもできる。順序ランダム化装置をデータベースシステムの中に構築することで、検索結果の規則性を無くすことができ、出力データの順番によるデータの特性がデータベースシステム外部に漏れることを防ぐことができる。

#### 【0092】

第4の実施形態を、図20を用いて説明する。図20は、第4の実施形態におけるシステム全体構成の一例を説明する構成図である。

#### 【0093】

図20において、データベースシステム31は、データベース管理システム3111と記憶部312を備えている。データベース管理システム3111は、コマンド処理部3111、および順序ランダム化装置3112を備えている。なお、図20において、データベースシステム31以外の他の装置は、既に説明した第1の実施形態～第3の実施形態と同じものについては同じ符号を付して説明を省略する。コマンド処理部3111は、順序ランダム化装置3112を介して記憶部312と接続されており、記憶部312からデータを読み出す際に順序ランダム化装置3112にてデータがランダム化される。ランダム化装置内部のプールの管理方法は第1の実施例から第3の実施例と同様である。

10

#### 【0094】

第4の実施形態においては、データベースの出力段階にプールを入れることにより、データベースの内部構造から分離した高速な出力が可能になる。これによりデータベースの内部構造がデータベースシステム31の外部からは見えなくなる。また、順序ランダム化装置3112にアクセス制限を設け、順序ランダム化装置3112の存在を意識することなくデータの出力が可能になり、データベースシステムのデータ出力特性を秘匿化することが可能になる。

20

#### 【0095】

また、第4の実施形態は、第1の実施形態から第3の実施形態との併用が可能である。両者の併用によって、シャッフルリングの効果を高めることができ、順序ランダム化装置3112にアクセス可能な場合は、それぞれの順序ランダム化装置にて、ランダム化の負荷を分散させることができる。

#### 【0096】

図21は、トークン管理装置1のコンピュータシステムの一例を示すブロック図である。図21に示すトークン管理装置1は、CPU12、記憶部13、インターフェース(I/F)14、入力装置15、及び表示部16がバス17により接続された構成を有する。尚、CPU12と、記憶部13、I/F14、入力装置15、及び表示部16との接続は、図21に示すバス接続に限定されるものではない。

30

#### 【0097】

図21において、CPU12は、記憶部13に格納されたプログラムを実行することによりトークン管理装置1全体を制御する。記憶部13は、半導体記憶装置、磁気記録媒体、光記録媒体、光磁気記録媒体等のコンピュータ読み取り可能な記憶媒体で形成可能であり、上記のプログラムや各種データを格納すると共に、CPU12が実行する演算の中間結果や演算結果等を一時的に格納する一時メモリとしても機能する。記憶部13は、図2に示す識別子発番対応表103の記憶部としても機能する。I/F14は、分析者端末20や分析システム20とネットワーク(図示せず)を介しての通信を行うインターフェースである。入力装置15は、キーボード等により形成可能である。表示部16は、ディスプレイ等により形成可能である。入力装置15及び表示部16は、タッチパネルのように入力装置と表示部の両方の機能を有する入出力装置で形成しても良い。

40

#### 【0098】

CPU12は、記憶部13に格納されたプログラムを実行することにより、コンピュータシステムをトークン管理装置1として機能させる。つまり、図2で説明したトークン化装置10の順序ランダム化装置101、104、トークン管理部102、識別子発番対応表103は、プログラムの機能として記憶部13の記憶領域に実装されて、CPU12により実行されるようにすることができる。

50

## 【0099】

以上、本発明を実施するための形態について詳述したが、本発明は斯かる特定の実施形態に限定されるものではなく、特許請求の範囲に記載された本発明の要旨の範囲内において、種々の変形・変更が可能である。

