

(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
2020年11月26日(26.11.2020)



(10) 国際公開番号

WO 2020/235014 A1

(51) 国際特許分類:
G06F 21/62 (2013.01)

(21) 国際出願番号: PCT/JP2019/020164

(22) 国際出願日: 2019年5月21日(21.05.2019)

(25) 国際出願の言語: 日本語

(26) 国際公開の言語: 日本語

(71) 出願人: 日本電信電話株式会社 (**NIPPON TELEGRAPH AND TELEPHONE CORPORATION**) [JP/JP]; 〒1008116 東京都千代田区大手町一丁目5番1号 Tokyo (JP).

(72) 発明者: 美原 義行 (**MIHARA, Yoshiyuki**); 〒1808585 東京都武蔵野市緑町3丁目9-1 1 N T T 知的財産センタ内 Tokyo (JP).

(74) 代理人: 伊東 忠重, 外 (**ITOH, Tadashige et al.**); 〒1000005 東京都千代田区丸の内二丁目1番1号 丸の内 M Y P L A Z A (明治安田生命ビル) 16階 Tokyo (JP).

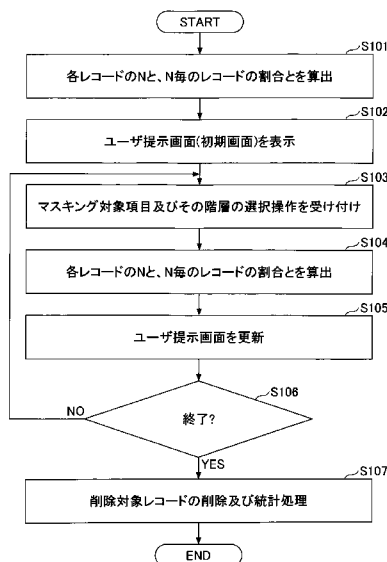
(81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, 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, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, 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, RS,

(54) **Title:** INFORMATION PROCESSING DEVICE, INFORMATION PROCESSING METHOD, AND PROGRAM

(54) 発明の名称: 情報処理装置、情報処理方法及びプログラム

[図7]



S101, S104 Calculate N of each record and percentage of records for each N
S102 Display user presentation screen (initial screen)
S103 Receive selected operation for item for masking and hierarchy thereof
S105 Update user presentation screen
S106 Completed?
S107 Perform statistical processing and deletion of records for deletion

(57) **Abstract:** An information processing device for anonymizing data constituted by records including one or more items by means of statistical processing, said information processing device being characterized by comprising: a calculation means for calculating the number N of records in an individual set and the percentage of records belonging to a set of N records by classifying each record constituting the data into one or more sets on the basis of items for masking, which indicate items that are to be masked, among the items, a dictionary in which categories of item values are expressed in a tree structure for each item for masking, a selection hierarchy indicating a hierarchy selected in the tree structure for each item for masking, and the number of records included in the data; and a display means for displaying the proportion of records belonging to set of N records as a UI for each N records in each hierarchy for each item for masking.

[続葉有]

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

添付公開書類：

一 国際調査報告（条約第21条(3)）

(57) 要約：1以上の項目が含まれるレコードで構成されるデータを統計加工によって匿名化する情報処理装置であって、前記項目のうち、マスキングの対象となる項目を示すマスキング対象項目と、前記マスキング対象項目毎に該項目値のカテゴリが木構造で表現された辞書と、前記マスキング対象項目毎に前記木構造で選択された階層を示す選択階層と、前記データに含まれるレコード数とに基づいて、前記データを構成する各レコードを1つ以上の集合に分類して、各集合のレコード数 N と、レコード数 N の集合に属するレコードの割合とを算出する算出手段と、前記マスキング対象項目毎の各階層において、前記レコード数 N 毎に、レコード数 N の集合に属するレコードの割合を U_i として表示する表示手段と、を有することを特徴とする。

明 細 書

発明の名称： 情報処理装置、情報処理方法及びプログラム

技術分野

[0001] 本発明は、情報処理装置、情報処理方法及びプログラムに関する。

背景技術

[0002] 近年、様々なデータ（例えば、購買データや宿泊データ、人流データ、医療データ、交通データ等）を収集及び分析して、事業活動や行政活動等に活用する取り組みが行われている。

[0003] これらのデータには、例えば、商品の購入者や宿泊者を特定可能な情報（個人情報）等が含まれる場合がある。このため、例えば、小売店やデパート等の商業施設が購買データをデータ収集・分析業者等に第三者提供したり、宿泊施設が宿泊データをデータ収集・分析業者等に第三者提供したりする際には、いわゆる個人情報保護法の規定を遵守する必要がある。個人情報保護法では、そのガイドラインにおいて、統計情報は特定の個人との対応関係が排斥されている限り、個人情報に該当しない旨を規定している。

[0004] また、個人の特定確率を $1/k$ 以下にするデータ加工手法として、 k -匿名化と呼ばれる手法が知られている（例えば、非特許文献1参照）。

先行技術文献

非特許文献

[0005] 非特許文献1：渡邊 奈津美，土井 洋，趙 晋輝，「 k -匿名化手法の効率向上に関する一提案」，情報処理学会 第75回全国大会講演論文集，2013(1)，519-520（2013-03-06）

発明の概要

発明が解決しようとする課題

[0006] しかしながら、第三者提供の対象となるデータに対して統計加工を行って、個人の特定確率を $1/k$ 以下にする場合、データ中のレコードのうち、個人の特定確率が $1/k$ より大きくなるレコードは削除する必要がある。他方

で、データ中のレコードの削除が多い場合（つまり、データの損失率が高い場合）、データ分析等の精度が低下する。

[0007] ここで、匿名化の粒度が粗いほどデータ中のレコードの削除が少なくなるが、1レコードあたりの情報の損失は大きくなる。一方で、匿名化の粒度が細かいほどデータ中のレコードの削除は多くなるが、1レコードあたりの情報の損失は小さくなる。このため、第三者提供の対象となるデータを可能な限り細かい粒度で匿名化して、匿名化による情報の損失は小さくしつつ、個人の特定確率が $1/k$ より大きくなるレコードの削除数を可能な限り少なくすることが好ましい。

[0008] 本発明は、上記の点に鑑みてなされたもので、データの匿名化粒度の決定を支援することを目的とする。

課題を解決するための手段

[0009] 上記目的を達成するため、本発明の実施の形態における情報処理装置は、1以上の項目が含まれるレコードで構成されるデータを統計加工によって匿名化する情報処理装置であって、前記項目のうち、マスキングの対象となる項目を示すマスキング対象項目と、前記マスキング対象項目毎に該項目値のカテゴリが木構造で表現された辞書と、前記マスキング対象項目毎に前記木構造で選択された階層を示す選択階層と、前記データに含まれるレコード数とに基づいて、前記データを構成する各レコードを1つ以上の集合に分類して、各集合のレコード数 N と、レコード数 N の集合に属するレコードの割合とを算出する算出手段と、前記マスキング対象項目毎の各階層において、前記レコード数 N 毎に、レコード数 N の集合に属するレコードの割合を U_i として表示する表示手段と、を有することを特徴とする。

発明の効果

[0010] データの匿名化粒度の決定を支援することができる。

図面の簡単な説明

[0011] [図1]本発明の実施の形態におけるデータ加工システムの全体構成の一例を示す図である。

[図2]本発明の実施の形態におけるデータ提供端末及びデータ分析装置のハードウェア構成の一例を示す図である。

[図3]対象データの一例を示す図である。

[図4A]分類辞書の一例を示す図である。

[図4B]分類辞書の一例を示す図である。

[図5]データ加工の一例を説明するための図である。

[図6]本発明の実施の形態におけるデータ加工処理部の機能構成の一例を示す図（実施例１）である。

[図7]本発明の実施の形態におけるデータ加工処理の一例を示すフローチャート（実施例１）である。

[図8A]ユーザ提示画面における階層選択の一例を説明するための図である。

[図8B]ユーザ提示画面における階層選択の一例を説明するための図である。

[図8C]ユーザ提示画面における階層選択の一例を説明するための図である。

[図8D]ユーザ提示画面における階層選択の一例を説明するための図である。

[図9A]N毎のレコードの割合の他の表示例を示す図である。

[図9B]N毎のレコードの割合の他の表示例を示す図である。

[図10]本発明の実施の形態におけるデータ加工処理部の機能構成の一例を示す図（実施例２）である。

[図11]本発明の実施の形態におけるデータ加工処理の一例を示すフローチャート（実施例２）である。

[図12]本発明の実施の形態におけるデータ加工処理部の機能構成の一例を示す図（実施例３）である。

[図13]本発明の実施の形態におけるデータ加工処理の一例を示すフローチャート（実施例３）である。

[図14]ユーザ提示画面の一例を示す図（実施例３）である。

[図15]クロス率の算出の一例を説明するための図（その１）である。

[図16]クロス率の算出の一例を説明するための図（その２）である。

[図17]本発明の実施の形態におけるデータ加工処理部の機能構成の一例を示

す図（実施例４）である。

[図18]本発明の実施の形態におけるデータ加工処理の一例を示すフローチャート（実施例４）である。

[図19]本発明の実施の形態におけるデータ加工処理部の機能構成の一例を示す図（実施例５）である。

[図20]本発明の実施の形態におけるデータ加工処理の一例を示すフローチャート（実施例５）である。

[図21]本発明の実施の形態における統計量の減算処理の一例を示すフローチャート（実施例５）である。

[図22]本発明の実施の形態におけるデータ加工処理部の機能構成の一例を示す図（実施例６）である。

[図23]本発明の実施の形態におけるデータ加工処理の一例を示すフローチャート（実施例６）である。

[図24]マスキング対象項目の削除の一例を説明するための図である。

[図25]本発明の実施の形態におけるデータ加工処理部の機能構成の一例を示す図（実施例７）である。

[図26A]分類辞書の修正の一例を説明するための図（その１）である。

[図26B]分類辞書の修正の一例を説明するための図（その１）である。

[図27A]分類辞書の修正の一例を説明するための図（その２）である。

[図27B]分類辞書の修正の一例を説明するための図（その２）である。

[図28]本発明の実施の形態におけるデータ加工処理の一例を示すフローチャート（実施例７）である。

[図29]ユーザ提示画面及び分類辞書修正画面の一例を示す図（実施例７）である。

発明を実施するための形態

[0012] 以下、本発明の実施の形態について説明する。以降で説明する本発明の実施の形態では、第三者提供の対象となるデータを統計加工によって匿名化するデータ加工システム１について説明する。

[0013] なお、本発明の実施の形態では、第三者提供の対象となるデータには何等かの個人情報が含まれていることを想定するが、必ずしも個人情報が含まれていなくてもよい。また、第三者提供の対象となるデータは任意のデータとしてよいが、例えば、小売店やデパート等の商業施設における購買データ、宿泊施設における宿泊データ、飲食店における顧客データ等が挙げられる。これら以外にも、第三者提供の対象となるデータとしては、例えば、人口データ、人流データ、水道使用量データ、医療データ、交通データ等も挙げられる。

[0014] [全体構成]

まず、本発明の実施の形態におけるデータ加工システム１の全体構成について、図１を参照しながら説明する。図１は、本発明の実施の形態におけるデータ加工システム１の全体構成の一例を示す図である。

[0015] 図１に示すように、本発明の実施の形態におけるデータ加工システム１には、１以上のデータ提供端末１０と、データ分析装置２０とが含まれる。各データ提供端末１０とデータ分析装置２０とは、例えばインターネット等の通信ネットワークＮを介して通信可能に接続されている。

[0016] データ提供端末１０は、データ提供者（例えば、商業施設等）が利用する情報処理装置（コンピュータ）である。データ提供端末１０は、データ提供者の操作に応じて、例えば購買データ等のデータをデータ分析装置２０に送信する。このとき、データ提供端末１０は、統計加工によってデータを匿名化した上で、この匿名化後のデータ（以降、「統計加工後データ」とも表す。）をデータ分析装置２０に送信する。

[0017] ここで、データ提供端末１０は、データ加工処理部１００と、分類辞書記憶部２００とを有する。データ加工処理部１００は、分類辞書記憶部２００に記憶されている分類辞書を参照して、統計加工によってデータを匿名化する処理（データ加工処理）を行う。分類辞書とは、各データ提供端末１０でデータを匿名化する際に用いられる木構造の辞書情報（つまり、階層構造を有する辞書情報）のことである。データを構成する各レコードを分類辞書に

よって1つ以上の集合に分類した上で、レコード数がk個未満の集合に属する各レコードを削除すると共に、レコード数がk個以上の集合に属する各レコードに対して統計加工を施すことで、当該データが匿名化される。なお、分類辞書の具体例については後述する。

[0018] データ提供端末10としては、例えば、PC（パーソナルコンピュータ）やスマートフォン、タブレット端末等を用いることができる。なお、以降では、複数のデータ提供端末10の各々を区別する場合は、「データ提供端末10A」、「データ提供端末10B」等と表す。この場合、本発明の実施の形態では、データ提供端末10Aとデータ提供端末10Bとは異なるデータ提供者が利用する端末であるものとする。例えば、データ提供端末10AはデパートAが利用する端末であり、データ提供端末10BはデパートBが利用する端末であるものとする。

[0019] データ分析装置20は、データ収集・分析業者（例えば、データの収集及び分析を行う事業者や自治体等）が利用又は管理する情報処理装置（コンピュータ）又は情報処理システム（コンピュータシステム）である。データ分析装置20は、各データ提供端末10から収集したデータ（つまり、統計加工後データ）を所定の目的に応じて分析（例えば、事業活動や行政活動のための購買分析等）する。

[0020] ここで、データ分析装置20は、データ分析処理部300と、マスターデータ記憶部400とを有する。データ分析処理部300は、統計加工後データを受信すると、この統計加工後データをマスターデータとしてマスターデータ記憶部400に記憶する。また、データ分析処理部300は、マスターデータ記憶部400に記憶されているマスターデータを所定の目的に応じて分析する。これにより、各データ提供端末10から収集したデータが分析される。

[0021] なお、図1に示すデータ加工システム1の全体構成は一例であって、他の構成であってもよい。例えば、データ加工システム1には、データ分析装置20での分析結果を閲覧可能な端末が含まれていてもよい。

[0022] [ハードウェア構成]

次に、本発明の実施の形態におけるデータ提供端末１０及びデータ分析装置２０のハードウェア構成について、図２を参照しながら説明する。図２は、本発明の実施の形態におけるデータ提供端末１０及びデータ分析装置２０のハードウェア構成の一例を示す図である。なお、データ提供端末１０及びデータ分析装置２０は同様のハードウェア構成で実現可能であるため、以降では、主に、データ提供端末１０のハードウェア構成について説明する。

[0023] 図２に示すように、本発明の実施の形態におけるデータ提供端末１０は、ハードウェアとして、入力装置１１と、表示装置１２と、外部Ｉ／Ｆ１３と、ＲＡＭ（Random Access Memory）１４と、ＲＯＭ（Read Only Memory）１５と、プロセッサ１６と、通信Ｉ／Ｆ１７と、補助記憶装置１８とを有する。これら各ハードウェアは、それぞれがバス１９を介して通信可能に接続されている。

[0024] 入力装置１１は、例えばキーボードやマウス、タッチパネル等であり、ユーザが各種の入力操作を行うのに用いられる。表示装置１２は、例えばディスプレイ等であり、データ提供端末１０の処理結果等を表示する。なお、データ分析装置２０は、入力装置１１及び表示装置１２の少なくとも一方を有していなくてもよい。

[0025] 外部Ｉ／Ｆ１３は、外部装置とのインタフェースである。外部装置には、記録媒体１３ａ等がある。データ提供端末１０は、外部Ｉ／Ｆ１３を介して、記録媒体１３ａの読み取りや書き込み等を行うことができる。記録媒体１３ａには、例えば、データ加工処理部１００を実現する１以上のプログラムやデータ分析処理部３００を実現する１以上のプログラム等が記録されていてもよい。

[0026] 記録媒体１３ａとしては、例えば、フレキシブルディスク、ＣＤ（Compact Disc）、ＤＶＤ（Digital Versatile Disk）、ＳＤメモリカード（Secure Digital memory card）、ＵＳＢ（Universal Serial Bus）メモリカード等がある。

[0027] ＲＡＭ１４は、プログラムやデータを一時保持する揮発性の半導体メモリ

である。ROM 15は、電源を切ってもプログラムやデータを保持することができる不揮発性の半導体メモリである。ROM 15には、例えば、OS (Operating System) に関する設定情報や通信ネットワークNに関する設定情報等が格納されている。

[0028] プロセッサ16は、例えばCPU (Central Processing Unit) 等であり、ROM 15や補助記憶装置18等からプログラムやデータをRAM 14上に読み出して処理を実行する演算装置である。データ加工処理部100は、ROM 15や補助記憶装置18等に格納されている1以上のプログラムをRAM 14上に読み出してプロセッサ16が処理を実行することで実現される。同様に、データ分析処理部300は、ROM 15や補助記憶装置18等に格納されている1以上のプログラムをRAM 14上に読み出してプロセッサ16が処理を実行することで実現される。

[0029] 通信I/F 17は、データ提供端末10を通信ネットワークNに接続するためのインタフェースである。データ加工処理部100を実現する1以上のプログラムやデータ分析処理部300を実現する1以上のプログラムは、通信I/F 17を介して、所定のサーバ装置等から取得（ダウンロード）されてもよい。

[0030] 補助記憶装置18は、例えばHDD (Hard Disk Drive) やSSD (Solid State Drive) 等であり、プログラムやデータを格納している不揮発性の記憶装置である。補助記憶装置18に格納されているプログラムやデータには、例えば、OS、当該OS上で各種機能を実現するアプリケーションプログラム等がある。また、データ提供端末10の補助記憶装置18には、データ加工処理部100を実現する1以上のプログラムが格納されている。同様に、データ分析装置20の補助記憶装置18には、データ分析処理部300を実現する1以上のプログラムが格納されている。

[0031] また、分類辞書記憶部200は、例えば、データ提供端末10の補助記憶装置18を用いて実現可能である。同様に、マスタデータ記憶部400は、例えば、データ分析装置20の補助記憶装置18を用いて実現可能である。

なお、分類辞書記憶部 200 は、データ提供端末 10 と通信ネットワーク N 等を介して接続される記憶装置等を用いて実現されていてもよい。同様に、マスタデータ記憶部 400 は、データ分析装置 20 と通信ネットワーク N 等を介して接続される記憶装置等を用いて実現されていてもよい。

[0032] 本発明の実施の形態におけるデータ提供端末 10 は、図 2 に示すハードウェア構成を有することにより、後述する各種処理を実現することができる。同様に、本発明の実施の形態におけるデータ分析装置 20 は、図 2 に示すハードウェア構成を有することにより、後述する各種処理を実現することができる。

[0033] なお、図 2 に示す例では、本発明の実施の形態におけるデータ提供端末 10 及びデータ分析装置 20 がそれぞれ 1 台の装置（コンピュータ）で実現されている場合を示したが、これに限られない。本発明の実施の形態におけるデータ提供端末 10 及びデータ分析装置 20 の少なくとも一方が、複数台の装置（コンピュータ）で実現されていてもよい。また、1 台の装置（コンピュータ）には、複数のプロセッサ 16 や複数のメモリ（RAM 14 や ROM 15、補助記憶装置 18 等）が含まれていてもよい。

[0034] [実施例 1]

最初に、実施例 1 として、データ提供端末 10 で対象データを統計加工によって匿名化する際に、ユーザによる適切な匿名化粒度の決定を支援する UI（ユーザインタフェース）を提供する場合について説明する。対象データとは統計加工の対象となるデータのことであり、例えば、第三者提供の対象となるデータそのもの（つまり、生データ）であってもよいし、第三者提供の対象となるデータを構成する各レコードに対して所定の匿名加工を施したデータであってもよい。

[0035] ここで、匿名化の粒度が細かすぎると、対象データ中の多くのレコードが削除されることで、対象データ全体の情報の損失（つまり、レコード削除に伴う対象データ全体の情報量の損失）が大きくなる、一方で、匿名化の粒度が粗すぎると、対象データ中のレコードの削除は少なくなるものの、1レコ

ードあたりの情報の損失（つまり、対象データを構成する各レコードの情報量の損失）が大きくなる。このため、 k -匿名性を満たしつつ、情報の損失をできるだけ抑えるためには、適切な匿名化粒度を決定する必要がある。

[0036] なお、匿名化の粒度が細かすぎて対象データ中の多くのレコードが削除されると、匿名化後の対象データを分析する際の精度（正確さ）に影響する。すなわち、レコードの削除数が多い場合、対象データ中のレコードの分布が歪み、分析結果が意味を持たなくなってしまう可能性がある。同様に、匿名化の粒度が粗すぎて1レコードあたりの情報量の損失が多い場合も、匿名化後の対象データを分析する際の精度（詳細さ）に影響する。すなわち、1レコードあたりの情報量の損失が大きい場合、大まかな分析しかできず、有用な情報（例えば、集団間の差異等）が発見できなくなる可能性がある。

[0037] 匿名加工とは、第三者提供の対象となるデータを構成する各レコードに含まれる各項目（項目は「フィールド」又は「属性」等と称されてもよい。）のうち、個人を識別可能な情報が設定される項目を削除したり、置き換えたりする処理等のことである。具体的には、第三者提供の対象となるデータが免税店における購買データである場合、購買データを構成する各レコードから項目「パスポート番号」を削除する処理が挙げられる。同様に、例えば、第三者提供の対象となるデータが宿泊施設における宿泊データである場合、宿泊データを構成する各レコードから項目「宿泊者名」を削除したデータ等が挙げられる。

[0038] 以降では、対象データは、第三者提供の対象となるデータを構成する各レコードに対して所定の匿名加工を施したデータであるものとする。

[0039] （対象データ）

まず、対象データの一例として、或る商業施設の購買データを構成する各レコードに対して匿名加工を施したデータについて、図3を参照しながら説明する。図3は、対象データの一例を示す図である。

[0040] 図3に示すように、対象データは複数のレコードで構成されており、各レコードには、少なくとも当該対象データ中で各レコードを一意に識別可能な

項目「レコードID」が含まれている。また、図3に示す例では、各レコードには、項目「住所」や項目「年代」、項目「性別」、項目「金額」が含まれている。例えば、レコードID「1」のレコードには、住所「東京都武蔵野市緑町3丁目」、年代「10代」、性別「男」、金額「500円」が含まれている。これは、例えば、東京都武蔵野市緑町3丁目の店舗（商業施設）にて、10代の男が500円分の商品を購入したことを表している。ただし、図3に示す対象データの各レコードには、これら以外にも、例えば、項目「商品名」や項目「購入個数」、項目「購入日時」、項目「業種」等が含まれていてもよい。

[0041] なお、対象データを構成する各レコードには少なくとも項目「レコードID」が含まれるが、項目「レコードID」以外にどのような項目が各レコードに含まれるかは、対象データの種類（又は対象データの基となったデータの種類）によっても異なり得るし、データ提供者によっても異なり得る。すなわち、例えば、購買データと宿泊データとでは各レコードに含まれる項目は異なり得るし、商業施設Aの購買データと商業施設Bの購買データとでも各レコードに含まれる項目は異なり得る。

[0042] また、図3に示す例では対象データを構成するレコード数が5レコードであるが、これは一例であって、対象データを構成するレコード数は任意である。データ提供者の規模等によっても異なるが、例えば、データ収集・分析業者に対して月次で対象データを提供するような場合、一般には、数千や数万、数十万レコード等といったレコード数になることが想定される。

[0043] （分類辞書）

次に、データ提供端末10の分類辞書記憶部200に記憶されている分類辞書の一例として、図3に示す対象データを提供するデータ提供端末10の分類辞書記憶部200に記憶されている分類辞書について、図4を参照しながら説明する。図4は、分類辞書の一例を示す図である。分類辞書は、例えば、対象データを構成する各レコードに含まれる項目毎に、分類辞書記憶部200に記憶されている。図4では、一例として、項目「住所」の分類辞書

と、項目「年代」の分類辞書とを示す。

[0044] 図4 Aは、項目「住所」の分類辞書の一例である。図4 Aに示すように、項目「住所」の分類辞書はカテゴリ（この例の場合、地域名を表すカテゴリ）の木構造（階層構造）になっており、階層が低いほどより詳細な情報（つまり、より詳細な住所）が表現できるようになっている。例えば、図4 Aに示す例では、「1丁目」、「2丁目」、「緑町」、「武蔵野市」、「三鷹市」、「東京都」等のそれぞれがカテゴリである。後述するように、ユーザによって階層が選択された場合、該当の項目において、この選択された階層未満の階層で表現される情報にマスキングされる。

