

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第4563558号  
(P4563558)

(45) 発行日 平成22年10月13日(2010.10.13)

(24) 登録日 平成22年8月6日(2010.8.6)

(51) Int. Cl.		F I			
<b>G06F 17/30</b>	<b>(2006.01)</b>	G06F 17/30	240C		
<b>G06F 12/00</b>	<b>(2006.01)</b>	G06F 17/30	180D		
		G06F 17/30	414B		
		G06F 12/00	511C		
		G06F 12/00	512		

請求項の数 6 (全 27 頁)

(21) 出願番号 特願2000-231029 (P2000-231029)  
 (22) 出願日 平成12年7月31日(2000.7.31)  
 (65) 公開番号 特開2002-41551 (P2002-41551A)  
 (43) 公開日 平成14年2月8日(2002.2.8)  
 審査請求日 平成19年7月10日(2007.7.10)

(73) 特許権者 502369012  
 株式会社ターボデータラボラトリー  
 神奈川県横浜市中区元浜町3丁目21番地  
 2号 ヘリオス関内ビル3階  
 (74) 代理人 100099715  
 弁理士 吉田 聡  
 (72) 発明者 古庄 晋二  
 神奈川県横浜市神奈川区松見町4丁目11  
 01番地7コートハウス菊名804号  
 審査官 波内 みさ

最終頁に続く

(54) 【発明の名称】 データのコンパイル方法、および、コンパイル方法を記憶した記憶媒体

(57) 【特許請求の範囲】

【請求項1】

各々が、項目とこれに含まれる項目値とを含むレコードの配列として表わされる表形式データを、コンピュータを用いて、項目値が配置された値リストと、レコード番号に対応して、前記値リスト中の項目値を特定するための位置指定番号を収容した位置指定配列とを有する、コンピュータの記憶装置に記憶された、情報ブロックの形態に変換するデータのコンパイル方法であって、

コンピュータのCPUが、前記表形式データにおいて、レコード番号と対応つけられた項目値からなる項目値配列と、項目値配列中の項目値の位置を特定する順位指定配列と、順位指定配列中の要素を特定する位置指定配列とからなる部分中間ブロックであって、初期的には、単一のレコードと対応つけられた部分中間ブロックをコンピュータの記憶装置に生成するステップと、

コンピュータのCPUが、隣接するレコード番号をもつレコードと対応つけられた隣接する部分中間ブロックを選択して、第1の部分中間ブロックおよび第2の部分中間ブロックからなる対を生成するステップと、

コンピュータのCPUが、前記対に関して、項目値配列を併合して新たな項目値配列をコンピュータの記憶装置に生成するステップと、

コンピュータのCPUが、前記対に関して、第1の部分中間ブロックに属する項目値と第2の部分中間ブロックに属する項目値とを比較することで、併合された新たな項目値配列において、所定の順序にて項目値を特定するように、新たな順位指定配列の要素を決定

するステップと、

コンピュータのCPUが、前記決定された前記新たな順位指定配列の要素が前記新たな順位指定配列の中で格納されている位置番号を新たな位置指定配列の要素に設定するステップと、

コンピュータのCPUが、隣接する部分中間ブロックを選択して対を生成し、新たな項目値配列、要素が決定された新たな順位指定配列および新たな位置指定配列からなる新たな部分中間ブロックを作成することを繰り返して、最終的に一つの部分中間ブロックをコンピュータの記憶装置に作成するステップとを備え、

コンピュータのCPUが、前記最終的に作成された部分中間ブロックを前記情報ブロックとし、かつ、前記新たな順位指定配列および前記新たな項目値配列により前記値リストを構成することを特徴とするデータのコンパイル方法。

10

【請求項2】

コンピュータのCPUが、さらに、最終的に一つの部分中間ブロックが生成された後に、前記新たな順位指定配列の上位の位置から順番に前記新たな順位指定配列の要素を取り出し、前記新たな項目値配列の中で前記取り出された前記順位指定配列の要素によって特定される項目値をコンピュータの記憶装置に確保されたさらに新たな項目値配列の上位の位置から順番に格納することにより、前記項目値が重複なく所定の順序に配置された前記さらに新たな項目値配列が前記値リストを構成することを特徴とする請求項1に記載のデータのコンパイル方法。

20

【請求項3】

各々が、項目とこれに含まれる項目値とを含むレコードの配列として表わされる表形式データを、コンピュータを用いて、項目値が配置された値リストと、レコード番号に対応して、前記値リスト中の項目値を特定するための位置指定番号を収容した位置指定配列とを有する、コンピュータの記憶装置に記憶された、情報ブロックの形態に変換するデータのコンパイル方法を記憶したコンピュータにより読み取り可能な記憶媒体であって、

コンピュータのCPUが、前記表形式データにおいて、レコード番号と対応つけられた項目値からなる項目値配列と、項目値配列中の項目値の位置を特定する順位指定配列と、順位指定配列中の要素を特定する位置指定配列とからなる部分中間ブロックであって、初期的には、単一のレコードと対応つけられた部分中間ブロックをコンピュータの記憶装置に生成するステップと、

30

コンピュータのCPUが、隣接するレコード番号をもつレコードと対応つけられた隣接する部分中間ブロックを選択して、第1の部分中間ブロックおよび第2の部分中間ブロックからなる対を生成するステップと、

コンピュータのCPUが、前記対に関して、項目値配列を併合して新たな項目値配列をコンピュータの記憶装置に生成するステップと、

コンピュータのCPUが、前記対に関して、第1の部分中間ブロックに属する項目値と第2の部分中間ブロックに属する項目値とを比較することで、併合された新たな項目値配列において、所定の順序にて項目値を特定するように、新たな順位指定配列の要素を決定するステップと、

コンピュータのCPUが、前記決定された前記新たな順位指定配列の要素が前記新たな順位指定配列の中で格納されている位置番号を新たな位置指定配列の要素に設定するステップと、

40

コンピュータのCPUが、隣接する部分中間ブロックを選択して対を生成し、新たな項目値配列、要素が決定された新たな順位指定配列および新たな位置指定配列からなる新たな部分中間ブロックを作成することを繰り返して、最終的に一つの部分中間ブロックをコンピュータの記憶装置に作成するステップとを備え、

コンピュータのCPUが、前記最終的に作成された部分中間ブロックを前記情報ブロックとし、かつ、前記新たな順位指定配列および前記新たな項目値配列により前記値リストを構成することを特徴とするデータのコンパイル方法を記憶した記憶媒体。

【請求項4】

50

コンピュータのCPUが、さらに、最終的に一つの部分中間ブロックが生成された後に、前記新たな順位指定配列の上位の位置から順番に前記新たな順位指定配列の要素を取り出し、前記新たな項目値配列の中で前記取り出された前記新たな順位指定配列の要素によって特定される項目値をコンピュータの記憶装置に確保されたさらに新たな項目値配列の上位の位置から順番に格納することにより、前記項目値が重複なく所定の順序に配置された前記さらに新たな項目値配列が前記値リストを構成することを特徴とする請求項3に記載のデータのコンパイル方法を記憶した記憶媒体。

【請求項5】

各々が、項目とこれに含まれる項目値とを含むレコードの配列として表わされる表形式データを、項目値が配置された値リストと、レコード番号に対応して、前記値リスト中の項目値を特定するための位置指定番号を収容した位置指定配列とを有する情報ブロックの形態に変換するデータ処理装置であって、

前記表形式データにおいて、レコード番号と対応つけられた項目値からなる項目値配列と、項目値配列中の項目値の位置を特定する順位指定配列と、順位指定配列中の要素を特定する位置指定配列とからなる部分中間ブロックであって、初期的には、単一のレコードと対応つけられた部分中間ブロックを生成する手段と、

隣接するレコード番号をもつレコードと対応つけられた隣接する部分中間ブロックを選択して、第1の部分中間ブロックおよび第2の部分中間ブロックからなる対を生成する手段と、

前記対に関して、項目値配列を併合して新たな項目値配列を生成する手段と、

前記対に関して、第1の部分中間ブロックに属する項目値と第2の部分中間ブロックに属する項目値とを比較することで、併合された新たな項目値配列において、所定の順序にて項目値を特定するように、新たな順位指定配列の要素を決定する手段と、

前記決定された前記新たな順位指定配列の要素が前記新たな順位指定配列の中で格納されている位置番号を新たな位置指定配列の要素に設定する手段と、

隣接する部分中間ブロックを選択して対を生成し、新たな項目値配列、要素が決定された新たな順位指定配列および新たな位置指定配列からなる新たな部分中間ブロックを作成することを繰り返して、最終的に一つの部分中間ブロックを作成する手段とを備え、

前記最終的に作成された部分中間ブロックを前記情報ブロックとし、かつ、前記新たな順位指定配列および前記新たな項目値配列により前記値リストを構成することを特徴とするデータ処理装置。

【請求項6】

さらに、最終的に一つの部分中間ブロックが生成された後に、前記新たな順位指定配列の上位の位置から順番に前記新たな順位指定配列の要素を取り出し、前記新たな項目値配列の中で前記取り出された前記新たな順位指定配列の要素によって特定される項目値をさらに新たな項目値配列の上位の位置から順番に格納することにより、前記項目値が重複なく所定の順序に配置された前記さらに新たな項目値配列が前記値リストを構成することを特徴とする請求項5に記載のデータ処理装置。

【発明の詳細な説明】

【0001】

【産業上の技術分野】

本発明は、コンピュータのような情報処理装置を用いて大量のデータを処理するデータ処理方法およびデータ処理装置に関し、より詳細には、任意の表形式データを、高速処理に適した所定の形式に変換する方法および装置に関する。

【0002】

【従来の技術】

データベースなど大量のデータを、検索、集計、ソートするため、或いは、これらのデータをジョイン、更新等するためには多大な時間を要する。これを解決すべく、本発明者は、表形式のデータを極めて拘束に検索、集計およびソートするための手法、並びに、表形式のデータをジョインし、更新し、或いは、トランザクション処理する手法を提案してい

10

20

30

40

50

る(PCT/J P 99/04300、特願平11-151156号、特願平11-215450号)。

これらに提示された一連の方法は、一切のインデックスを用いず、統一的な処理を実行することができ、かつ、部分集合に対しても効率を落とすことなく処理を実行することが可能であるという点で革新的である。

#### 【0003】

上記高速な検索、集計、ソート、ジョイン、更新等を実現するためには、与えられた表形式のデータを、所定の形式に変換する処理(以下、これを「コンパイル」ないし「コンパイル処理」と称する。)を実行する必要がある。この形式においては、図2に示すように、ある項目に関して、項目値が重複無かつ所定の順序にて配置された値リストと、レコード番号に対応して要素が設けられ、各々の要素が、レコード番号に対応した項目値の配置された、値リスト中の位置を指定する配列(ポインタ配列)とが設けられている。このようなある項目に対して所定の形式となっているデータ群を情報ブロックと称する。

10

#### 【0004】

コンパイル処理は、以下のような場合に必要となる。

- (1)新規に与えられた表形式のデータを所定の形式に変換して記憶すべき場合、
- (2)1または複数の項目から新たな項目を算出する場合、および、
- (3)ある表形式のデータに、他の表形式のデータを追加すべき場合、たとえば、1月分の表に、2月分の表を追加する場合。

このように、コンパイル処理を実行すべき場合には、従来、既知のソート手法(たとえば、クイックソート)を用いて、値リスト中の項目値を並べ替えるなどの処理を実行していた。

