

(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
2017年2月2日(02.02.2017)



(10) 国際公開番号
WO 2017/017748 A1

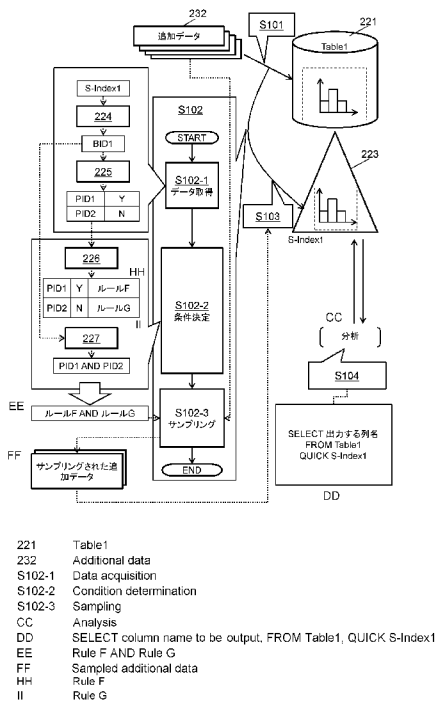
- (51) 国際特許分類:
G06F 17/30 (2006.01)
- (21) 国際出願番号: PCT/JP2015/071196
- (22) 国際出願日: 2015年7月27日(27.07.2015)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (71) 出願人: 株式会社日立製作所 (HITACHI, LTD.) [JP/JP]; 〒1008280 東京都千代田区丸の内一丁目6番6号 Tokyo (JP).
- (72) 発明者: 山田 佑樹(YAMADA, Yuki); 〒1008280 東京都千代田区丸の内一丁目6番6号 株式会社日立製作所内 Tokyo (JP). 濱田 信輔(HAMADA, Shinsuke); 〒1008280 東京都千代田区丸の内一丁目6番6号 株式会社日立製作所内 Tokyo (JP). 河井 渉(KAWAI, Wataru); 〒1008280 東京都千代田区丸の内一丁目6番6号 株式会社日立製作所内 Tokyo (JP). 山根 康仁(YAMANE, Yasuhito); 〒1008280 東京都千代田区丸の内一丁目6番6号 株式会社日立製作所内 Tokyo (JP). 藤原 真二(FUJIWARA, Shinji); 〒1008280 東京都千代田区丸の内一丁目6番6号 株式会社日立製作所内 Tokyo (JP).
- (74) 代理人: 特許業務法人ウィルフォート国際特許事務所 (WILLFORT INTERNATIONAL PATENT FIRM); 〒1030016 東京都中央区日本橋小網町1-9-7 日本橋TCビル 1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, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, 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,

[続葉有]

(54) Title: COMPUTER SYSTEM AND SAMPLING METHOD

(54) 発明の名称: 計算機システム及びサンプリング方法

【図1】



(57) Abstract: Provided is a computer system that identifies work associated with a sampling candidate data group (one or more sampling candidate data items). On the basis of management information, which is information that includes information expressing a relationship between work including a data search, the work properties and the sampling rule, the computer system determines a sampling condition in accordance with a sampling rule corresponding to the properties of the identified work. Then, the computer system samples, from the sampling candidate data group, data items which meet the determined sampling condition. A sampling result is referenced at the time of the data search performed in the course of the identified work. A sampling data group, which is one or more sampling data items that include the sampled data, and/or a sampling index, which is an index of the sampling data group, is the sampling result which is referenced at the time of the data search performed in the course of the identified work.

(57) 要約: 計算機システムは、サンプリング候補データ群（1以上のサンプリング候補データ）に関連付けられた業務を特定する。計算機システムは、その特定した業務の特性に対応したサンプリングルールに従うサンプリング条件を、データ検索を含んだ業務と業務の特性とサンプリングルールとの関係を表す情報を含んだ情報である管理情報を基に決定する。そして、計算機システムは、決定したサンプリング条件に適合するデータをサンプリング候補データ群からサンプリングする。特定された業務において行われるデータ検索の際に、サンプリング結果が参照される。サンプリングされたデータを含んだ1以上のサンプリングデータであるサンプリングデータ群と、そのサンプリングデータ群のインデクスであるサンプリングインデクスとのうちの少なくとも1つが、上記特定された業務において行われるデータ検索の際に参照され

るサンプリング結果である。

WO 2017/017748 A1

MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, 添付公開書類:
TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, — 國際調查報告 (條約第 21 條(3))
KM, ML, MR, NE, SN, TD, TG).

明 細 書

発明の名称： 計算機システム及びサンプリング方法

技術分野

[0001] 本発明は、概して、コンピュータを用いたサンプリング技術に関する。

背景技術

[0002] 膨大な量のデータ群（母集団）の全てを検索対象（検索範囲）とすると、検索に長い時間を要する。そこで、母集団から予めサンプリングされたデータのみで構成されたサンプリングデータ群を検索対象とすることで、検索時間を短縮する方法が考えられる。

[0003] 例えば、特許文献1には、データベースシステムに設けられた層化抽出機構が記載されている。特許文献1によれば、層化抽出機構は、層化抽出の依頼を意味する句をクエリに定義する。ソース表からのデータは、クエリにおける層化条件に基づき異なるサブグループに分類される。サンプリングが、サブグループ別に行われる。

先行技術文献

特許文献

[0004] 特許文献1：US7,069,264

発明の概要

発明が解決しようとする課題

[0005] 母集団より小さいサンプリングデータ群を検索対象とすると、母集団全体が検索対象であることに比べて業務精度（例えば分析精度）が大きく落ちるおそれがある。この問題は、母集団がデータベースである場合に限られない。

課題を解決するための手段

[0006] 計算機システムは、1以上のサンプリング候補データであるサンプリング候補データ群に関連付けられた業務を特定する。計算機システムは、その特定した業務の特性に対応したサンプリングルールに従うサンプリング条件を

、データ検索を含んだ業務と業務の特性とサンプリングルールとの関係を表す情報を含んだ情報である管理情報を基に決定する。そして、計算機システムは、決定したサンプリング条件に適合するデータをサンプリング候補データ群からサンプリングする。特定された業務において行われるデータ検索の際に、サンプリング結果が参照される。サンプリングされたデータを含んだ1以上のサンプリングデータであるサンプリングデータ群と、そのサンプリングデータ群のインデクスであるサンプリングインデクスとのうちの少なくとも1つが、上記特定された業務において行われるデータ検索の際に参照されるサンプリング結果である。

発明の効果

[0007] サンプリング結果はサンプリング条件に依存している。サンプリング条件は、データ検索が行われることになる業務の特性に対応したサンプリングルールに従う条件である。このため、サンプリング結果を検索対象とした検索を含んだ業務の精度として高い精度が期待される。

図面の簡単な説明

- [0008] [図1]実施形態の概要の一例を示す。
- [図2]システム全体の構成例を示す。
- [図3]業務管理情報の構成例を示す。
- [図4]特性管理情報の構成例を示す。
- [図5]サンプリングインデクスの構成例を示す。
- [図6]ルール管理情報の構成例を示す。
- [図7]関係管理情報の構成例を示す。
- [図8]特性入力GUIの一例を示す。
- [図9]データ追加処理の流れの一例を示す。
- [図10]サンプリング処理の流れの一例を示す。
- [図11]サンプリング条件決定処理の流れの一例を示す。
- [図12]メンテナンス処理の流れの一例を示す。
- [図13]間引き処理の流れの一例を示す。

[図14]検索処理の流れの一例を示す。

[図15]サンプリングインデクスが指定された検索クエリの構成例を示す。

[図16]定義処理の流れの一例を示す。

[図17]サンプリングインデクスが指定された定義クエリの構成例を示す。

発明を実施するための形態

[0009] 以下の説明では、「データベース」が適宜「DB」と略記される。

[0010] また、以下の説明では、「bbb部」（又はbbbアクセラレータ）を主語として説明を行う場合があるが、これら機能部は、プロセッサによって実行されることで定められた処理をメモリ及び通信ポート（ネットワークI/F）を用いながら行うことができるため、プロセッサを主語とした説明としてもよい。プロセッサは、典型的には、マイクロプロセッサ（例えばCPU（Central Processing Unit））を含んでおり、更に、専用ハードウェア（例えばASIC（Application Specific Integrated Circuit）又はFPGA（Field-Programmable GateArray））を含んでもよい。また、これら機能部を主語として開示された処理は、計算機が行う処理としてもよい。また、これら機能部の一部又は全ては、専用ハードウェアによって実現されてもよい。また、各種機能部は、プログラム配布サーバや、計算機が読み取り可能な記憶メディアによって各計算機にインストールされてもよい。また、各種機能部及びサーバは1つの計算機にインストールされ実行されても良いし、複数の計算機にインストールされ実行されても良い。プロセッサは、制御部の一例であり、処理の一部又は全部を行うハードウェア回路を含んでもよい。プログラムは、プログラムソースから計算機のような装置にインストールされてもよい。プログラムソースは、例えば、プログラム配布サーバ又は計算機が読み取り可能な記憶メディアであってもよい。プログラムソースがプログラム配布サーバの場合、プログラム配布サーバはプロセッサ（例えばCPU）と記憶部を含み、記憶部はさらに配布プログラムと配布対象であるプログラムとを記憶してよい。そして、プログラム配布サーバのプロセッサが配布プログラムを実行することで、プログラム配布サーバのプロセッサは配布対

象のプログラムを他の計算機に配布してよい。また、以下の説明において、2以上のプログラムが1つのプログラムとして実現されてもよいし、1つのプログラムが2以上のプログラムとして実現されてもよい。

[0011] また、以下の説明において、「記憶部」は、メモリを含んだ1以上の記憶デバイスでよい。例えば、記憶部は、主記憶デバイス（典型的には揮発性のメモリ）及び補助記憶デバイス（典型的には不揮発性の記憶デバイス）のうちの少なくとも主記憶デバイスでよい。

[0012] また、以下の説明において、「母集団」は、データ群（1以上のデータ）である。「追加データ群」は、1以上の追加データである。「追加データ」は、母集団に追加されるデータである。「サンプリング」とは、母集団及び追加データ群のうちの少なくとも1つからデータを抽出することである。「サンプリングデータ」とは、母集団及び追加データ群のうちの少なくとも1つから抽出されたデータ、つまり、サンプリングされたデータである。「サンプリングデータ群」は、1以上のサンプリングデータである。

[0013] 以下、図面を参照しながら、一実施形態を説明する。

[0014] 図1は、実施形態の概要の一例を示す。

[0015] 表221（表名“Table1”）が、母集団の一例としてある。表221に、1以上の追加データ232である追加データ群が繰り返し（例えば定期的に又は不定期的に）追加される。追加データ群を構成する追加データ232の数は、一定でなくてよい。表221及び追加データ群のうちの少なくとも表221が、いわゆるビッグデータでよい。

