

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2008-15810

(P2008-15810A)

(43) 公開日 平成20年1月24日(2008.1.24)

(51) Int. Cl.	F I	テーマコード (参考)
<b>G06F 12/00 (2006.01)</b>	G06F 12/00 520A	5B075
<b>G06F 17/30 (2006.01)</b>	G06F 17/30 180D	5B082

審査請求 未請求 請求項の数 2 O L (全 9 頁)

(21) 出願番号	特願2006-186648 (P2006-186648)	(71) 出願人	000233055 日立ソフトウェアエンジニアリング株式会社 東京都品川区東品川四丁目12番7号
(22) 出願日	平成18年7月6日(2006.7.6)	(74) 代理人	100096954 弁理士 矢島 保夫
		(72) 発明者	山崎 修治 東京都品川区東品川4丁目12番7号 日立ソフトウェアエンジニアリング株式会社内
		(72) 発明者	笠尾 英明 東京都品川区東品川4丁目12番7号 日立ソフトウェアエンジニアリング株式会社内
		Fターム(参考)	5B075 QT06 5B082 BA03 EA05 GA15

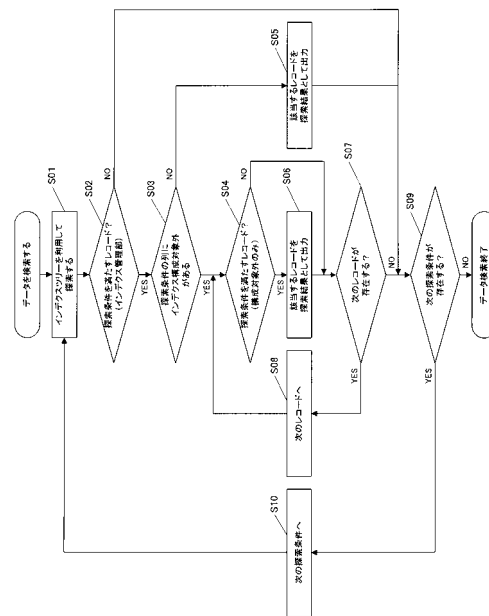
(54) 【発明の名称】 複数列の表データにおけるインデクス分割管理方法

(57) 【要約】

【課題】 複数のフィールドで構成された表データにおいて、それらのフィールドの更新が多い場合であっても、検索効率及びデータ格納効率を維持することができるインデクス管理の技術を提供することを目的とする。

【解決手段】 コンピュータ上で複数のフィールド(列)を備える表データを管理する際に、管理対象の表データに対し、更新頻度の高い列についてはインデクス作成の対象から除外し、それ以外の更新頻度の低い列についてはインデクス作成対象として、インデクスを作成する。検索の際には、指示された探索条件が更新頻度の低い列についての探索条件を含むかどうかチェックし、含む場合は、そのインデクスを利用した探索を行なうようにする。インデクスを利用した検索で高速にレコードを絞り込み、その絞り込んだレコードを対象として、インデクス構成から除外した更新頻度の高い列に関する探索条件で順検索により探索を行なう。

【選択図】 図6



**【特許請求の範囲】****【請求項 1】**

コンピュータ上で複数のフィールド（列）を備える表データを管理するためのインデクス分割管理方法であって、

コンピュータが備えるインデクス作成手段が、前記管理対象の表データに対し、所定の基準値以上に更新頻度の高いフィールドについてはインデクス作成の対象から除外し、それ以外の所定の基準値未満に更新頻度の低いフィールドについてはインデクス作成対象として、インデクスを作成するステップと、

コンピュータが備える検索手段が、指示された探索条件が前記所定の基準値未満に更新頻度の低いフィールドについての探索条件を含む場合は、前記インデクスを利用した探索を行なうステップと