## 【0100】

本発明は、以下に記載する付記のような構成が考えられる。

## (付記1)

識別情報を含むデータが入力されて、入力された前記データをランダムの順序で出力する第1のランダム化装置と、

前記第1のランダム化装置から出力された前記データの前記識別情報をトークンで置換するトークン化部と、を備えた情報秘匿化装置。 10

## (付記2)

前記第1のランダム化装置は、

所定の格納数の格納場所を有して、入力された前記データを前記格納場所に逐次格納する第1の格納部と、

前記第1の格納部の格納数に応じた第1の乱数を生成させる第1の乱数生成部と、を備え、

前記第1の乱数によってランダムに特定される前記第1の格納部の格納場所から格納された前記データを取り出して出力する付記2に記載の情報秘匿化装置。 20

## (付記3)

前記第1の格納部の格納場所は、入力順序が早い入力データに対して入力順序が遅い入力データを優先的に取り出す付記1又は2に記載の情報秘匿化装置。

## (付記4)

前記トークンを発生させるトークン発生器と、

前記トークン発生器で発生された前記トークンが入力されて、入力された前記トークンをランダムの順序で出力する第2のランダム化装置と、をさらに備えた付記1乃至3のいずれかに記載の情報秘匿化装置。 25

## (付記5)

前記第2のランダム化装置は、

所定の格納数の格納場所を有して、入力された前記トークンを前記格納場所に逐次格納する第2の格納部と、

前記第2の格納部の格納数に応じた第2の乱数を生成させる第2の乱数生成部と、を備え、

前記第2の乱数によってランダムに特定される前記第2の格納部の格納場所から格納された前記トークンを取り出して出力する付記4に記載の情報秘匿化装置。 30

## (付記6)

前記第2の格納部の格納場所は、入力順序が早い入力データに対して入力順序が遅い入力データを優先的に取り出す付記4又は5に記載の情報秘匿化装置。

## (付記7)

識別情報を含むデータが入力されて、入力された前記データをランダムの順序で出力する第1のランダム化処理と、

前記第1のランダム化処理で出力された前記データの前記識別情報をトークンで置換するトークン化処理と、をコンピュータが実行する識別情報の情報秘匿化方法。 40

## (付記8)

前記第1のランダム化処理は、

所定の格納数の格納場所を有して、入力された前記データを前記格納場所に逐次格納する第1の格納処理と、

前記第1の格納処理の格納数に応じた第1の乱数を生成させる第1の乱数生成処理と、を備え、

前記第1の乱数によってランダムに特定される前記第1の格納処理で格納される格納場 50

所から格納された前記データを取り出して出力する処理をコンピュータが実行する付記7に記載の識別情報の情報秘匿化方法。

(付記9)

前記第1の格納処理は、入力順序が早い入力データに対して入力順序が遅い入力データを優先的に取り出す付記7又は8に記載の情報秘匿化方法。

(付記10)

前記トークンを発生させるトークン発生処理と、

前記トークン発生処理で発生された前記トークンが入力されて、入力された前記トークンをランダムの順序で出力する第2のランダム化処理と、をさらに備えた付記7乃至9のいずれか一に記載の情報秘匿化方法。

10

(付記11)

前記第2のランダム化処理は、

所定の格納数の格納場所を有して、入力された前記トークンを前記格納場所に逐次格納する第2の格納処理と、

前記第2の格納処理の格納数に応じた第2の乱数を生成させる第2の乱数生成処理と、を備え、

前記第2の乱数によってランダムに特定される前記第2の格納処理で格納される格納場所から前記トークンを取り出して出力する付記10に記載の情報秘匿化方法。

(付記12)

前記第2の格納処理は、入力順序が早い入力データに対して入力順序が遅い入力データを優先的に取り出す付記4又は5に記載の情報秘匿化方法。

20

(付記13)