20

#### 【0005】

##### 【発明が解決しようとする課題】

しかしながら、上記既知のソート方法を用いて情報ブロックを作成する場合には、以下の問題が生じる。

- (1)一般に、 $O(n \cdot \log(n))$ の比較演算が必要とされ、多大な処理時間を要する。
- (2)特にクイックソートを用いた場合に、データの分布により極端に処理効率が悪化する場合がある。
- (3)複数のプロセッサを利用してソートを高速化することが困難である。

30

本発明は、高速な処理時間と、データの分布によっても処理効率が極端に悪化しないことが予測できるデータのコンパイル方法を提供することを目的とする。

また、本発明は、シングルプロセッサでも効果的であるが、特に、末子部パラレルシステムにおいてパフォーマンスを劇的に改善できるデータのコンパイル方法を提供することを目的とする。

#### 【0006】

##### 【課題を解決するための手段】

本発明の目的は、各々が、項目とこれに含まれる項目値とを含むレコードの配列として表わされる表形式データを、項目値が所定の順序でかつ重複なしに配置された値リストと、レコード番号に対応して、前記値リスト中の項目値を特定するための位置指定番号を収容した位置指定配列とを有する情報ブロックの形態に変換するデータのコンパイル方法であって、

40

前記表形式データにおいて、レコード番号と対応つけられた項目値からなる項目値配列と、項目値配列中の項目値の位置を特定する順位指定配列と、順位指定配列中の要素を特定する位置指定配列とからなる部分中間ブロックであって、初期的には、単一のレコードと対応つけられた部分中間ブロックを生成するステップと、

隣接する部分中間ブロックを選択して、第1の部分中間ブロックおよび第2の部分空間ブロックからなる対を生成するステップと、

前記対に関して、項目値配列を併合して新たな項目値配列を生成するステップと、

前記対に関して、第1の部分中間ブロックに属する項目値と第2の部分中間ブロックに属

50

する項目値とを比較することで、併合された新たな項目値配列において、所定の順序にて項目値を特定するように、順位指定配列の要素を決定するステップと、  
前記順位指定配列の要素の決定に伴って、併合前の項目値を特定できるように、前記位置指定配列の要素を決定するステップと、  
隣接する部分中間ブロックを選択して対を生成し、併合した項目値配列、要素が決定された順位指定配列および位置指定配列からなる新たな部分中間ブロックを作成することを繰り返して、最終的に一つの部分中間ブロックを作成するステップとを備え、  
前記最終的に作成された部分中間ブロックを情報ブロックとし、かつ、前記順位指定配列および項目値配列により値リストを構成することを特徴とするデータのコンパイル方法により達成される。

10

**【0007】**

本発明においては、部分中間ブロックの対を生成して、各部分中間ブロックの項目値の比較を行なうことで、対における項目値の順位を決定している。また、順次隣接する部分中間ブロックを対にして新たな部分中間ブロックを生成するような、いわゆるトーナメントの形態で、より大きな部分中間ブロックを生成している。本発明においては、「 $n$ 」個のレコードに対して全ての項目値が異なっている場合には、比較の回数が、約「 $n * (\log_2(n) - 1)$ 」となり、いわゆる $O(n \cdot \log(n))$ となる。しかしながら、「 $m$  ( $n > m$ )」個の項目値のみが異なり、これが「 $n$ 」と比較して十分に小さい場合には、比較の回数は略「 $n$ 」回にまで減じることができる。また、「 $m$ 」が比較的大きな場合であっても、比較回数を「 $m * (\log_2(n) - 1)$ 」とすることができる。すなわち、処理効率を予測することができ、かつ、処理効率が最悪の場合でも、せいぜい、 $O(n \cdot \log(n))$ とすることが可能となる。

20

**【0008】**

また、いわゆるトーナメントの形態を利用しているため、部分中間ブロックの対の比較を、複数のプロセッサにてパラレル処理することに適している。これにより、処理時間を著しく減じることが可能となる。

**【0009】**

本発明の好ましい実施態様においては、順位指定配列の要素を決定するステップが、新たな順位指定配列を生成するステップと、第1の部分中間ブロックの項目値と第2の部分中間ブロックの項目値とを比較して、より高位の順位となるものを決定し、前記新たな順位指定配列の上位から順次、当該項目値を指定する第1の部分中間ブロック或いは第2の部分中間ブロック中の順位指定配列の値を配置するステップとを有し、  
位置指定配列の要素を決定するステップが、前記新たな順位指定配列を特定するための位置指定再定義配列を生成するステップと、前記順位指定配列の値が配置される際に、第1の部分中間ブロック或いは第2の部分中間ブロックにおいて、位置指定再定義配列中の対応する位置に、前記新たな順位指定配列にて配置された値を特定できるように要素を配置するステップとを有し、  
さらに、前記位置指定再定義配列にて、前記位置指定配列の値を変換して、新たな位置指定配列を求めるステップを備えている。

30

**【0010】**

本発明において、情報ブロックの値リストは、順位指定配列と項目値配列の形態となっても良い。或いは、他の形態においては、さらに、最終的に一つの部分中間ブロックが生成された後に、前記順位指定配列中の上位に位置している要素にて特定される項目値を、順次配置することにより、重複なくかつ所定の順序にて配置された新たな項目値配列を作成するように構成しても良い。  
また、本発明の目的は、上記ステップを備えたプログラムを記憶した、コンピュータにより読み取り可能な記憶媒体によっても達成される。

40

**【0011】****【発明の実施の形態】**

以下、添付図面を参照して、本発明の実施の形態につき説明を加える。図1は、本発明の

50

実施の形態にかかる検索、集計およびサーチ方法を実現できるコンピュータシステムのハードウェア構成を示すブロックダイヤグラムである。図1に示すように、このコンピュータシステム10は、通常のものと同様の構成であり、プログラムを実行することにより、システム全体および個々の構成部分を制御するCPU12、ワークデータなどを記憶するRAM(Random Access Memory)14、プログラム等を記憶するROM(Read Only Memory)16、ハードディスク等の固定記憶媒体18、CD-ROM19をアクセスするためのCD-ROMドライバ20、CD-ROMドライバ20や外部ネットワーク(図示せず)と接続された外部端子との間に設けられたインタフェース(I/F)22、キーボードやマウスからなる入力装置24、CRT表示装置26を備えている。CPU12、RAM14、ROM16、外部記憶媒体18、I/F22、入力装置24および表示装置26は、バス28を介して相互に接続されている。

10

#### 【0012】

本実施の形態にかかる、任意の表形式データをコンパイルするプログラム、コンパイルした表形式データから所定の項目の表(ビュー)を作成するプログラム、検索プログラム、集計プログラムおよびソートプログラムは、CD-ROM19に収容され、CD-ROMドライバ20に読取られても良いし、ROM16に予め記憶されていても良い。また、いったんCD-ROM19から読み出したものを、外部記憶媒体18の所定の領域に記憶しておいても良い。或いは、上記プログラムは、ネットワーク(図示せず)、外部端子およびI/F22を経て外部から供給されるものであっても良い。

本発明においては、任意の表形式データをコンパイルして、高速な検索、集計およびソートが可能な情報ブロックの集合体の形式を有するデータを構築する。

20

このデータを構築する際の高速なコンパイル方法が本発明の本質的事項となる。

#### 【0013】

本発明者は、情報ブロックの集合体の形式を有するデータを利用した検索、集計およびソート方法を考案した(特願平10-227278号)。また、上記形式を有するデータを利用したデータのジョイン方法およびデータの削除/追加/更新方法を考案した(特願平11-151156号、特願平11-215450号)。まず、これら情報ブロックの構造、および、これを利用した検索について説明し、その後、上記情報ブロックの作成(コンパイル)につき説明を加える。

#### 【0014】

30

図2は、本実施の形態にて用いる情報ブロックを示す図である。図2に示すように、情報ブロック100は、値リスト110と値リストへのポインタ配列120とを含んでいる。値リスト110は、表形式データの各項目に対して、その項目に属する項目値が順序付け(整数化)された項目値番号の順番に、上記項目値番号に対応した項目値111が格納されたテーブルである。値リストへのポインタ配列120は、表形式データのある列(すなわち項目)の項目値番号、つまり値リスト110へのポインタが表形式データのレコード番号順に格納された配列である。

#### 【0015】

上記値リストへのポインタ配列120と値リスト110とを組み合わせることにより、あるレコード番号が与えられたときに、所定の項目に関する値リストへのポインタ配列120からそのレコード番号に対応して格納された項目値番号を取り出し、次いで、値リスト110内でその項目値番号に対応して格納された項目値を取り出すことにより、レコード番号から項目値を得ることができる。したがって、従来のデータ表と同様に、レコード番号(行)と項目(列)という座標を用いてすべてのデータ(項目値)を参照することができる。

40

#### 【0016】

たとえば、図3(a)に示す表形式データを考える。この例では、顧客ID、顧客名、電話番号という項目に種々の項目値が与えられている。本実施の形態においては、このような表形式データを、図3(b)ないし(d)に示す形式の情報ブロックとして保持している。たとえば、図3(b)において、ポインタ配列120-1は、顧客IDを示す項目値

50

を格納した値リスト 1 1 0 - 1 に関連付けられている。すなわち、先頭レコード（レコード番号「0」）のポインタ配列のポインタ値は 0 であり、これに対応して、顧客 ID を示す項目値「1」が得られる。図 3（b）において、ポインタ配列 1 2 0 - 2 は、顧客名を示す項目値を格納した値リスト 1 1 0 - 2 に関連付けられている。たとえば先頭レコード（レコード番号「0」）のポインタ配列におけるポインタ値は「5」であり、これに対応して、顧客名を示す項目値「山田 男」が得られる。図 3（c）においても、同様に、ポインタ配列 1 2 0 - 3 が、電話番号を示す項目値を格納した値リスト 1 1 0 - 3 に関連付けられていることが理解できよう。また、各値リストにおいては、項目値が順序付けられて（この例では昇順）いることが理解できよう。

#### 【0017】

さらに、本実施の形態においては、情報ブロック 1 0 0 の値管理テーブルは、値リスト 1 1 0 のほか、検索や集計のために用いる分類番号フラグ配列、項目値に対応するポインタを格納すべきメモリ空間の先頭アドレスを示す開始位置配列、および、存在数配列が含まれている。分類番号フラグ配列の各フラグ、および、存在数配列の各存在数は、項目値の各々に対応付けられている。分類番号フラグのフラグ値は、通常「0」であり、検索や集計の際に見出すべき項目値に対応して「1」にセットされる。また、存在数は、その項目値を有するレコードの個数に対応する。なお、開始位置は、対応するポインタ値よりも小さなポインタ値に対応する存在数を加算したのに対応するため、必ずしも設ける必要はない。

図 4（a）は、表形式データの他の例を示す図、図 4（b）および（c）は、それぞれ、「性別」および「年令」に関する情報ブロックを示す図である。図 4（b）に示すように、性別に関する情報ブロック 2 0 0 - 1 の値管理テーブル 2 1 0 - 1 には、ポインタ配列 2 2 0 の各ポインタ値に対応する項目値（「男性」および「女性」）と、各項目値に対応する分類番号、開始位置および存在数が示されている。たとえば、ポインタ値が「0」（つまり、値リストの項目値が「男性」）であるようなレコードの数は 6 3 2 5 6 4 個であり、その一方、ポインタ値が「1」（つまり、値リストの項目値が「女性」）であるようなレコードの数は 3 6 7 4 2 6 個となっている。また、各項目値に対応する開始位置は、後述するレコードへのポインタ配列 2 3 0 - 1 の先頭アドレスを示している。図 4（c）においても、同様のことが理解できよう。