10

を備えることを特徴とする複数列の表データにおけるインデクス分割管理方法。

**【請求項 2】**

請求項 1 に記載の複数列の表データにおけるインデクス分割管理方法において、

コンピュータが備える更新手段が、

指示された更新条件が前記所定の基準値未満に更新頻度の低いフィールドについての更新条件を含む場合は、前記フィールドの更新を行なった後に前記インデクスを再作成するステップと、

前記指示された更新条件が前記所定の基準値以上に更新頻度の高いフィールドについての更新条件のみを含む場合は、前記フィールドの更新のみを行なうステップと、

20

を備えることを特徴とする複数列の表データにおけるインデクス分割管理方法。

**【発明の詳細な説明】****【技術分野】****【0001】**

本発明は、複数の列で構成された表のデータを、該表に格納されているデータに対して定義されたインデクスで、データの更新頻度別に分割管理する技術に関するものである。

**【背景技術】****【0002】**

従来データベース管理システム（DataBase Management System: DBMS）では、データを表に登録して管理する。その表は、レコード（行）とフィールド（列）で構成される。指定されたデータを検索する場合、レコードを格納した順に検索を行うため、登録したレコード件数が大量の場合は容易に抽出できない。そこで、大量に登録されているデータの検索効率を向上させるため、表に登録されているデータに対して索引（インデクス）を作成する方法が利用されている。

30

**【0003】**

なお、本発明に関連する公知技術文献としては下記の特許文献 1 がある。該文献には、CSVファイル等の被検索ファイルが大規模なものとなった場合でも効率的かつ高速にデータ抽出を行うためにインデクスファイルを用いて検索する技術が開示されている。

**【特許文献 1】特開 2003 - 162545****【発明の開示】**

40

**【発明が解決しようとする課題】****【0004】**

インデクスを作成するにあたり、上記特許文献 1 によると、表に登録されているデータは予め整列し、各フィールドごとに対してインデクスを作成する必要がある。しかし、表に登録されているレコード及びデータは絶えず更新されるものであるため、インデクス管理情報も頻繁に変更されることとなり、そのためにインデクス情報を使った検索とそのデータの格納効率が下がり、有用な効果が得られない。

**【0005】**

そこで、B-Tree（Binary-Tree）によるインデクス探索手法が考えられている。この手法を用いることで、表のレコードの格納順序に関係なく、整列されたデータとして探索す

50

ることができる。B-Tree ( Binary-Tree ) によるインデクス探索手法で用いる木構造のインデクスのことをインデクスツリーと呼ぶ。インデクスツリーは、複数のフィールドを対象にして構成することができ、これによりデータの検索効率を上げることができる。

【 0 0 0 6 】

ところが、前述したように、表に登録されているレコードは絶えず更新されているため、それに伴いインデクスツリーも絶えず更新されることになる。このため、インデクスツリー構造が乱れ、指定したデータを検索するときの効率が低下してしまう。さらに、データの格納領域の断片化が発生するため、データの格納領域の再構成が必要となる頻度が高くなる。データの格納領域を再構成する場合は、業務を一度中断することになるため、業務の中断期間が長くなってしまふ。

10

【 0 0 0 7 】

本発明の目的は、複数のフィールドで構成された表データにおいて、それらのフィールドの更新が多い場合であっても、検索効率及びデータ格納効率を維持することができるインデクス管理の技術を提供することにある。