[0045] 例えば、或るレコードの住所が「東京都武蔵野市緑町3丁目」である場合に、ユーザによって第2階層が選択されると、当該住所が「東京都武蔵野市緑町」とマスキングされる。したがって、この場合、「3丁目」という情報が表現できなくなり、項目「住所」の情報が抽象化される。同様に、例えば、ユーザによって第3階層が選択されると、当該住所が「東京都武蔵野市」とマスキングされる（この場合、「緑町3丁目」という情報が表現できなくなる。）。また、同様に、例えば、ユーザによって第4階層が選択されると、当該住所が「東京都」とマスキングされる（この場合、「武蔵野市緑町3丁目」という情報が表現できなくなる。）。一方で、ユーザによって第1階層が選択された場合には、マスキング前後で当該住所は「東京都武蔵野市緑町3丁目」である。

[0046] 図4 Bは、項目「年代」の分類辞書の一例である。図4 Bに示すように、項目「年代」の分類辞書はカテゴリ（この例の場合、年代の数値幅を表すカテゴリ）の木構造（階層構造）になっており、階層が低いほどより詳細な情報（つまり、より詳細な年代）が表現できるようになっている。例えば、図4 Bに示す例では、「0代」、「10代」、「20代」、「30代」、「0～10代」、「20～30代」、「0～30代」等のそれぞれがカテゴリである。後述するように、ユーザによって階層が選択された場合、該当の項目において、この選択された階層未満の階層で表現される情報にマスキングさ

れる。例えば、或るレコードの年代が「10代」である場合に、ユーザによって第2階層が選択されると、当該年代が「0～10代」にマスキングされる。したがって、この場合、項目「年代」によって表現可能な年齢幅が広がるため、項目「年代」の情報が抽象化される。同様に、ユーザによって第3階層が選択されると、当該年代が「0～30代」にマスキングされる。一方で、ユーザによって第1階層が選択された場合には、マスキング前後で年代は「10代」である。

[0047] より高い階層でマスキングすることで該当の項目の情報を抽象化することができる。このため、これらの項目の情報が互いに一致するレコード同士を同一集合に分類した上で、レコード数が k 個以上の集合に属する各レコードを集合毎に1つのレコードに集約する統計加工を行って、 k -匿名性を満たすようなレコードを作成することが可能となる。一方で、レコード数が k 個未満の集合に属する各レコードによっては統計加工により k -匿名性を満たすようなレコードを作成することはできないため、レコード数が k 個未満の集合に属するレコードは削除する必要がある。

[0048] したがって、データ分析装置20における分析精度を考慮すると、ユーザは、 k -匿名性を満たしつつ、削除されるレコード数を減らすように、該当の項目（この項目を以降では「マスキング対象項目」とも表す。）の階層を選択する必要がある。すなわち、ユーザは、 k -匿名性を満たしつつ、匿名化の粒度が可能な限り細くなるように、マスキング対象項目の階層を選択する必要がある。

[0049] なお、どのような分類辞書が分類辞書記憶部200に記憶されているかは、対象データの種類（又は対象データの基となったデータの種類）によっても異なり得るし、データ提供者によっても異なり得る。すなわち、例えば、購買データのマスキングに用いられる分類辞書と宿泊データのマスキングに用いられる分類辞書とは異なり得るし、商業施設Aの購買データのマスキングに用いられる分類辞書と商業施設Bの購買データのマスキングに用いられる分類辞書とは異なり得る。

[0050] 例えば、上述した項目「住所」や項目「年代」以外にも、項目「業種」の分類辞書が挙げられる。項目「業種」の分類辞書としては、例えば、第4階層として「小売り」や「飲食」、第4階層「小売り」の第3階層として「電気店」や「デパート」、第3階層「デパート」の第2階層として「デパートA」や「デパートB」、第2階層「デパートA」の第1階層として「〇〇店」や「××店」等とすればよい。

[0051] (データ加工の概略)

次に、マスキング対象項目を項目「住所」及び項目「年代」として、図4に示す分類辞書によって図3に示す対象データを統計加工して、匿名化(k-匿名化)するデータ加工の概略について説明する。図5は、データ加工の一例を説明するための図である。なお、図5に示す例では、 $k = 2$ であるものとして説明する。

[0052] Step 1) データ加工処理部100は、対象データを構成する各レコードのマスキング対象項目を、選択された階層(以降、「選択階層」とも表す。)でマスキングする。ここで、一例として、項目「住所」の選択階層を第3階層、項目「年代」の選択階層を第3階層としてマスキングしたものとする。

[0053] Step 2) データ加工処理部100は、マスキング後の対象データを構成する各レコードについて、各マスキング対象項目の情報(つまり、項目「住所」の項目値と項目「年代」の項目値。以降、項目の情報(又は項目に設定されている情報)を「項目値」とも表す。)が互いに一致するレコード同士で分類した上で、集合毎に、同一集合に属するレコードの数Nを算出する。そして、データ加工処理部100は、N毎に、Nが同一であるレコードの割合を算出する。なお、割合とは、対象データを構成する全レコード数に対してNが同一であるレコード数の割合のことであり、例えば、「比率」等と称されてもよい。

[0054] 図5に示す例では、レコードID「1」～レコードID「3」の各レコードは、第3階層の項目「住所」の項目値と第3階層の項目「年代」の項目値

とが一致している。このため、これらのレコードは同一集合に分類され、この集合に属するレコードのNの値は $N = 3$ となる。

[0055] 一方で、レコードID「4」のレコード及びレコードID「5」のレコードは、第3階層の項目「住所」の項目値と第3階層の項目「年代」の項目値とが一致する他のレコードが存在しない。このため、レコードID「4」のレコードが分類される集合には、このレコードのみが属することにより、そのNは $N = 1$ となる。同様に、レコードID「5」のレコードのNも $N = 1$ となる。

[0056] また、 $N = 3$ であるレコードの割合は $3 / 5 \times 100 = 60$ （％）となり、 $N = 1$ であるレコードの割合は $2 / 5 \times 100 = 40$ （％）となる。なお、後述するように、N毎のレコードの割合は、例えば、ユーザに提示される。この割合を参照することで、ユーザは、マスキング対象項目に対する適切な階層を選択することができるようになる。なお、Nがk未満のレコードの割合の合計（つまり、 $N (< k)$ であるレコードが属する集合のレコード数の割合の合計）が、削除されるレコードの割合を表す。この割合がより小さくなるように、ユーザはUIを確認しながら選択階層を設定する。

[0057] Step 3) データ加工処理部100は、対象データを構成する各レコードのうち、Nがk未満であるレコードを削除すると共に、Nがk以上である各レコードを同一集合内で統計加工する。

[0058] 図5に示す例では、レコードID「1」～レコードID「3」のレコードの項目「性別」を削除した上で、人数（つまり、レコード数又はヒット数）をカウントして項目「人数」の項目値とすると共に、項目「金額」の項目値を合計する統計加工を行っている。これにより、k-匿名性を満たすレコードが作成される。なお、この統計加工は一例であって、任意の統計加工（例えば、平均値の計算や中央値の計算等）を行ってもよい。

[0059] なお、上記の統計加工は、Nがk以上であるレコードが属する集合毎に行われる。例えば、Nがk以上であるレコードが属する集合として第1の集合と第2の集合とが存在する場合、第1の集合内で各レコードを統計加工する

と共に、第2の集合内で各レコードを統計加工する。これにより、 k -匿名性を満たすレコードとして、第1の集合に対応するレコードと、第2の集合に対応するレコードとが作成する。

[0060] (データ加工処理部100の機能構成)

まず、実施例1におけるデータ加工処理部100の機能構成について、図6を参照しながら説明する。図6は、本発明の実施の形態におけるデータ加工処理部100の機能構成の一例を示す図(実施例1)である。

[0061] 図6に示すように、実施例1におけるデータ加工処理部100には、算出部101と、UI提供部102と、データ加工部103とが含まれる。

[0062] 算出部101は、予め設定されたマスキング対象項目と、分類辞書記憶部200に記憶されている分類辞書と、各マスキング対象項目の階層と、対象データを構成するレコード数とに基づいて、対象データを構成する各レコードを分類して、これら各レコードが分類された集合毎に、同一集合に属するレコードの数 N を算出する。そして、算出部101は、 N 毎に、 N が同一であるレコードの割合を算出する。ここで、上述したように、算出部101は、該当の階層でマスキングされた各マスキング対象項目の項目値が互いに一致するレコード同士を同一集合に分類する。

[0063] UI提供部102は、算出部101により算出された N 毎のレコードの割合が含まれるユーザ提示画面を表示する。また、UI提供部102は、ユーザ提示画面におけるユーザの各種操作(例えば、階層の選択操作)を受け付ける。

[0064] データ加工部103は、UI提供部102により表示されたユーザ提示画面におけるユーザ操作に応じて、同一集合に属するレコード数 N が k 未満のレコードを削除すると共に、 N が k 以上である各レコードを同一集合内で統計加工する。

[0065] (データ加工処理)

次に、データ提供端末10で対象データを統計加工して、匿名化(k -匿名化)するデータ加工処理について、図7を参照しながら説明する。図7は

、本発明の実施の形態におけるデータ加工処理の一例を示すフローチャート（実施例１）である。なお、対象データは、データ提供端末１０の補助記憶装置１８に記憶されていてもよいし、データ提供端末１０とローカルな通信ネットワーク（例えば、社内ネットワーク等）を介して接続される記憶装置等に記憶されていてもよい。また、以降では、 $k = 5$ であるものとする。

[0066] まず、算出部１０１は、予め設定されたマスキング対象項目と、分類辞書記憶部２００に記憶されている分類辞書と、各マスキング対象項目の階層と、対象データを構成するレコード数とに基づいて、対象データを構成する各レコードを分類した場合に同一集合に属するレコードの数 N （つまり、集合毎のレコード数 N ）と、 N 毎のレコードの割合とを算出する（ステップＳ１０１）。ここで、ステップＳ１０１では、算出部１０１は、各マスキング対象項目の選択階層が「第１階層」であるものとして、選択階層での集合毎のレコード数 N 及び N 毎のレコードの割合と、１つのマスキング対象項目のみ階層を上げた場合における集合毎のレコード数 N 及び N 毎のレコードの割合とを算出する。

[0067] 例えば、マスキング対象項目を項目「住所」及び項目「年代」とした場合、算出部１０１は、以下の集合毎のレコード数 N 及び N 毎のレコードの割合を算出する。

[0068] ・項目「住所」の階層が「第１階層」、かつ、項目「年代」の階層が「第１階層」である場合における集合毎のレコード数 N 及び N 毎のレコードの割合

・項目「住所」の階層が「第２階層」、かつ、項目「年代」の階層が「第１階層」である場合における集合毎のレコード数 N 及び N 毎のレコードの割合

・項目「住所」の階層が「第３階層」、かつ、項目「年代」の階層が「第１階層」である場合における集合毎のレコード数 N 及び N 毎のレコードの割合

・項目「住所」の階層が「第４階層」、かつ、項目「年代」の階層が「第

1 階層」である場合における集合毎のレコード数N及びN毎のレコードの割合

・項目「住所」の階層が「第1階層」、かつ、項目「年代」の階層が「第2階層」である場合における集合毎のレコード数N及びN毎のレコードの割合

・項目「住所」の階層が「第1階層」、かつ、項目「年代」の階層が「第3階層」である場合における集合毎のレコード数N及びN毎のレコードの割合

・項目「住所」の階層が「第1階層」、かつ、項目「年代」の階層が「第4階層」である場合における集合毎のレコード数N及びN毎のレコードの割合

このように、算出部101は、まず、各マスキング対象項目の選択階層が「第1階層」であるものとして、1つのマスキング対象項目のみ階層を上げた場合における集合毎のレコード数NとN毎のレコードの割合とをそれぞれ算出する。

[0069] ここで、上述したように、算出部101は、該当の階層でマスキングされた各マスキング対象項目の項目値が互いに一致するレコード同士を同一集合に分類する。例えば、項目「住所」の階層が「第1階層」、かつ、項目「年代」の階層が「第1階層」である場合、算出部101は、「第1階層」でマスキングされた項目「住所」の項目値と、「第1階層」でマスキングされた項目「年代」の項目値との両方が一致するレコード同士を同一集合に分類する。同様に、例えば、項目「住所」の階層が「第2階層」、かつ、項目「年代」の階層が「第1階層」である場合、算出部101は、「第2階層」でマスキングされた項目「住所」の項目値と、「第1階層」でマスキングされた項目「年代」の項目値との両方が一致するレコード同士を同一集合に分類する。同様に、例えば、項目「住所」の階層が「第3階層」、かつ、項目「年代」の階層が「第1階層」である場合、算出部101は、「第3階層」でマスキングされた項目「住所」の項目値と、「第1階層」でマスキングされた

項目「年代」の項目値との両方が一致するレコード同士を同一集合に分類する。以降も同様である。

[0070] 以降では、一例として、マスキング対象項目は項目「住所」及び項目「年代」であるものとして説明を続ける。なお、本実施例ではマスキング対象項目が予め設定されているものとするが、マスキング対象項目はユーザにより選択及び設定されてもよい。

[0071] 次に、UI提供部102は、上記のステップS101で算出されたN毎のレコードの割合が含まれるユーザ提示画面を表示する（ステップS102）。すなわち、UI提供部102は、例えば、図8Aに示すユーザ提示画面G100を表示する。

[0072] 図8Aに示すユーザ提示画面G100は、データ加工のための階層をユーザが選択する際に表示される初期画面であり、ユーザ提示情報表示欄G110と、決定ボタンG120とが含まれる。

[0073] 図8Aに示すユーザ提示画面G100のユーザ提示情報表示欄G110では、選択階層が網掛けで表示されている。また、図8Aに示すユーザ提示画面G100のユーザ提示情報表示欄G110には、上記のステップS101で算出されたN毎のレコードの割合が、マスキング対象項目の階層を変化させた場合におけるN毎のレコードの割合として表示される。

[0074] 図8Aに示す例では、項目「住所」及び項目「年代」の選択階層は共に「第1階層」であり、この場合の各集合のレコード数は $N=1$ で、 $N=1$ のレコードの割合は100%（つまり、レコード数が $N=1$ の集合に属するレコードの割合は100%）であることが表示されている。

[0075] また、このとき、項目「住所」のみを「第2階層」に上げた場合、レコード数が $N=2$ の集合に属するレコードの割合は40%、レコード数が $N=1$ の集合に属するレコードの割合は60%になることが表示されている。同様に、項目「住所」のみを「第3階層」に上げた場合、レコード数が $N=3$ の集合に属するレコードの割合は60%、レコード数が $N=1$ の集合に属するレコードの割合は40%になることが表示されている。同様に、項目「住所

」のみを「第4階層」に上げた場合、レコード数が $N=3$ の集合に属するレコードの割合は60%、レコード数が $N=1$ の集合に属するレコードの割合は40%になることが表示されている。一方で、項目「年代」のみを「第2階層」以上に上げた場合、レコード数が $N=1$ の集合に属するレコードの割合は100%のままであることが表示されている。

[0076] ユーザは、ユーザ提示情報表示欄G110に表示されている N の値とその割合とを確認することで、どのマスキング対象項目の階層を上げればよいかを知ることができる。例えば、図8Aに示す例の場合、項目「年代」の階層を上げて N の値とその割合とが変化しないため、匿名化の粒度を変化させることはできないと知ることができる。一方で、例えば、項目「住所」の階層を2つ上げることで、「 $N=1:100\%$ 」から「 $N=3:60\%$ 、 $N=1:40\%$ 」に変化させることができると知ることができる。なお、決定ボタンG120がユーザによって押下されることで、選択階層で対象データを構成する各レコードをデータ加工することができる。

[0077] 以降では、ユーザは、項目「住所」の階層を「第3階層」にする選択操作を行ったものとして説明を続ける。なお、ユーザは、例えば、ユーザ提示情報表示欄G110において、所望のマスキング対象項目と所望の階層とが交差するセルを押下することで、所望のマスキング対象項目に対する階層の選択操作を行うことができる。

[0078] 次に、UI提供部102は、マスキング対象項目に対する階層の選択操作を受け付ける（ステップS103）。上述したように、項目「住所」に対する「第3階層」の選択操作がユーザにより行われたものとして、UI提供部102は、この選択操作を受け付けたものとする。

[0079] 次に、算出部101は、上記のステップS101と同様に、集合毎のレコード数 N と、 N 毎のレコードの割合とを算出する（ステップS104）。ここで、ステップS104では、算出部101は、各マスキング対象項目の選択階層での集合毎のレコード数 N 及び N 毎のレコードの割合と、1つのマスキング対象項目のみ階層を上げた場合における集合毎のレコード数 N 及び N

毎のレコードの割合とを算出する。

[0080] 例えば、項目「住所」の階層として「第3階層」、項目「年代」の階層として「第1階層」が選択されている場合、算出部101は、以下の集合毎のレコード数N及びN毎のレコードの割合を算出する。

[0081] ・項目「住所」の階層が「第3階層」、かつ、項目「年代」の階層が「第1階層」である場合における集合毎のレコード数N及びN毎のレコードの割合

・項目「住所」の階層が「第1階層」、かつ、項目「年代」の階層が「第1階層」である場合における集合毎のレコード数N及びN毎のレコードの割合

・項目「住所」の階層が「第2階層」、かつ、項目「年代」の階層が「第1階層」である場合における集合毎のレコード数N及びN毎のレコードの割合

・項目「住所」の階層が「第4階層」、かつ、項目「年代」の階層が「第1階層」である場合における集合毎のレコード数N及びN毎のレコードの割合

・項目「住所」の階層が「第3階層」、かつ、項目「年代」の階層が「第2階層」である場合における集合毎のレコード数N及びN毎のレコードの割合

・項目「住所」の階層が「第3階層」、かつ、項目「年代」の階層が「第3階層」である場合における集合毎のレコード数N及びN毎のレコードの割合

・項目「住所」の階層が「第3階層」、かつ、項目「年代」の階層が「第4階層」である場合における集合毎のレコード数N及びN毎のレコードの割合

このように、算出部101は、各マスキング対象項目のうちの1つのマスキング対象項目の階層のみを、選択階層から変化させた場合における集合毎のレコード数NとN毎のレコードの割合とをそれぞれ算出する。

[0082] 次に、UI提供部102は、上記のステップS102で表示されたユーザ提示画面を更新して、上記のステップS104で算出されたN毎のレコードの割合が含まれるユーザ提示画面を表示する（ステップS105）。すなわち、UI提供部102は、例えば、図8Aに示すユーザ提示画面G100のユーザ提示情報表示欄G110を更新して、図8Bに示すユーザ提示画面G100を表示する。

[0083] 図8Bに示すユーザ提示画面G100のユーザ提示情報表示欄G110では、選択階層が網掛けで表示されている。図8Bに示す例では、項目「住所」の選択階層は「第3階層」であり、項目「年代」の選択階層は「第1階層」である。

[0084] また、図8Bに示すユーザ提示画面G100のユーザ提示情報表示欄G110には、上記のステップS104で算出されたN毎のレコードの割合が、マスキング対象項目の階層を変化させた場合におけるN毎のレコードの割合として表示される。

[0085] 図8Bに示す例では、項目「住所」及び項目「年代」の選択階層において、レコード数がN=3の集合に属するレコードの割合は60%、レコード数がN=1の集合に属するレコードの割合は40%であることが表示されている。

[0086] また、このとき、項目「住所」のみを「第4階層」に上げた場合、レコード数がN=3の集合に属するレコードの割合は60%、レコード数がN=1の集合に属するレコードの割合は40%のままであることが表示されている。同様に、項目「住所」のみを「第2階層」に下げた場合、レコード数がN=2の集合に属するレコードの割合は40%、レコード数がN=1の集合に属するレコードの割合は60%になることが表示されている。同様に、項目「住所」のみを「第1階層」に上げた場合、レコード数がN=1の集合に属するレコードの割合は100%になることが表示されている。一方で、項目「年代」のみを「第2階層」以上に上げた場合、レコード数がN=3の集合に属するレコードの割合は60%、レコード数がN=1の集合に属するレコ

ードの割合は40%のままであることが表示されている。

[0087] ユーザは、ユーザ提示情報表示欄G110に表示されているNの値とその割合とを確認することで、どのマスキング対象項目の階層を上げればよいかを知ることができる。例えば、図8Bに示す例の場合、項目「年代」の階層を上げてもNの値とその割合とが変化しないため、匿名化の粒度を変化させても、匿名化可能なレコード数を増やす（つまり、削除されるレコード数を減らす）ことはできないと知ることができる。したがって、図8Bに示す例の場合、ユーザは、項目「住所」の階層を1つ上げる操作を行うことが考えられる。

[0088] 次に、UI提供部102は、マスキング対象項目の階層選択を終了するかどうかを判定する（ステップS106）。ここで、UI提供部102は、例えば、ユーザによって決定ボタンG120が押下された場合に、マスキング対象項目の階層選択を終了すると判定すればよい。

[0089] ステップS106でマスキング対象項目の階層選択を終了すると判定されなかった場合、データ加工処理部100は、ステップS103に戻る。これにより、マスキング対象項目の階層選択が終了するまで、上記のステップS103～ステップS105が繰り返し実行される。

[0090] 例えば、図8Bに示すユーザ提示画面G100において項目「住所」の階層として「第4階層」がユーザによって選択された場合、UI提供部102により、図8Cに示すユーザ提示画面G100が表示される。図8Cに示すユーザ提示画面G100では、項目「住所」の選択階層として「第4階層」が、項目「年代」の選択階層として「第1階層」が選択されている。ユーザは、図8Cに示すユーザ提示画面G100のユーザ提示情報表示欄G110に表示されているNの値とその割合とを確認することで、例えば、項目「年代」の階層を「第3階層」まで上げることで、k-匿名性を確保しつつ（つまり、レコードの削除数を最低限に抑えたまま）、匿名化の粒度を最も細くすることができることを知ることができる。

[0091] 例えば、図8Cに示すユーザ提示画面G100において項目「年代」の階

層として「第3階層」がユーザによって選択された場合、UI提供部102により、図8Dに示すユーザ提示画面G100が表示される。図8Dに示すユーザ提示画面G100では、項目「住所」の選択階層として「第4階層」が、項目「年代」の選択階層として「第3階層」が選択されている。ユーザは、図8Dに示すユーザ提示画面G100のユーザ提示情報表示欄G110に表示されているNの値とその割合とを確認することで、例えば、項目「住所」及び項目「年代」の選択階層にて、k-匿名性を確保しつつ（つまり、レコードの削除数を最低限に抑えたまま）、匿名化の粒度を最も細かくすることができると知ることができる。

[0092] このように、ユーザは、ユーザ提示情報表示欄G110に表示されているNの値とその割合とを確認することで、N毎のレコードの割合を確認することができるため、Nがk以上となるレコードの割合を知ることができる。これにより、ユーザは、例えば、各マスキング対象項目の階層をできるだけ低くしつつ、Nがk以上となるレコードの割合が高くなるようにすることで、k-匿名性を確保しつつ、可能な限り細かい粒度で多くのレコードを匿名化することが可能となる。すなわち、ユーザは、Nの値とその割合とを確認することで、適切な匿名化粒度を決定することができるようになる。

[0093] 一方、ステップS106でマスキング対象項目の階層選択を終了すると判定された場合、データ加工部103は、同一集合に属するレコード数Nがk未満のレコードを削除すると共に、Nがk以上である各レコードを同一集合内で統計加工する（ステップS107）。これにより、k-匿名性を有するレコードが作成され、これらのレコードで構成される統計加工後データが得られる。なお、統計加工の処理内容については、対象データの種類（又は対象データの基となったデータの種類）によって異なる。例えば、対象データの基となったデータが購買データである場合、統計加工の処理としては、金額の合計の算出、購入個数の合計の算出、購入者数の合計の算出、不要な項目（例えば、性別等）の削除等が挙げられる。

[0094] 上記のステップS107で作成された統計加工後データは、データ加工処

理部100により、データ分析装置20に送信される。そして、データ分析装置20のデータ分析処理部300は、受信した統計加工後データをマスタデータ記憶部400に記憶する。これにより、マスタデータ記憶部400にはマスタデータが蓄積され、データ分析処理部300は、これらのマスタデータを所定の目的に応じて分析することが可能となる。

[0095] なお、本実施例では、図8A～図8Dに示すように、ユーザ提示画面G100を遷移させたが、ユーザによる階層選択を戻す（取り消す）ことで、画面遷移を戻すことができてもよい。例えば、図8Bに示すユーザ提示画面G100から図8Aに示すユーザ提示画面G100に戻ることができてもよい。この場合、例えば、画面遷移を戻すための「戻る」ボタンやリンク等がユーザ提示画面G100に含まれており、ユーザが「戻る」ボタンやリンク等を押下することで、画面遷移を戻すことができてもよい。