#### 【0018】

このようなデータ構造を有する情報ブロックを用いた検索の一例につき以下に説明を加える。図 5 は、単一項目に関する検索手法を示すフローチャートである。この処理は、CPU 1 2（図 1 参照）が所定の検索プログラムを実行することにより実現される。この例では、「年令」の項目値が 1 6 歳または 1 9 歳であるレコードが検索される。まず、表形式データに関する情報ブロックのうち、図 4（c）に示す「年令」に関する情報ブロック 2 0 0 - 2 が特定される（ステップ 5 0 1）。

#### 【0019】

次いで、特定された情報ブロック（以下、「特定情報ブロック」と称する。）の値リスト 2 1 0 - 2 において、項目値が上記検索条件に合致するもの（1 6 歳または 1 9 歳）に対応する行の分類番号が「1」にセットされる（ステップ 5 0 2）。本例の場合には、項目値番号「0」および項目値番号「3」に対応する行の分類番号が 1 にセットされる。次いで、分類番号が「1」にセットされている行に対応した開始位置および存在数が取得される（ステップ 5 0 3）。これら情報をポインタ取り出し情報と称する。レコードへのポインタ配列において、ステップ 5 0 3 にて取得されたポインタ取り出し情報に基づき、検索条件に合致したレコードへのポインタを示すレコード番号が取り出される（ステップ 5 0 4）。

本例においては、項目値番号「0」に対応したレコードのポインタは、レコードへのポインタ配列の開始位置「0」すなわち先頭から、4 5 8 9 8 個目までの領域に格納され、その一方、項目値番号「3」に対応したレコードのポインタは、レコードへのポインタ配列の 2 3 8 3 1 3 7 番目から 1 8 9 6 5 3 個分の領域に格納されていることがわかる。最後

10

20

30

40

50

に、後の処理にて利用できるようにするために、取り出されたレコード番号の配列が、結果集合として作成され、これが保持される（ステップ505）。

また、集計およびソートも、分類番号、開始位置および存在数を利用することにより実現することができる。

#### 【0020】

次に、上記情報ブロックを生成するためのコンパイル処理について説明を加える。このコンパイル処理においては、ある項目について無作為に配置されたデータを、重複なくかつ所定の順序（たとえば五十音順やアルファベット順）でソートされた値リストを作り出している。図6(a)は、表形式データにおける、レコードとある項目（姓）の項目値とを示す図である。本実施の形態にかかるコンパイル処理においては、この図6(a)に示す

10

ような、項目値がランダムに並べられた項目に関して、図6(b)に示すような形式の情報ブロックを生成する。図6(b)において、情報ブロック600には、値リストへのポインタ配列601（以下、「PV」と称する。）、および、値リスト602が含まれる。コンパイル処理の直後には、値リスト602は、元の項目値が収容された配列（オリジナル項目値配列）603と、当該オリジナル項目値配列603の格納位置番号を示す配列604（以下、「VLP」と称する。）とから構成される。上記ポインタ配列（PV）は、値リスト中の項目値を指定するという意味で、位置指定配列とすることができる。また、VLPは、項目値の順序を表しているという意味で、順位指定配列とすることができる。コンパイル処理の直後の

20

情報ブロックは、上述した図3(b)に示す情報ブロックとは若干形式が異なっているが、後に述べる処理により同等の形式とすることができる。図7は、この実施の形態にかかるコンパイル処理の概略を示すフローチャートである。図7に示すように、コンパイル処理においては、コンピュータシステム10は、固定記憶媒体18等に記憶された表形式データを読み出し（ステップ701）、情報ブロックを作成する必要がある項目を特定する（ステップ702）。たとえば、全ての項目について情報ブロックを作成しても良いし、指定された任意の項目に関する情報ブロックのみを作成しても良い。また、読み出した表形式データは、RAM14上に一時的に保存しておくのが、処理時間を考慮した場合に望ましい。

#### 【0021】

次いで、コンピュータシステム10は、表形式データから所定の項目を選択して、当該項目にかかるデータ（項目値）をレコード番号と関連付けて切り出す（ステップ703）。これにより、たとえば、図6(a)に示す形式のデータを取り出すことができる。次いで、当該項目に関する情報ブロックが作成される（ステップ704）。この処理については後に詳述する。ステップ703、704の処理は、指定された全項目について実行される（ステップ705、706参照）。これにより、所定の項目に関する情報ブロックが完成する。

30

#### 【0022】

次に、ステップ704の処理につきより詳細に説明を加える。ステップ704の処理は、隣接するレコードを順次組み合わせて、当該組み合わせにおいてPVやVLP中の値を決定することにより実現される。たとえば、レコード番号が $0 \sim 4n$ （レコード数 =  $4n + 1$ ）である項目について、図8(a)に示すように、初めには、 $4n + 1$ 個の部分中間ブロック $801 - 0 \sim 801 - (4n)$ を考える。そして、隣接する二つの部分中間ブロック（たとえば、部分中間ブロック $801 - 0$ と部分中間ブロック $801 - 1$ ）を組み合わせ、図8(b)に示す新たな部分中間ブロック（たとえば、部分中間ブロック $811 - 0$ ）を考え、これら新たな部分中間ブロックにおいてPVやVLPの値を更新する。図8(b)に示すように、この時点で、 $(2n + 1)$ 個の新たな部分中間ブロックが作られる。さらに、隣接する二つの部分中間ブロック（たとえば、部分中間ブロック $811 - 0$ と $811 - 1$ ）とを組み合わせ、図8(c)に示すように新たな部分中間ブロック（たとえば、部分中間ブロック $821 - 0$ ）を考え、これら新たな部分中間ブロックにおいてPVやVLPの値を更新する。

40

50

## 【 0 0 2 3 】

このような処理を繰り返すことで、最終的には、2つの部分中間ブロックを組み合わせたブロックを作り出すことができる。このブロックが情報ブロックとなる。上記処理をフローチャートにて表したものを図9に示す。

図9のステップ901にて生成される初期的な部分中間ブロックには、たとえば、図8(a)の部分中間ブロックが対応する。また、末尾のブロックまで、隣接する部分中間ブロックが選ばれて(ステップ902)、これらが合併されて、変換配列(以下、「TR」と称する。)や合併後の新たな格納位置番号を示す配列(以下、「VLP」と称する。)のための領域が確保される(ステップ903)。次いで、合併された形態のPVやVLPの値が更新される(ステップ904、905)。上記TRは、PVを再定義するために利用されるため、位置指定再定義配列といえることができる。

10

## 【 0 0 2 4 】

さらに、あるレベルの部分中間ブロック群において末尾の部分中間ブロックまで、合併の処理が終了すると、合併された部分中間ブロックの数が1つであるか否かが判断される(ステップ906)。ここで「イエス(Yes)」と判断されれば、当該1つの部分中間ブロックが完成した情報ブロックとなる。その一方、「ノー(No)」と判断されれば、合併後の複数の部分中間ブロックを、処理対象として再度ステップ902~905の処理が進められる。

図6(a)に示すレコード番号と、「姓」という項目の対応する項目値とを例にとって、図9の処理、特にステップ902~904の処理を詳細に説明する。

20

## 【 0 0 2 5 】

まず、図17に示すように、9つの初期的な部分中間ブロックが生成される(図9のステップ901参照)。ここで、冒頭の部分中間ブロック1701-0および部分中間ブロック1701-1が合併され、PVやVLPが更新される(図9のステップ904参照)。図10は、ステップ904の処理をより詳細に示す図である。この処理において、合併される二つの部分中間ブロックのうち、より小さなレコード番号に関するものを第1の部分中間ブロック、より大きなレコード番号に関するものを第2の部分中間ブロックと称する。また、第1の部分中間ブロックにおいて、比較対象となる項目値を示すために、VLPの値を指すように配置されたポイントを第1の比較位置ポイントと称し、第2の部分中間ブロックにおいて、比較対象となる項目値を示すために、VLPの値を指すように配置されたポイントを第2の比較位置ポイントと称する。この比較位置ポイントは、図10の処理の冒頭では、各部分中間ブロックにおいて最も小さいレコード番号を有するVLPの値を指している。さらに、この処理において生成されるVLP'において値を格納すべき位置を示すポイントを格納位置ポイントと称する。これら第1の比較位置ポイント、第2の比較位置ポイントおよび格納位置ポイントは、図10の処理に先立って、初期位置(各VLPにおける先頭の位置)に設定される。

30

## 【 0 0 2 6 】

図10に示すように、コンピュータシステム10は、第1の比較位置ポイントにて指示される項目値(第1の指示値)と、第2の比較位置ポイントにて指示される項目値(第2の指示値)とを取り出して、これらの値を比較する(ステップ1001~1003)。以下、第1の指示値および第2の指示値の大小、或いは、第1の比較ポイントおよび第2の比較ポイントの位置により、異なる処理に進む(ステップ1004~ステップ1014参照)。

40

図15に示す部分中間ブロック1501-0、1501-1に着目すると、初期的には、第1の比較位置ポイントは、部分中間ブロック1501-0における「たなか」を指示し、その一方、第2の比較位置ポイントは、部分中間ブロック1501-1における「あべ」を指示する。したがって、これらを比較すると(図10のステップ1003参照)、「第1の指示値>第2の指示値」となる(ステップ1008参照)。この場合には、ステップ1009に進み、図13に示す処理が実行される。

## 【 0 0 2 7 】

50

より詳細には、第2の比較位置ポイントが示す項目値「あべ」と対応するVLPの値「1」が特定され(ステップ1301)、当該VLPの値「1」が、VLP'中の格納位置ポイントにて示される位置(先頭つまり格納位置番号「0」の位置)に格納される(ステップ1302)。次いで、第2の比較位置ポイントの位置に対応するTRの値として、格納位置ポイントの位置している格納位置番号(この場合には「0」)が与えられる(ステップ1303)。

その後、コンピュータシステム10は、第2の比較位置ポイントを一つ下流側(格納位置番号が大きくなる側)に移動させる(ステップ1304)とともに、格納位置ポイントも一つ下流側に移動させる(ステップ1305)。

#### 【0028】

このような処理の後、再度ステップ1001~ステップ1003の処理が実行される。しかしながら、上記例では第2の比較位置ポイントを一つ下流側に移動させると、第2の比較位置ポイントが指し示す項目値が存在しない。このため、ここではステップ1012にてイエス(Yes)と判断され、図12の処理が実行される。

#### 【0029】

より詳細には、第1の比較位置ポイントに示す項目値「たなか」と対応するVLPの値「0」が特定され(ステップ1201)、これが、VLP'中、格納位置ポイントが示す位置(2番目の位置つまり格納位置番号「1」にて示される位置)に格納される(ステップ1202)。次いで、第1の比較位置ポイントの位置に対応するTRの値として、格納位置ポイントの位置している格納位置番号(この場合には「1」)を与える(ステップ1203)。その後、第2の比較位置ポイントを一つ下流側(格納位置番号が大きくなる側)に移動させる(ステップ1204)とともに、格納位置ポイントも一つ下流側に移動させる(ステップ1205)。

さらに、再度ステップ1001~ステップ1003が実行されるが、第1の比較位置ポイントおよび第2の比較位置ポイントとも、指示する項目値が存在しない。このため、ここではステップ1004~ステップ1012においてノー(No)と判断されてステップ1014に進む。