[0016] 表221に、サンプリングインデクス223が関連付けられている。サンプリングインデクス223は、母集団全体ではなく、母集団の一部（サンプリングデータ群）のインデクスである。具体的には、例えば、サンプリングインデクス223は、サンプリングデータ群を構成する複数のサンプリングデータをそれぞれ指す複数のデータID（例えばポインタ）の集合でよい。サンプリングインデクス223のデータ構成比は、例えば、表221のデータ構成比となるべく同じになるよう維持される。「データ構成比」とは、複

数の値範囲にそれぞれ対応した複数のデータ比率であり、例えば、度数分布でよい。「値範囲」は、最小値と最大値とで構成された範囲でよいが、1つの値であってもよい。値範囲のデータ比率は、ソースデータ総数に対する対象データ数の比率である。「ソースデータ総数」とは、ソースデータ群を構成するデータの数（例えば、サンプリングインデクス223を構成するデータIDの数、又は、表221を構成するデータ数（例えばレコード数））である。値範囲の「対象データ数」は、その値範囲に属する値を含んだデータ（又はそのデータのデータID）の数である。

- [0017] 計算機が、データ追加処理（S101）、サンプリング処理（S102）、メンテナンス処理（S103）、及び、検索処理（S104）を行う。
- [0018] データ追加処理（S101）では、例えば次の処理が行われる。すなわち、計算機が、追加データ群を表221に追加する。そして、データ追加処理において、サンプリング処理（S102）及びメンテナンス処理（S103）が行われる。
- [0019] サンプリング処理（S102）は、例えば、データ取得（S102-1）、条件決定（S102-2）、及びデータサンプリング（S102-3）を含む。
- [0020] データ取得（S102-1）では、例えば次の処理が行われる。すなわち、計算機が、追加先の表221のサンプリングインデクス223のサンプリングインデクス名“S-Index1”に対応した業務ID“BID1”を、サンプリングインデクス名と業務IDの関係を表す業務管理情報224から取得する。また、計算機が、業務ID“BID1”に対応した特性詳細“PID1/Y”及び“PID2/N”を、業務IDと特性詳細の関係を表す特性管理情報225から取得する。
- [0021] 条件決定（S102-2）では、例えば次の処理が行われる。すなわち、計算機が、特性“PID1/Y”及び“PID2/N”にそれぞれ対応したサンプリングルール“ルールF”及び“ルールG”を、特性とサンプリングルールとの関係を表すルール管理情報226から取得する。また、計算機が、業務ID“B

ID1”の業務種別に対応した特性間関係“PID1 AND PID2”を、業務種別と特性間関係との関係を表す関係管理情報227から取得する。特性間関係は、特性間の関係であり、例えば、特性IDと論理演算子との組合せである。計算機が、特性“PID1/Y”及び“PID2/N”にそれぞれ対応したサンプリングルール“ルールF”及び“ルールG”を、特性間関係“PID1 AND PID2”に従い組み合わせることで、サンプリング条件“ルールF AND ルールG”を決定する。すなわち、特性間関係“PID1 AND PID2”における複数の特性IDの各々が、その特性IDを含んだ特性詳細に対応したサンプリングルールに差し替えられ、結果として、サンプリング条件“ルールF AND ルールG”が生成される。「サンプリング条件」とは、サンプリングされるデータの条件である。

- [0022] データサンプリング（S102-3）では、例えば次の処理が行われる。すなわち、計算機が、サンプリング条件“ルールF AND ルールG”に適合するデータを追加データ群からサンプリングする。
- [0023] メンテナンス処理（S103）では、例えば次の処理が行われる。すなわち、計算機が、表221に追加されるサンプリングデータ（追加データ群からサンプリングされたデータ）のデータID（例えばポインタ）を、表221のサンプリングインデクス223に追加する。つまり、計算機が、サンプリング処理（S102）の結果を基に、サンプリングデータの追加先の表221に関連付けられたサンプリングインデクス223をメンテナンスする。
- [0024] 以上で、サンプリング処理（S102）及びメンテナンス処理（S103）を含んだデータ追加処理（S101）が終了する。
- [0025] 検索処理（S104）では、例えば次の処理が行われる。すなわち、計算機が、クエリ発行元（例えばクライアント計算機）からの検索クエリで、サンプリングインデクス223が指定されているか否かを判断する。検索クエリに、サンプリングインデクス223を意味する記述（例えば、“QUICK S-Index1”）があれば、判断結果が肯定となる。判断結果が肯定の場合、計算機は、検索クエリで指定された検索条件に適合するデータを、その検索クエリ

で指定された表 2 2 1 のサンプリングインデクス 2 2 3 を通じて表 2 2 1 の一部（サンプリングデータ群）のみから探す。判断結果が否定の場合、計算機は、検索クエリで指定された検索条件に適合するデータを、その検索クエリで指定された表 2 2 1 全体から探す。

[0026] 以上が、本実施形態の概要である。

[0027] 本実施形態の適用範囲として様々な範囲が期待できる。例えば、所定地域の全世帯の平均消費電力量を算出する分析業務に本実施形態を適用することで、その分析業務を短時間で実行することが期待できる。具体的には、例えば、本実施形態に係る計算機が、電力会社の計算機に適用される。電力会社の計算機が、各世帯の配電盤に設置されているセンサから、第 1 周期で（例えば 1 分毎に）、消費電力量データ（消費電力量を表すデータ）を受信する。計算機が、受信した消費電力量データを、第 1 周期と同じ又は異なる第 2 周期で（例えば 5 分毎に）、DB の表に蓄積する。このため、表が、母グループの一例であり、表に格納される消費電力量データ群が、追加データ群の一例である。消費電力量データ群の量は、世帯数に依存するが、一般には大きい。消費電力量データ群の全てを用いて直近の平均消費電力量を算出するには、長い時間が必要とされる。そこで、表 2 2 1 の一部に対応したサンプリングインデクスが用意され、計算機が、消費電力量データ群の追加時に、上記の分析業務の特性に応じて消費電力量データ群からデータをサンプリングし、サンプリングされた消費電力量データを基に、サンプリングインデクスをメンテナンスする。分析業務における検索では、サンプリングインデクスが参照される。消費電力量データ群の追加の際にサンプリングインデクスが更新されるので、参照されるサンプリングインデクスは常に最新の状態が維持されていると期待される。また、サンプリングデータ群に代えてサンプリングインデクスが使用されるため、使用容量（消費される記憶容量）を抑えることができる。そして、サンプリングインデクスは、分析業務の特性に応じてサンプリングされたデータを基にメンテナンスされているので、サンプリングインデクスを使用した検索を含んだ業務の精度が高いことが期待される。

- 。
- [0028] なお、本実施形態によれば、追加データ群が、サンプリング候補データ群（1以上のサンプリング候補データ）の一例である。サンプリング候補データ群の一例は、追加データ群に代えて又は加えて、表221のような母集団が採用されてよい。
- [0029] また、サンプリングインデクス223が、サンプリング結果の一例である。サンプリング結果の一例は、サンプリングインデクス223に代えて、表221の一部のようなサンプリングデータ群が採用されてよい。すなわち、サンプリングインデクス223が無くてもよい。しかし、その場合、表221に加えてサンプリングデータ群が、記憶装置に格納されることになる。結果として、記憶装置の使用容量が大きい。この問題は、表221がビッグデータの場合には一層大きい。このため、本実施形態のように、サンプリング結果として、サンプリングデータ群ではなくサンプリングインデクス223が採用されることが好ましい。
- [0030] 以下、本実施形態を詳細に説明する。
- [0031] 図2は、システム全体の構成例を示す。
- [0032] 計算機209に、端末装置201と外部記憶装置220が接続されている。
- 。
- [0033] 端末装置201は、クライアントシステムの一例である。端末装置201は、I/F（インターフェースデバイス）11、入力デバイス202、表示デバイス203、メモリ12及びそれらに接続されたCPU207を有する。I/F11は、第1通信ネットワーク208を介して計算機209に接続される。入力デバイス202は、キーボード及びポインティングデバイス等でよい。表示デバイス203は、液晶ディスプレイ等でよい。入力デバイス202及び表示デバイス203は、タッチパネルのように一体でもよい。メモリ12は、OS（Operating System）206及び業務プログラム205を記憶する。CPU207が、それらのプログラムを実行する。業務プログラム205は、業務のために検索クエリ等のクエリを計算機209に発行する

- 。クエリは、例えば、SQL (Structured Query Language) で記述される。
- [0034] 計算機209は、計算機システムの一例である。計算機209は、F E - I / F (フロントエンドインターフェースデバイス) 210、B E - I / F (バックエンドインターフェースデバイス) 241、メモリ200、及びそれらに接続されたCPU219を有する。F E - I / F 210及びB E - I / F 241が、インターフェースデバイスの一例である。メモリ200が、記憶部の一例である。CPU219が、プロセッサの一例である。F E - I / F 210に、第1通信ネットワーク208を介して端末装置201が接続される。B E - I / F 241に、第2通信ネットワーク (図示せず) を介して外部記憶装置220が接続される。第1及び第2通信ネットワークのうちの少なくとも1つが無くてもよいし、それらが同一の通信ネットワークであってもよい。メモリ200は、OS211及びDBMS (Database Management System) 212を記憶する。CPU219が、それらのプログラムを実行する。
- [0035] DBMS212は、クエリ制御部213、検索部214、定義部215、データ追加部216、サンプリング部217及びメンテナンス部218を有する。クエリ制御部213は、クエリを受け付け、そのクエリを処理する。クエリの処理のために、クエリ制御部213は、必要に応じて、検索部214、定義部215、データ追加部216、サンプリング部217及びメンテナンス部218のうちの少なくとも1つを呼び出すことができる。検索部214は、サンプリングインデクス223を参照した検索を実行できる。定義部215は、サンプリングインデクス223の定義を実行できる。データ追加部216は、追加データ群の追加、具体的には、上述したデータ追加処理 (S101) を実行する。サンプリング部217は、追加データ群からのデータのサンプリング、具体的には、上述したサンプリング処理 (S102) を実行する。メンテナンス部218は、サンプリングインデクス223のメンテナンス、具体的には、上述したメンテナンス処理 (S103) を実行する。

