

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7462191号
(P7462191)

(45)発行日 令和6年4月5日(2024.4.5)

(24)登録日 令和6年3月28日(2024.3.28)

(51)国際特許分類

F I

G 0 6 F 16/22 (2019.01)

G 0 6 F 16/22

請求項の数 7 (全17頁)

(21)出願番号	特願2021-575586(P2021-575586)	(73)特許権者	523066680
(86)(22)出願日	令和2年2月7日(2020.2.7)		株式会社 E S S ホールディングス
(86)国際出願番号	PCT/JP2020/004997		千葉県流山市前平井 6 1 番地 E S S ビル
(87)国際公開番号	WO2021/157092	(74)代理人	100107766
(87)国際公開日	令和3年8月12日(2021.8.12)		弁理士 伊東 忠重
審査請求日	令和4年12月26日(2022.12.26)	(74)代理人	100070150
			弁理士 伊東 忠彦
		(72)発明者	古庄 晋二
			神奈川県横浜市中区元浜町 3 - 2 1 - 2
			ヘリオス関内ビル
		審査官	甲斐 哲雄

最終頁に続く

(54)【発明の名称】 検索方法及び検索装置

(57)【特許請求の範囲】

【請求項 1】

複数の検索対象パラメータが含まれる多次元データの集合を表すテーブルから検索条件を満たす多次元データを検索する検索方法であって、

前記テーブルに含まれる多次元データを用いて、前記検索対象パラメータの値からレコード番号を特定するための転置ブロックで構成される多次元インデックスを作成するインデックス作成手順と、

前記多次元インデックスを用いて、前記テーブルを構成する多次元データのうち、前記検索条件を満たす多次元データのレコード番号を特定する検索手順と、

をコンピュータが実行し、

前記インデックス作成手順は、

所定の順序で、前記検索対象パラメータ毎に、前記検索対象パラメータの値から 1 つ前の転置ブロックのレコード番号又は前記テーブルのレコード番号を特定するための転置ブロックを作成し、

前記検索対象パラメータ毎に作成された各転置ブロックで構成される前記多次元インデックスを作成する、ことを特徴とする検索方法。

【請求項 2】

前記インデックス作成手順は、

前記検索対象パラメータの値の昇順又は降順にアクセスして、前記検索対象パラメータの値から 1 つ前の転置ブロックのレコード番号又は前記テーブルのレコード番号を特定す

ることが可能な転置ブロックを作成する、ことを特徴とする請求項 1 に記載の検索方法。

【請求項 3】

前記検索手順は、

前記所定の順序の逆順で、前記多次元インデックスを構成する各転置インデックスを用いて、前記検索条件を満たす多次元データのレコード番号を特定する、ことを特徴とする請求項 1 又は 2 に記載の検索方法。

【請求項 4】

前記検索条件には、前記複数の検索対象パラメータのうちの少なくとも 1 つの検索対象パラメータの取り得る値の範囲が指定され、

前記検索手順は、

前記所定の順序の逆順で前記各転置インデックスを用いて、前記検索条件に指定されている前記検索対象パラメータの取り得る値の範囲に対応するレコード番号の範囲を特定し、特定したレコード番号の範囲を用いて、1 つ前の転置インデックスで特定されたレコード番号のうち前記範囲に含まれないレコード番号を除外することで、前記検索条件を満たす多次元データのレコード番号を特定する、ことを特徴とする請求項 3 に記載の検索方法。

【請求項 5】

前記検索手順は、

前記レコード番号の範囲を複数の範囲に分割し、

前記複数の範囲の各々で、前記 1 つ前の転置インデックスにおけるレコード番号の特定及び前記範囲に含まれないレコード番号の除外を並列に行う、ことを特徴とする請求項 4 に記載の検索方法。

【請求項 6】

前記検索手順は、

パイプライン処理により、前記各転置インデックスにおけるレコード番号の特定及び前記範囲に含まれないレコード番号の除外を並列に行う、ことを特徴とする請求項 4 に記載の検索方法。

【請求項 7】

複数の検索対象パラメータが含まれる多次元データの集合を表すテーブルから検索条件を満たす多次元データを検索する検索装置であって、

前記テーブルに含まれる多次元データを用いて、前記検索対象パラメータの値からレコード番号を特定するための転置ブロックで構成される多次元インデックスを作成するインデックス作成部と、

前記多次元インデックスを用いて、前記テーブルを構成する多次元データのうち、前記検索条件を満たす多次元データのレコード番号を特定する検索部と、

を有し、

前記インデックス作成部は、

所定の順序で、前記検索対象パラメータ毎に、前記検索対象パラメータの値から 1 つ前の転置ブロックのレコード番号又は前記テーブルのレコード番号を特定するための転置ブロックを作成し、

前記検索対象パラメータ毎に作成された各転置ブロックで構成される前記多次元インデックスを作成する、ことを特徴とする検索装置。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、検索方法及び検索装置に関する。

【背景技術】

【0002】

膨大なデータ群（例えば、科学技術計算に用いられる観測データ等のビッグデータ）の中から必要な部分を検索して提供するサービスのニーズがある。このようなサービスで検索対象となるデータ群は多次元データの集合であり、検索条件として多次元の範囲が指定

10

20

30

40

50

されることが多い。また、検索条件の中には、文字列等の非数値が混在することもある。例えば、 n を観測対象の名称、 (x, y, z) を観測対象の周囲の空間座標として、 $n =$ “観測名 A ” , かつ , $x_1 \leq x \leq x_2$, かつ , $y_1 \leq y \leq y_2$, かつ , $z_1 \leq z \leq z_2$ を満たす範囲の観測データを検索するような場合である。

【先行技術文献】

【非特許文献】

【0003】

【文献】Shaohua Wan, Yu Zhao, Tian Wang, Zonghua Gu, Qammer H. Abbasi, Kim-Kwang Raymond Choo, “ Multi-dimensional data indexing and range query processing via voronoi diagram for internet of things ” , Future Generation Computer System. Volume 91, February 2019, Pages 382-391.

10

【文献】Chisato Yamauchi, “ Development of 2MASS Catalog Server Kit ” , Publications of the Astronomical Society of the Pacific, Volume 123, (2011), pp.1324-1333.