#### 【0030】

より詳細には、第1の部分中間ブロック(図15の符号1501-0参照)において、PVの値を、それぞれ、対応するTRの値で変換して、新たなPVを生成する(ステップ1401)とともに、第2の部分中間ブロック(図15の符号1501-1参照)において、PVの値を、それぞれ、対応するTRの値で変換して、新たなPVを生成する(ステップ1402)。図16に、新たなPVやVLPの生成についてより具体的に説明する。たとえば、第1の部分中間ブロック1501-0中のPV(旧PV)1601の要素(値)「0」を取り出し、この要素にて示される格納位置番号のTR1602の要素(値)を特定する。この要素(値)は「1」であるため、新たなPV(新PV)1603において、当該TRの要素にて示される格納位置番号に示される位置に、旧PVの要素「0」を収容すればよい。第2の部分中間ブロックにおいても同様の処理が進められ、これにより、新PVを得ることが可能となる。

また、図10のステップ1001~1013の処理を繰り返すことによって作り出されたVLP'が、新たなVLPとなる(ステップ1403、図16の符号1604参照)。

#### 【0031】

他の部分中間ブロックの組でも図10の処理が実行されて、新PVおよび新VLPが作り出される。図10~図14の処理においてまだ説明していないものを中心に、他の部分中間ブロックの組に関する処理につき説明する。図15に示す部分中間ブロック1501-4、1501-5に着目すると、初期的には、第1の比較位置ポイントは、部分中間ブロック1501-4における「あべ」を指示し、その一方、第2の比較位置ポイントは、部分中間ブロック1501-5における「たなか」を指示する。したがって、これらと比較すると、「第1の指示値<第2の指示値」となり(ステップ1006でイエス(Yes))、ステップ1007に進み、図12の処理が実行される。

10

20

30

40

50

## 【 0 0 3 2 】

図 1 2 においては、第 1 の比較位置ポインタが示す項目値「あべ」と対応する V L P の値「4」が特定され(ステップ 1 2 0 1)、当該 V L P の値「4」が、V L P ' 中の格納位置ポインタにて示される位置(格納位置番号「4」の位置)に格納される(ステップ 1 2 0 2)。次いで、第 1 の比較位置ポインタの位置に対応する T R の値として、格納位置ポインタの位置している格納位置番号(この場合には「4」)が与えられる(ステップ 1 2 0 3)。

その後、コンピュータシステム 1 0 は、第 1 の比較位置ポインタを一つ下流側(格納位置番号が大きくなる側)に移動させる(ステップ 1 2 0 4)とともに、格納位置ポインタも一つ下流側に移動させる(ステップ 1 2 0 5)。

10

## 【 0 0 3 3 】

このような処理の後、再度ステップ 1 0 0 1 ~ ステップ 1 0 0 3 の処理が実行される。しかしながら、この例では第 1 の比較位置ポインタを一つ下流側に移動させると、第 1 の比較位置ポインタが指し示す項目値が存在しない。このため、ここでは、ステップ 1 0 1 0 にてイエス(Yes)と判断され、図 1 3 に示す処理に進む。

## 【 0 0 3 4 】

図 1 3 の処理では、第 2 の比較位置ポインタが示す項目値「たなか」と対応する V L P の値「5」が特定され、これが、V L P ' 中、格納位置番号「5」にて示される位置に格納される(ステップ 1 3 0 1、1 3 0 2)。また、第 2 の比較位置ポインタに対応する T R の値として、格納位置番号「5」が与えられる(ステップ 1 3 0 3)。次いで、第 2 の比較位置ポインタおよび格納位置ポインタが一つ下流側に移動される(ステップ 1 3 0 4、1 3 0 5)。

20

この時点で、第 1 の比較位置ポインタおよび第 2 の比較位置ポインタの双方とも、指示する項目値が存在しないため、ステップ 1 0 1 4 に進み、新 P V および新 V L P が作り出される(ステップ 1 4 0 1 ~ 1 4 0 3 参照)。

## 【 0 0 3 5 】

次に、図 1 5 に示す部分中間ブロック 1 5 0 1 - 6、1 5 0 1 - 7 に着目すると、初期的には、第 1 の比較位置ポインタは、部分中間ブロック 1 5 0 1 - 7 における「いとう」を指示し、その一方、第 2 の比較位置ポインタは、部分中間ブロック 1 5 0 1 - 7 における「いとう」を指示する。したがって、これらと比較すると、「第 1 の指示値 = 第 2 の指示値」となり(ステップ 1 0 0 4 でイエス(Yes))、ステップ 1 0 0 5 に進む。

30

図 1 1 においては、第 1 の比較位置ポインタが示す項目値「いとう」と対応する V L P の値「6」が特定され(ステップ 1 1 0 1)、当該 V L P の値「6」が、V L P ' 中の格納位置ポインタにて示される位置(格納位置番号「6」の位置)に格納される(ステップ 1 1 0 2)。次いで、第 1 の比較位置ポインタに対応する T R の値として、格納位置ポインタの位置している格納位置番号「6」が与えられる(ステップ 1 1 0 3)。その後、第 1 の比較位置ポインタが一つ下流側に移動される。

## 【 0 0 3 6 】

同様に、第 2 の比較位置ポインタの位置に対応する T R の値として、格納位置ポインタが位置している格納位置番号「6」が与えられ(ステップ 1 1 0 5)、第 2 の比較位置ポインタが一つ下流側に移動される(ステップ 1 1 0 6)。

40

このように、双方の部分中間ブロックにおいて、同じ項目値が存在する場合には、その重複を排除する形で、V L P ' に値が収容される。

その後の処理では、第 1 の比較位置ポインタ、第 2 の比較位置ポインタとも、指示する項目値が存在しないため、図 1 4 に進み、新 P V および新 V L P が作られる。

## 【 0 0 3 7 】

このようにして、図 1 5 の符号 1 5 1 1 - 0 ~ 1 5 1 1 - 4 に示すような新たな部分中間ブロックを得ることができる。これら部分中間ブロックについて、図 9 ~ 図 1 4 の処理が実行される。ここでは、図 1 7 に示すように、部分中間ブロック 1 5 1 1 - 0 をそのままにして、部分中間ブロック 1 5 1 1 - 1、1 5 1 1 - 2 の組、部分中間ブロック 1 5 1 1

50

- 3、1511-4の組を作り、部分中間ブロックの組のそれぞれに関する処理を実行している。

部分中間ブロック1511-1、1511-2の組に関する処理について、再度、フローチャート等を参照して説明を加える。図18(a)の矢印A、BおよびCに示すように、初期的には、第1の比較位置ポインタ、第2の比較位置ポインタおよび第3の比較位置ポインタは、それぞれ、VLP、VLP'の先頭に配置される。

【0038】

図10にしたがって、第1の比較位置ポインタ(図18(a)の矢印A)が指し示すVLPの値「3」にて特定される項目値「いとう」が取り出され(ステップ1011参照)、その一方、第2の比較位置ポインタ(同図の矢印B)が指し示すVLPの値「4」にて特定される項目値「あべ」が取り出される(ステップ1012)。

第1の指示値「いとう」と第2の指示値「あべ」とを比較すると、第1の指示値>第2の指示値となるため(ステップ1008でイエス(Yes))、図13の処理が実行される。これにより、図18(b)に示すように、格納位置ポインタ(矢印C)にて示される位置に、VLPの値「4」が配置されるとともに、所定のTR中の値も配置される。その後、第2の比較位置ポインタおよび格納位置ポインタが、一つ下流側に移動される(矢印B'および矢印C')参照。

【0039】

次いで、第1の比較位置ポインタ(図18(b)の矢印A)が指し示すVLPの値「3」にて特定される項目値「いとう」と、第2の比較位置ポインタ(同図の矢印B')が指し示すVLPの値「5」にて特定される項目値「たなか」とが比較される(図10のステップ1001~1003)。この場合には、第1の指示値<第2の指示値となるため、ステップ1007に進み、図18(c)に示すように、格納位置ポインタ(同図矢印C)にて示される位置に、第1の比較位置ポインタにて示されるVLPの値「3」が配置されるとともに、所定のTRの値も配置される。また、第1の比較位置ポインタおよび格納位置ポインタが一つ下流側に移動される(矢印A'および矢印C')参照。

【0040】

さらに、第1の比較位置ポインタ(図18(c)の矢印A')が指し示すVLPの値「2」にて特定される項目値「たなか」と、第2の比較位置ポインタ(同図の矢印B)が指し示すVLPの値「5」にて特定される項目値「たなか」とが比較される(図10のステップ1001~1003参照)。この場合には、第1の指示値=第2の指示値となるため、ステップ1005に進む。ここでは、図18(d)に示すように、格納位置ポインタ(矢印C)にて示される位置に、第1の比較位置ポインタにて示されるVLPの値「2」が配置される。また、第1の比較位置ポインタおよび第2の比較位置ポインタにて示される位置に対応するTRの値として、格納位置ポインタが指し示す格納位置番号「4」が配置される。

【0041】

このような処理の後、再度図10の処理が進められると、ステップ1014に進む。ここでは、旧PVをTRにて変換して新たなPVを作る処理が実行される。旧PVの値が取りだされ、当該値が指し示す位置のTRの値を特定して、当該TRの値を、新PVにおいて前記旧PVを取り出した値の位置と対応する位置に配置する。この処理を繰り返すことで、図20(b)に示すような新PV2001を得ることが可能となる。

【0042】

次に、部分中間ブロック1511-3、1511-4からなる組に関する処理についても簡単に説明する。ここでも、先の部分中間ブロックの組と同様の処理が進められる。たとえば、最初に図10の処理を施すことにより、図19(b)に示すように、格納位置ポインタ(図19(b)の矢印C)の位置に、第2の格納位置ポインタ(同図の矢印B)にて指し示されるVLPの値「8」が配置されるとともに、対応するTRの値「6」が決定される。また、次に図10の処理を施すことにより、図19(c)に示すように、格納位置ポインタ(図19(c)の矢印C)の位置に、第1の比較位置ポインタ(図19(c)の

矢印 A) にて指し示される V L P の値「6」が配置されるとともに、所定の T R の値（第 1 の比較位置ポイントおよび第 2 の比較位置ポイントが指し示す位置に対応する T R の値）が決定される。

【0043】

最後に、もう一度図 10 の処理を施すことにより、ステップ 1014 において新たな P V が作られる（図 20 (b) 参照）。

これにより、図 17 の符号 1701 - 1、1701 - 2 に示すような部分中間ブロックを作り出すことができる。

【0044】

さらに、この例では、部分中間ブロック 1701 - 0（実際には、部分中間ブロック 1501 - 0 と同じ）と部分中間ブロック 1701 - 1 とを併合している。ここでも、図 10 に示す処理が繰り返され、部分中間ブロック 1701 - 0 と部分中間ブロック 1701 - 1 との間で、比較位置ポイントにて指し示す V L P の値により特定される項目値の比較結果にしたがって、V L P ' の値や T R の値が順次決定される。

【0045】

図 21 に示すように、部分中間ブロック 1701 - 0 と部分中間ブロック 1701 - 1 に関して、図 10 の処理が施される。図 21 (a) に示すように、最初に行われる図 10 の処理では、第 1 の比較位置ポイント（図 21 (a) の矢印 A）が示す V L P の値「1」にて特定される項目値「あべ」と、第 2 の比較位置ポイント（図 21 (b) の矢印 B）が示す V L P の値「4」にて特定される項目値「あべ」とが比較される。この比較結果に基づき図 11（ステップ 1005）の処理が実行されて、格納位置ポイント（図 21 (a) の矢印 C）に示す V L P ' 中の値「1」が決定されるとともに、第 1 および第 2 の比較位置ポイントの位置に対応する T R 中の値「0」が決定される。