[0036] 外部記憶装置 220 は、例えば、複数の記憶デバイスを有するストレージシステムでよい。複数の記憶デバイスにより、1以上のRAID (Redundant Array of Independent (or Inexpensive) Disks) グループが構成されてもよい。RAIDグループにおける各記憶デバイスは、不揮発性の記憶デバイスであり、例えば、HDD (Hard Disk Drive) であってもよいしSSD (Solid State Drive) であってもよい。外部記憶装置 220 は、DB 242、業務管理情報 224、特性管理情報 225、ルール管理情報 226、関係管理情報 227、DB管理情報 228、及び特性GUI (Graphical User Interface) 情報 230を記憶する。DB 242、業務管理情報 224、特性管理情報 225、ルール管理情報 226、関係管理情報 227、DB管理情報 228及び特性GUI情報 230のうちの少なくとも1つが、計算機 209のメモリ 200に記憶されてもよい。すなわち、例えば、計算機 209が、インメモリデータベースシステムでもよい。また、例えば、計算機 209と、外部記憶装置 220の少なくとも一部との組合せが、計算機システムの一例でもよい。また、例えば、外部記憶装置 220の少なくとも一部が、記憶部の一部でもよい。また、上述した異なる種類の情報が、同一の情報に統合されてもよいし、異なる異なる情報に分割されてもよい。例えば、特性管理情報 225 (図4参照) と関係管理情報 227 (図7参照) は1つに統合されていてもよい。

[0037] DB 242 は、表 221、通常インデクス 222、及びサンプリングインデクス 223を含む。表 221 は、典型的には、複数のカラムと複数のレコードで構成される。レコードの少なくとも一部が、表 221 を構成するデータの一例である。表 221 に、1以上の通常インデクス 222 が関連付けられている。また、少なくとも1つの表 221 に、1以上のサンプリングインデクス 223 が関連付けられている。表 221 に関連付けられている全ての通常インデクス 222 が、その表 221 全体を指す。表 221 に関連付けられている全てのサンプリングインデクス 223 が、表 221 の一部を指す。通常インデクス 222 は、サンプリングインデクス 223 以外のいずれかの

インデクスであり、典型的には一般的なインデクスである。サンプリングインデクス 2 2 3 は、サンプリングされたデータのデータ ID を有するインデクスである。通常インデクス 2 2 2 及びサンプリングインデクス 2 2 3 のいずれも、データ構造は限定されない。すなわち、通常インデクス 2 2 2 及びサンプリングインデクス 2 2 3 のいずれも、リスト構造でもよいし、B 木構造でもよい（図 5 が、サンプリングインデクス 2 2 3 の構成例を示す）。

[0038] 業務管理情報 2 2 4、特性管理情報 2 2 5、ルール管理情報 2 2 6、関係管理情報 2 2 7 及び特性 GUI 情報 2 3 0 については、後述する。業務管理情報 2 2 4、特性管理情報 2 2 5、ルール管理情報 2 2 6 及び関係管理情報 2 2 7 を含んだ情報が、管理情報の一例である。管理情報に、特性 GUI 情報 2 3 0 が含まれてもよい。管理情報に、更に、DB 管理情報 2 2 8 が含まれてもよい。

[0039] DB 管理情報 2 2 8 は、DB 2 4 2 に関する情報を含む。例えば、DB 管理情報 2 2 8 は、サンプリングインデクス 2 2 3 毎にサンプリングインデクス 2 2 3 が持つデータ ID 数（例えばポインタ数）を表す情報と、表 2 2 1 とインデクス（通常インデクス 2 2 2 及びサンプリングインデクス 2 2 3）との関係を表す情報であるインデクス管理情報とを含んでよい。

[0040] 図 3 は、業務管理情報 2 2 4 の構成例を示す。

[0041] 業務管理情報 2 2 4 は、サンプリングインデクス 2 2 3 と業務との関係を表す。具体的には、例えば、業務管理情報 2 2 4 は、サンプリングインデクス 2 2 3 にそれぞれ対応したエントリを有する。各エントリが、サンプリングインデクス名 3 0 1 及び業務 ID 3 0 2 という情報を含む。サンプリングインデクス名 3 0 1 は、サンプリングインデクス 2 2 3 の名称（他種の ID でもよい）を表す。業務 ID 3 0 2 は、業務の ID である。

[0042] 図 4 は、特性管理情報 2 2 5 の構成例を示す。

[0043] 特性管理情報 2 2 5 は、業務と業務の特性との関係を表す。具体的には、例えば、特性管理情報 2 2 5 は、業務にそれぞれ対応したエントリを有する。各エントリが、業務 ID 4 0 1、特性 ID 4 0 2 及び入力値 4 0 3 という

情報を含む。特性ID402及び入力値403の組が、特性詳細である。特性ID402が、特性のIDである。入力値403が、特性に関して入力された値である。エントリには、対応する業務に属する全ての特性詳細が含まれている。言い換えれば、業務は、1以上の特性詳細で表現されてよい。1つの業務につき、1以上の特性がある。

[0044] 図6は、ルール管理情報226の構成例を示す。

[0045] ルール管理情報226は、特性（特性詳細）とサンプリングルールとの関係を表す。具体的には、例えば、ルール管理情報226は、特性にそれぞれ対応したエントリを有する。各エントリが、特性ID601、入力値602、サンプリングルール603、外部入力604及び取得元605という情報を含む。サンプリングルール603は、データのサンプリングのルールを表す。サンプリングルール603において、「データ生成ID」は、データを生成する主体を一意に特定するID（例えば、センサデータを生成するセンサを一意に特定するセンサIDなど）であり、「サンプリング割合」は、追加データ総数（追加データ群を構成する追加データ232の数）に対するサンプリング追加データ数（サンプリングされた追加データ232の数）の割合である。外部入力604は、サンプリング条件決定の際にサンプリング部217（DBMS212）の外部から入力される値を表す。取得元605は、外部入力604が表す値の取得元605（入力元）を表す。

[0046] 図6の例によれば、特性詳細とサンプリングルール603が1対1で対応している。図6の例によれば、例えば次のことが言える。特性詳細“PID4/Y”に対応したサンプリングルール603によれば、サンプリングインデクス223の構成比が、そのサンプリングインデクス223が関連付けられた表221の構成比となるべく同じになるよう、データがサンプリングされる（サンプリングインデクス223がメンテナンスされる）ことになる。また、特性詳細“PID5/Y”に対応したサンプリングルール603、外部入力604及び取得元605によれば、追加データ232の優先度がユーザにより入力され、その入力された優先度が高いデータほど優先的にサンプリングされる

ことになる。なお、本実施形態で言う「ユーザ」は、端末装置 201 のユーザであるが、計算機 209 の管理者であってもよい。また、取得元 605 の「外部記憶」は、外部記憶装置 220 を意味する。

[0047] 図 6 の例によれば、特性毎に、特性詳細が“Y”又は“N”の 2 つに分類されるが、少なくとも 1 つの特性について、特性詳細の数は、2 に限られず、1 でも 3 以上の整数でもよい。

[0048] 図 7 は、関係管理情報 227 の構成例を示す。

[0049] 関係管理情報 227 は、業務と特性間関係との関係を表す。具体的には、例えば、関係管理情報 227 は、業務種別にそれぞれ対応したエントリを有する。各エントリは、業務種別 701、業務 ID 702、特性 ID 703 及び特性間関係 704 という情報を含む。業務種別 701 は、業務種別 701 を表す。業務 ID 702 は、対応する業務種別 701 に属する業務の ID である。特性 ID 703 は、対応する業務種別 701 に属する特性の ID である。特性間関係 704 は、対応する業務種別 701 に属する特性間の関係を表す。

[0050] 図 7 の例によれば、例えば次のことが言える。すなわち、業務種別によって、業務を構成する特性と特性間の関係が異なる。特性 ID が 1 つであれば、特性間関係 704 は、その 1 つの特性 ID である。一方、特性 ID が複数あれば、特性間関係 704 は、1 以上の論理演算子（例えば、“AND”又は“OR”）で直列に連結された複数の特性 ID である。特性間関係 704 は、論理演算子による複数の特性 ID の連結に限らないが、特性間関係 704 のこのような構成によれば、論理演算子及び特性 ID のような要素を追加、削除又は変更するだけで特性間関係 704 の変更が済むので、利便性が高いと考えられる。

[0051] 図 7 の例によれば、例えば次のことが言える。すなわち、業務 ID “BID2” が特定された場合、DBMS 212 が、関係管理情報 227 から、業務 ID “BID2” が属する業務種別 “センサデータ分析” に対応した特性間関係 “PID1 AND PID2 AND PID5 AND PID6” を特定する。また、DBMS 212 が、

業務ID “PID2” に対応する特性詳細 “PID1/Y” 、 “PID2/Y” 、 “PID5/Y” 及び “PID6/Y” を、特性管理情報 225 から特定したとする。この場合、サンプリング条件は、“PID1/Yに対応したサンプリングルール AND PID2/Yに対応したサンプリングルール AND PID5/Yに対応したサンプリングルール AND PID6/Yに対応したサンプリングルール” となる。つまり、特性間関係における各特性IDが、その特性IDを含んだ特性詳細に対応するサンプリングルール（ルール管理情報 226 から特定されたサンプリングルール 603）に差し替えられる。結果として、サンプリング条件は、特性間関係に従うサンプリングルール集合である。

[0052] 図8は、特性入力GUIの一例を示す。

[0053] 特性入力GUI 800は、そのGUIに対応する特性GUI情報 230に基づき生成され表示されたGUIである。特性入力GUI 800は、業務の特性の入力を受け付けるUI（ユーザインタフェース）の一例である。特性入力GUI 800は、例えば、DBMS 212により端末装置 201に提供され、端末装置 201の表示デバイス 203に表示される。特性入力GUI 800は、計算機 209の管理者が使用する端末装置 201に提供され表示されてもよい。

[0054] 特性入力GUI 800（特性GUI情報 230）によれば、特性詳細サブUI 802が表示される。そのサブUI 802によれば、少なくとも1つの特性について、特性IDと、特性の内容と、入力値と、補足事項とが表示される。入力値について、サブUIがある。入力値についてのサブUIは、選択可能な複数の入力値にそれぞれ対応した複数のラジオボタンであってよい。補足事項は、少なくとも1つの特性についての入力可否を提供することの一例である。すなわち、補足事項は、その補足事項に対応する特性について、ユーザが指定する業務種別に応じた入力可否を表す。なお、補足事項は、更に、具体例（例えば、特性ID “PID4” について、“男性のデータ数20%、女性のデータ数80%という構成比を維持する”）を表す文章を含んでもよい。

[0055] 特性入力GUI 800は、ユーザから業務種別を受け付けるサブUI 80

1を有してもよい。DBMS 212は、サブUI 801に業務種別が入力された場合、入力必須の特性についてのみ（例えば、入力された業務種別に対応した特性IDについてのみ）、入力をイネーブルとし、入力不要の特性については（例えば、入力された業務種別に対応していない特性IDについては）、入力をディスエーブルとしてもよい（例えば自動的に「入力なし」が選択されてもよい）。入力の自動的なイネーブル／ディスエーブルも、少なくとも1つの特性についての入力可否を提供することの一例でよい。