【発明の概要】

【発明が解決しようとする課題】

【0004】

しかしながら、従来技術では、上記のサービスで検索対象となる多次元データ集合を効率的に検索することができなかった。

【0005】

20

例えば、既存のインデックス技術は値の大小関係（つまり、1次元の大小関係）に基づいてインデックスが作成されており、多次元データ集合を効率的に検索することはできない。また、例えば、多次元空間をより小さい次元数の空間にマッピングして、この次元数の小さい空間の中で多次元データ集合を検索する技術も知られているが、文字列等の非数値が混在している場合には多次元空間をより小さい次元数の空間にマッピングすることができない。また、例えば、ハッシュ関数を利用して多次元データ集合を検索する技術も知られているが、検索条件が多次元の範囲で指定されている場合にはハッシュ関数を利用することができない。

【0006】

本発明の一実施形態は、上記の点に鑑みてなされたもので、多次元データ集合を効率的に検索することを目的とする。

30

【課題を解決するための手段】

【0007】

上記目的を達成するため、一実施形態に係る検索方法は、複数の検索対象パラメータが含まれる多次元データの集合を表すテーブルから検索条件を満たす多次元データを検索する検索方法であって、前記テーブルに含まれる多次元データを用いて、前記検索対象パラメータの値からレコード番号を特定するための転置ブロックで構成される多次元インデックスを作成するインデックス作成手順と、前記多次元インデックスを用いて、前記テーブルを構成する多次元データのうち、前記検索条件を満たす多次元データのレコード番号を特定する検索手順と、をコンピュータが実行することを特徴とする。

40

【発明の効果】

【0008】

多次元データを効率的に検索することができる。

【図面の簡単な説明】

【0009】

【図1】本実施形態に係るデータ検索システムの全体構成の一例を示す図である。

【図2】データベースに格納されているテーブルの一例を示す図である。

【図3】コンピュータのハードウェア構成の一例を示す図である。

【図4】本実施形態に係る多次元インデックス作成処理の一例を示すフローチャートである。

50

【図 5】多次元インデックス作成用テーブルの一例を示す図である。

【図 6 A】転置ブロック作成の一例を説明するための図である。

【図 6 B】転置ブロック作成の一例を説明するための図である。

【図 6 C】転置ブロック作成の一例を説明するための図である。

【図 7】多次元インデックスの一例を示す図である。

【図 8】本実施形態に係る検索処理の一例を示すフローチャートである。

【図 9】検索条件を満たす閉区間の一例を説明するための図である。

【図 10 A】パーティショニングの一例を説明するための図である。

【図 10 B】パーティショニングの一例を説明するための図である。

【図 10 C】パーティショニングの一例を説明するための図である。

10

【発明を実施するための形態】

【0010】

以下、本発明の一実施形態について説明する。本実施形態では、膨大な多次元データの集合の中から所望の多次元データを効率的に（つまり、少メモリで、かつ、高速に）検索することが可能なデータ検索システム 1 について説明する。多次元データとは、少なくとも検索条件に指定されるパラメータ（以下、「検索対象パラメータ」という。）が多次元であるデータのことである。すなわち、多次元データとは、検索対象パラメータが N （ただし、 N は 2 以上の整数）個のパラメータ（ $x_0, x_1, \dots, x_{N-1}$ ）で表されるデータのことである。ここで、各検索対象パラメータ x_i （ $i = 0, 1, \dots, N-1$ ）は、順序関係が定義可能な任意の値（例えば、数値や文字列等）を取るものとする。

20

【0011】

本実施形態では、一例として、 $N = 3$ であるものとして、 $x_0 = x, x_1 = y, x_2 = z$ である場合について説明する。

【0012】

< 全体構成 >

まず、本実施形態に係るデータ検索システム 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、本実施形態に係るデータ検索システム 1 の全体構成の一例を示す図である。

【0013】

図 1 に示すように、本実施形態に係るデータ検索システム 1 には、端末装置 10 と、データ検索装置 20 とが含まれる。また、端末装置 10 及びデータ検索装置 20 は、通信ネットワーク N を介して通信可能に接続される。

30

【0014】

端末装置 10 は、膨大な多次元データの中から検索された所望の多次元データを利用する利用側のユーザが操作する PC 等の情報処理装置（コンピュータ）である。ユーザは、端末装置 10 を操作して、データ検索装置 20 で管理されているデータの検索及び表示等を行うことができる。

【0015】

端末装置 10 は、データ検索装置 20 で管理されている多次元データを検索する際の検索条件を設定するための条件設定部 101 を有する。条件設定部 101 は、例えば、端末装置 10 にインストールされた 1 以上のプログラムが CPU（Central Processing Unit）に実行させる処理により実現される。なお、端末装置 10 は、PC に限られず、例えば、スマートフォンやタブレット端末等の種々の情報処理装置であってもよい。

40

【0016】

データ検索装置 20 は、多次元データの提供側が管理するデータベースサーバ等の情報処理装置（コンピュータ）である。

【0017】

データ検索装置 20 は、多次元データを検索するためのインデックスである多次元インデックスを作成するインデックス作成部 201 と、多次元インデックスを用いて検索条件を満たす多次元データを検索する検索部 202 と、多数の多次元データで構成されるデー

50

ブル 1 0 0 0 が格納されているデータベース 2 0 3 とを有する。インデックス作成部 2 0 1 及び検索部 2 0 2 は、例えば、データ検索装置 2 0 にインストールされた 1 以上のプログラムが C P U に実行させる処理により実現される。また、データベース 2 0 3 は、例えば、H D D (Hard Disk Drive) や S S D (Solid State Drive) 等の補助記憶装置により実現可能である。

【 0 0 1 8 】

なお、図 1 に示すデータ検索システム 1 の構成は一例であって、他の構成であってもよい。例えば、データ検索システム 1 には、複数の端末装置 1 0 が含まれていてもよいし、複数のデータ検索装置 2 0 が含まれていてもよい。また、データ検索装置 2 0 が有するデータベース 2 0 3 は、例えば、N A S (Network Attached Storage) 等により実現されていてもよい。

10

【 0 0 1 9 】

< データベース 2 0 3 に格納されているテーブル 1 0 0 0 >

ここで、データベース 2 0 3 に格納されているテーブル 1 0 0 0 について、図 2 を参照しながら説明する。図 2 は、データベース 2 0 3 に格納されているテーブル 1 0 0 0 の一例を示す図である。

【 0 0 2 0 】

図 2 に示すように、テーブル 1 0 0 0 は複数の多次元データ（つまり、複数のレコード）で構成されている。また、各多次元データは、検索対象パラメータを含む複数のパラメータで構成されている。図 2 に示す例では、各多次元データは、6 つのパラメータ（ x , y , z , m_0 , m_1 , m_2 ）で構成されている。これら 6 つのパラメータのうち、 x , y , z が検索対象パラメータである。なお、パラメータ m_0 , m_1 , m_3 は検索対象とならないパラメータであり、例えば、科学技術計算に用いられる観測値等を取るパラメータである。ただし、多次元データを構成するパラメータのうち、どのパラメータを検索対象パラメータとするかは、多次元データの種類やユーザが多次元データを利用する目的等によって異なり得る。

20

【 0 0 2 1 】

また、各多次元データには、テーブル 1 0 0 0 中の位置を示すレコード番号が付与されている。例えば、 $(x, y, z, m_0, m_1, m_2) = (11, 22, 32, m_{00}, m_{10}, m_{20})$ の多次元データには、レコード番号 0 が付与されている。同様に、 $(x, y, z, m_0, m_1, m_2) = (10, 22, 32, m_{01}, m_{11}, m_{21})$ の多次元データには、レコード番号 1 が付与されている。他の多次元データについても同様である。なお、以降では、テーブル 1 0 0 0 のことを「原始テーブル 1 0 0 0」とも表す。

30

【 0 0 2 2 】

< ハードウェア構成 >

次に、本実施形態に係る端末装置 1 0 及びデータ検索装置 2 0 を実現する 1 以上のコンピュータのハードウェア構成について、図 3 を参照しながら説明する。図 3 は、コンピュータ 5 0 0 のハードウェア構成の一例を示す図である。