【0046】

また、図 21 (b) に示すように、次に実行される図 10 の処理では、第 1 の比較位置ポイント（図 21 (b) の矢印 A）が示す V L P の値「0」にて特定される項目値「たなか」と、第 2 の比較位置ポイント（図 21 (b) の矢印 B）が示す V L P の値「3」にて特定される項目値「いとう」とが比較される。この比較結果に基づき、図 13 の処理（ステップ 1009）が実行されて、格納位置ポイント（図 21 (b) の矢印 C）に示す V L P ' 中の値「3」が決定されるとともに、第 2 の比較位置ポイントの位置に対応する T R 中の値「1」が決定される。

【0047】

さらに、図 21 (c) に示すように、次に実行される図 10 の処理では、第 1 の比較位置ポイント（図 21 (c) の矢印 A）が示す V L P の値「0」にて特定される項目値「たなか」と、第 2 の比較位置ポイント（図 21 (c) の矢印 B）が示す V L P の値「2」にて特定される項目値「たなか」とが比較される。この比較結果に基づき、図 11 の処理（ステップ 1005）が実行されて、格納位置ポイント（図 21 (c) の矢印 C）に示す V L P ' の値「0」が決定されるとともに、第 1 および第 2 の比較位置ポイントの位置に対応する T R 中の値「2」が決定される。この後に実行される図 10 の処理では、第 1 の比較位置ポイントおよび第 2 の比較位置ポイントが示す V L P の値が存在しないため、図 14 の処理（ステップ 1014）が実行され、P V（旧 P V）の値が、関連する T R にて変換されて、新たな P V（新 P V）を得ることができる（図 23 (a) 参照）。

このような処理により、図 17 の符号 1711 - 0 にて示すような部分中間ブロックが生成される。最後に、部分中間ブロック 1711 - 0 と部分中間ブロック 1711 - 1（実際には部分中間ブロック 1701 - 2 と同じ）との組を作り、この部分中間ブロックの組について所定の処理を実行する。

【0048】

上記部分中間ブロック 1711 - 0、1711 - 1 の組に関する処理について、フローチャート等を参照して説明を加える。たとえば、最初に図 10 の処理においては、図 22 (a) に示すように、第 1 の比較位置ポイント（図 22 (a) の矢印 A）が示す V L P の値

10

20

30

40

50

「1」にて特定される項目値「あべ」と、第2の比較位置ポインタ(図22(b)の矢印B)が示すVLPの値「8」にて特定される項目値「あべ」とが比較される。この比較結果に基づき図11(ステップ1005)の処理が実行されて、格納位置ポインタ(図22(a)の矢印C)に示すVLP'中の値「1」が決定されるとともに、第1および第2の比較位置ポインタの位置に対応するTR中の値「0」が決定される。

【0049】

また、図22(b)に示すように、次に実行される図10の処理では、第1の比較位置ポインタ(図22(b)の矢印A)が示すVLPの値「3」にて特定される項目値「いとう」と、第2の比較位置ポインタ(図22(b)の矢印B)が示すVLPの値「3」にて特定される項目値「いとう」とが比較される。この比較結果に基づき、図11の処理(ステップ1005)が実行されて、格納位置ポインタ(図22(b)の矢印C)に示すVLP'中の値「3」が決定されるとともに、第1および第2の比較位置ポインタの位置に対応するTRの値「1」が決定される。

【0050】

さらに、図22(c)に示すように、次に実行される図10の処理では、第1の比較位置ポインタ(図22(c)の矢印A)が示すVLPの値「0」は存在するが、第2の比較位置ポインタ(図22(c)の矢印B)が示すVLPの値が存在しないため、図12の処理(ステップ1013)に進み、格納位置ポインタ(図22(c)の矢印C)に示すVLP'の値「0」が決定されるとともに、第1の比較位置ポインタの位置に対応するTR中の値「2」が決定される。この後に実行される図10の処理では、第1の比較位置ポインタおよび第2の比較位置ポインタが示すVLPの値が存在しないため、図14の処理(ステップ1014)が実行され、PV(旧PV)の値が、関連するTRにて変換されて、新たなPV(新PV)を得ることができる(図23(b)参照)。

【0051】

このようにして、図6(b)に示すように、VP601と、VLP604およびオリジナル項目値配列603からなる値リスト602を有する情報ブロック600を得ることができる。ここで、オリジナル項目値配列においては、実際には、格納位置番号「0」、「1」および「3」に対応する要素(項目値)「たなか」、「あべ」および「いとう」(符号610、611、613参照)のみが利用され、他の要素(項目値)(たとえば、符号612、614参照)が利用されていない、つまり、VLPの値に指し示されていないことが理解できる。すなわち、オリジナル項目値配列において、項目値が重複している場合には、もっとも格納位置番号の小さなもののみが利用され、他のものは利用されない。

【0052】

この形式であっても、図3(b)に示す情報ブロックと同等であることにつき以下に説明を加える。たとえば、図6(a)においては、レコード番号「0」に対応して項目値「たなか」が特定されている。図6(b)においても、レコード番号「0」に対応して、PV中、格納位置番号「0」の要素(値)「2」が特定される。この要素「2」を格納位置番号とするように、VLP中の要素(値)を特定すると、「0」を得ることができ、さらに、当該VLPの要素(値)「0」を格納位置番号とするオリジナル項目値配列の項目値「たなか」を特定することが可能となる。他のレコード番号についても同様に、項目値を得られることは明らかである。

【0053】

したがって、図6(b)に示すVLP604およびオリジナル項目値配列603からなる値リスト602を利用して、検索、集計およびソートを実現することも可能である。ただし、この場合には、オリジナル項目値配列をそのまま利用することになる。図6(b)から明らかなように、オリジナル項目値配列においては、項目値が重複する場合には、格納位置番号が最も小さいもののみが利用され、残りの項目値が利用されない。したがって、重複が多い場合には、オリジナル項目値配列において利用されない部分が多く出現する。そこで、VLPの値にて示す項目値を、新たな配列に収容することで、重複のまったくない項目値配列(圧縮型値リスト)を作成することが可能となる。

10

20

30

40

50

## 【 0 0 5 4 】

図 2 4 は、圧縮型値リストを生成する処理を示すフローチャートである。図 2 4 に示すように、コンピュータシステム 1 0 は、格納位置番号を初期化するとともに、圧縮型値リストのための領域を確保する（ステップ 2 4 0 1）。この領域に配置されるデータ（項目値）のレコード数は、V L P に含まれる値の数と対応するため、この値の数に対応して領域を確保すればよい。

## 【 0 0 5 5 】

次いで、格納位置番号が示す V L P 中の値が取り出され（ステップ 2 4 0 2）、当該 V L P の値が示す、オリジナル項目値配列中の項目値が特定される（ステップ 2 4 0 3）。その後、特定された項目値が、圧縮型値リスト中、格納位置番号が示す位置に配置される（ステップ 2 4 0 4）。この処理は、全ての V L P の値に関して実行される（ステップ 2 4 0 5、2 4 0 6 参照）。たとえば、図 6（b）に示す値リスト 6 0 2 に関して、図 2 4 の処理を施すと、図 2 5（a）に示すように、V L P の値（たとえば、格納位置番号「0」にて示される値「1」）によりオリジナル項目値配列中の項目値（たとえば「あべ」）が特定され（ステップ 2 4 0 2、2 4 0 3）る。そして、この項目値が、圧縮型値リストの格納位置番号にて示される位置（たとえば「0」の位置）に配置される。

10

## 【 0 0 5 6 】

このような処理の結果、図 6（b）に示す情報ブロック 6 0 0 は、図 2 5（b）に示すように、P V 6 0 1 と圧縮型値リスト 2 5 0 1 とを有する情報ブロック 2 5 0 0 に変換される。新たになに得られた情報ブロック 2 5 0 0 は、図 3（b）に示す情報ブロックと同じ形式であることが理解できるであろう。図 2 4 の処理の結果得られた情報ブロック 2 5 0 0 に基づき、検索、集計およびソートの処理を実行できることは言うまでもない。

20

## 【 0 0 5 7 】

上記コンパイル処理において、図 1 0 に示す処理を繰り返して二つの部分中間ブロックを併合することにより得られた配列につき再度考察を加える。たとえば、図 1 8（d）の V L P ' および項目値配列を参照する。新たな V L P（新 V L P）である V L P '（符号 1 8 0 1 参照）中、格納位置番号「2」の位置に収容された値「4」にて特定される、項目値配列（符号 1 8 0 2 参照）の要素（値）は、「あべ」である。また、新 V L P 1 8 0 1 中、格納位置番号「3」の位置に収容された値「3」にて特定される、項目値配列 1 8 0 2 の要素（値）は「いとう」である。さらに、新 V L P 1 8 0 1 中、格納位置番号「4」の位置に収容された値「2」にて特定される、項目値配列 1 8 0 2 の要素（値）は「たなか」である。新 V L P および項目値配列により、併合された部分中間ブロックに関する、項目値がソートされた状態の値リストが作られていることが理解できる。すなわち、ある二つの部分中間ブロックに関して図 1 0 の処理を繰り返し実行することにより、上記二つの部分中間ブロックを併合した部分中間ブロックに関して、項目値が所定の順序で並べられかつ重複が排除された値リストが作られる。

30

## 【 0 0 5 8 】

上記構成のデータのコンパイル方法における処理速度について考察を加える。ここでは説明を容易にするために、値リストを、V L P および項目値配列の形式ではなく、V L P の値により項目値配列の値を特定し、これを並べた配列（圧縮型値リスト）の形式として説明する。

40

図 2 6 に示すように、8 つのレコードに、0 ~ 7 の何れかの数字が項目値として割り当てられていると考える。ここで、各レコードに対応する項目値は全てことなると考える。この場合に、本実施の形態においては、レコード番号が「0」および「1」の値リスト（符号 2 6 0 1 - 0、2 6 0 1 - 1 参照）に関する部分中間ブロックの組、レコード番号「2」および「3」の値リストに関する部分中間ブロックの組など、4 つの部分中間ブロックの組（図 2 6（b）参照）のそれぞれに関して図 1 0 の処理が繰り返し実行される。

## 【 0 0 5 9 】

上記例では全ての項目値が異なるため、4 回の項目値の比較、すなわち、「2」と「6」との比較、「4」と「0」との比較、「3」と「7」との比較および「5」と「1」との

50

比較が行なわれる。これを一般化すると、要素が「 $n$ （ただし $n$ は2の累乗であるのが望ましい）」であれば、最初の処理においては、「 $n / 2^1$ 」回の比較が実行されることになる。

【0060】

次の部分中間ブロックの併合により4つの部分中間ブロックから、2つの部分中間ブロックが作られる（図26（b）および（c）参照）。ここでは、部分中間ブロック2602-0、2602-1に関して、3回の項目値の比較、つまり、「2」と「0」との比較、「2」と「4」との比較、および、「6」と「4」との比較が実行され、残りの情報ブロックに関して、3回の項目値の比較、つまり、「3」と「1」との比較、「3」と「5」との比較、および、「7」と「5」との比較が実行される。これを一般化すると、要素が「 $n$ （ただし $n$ は2の累乗であるのが望ましい）」であれば、最大で「 $(n / 2^2) * (2^2 - 1)$ 」回の比較が実行されることになる。