【 課題を解決するための手段 】

【 0 0 0 8 】

上記の目的を達成するため、本発明のインデクス分割管理方式は、コンピュータ上で複数のフィールド ( 列 ) を備える表データを管理する際に、管理対象の表データに対し、所定の基準値以上に更新頻度の高いフィールドについてはインデクス作成の対象から除外し、それ以外の所定の基準値未満に更新頻度の低いフィールドについてはインデクス作成対象として、インデクスを作成することを特徴とする。そして、検索の際には、指示された探索条件が前記所定の基準値未満に更新頻度の低いフィールドについての探索条件を含むかどうかチェックし、含む場合は、そのインデクスを利用した探索を行なうようにする。具体的には、指示された探索条件が所定の基準値未満に更新頻度の低いフィールドについての探索条件のみからなる場合は、インデクスを利用した探索で、指示された探索条件を満たすレコードが探索できる。また、指示された探索条件が所定の基準値未満に更新頻度の低いフィールドについての探索条件と所定の基準値以上に更新頻度の高いフィールドについての探索条件の両方を含む場合は、まずインデクスを利用した探索で更新頻度の低いフィールドについての探索条件を満たすレコードに絞り込み、絞り込んだレコードを順検索して更新頻度の高いフィールドについての探索条件を満たすレコードを探索する。

20

30

【 0 0 0 9 】

さらに、データの更新時には、指示された更新条件が前記所定の基準値未満に更新頻度の低いフィールドについての更新条件を含む場合は、前記フィールドの更新を行なった後に前記インデクスを再作成し、指示された更新条件が前記所定の基準値以上に更新頻度の高いフィールドについての更新条件のみを含む場合は、前記フィールドの更新のみを行なうようにする。

【 0 0 1 0 】

上記インデクスは、例えば、インデクスの作成対象のフィールドをキーとする木構造をもつようなものである。インデクスを利用した探索は、例えば木構造のB-Treeサーチ ( Binary-Tree Search ) の方式などを用いる。なお、表データの更新によりインデクス格納容量が不足する場合があるが、その場合は、自動的に更新フィールドをインデクス作成の対象から除外するようにしてもよい。

40

【 発明の効果 】

【 0 0 1 1 】

本発明の複数列に対するインデクスにおけるインデクス分割管理方式によれば、次のような効果がある。

( 1 ) データの更新頻度の高いフィールドをインデクス構成対象から除外しているため、当該フィールドを更新した場合はインデクスに変更がないため、インデクスのデータ格納効率を維持することができる。また、検索時には、インデクスを利用した高速な検索が行えるので、データの検索効率の維持・向上を図ることができる。

50

(2) データの更新頻度の高いフィールドをインデクス構成対象から除外しているため、データの更新頻度の高いフィールドのデータ格納領域が不足し、別データ格納領域に分割格納する場合でもインデクスに変更がないため、データの検索効率の維持・向上、及びインデクスのデータ格納効率を維持することができる。

【発明を実施するための最良の形態】

【0012】

以下、本発明を適用した複数列の表データに対するインデクスにおけるインデクス分割管理方式の一実施の形態について説明する。

【0013】

図1は、本発明の実施の形態であるデータベース管理システムの構成例を示すブロック図である。データベース管理システム装置D1は、データベースアクセス部D3及びデータ管理部D4を備える。アプリケーションD2は、入力情報として表名や探索条件などを指定して、データベースアクセス部D3に処理を依頼する。データベースアクセス部D3は、アプリケーションD2で指定された入力情報を受け取るデータ入力部D31、データ管理部D4の情報を抽出・更新するデータ操作部D32、及び、結果情報をアプリケーションD2に受け渡すデータ出力部D33を備える。データ管理部D4は、様々なデータを管理する外部記憶装置で、各データは表D41に格納されて管理され、表D41のデータを管理するためにインデクスD42が作成されている。データ管理部D4は、複数の外部記憶装置で構成することもできる。なお、図1のデータベース管理システム装置D1やアプリケーションD2は、具体的にはコンピュータ上で所定のソフトウェアを実行することにより実現されている。

【0014】

図2は、図1のシステムで管理する表の一例を示すデータ構成図である。表T1は、フィールド1(T11)とフィールド2(T12)とフィールド3(T13)の3つのフィールドで構成されており、データは登録した順に格納されている。フィールド1(T11)にはひらがなの文字が格納され、フィールド2(T12)には英字が格納され、フィールド3(T13)には数字が格納されるものとする。また、表T1に格納されているデータの更新頻度は、フィールド1(T11)とフィールド2(T12)は低く、フィールド3(T13)は高いとする。