[0056] 以下、本実施形態で行われる処理の一例を説明する。

[0057] 図9は、データ追加処理（S101）の流れの一例を示す。

[0058] データ追加部216が、追加データ群を表221に追加する（S901）。データ追加部216が、DB管理情報228内のインデクス管理情報を参照し、追加先の表221に関連付けられている全てのサンプリングインデクス名を取得する（S902）。取得されたサンプリングインデクス名毎に（サンプリングインデクス223毎に）、サンプリング処理（S102）及びメンテナンス処理（S103）が行われる。

[0059] 以下、サンプリング処理（S102）及びメンテナンス処理（S103）を説明する。その際、1つのサンプリングインデクス名に対応したサンプリングインデクス223を例に取る。例として採用されたサンプリングインデクス223を、図10～図13の説明において、「対象サンプリングインデクス223」と呼ぶ。

[0060] 図10は、サンプリング処理（S102）の流れの一例を示す。サンプリング処理（S102）は、例えば、データ追加処理（S101）においてデータ追加部216によりサンプリング部217が呼び出されることにより行われてよい。

[0061] サンプリング部217は、業務管理情報224を参照し、対象サンプリングインデクス名301に関連付いた業務ID302を取得する（S1001）。

[0062] サンプリング部217は、特性管理情報225を参照し、S1001で取

得した業務ID302に関連付いた特性詳細（特性ID402と入力値403との組）を取得する（S1002）。

[0063] サンプリング部217は、S1002で取得された特性詳細を用いてサンプリング条件決定処理を実行する（S1003）。

[0064] サンプリング部217は、追加データ群に含まれる追加データ毎に、次の処理を行う。すなわち、サンプリング部217は、S1003で決定されたサンプリング条件に、追加データが適合するか否かを判断する（S1011）。S1011の判断結果が肯定の場合、サンプリング部217は、その追加データを追加データ群からサンプリングする（S1012）。S1012でサンプリングされた追加データが、サンプリングデータである。

[0065] 図11は、サンプリング条件決定処理（S1003）の流れの一例を示す。

[0066] サンプリング部217は、ルール管理情報226を参照し、S1002で取得された特性詳細にそれぞれ関連付いたサンプリングルール603を取得する（S1101）。

[0067] サンプリング部217は、関係管理情報227を参照し、S1001で取得された業務IDが属する業務種別701に関連付いた特性間関係704を取得する（S1102）。

[0068] サンプリング部217は、S1101で取得したサンプリングルール603を、S1102で取得した特性間関係704に従って組み合わせることで、サンプリング条件を生成（決定）する（S1103）。具体的には、例えば、S1102で取得した特性間関係704における各特性IDが、S1101で取得したサンプリングルール603のうちの、その特性IDを含んだ特性詳細に関連付いたサンプリングルール603に差し替えられる。その結果が、サンプリング条件である。

[0069] 図12は、メンテナンス処理（S103）の流れの一例を示す。メンテナンス処理（S103）は、例えば、データ追加処理（S101）においてデータ追加部216によりメンテナンス部218が呼び出されることにより行

われてよい。

- [0070] メンテナンス部 218 は、サンプリングされたデータ毎に、データ ID (例えばポインタ) を取得する (S1201)。メンテナンス部 218 は、S1201 で取得したデータ ID を、対象サンプリングインデクス 223 に追加する (S1202)。そして、メンテナンス部 218 は、間引き処理を実行する (S1203)。
- [0071] 図 13 は、間引き処理 (S1203) の流れの一例を示す。
- [0072] メンテナンス部 218 は、S1202 後の対象サンプリングインデクス 223 が持つデータ ID の数が一定数より多いか否かを判断する (S1301)。対象サンプリングインデクス 223 が持つデータ ID の数は、例えば DB 管理情報 228 から特定されてもよいし、対象サンプリングインデクス 223 をスキャンすることで特定されてもよい。
- [0073] S1301 の判断結果が否定の場合 (S1301: NO)、対象サンプリングインデクス 223 からデータ ID を削除する必要は無いため、間引き処理が終了する。
- [0074] 一方、S1301 の判断結果が肯定の場合 (S1301: YES)、メンテナンス部 218 は、超過数分のデータ ID を対象サンプリングインデクス 223 から削除することで、対象サンプリングインデクス 223 が持つデータ ID の数を、一定数以下にする (S1302)。メンテナンス部 218 は、DB 管理情報 228 のうち、対象サンプリングインデクス 223 が持つデータ ID の数を表す情報を、S1302 後のデータ ID 数に更新する (S1303)。
- [0075] S1302 において削除されるデータ ID 数は、超過数分のデータ ID より多くのデータ ID が削除されてもよい。
- [0076] また、S1302 では、削除対象候補及び削除条件のうち少なくとも 1 つについて幾つかのバリエーションがあり得る。
- [0077] 例えば、削除対象候補 (削除され得るデータ ID) は、対象サンプリングインデクス 223 におけるいずれのデータ ID であってよい。また、対象サ

ンプリングインデクス 223 のうち、直前の S 1203 で追加されたデータ ID (すなわち、最新のサンプリング条件に適合する追加データ (サンプリングデータ) のデータ ID) は、削除対象候補から除外されてもよい。

[0078] 一方、削除条件は、例えば、最新のサンプリング条件に適合しなかったデータを指すデータ ID であってもよいし、古いデータ ID であってもよい。より具体的には、例えば、メンテナンス部 218 は、最新のサンプリング条件に適合しないデータのデータ ID を対象サンプリングインデクス 223 から優先的に削除し、それでも未だ対象サンプリングインデクス 223 のデータ ID 数が一定数を超過していれば、古いデータ ID ほど優先的に対象サンプリングインデクス 223 から削除してよい。最新のサンプリング条件に適合しないデータのデータ ID が優先的に削除されるので、結果として、最新のサンプリングデータのデータ ID が自動的に削除対象から除外されると共に、最新のサンプリング条件に適合する既存データのデータ ID を優先的に残すことができる。

[0079] なお、「最新のサンプリング条件」とは、今回のデータ追加処理 (S 101) のサンプリング処理 (S 102) において決定されたサンプリング条件である。特性間関係 704 とサンプリングルール 603 のうちの少なくとも 1 つが、データ追加処理間に変更されることがある。そのため、業務が同じであっても、決定されるサンプリング条件が毎回同じとは限らない。そこで、削除対象候補及び削除条件のうち少なくとも 1 つにおいて、「最新のサンプリング条件」が過去のサンプリング条件よりも優先されてよい。

[0080] 図 14 は、検索処理の流れの一例を示す。

[0081] 例えば、端末装置 201 から計算機 209 にクエリが入力されると、クエリ制御部 213 が、そのクエリを解析する。入力されたクエリが検索クエリの場合は、クエリ制御部 213 が検索部 214 を呼び出し、それにより、検索処理が開始される。検索処理では、検索クエリで指定された検索条件に適合したデータが結果集合として取得され、検索結果が端末装置 201 に返される。検索結果は、取得された結果集合それ自体でもよいし、取得された結

果集合が分類又は集約された結果でもよい。

[0082] 検索部 214 は、検索クエリにサンプリングインデクス 223 を使用することを示すヒントが指定されているか否かを判断する (S1401)。ここで言う「ヒント」は、サンプリングインデクス 223 の使用を意味する記述の一例である。図 15 が、検索クエリの構成例を示すが、その例によれば、ヒントとして、「QUICK S-Index1」がある。検索クエリでは、使用されるサンプリングインデクス 223 が個々に指定されていてもよいし、サンプリングインデクス 223 の使用だけが指定されていてもよい。後者の場合、検索対象の表 221 に関連付いた全てのサンプリングインデクス 223 が使用されることになる。検索クエリは、業務プログラム 205 のようなクエリ発行元により生成されたクエリである。

[0083] S1401 の判断結果が肯定の場合 (S1401: YES)、検索部 214 は、検索クエリからサンプリングインデクス 223 を特定し (S1402)、特定したサンプリングインデクス 223 を通じて、検索クエリで指定された検索条件に適合するデータを結果集合として取得する (S1403)。具体的には、例えば、検索部 214 は、特定したサンプリングインデクス 223 のデータ ID 毎に、そのデータ ID から特定されるデータ (表 221 におけるデータ) を参照し、そのデータが検索条件に適合するか否かを判断する。検索条件に適合するデータの集合が、結果集合である。検索部 214 は、結果集合を返す (S1405)。

[0084] S1401 の判断結果が否定の場合 (S1401: NO)、検索部 214 は、検索クエリで指定された検索条件に適合するデータを、検索クエリから特定された表 221 に関連付いている通常インデクス 222 を使用して、結果集合として取得する (S1404)。検索部 214 は、結果集合を返す (S1405)。

[0085] 図 16 は、定義処理の流れの一例を示す。

[0086] 例えば、端末装置 201 から計算機 209 にクエリが入力されると、クエリ制御部 213 が、そのクエリを解析する。入力されたクエリが定義クエリ

の場合は、クエリ制御部 213 が定義部 215 を呼び出し、それにより、定義処理が開始される。

[0087] 定義部 215 は、定義クエリに記述されている表 221 名、インデクス名及びインデクス種別を、DB 管理情報 228（例えばインデクス管理情報）に追加する（S1601）。

[0088] 定義部 215 は、定義クエリに、サンプリングインデクス 223 の定義であることを示すヒントが指定されているか否かを判断する（S1602）。ヒントは、例えば図 17 に示す定義クエリの一例によれば、“ON Table1 QUICK” である。

[0089] S1602 の判断結果が肯定の場合（S1602: YES）、定義部 215 は、端末装置 201 に、特性入力 GUI 800 を提供（表示）する（S1603）。

[0090] 定義部 215 は、提供された特性入力 GUI 800 経由で、業務の特性（入力値）を受け（S1604）、業務 ID と特性詳細（例えば「入力なし」以外の入力値が入力された特性）とを、特性管理情報 225 に登録する（S1605）。業務 ID 702 は、定義クエリで指定されていてもよいし、定義部 215 により自動で割り振られた ID でもよい。定義部 215 は、業務 ID とサンプリングインデクス名（定義クエリで指定されたインデクス名）とを、業務管理情報 224 に登録する（S1606）。

[0091] 以下、実施形態を総括する。なお、総括の説明において、適宜、実施形態の変形例等の新たな記載を追加することができる。

[0092] 計算機 209 が、データ検索を含んだ業務と業務の特性とサンプリングルールとの関係を表す情報を含んだ情報である管理情報を記憶するメモリ 200 と、メモリ 200 に接続された CPU 219 とを有する。CPU 219 が、（A）追加データ群の追加先の表 221 に関連付けられた業務を特定し、（B）管理情報を基に、（A）で特定した業務の特性に対応したサンプリングルールに従うサンプリング条件を決定し、（C）（B）で決定したサンプリング条件に適合するデータを追加データ群からサンプリングする。サンプ