クエリを受け付けるコマンド処理部と、

前記コマンド処理部からの指示によって記憶しているデータを出力する記憶部と、

ランダム化装置と、を備えたデータベースシステムであって、

ランダム化装置は、

所定の格納数の格納場所を有して、前記記憶部から出力されたデータを前記格納場所に逐次格納する格納部と、

前記格納部の格納数に応じた乱数を生成させる乱数生成部と、を備えたデータベースシステム。

30

【符号の説明】

【0101】

1 トークン管理装置

10 トークン化装置

101、104 順序ランダム化装置

1011、1041 格納部

1012, 1042 乱数発生部

102 トークン管理部

1021 トークン化部

1022 トークン登録部

40

103 識別子発番対応表

105 トークン発生器

11 復元装置

12 C P U

13 記憶部

14 インタフェース

15 入力装置

16 表示部

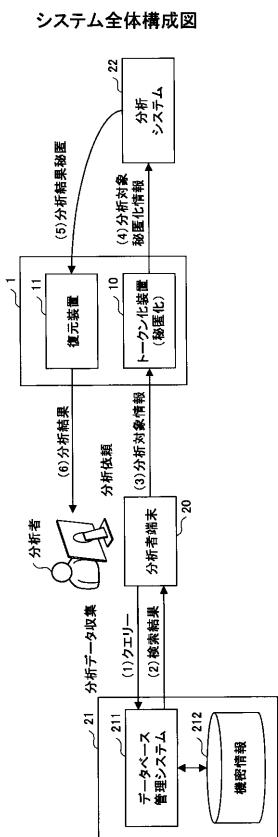
17 バス

20 分析者端末

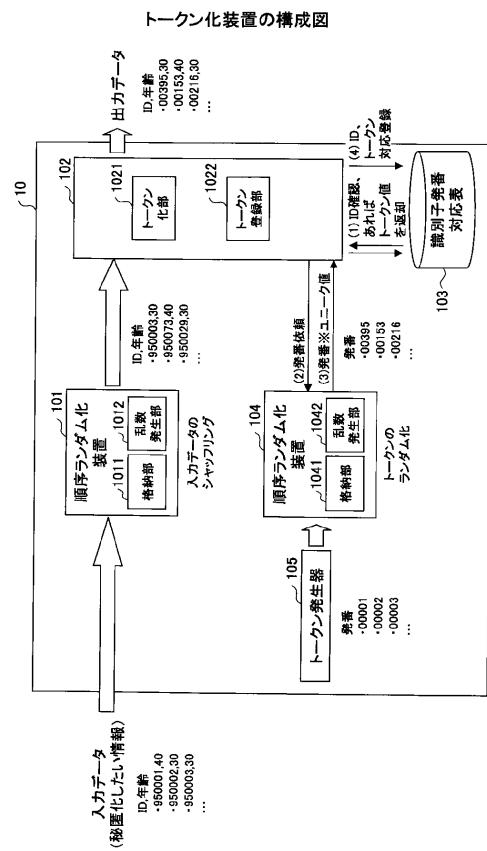
50

2 1、3 1 データベースシステム  
 2 1 1、3 1 1 データベース管理システム  
 2 1 2、3 1 2 記憶部  
 2 2 分析システム  
 3 1 1 1 コマンド処理部  
 3 1 1 2 順序ランダム化装置