【0015】

本実施形態では、表T1の検索効率を向上させるため、表T1に対して、図3のようなインデクスI1を作成する。インデクスI1は、表T1に格納されているデータをもとにインデクスツリーI11を作成し、各データの位置を特定できるインデクス管理部I12を、データの更新頻度の低いフィールド1(T11)とフィールド2(T12)の2つのフィールドに対して作成したものである。データの更新頻度の高いフィールド3(T13)については、インデクス管理部I12の管理対象に含めない構成にする。従って、フィールド3の検索を行なうときには、インデクスを利用することなく、表の各レコードのフィールド3のデータそのものを順番に参照しながら検索する順検索の手法を採る必要がある。

【0016】

表T1にインデクスI1が定義されている状態で、表T1のレコードT14(図2)のフィールド3(T13)の値“5”を“11”に更新したとする。このとき、インデクスI1は、図4のようなインデクスI2の状態となる。図4から分るように、フィールド3(T13)はインデクスツリー構成対象から除外してあるため、インデクス管理部I12に対する変更はなく、インデクス管理部I11とI12は同じである。また、インデクスキー情報I13の値がI23に更新されるのみで別領域を使用しないため、インデクス管理部におけるデータ格納効率を維持することができる。また、データ更新前のインデクスI1とデータ更新後のインデクスI2は全く変更がないため、今までの検索効率を維持することができる。

【0017】

10

20

30

40

50

図5は、本実施形態におけるインデクス管理方法での、インデクス作成時のフローチャート図である。表に対してインデクスを作成する場合、まずインデクスツリーの構成対象としようとするフィールドの更新頻度が高いかどうかを判断する(ステップC01)。データの更新頻度の高い列である場合は、インデクスツリー構成対象に含めずに終了する。データの更新頻度の低い列である場合は、その列に格納されているデータを昇順にしてインデクス管理部としてキー情報を登録し(ステップC02)、インデクスツリーを構成する(ステップC03)。

#### 【0018】

図6は、本実施形態におけるインデクス管理方法での、データ検索時のフローチャート図である。データ検索時に与えられる命令は、例えば、図2の表に対して「{(列1が『う』)かつ(列2が『B』)かつ(列3が『8』)}または{(列1が『お』)かつ(列2が『C』)かつ(列3が『7』)}であるレコードを探索せよ」というようなものである。図6の処理に入る前の事前の処理により、与えられた命令を整理して、何れかの1つのフィールド(列)についての探索条件(例えば「列1が『う』」)を「かつ」で接続したものを単位として、幾つかの探索条件に区分けしておく。区分けした各探索条件ごとにステップS01~S09の処理を行なう。

10

#### 【0019】

まず初めの処理対象の探索条件が、データの更新頻度の低いフィールドについての探索条件を含むものである場合、その更新頻度の低いフィールドについての探索条件で、インデクスツリーを利用して、探索を行う(ステップS01)。インデクスツリーは更新頻度の低いフィールドについて作成されており、これを利用することにより高速に探索を行なうことができる。そして、インデクス管理部でその探索条件を満たしているレコードがあるかどうかを判断し(ステップS02)、無ければ上記探索条件を満たすレコードが無いということであるから、ステップS09に進んで、次の探索条件(上記事前の処理により区分けしたもの)があるか判断する。あればその探索条件を処理対象として(ステップS10)、ステップS01から繰り返す。