[0096] また、画面遷移が戻った場合にはN毎のレコードの割合が算出部101によって再度算出されてもよいが、例えば、画面遷移を戻す場合のために補助記憶装置18等に予め履歴としてN毎のレコードの割合を記憶させておき、画面遷移が戻った場合には、履歴として記憶されているN毎のレコードの割合を用いてもよい。同様に、例えば、過去に選択されたことがある階層が再度選択された場合にも、履歴として記憶されているN毎のレコードの割合が用いられてもよい。

[0097] 適切な匿名化粒度を決定するために、ユーザは、UI上で選択階層を頻繁に変更しながら試行錯誤を行うことが予想される。このため、上記のように履歴として記憶されている情報を用いることで、選択階層の変更や画面遷移の際の処理時間を短縮させることが可能となる。このような処理時間の短縮は、対象データの規模が大きくなるほど（つまり、対象データを構成するレコード数が多くなるほど）顕著になる。

[0098] （ユーザ提示情報の他の表示例）

本実施例では、ユーザ提示情報表示欄G110にてN毎のレコードの割合を表示する例を示したが、これ以外にも種々の表示方法にてN毎のレコード

の割合が表示されてもよい。

[0099] 例えば、図9Aに示すように、円グラフにてN毎のレコードの割合が表示されてもよい。図9Aに示す例では、 $N=1$ であるレコードの割合は68%、 $N=2$ であるレコードの割合は14%、 $N=3$ であるレコードの割合は6%、 $N=4$ であるレコードの割合は3%、 $N=5$ であるレコードの割合は2%等と円グラフで表示されている。また、図9Aに示す例では、 $N=1$ であるレコード数は14334件、 $N=2$ であるレコード数は2959件等と、N毎のレコード数も表示されている。

[0100] また、例えば、図9Bに示すように、棒グラフにてN毎のレコード数が表示されてもよい。図9Bに示す例では、 $N=1$ であるレコード数は14件、 $N=2$ であるレコード数は9件、 $N=3$ であるレコード数は4件、 $N=4$ であるレコード数は3件、 $N \geq 5$ であるレコード数は2件と棒グラフで表示されている。

[0101] なお、上記の図9A及び図9B以外にも、例えば、積み上げ棒グラフや折れ線グラフ等の種々のグラフにてN毎のレコードの割合（又はN毎のレコード数）が表示されてもよい。

[0102] また、N毎のレコードの割合が表示される代わりに、例えば、Nがk以上のレコードの割合と、Nがk未満のレコードの割合とが表示されてもよい。これにより、ユーザは、削除されるレコード（つまり、Nがk未満のレコード）の割合を容易に把握することができるようになる。

[0103] [実施例2]

次に、実施例2として、データ提供端末10で対象データを統計加工によって匿名化する際に、自動的に適切な匿名化粒度を決定する場合について説明する。なお、実施例2では、実施例1と同一の構成要素についてはその説明を省略する。

[0104] （データ加工処理部100の機能構成）

まず、実施例2におけるデータ加工処理部100の機能構成について、図10を参照しながら説明する。図10は、本発明の実施の形態におけるデー

タ加工処理部１００の機能構成の一例を示す図（実施例２）である。

[0105] 図１０に示すように、実施例２におけるデータ加工処理部１００には、算出部１０１と、データ加工部１０３と、選択部１０４と、終了条件判定部１０５とが含まれる。また、実施例２におけるデータ加工処理部１００には、ＵＩ提供部１０２が含まれていてもよいし、ＵＩ提供部１０２が含まれていなくてもよい。

[0106] 選択部１０４は、算出部１０１による算出結果と、マスキング対象項目の優先度とに基づいて、各マスキング対象項目の階層を選択する。ここで、マスキング対象項目の優先度とは、階層を上げるマスキング対象項目を選択するための値である。選択部１０４は、例えば、優先度が低いマスキング対象項目の階層を上げるように、各マスキング対象項目の階層を選択する。なお、優先度としては、ユーザによって設定された数値等が用いられてもよいし、任意の方法によって算出された各種スコアが用いられてもよい。各種スコアとしては、例えば、後述するクロス率や損失率、集約率、分離率、カバー率等を用いることができる。また、複数のスコアを用いる場合には、スコア間の優先度が設定されてもよいし、スコアの和や重み付き和が用いられてもよい。

[0107] なお、スコアの種類によっては、スコアの値が高いほど良い場合とスコアの値が低いほど良い場合とがある。複数のスコアを用いる場合に、このようなスコアが混在している場合には、適宜、逆数をとったり、負数をとったりすればよい。

[0108] 終了条件判定部１０５は、所定の終了条件を満たしたか否かを判定する。終了条件とは、算出部１０１による算出と、選択部１０４による階層選択との繰り返しを終了させるための条件のことである。したがって、この終了条件を満たすまで、算出部１０１による算出と、選択部１０４による階層選択とが繰り返し実行される。

[0109] （データ加工処理）

次に、データ提供端末１０で対象データを統計加工して、匿名化（ k -匿

名化)するデータ加工処理について、図11を参照しながら説明する。図11は、本発明の実施の形態におけるデータ加工処理の一例を示すフローチャート(実施例2)である。

[0110] まず、算出部101は、図7のステップS101と同様に、予め設定されたマスキング対象項目と、分類辞書記憶部200に記憶されている分類辞書と、各マスキング対象項目の階層と、対象データを構成するレコード数とに基づいて、対象データを構成する各レコードを分類した場合に、同一集合に属するレコードの数 N と、 N 毎のレコードの割合とを算出する(ステップS201)。なお、上述したように、算出部101は、各マスキング対象項目は「第1階層」が選択されているものとして、同一集合に属するレコードの数 N と、 N 毎のレコードの割合とを算出する。

[0111] 次に、選択部104は、算出部101による算出結果と、マスキング対象項目の優先度とに基づいて、各マスキング対象項目の階層を選択する(ステップS202)。ここで、選択部104は、以下の(選択条件1)及び(選択条件2)により各マスキング対象項目の階層を選択する。

[0112] (選択条件1) 階層を1つ上げることで N が k 以上のレコードの割合が向上するマスキング対象項目が存在する場合には、当該マスキング対象項目の1つ上の階層を選択する。ここで、 N 毎のレコードの割合が向上するとは、階層を1つ上げることで、 N の値が大きくなり、かつ、当該 N のレコードの割合が大きくなることを意味する。

[0113] (選択条件2) 階層を1つ上げることで N 毎のレコードの割合が向上するマスキング対象項目が存在しない場合には、最も優先度が低いマスキング対象項目の1つ上の階層を選択する。

[0114] なお、上記の(選択条件1)及び(選択条件2)は一例であって、選択部104は、他の方法により各マスキング対象項目の階層を選択してもよい。例えば、選択部104は、マスキング対象項目の階層を1つ上げることで N 毎のレコードの割合が向上する度合いと、当該マスキング対象項目の優先度との和や積、重み付き積等により、どのマスキング対象項目の階層を1つ上

げるかを選択してもよい。

[0115] 次に、算出部101は、上記のステップS201と同様に、集合毎のレコード数Nと、N毎のレコードの割合とを算出する（ステップS203）。なお、上述したように、算出部101は、各マスキング対象項目の選択階層での集合毎のレコード数N及びN毎のレコードの割合と、1つのマスキング対象項目のみ階層を上げた場合における集合毎のレコード数N及びN毎のレコードの割合とを算出する。

[0116] 次に、終了条件判定部105は、所定の終了条件を満たしたか否かを判定する（ステップS204）。ここで、終了条件としては、例えば、以下の（終了条件1）～（終了条件3）のいずれかが挙げられる。

[0117] （終了条件1）対象データを構成する全てのレコードのNがk以上となる。

[0118] （終了条件2）後述するステップS205でデータ加工部103によって削除されるレコードが所定の割合（又は所定の件数）以下となる。これは、言い換えれば、Nがk未満であるレコードが所定の割合（又は所定の件数）以下であることを意味する。

[0119] （終了条件3）各マスキング対象項目の階層が、予め設定された上限の階層となる。例えば、項目「住所」の階層は上限が「第3階層」、項目「年代」の階層は上限が「第2階層」と設定されている場合に、項目「住所」の階層が「第3階層」となり、かつ、項目「年代」の階層が「第2階層」となったときである。

[0120] なお、上記以外にも、例えば、終了条件として、繰り返し回数が所定の回数に達したこと等が用いられてもよい。又は、例えば、ユーザによって設定された任意の終了条件が用いられてもよい。

[0121] ステップS204で終了条件を満たすと判定されなかった場合、データ加工処理部100は、ステップS202に戻る。これにより、終了条件を満たすまで、上記のステップS202～ステップS203が繰り返し実行される。なお、例えば、UI提供部102は、適宜、ユーザ提示画面を表示して、

マスキング対象項目の階層をユーザに選択させるようにしてもよい。

[0122] 一方、ステップS204で終了条件を満たすと判定された場合、データ加工部103は、図7のステップS107と同様に、同一集合に属するレコード数 N が k 未満のレコードを削除すると共に、 N が k 以上である各レコードを同一集合内で統計加工する（ステップS205）。これにより、 k -匿名性を有するレコードが作成され、これらのレコードで構成される統計加工後データが得られる。

[0123] このように、実施例2では、各マスキング対象項目の階層が自動的に選択されることで、 k -匿名性を確保しつつ、可能な限り細かい粒度で多くのレコードを匿名化することが可能となる。しかも、実施例2では、ユーザは、マスキング対象項目の階層を選択する必要がないため、対象データを構成する各レコードの匿名化を容易に行うことが可能となる。

[0124] [実施例3]

次に、実施例3として、実施例1と同様のデータ加工を行う際に、指標値の1つであるクロス率を算出した上で、ユーザに提示する場合について説明する。クロス率とは、2以上のデータ集合間で、同一項目で同一情報（つまり、同一項目値）を有するデータ数を表す指標値のことであり、2つ以上の集合間の共通度を表す。本実施例では、対象データを構成する各レコード（第1のレコード集合）と、マスタデータ記憶部400に記憶されているマスタデータを構成する各レコード（第2のレコード集合）との間で、同一項目で同一情報（つまり、同一項目値）を有するレコード数を表す指標値としてクロス率を定義する。クロス率をユーザに提示することで、例えば、当該ユーザは、統計加工後データ（マスタデータ）がクロス分析に用いられることも考慮して、マスキング対象項目の階層を選択することが可能となる。

[0125] ここで、クロス分析を行う際には、第1のレコード集合と第2のレコード集合との間で、分析対象項目における同一項目の項目値の粒度（つまり、当該項目の階層）を揃える必要がある。このため、例えば、対象データを匿名化する際にレコード数を犠牲にして細かい粒度で匿名化を行ったとしても、

マスタデータを構成する各レコードの粒度が粗い場合には、匿名化後の対象データを構成する各レコードの粒度を、マスタデータを構成する各レコードの粒度に揃える必要がある。なお、分析対象項目とは、クロス分析で分析の対象とする項目のことである。

[0126] また、クロス分析の分析対象項目間で、同一項目で共通する項目値（後述する共通値）が或る程度存在していないと、有用なクロス分析を行うことができない。このため、或る程度の共通値が生じるように、粒度を調整する必要がある。例えば、或る2つの会社（A社及びB社）間でチョコレートの購買金額比率を比較したい場合には、A社の購買データと、B社の購買データとの間で、例えば、同一項目「商品種別」で共通する項目値「チョコレート」が含まれるレコードが存在する必要がある。

[0127] なお、実施例3では、実施例1と同一の構成要素についてはその説明を省略する。

[0128] （データ加工処理部100の機能構成）

まず、実施例3におけるデータ加工処理部100の機能構成について、図12を参照しながら説明する。図12は、本発明の実施の形態におけるデータ加工処理部100の機能構成の一例を示す図（実施例3）である。

[0129] 図12に示すように、実施例3におけるデータ加工処理部100には、算出部101と、UI提供部102と、データ加工部103と、マスタデータ取得部106とが含まれる。

[0130] マスタデータ取得部106は、データ分析装置20のマスタデータ記憶部400に記憶されているマスタデータを取得する。マスタデータ取得部106は、例えば、マスタデータの取得要求をデータ分析装置20に送信して、この取得要求の応答として、マスタデータを取得することができる。

[0131] また、実施例3における算出部101は、更に、マスタデータ取得部106により取得されたマスタデータと、対象データとに基づいて、指標値の1つであるクロス率を算出する。

[0132] （データ加工処理）

次に、データ提供端末 10 で対象データを統計加工して、匿名化（k-匿名化）する際に、クロス率もユーザに提示する場合のデータ加工処理について、図 13 を参照しながら説明する。図 13 は、本発明の実施の形態におけるデータ加工処理の一例を示すフローチャート（実施例 3）である。

[0133] まず、マスタデータ取得部 106 は、データ分析装置 20 のマスタデータ記憶部 400 に記憶されているマスタデータを取得する（ステップ S301）。ここで、マスタデータ取得部 106 は、マスタデータを構成する全てのレコードを取得してもよいし、マスタデータを構成する各レコードのうち、所定の条件を満たすレコードのみを取得してもよい、所定の条件としては、例えば、「マスキング対象項目を全て含むレコード」等が挙げられる。

[0134] また、マスタデータ取得部 106 により取得されたマスタデータを構成する各レコードのうち、対象データを構成する各レコードとの間で共通の項目が 1 つも含まれないレコードは、当該マスタデータから削除される。このような削除は、マスタデータ取得部 106 によって行われてもよいし、算出部 101 によって行われてもよい。

[0135] 次に、算出部 101 は、予め設定されたマスキング対象項目と、分類辞書記憶部 200 に記憶されている分類辞書と、各マスキング対象項目の階層と、対象データを構成するレコード数とに基づいて、対象データを構成する各レコードを分類した場合に同一集合に属するレコードの数 N （つまり、集合毎のレコード数 N ）と、 N 毎のレコードの割合と、クロス率とを算出する（ステップ S302）。なお、集合毎のレコード数 N 及び N 毎のレコードの割合は実施例 1 と同様である。また、クロス率についても、各マスキング対象項目は「第 1 階層」が選択されているものとして、クロス率を算出する。クロス率の算出方法については後述する。

[0136] 次に、UI 提供部 102 は、上記のステップ S302 で算出された N 毎のレコードの割合とクロス率とが含まれるユーザ提示画面を表示する（ステップ S303）。すなわち、UI 提供部 102 は、例えば、図 14 に示すユーザ提示画面 G100 を表示する。

- [0137] 図14に示すユーザ提示画面G100のユーザ提示情報表示欄G110には、N毎のレコードの割合に加えて、マスキング対象項目の階層を変化させた場合におけるクロス率が表示されている。ユーザは、ユーザ提示情報表示欄G110に表示されているクロス率も確認することで、クロス分析を考慮した場合に、どのマスキング対象項目の階層を上げればよいかを知ることができる。
- [0138] 次に、UI提供部102は、マスキング対象項目に対する階層の選択操作を受け付ける（ステップS304）。
- [0139] 次に、算出部101は、上記のステップS302と同様に、集合毎のレコード数Nと、N毎のレコードの割合と、クロス率とを算出する（ステップS305）。ここで、ステップS305では、算出部101は、各マスキング対象項目の選択階層での集合毎のレコード数N、N毎のレコードの割合及びクロス率と、1つのマスキング対象項目のみ階層を上げた場合における集合毎のレコード数N、N毎のレコードの割合及びクロス率とを算出する。なお、クロス率の算出方法については後述する。
- [0140] 次に、UI提供部102は、ユーザ提示画面を更新して、上記のステップS305で算出されたN毎のレコードの割合とクロス率とが含まれるユーザ提示画面を表示する（ステップS306）。
- [0141] 次に、UI提供部102は、図7のステップS106と同様に、マスキング対象項目の階層選択を終了するか否かを判定する（ステップS307）。
- [0142] ステップS307でマスキング対象項目の階層選択を終了すると判定されなかった場合、データ加工処理部100は、ステップS304に戻る。これにより、マスキング対象項目の階層選択が終了するまで、上記のステップS304～ステップS306が繰り返し実行される。
- [0143] 一方、ステップS306でマスキング対象項目の階層選択を終了すると判定された場合、データ加工部103は、図7のステップS107と同様に、同一集合に属するレコード数Nがk未満のレコードを削除すると共に、Nがk以上である各レコードを同一集合内で統計加工する（ステップS308）

。これにより、k-匿名性を有するレコードが作成され、これらのレコードで構成される統計加工後データが得られる。

[0144] (クロス率の算出方法)

ここで、上記のステップS302及びステップS305におけるクロス率の算出方法について説明する。以降では、単に「マスタデータ」と記載した場合には、マスタデータ取得部106により取得されたマスタデータを構成する各レコードのうち、対象データを構成する各レコードとの間で共通の項目が1つも含まれないレコードを削除したデータを指すものとする。

[0145] なお、クロス分析では、2つの分析対象項目を設定する必要がある。例えば、分析対象項目を「業種」及び「商品種別」と設定する等である。この場合、クロス分析では、例えば、同じ商品種別の商品が、複数の業種の業者から購入されていることが確認可能となるまで分析対象項目の項目値が抽象化されている必要がある。このため、対象データをクロス分析に用いる場合には、対象データのマスクキング対象項目の階層が低い（つまり、中抽象度が低い）方が良いとは必ずしも限らず、クロス率が低い場合には階層を高く（つまり、抽象度を高く）した方が良いこともある。

[0146] 一般に、クロス分析の分析対象項目の設定する際には、以下の2つのパターンが考えられる。

[0147] (パターン1) 1つのデータ（対象データ、マスタデータ、又は対象データとマスタデータとを統合したデータ）内に分析対象項目が2つとも存在する場合

例えば、分析対象項目が「業種」及び「商品種別」として、1つのデータを構成する各レコードには項目「業種」と項目「商品種別」とが含まれる場合である。

[0148] (パターン2) 分析対象項目の1つがデータ（対象データ、マスタデータ）で決まる場合

例えば、分析対象項目が「業種」及び「商品種別」として、対象データが「A社の購買データ」、マスタデータが「B社の購買データ」であり

、対象データ及びマスタデータをそれぞれ構成する各レコードには項目「商品種別」が含まれる場合である。なお、この場合は、例えば、対象データを構成する各レコードに対して項目「業種」及び項目値「A社」を追加すると共に、マスタデータを構成する各レコードに対して項目「業種」及び項目値「B社」を追加することで、パターン1と同様に扱うことが可能となる。

[0149] ・クロス率の算出方法（その1）

一例として、図15に示す対象データ及びマスタデータを用いて、クロス率の算出方法（その1）を説明する。図15に示す対象データ及びマスタデータをそれぞれ構成する各レコードには項目「商品種別」が共通に含まれており、この項目「商品種別」がマスキング対象項目であるものとする。すなわち、1つの分析対象項目が「商品種別」、もう1つの分析対象項目が対象データ及びマスタデータで決まる場合（上記のパターン2）のクロス率の算出方法を説明する。以降では、対象データを構成する各レコードと、マスタデータを構成する各レコードとの間で共通に含まれるマスキング対象項目を「共通項目」と表す。また、対象データを構成する各レコードと、マスタデータを構成する各レコードとの間で、共通項目中の同一情報（同一項目値）を「共通値」と表す。図15に示す例では、共通項目「商品種別」中の共通項目値は、「チョコレート」及び「飴」である。

[0150] クロス率の算出方法（その1）では、以下の（式1）によりクロス率を算出する。

[0151] $\text{クロス率} = (\text{該当の階層における共通値の個数}) / (\text{該当の階層における対象データの共通項目中で異なる情報（項目値）の個数}) \times 100 \dots$
（式1）

例えば、図15に示す対象データ及びマスタデータが既に該当の階層でマスキング済みであるとすれば、上記の（式1）に示す定義の分数部分の分子については、共通値は「チョコレート」及び「飴」であるため、「2」となる。一方で、分母については、対象データの共通項目中で異なる項目値は「チョコレート」、「飴」及び「扇風機」であるため、「3」となる。したが

って、上記の（式１）に示す定義では、クロス率＝ $2/3 \times 100 = \text{約} 66$ （％）と算出される。

[0152] なお、上記の（式１）に示す定義の分数部分の分母は、「該当の階層におけるマスタデータの共通項目中で異なる情報（項目値）の個数」としてもよいし、「該当の階層における対象データ及びマスタデータの和集合で表されるデータの共通項目中で異なる情報（項目値）の個数」としてもよい。なお、該当の階層における対象データ及びマスタデータの和集合で表されるデータとは、該当の階層で、対象データ及びマスタデータをマージすることで得られるデータのことである。

[0153] また、上記の（式１）に示す定義の代わりに、以下の（式２）に示す定義によりクロス率が算出されてもよい。

[0154] クロス率＝（該当の階層における対象データで共通値を持つレコード数）
／（対象データのレコード数） $\times 100$ ・・・（式２）

この場合、上記の（式２）に示す定義の分数部分の分子については「３」、分母部分については「４」であるため、クロス率＝ $3/4 \times 100 = 75$ （％）と算出される。

[0155] 更に、上記の（式２）に示す定義の代わりに、以下の（式３）又は（式４）を用いてクロス率が算出されてもよい。

[0156] クロス率＝（該当の階層におけるマスタデータで共通値を持つレコード数）
／（マスタデータのレコード数） $\times 100$ ・・・（式３）

この場合、上記の（式３）に示す定義の分数部分の分子については「３」、分母部分については「５」であるため、クロス率＝ $3/5 \times 100 = 60$ （％）と算出される。

[0157] クロス率＝（該当の階層における対象データ及びマスタデータの和集合で表されるデータで共通値を持つレコード数）
／（該当の階層における対象データ及びマスタデータの和集合で表されるデータのレコード数） $\times 100$
・・・（式４）

この場合、上記の（式４）に示す定義の分数部分の分子については「７」

、分母部分については「9」であるため、クロス率＝ $7 / 9 \times 100 \div 77$ (%)と算出される。

[0158] ・クロス率の算出方法（その2）

一例として、図16に示す対象データ及びマスタデータを用いて、クロス率の算出方法（その2）を説明する。図16に示す対象データ及びマスタデータをそれぞれ構成する各レコードには、共通項目「商品種別」と「業種」が含まれている。すなわち、2つの分析対象項目「商品種別」及び「業種」が対象データ及びマスタデータに含まれる場合（上記のパターン1）のクロス率の算出方法を説明する。なお、これらの項目「商品種別」及び「業種」はマスキング対象項目である。

[0159] このとき、図16に示すように、算出部101は、該当の階層において、対象データとマスタデータとを或る共通項目で集計処理して、集計データを作成する。図16に示す例では、共通項目「商品種別」で集計処理して、集計データを作成した場合を示している。なお、ヒット数とは、対象データ及びマスタデータで、同一商品種別であるレコード数の合計である。

[0160] そして、クロス率の算出方法（その2）では、以下の（式5）又は（式6）によりクロス率を算出する。

[0161] $\text{クロス率} = (\text{集計データにおいて、特定の項目の項目値が所定の値以上のレコード数}) / (\text{集計データを構成するレコード数}) \times 100 \dots (\text{式5})$

$\text{クロス率} = (\text{集計データにおいて、特定の項目の項目値が所定の値以上のヒット数の合計}) / (\text{集計データを構成する各レコードのヒット数の合計}) \times 100 \dots (\text{式6})$

例えば、特定の項目を「業種数」、所定の値を「3」とした場合、上記の（式5）に示す定義では、クロス率＝ $1 / 3 \times 100 \div 33$ (%)と算出される。一方で、上記の（式6）に示す定義では、クロス率＝ $4 / 8 \times 100 = 50$ (%)と算出される。なお、集計データを構成する各レコードの項目のうちのどの項目を特定の項目とするかは、例えば、ユーザ等によって予め

設定される。同様に、所定の値についても、例えば、ユーザ等によって予め設定される。

[0162] (クロス率の他の算出方法)

ここで、統計加工によってNがk未満のレコードは対象データから削除されるため、統計加工の前後でクロス率が変わる可能性がある。このため、統計加工後のクロス率（つまり、統計加工後データをデータ分析装置20に送信（アップロード）した後のクロス率）を確認したい場合もある。

[0163] そこで、統計加工後のクロス率の算出方法として、以下の（式7）又は（式8）のいずれかが用いられてもよい。なお、以降では、該当の階層において、対象データを構成する各レコードのうち、Nがk以上のレコード（すなわち、Nがk未満のレコードを除外した対象データ）と、マスタデータを構成する各レコードとを或る共通項目で集計処理して作成された集計データを「除外集計データ」と表す。