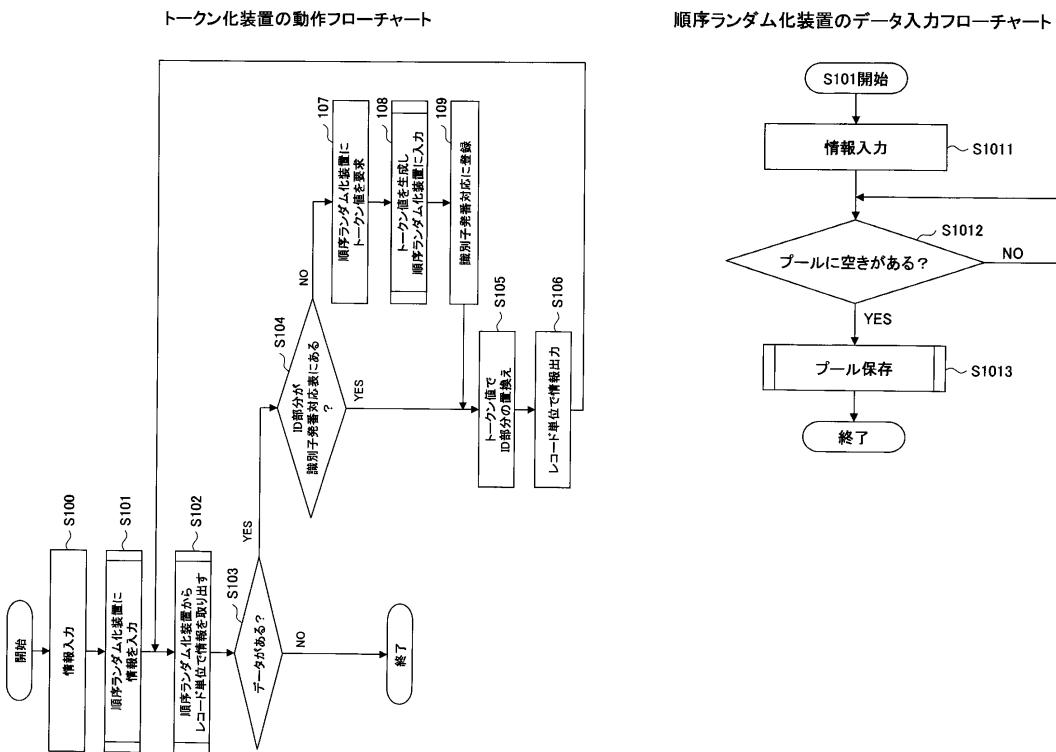
【図1】



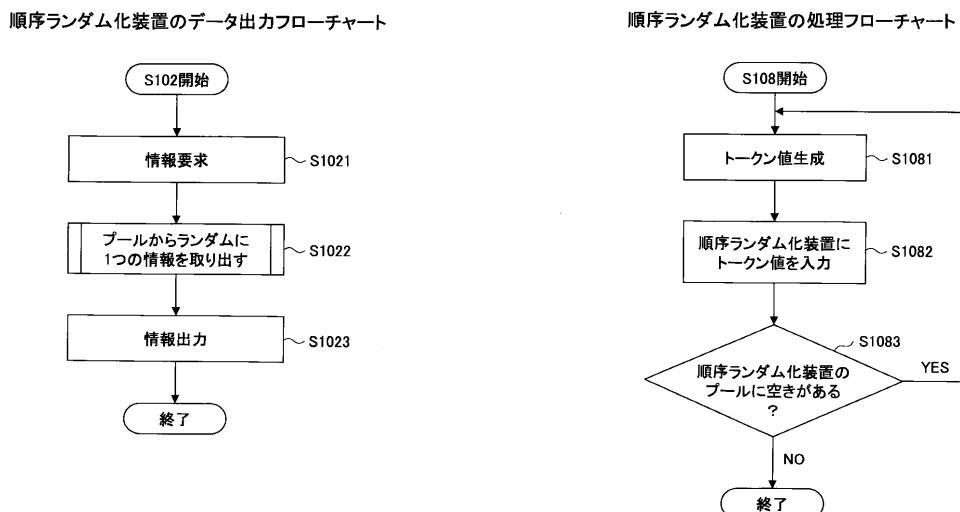
【図2】



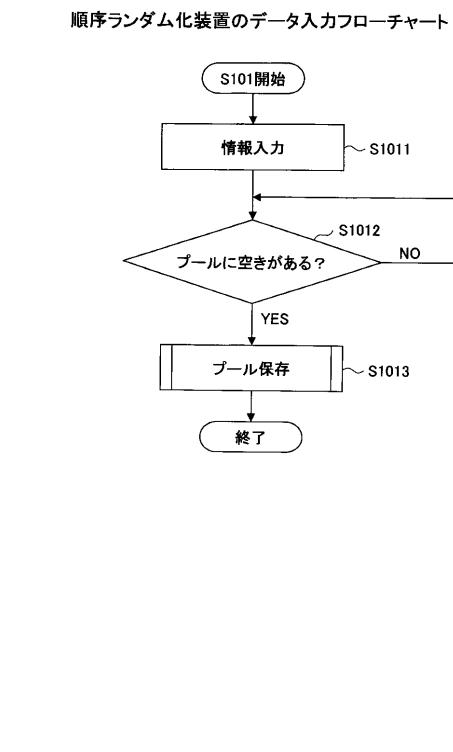
【図3】



【図5】



【図4】



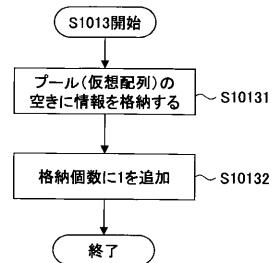
【図6】

【図7】

識別子発番対応表

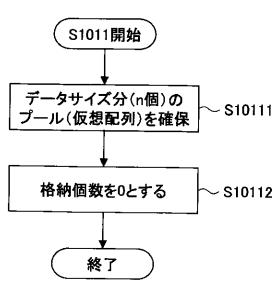
識別子(ID)	トークン値
20120001	39275831
20120002	87299017
20120003	18644819
20120004	69596127
20120005	91724012
20120006	58363108
...	...