20

#### 【0020】

ステップS02で上記更新頻度の低いフィールドについての探索条件を満たすレコードがあった場合は、更新頻度の低いフィールドについての探索条件でレコードが絞り込まれたということであるから、当該絞り込まれたレコードを探索対象として、ステップS03以降の処理を行なう。まず、処理対象の探索条件が、インデクス構成対象外のフィールド(すなわち、データの更新頻度の高いフィールド)についての探索条件を含むか判断する(ステップS03)。そのような探索条件を含む場合は、上記更新頻度の低いフィールドについての探索条件で絞り込んだレコードを探索対象として、上記更新頻度の高いフィールドについての探索条件を満たすレコードを順検索し(ステップS04、ステップS06~S08)、探索条件と合致するレコードを全て探索結果として出力した段階でステップS09に進む。ステップS03でインデクス構成対象外のフィールドについての探索条件を含まない場合は、ステップS01、S02の探索結果が、処理対象の探索条件を満たすレコードであるから、それらのレコードを探索結果として出力し(ステップS05)、ステップS09に進む。

30

40

#### 【0021】

なお、処理対象の探索条件が更新頻度の高いフィールドについての探索条件のみから構成される場合は、ステップS01、S02のインデクスを利用した探索の処理はスキップして、表内の全レコードを探索対象としてステップS03以降の処理を行なうものとする。

#### 【0022】

上記のデータ検索命令「{(列1が『う』)かつ(列2が『B』)かつ(列3が『8』)}または{(列1が『お』)かつ(列2が『C』)かつ(列3が『7』)}であるレコードを探索せよ」が与えられたとして、図6の流れを説明する。図6の処理に入る前に、事前の処理により、上記命令を「(列1が『う』)かつ(列2が『B』)かつ(列3が『

50

8』)」と「(列1が『お』)かつ(列2が『C』)かつ(列3が『7』)」の2つの探索条件に区分けしておく。

【0023】

まず探索条件「(列1が『う』)かつ(列2が『B』)かつ(列3が『8』)」を処理対象として、ステップS01からの処理を行なう。この探索条件のうち、インデクスが作成されている列1と列2についての探索条件「(列1が『う』)かつ(列2が『B』)」についてはインデクスを利用して高速に検索できる。そこで、ステップS01, S02で探索条件「(列1が『う』)かつ(列2が『B』)」でインデクスを利用した探索を行なう。これにより、図2の「う B 5」と「う B 8」のレコードに絞り込まれる。上記処理対象の探索条件は、インデクスが作成されていない列3についての探索条件「列3が『8』」も含むから、ステップS03からS04に進み、ステップS04, S06~S08で上記絞り込まれた2つのレコードを探索対象として、探索条件「列3が『8』」を満たすレコードを順検索する。これにより、探索結果として「う B 8」のレコードが出力される。この結果は、探索条件「(列1が『う』)かつ(列2が『B』)かつ(列3が『8』)」の探索結果として、所定のバッファに記憶しておく。

10

【0024】

次にステップS09からS10に進み、次の探索条件「(列1が『お』)かつ(列2が『C』)かつ(列3が『7』)」についてステップS01からの処理を行なう。上記と同様にして、インデクスを利用した探索で絞り込み、順検索を行ない、探索結果として「お C 7」のレコードを得る。その後、ステップS09では次の探索条件は無いので、データ検索を終了する。なお、上記2つの探索条件は「または」で結ばれていたもので、上記探索結果を「または」で結んだ「(う B 8)または(お C 7)」が最終的な結果となる。

20

【0025】

図7は、本実施形態におけるインデクス管理方法での、データ更新時のフローチャート図である。なお、データ更新時には、更新すべきレコードを特定するための探索条件が与えられると共に、その探索条件で特定されるレコードの更新対象のフィールドを指定する情報と、そのフィールドに設定する値が与えられるものとする。

【0026】

まず、更新対象のフィールドがインデクスツリー構成対象のフィールドか判定する(ステップU01)。

30

【0027】

インデクスツリー構成対象のフィールドである場合は、与えられた探索条件に応じてインデクスツリーを利用して更新対象のレコードの探索を行ない(ステップU02)、該当するインデクス管理レコードをインデクスツリーから削除して(ステップU03)、該当するレコードの該当するフィールドのデータを更新する(ステップU04)。次に、更新後のインデクス管理レコードを挿入すべき新たな位置を探索するために、再度、インデクスツリーを利用して探索を行い(ステップU05)、該当するインデクス管理レコードをインデクスツリーに追加し、インデクスツリーを再作成する(ステップU06)。