【 0 0 2 3 】

図 3 に示すコンピュータ 5 0 0 は、入力装置 5 0 1 と、表示装置 5 0 2 と、外部 I / F 5 0 3 と、通信 I / F 5 0 4 と、プロセッサ 5 0 5 と、メモリ装置 5 0 6 とを有する。これら各ハードウェアは、それぞれがバス 5 0 7 を介して通信可能に接続されている。

40

【 0 0 2 4 】

入力装置 5 0 1 は、例えば、キーボードやマウス、タッチパネル等である。表示装置 5 0 2 は、例えば、ディスプレイ等である。なお、データ検索装置 2 0 は、入力装置 5 0 1 及び表示装置 5 0 2 のうちの少なくとも一方を有していなくてもよい。

【 0 0 2 5 】

外部 I / F 5 0 3 は、外部装置とのインタフェースである。外部装置には、例えば、記録媒体 5 0 3 a 等がある。端末装置 1 0 及びデータ検索装置 2 0 は、外部 I / F 5 0 3 を介して、記録媒体 5 0 3 a の読み取りや書き込み等を行うことができる。

50

【 0 0 2 6 】

記録媒体 5 0 3 a には、例えば、フレキシブルディスク、C D (Compact Disc)、D V D (Digital Versatile Disk)、S D メモリカード (Secure Digital memory card)、U S B (Universal Serial Bus) メモリカード等がある。

【 0 0 2 7 】

通信 I / F 5 0 4 は、端末装置 1 0 及びデータ検索装置 2 0 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。端末装置 1 0 及びデータ検索装置 2 0 は、通信 I / F 5 0 4 を介して、他の装置又は機器とデータ通信を行うことができる。

【 0 0 2 8 】

プロセッサ 5 0 5 は、例えば、C P U や G P U (Graphics Processing Unit) 等の各種演算装置である。メモリ装置 5 0 6 は、例えば、R A M (Random Access Memory) や R O M (Read Only Memory)、H D D、S S D、フラッシュメモリ等の各種記憶装置である。

10

【 0 0 2 9 】

本実施形態に係る端末装置 1 0 及びデータ検索装置 2 0 は、図 3 に示すコンピュータ 5 0 0 のハードウェア構成により、後述する各種処理を実現することができる。なお、図 3 に示すコンピュータ 5 0 0 のハードウェア構成は一例であって、他のハードウェア構成であってもよい。例えば、図 3 に示すコンピュータ 5 0 0 は、複数のプロセッサ 5 0 5 を有していてもよいし、複数のメモリ装置 5 0 6 を有していてもよい。

【 0 0 3 0 】

20

< 多次元インデックス作成処理 >

次に、データベース 2 0 3 に格納されている原始テーブル 1 0 0 0 を構成する多次元データの中から検索条件を満たす多次元データを検索する際に用いられる多次元インデックスを作成するための多次元インデックス作成処理について、図 4 を参照しながら説明する。図 4 は、本実施形態に係る多次元インデックス作成処理の一例を示すフローチャートである。以降では、一例として、図 2 に示す原始テーブル 1 0 0 0 がデータベース 2 0 3 に格納されているものとして説明する。

【 0 0 3 1 】

まず、データ検索装置 2 0 のインデックス作成部 2 0 1 は、図 2 に示す原始テーブル 1 0 0 0 を構成する各多次元データの検索対象パラメータを抽出することで、図 5 に示す多次元インデックス作成用テーブル 1 1 0 0 を作成する (ステップ S 1 0 1)。図 5 に示すように、多次元インデックス作成用テーブル 1 1 0 0 は、図 2 に示す原始テーブル 1 0 0 0 を構成する各多次元データから検索対象パラメータを抽出したデータ (レコード) で構成される。

30

【 0 0 3 2 】

次に、データ検索装置 2 0 のインデックス作成部 2 0 1 は、多次元インデックス作成用テーブル 1 1 0 0 を用いて、所定の順序で検索対象パラメータの転置ブロックを作成する (ステップ S 1 0 2)。ここで、所定の順序とは、予め決められた検索対象パラメータの順序のことである。このような順序としては、任意の順序を用いることが可能である。以降では、一例として、x , y , z の順序で検索対象パラメータの転置ブロックを作成する場合について説明する。

40

【 0 0 3 3 】

S t e p 1 - 1 : まず、図 6 A に示すように、インデックス作成部 2 0 1 は、多次元インデックス作成用テーブル 1 1 0 0 を構成する各レコードに対して R e c N o 0 を追加して、テーブル 1 2 0 0 を作成する。このとき、各 R e c N o 0 には、多次元インデックス作成用テーブル 1 1 0 0 を構成する各レコードのレコード番号がそれぞれ格納される。

【 0 0 3 4 】

S t e p 1 - 2 : 次に、図 6 A に示すように、インデックス作成部 2 0 1 は、テーブル 1 2 0 0 を構成する各レコードを検索対象パラメータ x でソートして、テーブル 1 3 0 0 を作成する。

50

【 0 0 3 5 】

なお、図 6 A に示す例では、検索対象パラメータ x の昇順にソートしているが、降順にソートしてもよい。また、図 6 A に示す例では、検索対象パラメータ x でソートすることで各レコードの位置（レコード番号）を入れ替えているが、必ずしもレコードの位置を入れ替える必要はなく、各レコードが検索対象パラメータ x のソート順にアクセスできればよい。例えば、レコードの位置を入れ替えずに、既存のインデックス技術を利用して検索パラメータ x のソート順にアクセスできるようにしてもよい。

【 0 0 3 6 】

Step 1 - 3 : そして、図 6 A に示すように、インデックス作成部 2 0 1 は、テーブル 1 3 0 0 を構成する各レコードから RecNo 0 及び検索対象パラメータ x を抽出することで、x の転置ブロック 1 4 0 0 を作成する。また、インデックス作成部 2 0 1 は、テーブル 1 3 0 0 を構成する各レコードから残りの検索対象パラメータ（つまり、検索対象パラメータ y 及び z ）を抽出することで、テーブル 1 5 0 0 を作成する。

10

【 0 0 3 7 】

Step 2 - 1 : 続いて、図 6 B に示すように、インデックス作成部 2 0 1 は、テーブル 1 5 0 0 を構成する各レコードに対して RecNo 1 を追加して、テーブル 1 2 1 0 を作成する。このとき、各 RecNo 1 には、テーブル 1 5 0 0 を構成する各レコードのレコード番号がそれぞれ格納される。

【 0 0 3 8 】

Step 2 - 2 : 次に、図 6 B に示すように、インデックス作成部 2 0 1 は、テーブル 1 2 1 0 を構成する各レコードを検索対象パラメータ y でソートして、テーブル 1 3 1 0 を作成する。

20

【 0 0 3 9 】

なお、図 6 B に示す例では、検索対象パラメータ y の昇順にソートしているが、降順にソートしてもよい。また、図 6 B に示す例では、検索対象パラメータ y でソートすることで各レコードの位置（レコード番号）を入れ替えているが、必ずしもレコードの位置を入れ替える必要はなく、各レコードが検索対象パラメータ y のソート順にアクセスできればよい。例えば、レコードの位置を入れ替えずに、既存のインデックス技術を利用して検索パラメータ y のソート順にアクセスできるようにしてもよい。