リングされたデータのインデクスであるサンプリングインデクス 223 が、上記特定された業務において行われるデータ検索の際に参照される。サンプリングインデクス 223 は、サンプリング条件に依存している。サンプリング条件は、データ検索が行われることになる業務の特性に対応したサンプリングルールに従う条件である。このため、サンプリングインデクス 223 を検索対象とした検索を含んだ業務（例えば分析）の精度として高い精度が期待される。また、追加データ群からサンプリングされるため、サンプリング処理において参照されるデータ総量は、表 221 からサンプリングすることに比べれば典型的には小さくて済む。また、データ検索では、サンプリングインデクス 223 が参照されるため、表 221 の一部であるサンプリングデータ群の複製を格納することに比べて、使用容量を抑えることができる。

[0093] CPU 219 は、(D) (C) でサンプリングされたデータを指すデータをサンプリングインデクス 223 に追加する処理を含んだメンテナンスを実行する。これにより、サンプリングの都度にサンプリングインデクス 223 を最新の状態にすることができる。

[0094] CPU 219 は、追加データ群が表 221 に追加されるときに、(A) 乃至 (D) を実行する。データ追加処理のときにサンプリング及びメンテナンスが行われるので、サンプリングインデクス 223 を最新の状態に維持することが期待できる。

[0095] 表 221 及び追加データ群のうちの少なくとも表 221 は、ビッグデータである。上述の実施形態は、表 221 のサイズが大きいほど効果的であると考えられる。

[0096] (B) において、CPU 219 は、(b1) 特定した業務の特性にそれぞれ対応したサンプリングルールと、特定した業務の特性間関係を管理情報を基に特定し、(b2) 特定した特性間関係に従い、特定したサンプリングルールを組み合わせる。特定したサンプリンググループの組合せが、サンプリング条件である。これにより、業務の特性間関係に最適なサンプリング条件が期待できる。

- [0097] 特性間関係は、特性と論理演算子との組合せである。このため、特性及び論理演算子の少なくとも1つを変更するだけで特性間関係の構成を変更でき、利便性が高い。
- [0098] 特定したサンプリングルールの組合せは、特定した特性間関係に含まれる特性にそれぞれ対応したサンプリングルールと、特性間関係に含まれる論理演算子との組合せである。特性毎にサンプリングルールが用意されていれば、特性間関係に従い、特性間関係における特性をその特性に対応したサンプリングルールに差し替えるという簡単な処理で、サンプリング条件を生成することができる。
- [0099] CPU 219が、業務の特性の入力を受け付ける特性入力GUI 800を提供する。特性入力GUI 800が、少なくとも1つの特性について、ユーザが指定する業務種別に応じた入力可否を提供する。ユーザは、入力値の入力が必要とされる特性をすぐに特定でき、特定された特性について、特性の入力値を入力できる。
- [0100] 特定された業務の少なくとも1つの特性は、サンプリングインデクス223が表すデータ構成比が、表221のデータ構成比と同じになることである。これにより、サンプリングインデクス223を使用した検索を含む業務の精度が、表221全体を参照する検索を含んだ業務の精度に近い（例えば同等である）ことが期待できる。
- [0101] CPU 219は、(E) (D)のメンテナンス後のサンプリングインデクス223を構成するデータIDの数が一定数を超過している場合、サンプリングインデクス223のデータID数が一定数以下になるようサンプリングインデクス223からデータIDを削除する。これにより、サンプリングインデクス223のサイズを一定サイズ以下に抑えることができる。
- [0102] (E)において、CPU 219は、(e1) (B)で決定されたサンプリング条件に適合しないデータのデータIDをサンプリングインデクス223から優先的に削除し、(e2) (e1)の後のサンプリングインデクス223のデータID数が未だ一定数を超過している場合、古いデータIDほど優先

的にサンプリングインデクス 2 2 3 から削除する。最新のサンプリング条件に適合しないデータのデータ I D が優先的に削除されるので、結果として、最新のサンプリングデータのデータ I D が自動的に削除対象から除外されると共に、最新のサンプリング条件に適合する既存データのデータ I D を優先的にサンプリングインデクス 2 2 3 に残すことができる。

[0103] 以上、一実施形態を説明したが、本発明は、この実施形態に限定されるものでなく、その要旨を逸脱しない範囲で種々変更可能であることはいうまでもない。例えば、母集団は、データベースの表 2 2 1 に限定されない。

符号の説明

[0104] 2 0 9 … 計算機

請求の範囲

- [請求項1] データ検索を含んだ業務と業務の特性とサンプリングルールとの関係を表す情報を含んだ情報である管理情報を記憶する記憶部と、
前記記憶部に接続されたプロセッサと
を有し、
前記プロセッサが、
(A) 1以上のサンプリング候補データであるサンプリング候補データ群に関連付けられた業務を特定し、
(B) 前記管理情報を基に、前記特定した業務の特性に対応したサンプリングルールに従うサンプリング条件を決定し、
(C) 前記決定したサンプリング条件に適合するデータを前記サンプリング候補データ群からサンプリングし、
前記サンプリングされたデータを含んだ1以上のサンプリングデータであるサンプリングデータ群と前記サンプリングデータ群のインデクスであるサンプリングインデクスとのうちの少なくとも1つが、前記特定された業務において行われるデータ検索の際に参照されるサンプリング結果である、
計算機システム。
- [請求項2] 前記サンプリング候補データ群は、1以上のデータである母集団に追加される1以上のデータであり、
前記サンプリング候補データ群に関連付けられた業務は、前記サンプリング候補データ群の追加先の母集団に関連付けられた業務である、
請求項1記載の計算機システム。
- [請求項3] 前記サンプリング結果は、前記サンプリングインデクスであり、
前記サンプリングインデクスは、前記母集団のうちのサンプリングデータ群のインデクスである、
請求項2記載の計算機システム。

- [請求項4] 前記プロセッサは、
(D) (C) でサンプリングされたデータを指すデータを前記サンプリングインデクスに追加する処理を含んだメンテナンスを実行する、
請求項3記載の計算機システム。
- [請求項5] 前記プロセッサは、前記サンプリング候補データ群が前記母集団に追加されるときに、(A)乃至(D)を実行する、
請求項4記載の計算機システム。
- [請求項6] 前記母集団及び前記サンプリング候補データ群のうちの少なくとも前記母集団は、ビッグデータである、
請求項5記載の計算機システム。
- [請求項7] 前記プロセッサは、(B)において、
(b1) 前記管理情報を基に、前記特定した業務の特性にそれぞれ対応したサンプリングルールと、前記特定した業務の特性間との関係とを特定し、
(b2) 前記特定した特性間関係に従い前記特定したサンプリングルールを組み合わせ、
前記特定したサンプリンググループの組合せが、前記サンプリング条件である、
請求項1記載の計算機システム。
- [請求項8] 前記特性間関係は、特性と論理演算子との組合せである、
請求項7記載の計算機システム。
- [請求項9] 前記特定したサンプリングルールの組合せは、前記特定した特性間関係に含まれる特性にそれぞれ対応したサンプリングルールと、前記特性間関係に含まれる論理演算子との組合せである、
請求項8記載の計算機システム。
- [請求項10] 前記管理情報が、業務種別と特性間関係との関係を表す情報を含み、

前記プロセッサが、業務の特性の入力を受け付けるユーザインタフェースを提供し、

前記ユーザインタフェースが、少なくとも1つの特性について、ユーザが指定する業務種別に応じた入力可否を提供する、
請求項7記載の計算機システム。

[請求項11] 前記特定された業務の少なくとも1つの特性は、前記サンプリング結果が表すデータ構成比が、複数のデータである母集団のデータ構成比と同じになることである、
請求項1記載の計算機システム。

[請求項12] 前記サンプリング結果は、前記サンプリングインデクスであり、
前記サンプリングインデクスは、前記母集団のうちのサンプリングされたデータのインデクスである、
請求項1記載の計算機システム。

[請求項13] 前記プロセッサは、
(D) (C)の結果を基に前記サンプリング結果を更新するメンテナンスを実行し、
(E) (D)のメンテナンス後のサンプリング結果を構成するデータ数が一定数を超えている場合、前記サンプリング結果を構成するデータ数が前記一定数以下になるよう前記サンプリング結果からデータを削除する、
請求項1記載の計算機システム。

[請求項14] 前記サンプリング候補データ群は、1以上のデータである母集団に追加される1以上のデータであり、
前記サンプリング候補データ群に関連付けられた業務は、前記サンプリング候補データ群の追加先の母集団に関連付けられた業務であり、
前記プロセッサは、(E)において、

(e1) (B)で決定されたサンプリング条件に適合しないデー

タを前記サンプリング結果から優先的に削除し、

(e 2) (e 1) の後のサンプリング結果を構成するデータ数が未だ一定数を超えている場合、古いデータほど優先的に前記サンプリング結果から削除する、

請求項 1 3 記載の計算機システム。

[請求項15]

(A) 1 以上のサンプリング候補データであるサンプリング候補データ群に関連付けられた業務を特定し、

(B) データ検索を含んだ業務と業務の特性とサンプリングルールとの関係を表す情報を含んだ情報である管理情報を基に、前記特定した業務の特性に対応したサンプリングルールに従うサンプリング条件を決定し、