[0164] $\text{クロス率} = (\text{除外集計データにおいて、特定の項目の項目値が所定の値以上のレコード数}) / (\text{除外集計データを構成するレコード数}) \times 100 \quad \dots$
（式7）

$\text{クロス率} = (\text{除外集計データにおいて、特定の項目の項目値が所定の値以上のレコードのヒット数}) / (\text{除外集計データを構成する各レコードのヒット数の合計}) \times 100 \quad \dots$ （式8）

[0165] また、マスタデータを考慮せずに、クロス率が算出されてもよい。この場合は、クロス率の他の算出方法として、以下の（式9）又は（式10）のいずれかが用いられてもよい。

[0166] $\text{クロス率} = (\text{該当の階層における対象データにおいて、特定の項目の項目値が所定の値以上のレコード数}) / (\text{該当の階層における対象データを構成するレコード数}) \times 100 \quad \dots$ （式9）

$\text{クロス率} = (\text{該当の階層における対象データにおいて、特定の項目の項目値が所定の値以上の項目値の個数}) / (\text{該当の階層における対象データにおいて、特定の項目の項目値の個数}) \times 100 \quad \dots$ （式10）

[0167] (他の指標値)

ここで、本実施例において、クロス率に代えて又はクロス率と共に、指標値の1つとして損失率が用いられてもよい。ユーザは、例えば、ユーザ提示画面に表示された損失率を確認することで、損失率も考慮して、マスキング対象項目の階層を選択することができるようになる。損失率とは、対象データとマスタデータとを統合した後に行う分析（例えば、クロス分析）において、削除されるレコード又はカテゴリの粒度が合わないために使用できないレコードの割合を表す指標値のことである。

[0168] ・マスタデータの損失率

マスタデータの損失率とは、マスタデータを構成するレコードのうち、クロス率の算出に用いることができないレコードの割合のことである。マスタデータの損失率は、マスキング対象項目毎に、以下の（式11）により算出される。

[0169] マスタデータの損失率＝（マスタデータを構成する各レコードのうち、対象データを構成する各レコードとの間で共通の項目値が1つも含まれないレコードの数）／（マスタデータを構成するレコード数）×100・・・（式11）

[0170] なお、前記した「クロス率の算出に用いることができないレコード」は、「マスタデータの項目値の粒度が対象データの項目値の粒度と合わないために、クロス分析に用いることができないレコード」でもある。例えば、マスタデータの項目「住所」が第3階層の粒度であるレコードが80%、第4階層であるレコードが20%であり、対象データの項目「住所」を第3階層で匿名化した上で、マスタデータと、匿名化後の対象データとを統合したデータを用いた分析（クロス分析等）を行う場合を考える。このとき、マスタデータに由来する20%のレコードは第4階層の情報しか持たない。よって、統合後のデータの「住所」の第3階層の情報をを用いた分析において、前記した20%のレコードは分析に用いることができない。

[0171] ・対象データの損失率

対象データの損失率とは、対象データを構成するレコードのうち、データ加工によって削除されるレコードの割合のことである。対象データの損失率は、以下の（式１２）又は（式１３）により算出される。

[0172] 対象データの損失率＝（該当の階層において、対象データを構成する各レコードのうち、Nがk未満であるレコードの数）／（対象データを構成するレコード数）×１００・・・（式１２）

対象データの損失率＝（該当の階層において、対象データを構成する各レコードの該当のマスキング対象項目の項目値のうち、Nがk未満のレコードの当該項目値の個数）／（該当の階層において、対象データを構成する各レコードの該当のマスキング対象項目の項目値の個数）×１００・・・（式１３）

[0173] 本実施例により指標値を算出することで、対象データだけではなく、統計加工後データをデータ分析装置２０に送信（アップロード）した後の分析も考慮した指標値をユーザに提示することが可能となる。これにより、ユーザは、例えば、最終的な分析（例えば、クロス分析）の際に使用することができないレコード数を最低限に抑えたり、階層を可能な限り低く保ったりしながら、対象データの匿名化を行うことが可能となる。

[0174] [実施例４]

次に、実施例４として、データ提供端末１０で対象データを統計加工によって匿名化する際に、指標値の１つであるクロス率を算出すると共に自動的に適切な匿名化粒度を決定する場合について説明する。なお、実施例４では、実施例２や実施例３と同一の構成要素についてはその説明を省略する。

[0175] （データ加工処理部１００の機能構成）

まず、実施例４におけるデータ加工処理部１００の機能構成について、図１７を参照しながら説明する。図１７は、本発明の実施の形態におけるデータ加工処理部１００の機能構成の一例を示す図（実施例４）である。

[0176] 図１７に示すように、実施例４におけるデータ加工処理部１００には、算出部１０１と、データ加工部１０３と、選択部１０４と、終了条件判定部１

05と、マスタデータ取得部106とが含まれる。また、実施例4におけるデータ加工処理部100には、UI提供部102が含まれていてもよいし、UI提供部102が含まれていなくてもよい。なお、これら各部の機能は実施例2や実施例3と同様であるため、その説明を省略する。ただし、実施例4における選択部は、更に、クロス率等の指標値にも基づいて、各マスキング対象項目の階層を選択する。

[0177] (データ加工処理)

次に、データ提供端末10で対象データを統計加工して、匿名化(k-匿名化)する際に、クロス率も算出するデータ加工処理について、図18を参照しながら説明する。図18は、本発明の実施の形態におけるデータ加工処理の一例を示すフローチャート(実施例4)である。

[0178] まず、マスタデータ取得部106は、図13のステップS301と同様に、データ分析装置20のマスタデータ記憶部400に記憶されているマスタデータを取得する(ステップS401)。

[0179] 次に、算出部101は、図13のステップS302と同様に、予め設定されたマスキング対象項目と、分類辞書記憶部200に記憶されている分類辞書と、各マスキング対象項目の階層と、対象データを構成するレコード数とに基づいて、対象データを構成する各レコードを分類した場合に同一集合に属するレコードの数N(つまり、集合毎のレコード数N)と、N毎のレコードの割合と、クロス率とを算出する(ステップS402)。

[0180] 次に、選択部104は、算出部101による算出結果と、マスキング対象項目の優先度と、クロス率等の指標値とに基づいて、各マスキング対象項目の階層を選択する(ステップS403)。ここで、選択部104は、例えば、図11のステップS202における(選択条件1)及び(選択条件2)に代えて、以下の(選択条件1')及び(選択条件2')により各マスキング対象項目の階層を選択すればよい。

[0181] (選択条件1') 階層を1つ上げることでN毎のレコードの割合が向上し、かつ、クロス率も高くなるマスキング対象項目が存在する場合には、当該

マスキング対象項目の1つ上の階層を選択する。

[0182] (選択条件2') 階層を1つ上げることでN毎のレコードの割合が向上し、かつ、クロス率も高くなるマスキング対象項目が存在しない場合には、最も優先度が低いマスキング対象項目の1つ上の階層を選択する。

[0183] 次に、算出部101は、図13のステップS305と同様に、集合毎のレコード数Nと、N毎のレコードの割合と、クロス率とを算出する(ステップS404)。

[0184] 次に、終了条件判定部105は、図11のステップS204と同様に、所定の終了条件を満たしたか否かを判定する(ステップS405)。

[0185] ステップS405で終了条件を満たすと判定されなかった場合、データ加工処理部100は、ステップS403に戻る。これにより、終了条件を満たすまで、上記のステップS403～ステップS404が繰り返し実行される。なお、例えば、UI提供部102は、適宜、ユーザ提示画面を表示して、マスキング対象項目の階層をユーザに選択させるようにしてもよい。

[0186] 一方、ステップS405で終了条件を満たすと判定された場合、データ加工部103は、図13のステップS308と同様に、同一集合に属するレコード数Nがk未満のレコードを削除すると共に、Nがk以上である各レコードを同一集合内で統計加工する(ステップS406)。これにより、k-匿名性を有するレコードが作成され、これらのレコードで構成される統計加工後データが得られる。

[0187] なお、本実施例でも、実施例3と同様に、クロス率に代えて又はクロス率と共に、指標値の1つとして損失率が算出されてもよい。損失率が算出された場合には、上記のステップS403では、選択部104は、損失率にも基づいて、各マスキング対象項目の階層を選択する。

[0188] [実施例5]

次に、実施例5として、対象データとマスタデータの全部又は一部とをマージしたデータをデータ加工する場合について説明する。ここで、例えば、比較的小規模な小売店等の商業施設では十分なレコード数の対象データを準

備することができない場合がある。レコード数が少ない場合には、マスキング対象項目の階層を高くしないとNがk未満となるレコード数が多くなってしまう。したがって、マスキング対象項目の階層を比較的低くした場合には、対象データ中の多くのレコードが削除され、統計加工後データに含まれるレコードが少なくなってしまう、データ分析の精度（正確さ）が低下してしまう。一方で、マスキング対象項目の階層を比較的高くした場合には、統計加工後データには多くのレコードを残すことができるものの、マスキング対象項目の情報の抽象度が上がってしまい、データ分析の精度（詳細さ）が低下してしまう。

[0189] そこで、実施例5では、対象データとマスタデータの全部又は一部とをマージしたデータをデータ加工することで、対象データ中のレコード数が少ない場合であっても、削除されるレコード数を減らすことでデータ分析の精度（正確さ及び詳細さ）の低下を防止する。なお、実施例5では、実施例1や実施例3と同一の構成要素についてはその説明を省略する。

[0190] （データ加工処理部100の機能構成）

まず、実施例5におけるデータ加工処理部100の機能構成について、図19を参照しながら説明する。図19は、本発明の実施の形態におけるデータ加工処理部100の機能構成の一例を示す図（実施例5）である。

[0191] 図19に示すように、実施例5におけるデータ加工処理部100には、算出部101と、UI提供部102と、データ加工部103と、マスタデータ取得部106と、マージ部107とが含まれる。なお、実施例5におけるデータ加工処理部100には、UI提供部102が含まれていなくてもよい。

[0192] マージ部107は、マスタデータ取得部106により取得されたマスタデータと、対象データとをマージしたデータを作成する。

[0193] また、実施例5における算出部101は、マージ部107により作成されたデータ（つまり、マスタデータと対象データとをマージしたデータ）を用いて、このデータを構成する各レコードを分類して、これら各レコードが分類された集合毎に、同一集合に属するレコードの数Nを算出する。そして、

算出部 101 は、N 毎に、N が同一であるレコードの割合を算出する。言い換えれば、実施例 5 における算出部 101 は、実施例 1 の「対象データ」の代わりに、「マスタデータと対象データとをマージしたデータ」を用いて、N 毎に、N が同一であるレコードの割合を算出する。

[0194] (データ加工処理)

次に、マスタデータと対象データとをマージしたデータ（以降、「マージ対象データ」とも表す。）を作成した上で、データ提供端末 10 でマージ対象データを統計加工して、匿名化（k-匿名化）するデータ加工処理について、図 20 を参照しながら説明する。図 20 は、本発明の実施の形態におけるデータ加工処理の一例を示すフローチャート（実施例 5）である。

[0195] まず、マスタデータ取得部 106 は、データ分析装置 20 のマスタデータ記憶部 400 に記憶されているマスタデータを取得する（ステップ S501）。ここで、マスタデータ取得部 106 は、マスタデータ記憶部 400 に記憶されているマスタデータを構成する各レコードの全部を取得してもよいし、一部のレコードのみを取得してもよい。なお、マスタデータの全レコードを取得する場合に、これらのレコードの中に不足する項目（つまり、対象データを構成するレコード中には含まれる一方で、マスタデータを構成するレコード中には含まれない項目）が存在するときには、当該項目に任意の値を代入してもよい。これは、後述する「統計量の減算処理」のステップ S602 において、当該項目の項目値が、統計加工後データを構成する各レコードの当該項目の統計量から減算されるため、最終的な統計量には影響を与えないためである。

[0196] 一部のレコードのみを取得する場合は、マスタデータ取得部 106 は、例えば、取得条件を指定した取得要求をデータ分析装置 20 に送信すればよい。これにより、例えば、データ分析処理部 300 によってマスタデータ記憶部 400 が検索され、取得条件を満たすレコードで構成されるマスタデータがデータ提供端末 10 に返信される。

[0197] このような取得条件としては、例えば、マスキング対象項目の項目値を指

定すればよい。例えば、マスキング対象項目が項目「住所」及び項目「年代」である場合、取得条件としては、『住所＝「東京都武蔵野市緑町」、かつ、年代＝「１０代」』等とすればよい。又は、例えば、マスキング対象項目が項目「住所」、項目「年代」及び項目「業種」である場合、取得条件としては、『住所＝「東京都武蔵野市緑町」、かつ、年代＝「１０代」、かつ、業種＝「電気店」』等とすればよい。これら以外にも、取得条件として、例えば、マスキング対象項目の項目名のみが指定されてもよい。このような取得条件は、例えば、マージ対象データの損失率（つまり、マージ対象データを構成するレコードのうち、データ加工によって削除されるレコードの割合）が所望の値よりも小さくなるようにユーザによって決定される。

[0198] 次に、マージ部１０７は、上記のステップＳ５０１で取得されたマスタデータと、対象データとをマージしたマージ対象データを作成する（ステップＳ５０２）。

[0199] 次に、データ加工処理部１００は、実施例１又は実施例２で「対象データ」の代わりに「マージ対象データ」を用いたデータ加工処理を行う（ステップＳ５０３）。これにより、マージ対象データから統計加工後データが作成され、データ分析装置２０に送信される。

[0200] （統計量の減算処理）

ここで、上記の統計加工後データを構成する各レコードの統計量（例えば、金額の合計や購入個数の合計、購入者数の合計等）の算出には、上記のステップＳ５０１で取得されたマスタデータに含まれるレコードの情報も用いられている。このため、統計加工後データをマスタデータ記憶部４００に記憶させる前に、当該統計加工後データを構成する各レコードの統計量を減算する必要がある。そこで、この統計量の減算処理について、図２１を参照しながら説明する。図２１は、本発明の実施の形態における統計量の減算処理の一例を示すフローチャート（実施例５）である。

[0201] まず、データ分析処理部３００は、データ提供端末１０から統計加工後データを受信する（ステップＳ６０１）。

[0202] 次に、データ分析処理部300は、当該統計加工後データを構成する各レコードの統計量から、当該データ提供端末10に送信したマスタデータの該当のレコードの項目値を減算する（ステップS602）。

[0203] 例えば、統計加工後データに含まれる或るレコードの統計量が合計金額であり、この合計金額が、対象データのレコードA、レコードB及びレコードCと、マスタデータのレコードD及びレコードEとで、項目「購入金額」の項目値を合計したものとする。この場合、当該合計金額から、レコードDの項目「購入金額」の項目値と、レコードEの項目「購入金額」の項目値とを減算する。これにより、統計加工後データを構成する各レコードの統計量を、対象データを構成する各レコードから算出される統計量と一致させることができる。

[0204] [実施例6]

次に、実施例6として、対象データを構成する各レコードのマスキング対象項目のうちの一部のマスキング対象項目を削除したデータをデータ加工する場合について説明する。実施例5と同様に、例えば、比較的小規模な小売店等の商業施設のように十分なレコード数の対象データを準備することができない場合に、一部のマスキング対象項目を削除することで、データ分析の精度（正確さ）の低下を防止することができる。なお、実施例6では、実施例1や実施例3と同一の構成要素についてはその説明を省略する。

[0205] （データ加工処理部100の機能構成）

まず、実施例6におけるデータ加工処理部100の機能構成について、図22を参照しながら説明する。図22は、本発明の実施の形態におけるデータ加工処理部100の機能構成の一例を示す図（実施例6）である。

[0206] 図22に示すように、実施例6におけるデータ加工処理部100には、算出部101と、UI提供部102と、データ加工部103と、項目削除部108とが含まれる。なお、実施例6におけるデータ加工処理部100には、UI提供部102が含まれていなくてもよい。

[0207] 項目削除部108は、対象データを構成する各レコードのマスキング対象

項目のうち、一部のマスキング対象項目を削除したデータを作成する。

[0208] また、実施例6における算出部101は、項目削除部108により作成されたデータ（つまり、対象データを構成する各レコードのマスキング対象項目のうち、一部のマスキング対象項目を削除したデータ）を用いて、このデータを構成する各レコードを分類して、これら各レコードが分類された集合毎に、同一集合に属するレコードの数Nを算出する。そして、算出部101は、N毎に、Nが同一であるレコードの割合を算出する。言い換えれば、実施例6における算出部101は、実施例1の「対象データ」の代わりに、「対象データの一部のマスキング対象項目を削除したデータ」を用いて、N毎に、Nが同一であるレコードの割合を算出する。

[0209] （データ加工処理）

次に、対象データを構成する各レコードから一部のマスキング対象項目を削除したデータ（以降では、「項目削除後データ」とも表す。）を作成した上で、データ提供端末10で項目削除後データを統計加工して、匿名化（k-匿名化）するデータ加工処理について、図23を参照しながら説明する。図23は、本発明の実施の形態におけるデータ加工処理の一例を示すフローチャート（実施例6）である。なお、以降では、一例として、マスキング対象項目は「住所」及び「年代」であるものとする。

[0210] まず、項目削除部108は、対象データを構成する各レコードのマスキング対象項目のうち、一部のマスキング対象項目を削除して、項目削除後対象データを作成する（ステップS701）。例えば、図24に示す対象データから一部のマスキング対象項目を削除する場合、項目削除部108は、図24に示す対象データの項目「年代」を削除した年代削除後対象データを項目削除後対象データとして作成してもよいし、項目「住所」を削除した住所削除後対象データを項目削除後対象データとして作成してもよい。又は、項目削除部108は、年代削除後対象データと住所削除後対象データとの両方を項目削除後対象データとして作成してもよい（すなわち、項目削除部108は、複数の項目削除後対象データを作成してもよい。）。複数の項目削除後

対象データを作成することを、「対象データを分割する」と称されてもよい。

[0211] 次に、データ加工処理部 100 は、実施例 1 又は実施例 2 で「対象データ」の代わりに「項目削除後対象データ」を用いたデータ加工処理を行う（ステップ S 702）。これにより、項目削除後対象データから統計加工後データが作成され、データ分析装置 20 に送信される。なお、上記のステップ S 702 で複数の項目削除後対象データが作成された場合には、各項目削除後対象データをそれぞれ用いて、実施例 1 又は実施例 2 のデータ加工処理を行えばよい。

[0212] なお、本実施例では、対象データを構成するレコード数が少ないことを前提として、対象データから項目削除後対象データを作成したが、対象データを構成するレコード数の多寡に限られず、項目削除後対象データを作成してもよい。例えば、ユーザ提示画面 G 100 上で、対象データから一部のマスキング対象項目を削除するか否かや対象データを分割するか否かをユーザに問い合わせた上で、この問い合わせに対して、削除操作や分割操作が行われた場合に項目削除後対象データを作成してもよい。特に、このような問い合わせは、マスキング対象項目の階層を上げて、N 毎のレコードの割合が向上したり、所定の指標値が向上したりしない場合に行われるようにしてもよい。

[0213] [実施例 7]

次に、実施例 7 として、分類辞書記憶部 200 に記憶されている分類辞書を修正する場合について説明する。ここで、上述したように、分類辞書は、対象データを構成するレコードのマスキング項目毎に、カテゴリの木構造で表現される。しかしながら、カテゴリの粒度が粗すぎたり、又はカテゴリの粒度が細かすぎたりする場合がある。このような場合、例えば、対象データ中の多くのレコードが削除されたり、マスキング対象項目の情報の抽象度が上がったりしてしまい、データ分析の精度（正確さ又は詳細さ）が低下してしまう。

[0214] そこで、実施例 7 では、分類辞書を修正可能とする場合について説明する。これにより、ユーザは、分類辞書を適切に修正することで、データ分析の精度（正確さ及び詳細さ）の低下を防止することができるようになる。なお、実施例 5 では、実施例 1 と同一の構成要素についてはその説明を省略する。

[0215] （データ加工処理部 100 の機能構成）

まず、実施例 7 におけるデータ加工処理部 100 の機能構成について、図 25 を参照しながら説明する。図 25 は、本発明の実施の形態におけるデータ加工処理部の機能構成の一例を示す図（実施例 7）である。

[0216] 図 25 に示すように、実施例 7 におけるデータ加工処理部 100 には、算出部 101 と、UI 提供部 102 と、データ加工部 103 と、分類修正部 109 とが含まれる。

[0217] 分類修正部 109 は、ユーザの操作に応じて、分類辞書記憶部 200 に記憶されている分類辞書を修正する。ここで、分類辞書の修正とは、木構造で表現される分類辞書に対してカテゴリを追加したり、分類辞書からカテゴリを削除したり、分類辞書のカテゴリ自体を変更したりすることである。

[0218] また、実施例 7 における算出部 101 は、更に、分類辞書記憶部 200 に記憶されている分類辞書と、マスキング対象項目と、対象データとに基づいて、指標値の 1 つである集約率を算出する。集約率とは、分類辞書によってマスキング対象項目の項目値をマスキングした場合に、同一集合に分類されるレコード数を表す指標値である。ユーザは、集約率を参考にして、分類辞書を修正するか否かやどのような修正を行ったらよいかを判断することができる。

[0219] ここで、低すぎる集約率は、カテゴリの粒度が細かすぎて、対象データ中のレコードがまとまっていない（つまり、各レコードがバラバラになっている）ことを表す。一方で、高すぎる集約率は、カテゴリの粒度が粗すぎて、対象データ中のレコードがまとまりすぎていることを表す。また、例えば、或る項目の分類辞書で階層毎に集約率を算出した場合、これらの集約率は、

階層が上がるに従ってなだらかに上昇していくことが望ましい。例えば、或る階層で集約率が急激に上昇する場合や集約率の上昇がほとんど無い場合、集約率が最初から高い場合等は、当該項目の各階層のカテゴリ（の粒度）が適切でないことを表す。したがって、階層毎の集約率をUI上に表示し、可視化することで、ユーザは、例えば、集約率の上昇度合い等を把握することができるようになる。また、このとき、集約率を参考にして分類辞書の修正を行ったり、修正後の分類辞書を用いた集約率を確認したりすることで、ユーザは、分類辞書の編集を容易に行うことが可能となる。

[0220] (集約率の算出方法)

集約率は以下の（式14）により算出される。

[0221] 集約率 = (該当の階層よりも1つ下の階層で対象データを構成する各レコードの該当の項目の項目値が属するカテゴリ数 - 該当の階層で対象データを構成する各レコードの該当の項目の項目値が属するカテゴリ数) / (該当の階層よりも1つ下の階層で対象データを構成する各レコードの該当の項目の項目値が属するカテゴリ数) × 100 …… (式14)

上記の（式14）の代わりに、以下の（式15）により集約率が算出されてもよい。

[0222] 集約率 = (該当の階層よりも1つ下の階層における該当の項目のカテゴリ数 - 該当の階層における該当の項目のカテゴリ数) / (該当の階層よりも1つ下の階層における該当の項目のカテゴリ数) × 100 …… (式15)

[0223] (分類辞書の修正)

ここで、一例として、図26Aに示す対象データを用いて、図26Bに示す分類辞書の修正を行う場合について説明する。

[0224] マスキング対象項目を、項目「レコードID」以外の全項目、 $k = 1$ 、集約率を算出する対象の階層を「第2階層」とした場合に、図26Aに示す対象データのマスキング対象項目「日時」の集約率を上記の（式14）により算出すると、80（%）となる。すなわち、図26Aに示す対象データを構成する各レコードは、マスキング対象項目「日時」の「第2階層」では1つ

のカテゴリ「17時」に属する。一方で、「第1階層」では、レコードID「1」及び「2」がカテゴリ「8時」、レコードID「3」がカテゴリ「9時」、レコードID「4」がカテゴリ「11時」、レコードID「5」がカテゴリ「17時」、レコードID「6」が「20時」の計5つのカテゴリにそれぞれ属する。したがって、集約率は、 $(5 - 1) / 5 \times 100 = 80$ (%)と算出される。なお、上記の(式15)により集約率を算出した場合、約96 (%)となる。

[0225] 集約率が高い場合には対象データに含まれる多くのレコードを1つのレコードに集約して匿名化を図ることができる一方で、情報の損失が多くなる。例えば、図26Bに示す分類辞書を用いて、図26Aに示す対象データのマスキング対象項目「日時」を「第2階層」とした場合、マスキング対象項目「日時」の項目値のうちの時刻情報(8時や9時、11時、17時、20時等)が失われてしまう。