【 0 0 4 0 】

30

Step 2 - 3 : そして、図 6 B に示すように、インデックス作成部 2 0 1 は、テーブル 1 3 1 0 を構成する各レコードから RecNo 1 及び検索対象パラメータ y を抽出することで、y の転置ブロック 1 4 1 0 を作成する。また、インデックス作成部 2 0 1 は、テーブル 1 3 1 0 を構成する各レコードから残りの検索対象パラメータ（つまり、検索対象パラメータ z ）を抽出することで、テーブル 1 5 1 0 を作成する。

【 0 0 4 1 】

Step 3 - 1 : 続いて、図 6 C に示すように、インデックス作成部 2 0 1 は、テーブル 1 5 1 0 を構成する各レコードに対して RecNo 2 を追加して、テーブル 1 2 2 0 を作成する。このとき、各 RecNo 2 には、テーブル 1 5 1 0 を構成する各レコードのレコード番号がそれぞれ格納される。

40

【 0 0 4 2 】

Step 3 - 2 : 次に、図 6 C に示すように、インデックス作成部 2 0 1 は、テーブル 1 2 2 0 を構成する各レコードを検索対象パラメータ z でソートして、テーブル 1 4 2 0 を作成する。

【 0 0 4 3 】

なお、図 6 C に示す例では、検索対象パラメータ z の昇順にソートしているが、降順にソートしてもよい。また、図 6 C に示す例では、検索対象パラメータ z でソートすることで各レコードの位置（レコード番号）を入れ替えているが、必ずしもレコードの位置を入れ替える必要はなく、各レコードが検索対象パラメータ z のソート順にアクセスできればよい。例えば、レコードの位置を入れ替えずに、既存のインデックス技術を利用して検索

50

パラメータ z のソート順にアクセスできるようにしてもよい。

【0044】

Step 3 - 3 : そして、図 6 C に示すように、インデックス作成部 201 は、テーブル 1420 を z の転置ブロック 1420 とする。

【0045】

図 4 に戻る。ステップ S102 に続いて、データ検索装置 20 のインデックス作成部 201 は、上記のステップ S102 で作成した転置ブロック 1400 ~ 1420 を多次元インデックス 1600 としてデータベース 203 に格納する (ステップ S103)。ここで、多次元インデックス 1600 を図 7 に示す。図 7 に示すように、多次元インデックス 1600 は、検索対象パラメータ x の転置ブロック 1400 と検索対象パラメータ y の転置ブロック 1410 と検索対象パラメータ z の転置ブロック 1420 とで構成される。

10

【0046】

以上により作成された多次元インデックス 1600 は、以下の性質 1 及び性質 2 を有する。

【0047】

性質 1 : 所定の順序の逆順により、転置ブロックのレコード番号から、原始テーブル 1000 のレコード番号を特定することができる。

【0048】

例えば、 $RecNo0$, $RecNo1$, $RecNo2$ をそれぞれ 1 次元配列とみなして、 i を 1 次元配列 $RecNo2$ のインデックスとする。図 7 に示す例では、 $RecNo2[0] = 0$, $RecNo2[1] = 1$, $RecNo2[2] = 3$, $\dots$, $RecNo2[10] = 9$, $RecNo2[11] = 11$ である。

20

【0049】

このとき、検索対象パラメータ z の転置ブロック 1420 の作成方法により、 $RecNo2[i]$ は検索対象パラメータ y の転置ブロック 1410 のレコード番号を与える。同様に、検索対象パラメータ y の転置ブロック 1410 の作成方法により、 $RecNo1[RecNo2[i]]$ は検索対象パラメータ x の転置ブロック 1400 のレコード番号を与える。同様に、検索対象パラメータ x の転置ブロック 1400 の作成方法により、 $RecNo0[RecNo1[RecNo2[i]]]$ は原始テーブル 1000 のレコード番号を与える。

30

【0050】

したがって、検索対象パラメータ z の転置ブロック 1420 のレコード番号 i が与えられた場合、 $RecNo0[RecNo1[RecNo2[i]]]$ により、原始テーブル 1000 のレコード番号を特定することができる。

【0051】

性質 2 : 検索対象パラメータの値がソートされているため、検索対象パラメータの値の範囲が与えられると、各転置ブロックのレコード番号の範囲を特定することができる。

【0052】

例えば、 $10 \leq x \leq 11$ が与えられた場合、図 7 に示す転置ブロック 1400 のレコード番号の範囲は閉区間 $[0, 6]$ となる。同様に、 $21 \leq y \leq 22$ が与えられた場合、図 7 に示す転置ブロック 1410 のレコード番号の範囲は閉区間 $[3, 11]$ となる。同様に、 $31 \leq z \leq 32$ が与えられた場合、図 7 に示す転置ブロック 1420 のレコード番号は閉区間 $[4, 11]$ となる。

40

【0053】

< 検索処理 >

次に、検索条件が与えられた場合に、多次元インデックス 1600 を用いて当該検索条件を満たす多次元データを検索するための検索処理について、図 8 を参照しながら説明する。図 8 は、本実施形態に係る検索処理の一例を示すフローチャートである。以降では、一例として、端末装置 10 の条件設定部 101 により検索条件「 $10 \leq x \leq 11$, and , $21 \leq y \leq 22$, and , $31 \leq z \leq 32$ 」が設定されたものとして説明する。なお、

50

and は and 条件を表す。

【0054】

まず、データ検索装置20の検索部202は、端末装置10の条件設定部101で設定された検索条件「10 x 11, 21 y 22, 31 z 32」を受け付ける（ステップS201）。

【0055】

次に、データ検索装置20の検索部202は、上述した性質2を利用して、多次元インデックス1600を構成する各転置ブロックで検索条件を満たす閉区間を特定する（ステップS202）。すなわち、検索部202は、図9に示すように、検索条件10 x 11を満たす閉区間 $S_0 = [0, 6]$ と、検索条件21 y 22を満たす閉区間 $S_1 = [3, 11]$ と、検索条件31 z 32を満たす閉区間 $S_2 = [4, 11]$ とを特定する。

10

【0056】

次に、データ検索装置20の検索部202は、上述した性質1を利用して、原始テーブル1000のレコード番号のうち、検索条件を満たす多次元データのレコード番号を特定する（ステップS203）。検索部202は、以下のStep4-1～Step6-2により検索条件を満たす多次元データのレコード番号を特定する。

【0057】

Step4-1：まず、検索部202は、検索対象パラメータzの転置ブロック1420の閉区間 $S_2 = [4, 11]$ 内のレコードのRecNo2{5, 7, 10, 2, 4, 8, 9, 11}を抽出する。

20

【0058】

Step4-2：上記のStep4-1で抽出したRecNo2は検索対象パラメータyの転置ブロック1410のレコード番号であるため、検索部202は、これらのRecNo2の中で閉区間 $S_1 = [3, 11]$ に含まれないものを除外する。これにより、RecNo2{5, 7, 10, 4, 8, 9, 11}が得られる。