10

さらに、2つの部分中間ブロックから最終的な情報ブロックが作られる（図26（c）および（d）参照）。ここでは、7回の項目値の比較、つまり、「0」と「1」、「2」と「1」、「2」と「3」、「4」と「3」、「4」と「5」、「6」と「5」、および、「6」と「7」の比較が実行される。これを一般化すると、要素が「 $n$ （ただし $n$ は2の累乗であるのが望ましい）」であれば、最大で「 $(n / 2^3) * (2^3 - 1)$ 」回の比較が実行されることになる。

【0061】

上記比較回数を検討すると、部分中間ブロックの併合において以下の回数だけ比較処理が実行される。

20

1回目                               :  $(n / 2^1)$  回  
 2回目                               :  $(n / 2^2) * (2^2 - 1)$  回  
 3回目                               :  $(n / 2^3) * (2^3 - 1)$  回  
 :  
 $\log_2(n)$  回目                       :  $(n / 2^{\log_2(n)}) * (2^{\log_2(n)} - 1)$  回

【0062】

情報ブロック生成に必要な比較回数は、上記各併合処理における比較回数の総和と考えられる。したがって、総比較回数は、以下の式1により得ることが可能である。

30

【数1】

数1

$$\log_2(n) \sum_{j=1} n * (2^j - 1) / 2^j \doteq n * (\log_2(n) - 1) \dots (1)$$

このように、項目値に重複のない場合（各項目値が異なる場合）には、最大で、約「 $n * \log_2(n) - 1$ 」回の比較処理が行なわれることが理解できる。

40

【0063】

これに対して、「 $n$ 」個の項目値のうち、「 $m$ 」個が異なる場合を考える。この場合には、部分中間ブロックの併合において以下の回数だけ比較処理が実行される。

1回目 :  $(n/2^1)$  回  
 2回目 :  $(n/2^2) * \text{Max}((m/n) * 2^2 - 1, 1)$  回  
 3回目 :  $(n/2^3) * \text{Max}((m/n) * 2^3 - 1, 1)$  回  
 :  
 $\log_2(n)$  回目 :  $(n/2^{\log_2(n)}) * \text{Max}((m/n) * 2^{\log_2(n)} - 1, 1)$  回

ここで、「 $n$ 」に対して「 $m$ 」が十分に小さい場合（たとえば、 $m = 1 \sim 4$  程度の場合）には、 10

1回目 :  $(n/2^1)$  回  
 2回目 :  $(n/2^2)$  回  
 3回目 :  $(n/2^3)$  回  
 :  
 $\log_2(n)$  回目 :  $(n/2^{\log_2(n)})$  回

と考えることができるため、総比較回数は約「 $n$ 」回と考えることができる。

【0064】

その一方、「 $n$ 」が「 $m$ 」に対して比較的大きい場合には、

1回目 :  $(n/2^1)$  回 20  
 2回目 :  $(n/2^2) * ((m/n) * 2^2 - 1)$  回  
 3回目 :  $(n/2^3) * ((m/n) * 2^3 - 1)$  回  
 :  
 $\log_2(n)$  回目 :  $(n/2^{\log_2(n)}) * ((m/n) * 2^{\log_2(n)} - 1)$  回

と考えることができるため、総比較回数は、以下の式2となる。

【数2】

数2

$$(n/2^1) + \sum_{j=2}^{\log_2(n)} m - \sum_{j=2}^{\log_2(n)} n/2^j$$

$$\doteq m * (\log_2(n) - 1) \dots (2)$$

【0065】

このように、「 $n$ 」個の項目値中、「 $m$ 」個が異なる場合には、「 $m$ 」の大きさにしたがって以下のように比較処理が実行される。 40

「 $m$ 」が十分に小さい場合：約「 $n$ 」回の比較

「 $m$ 」が比較的大きい場合：約「 $m * \log_2(n) - 1$ 」回の比較

したがって、特に項目値の重複が大きい場合には、ほぼ「 $n$ 」のオーダーで情報ブロックを生成することが可能となる。

【0066】

本実施の形態によれば、部分中間ブロックの対からなる新たな部分中間ブロックを作成して、この新たな部分空間ブロックにおける項目値の比較によって、部分中間ブロックにおける項目値配列を作成し、部分中間ブロックを漸次大きくすることにより、最終的に一つの部分中間ブロックを得て、これを情報ブロックとしている。したがって、対に関する処理を、複数のプロセッサによりパラレルに実施することが可能となる。 50

## 【 0 0 6 7 】

本発明は、以上の実施の形態に限定されることなく、特許請求の範囲に記載された発明の範囲内で、種々の変更が可能であり、それらも本発明の範囲内に包含されるものであることは言うまでもない。

前記実施の形態にかかる部分中間ブロックの対に対する処理は、シングルプロセッサにより実行されても良いし、パラレルプロセッサを有するシステムにおいて、ある対の処理をあるプロセッサに割り当てることにより、より高速化を図っても良いことは言うまでもない。

## 【 0 0 6 8 】

また、部分中間ブロックの対の選択は、前記実施の形態に示すものに限定されず、隣接する二つの部分中間ブロックであれば所望のように、対とすることが可能である。

さらに、前記実施の形態においては、一般のコンピュータシステム 10 内に、所定のプログラムを読み込み、当該プログラムを実行することにより、複数の表形式データのジョインおよびジョインされた表形式データに関する処理を実現しているが、本発明はこれに限定されるものではなく、パーソナルコンピュータ等のような一般のコンピュータシステムに、データベース処理専用のボードコンピュータを接続し、当該ボードコンピュータが上記処理を実行できるように構成しても良いことは言うまでもない。したがって、本明細書において、手段とは必ずしも物理的手段を意味するものではなく、各手段の機能が、ソフトウェアによって実現される場合も包含する。さらに、一つの手段の機能が、二つ以上の物理的手段により実現されても、若しくは、二つ以上の手段の機能が、一つの物理的手段により実現されてもよい。

## 【 0 0 6 9 】

## 【発明の効果】

本発明によれば、高速な処理時間と、データの分布によっても処理効率が極端に悪化しないことが予測できるデータのコンパイル方法を提供することが可能となる。

また、本発明によれば、特に、末子部パラレルシステムにおいてパフォーマンスを劇的に改善できるデータのコンパイル方法を提供することが可能となる。

## 【図面の簡単な説明】

【図 1】 図 1 は、本発明の実施の形態にかかる検索、集計およびサーチ方法を実現できるコンピュータシステムのハードウェア構成を示すブロックダイヤグラムである。

【図 2】 図 2 は、本実施の形態にて用いる情報ブロックを示す図である。

【図 3】 図 3 は、表形式データの例、および、当該表形式データに基づく情報ブロックの例を示す図である。

【図 4】 図 4 は、表形式データの他の例、および、当該表形式データに基づく情報ブロックの他の例を示す図である。

【図 5】 図 5 は、単一項目に関する検索手法を示すフローチャートである。

【図 6】 図 6 は、ある項目に関するレコード番号と項目値との関係、および、完成した本実施の形態にかかる情報ブロックを示す図である。

【図 7】 図 7 は、本実施の形態にかかるコンパイル処理の概略を示すフローチャートである。

【図 8】 図 8 は、本実施の形態にかかる部分中間ブロックの一例を示す図である。

【図 9】 図 9 は、本実施の形態において二つの部分中間ブロックを併合して新たな部分中間ブロックを生成する処理の概略を示すフローチャートである。

【図 10】 図 10 は、図 9 のステップ 904 の処理をより詳細に示すフローチャートである。

【図 11】 図 11 は、図 10 のステップ 1005 をより詳細に示すフローチャートである。

【図 12】 図 12 は、図 10 のステップ 1007、1013 をより詳細に示すフローチャートである。

【図 13】 図 13 は、図 10 のステップ 1009、1011 をより詳細に示すフローチャートである。

10

20

30

40

50

ャートである。

【図 1 4】 図 1 4 は、図 1 0 のステップ 1 0 1 4 をより詳細に示すフローチャートである。

【図 1 5】 図 1 5 は、初期的な部分中間ブロックが併合されて新たな部分空間ブロックが生成された例を示す図である。

【図 1 6】 図 1 6 は、本実施の形態において新たな P V を作成するための処理を示す図である。

【図 1 7】 図 1 7 は、本実施の形態において部分中間ブロックの併合の一例を示す図である。

【図 1 8】 図 1 8 は、本実施の形態において部分中間ブロックの併合処理を説明するための図である。

10

【図 1 9】 図 1 9 は、本実施の形態において部分中間ブロックの併合処理を説明するための図である。

【図 2 0】 図 2 0 は、本実施の形態において新たな P V を作成するための処理を示す図である。

【図 2 1】 図 2 1 は、本実施の形態において部分中間ブロックの併合処理を説明するための図である。

【図 2 2】 図 2 2 は、本実施の形態において部分中間ブロックの併合処理を説明するための図である。

【図 2 3】 図 2 3 は、本実施の形態において新たな P V を作成するための処理を示す図である。

20

【図 2 4】 図 2 4 は、本実施の形態において、圧縮型値リストを生成する処理を示すフローチャートである。

【図 2 5】 図 2 5 は、本実施の形態において、圧縮型値リストを生成する処理を説明するための図である。

【図 2 6】 図 2 6 は、本実施の形態にかかる処理速度を考察するために用いられた配列の一例を示す図である。

【符号の説明】

1 0	コンピュータシステム
1 2	C P U
1 4	R A M
1 6	R O M
1 8	固定記憶装置
2 0	C D - R O M ドライバ
2 2	I / F
2 4	入力装置
2 6	表示装置

30

【図1】

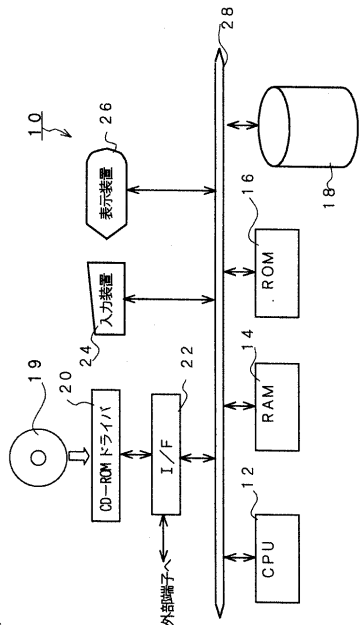
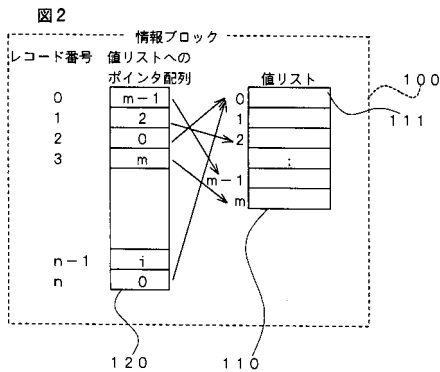