[0226] そこで、集約率が高すぎるような場合には、ユーザは、分類辞書に対して階層を追加することで、集約率を下げて、情報の損失を抑えることが可能となる。例えば、図27Aに示すように、図27Bに示す分類辞書に対して、「第2階層」以上の階層を1つ上の階層とした上で、新たな「第2階層」としてカテゴリ「午前」、「午後」を追加することで、上記の(式14)により算出されるマスキング対象項目「日時」の「第2階層」における集約率を60 (%)に下げることができる。すなわち、図26Aに示す対象データを構成する各レコードは、マスキング対象項目「日時」の「第2階層」では「午前」及び「午後」の2つのカテゴリに属する。一方で、「第1階層」では、レコードID「1」及び「2」がカテゴリ「8時」、レコードID「3」がカテゴリ「9時」、レコードID「4」がカテゴリ「11時」、レコードID「5」がカテゴリ「17時」、レコードID「6」が「20時」の計5つのカテゴリにそれぞれ属する。したがって、集約率は、 $(5 - 2) / 5 \times 100 = 60$ (%)と算出される。なお、上記の(式15)により集約率を算出した場合、約92 (%)となる。

[0227] これにより、集約率を下げることができ、情報の損失を抑えることが可能となる。例えば、図27Bに示す対象データでは、マスキング対象項目「日時」の時刻情報として、午前又は午後を残すことができる。したがって、データ分析装置20におけるデータ分析の精度低下を抑えることが可能となる。

[0228] なお、上記では、集約率が高すぎる場合に分類辞書に階層を追加する場合を説明したが、例えば、集約率が低すぎるような場合には分類辞書から階層を削除してもよい。また、既存の階層にカテゴリを追加したり、既存の階層のカテゴリ自体を修正したりしてもよい。

[0229] (データ加工処理)

次に、データ提供端末10で対象データを統計加工して、匿名化(k-匿名化)する際に、集約率もユーザに提示し、必要に応じて分類辞書の修正が可能なデータ加工処理について、図28を参照しながら説明する。図28は、本発明の実施の形態におけるデータ加工処理の一例を示すフローチャート(実施例7)である。

[0230] まず、算出部101は、予め設定されたマスキング対象項目と、分類辞書記憶部200に記憶されている分類辞書と、各マスキング対象項目の階層と、対象データを構成するレコード数とに基づいて、対象データを構成する各レコードを分類した場合に同一集合に属するレコードの数N(つまり、集合毎のレコード数N)と、N毎のレコードの割合と、集約率とを算出する(ステップS801)。なお、集合毎のレコード数N及びN毎のレコードの割合は実施例1と同様である。また、集約率についても、各マスキング対象項目は「第1階層」が選択されているものとして、上記の(式14)又は(式15)により集約率を算出する。なお、集約率の定義から、「第1階層」の集約率は算出されない。

[0231] 次に、UI提供部102は、上記のステップS801で算出されたN毎のレコードの割合と集約率とが含まれるユーザ提示画面を表示する(ステップS802)。すなわち、UI提供部102は、例えば、図29に示すユーザ

提示画面G100を表示する。

[0232] 図29に示すユーザ提示画面G100のユーザ提示情報表示欄G110には、N毎のレコードの割合に加えて、マスキング対象項目の階層を変化させた場合における集約率が表示されている。ユーザは、ユーザ提示情報表示欄G110に表示されている集約率を確認することで、分類辞書の修正をするか否かを判断することができる。ここで、図29に示すユーザ提示画面G100には、「分類辞書を修正」ボタンG130が含まれる。ユーザは、分類辞書の修正が必要と判断した場合には「分類辞書を修正」ボタンG130を押下して分類辞書の修正開始操作を行うことで、図29に示す分類辞書の修正画面G200を表示させることができる。以降では、ユーザは、マスキング対象項目に対する階層の選択操作又は分類辞書の修正開始操作のいずれかを行ったものとして説明を続ける。

[0233] 次に、UI提供部102は、階層の選択操作又は分類辞書の修正開始操作のいずれを受け付けたかを判定する（ステップS803）。

[0234] ステップS803で分類辞書の修正開始操作を受け付けたと判定した場合、UI提供部102は、例えば、図29に示す分類辞書の修正画面G200を表示する（ステップS804）。

[0235] 図29に示す分類辞書の修正画面G200は、分類辞書を修正するための画面である。なお、図29に示す分類辞書の修正画面G200は、例えば、図29に示すユーザ提示画面G100から画面遷移により表示されてもよいし、ポップアップにより表示されてもよい。

[0236] 図29に示す分類辞書の修正画面G200には、例えば、修正対象の分類辞書の項目を選択するためのマスキング対象項目選択欄G210と、修正方法（追加、削除、変更等）を選択するための修正方法選択欄G220と、修正対象の階層を選択するための階層選択欄G230とが含まれる。また、図29に示す分類辞書の修正画面G200には、現在の集約率（例えば、マスキング対象項目選択欄G210及び階層選択欄G230でそれぞれ選択されている項目及び階層の集約率）も表示される。更に、図29に示す分類辞書

の修正画面 G 2 0 0 には、修正方法が「追加」又は「変更」である場合に、追加されるカテゴリの内容や変更後のカテゴリの内容を入力するカテゴリ設定欄 G 2 5 0 が含まれる。

[0237] 加えて、図 2 9 に示す分類辞書の修正画面 G 2 0 0 には、スコア再計算ボタン G 2 7 0 が含まれる。スコア再計算ボタン G 2 7 0 が押下されることで、分類辞書が修正された後における該当の項目及び階層のスコア（例えば、集約率）が計算される。

[0238] ユーザは、マスキング対象項目選択欄 G 2 1 0、修正方法選択欄 G 2 2 0 及び階層選択欄 G 2 3 0 から項目、修正方法及び階層をそれぞれ選択した上で、必要に応じてカテゴリ設定欄 G 2 5 0 にカテゴリの内容を設定して、決定ボタン G 2 6 0 を押下することで、カテゴリ修正操作を行うことができる。カテゴリ修正操作が行われることで、分類修正部 1 0 9 により、当該修正操作で選択及び入力された内容で、分類辞書記憶部 2 0 0 に記憶されている該当の分類辞書が修正される。

[0239] 一方、ステップ S 8 0 3 で階層の選択操作を受け付けたと判定した場合又は上記のステップ S 8 0 4 に続いて、算出部 1 0 1 は、上記のステップ S 8 0 1 と同様に、集合毎のレコード数 N と、N 毎のレコードの割合と、集約率とを算出する（ステップ S 8 0 5）。ここで、ステップ S 8 0 3 では、算出部 1 0 1 は、各マスキング対象項目の選択階層での集合毎のレコード数 N、N 毎のレコードの割合及び集約率と、1 つのマスキング対象項目のみ階層を上げた場合における集合毎のレコード数 N、N 毎のレコードの割合及び集約率とを算出する。なお、このとき、上記のステップ S 8 0 4 で分類辞書が修正された場合には、修正後の分類辞書を用いて、集合毎のレコード数 N と、N 毎のレコードの割合と、集約率とを算出する。

[0240] 次に、UI 提供部 1 0 2 は、ユーザ提示画面を更新して、上記のステップ S 8 0 5 で算出された N 毎のレコードの割合と集約率とが含まれるユーザ提示画面を表示する（ステップ S 8 0 6）。

[0241] 次に、UI 提供部 1 0 2 は、図 7 のステップ S 1 0 6 と同様に、マスキニ

グ対象項目の階層選択を終了するか否かを判定する（ステップS807）。

[0242] ステップS807でマスキング対象項目の階層選択を終了すると判定されなかった場合、データ加工処理部100は、ステップS803に戻る。これにより、マスキング対象項目の階層選択が終了するまで、上記のステップS803～ステップS806が繰り返し実行される。

[0243] 一方、ステップS807でマスキング対象項目の階層選択を終了すると判定された場合、データ加工部103は、図7のステップS107と同様に、同一集合に属するレコード数Nがk未満のレコードを削除すると共に、Nがk以上である各レコードを同一集合内で統計加工する（ステップS808）。これにより、k-匿名性を有するレコードが作成され、これらのレコードで構成される統計加工後データが得られる。

[0244] （他の指標値）

ここで、本実施例において、集約率に代えて又は集約率と共に、指標値の1つとして分離率やカバー率が用いられてもよい。ユーザは、例えば、ユーザ提示画面に表示された分離率やカバー率を確認することで、これらの指標値も考慮して、分類辞書を修正するか否かを判断することができるようになる。

[0245] ・分離率

分離率とは、対象データを構成する各レコードのマスキング対象項目を分類辞書によってマスキングする際の細かさを表す指標値のことである。分離率が大きいほど、Nがk未満であるとしてデータ加工の際に削除され易くなる。分離率は以下の（式16）により算出される。

[0246] 分離率＝（該当の階層における対象データを構成する各レコードの各項目の項目値のうち、同一カテゴリに属する項目値の個数がM個以下の項目値の個数）／（該当の階層における対象データを構成する各レコードの各項目の項目値の個数）×100・・・（式16）

なお、Mとしては、例えば、M＝1やM＝2等とすることが考えられる。

[0247] ・カバー率

カバー率とは、対象データを構成する各レコードのマスキング対象項目を分類辞書によってマスキングする際における項目値が属するカテゴリの分布を表す指標値である。カバー率が低いと、マスタデータを機械学習の学習データとして利用する場合等に誤った学習を引き起しやすい。カバー率は以下の（式１７）により算出される。

[0248] カバー率＝（該当の階層で、対象データを構成する各レコードの各項目の項目値が属するカテゴリ数）／（該当の階層における各項目のカテゴリ数）
×１００　・・・（式１７）

[0249] 本発明は、具体的に開示された上記の実施形態に限定されるものではなく、特許請求の範囲から逸脱することなく、種々の変形や変更が可能である。また、上記の各実施例は、適宜、組み合わせて適用することが可能である。例えば、実施例１や実施例３に対して、実施例５～実施例７のうちの少なくとも１つの実施例を組み合わせることが可能である。同様に、例えば、実施例２や実施例４に対して、実施例５～実施例７のうちの少なくとも１つの実施例を組み合わせることが可能である。

符号の説明

[0250] １ データ加工システム
 １０ データ提供端末
 ２０ データ分析装置
 １００ データ加工処理部
 １０１ 算出部
 １０２ ＵＩ提供部
 １０３ データ加工部
 １０４ 選択部
 １０５ 終了条件判定部
 １０６ マスタデータ取得部
 １０７ マージ部
 １０８ 項目削除部

1 0 9	分類修正部
2 0 0	分類辞書記憶部
3 0 0	データ分析処理部
4 0 0	マスタデータ記憶部

請求の範囲

- [請求項1] 1以上の項目が含まれるレコードで構成されるデータを統計加工によって匿名化する情報処理装置であって、
- 前記項目のうち、マスキングの対象となる項目を示すマスキング対象項目と、前記マスキング対象項目毎に該項目値のカテゴリが木構造で表現された辞書と、前記マスキング対象項目毎に前記木構造で選択された階層を示す選択階層と、前記データに含まれるレコード数とに基づいて、前記データを構成する各レコードを1つ以上の集合に分類して、各集合のレコード数 N と、レコード数 N の集合に属するレコードの割合とを算出する算出手段と、
- 前記マスキング対象項目毎の各階層において、前記レコード数 N 毎に、レコード数 N の集合に属するレコードの割合を U として表示する表示手段と、
- を有することを特徴とする情報処理装置。
- [請求項2] 前記表示手段は、前記選択階層を特定可能な状態で前記 U 上に表示し、
- 前記算出手段は、
- 前記 U において、ユーザにより前記選択階層を変更する操作が行われた場合、前記辞書と、前記マスキング対象項目と、前記変更後の選択階層と、前記データに含まれるレコード数とに基づいて、前記各集合のレコード数 N と、前記レコード数 N の集合に属するレコードの割合とを再算出する、ことを特徴とする請求項1に記載の情報処理装置。
- [請求項3] 前記算出手段は、
- 前記変更前の選択階層に基づいて算出された算出結果を履歴情報として、所定の記憶部に記憶させ、
- 前記表示手段は、
- ユーザにより前記選択階層の変更を変更前に戻す操作が行われた場

合、前記履歴情報として前記記憶部に記憶されている前記算出結果を基に、前記UIの表示内容を変更前の内容に戻すと共に、前記選択階層が変更前の選択階層に戻ったことを特定可能な状態で前記UI上に表示する、ことを特徴とする請求項2に記載の情報処理装置。

[請求項4] 前記表示手段は、

前記レコード数Nの集合に属するレコードの割合が所定の条件を満たす場合に、前記データを分割するか否か又は前記データを構成する各レコードから所定の項目を削除するか否かをユーザに問い合わせるための情報を表示する、ことを特徴とする請求項1乃至3の何れか一項に記載の情報処理装置。

[請求項5] 前記問い合わせに応じて前記データを分割することがユーザにより選択された場合、前記データを複数のデータに分割する分割手段を有し、

前記情報処理装置は、

前記複数のデータの各々について、前記算出手段による算出と、前記表示手段による表示とを行う、ことを特徴とする請求項4に記載の情報処理装置。

[請求項6] 前記算出手段は、

更に、前記マスキング対象項目及び前記階層毎に、同一集合に分類されるレコード数を表す集約率を算出し、

前記表示手段は、

前記集約率もUIとして表示する、ことを特徴とする請求項1乃至5の何れか一項に記載の情報処理装置。

[請求項7] ユーザによって前記辞書を修正するための操作が行われた場合、該操作に応じて、前記辞書に対するカテゴリの追加と、前記辞書からのカテゴリの削除と、前記辞書のカテゴリの変更とを含む前記辞書の修正を行う辞書修正手段を有する、ことを特徴とする請求項1乃至6の何れか一項に記載の情報処理装置。

[請求項8] 前記算出手段は、

前記辞書の修正が行われた場合、修正後の辞書と、前記マスキング対象項目と、前記木構造におけるマスキング対象毎の階層と、前記データに含まれるレコード数とに基づいて、前記データを構成する各レコードを1つ以上の集合に分類して、各集合のレコード数 N と、レコード数 N の集合に属するレコードの割合とを再算出する、ことを特徴とする請求項7に記載の情報処理装置。

[請求項9] 1以上の項目が含まれるレコードで構成されるデータを統計加工によって匿名化するコンピュータが、

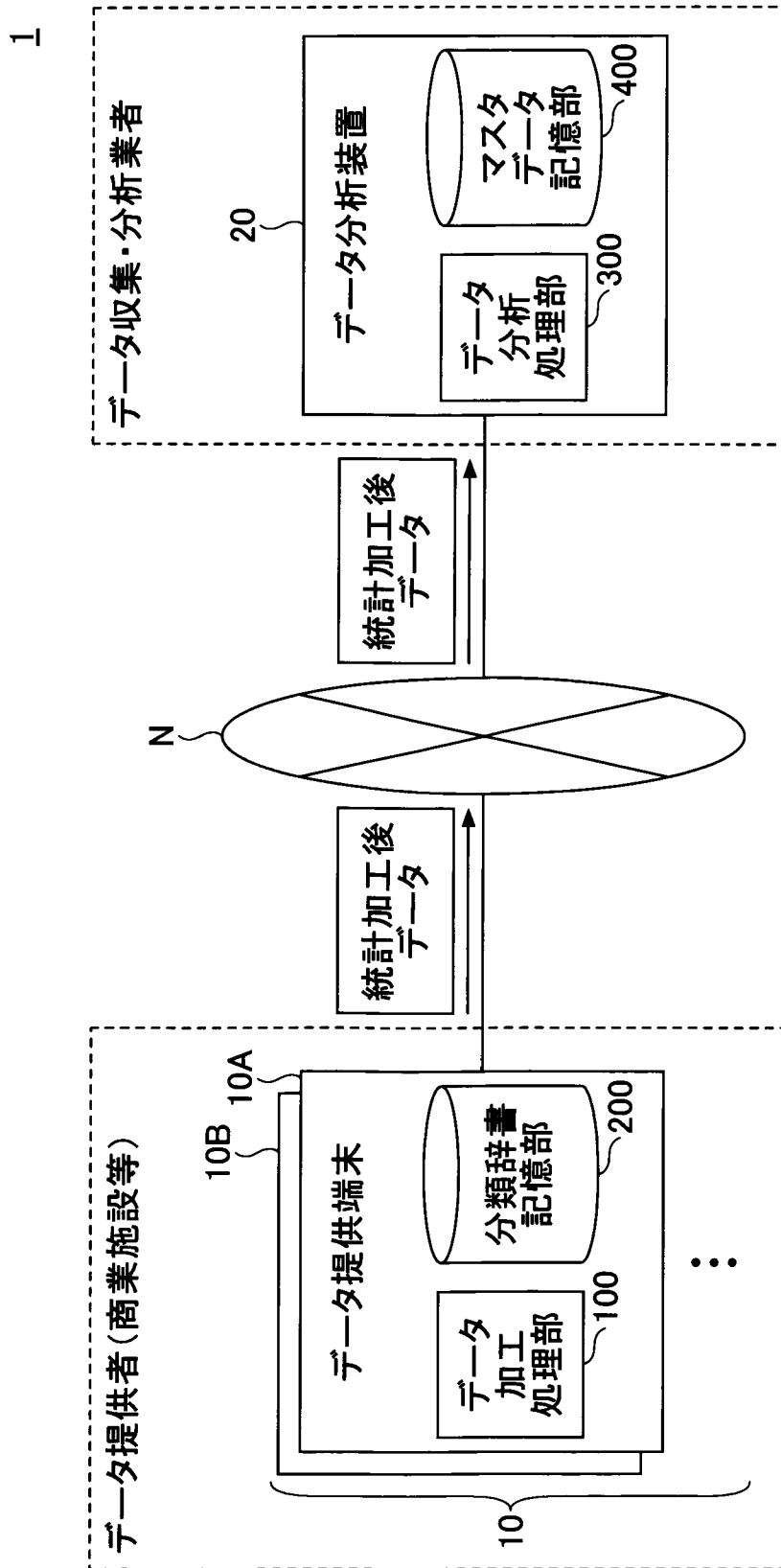
前記項目のうち、マスキングの対象となる項目を示すマスキング対象項目と、前記マスキング対象項目毎に該項目値のカテゴリが木構造で表現された辞書と、前記マスキング対象項目毎に前記木構造で選択された階層を示す選択階層と、前記データに含まれるレコード数とに基づいて、前記データを構成する各レコードを1つ以上の集合に分類して、各集合のレコード数 N と、レコード数 N の集合に属するレコードの割合とを算出する算出手順と、

前記マスキング対象項目毎の各階層において、前記レコード数 N 毎に、レコード数 N の集合に属するレコードの割合を U_i として表示する表示手順と、

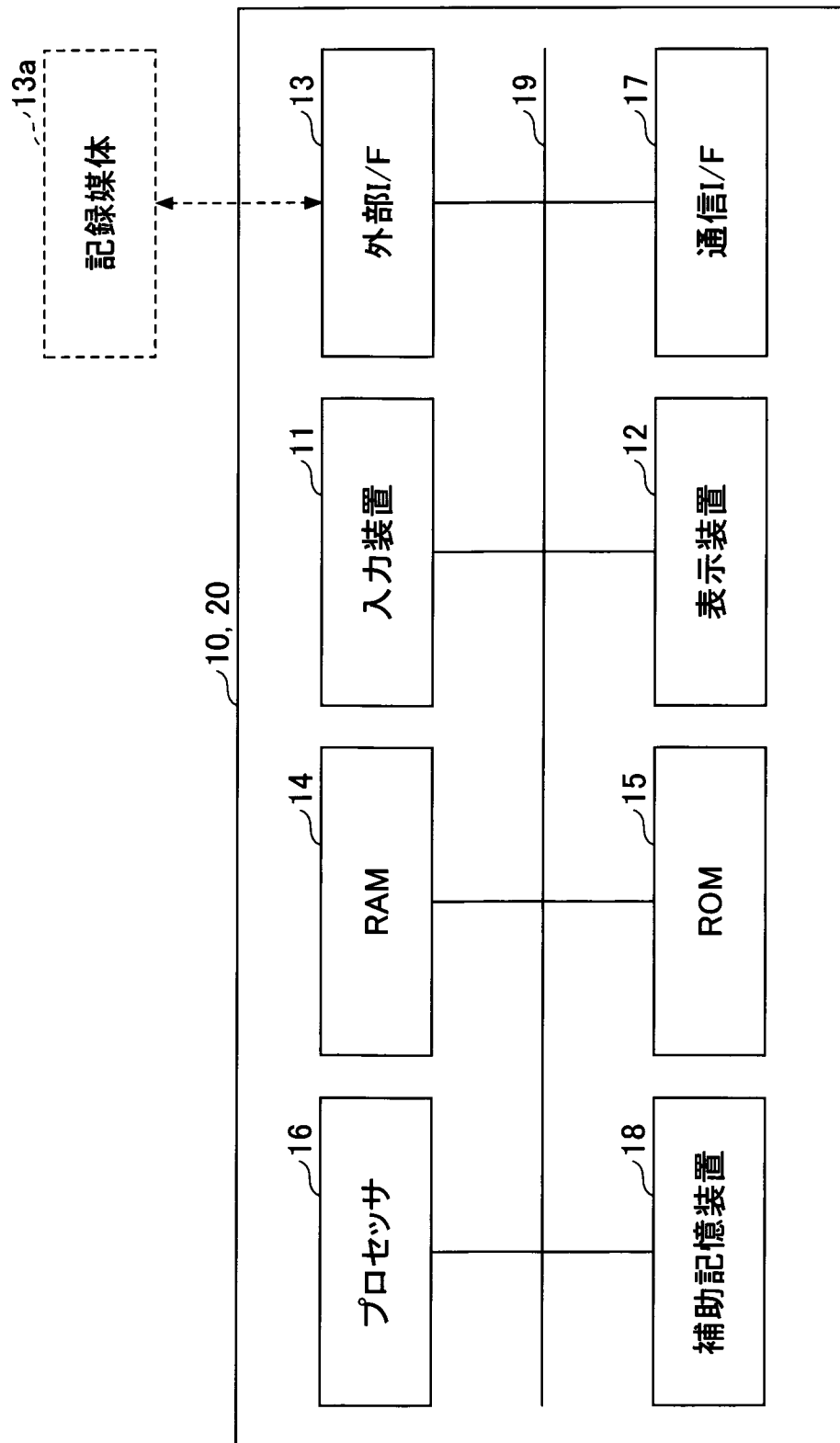
を実行することを特徴とする情報処理方法。

[請求項10] コンピュータを、請求項1乃至8の何れか一項に記載の情報処理装置における各手段として機能させるためのプログラム。

[図1]



[図2]

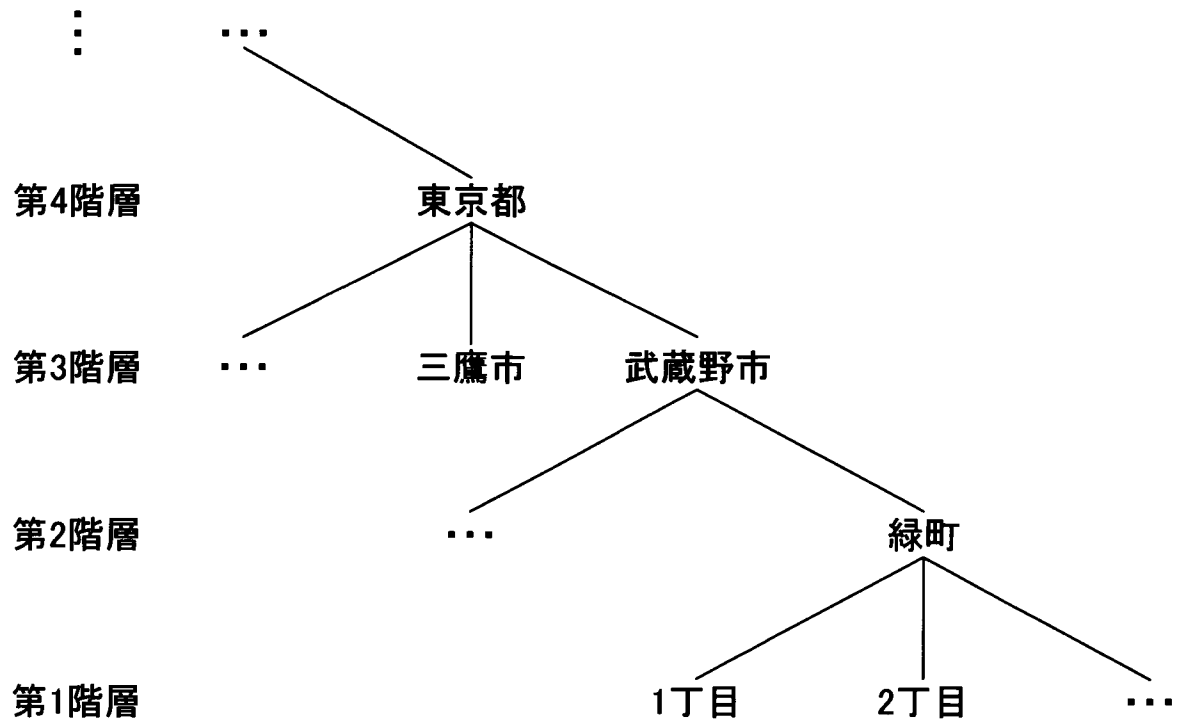


[図3]

レコードID	住所	年代	性別	金額
1	東京都武蔵野市緑町3丁目	10代	男	500円
2	東京都武蔵野市緑町1丁目	20代	男	1000円
3	東京都武蔵野市吉祥寺東町1丁目	10代	女	300円
4	東京都三鷹市下連雀1丁目	10代	男	2000円
5	東京都杉並区荻窪1丁目	30代	女	10000円

レコード

[図4A]

項目「住所」

[図4B]

項目「年代」

⋮

第4階層

...