【0059】

Step5-1：続いて、検索部202は、検索対象パラメータyの転置ブロック1410を構成するレコードの中で、RecNo2{5, 7, 10, 4, 8, 9, 11}に対応するレコードのRecNo1{5, 11, 9, 2, 0, 4, 10}を抽出する。すなわち、検索部202は、レコード番号「5」、「7」、「10」、「4」、「8」、「9」、「11」のレコードからそれぞれRecNo1の値を抽出して、RecNo1{5, 11, 9, 2, 0, 4, 10}を得る。

30

【0060】

Step5-2：上記のStep5-1で抽出したRecNo1は検索対象パラメータxの転置ブロック1400のレコード番号であるため、検索部202は、これらのRecNo1の中で閉区間 $S_0 = [0, 6]$ に含まれないものを除外する。これにより、RecNo1{5, 2, 0, 4}が得られる。

【0061】

Step6-1：続いて、検索部202は、検索対象パラメータxの転置ブロック1400を構成するレコードの中で、RecNo1{5, 2, 0, 4}に対応するレコードのRecNo0{3, 7, 1, 0}を抽出する。すなわち、検索部202は、レコード番号「5」、「2」、「0」、「4」のレコードからそれぞれRecNo0の値を抽出して、RecNo0{3, 7, 1, 0}を得る。

40

【0062】

Step6-2：上記のStep6-1で抽出したRecNo0は原始テーブル1000のレコード番号であるため、検索部202は、検索条件を満たす多次元データのレコード番号として{0, 1, 3, 7}を特定する。

【0063】

図8に戻る。データ検索装置20の検索部202は、上記のステップS203で特定したレコード番号のレコードを検索結果として出力する（ステップS204）。すなわち、

50

検索部 202 は、原始テーブル 1000 を構成する多次元データの中で、レコード番号「0」の多次元データと、レコード番号「1」の多次元データと、レコード番号「3」の多次元データと、レコード番号「7」の多次元データとを検索結果として出力する。これにより、検索条件を満たす多次元データが出力される。なお、これらの多次元データの出力先は端末装置 10 とすればよいが、端末装置 10 とは異なる他の装置又は機器に出力されてもよい。

【0064】

以上のように、本実施形態に係るデータ検索システム 1 は、予め作成された多次元インデックス 1600 を用いて、検索条件を満たす多次元データを検索することが可能となる。また、この際、転置ブロックを用いて、検索条件を満たす多次元データのレコード番号が順に絞り込まれるため、不要な検索が行なわれずに、効率的に検索を行うことが可能となる。

10

【0065】

なお、本実施形態に係るデータ検索システム 1 は、多次元データの検索を行う前に予め多次元インデックス 1600 を作成しておく必要があるため、データベース 203 は読取り専用であるか又はその更新頻度が比較的少ない（例えば、数週間から数か月毎の更新等）ことが好ましい。

【0066】

< 応用例 >

ここで、本実施形態に係るデータ検索システム 1 の応用例について、いくつか説明する。

20

【0067】

（応用例 1）

上記で説明した実施形態では、各検索対象パラメータの検索条件は値の範囲であるものとしたが、これに限られない。本実施形態は、検索対象パラメータ間は `and` 条件にする必要があるが、各検索対象パラメータの検索条件には任意の検索条件を指定することが可能である。例えば、検索条件として、`and`、`or`、`not`、`all` 等を指定することが可能である。

【0068】

具体的には、例えば、或る検索対象パラメータ x に関して、「 $x = all$ 」と指定したり、「 $x = 11 or 12$ 」と指定したり、「 $not(10 \leq x \leq 11)$ 」と指定したりすることが可能である。

30

【0069】

（応用例 2）

上記で説明した実施形態では、データ検索装置 20 が有する CPU がマルチコアである場合、上記のステップ S203 を各コアで並列処理を行うことが可能である。これにより、多次元データの検索をより効率的に行うことができる。

【0070】

例えば、閉区間 S_2 を任意の M 個の閉区間 $S_{21}, \dots, S_{2M}$ に分割した上で、各閉区間 $S_{21}, \dots, S_{2M}$ に関するステップ 4-1 ~ ステップ 6-2 を各コアで並列に処理することが可能である。具体的には、例えば、CPU のコアが 3 つである場合、閉区間 S_{21} に関するステップ 4-1 ~ ステップ 6-2 と、閉区間 S_{22} に関するステップ 4-1 ~ ステップ 6-2 と、閉区間 S_{23} に関するステップ 4-1 ~ ステップ 6-2 とをそれぞれコア 1 とコア 2 とコア 3 とで並列に処理することが可能である。

40

【0071】

また、例えば、パイプライン処理により、上記のステップ 4-1 ~ ステップ 4-2 と、上記のステップ 5-1 ~ ステップ 5-2 と、上記のステップ 6-1 ~ ステップ 6-2 とを 3 つのコアでそれぞれ並列に処理することも可能である。具体的には、例えば、コア 1 で閉区間 S_2 内のレコードの `RecNo2` の値の抽出と閉区間 S_1 に含まれない値の除外とを順に実行し、除外されなかった `RecNo2` の値をコア 2 に順次渡す。コア 2 では、コア 1 から渡された `RecNo2` の値に対応するレコードの `RecNo1` の値の抽出と閉区

50

間 S₀に含まれない値の除外とを順に実行し、除外されなかった R e c N o 1 の値をコア 3 に順次渡す。そして、コア 3 では、コア 2 から渡された R e c N o 1 の値に対応する R e c N o 0 の値の抽出を順に実行する。このようなパイプライン処理により、各転置ブロックでのレコード番号の抽出及びその除外をマルチコアで並列に処理することが可能となる。

【 0 0 7 2 】

(応用例 3)

本実施形態は、複数のデータベースに格納されている原始テーブル及びその多次元インデックスを結合又は仮想的に結合することが可能である。言い換えれば、パーティショニングされた原始テーブル及び多次元インデックスを結合又は仮想的に結合することが可能である。

10

【 0 0 7 3 】

例えば、図 1 0 A に示す原始テーブル 2 1 0 0 及びこの原始テーブル 2 1 0 0 の多次元インデックス 2 6 0 0 と、図 1 0 B に示す原始テーブル 2 1 1 0 及びこの原始テーブル 2 1 1 0 の多次元インデックス 2 6 1 0 とが異なる 2 つのデータベースにそれぞれ格納されているものとする。また、原始テーブル 2 1 0 0 及び 2 1 1 0 をそれぞれ構成する多次元データには、検索対象パラメータ (P l a n e t , x , y , z) が含まれているものとする。なお、これら 2 つのデータベースは、2 台のデータ検索装置 2 0 がそれぞれ有していてもよいし、端末装置 1 0 とデータ検索装置 2 0 とがそれぞれ有していてもよい。

【 0 0 7 4 】

20

このとき、原始テーブル 2 1 0 0 及び多次元インデックス 2 6 0 0 を前方、原始テーブル 2 1 1 0 及び多次元インデックス 2 6 1 0 を後方として、原始テーブル同士及び多次元インデックス同士を結合又は仮想的に結合すると、図 1 0 C に示す原始テーブル 3 1 0 0 及びこの原始テーブル 3 1 0 0 の多次元インデックス 3 6 0 0 が得られる。