第1の実施形態における順序ランダム化装置のプール保存フローチャート



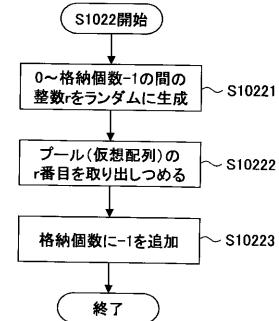
【図8】

第1の実施形態における順序ランダム化装置の初期処理フローチャート



【図10】

第1の実施形態における順序ランダム化装置のプール取出しフローチャート

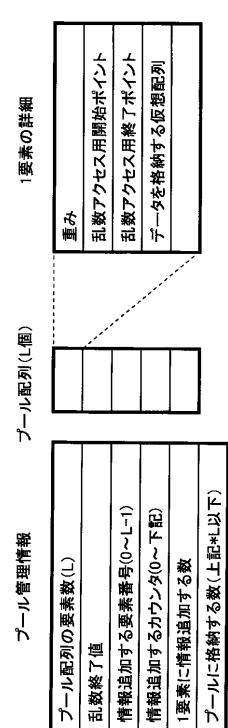


【図11】

第1の実施形態における仮想配列イメージ図

0	20120001
1	20120002
2	20120003
3	20120004
4	20120005
...	...
n	2012000n

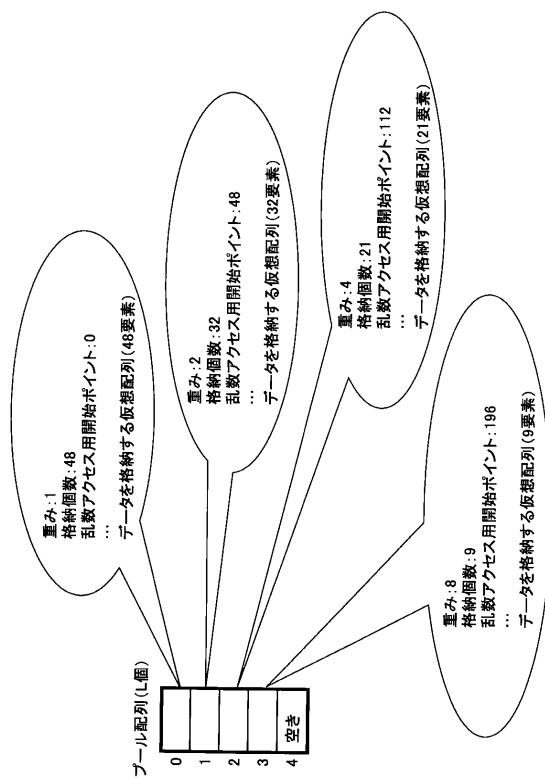
第2の実施形態における概要を説明する図



【図12】

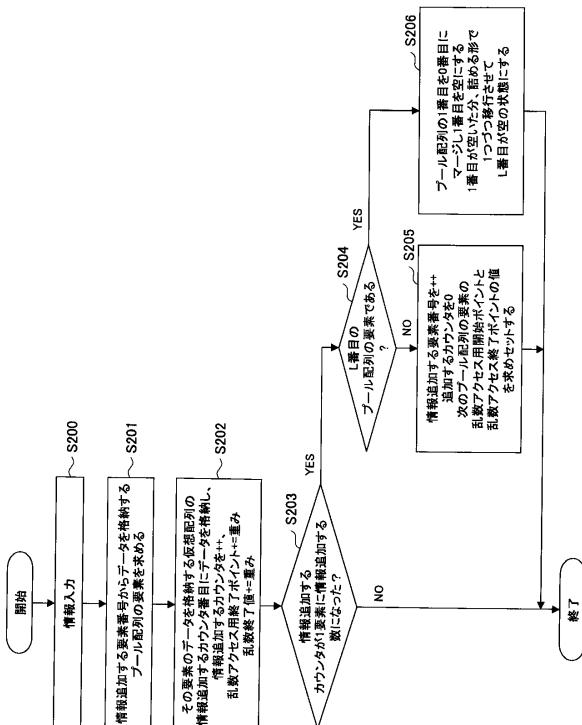
【図13】