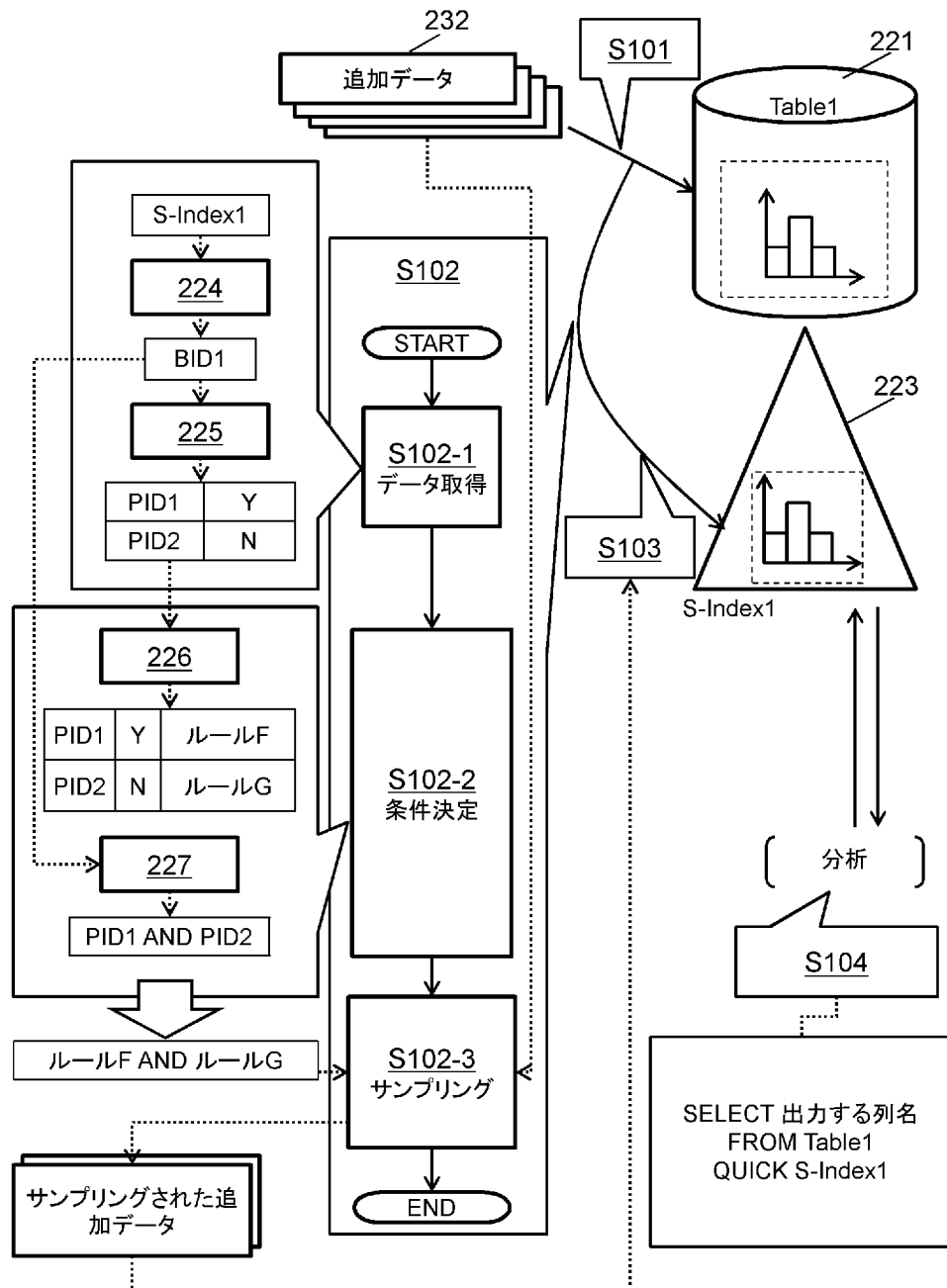
(C) 前記決定したサンプリング条件に適合するデータを前記サンプリング候補データ群からサンプリングし、

前記サンプリングされたデータを含んだ 1 以上のサンプリングデータであるサンプリングデータ群と前記サンプリングデータ群のインデクスであるサンプリングインデクスとのうちの少なくとも 1 つが、前記特定された業務において行われるデータ検索の際に参照されるであるサンプリング結果である、

サンプリング方法。

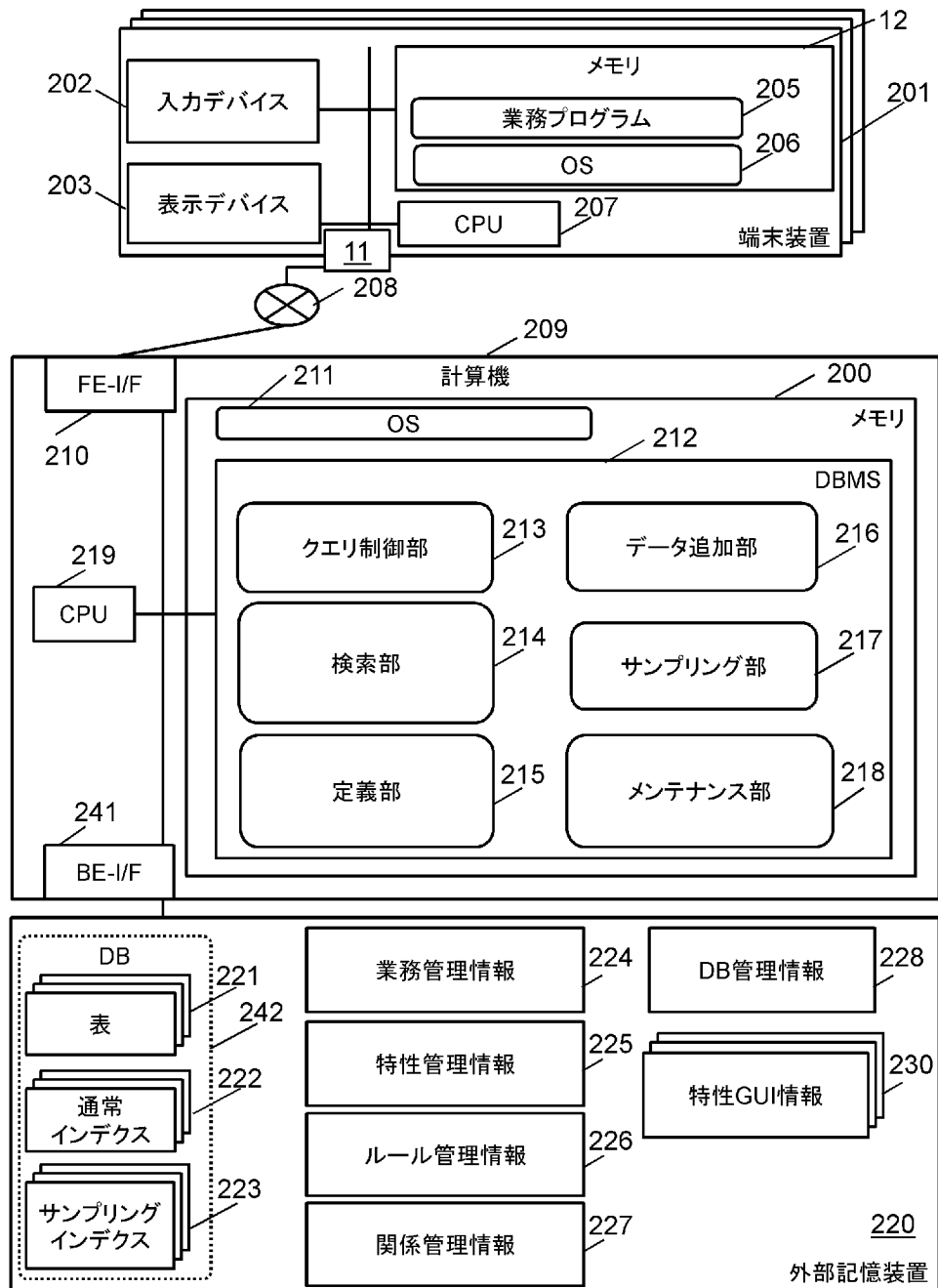
[図1]

FIG. 1



[図2]

FIG. 2



[図3]

FIG. 3業務管理情報
224

301	302
サンプリングインデクス名	業務ID
S-Index1	BID1
...	...

[図4]

FIG. 4特性管理情報
225

401	402	403
業務ID	特性ID	入力値
BID1	PID1	Y
	PID2	Y
	PID5	N
	PID6	N
...

[図5]

FIG. 5サンプリングインデクス
223

データID (ポインタ)
xxx
...

[図6]

FIG. 6

ルール管理情報
226

特性ID	601 入力値	602 サンプリングルール	603 外部入力	604 取得元
PID1	Y	データを静的情報でm個のグループに分ける。グループ毎にサンプリング。	静的情報の列名	ユーザ
	N	データを静的情報で分けない。	-	-
PID2	Y	データ生成IDが同じデータは、全てサンプリング又は全てサンプリングしない。	データ生成IDの列名	ユーザ
	N	サンプリングされるデータは、データ生成IDに依存しない。	-	-
PID3	Y	データを列の値でグループに分ける。グループ毎にサンプリング。	分割する列名	ユーザ
	N	データを列の値で分けない。	-	-
PID4	Y	各グループは、グループ内データ数と、全データ数に対するグループ内データ数の比率とを持つ。	-	-
	N	各グループは、グループ内データ数と、全データ数に対するグループ内データ数の比率とを持たない。	-	-
PID5	Y	優先度の高いデータの順にサンプリング。	データの優先度	ユーザ
	N	無作為にサンプリング。	-	-
PID6	Y	抽出する件数 = $\frac{[(追加データ数) \times (サンプリングインデクスのデータID数)]}{(表データ数)}$ (もし、サンプリングインデクスのデータID数、又は、表データ数が、0の場合、サンプリング割合は、ユーザにより指定された割合である。)	サンプリングインデクスのデータID数、及び、表データ数	外部記憶
	N	サンプリング割合は、ユーザにより指定された割合である。	サンプリング割合	ユーザ

[図7]

FIG. 7

関係管理情報
227

701 業務種別	702 業務ID	703 特性ID	704 特性間関係
センサデータ分析	BID2 BID5	PID1	PID1 AND PID2 AND PID5 AND PID6
		PID2	
		PID5	
		PID6	
行動履歴分析	BID4	PID2	PID2 AND PID3 AND PID4 AND PID5 AND PID6
		PID3	
		PID4	
		PID5	
		PID6	
...

[図8]