【0028】

なお、更新すべきレコードを特定するための探索条件がインデクスツリー構成対象外のフィールドについての探索条件のみからなる場合は、ステップU02でインデクスを利用した探索の代わりに順検索で更新対象のレコードの探索を行ない、ステップU03では当該更新対象のレコードのインデクス管理レコードを削除するようにすればよい。また、更新すべきレコードを特定するための探索条件がインデクスツリー構成対象のフィールドについての探索条件とインデクスツリー構成対象外のフィールドについての探索条件の両方を含むものである場合は、ステップU02のインデクスツリーを利用した探索だけでは絞り込みが行なえるだけであるので、さらにインデクスツリー構成対象外のフィールドについての探索条件で探索を行なう必要があるが、その探索はステップU02で行なえばよい。要するにU02~U06では、更新対象のレコードの更新対象のフィールドを更新する

40

50

処理に加えて、当該レコードのインデクスの作り直しを行なっている点が特徴である。

【0029】

ステップU01でインデクストリー構成対象外のフィールドの更新である場合は、まず与えられた探索条件に応じてインデクストリーを利用して更新対象のレコードの探索を行なう(ステップU07)。なお、ステップU07は、更新すべきレコードを特定するための探索条件がインデクストリー構成対象のフィールドについての探索条件を含む場合に、その探索条件でレコードを絞り込む処理を行なうものである。もし更新すべきレコードを特定するための探索条件がインデクストリー構成対象外のフィールドについての探索条件のみからなる場合は、ステップU07はスキップする。

【0030】

次に、更新すべきレコードを特定するための探索条件がインデクストリー構成対象外のフィールドについての探索条件を含む場合は、ステップU07で絞り込んだレコードを処理対象とし、該処理対象のレコードを1つ1つ取り出しながらステップU08～U11の処理を行なう。まず、取り出したレコードが、与えられているインデクストリー構成対象外のフィールドについての探索条件を満たすか判断する(ステップU08)。ステップU08で探索条件を満たさない場合は、ステップU10に進む。ステップU08で探索条件を満たす場合は、そのレコードが更新対象のレコードであるということであるから、そのレコードの更新対象のフィールドを指定されたとおりに更新し(ステップU09)、ステップU10に進む。次のレコードが存在するか判断し(ステップU10)、存在しなければステップU12に進み、存在すれば次のレコードを処理対象として(ステップU11)

10

20

【0031】

なお、更新すべきレコードを特定するための探索条件がインデクストリー構成対象外のフィールドについての探索条件を含まない場合は、既にステップU07の探索だけで更新すべきレコードが特定されているはずであり、ステップU08～U11のような順検索を行なう必要はない。従って、その場合は、ステップU07の後、更新対象のレコードの更新対象のフィールドを指定されたとおりに更新し、ステップU12に進むようにすればよい。要するにU07～U11では、更新対象のレコードの更新対象のフィールドを更新する処理を行なうだけで、当該レコードのインデクスについては全くそのままにしている点が特徴である。これにより、インデクストリー構造が変わらず、新たなインデクス管理領域を使用しないため、データ管理領域の格納効率を維持することができる。

30

【0032】

さらに、次のデータ更新が指示されている場合は(ステップU12)、そのデータ更新を行なうべく、次の探索条件を処理対象として(ステップU13)、ステップU01からの処理を繰り返す。指示された全てのデータ更新が終了した段階でデータ更新とインデクス管理情報の更新が終了する。