第2の実施形態におけるプール配列を説明する図



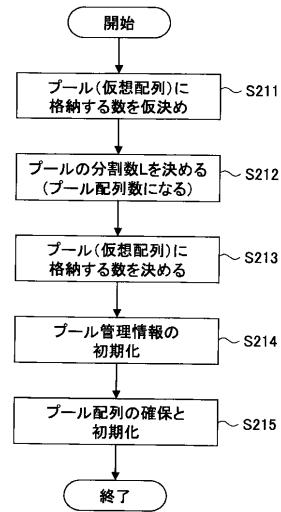
【図15】

第2の実施形態における順序ランダム化装置の動作フローチャート



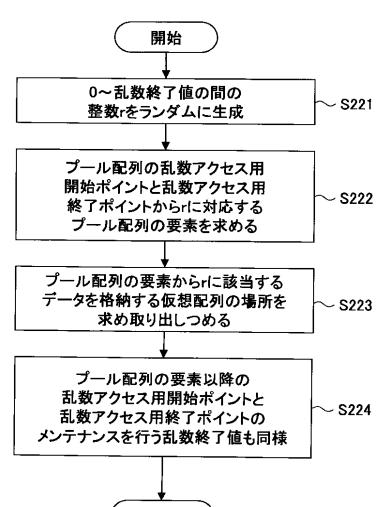
【図14】

第2の実施形態における順序ランダム化装置の初期処理フローチャート



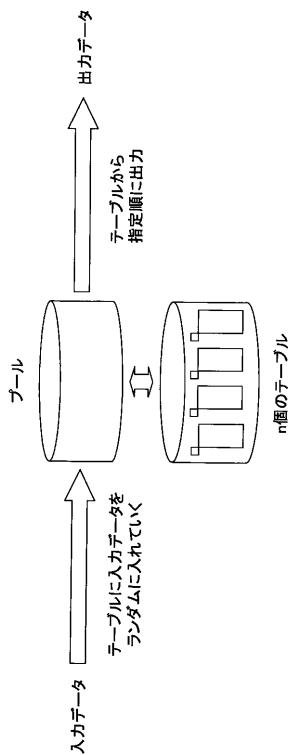
【図16】

第2の実施形態における順序ランダム化装置のプール取り出しひフローチャート



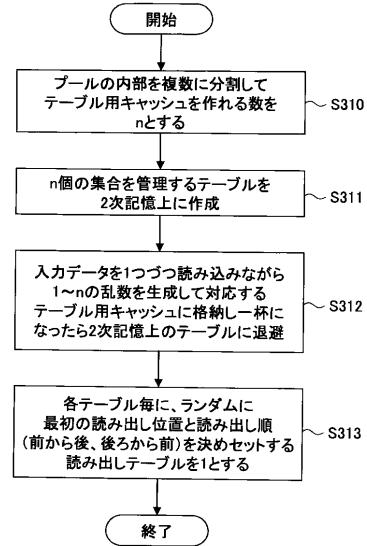
【図17】

第3の実施形態における概要を説明する図



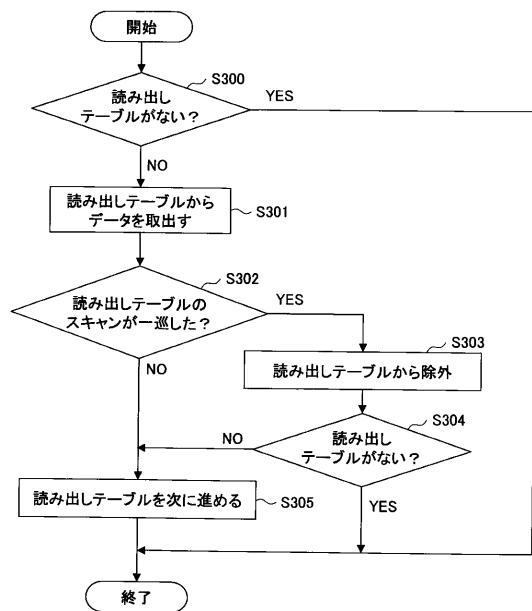
【図18】

第3の実施形態における順序ランダム化装置の初期処理フローチャート