FIG. 8

800

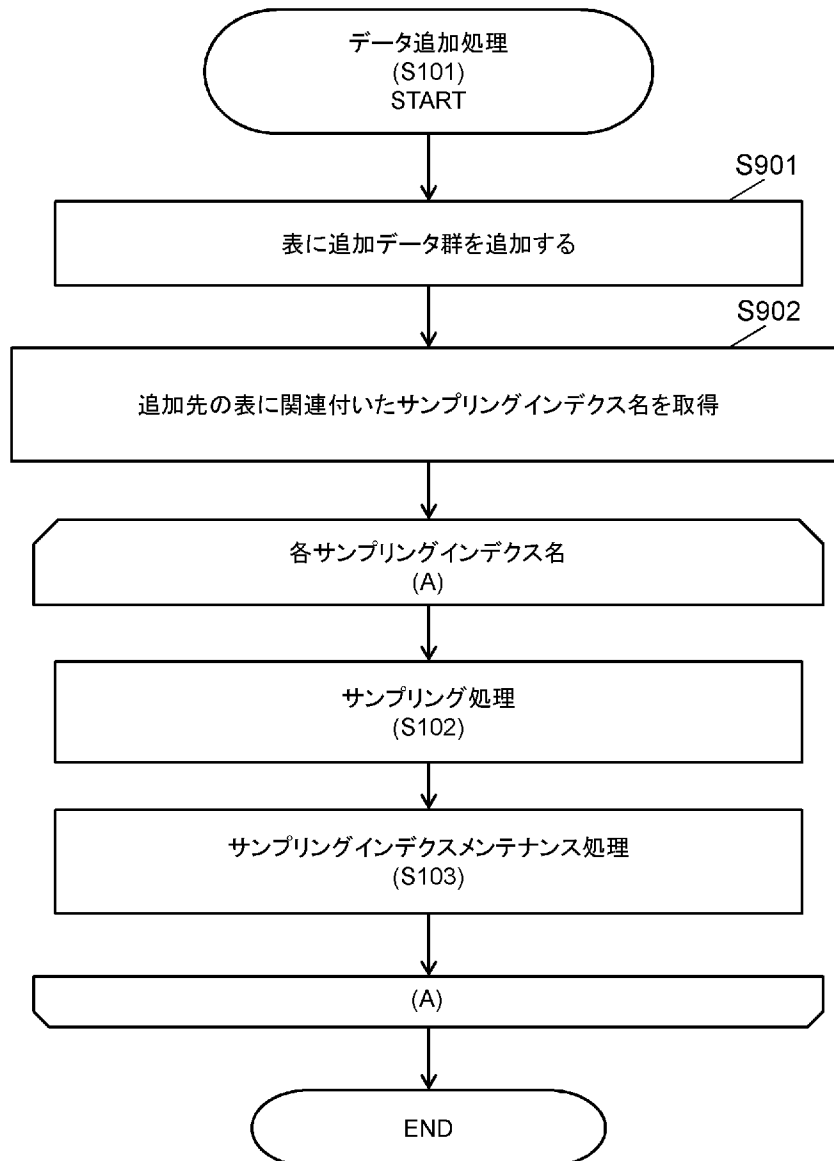
ID	特性	入力形式	補足事項
PID1	データは静的情報を持つ	<input type="radio"/> YES <input type="radio"/> NO <input checked="" type="radio"/> 入力なし	実施業務がセンサデータ分析の場合、本入力必須。
PID2	同じ主体が複数のデータを生成する	<input type="radio"/> YES <input type="radio"/> NO <input checked="" type="radio"/> 入力なし	実施業務が、センサデータ分析、又は、行動履歴分析の場合、本入力必須。
PID3	列の値で分割する	<input type="radio"/> YES <input type="radio"/> NO <input checked="" type="radio"/> 入力なし	実施業務が、行動履歴分析の場合、本入力必須。
PID4	データの構成比を維持する	<input type="radio"/> YES <input type="radio"/> NO <input checked="" type="radio"/> 入力なし	実施業務が、センサデータ分析、又は、行動履歴分析の場合、本入力必須。
PID5	データに優先度がある	<input type="radio"/> YES <input type="radio"/> NO <input checked="" type="radio"/> 入力なし	実施業務が、センサデータ分析、又は、行動履歴分析の場合、本入力必須。
PID6	表データ数に対するとサンプリングインデックスのデータID数の比率が一定	<input type="radio"/> YES <input type="radio"/> NO <input checked="" type="radio"/> 入力なし	実施業務が、センサデータ分析、又は、行動履歴分析の場合、本入力必須。

801 802

センサデータ分析 ▼ 決定 キャンセル

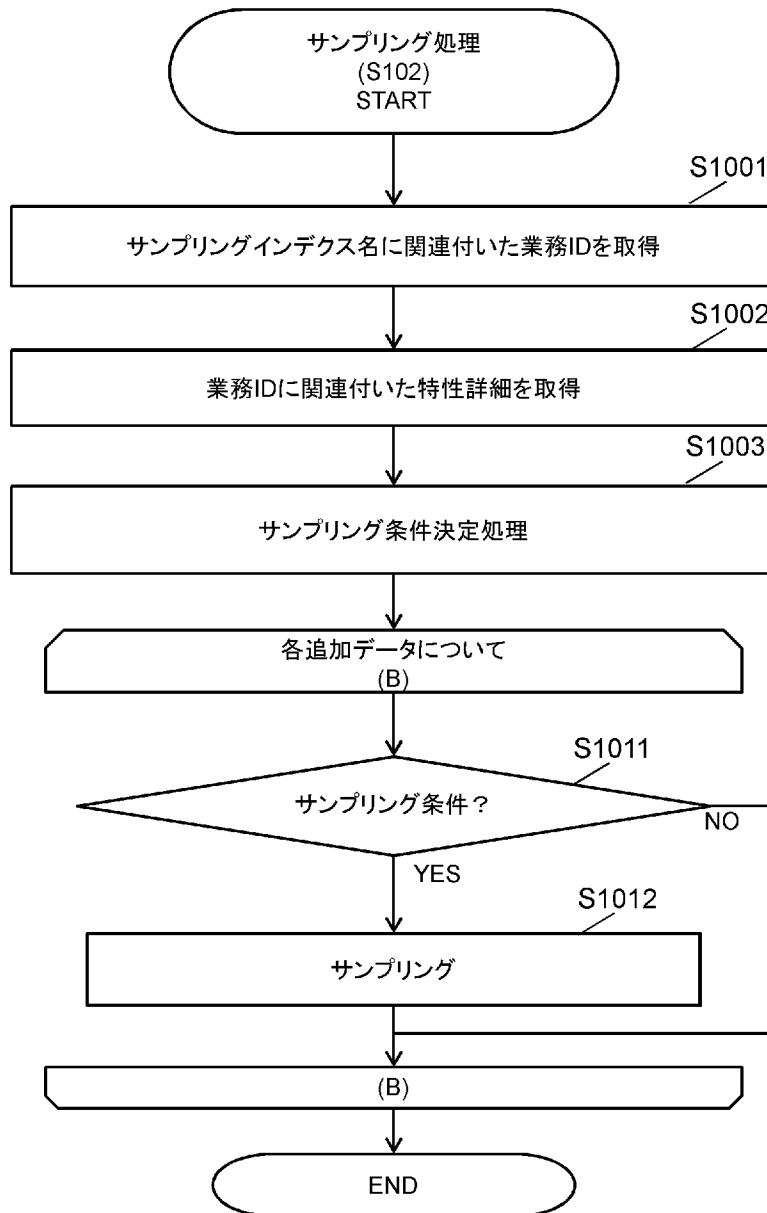
[図9]

FIG. 9



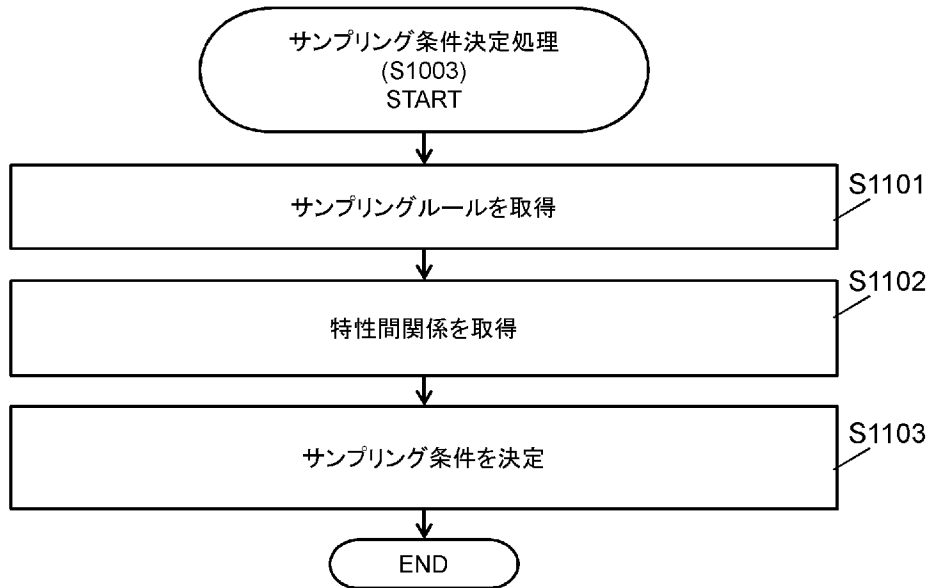
[図10]

FIG. 10



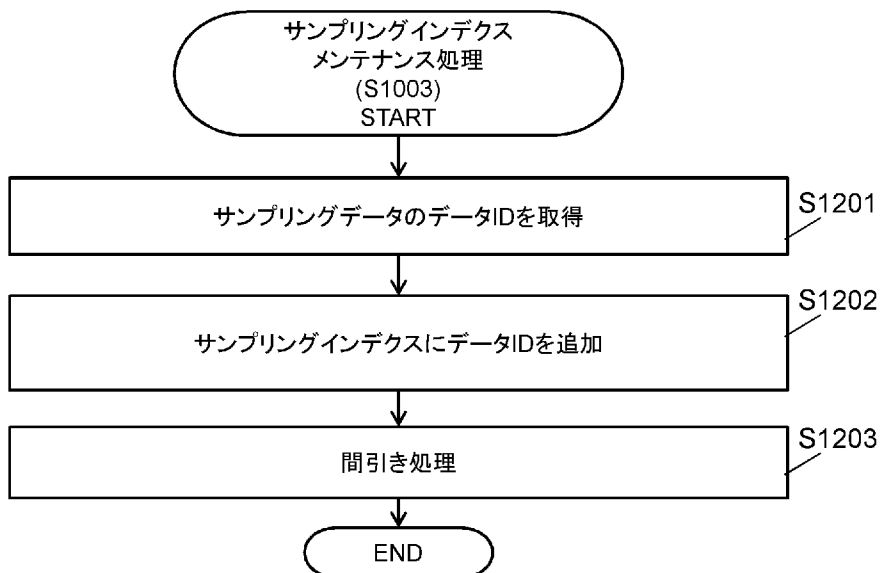
[図11]