【 0 0 7 5 】

図 1 0 C に示す原始テーブル 3 1 0 0 は、原始テーブル 2 1 0 0 を前方、原始テーブル 2 1 1 0 を後方として結合又は仮想的に結合したテーブルである。レコード番号「 0 」～「 1 1 」の多次元データは原始テーブル 2 1 0 0 を構成する多次元データであり、レコード番号「 1 2 」～「 2 3 」の多次元データは原始テーブル 2 1 1 0 を構成する多次元データである。ここで、レコード番号「 1 2 」～「 2 3 」は、原始テーブル 2 1 1 0 を構成する各多次元レコードのレコード番号に対して、原始テーブル 2 1 0 0 を構成する多次元データのデータ数 (レコード数) を加算した値である。これらのレコード番号「 1 2 」～「 2 3 」は、原始テーブル 2 1 0 0 と原始テーブル 2 1 1 0 とを結合した場合には実際に多次元データに付与されたものである。一方で、原始テーブル 2 1 0 0 と原始テーブル 2 1 1 0 とを仮想的に結合した場合には、これらのレコード番号「 1 2 」～「 2 3 」は、原始テーブル 2 1 1 0 を構成する各多次元データのレコード番号が読み替えられたものである。

30

【 0 0 7 6 】

また、図 1 0 C に示す多次元インデックス 3 6 0 0 は、原始テーブル 2 1 0 0 を前方、原始テーブル 2 1 1 0 を後方として結合又は仮想的に結合した場合の多次元インデックスである。ここで、レコード番号「 1 2 」～「 2 3 」は、多次元インデックス 2 6 1 0 を構成する各データのレコード番号に対して、多次元インデックス 2 6 0 0 を構成するデータのデータ数 (レコード数) を加算した値である。これらのレコード番号「 1 2 」～「 2 3 」は、原始テーブル 2 1 0 0 と原始テーブル 2 1 1 0 とを結合した場合には多次元インデックス 2 6 1 0 を構成する各データに実際に付与されたものである。一方で、原始テーブル 2 1 0 0 と原始テーブル 2 1 1 0 とを仮想的に結合した場合には、これらのレコード番号「 1 2 」～「 2 3 」は、多次元インデックス 2 6 1 0 を構成する各データのレコード番号が読み替えられたものである。

40

【 0 0 7 7 】

このように、本実施形態に係るデータ検索システム 1 では、複数の原始テーブルを結合又は仮想的に結合する際に、それらの多次元インデックスも容易に結合又は仮想的に結合

50

することができる。このため、例えば、データサイズの大きな原始テーブルを複数の原始テーブルに分割したり、複数の原始テーブルを結合又は仮想的に結合したりすることが容易にできるため、データ検索システム1の運用を容易にすることが可能となる。

【0078】

(応用例4)

本実施形態で作成した転置ブロックは、国際公開第2019/163610号に記載されているSVL、IND及びINVと同一の構造をしている。このため、この転置ブロックをSVL、IND及びINVの代わりに利用することが可能である(又は、逆に、SVL、IND及びINVを、本実施形態の転置ブロックの代わりに利用することも可能である。)。したがって、国際公開第2019/163610号に記載されているSVL、IND及びINV同様に、転置ブロックはコンパクトで、かつ、高速な検索が可能となる。

10

【0079】

(応用例5)

上記のステップS203で各転置ブロックから抽出されたRecNoをソート(ただし、厳密なソートでなくてもよい。)することで、キャッシュのミスヒットを低減することができる。

【0080】

具体的には、例えば、上記のStep4-2で得られたRecNo2{5, 7, 10, 4, 8, 9, 11}をソートして、RecNo2{4, 5, 7, 8, 9, 10, 11}を得る。次に、Step5-1では、このRecNo2{4, 5, 7, 8, 9, 10, 11}に対応するレコードのRecNo1{2, 5, 11, 0, 4, 9, 10}を抽出した後、このRecNo1{2, 5, 11, 0, 4, 9, 10}をソートしてRecNo1{0, 2, 4, 5, 9, 10, 11}を得る。そして、Step5-2では、このRecNo1{0, 2, 4, 5, 9, 10, 11}の中で閉区間S₀=[0, 6]に含まれないものを除外してRec1{0, 2, 4, 5}を得る。以降も同様に、転置ブロックから抽出されたRecNoをソートすればよい。

20

【0081】

なお、上述したように、厳密なソートである必要はなく、大まかなソートであってもよい(つまり、一部ソートされない部分が残っていてもよい。)。大まかなソートとすることで、ソート時間が短縮され、より高速な検索が可能となる。大まかなソートの具体例としては、例えば、SSDの1格納ユニットが4096バイトである場合に、レコード番号を64ビット整数として、512個のレコード番号が1格納ユニットに格納されることを利用し、レコード番号を512で割った番号でソートする、カウンティングソートやバケットソート等が挙げられる。

30

【0082】

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、請求の範囲の記載から逸脱することなく、種々の変形や変更、既知の技術との組み合わせ等が可能である。

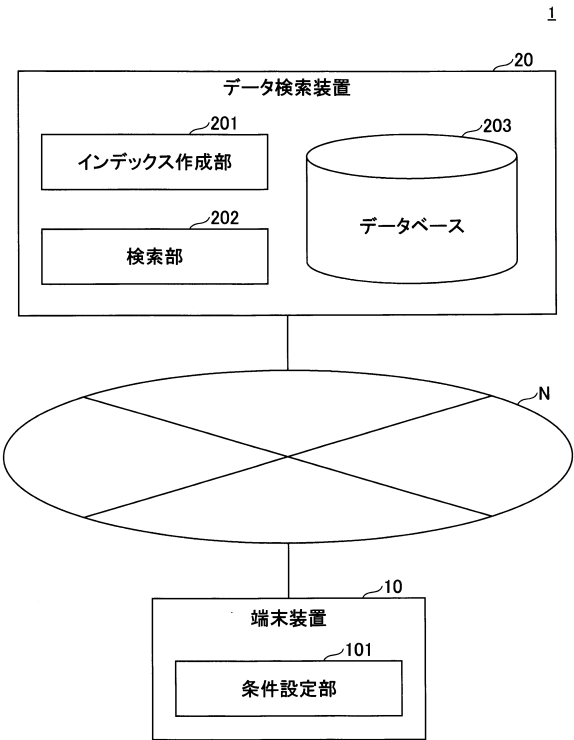
【符号の説明】

【0083】

- 1 データ検索システム
- 10 端末装置
- 20 データ検索装置
- 101 条件設定部
- 201 インデックス作成部
- 202 検索部
- 203 データベース