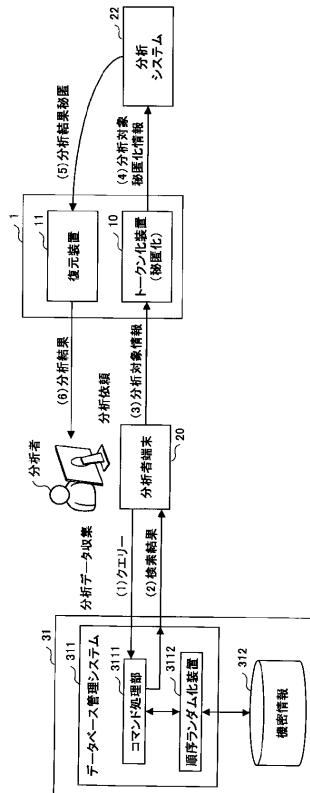
【図19】

第3の実施形態における順序ランダム化装置の動作フローチャート



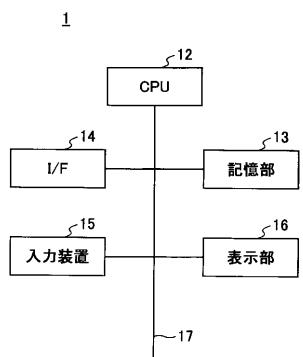
【図20】

第4の実施形態におけるシステム全体構成図



## 【図21】

トークン管理装置のハードウェア構成図



---

フロントページの続き

(72)発明者 伊藤 孝一

神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内

審査官 宮司 卓佳

(56)参考文献 特開2007-317075 (JP, A)

米国特許出願公開第2003/0220927 (US, A1)

特開2007-287102 (JP, A)

(58)調査した分野(Int.Cl., DB名)

G 06 F 21/62

G 06 F 17/30

G 09 C 1/00

H 04 L 9/32