FIG. 11



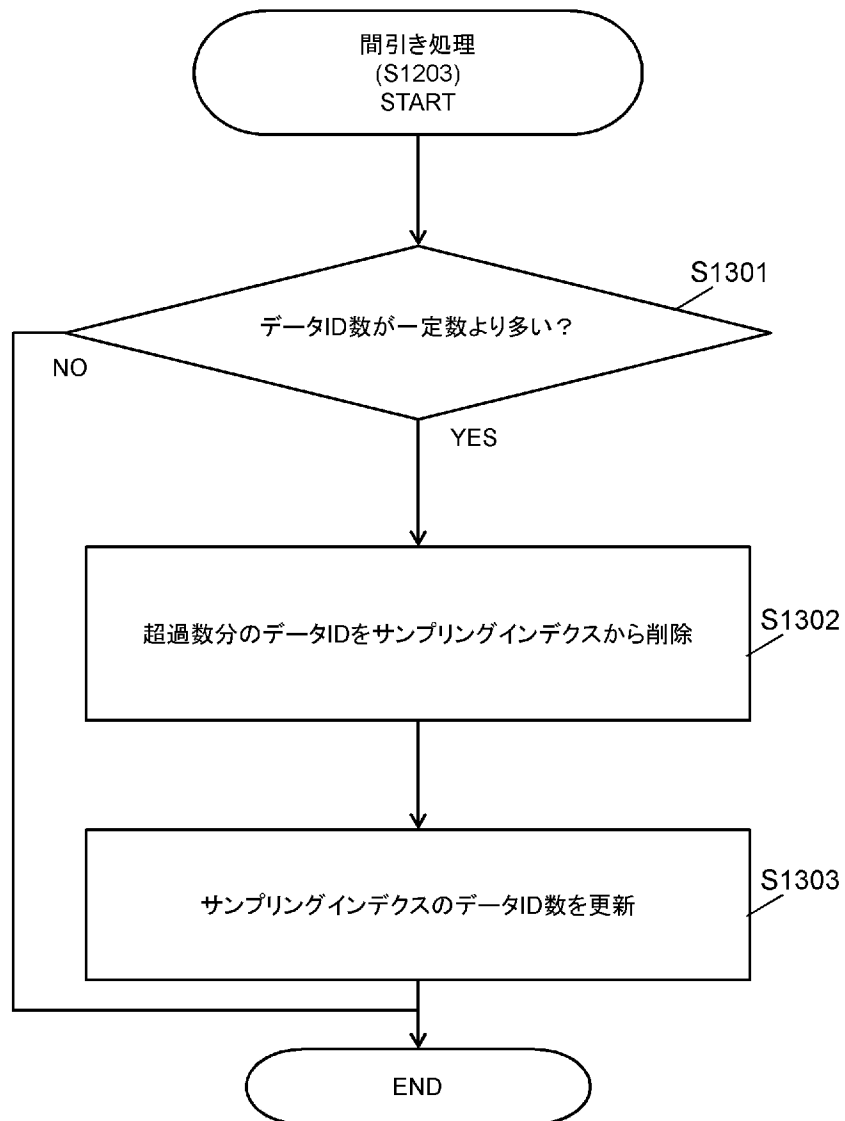
[図12]

FIG. 12



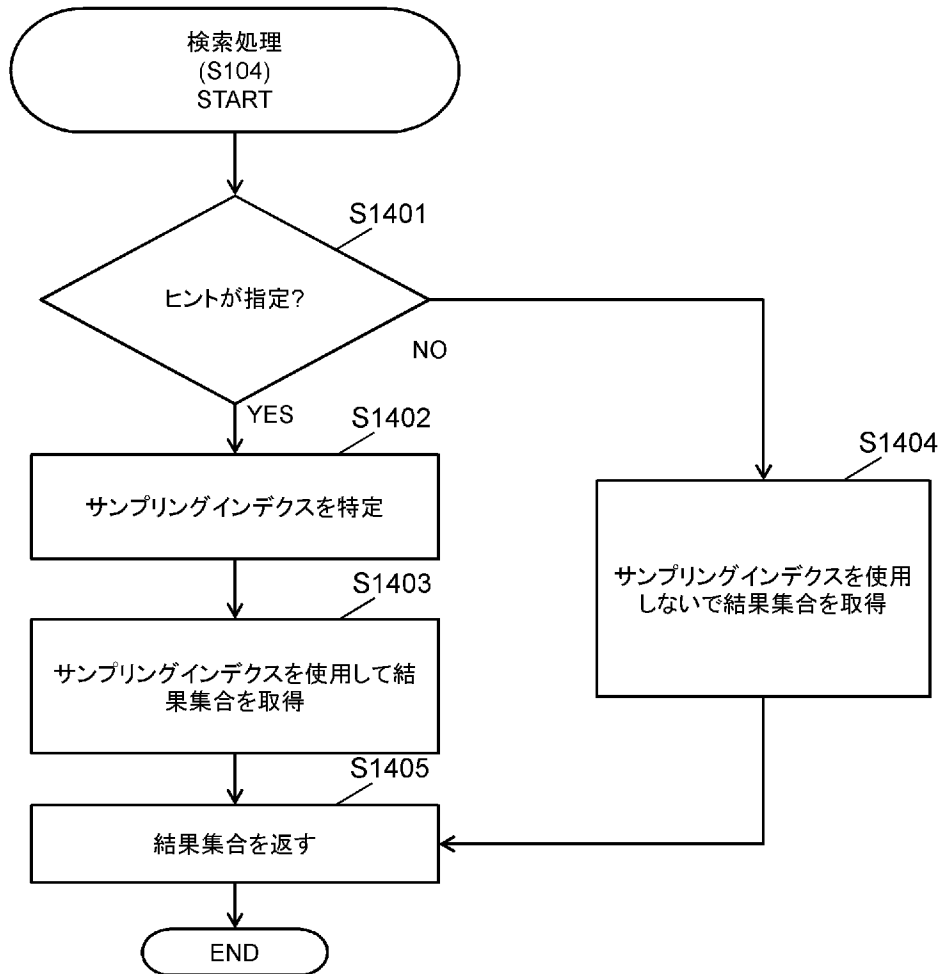
[図13]

FIG. 13



[図14]

FIG. 14



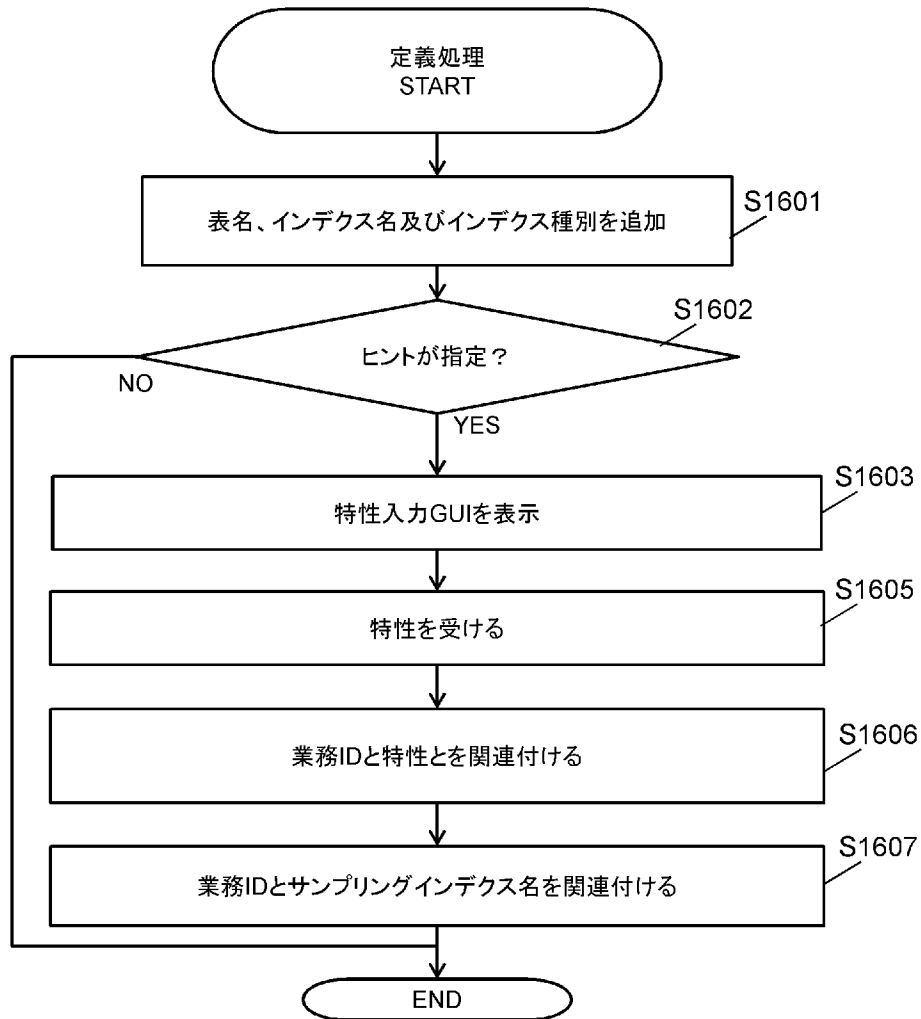
[図15]

FIG. 15

```
SELECT 出力する列名  
FROM Table1  
QUICK S-Index1
```

[図16]

FIG. 16



[図17]

FIG. 17

```
CREATE INDEX S-Index1
ON Table1 QUICK
```

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2015/071196

A. CLASSIFICATION OF SUBJECT MATTER
G06F17/30(2006.01) 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)
G06F17/30

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

Jitsuyo Shinan Koho	1922-1996	Jitsuyo Shinan Toroku Koho	1996-2015
Kokai Jitsuyo Shinan Koho	1971-2015	Toroku Jitsuyo Shinan Koho	1994-2015

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	WO 2010/095457 A1 (NEC Corp.), 26 August 2010 (26.08.2010), paragraphs [0022] to [0068]; fig. 1 to 11 & US 2011/0320650 A1 paragraphs [0052] to [0100]; fig. 1 to 11	1-15
A	JP 2012-194894 A (Fujitsu Ltd.), 11 October 2012 (11.10.2012), paragraphs [0012] to [0018]; fig. 1 to 2 & US 2012/0239599 A1 paragraphs [0054] to [0060]; fig. 1 to 2	1-15
A	US 2002/0198863 A1 (Vijayakumar Anjur et al.), 26 December 2002 (26.12.2002), entire text; all drawings & US 7069264 B2 & EP 1351165 A2	1-15

Further documents are listed in the continuation of Box C. See patent family annex.

* Special categories of cited documents:	"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
"A" document defining the general state of the art which is not considered to be of particular relevance	"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
"E" earlier application or patent but published on or after the international filing date	"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
"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)	"&" document member of the same patent family
"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	

Date of the actual completion of the international search 09 October 2015 (09.10.15)	Date of mailing of the international search report 27 October 2015 (27.10.15)
---	--

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. G06F17/30(2006.01)i		
B. 調査を行った分野 調査を行った最小限資料 (国際特許分類 (IPC)) Int.Cl. G06F17/30		
最小限資料以外の資料で調査を行った分野に含まれるもの 日本国実用新案公報 1922-1996年 日本国公開実用新案公報 1971-2015年 日本国実用新案登録公報 1996-2015年 日本国登録実用新案公報 1994-2015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	WO 2010/095457 A1 (日本電気株式会社) 2010. 08. 26, [0022]-[0068], 図 1-11 & US 2011/0320650 A1, [0052]-[0100], FIG. 1-11	1-15
A	JP 2012-194894 A (富士通株式会社) 2012. 10. 11, [0012]-[0018], 図 1-2 & US 2012/0239599 A1, [0054]-[0060], FIG. 1-2	1-15
A	US 2002/0198863 A1 (Vijayakumar Anjur et al.) 2002. 12. 26, 全 文、全図 & US 7069264 B2 & EP 1351165 A2	1-15
<input type="checkbox"/> C 欄の続きにも文献が列挙されている。 <input type="checkbox"/> 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 09. 10. 2015	国際調査報告の発送日 27. 10. 2015	
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) 山本 俊介 電話番号 03-3581-1101 内線 3599	5M 5087