40

【図面】
【図 1】

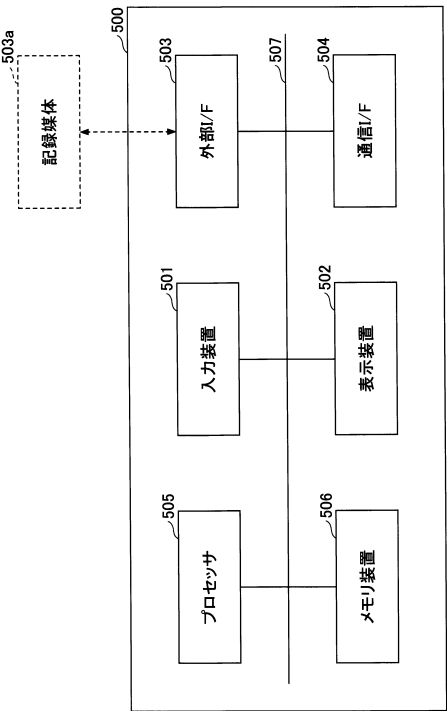


【図 2】

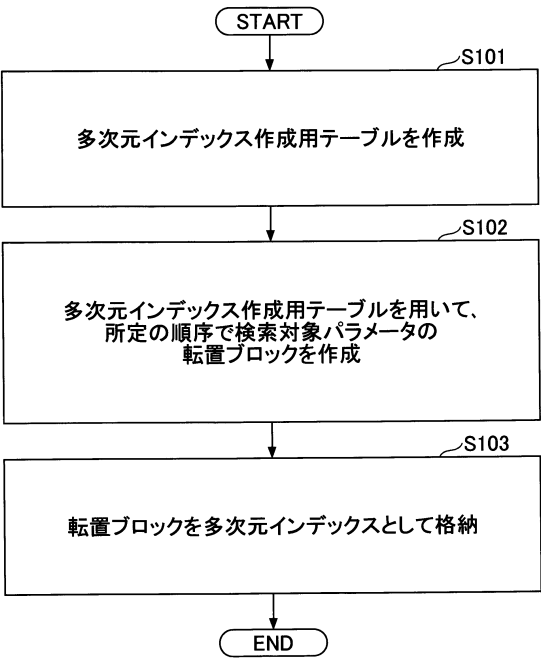
Figure 2 shows a multi-dimensional index table. The table is labeled "多次元データ" (Multi-dimensional Data) and "1000". The columns are labeled x, y, z, m0, m1, m2, m10, m11, m12, m13, m14, m15, m16, m17, m18, m19, m20, m21, m22, m23, m24, m25, m26, m27, m28, m29, m2a, m2b. The rows are labeled 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11.

	x	y	z	m0	m1	m2	m10	m11	m12	m13	m14	m15	m16	m17	m18	m19	m20	m21	m22	m23	m24	m25	m26	m27	m28	m29	m2a	m2b
0	11	22	32	m00	m1	m2	m10	m11	m12	m13	m14	m15	m16	m17	m18	m19	m20	m21	m22	m23	m24	m25	m26	m27	m28	m29	m2a	m2b
1	10	22	32	m01	m1	m2	m10	m11	m12	m13	m14	m15	m16	m17	m18	m19	m20	m21	m22	m23	m24	m25	m26	m27	m28	m29	m2a	m2b
2	12	20	30	m02	m1	m2	m10	m11	m12	m13	m14	m15	m16	m17	m18	m19	m20	m21	m22	m23	m24	m25	m26	m27	m28	m29	m2a	m2b
3	11	21	31	m03	m1	m2	m10	m11	m12	m13	m14	m15	m16	m17	m18	m19	m20	m21	m22	m23	m24	m25	m26	m27	m28	m29	m2a	m2b
4	10	21	30	m04	m1	m2	m10	m11	m12	m13	m14	m15	m16	m17	m18	m19	m20	m21	m22	m23	m24	m25	m26	m27	m28	m29	m2a	m2b
5	12	20	32	m05	m1	m2	m10	m11	m12	m13	m14	m15	m16	m17	m18	m19	m20	m21	m22	m23	m24	m25	m26	m27	m28	m29	m2a	m2b
6	12	22	31	m06	m1	m2	m10	m11	m12	m13	m14	m15	m16	m17	m18	m19	m20	m21	m22	m23	m24	m25	m26	m27	m28	m29	m2a	m2b
7	10	21	32	m07	m1	m2	m10	m11	m12	m13	m14	m15	m16	m17	m18	m19	m20	m21	m22	m23	m24	m25	m26	m27	m28	m29	m2a	m2b
8	12	22	32	m08	m1	m2	m10	m11	m12	m13	m14	m15	m16	m17	m18	m19	m20	m21	m22	m23	m24	m25	m26	m27	m28	m29	m2a	m2b
9	11	21	30	m09	m1	m2	m10	m11	m12	m13	m14	m15	m16	m17	m18	m19	m20	m21	m22	m23	m24	m25	m26	m27	m28	m29	m2a	m2b
10	12	21	31	m0a	m1	m2	m10	m11	m12	m13	m14	m15	m16	m17	m18	m19	m20	m21	m22	m23	m24	m25	m26	m27	m28	m29	m2a	m2b
11	10	20	30	m0b	m1	m2	m10	m11	m12	m13	m14	m15	m16	m17	m18	m19	m20	m21	m22	m23	m24	m25	m26	m27	m28	m29	m2a	m2b

【図 3】



【図 4】



10

20

30

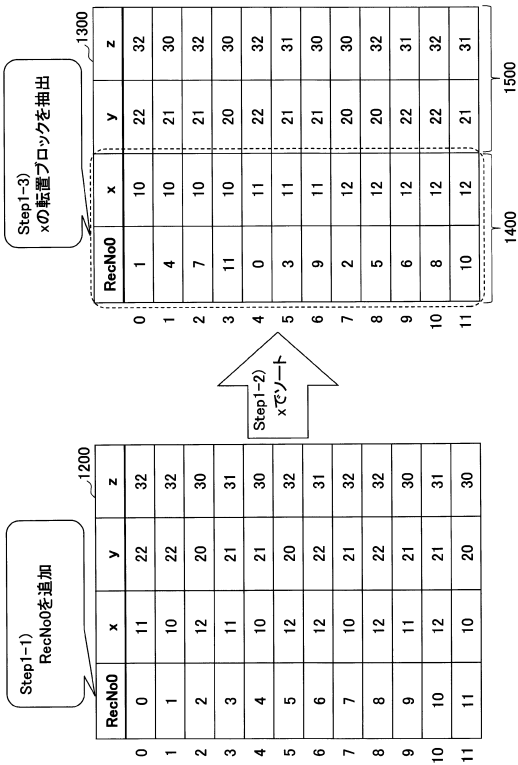
40

50

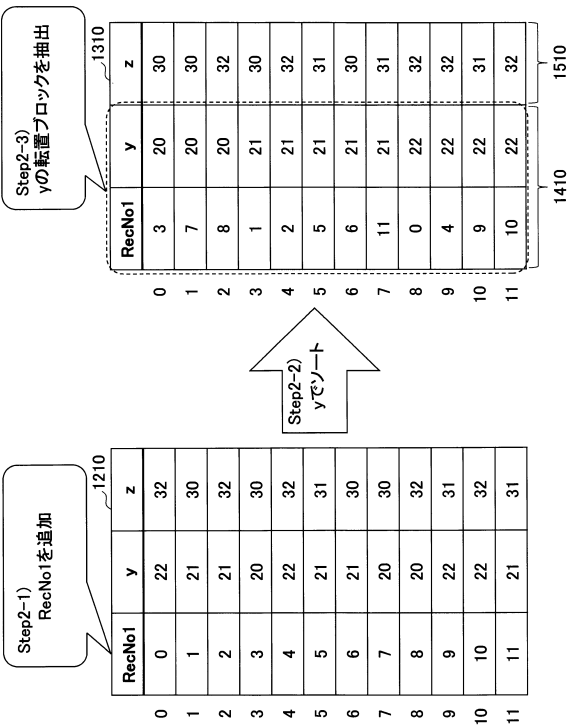
【図 5】

	x	y	z
0	11	22	32
1	10	22	32
2	12	20	30
3	11	21	31
4	10	21	30
5	12	20	32
6	12	22	31
7	10	21	32
8	12	22	32
9	11	21	30
10	12	21	31
11	10	20	30