第3階層

0～30代

第2階層

0～10代

20～30代

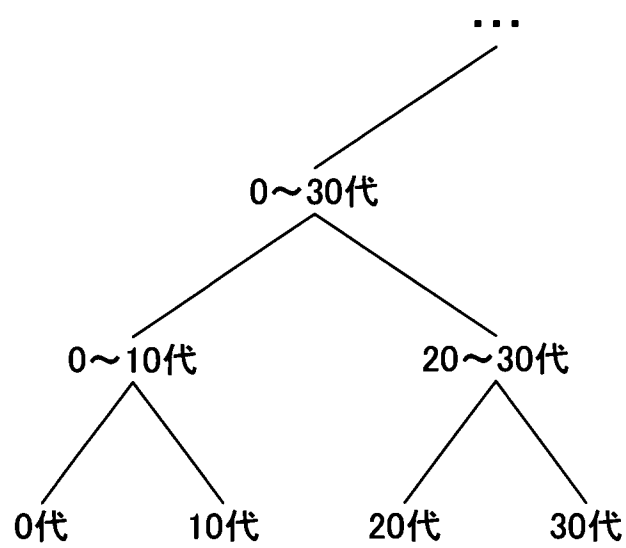
第1階層

0代

10代

20代

30代



[図5]

マスキング対象項目

Step1)

レコードID	住所	年代	性別	金額
1	東京都武蔵野市緑町3丁目	10代	男	500円
2	東京都武蔵野市緑町1丁目	20代	男	1000円
3	東京都武蔵野市吉祥寺東町1丁目	10代	女	300円
4	東京都三鷹市下連雀1丁目	10代	男	2000円
5	東京都杉並区荻窪1丁目	30代	女	10000円

Step2)



レコードID	住所	年代	性別	金額
1	東京都武蔵野市	0～30代	男	500円
2	東京都武蔵野市	0～30代	男	1000円
3	東京都武蔵野市	0～30代	女	300円
4	東京都三鷹市	0～30代	男	2000円
5	東京都杉並区	0～30代	女	10000円

各レコードについて、マスキング対象項目の
情報が一致するレコード同士を分類し、
各レコードのNと、Nが同一である
レコードの割合とを算出

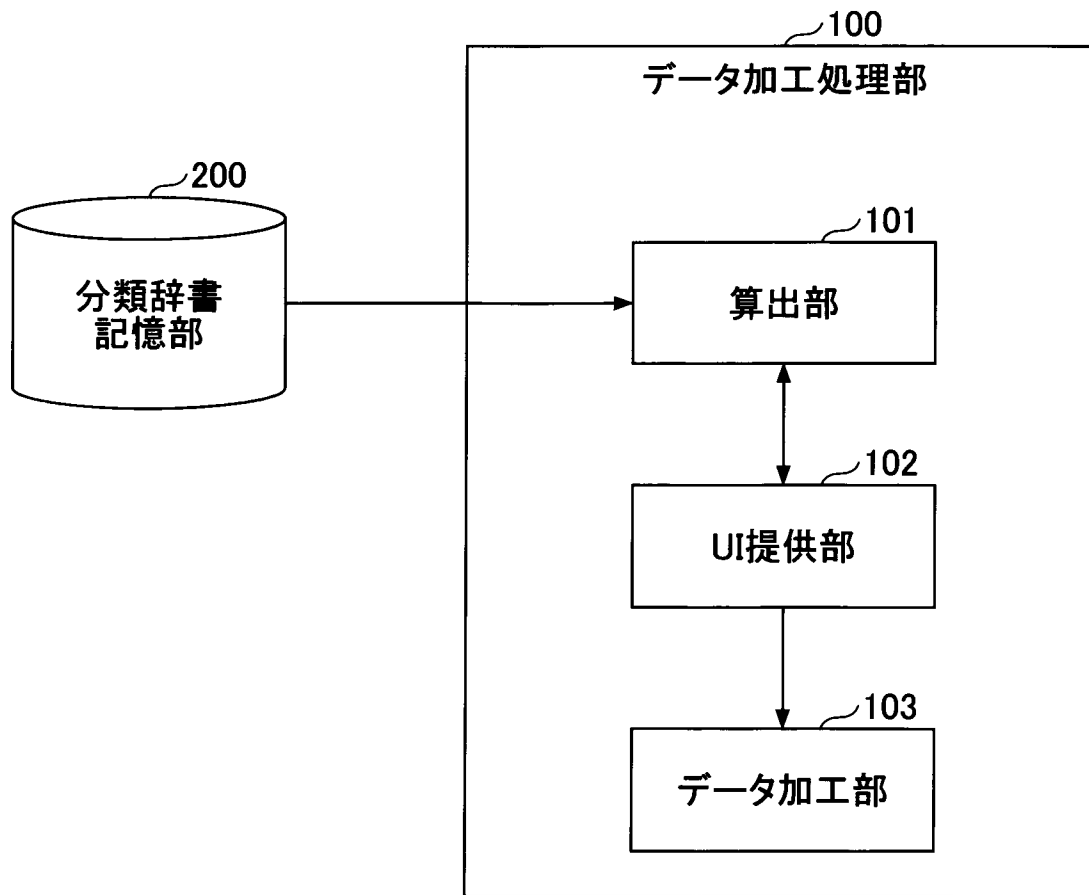
Nがk=2未満のレコードを削除対象とする

Step3)

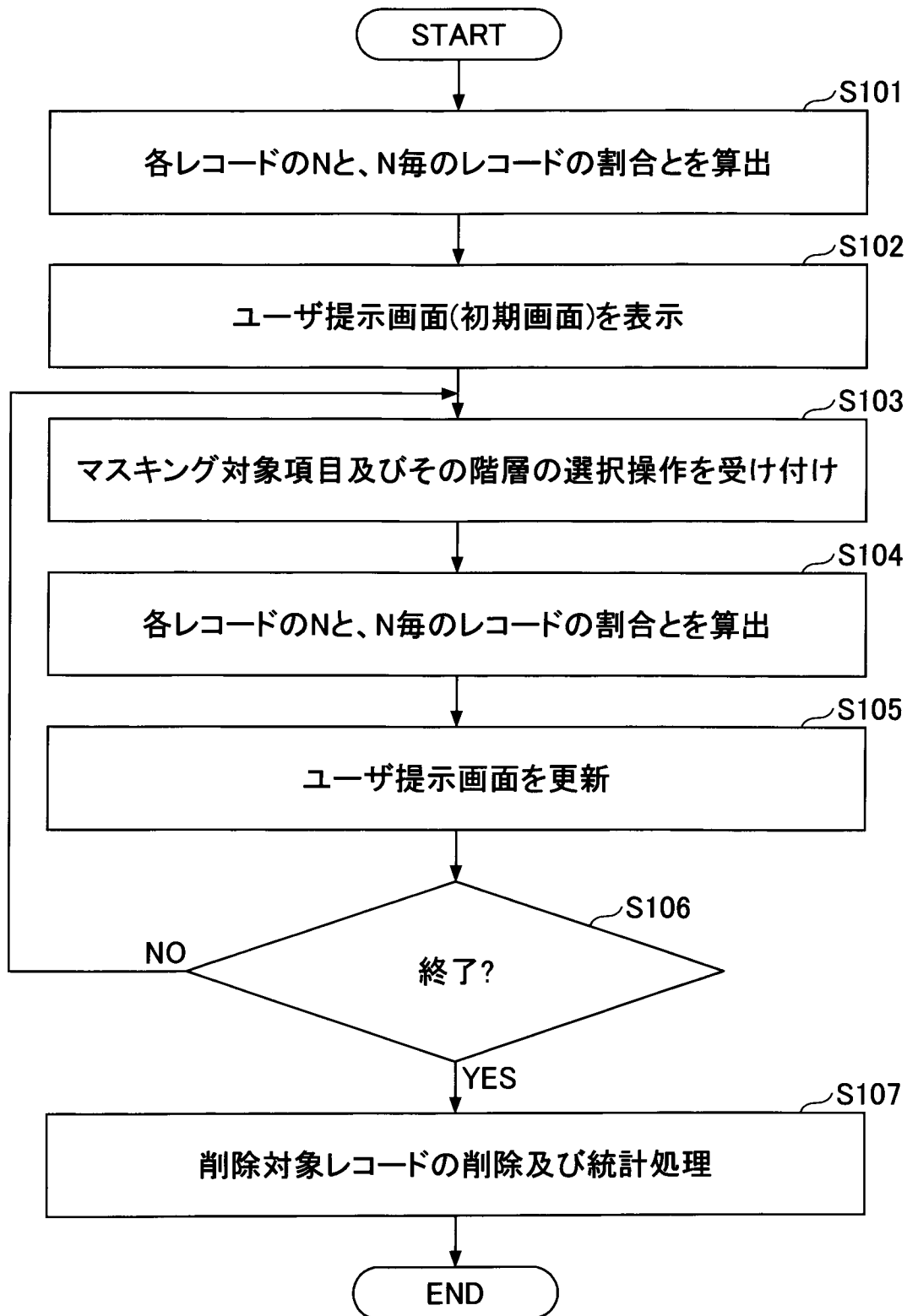


レコードID	住所	年代	人数	金額
1'	東京都武蔵野市	0～30代	3	1800円

[図6]



[図7]



[図8A]

G100

階層の選択

マスキング対象項目の階層を選択してください。

G110

	住所	年代
第4階層	N=3: 60%, N=1: 40%	N=1: 100%
第3階層	N=3: 60%, N=1: 40%	N=1: 100%
第2階層	N=2: 40%, N=1: 60%	N=1: 100%
第1階層	N=1: 100%	N=1: 100%

G120

キャンセル

決定

[図8B]

G100

階層の選択

マスキング対象項目の階層を選択してください。

G110

	住所	年代
第4階層	N=3: 60%, N=1: 40%	N=3: 60%, N=1: 40%
第3階層	N=3: 60%, N=1: 40%	N=3: 60%, N=1: 40%
第2階層	N=2: 40%, N=1: 60%	N=3: 60%, N=1: 40%
第1階層	N=1: 100%	N=3: 60%, N=1: 40%

G120

キャンセル

決定

N及びN毎の割合が更新される。

[図8C]

G100

階層の選択

マスキング対象項目の階層を選択してください。

G110

	住所	年代
第4階層	N=3: 60%, N=1: 40%	N=5: 100%
第3階層	N=3: 60%, N=1: 40%	N=5: 100%
第2階層	N=2: 40%, N=1: 60%	N=3: 60%, N=1: 40%
第1階層	N=1: 100%	N=3: 60%, N=1: 40%

G120

キャンセル

決定

N及びN毎の割合が更新される。

[図8D]

G100

階層の選択

マスキング対象項目の階層を選択してください。

G110

	住所	年代
第4階層	N=5: 100%	N=5: 100%
第3階層	N=3: 60%, N=1: 40%	N=5: 100%
第2階層	N=2: 40%, N=1: 60%	N=3: 60%, N=1: 40%
第1階層	N=1: 100%	N=3: 60%, N=1: 40%

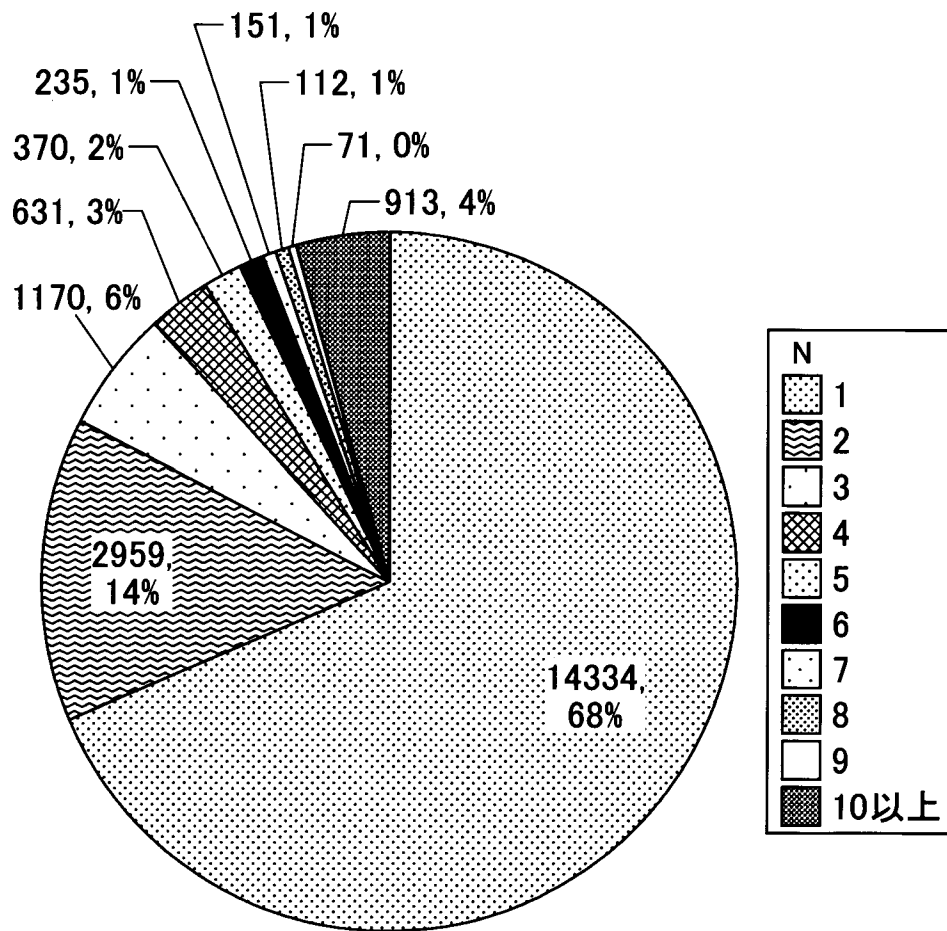
G120

キャンセル

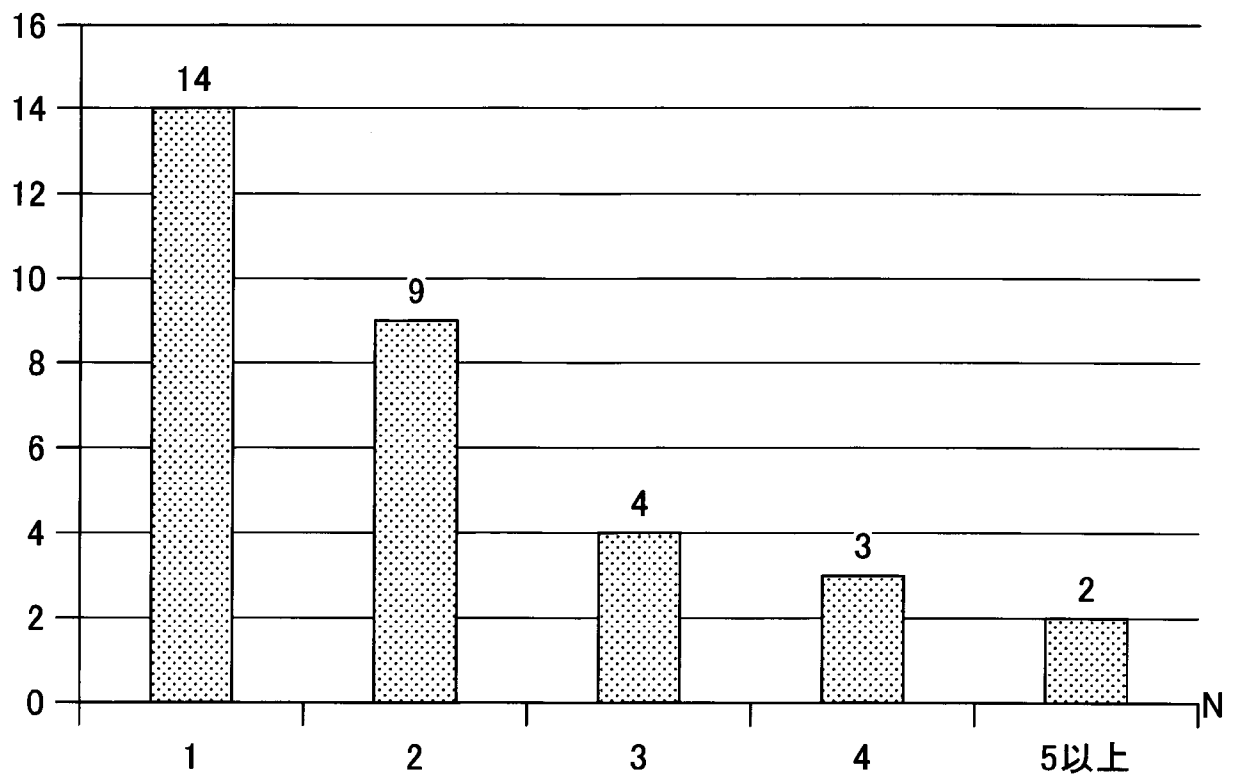
決定

N及びN毎の割合が更新される。

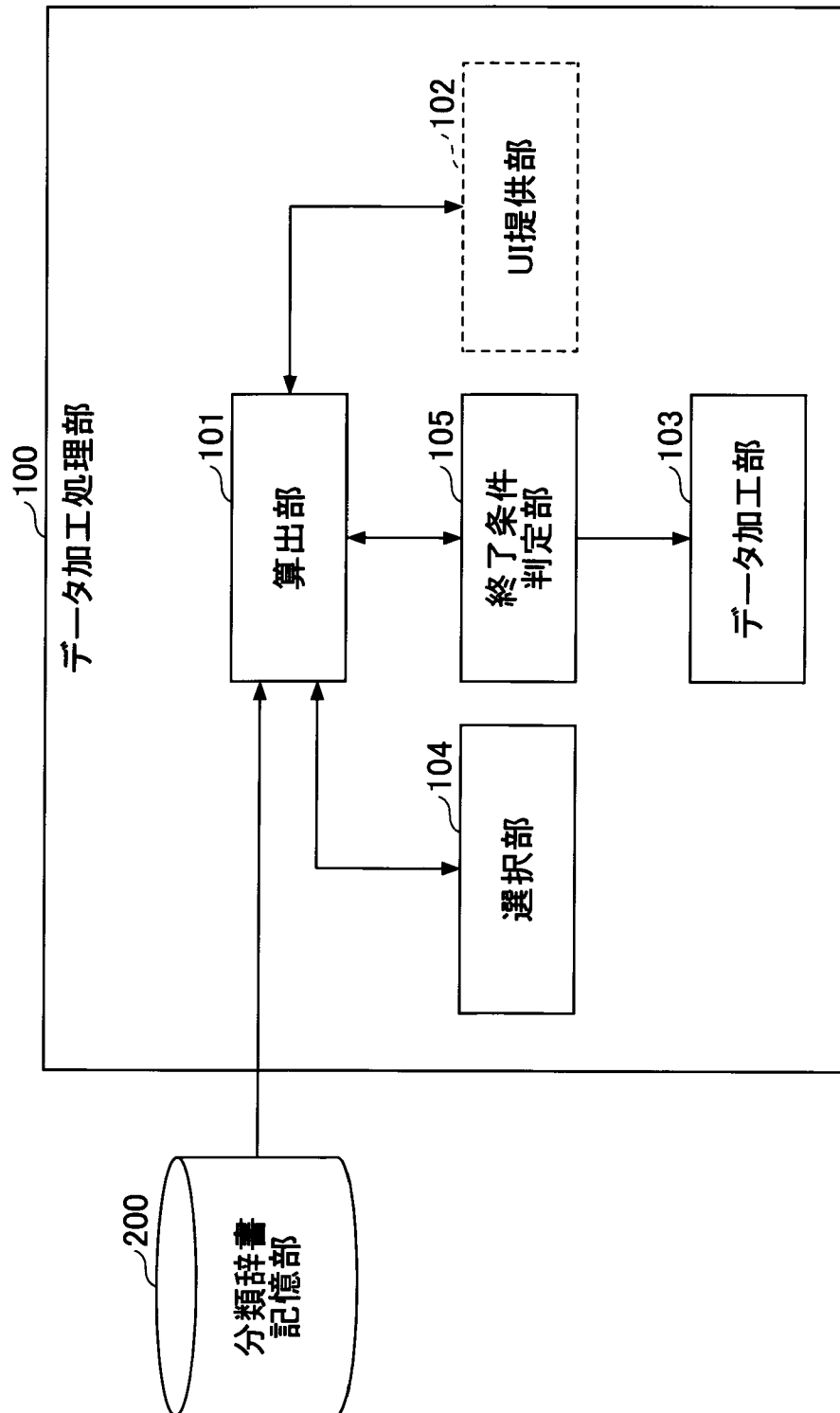
[図9A]



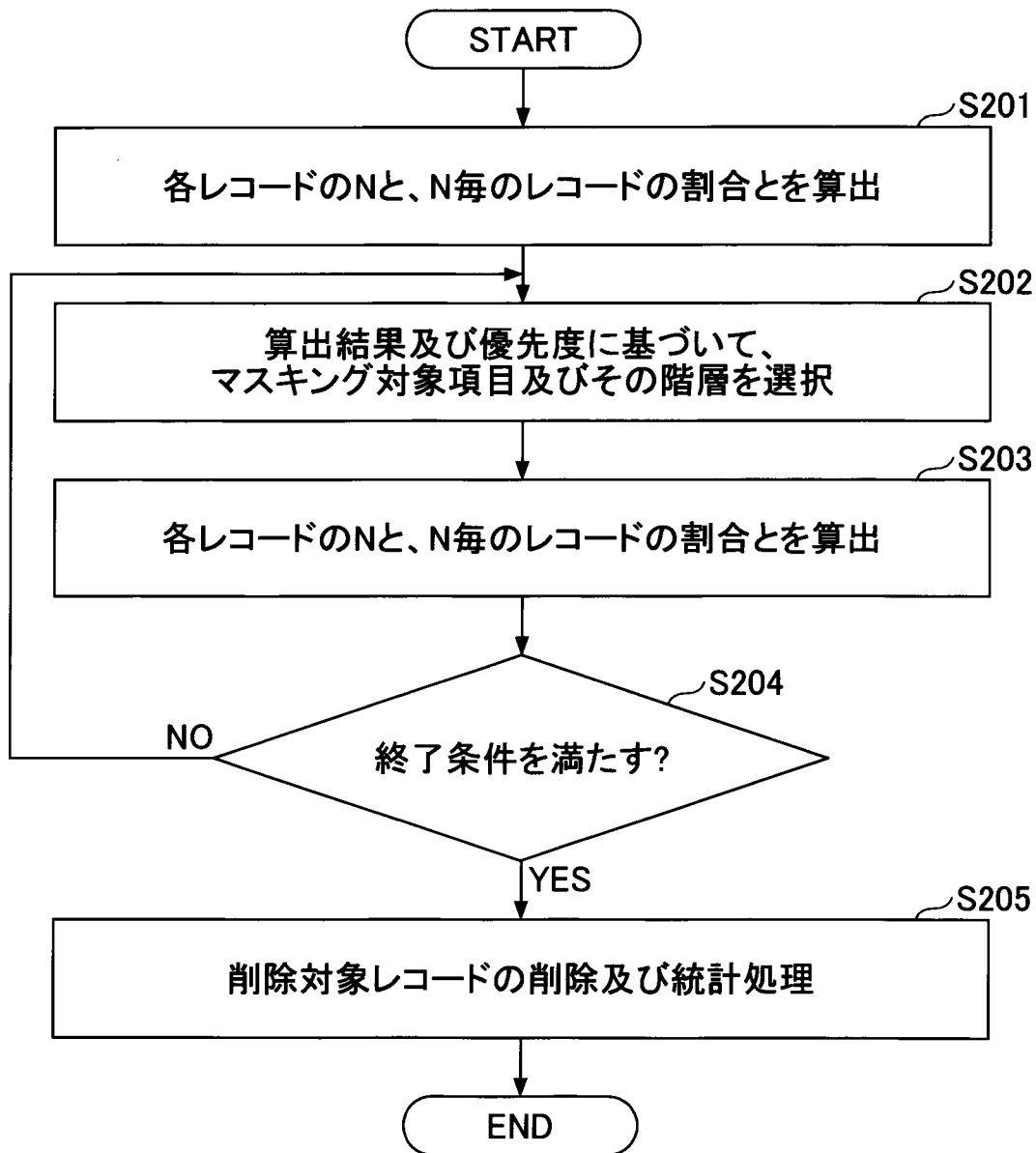
[図9B]