図1

【図2】



【図3】

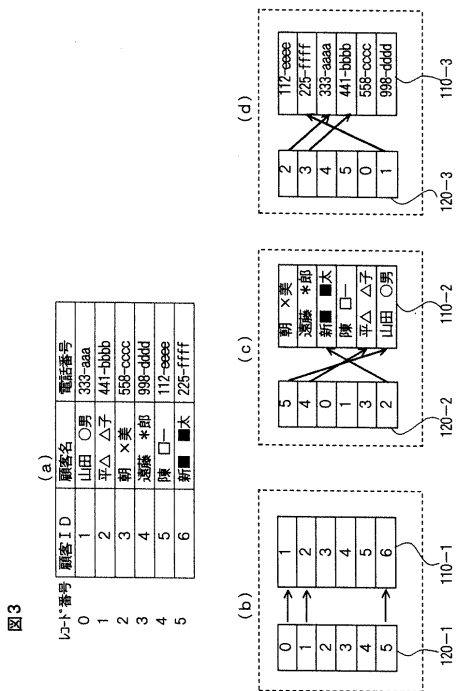


図3

【図4】

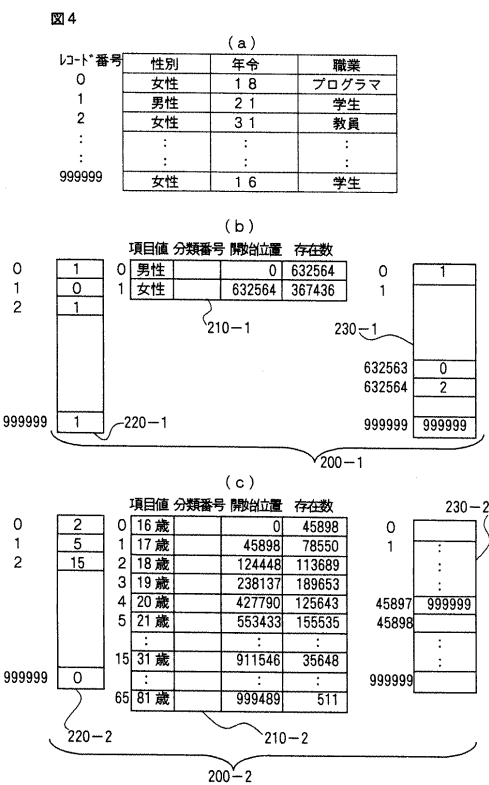
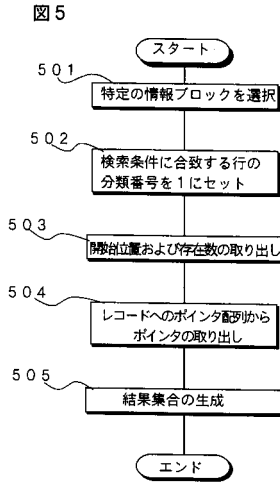
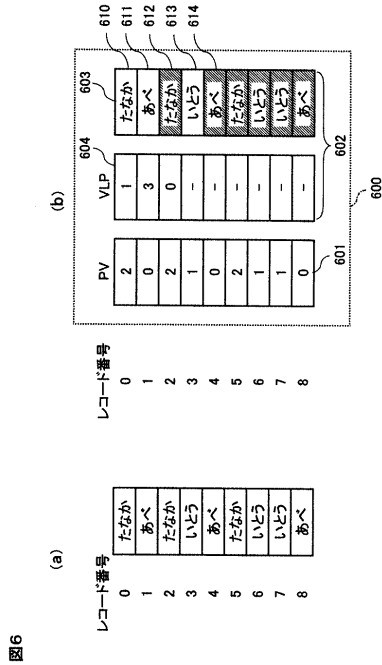


図4

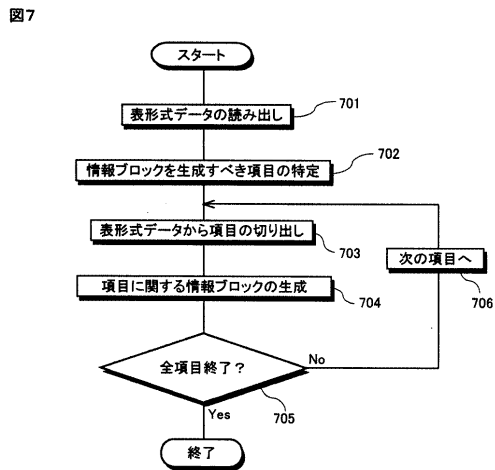
【図5】



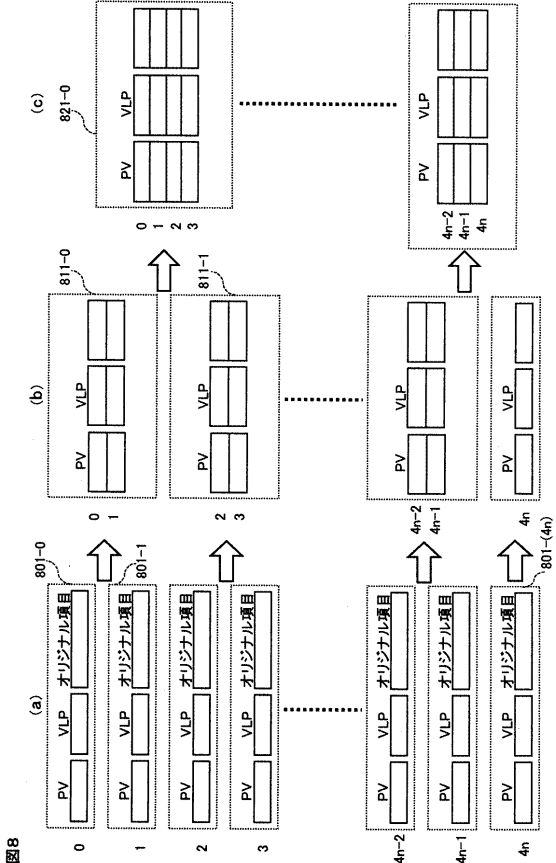
【図6】



【図7】



【図8】



【図9】

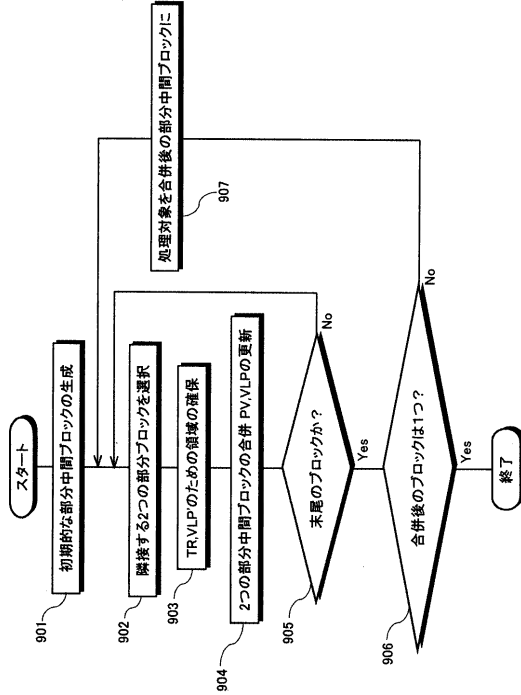
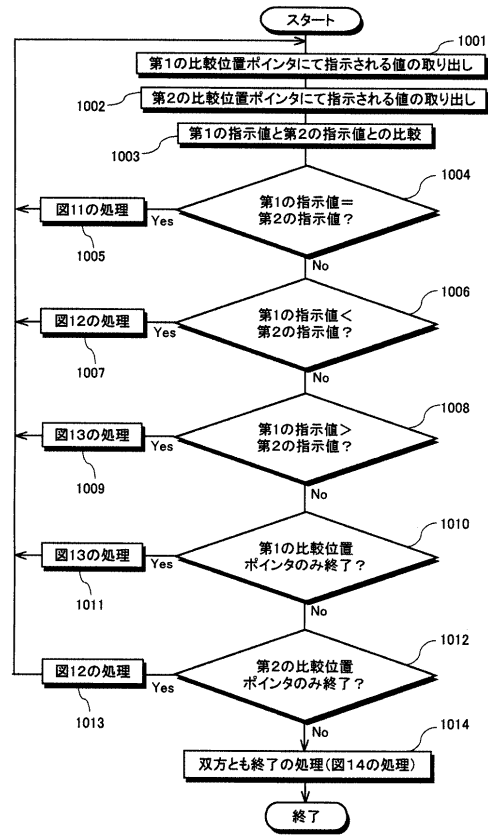