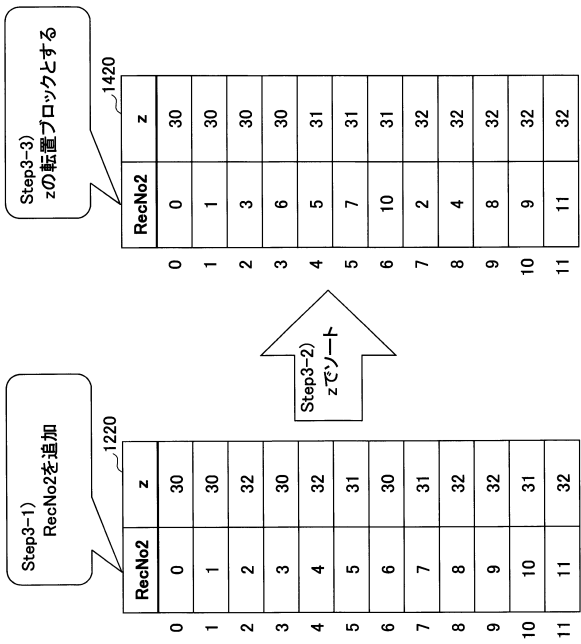
【図 6 A】



【図 6 B】



【図 6 C】



10

20

30

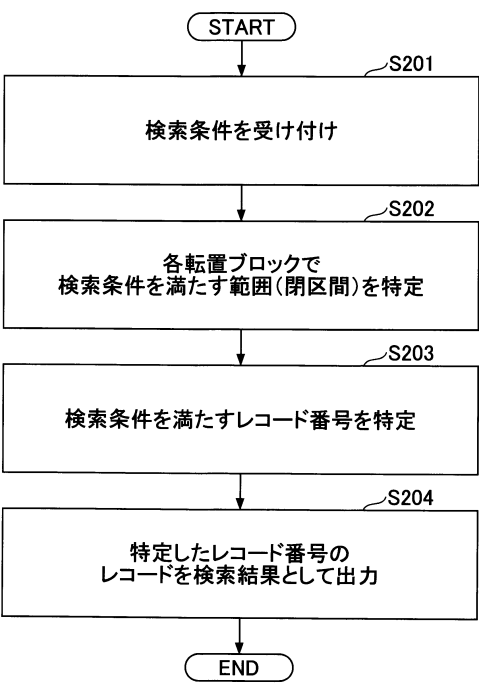
40

50

【図 7】

	RecNo0	x	RecNo1	y	RecNo2	z
0	1	10	3	20	0	30
1	4	10	7	20	1	30
2	7	10	8	20	3	30
3	11	10	1	21	6	30
4	0	11	2	21	5	31
5	3	11	5	21	7	31
6	9	11	6	21	10	31
7	2	12	11	21	2	32
8	5	12	0	22	4	32
9	6	12	4	22	8	32
10	8	12	9	22	9	32
11	10	12	10	22	11	32

【図 8】



【図 9】

	RecNo0	x	RecNo1	y	RecNo2	z
0	1	10	3	20	0	30
1	4	10	7	20	1	30
2	7	10	8	20	3	30
3	11	10	1	21	6	30
4	0	11	2	21	5	31
5	3	11	5	21	7	31
6	9	11	6	21	10	31
7	2	12	11	21	2	32
8	5	12	0	22	4	32
9	6	12	4	22	8	32
10	8	12	9	22	9	32
11	10	12	10	22	11	32

【図 10A】

	RecNo2	y	RecNo1	x	RecNo0	z
0	0	20	3	10	1	30
1	1	20	7	10	4	30
2	3	20	8	10	7	30
3	6	21	1	10	11	30
4	5	21	2	11	0	31
5	7	21	5	11	3	31
6	10	21	6	11	9	31
7	2	21	11	12	2	32
8	4	22	0	12	5	32
9	8	22	4	12	6	32
10	9	22	9	12	8	32
11	11	22	10	12	10	32

	Planet	x	y	z
0	Venus	11	22	32
1	Venus	10	22	32
2	Venus	12	20	30
3	Venus	11	21	31
4	Venus	10	21	30
5	Venus	12	20	32
6	Venus	12	22	31
7	Venus	10	21	32
8	Venus	12	22	32
9	Venus	11	21	30
10	Venus	12	21	31
11	Venus	10	20	30

10

20

30

40

50

【 1 0 B】

2610

	Planet	x	y	z	...
0	Mars	12	20	33	...
1	Mars	13	23	30	...
2	Mars	11	20	31	...
3	Mars	12	21	32	...
4	Mars	10	22	30	...
5	Mars	10	21	33	...
6	Mars	13	22	33	...
7	Mars	10	23	31	...
8	Mars	11	20	30	...
9	Mars	12	22	32	...
10	Mars	11	23	33	...
11	Mars	13	21	31	...

2110

	Planet	x	y	z	...
0	Mars	12	20	33	...
1	Mars	13	23	30	...
2	Mars	11	20	31	...
3	Mars	12	21	32	...
4	Mars	10	22	30	...
5	Mars	10	21	33	...
6	Mars	13	22	33	...
7	Mars	10	23	31	...
8	Mars	11	20	30	...
9	Mars	12	22	32	...
10	Mars	11	23	33	...
11	Mars	13	21	31	...

【 1 0 C】

3100				3600					
Planet	x	y	z	RecNo0	x	RecNo1	y	RecNo2	z
0 Venus	11	22	32	1	10	3	20	0	30
:	...	...	...	...	...	...	...	...	...
11 Venus	10	20	30	10	12	10	22	11	32
12 Mars	12	20	33	16	10	15	20	13	30
13 Mars	13	23	30	17	10	16	20	18	30
14 Mars	11	20	31	19	10	18	20	23	30
15 Mars	12	21	32	14	11	13	21	12	31
16 Mars	10	22	30	20	11	19	21	17	31
17 Mars	10	21	33	22	11	23	21	21	31
18 Mars	13	22	33	12	12	12	22	16	32
19 Mars	10	23	31	15	12	20	22	19	32
20 Mars	11	20	30	21	12	22	22	14	33
21 Mars	12	22	32	13	13	14	23	15	33
22 Mars	11	23	33	18	13	17	23	20	33
23 Mars	13	21	31	23	13	21	23	22	33

10

20

30

40

50

フロントページの続き

- (56)参考文献 国際公開第 2 0 1 4 / 0 5 0 1 9 2 (W O , A 1)
特開 2 0 1 5 - 1 0 6 3 4 7 (J P , A)
特開平 1 1 - 2 9 6 5 2 4 (J P , A)
- (58)調査した分野 (Int.Cl., D B 名)
G 0 6 F 1 6 / 0 0 - 1 6 / 9 5 8