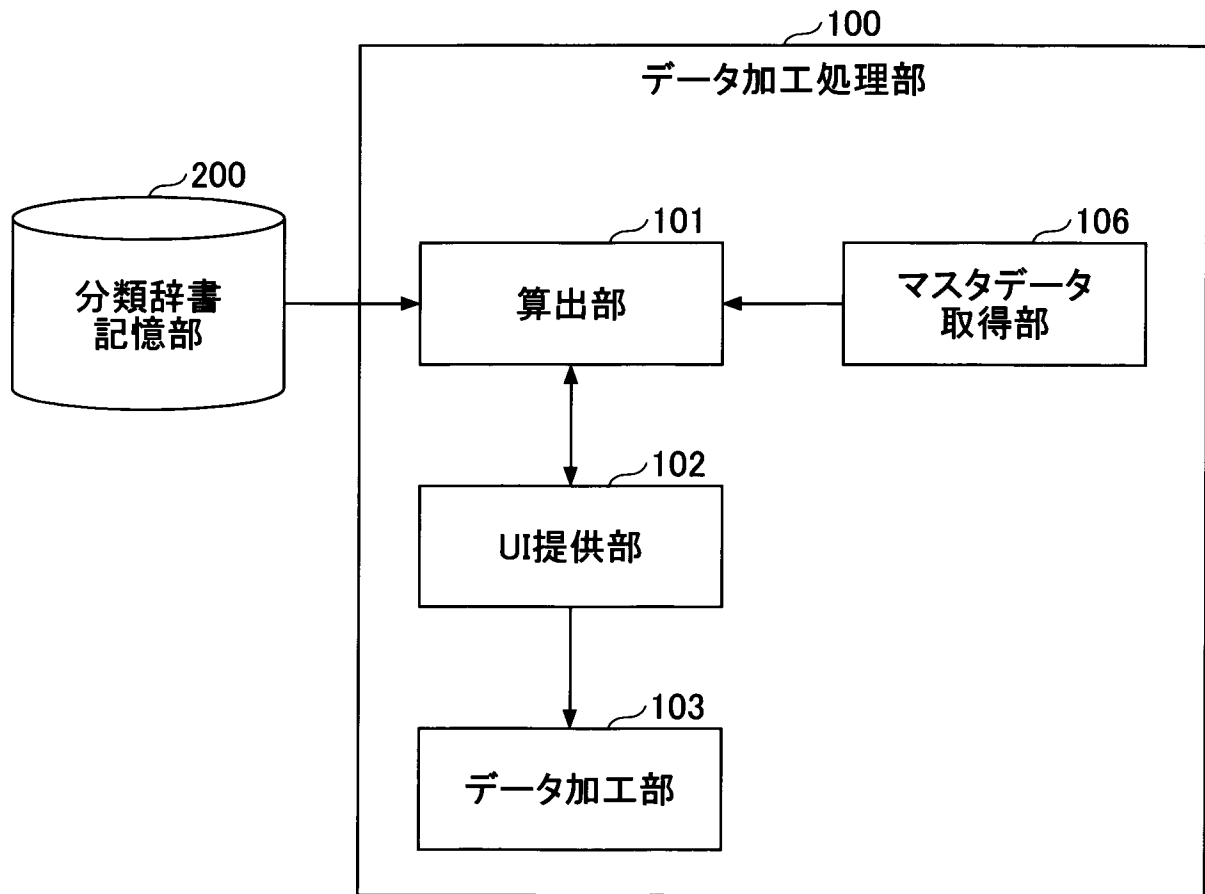
[図10]



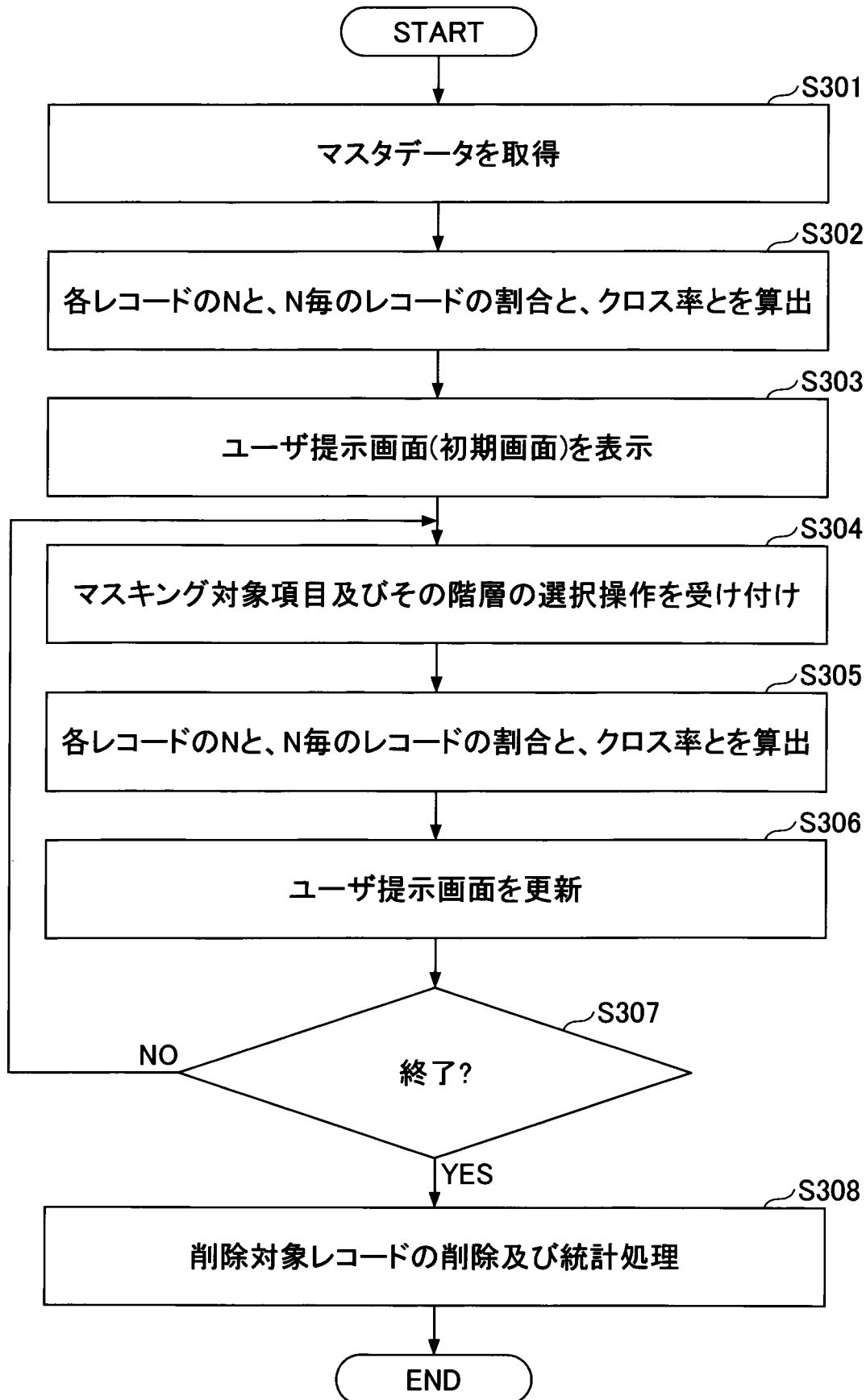
[図11]



[図12]



[図13]



[図14]

G100

階層の選択

マスキング対象項目の階層を選択してください。

G110

	住所	年代
第4階層	N=3: 60%, N=1: 40% クロス率: ...	N=1: 100% クロス率: ...
第3階層	N=3: 60%, N=1: 40% クロス率: ...	N=1: 100% クロス率: ...
第2階層	N=2: 40%, N=1: 60% クロス率: ...	N=1: 100% クロス率: ...
第1階層	N=1: 100% クロス率: ...	N=1: 100% クロス率: ...

G120

キャンセル 決定

[図15]

対象データ

レコードID	商品種別	...
1	チョコレート	...
2	飴	...
3	扇風機	...
4	飴	...

マスターデータ

レコードID	商品種別	...
101	チョコレート	...
102	チョコレート	...
103	チョコレート	...
104	飴	...
105	掃除機	...

[図16]

対象データ

レコードID	商品種別	業種
1	チョコレート	Aデパート〇〇店
2	飴	Aデパート〇〇店
3	扇風機	C電気店

マスタデータ

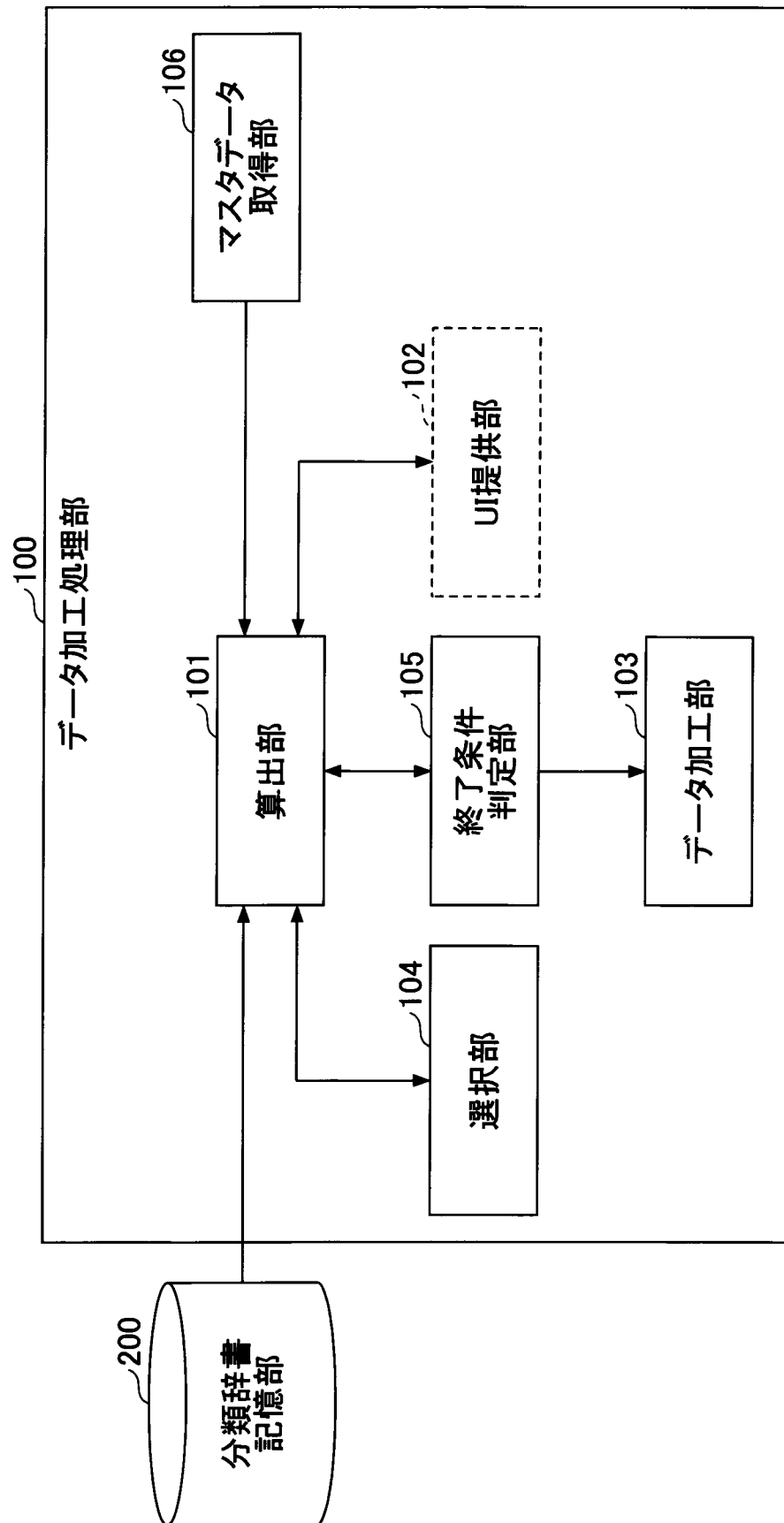
レコードID	商品種別	業種
101	チョコレート	Aデパート〇〇店
102	チョコレート	Aデパート××店
103	チョコレート	Bデパート△△店
104	飴	Aデパート〇〇店
105	扇風機	C電気店

↓ 集計処理

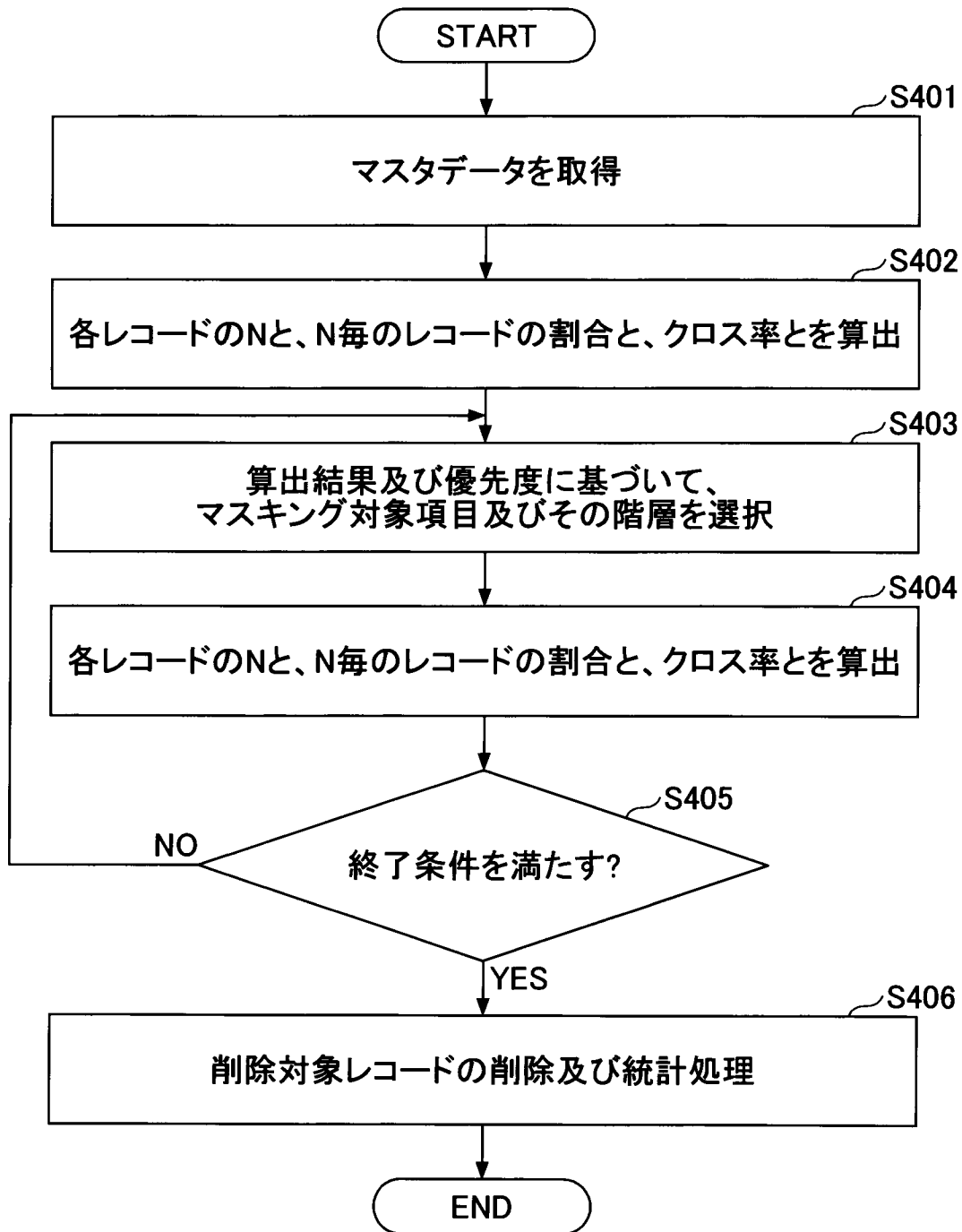
集計データ

レコードID	商品種別	業種	業種数	ヒット数
1'	チョコレート	Aデパート〇〇店、 Aデパート××店、 Bデパート△△店	3	4
2'	飴	Aデパート〇〇店	1	2
3'	扇風機	C電気店	1	2

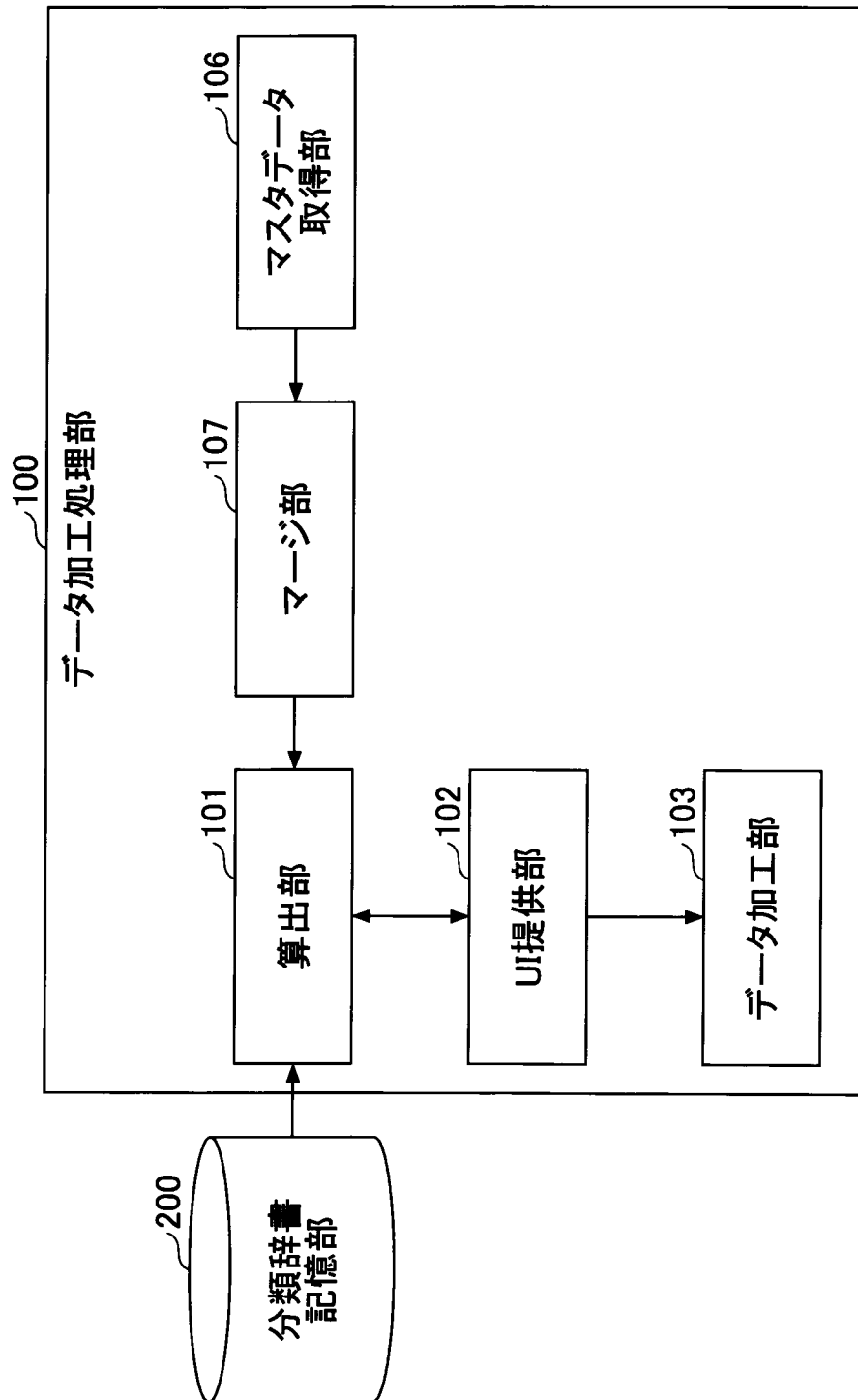
[図17]



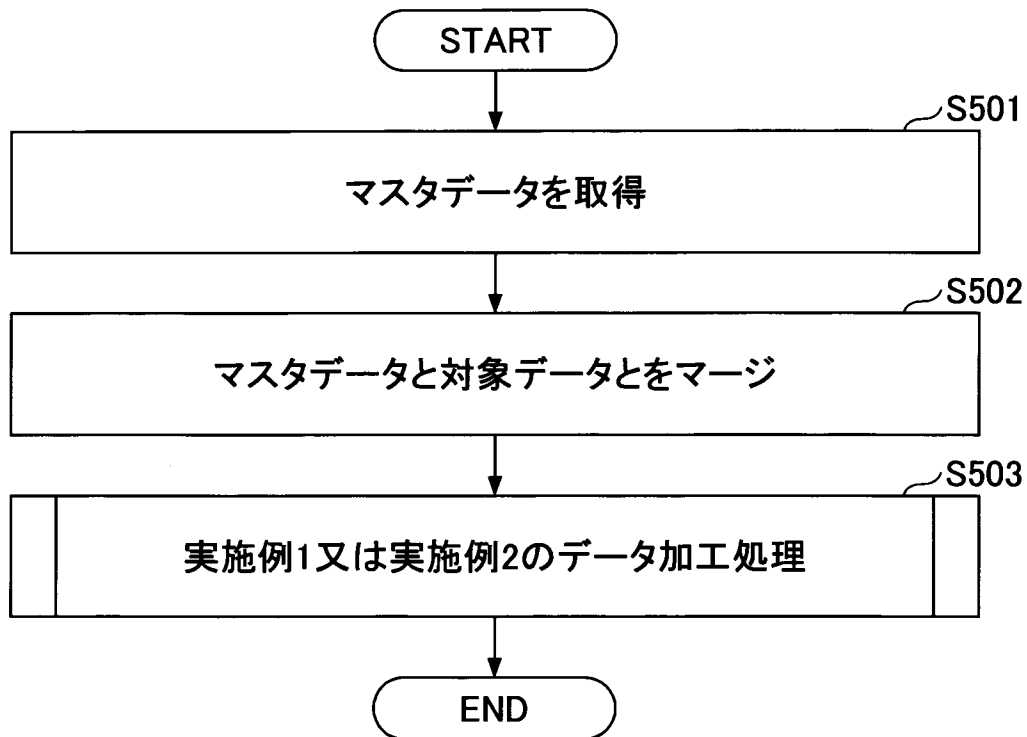
[図18]



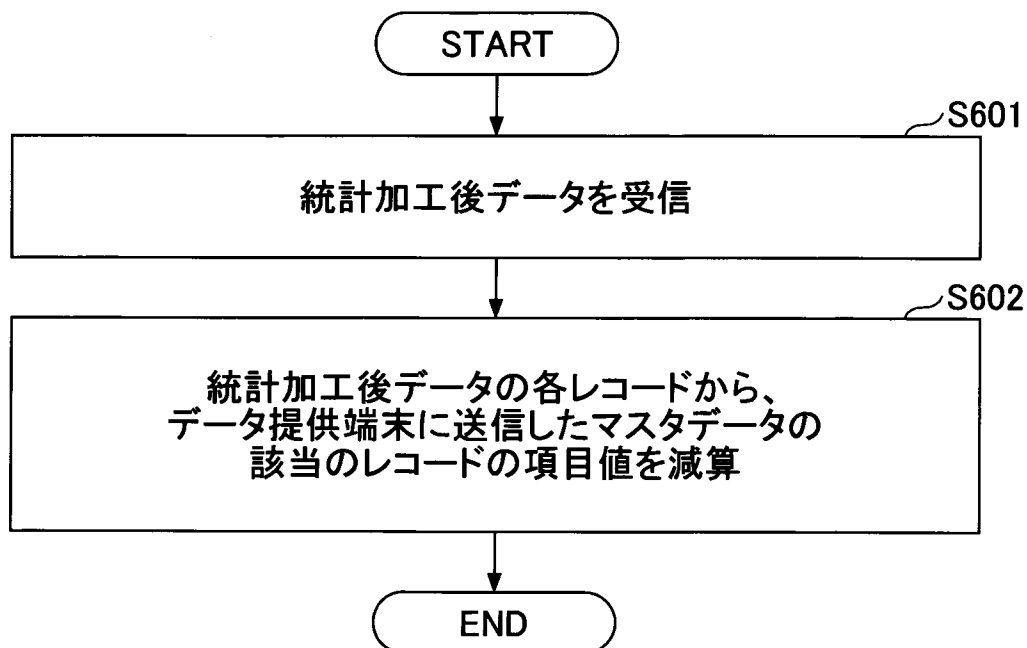
[図19]