図9

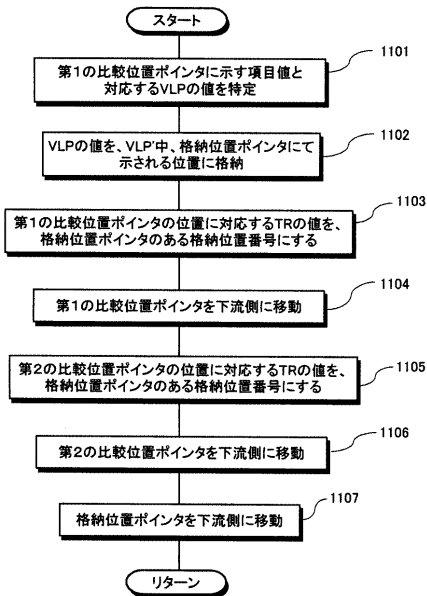
【図10】

図10



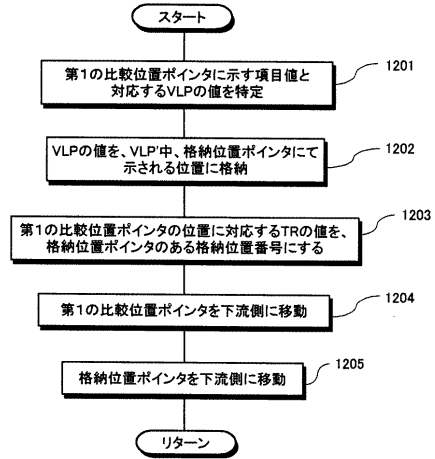
【図11】

図11



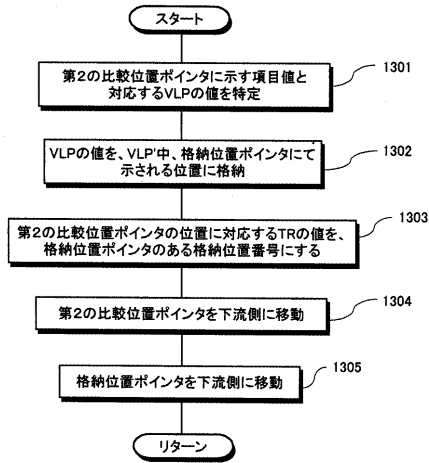
【図12】

図12



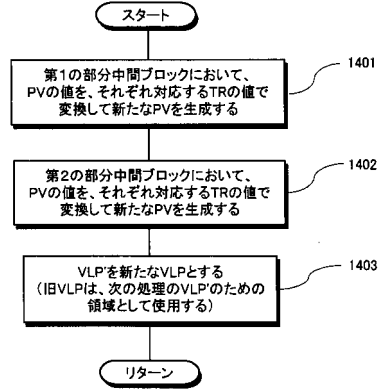
【図13】

図13



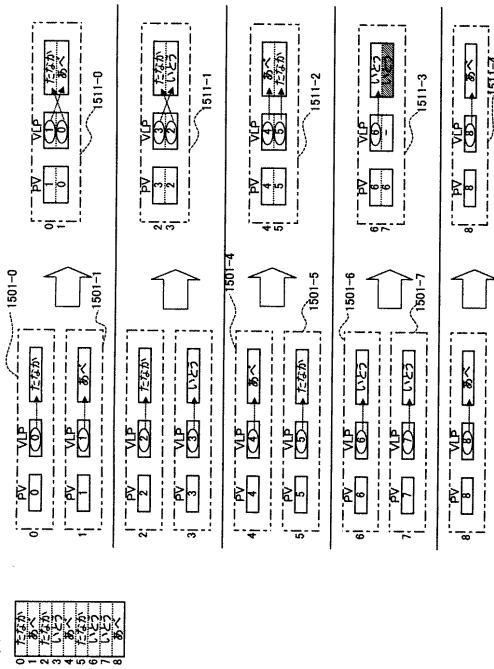
【図14】

図14



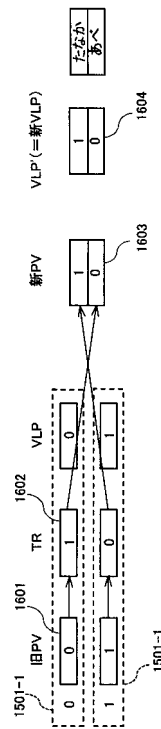
【図15】

図15

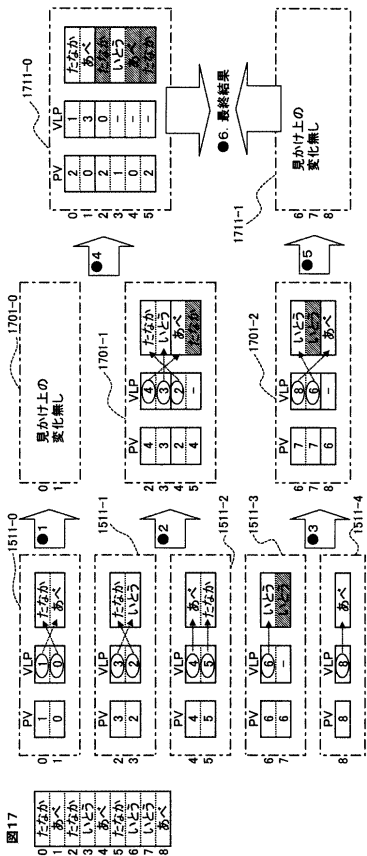


【図16】

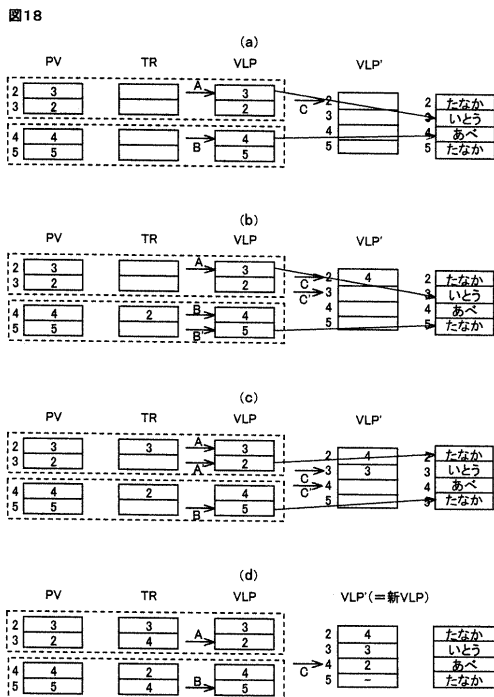
図16



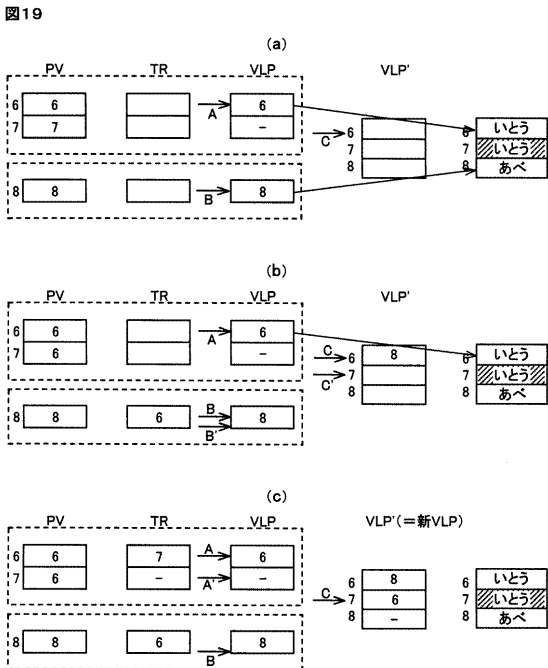
【図17】



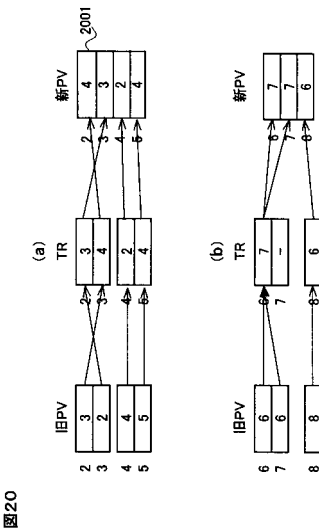
【図18】



【図19】



【図20】



【図 2 1】

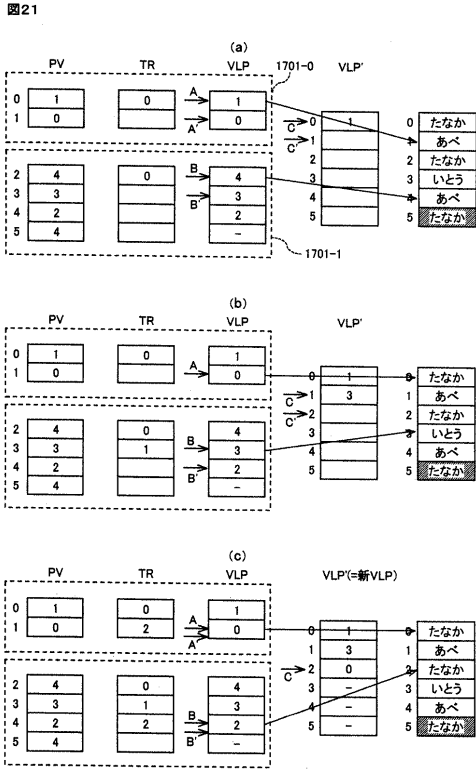
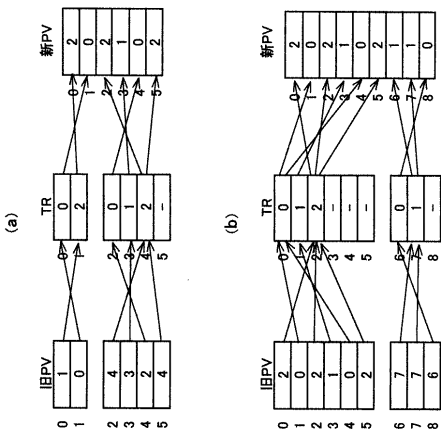
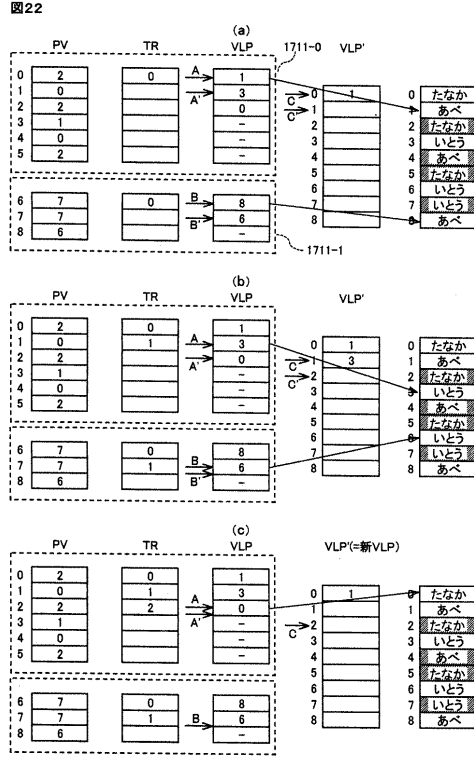


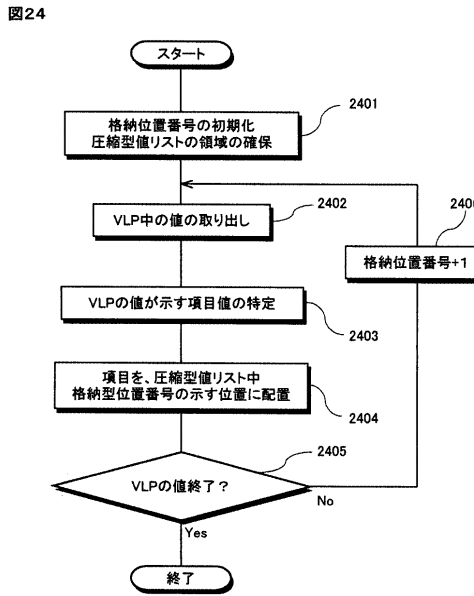
図23



【図 2 2】

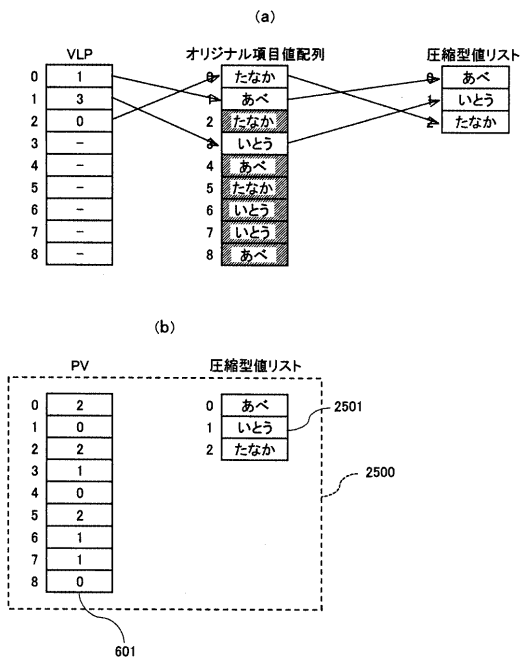


【図 2 4】



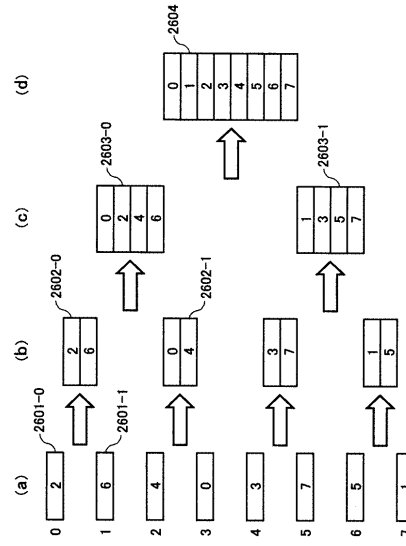
【図 25】

図25



【図 26】

図26



---

フロントページの続き

(56)参考文献 国際公開第00/010103(WO, A1)

特開平05-128164(JP, A)

特開平05-080977(JP, A)

特開昭63-298626(JP, A)

特開昭60-247732(JP, A)

鷓木 昌行, Sybase IQ「独自データ構造による、データウェアハウスへのアプローチ」, 電子情報通信学会技術研究報告 Vol. 97 No. 415, 日本, 社団法人電子情報通信学会, 1997年12月 2日, 第97巻 第415号, 51~56

(58)調査した分野(Int.Cl., DB名)

G06F 17/30

G06F 12/00