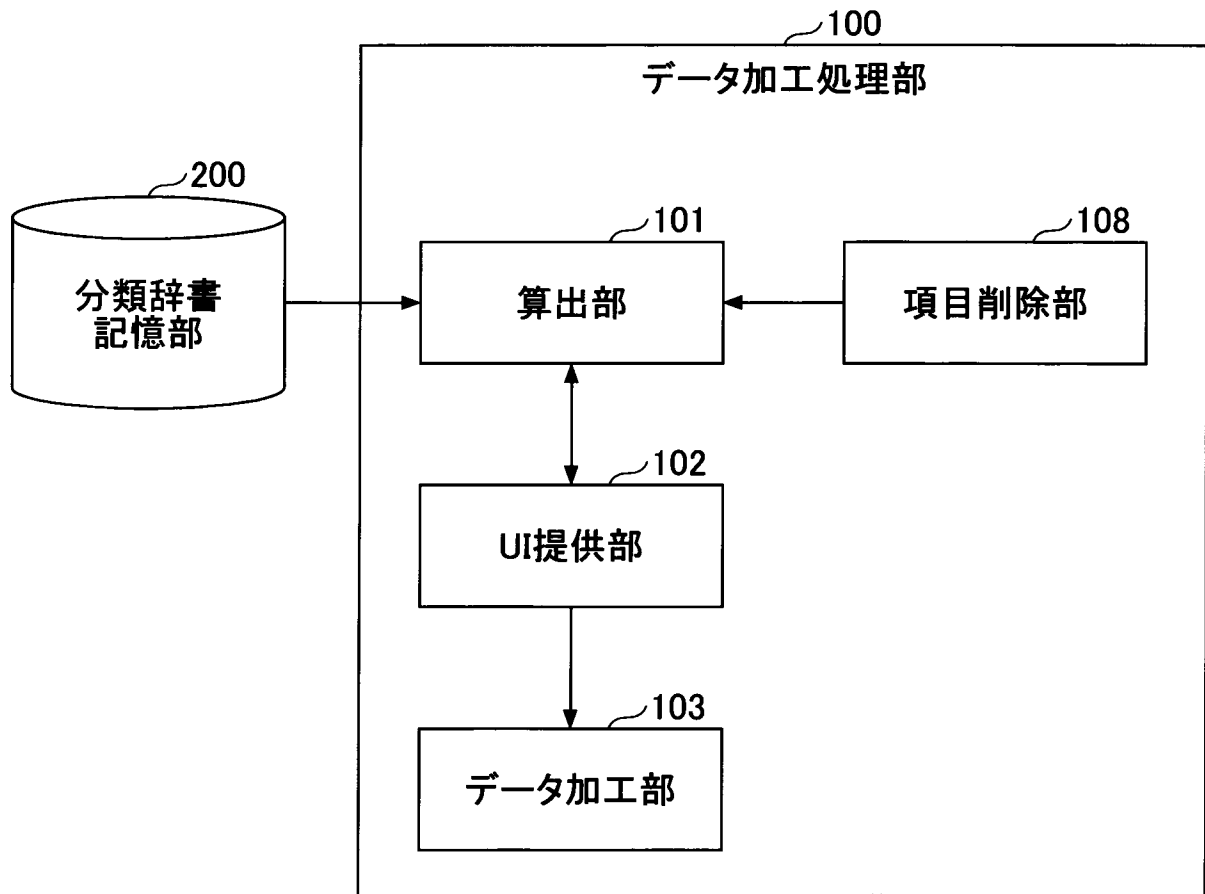
[図20]



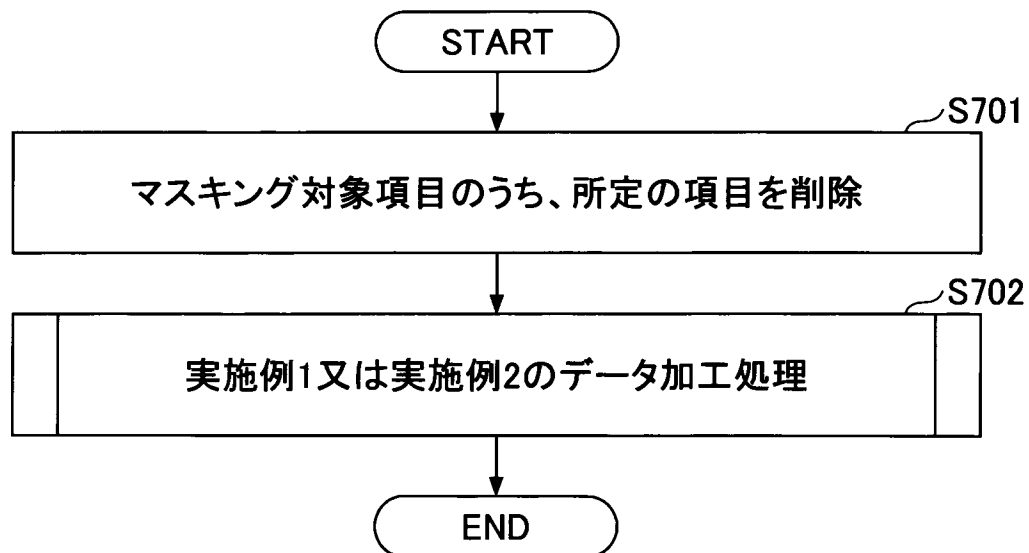
[図21]



[図22]



[図23]



[図24]

対象データ

レコードID	住所	年代	業種	金額
1	東京都武蔵野市緑町	10代	Aデパート	500円
2	東京都武蔵野市緑町	10代	Aデパート	1000円
3	東京都武蔵野市吉祥寺東町	10代	Bデパート	300円
4	東京都武蔵野市緑町	20代	Aデパート	2000円
5	東京都武蔵野市緑町	30代	C電気店	10000円

項目「年代」を削除

項目「住所」を削除

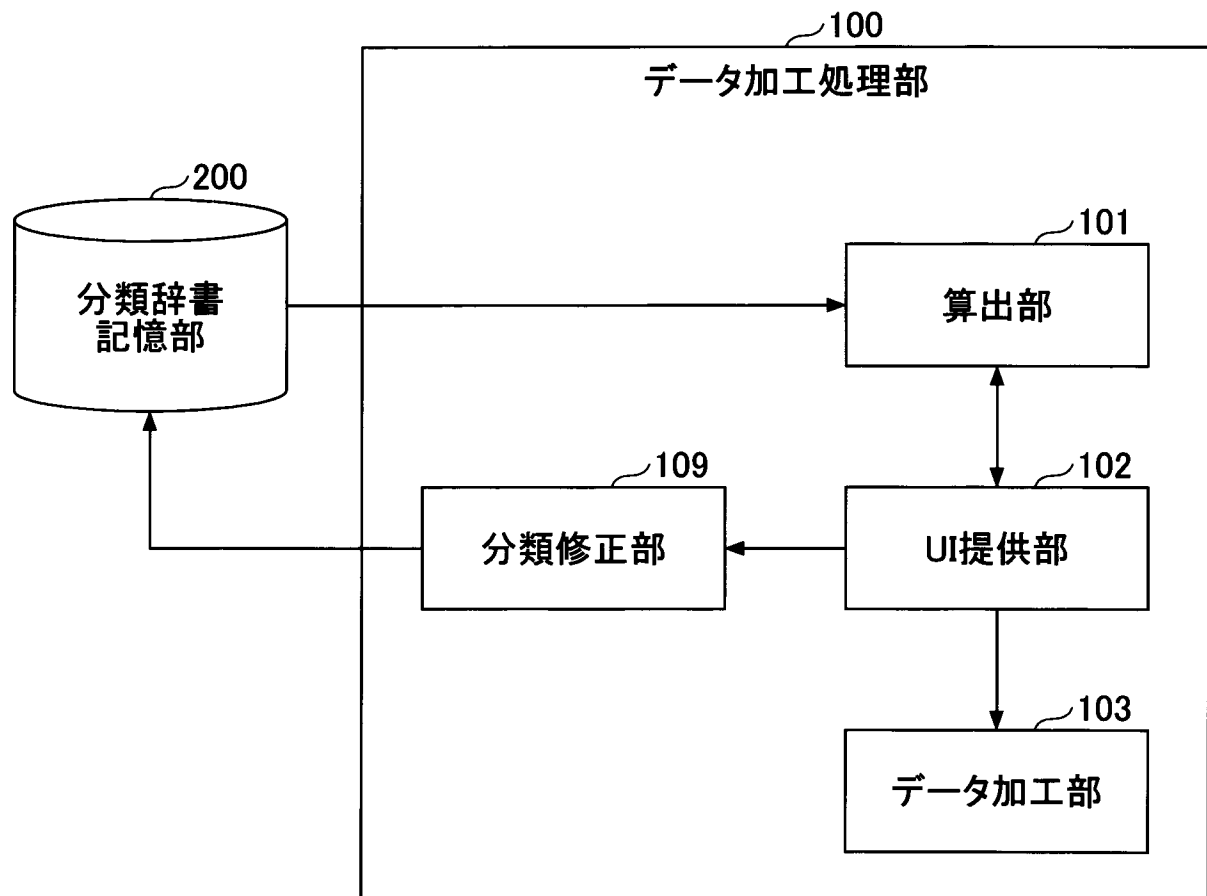
年代削除後対象データ

レコードID	住所	業種	金額
1	東京都武蔵野市緑町	Aデパート	500円
2	東京都武蔵野市緑町	Aデパート	1000円
3	東京都武蔵野市吉祥寺東町	Bデパート	300円
4	東京都武蔵野市緑町	Aデパート	2000円
5	東京都武蔵野市緑町	C電気店	10000円

住所削除後対象データ

レコードID	年代	業種	金額
1	10代	Aデパート	500円
2	10代	Aデパート	1000円
3	10代	Bデパート	300円
4	20代	Aデパート	2000円
5	30代	C電気店	10000円

[図25]



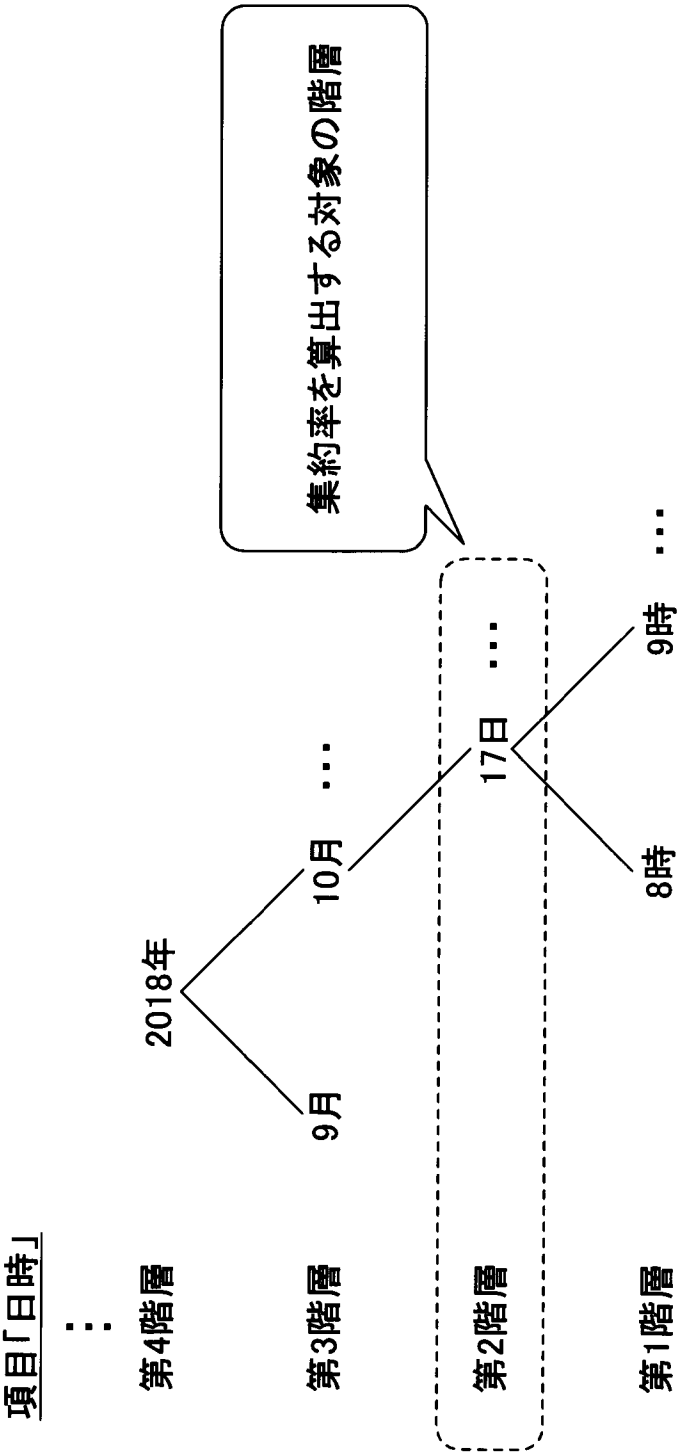
[図26A]

対象データ

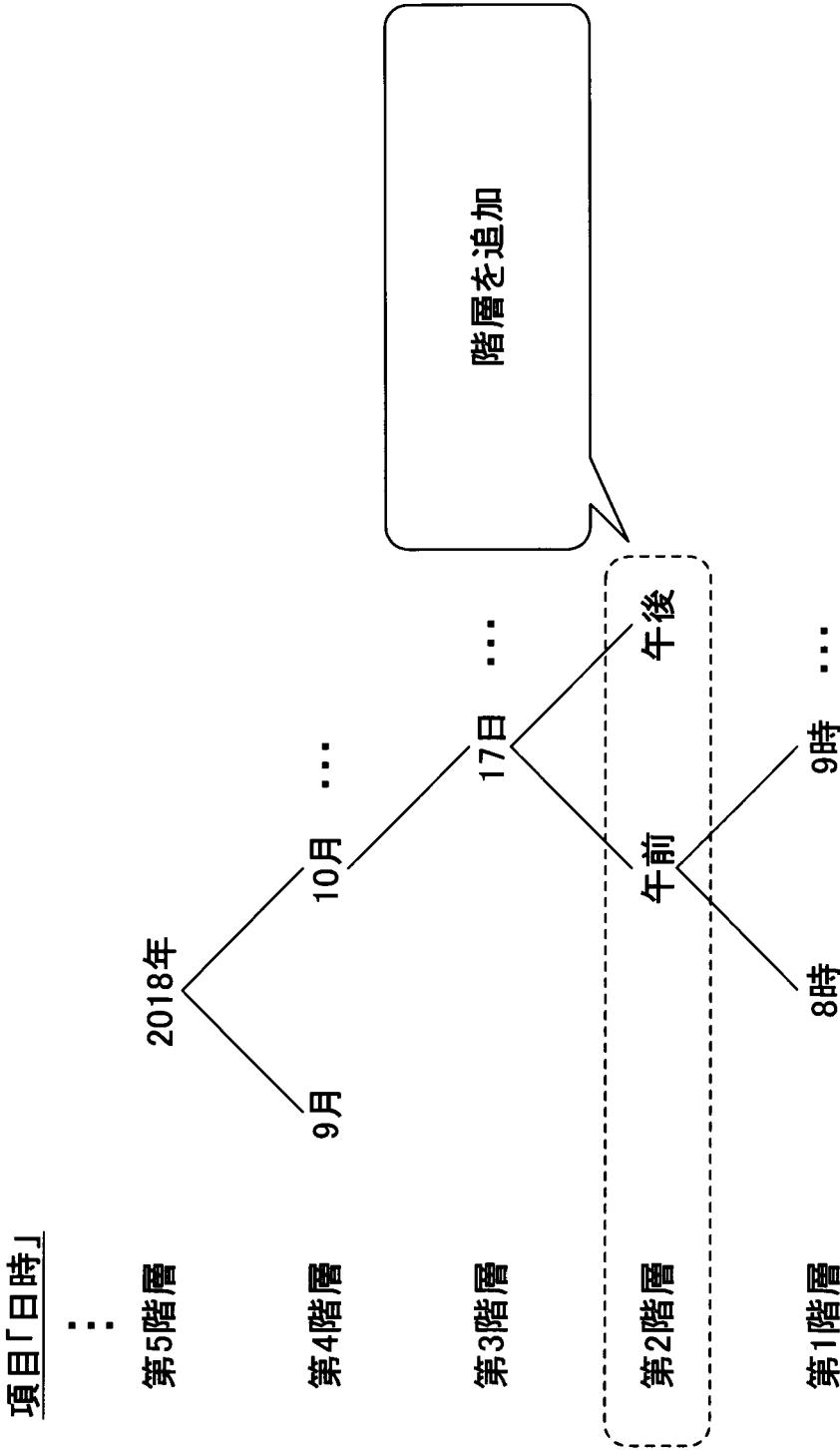
レコードID	日時	年代	金額
1	2018年10月17日 8時	10代	1000円以下
2	2018年10月17日 8時	20代	1000円以下
3	2018年10月17日 9時	20代	1000円以下
4	2018年10月17日 11時	10代	1000円以下
5	2018年10月17日 17時	30代	2000円以下
6	2018年10月17日 20時	30代	2000円以下

マスキング対象項目を、「レコードID」以外の全項目、 $k=1$ とした場合、項目「日時」の集約率は80%

[図26B]



[図27A]



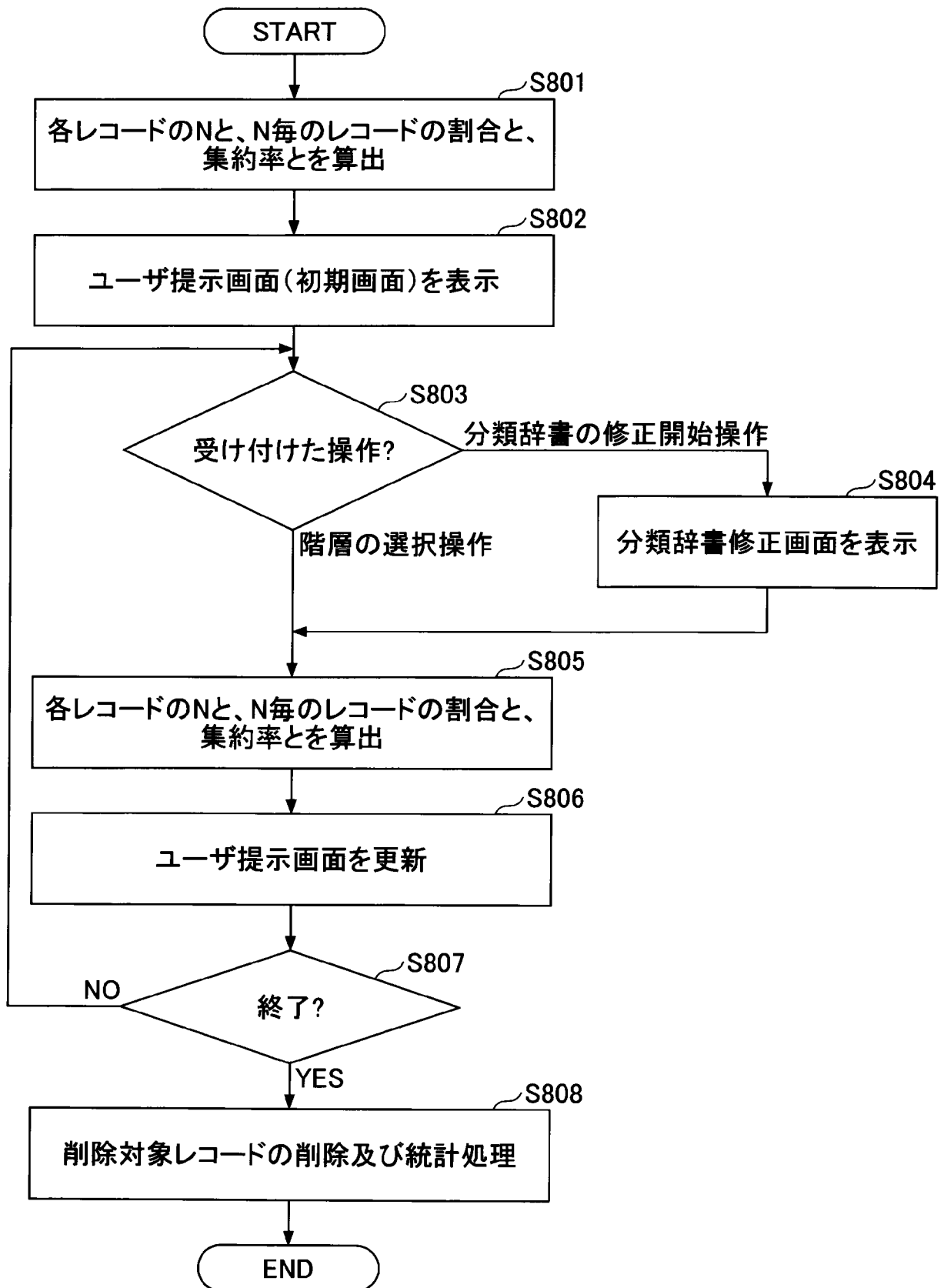
[図27B]

対象データ

レコードID	日時	年代	金額
1	2018年10月17日 午前	10代	1000円以下
2	2018年10月17日 午前	20代	1000円以下
3	2018年10月17日 午前	20代	1000円以下
4	2018年10月17日 午前	10代	1000円以下
5	2018年10月17日 午後	30代	2000円以下
6	2018年10月17日 午後	30代	2000円以下

マスキング対象項目を、「レコードID」以外の全項目、
k=1とした場合、項目「日時」の
集約率は60%
また、時刻情報として午前又は
午後を残せる

[図28]



G100

G110

決定

G120

決定

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2019/020164

A. CLASSIFICATION OF SUBJECT MATTER

Int. Cl. G06F21/62 (2013.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Int. Cl. G06F21/62

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Published examined utility model applications of Japan 1922-1996

Published unexamined utility model applications of Japan 1971-2019

Registered utility model specifications of Japan 1996-2019

Published registered utility model applications of Japan 1994-2019

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2015-046030 A (HITACHI SOLUTIONS, LTD.) 12 March 2015, paragraphs [0048], [0049], fig. 2 (Family: none)	1-10
A	JP 2017-49693 A (FUJITSU LTD.) 09 March 2017, paragraph [0099], fig. 8 & US 2017/0061156 A1, paragraph [0122], fig. 8 & EP 3136284 A1	1-10

☐ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search
14.08.2019

Date of mailing of the international search report
17.09.2019

Name and mailing address of the ISA/
Japan Patent Office
3-4-3, Kasumigaseki, Chiyoda-ku,
Tokyo 100-8915, Japan

Authorized officer

Telephone No.

A. 発明の属する分野の分類 (国際特許分類 (IPC))

Int.Cl. G06F21/62 (2013.01) 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. G06F21/62

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	1971-2019年
日本国実用新案登録公報	1996-2019年
日本国登録実用新案公報	1994-2019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)

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 2015-046030 A (株式会社日立ソリューションズ) 2015.03.12, 段落[0048]-[0049], 図2 (ファミリーなし)	1-10
A	JP 2017-49693 A (富士通株式会社) 2017.03.09, 段落[0099], 図8 & US 2017/0061156 A1, par. [0122], fig. 8 & EP 3136284 A1	1-10

C欄の続きにも文献が列挙されている。

パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

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

の日の後に公表された文献

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

国際調査を完了した日

14.08.2019

国際調査報告の発送日

17.09.2019

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)

平井 誠

5 S

9071

電話番号 03-3581-1101 内線 3546