【0033】

なお、上記実施の形態では、図5のステップC01において、フィールドの更新頻度が高いかどうかによりインデクスを作成するかしないかを決定しているが、この更新頻度の高低の判断は任意である。例えば、所定期間における各フィールドの更新回数を集計して

40

【0034】

以上のように、本実施の形態の複数列に対するインデクスにおけるインデクス分割管理方式によれば、次のような効果がある。

(1) データの更新頻度の高いフィールドをインデクストリー構成対象から除外しており、レコード更新を行った場合においてもインデクストリーに変更がないため、検索効率を維持・向上させることができる。

(2) データの更新頻度の高いフィールドをインデクストリー構成対象から除外している

50

ため、データの更新頻度の高いフィールドのデータ格納領域が不足し、別のデータ格納領域に分割格納した場合でもインデクストリーに変更がないため、検索効率を維持・向上させることができる。

【図面の簡単な説明】

【0035】

【図1】データベース管理システム構成の一例を示すブロック図である。

【図2】本発明の実施形態を説明する表の一例を示すデータ構成図である。

【図3】図2の表に対するインデクスを示すデータ構成図である。

【図4】本発明によるインデクス管理方法での、データの更新を行った場合の図3のインデクスを示すデータ構成図である。

10

【図5】本発明におけるインデクス管理方法での、インデクス作成時のフローチャート図である。

【図6】本発明におけるインデクス管理方法での、データ検索時のフローチャート図である。

【図7】本発明におけるインデクス管理方法での、データ更新時のフローチャート図である。

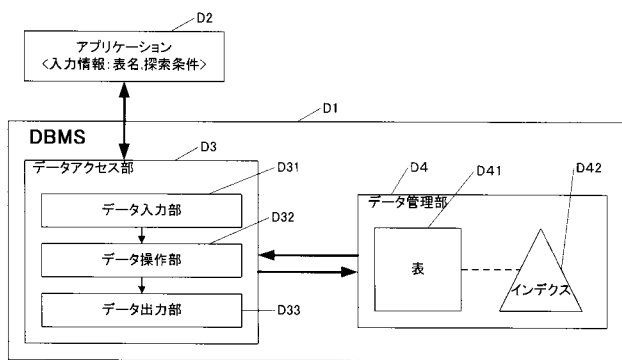
【符号の説明】

【0036】

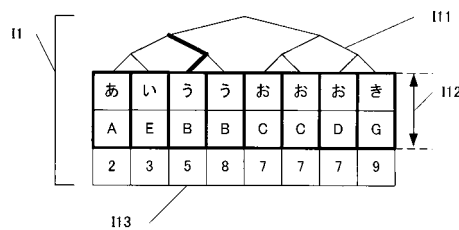
D1...データベース管理システム(DBMS: DataBase Management System)、D2...アプリケーション、D3...データアクセス部、D31...データ入力部、D32...データ操作部、D33...データ出力部、D4...データ管理部、D41...表、D42...インデクス、T1...表、I1, I2...インデクス、I11...インデクストリー、I12...インデクストリー構成対象。

20

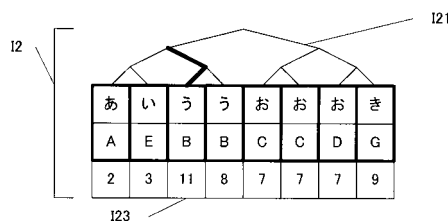
【図1】



【図3】



【図4】

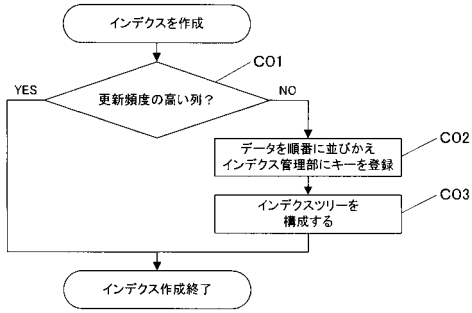


【図2】

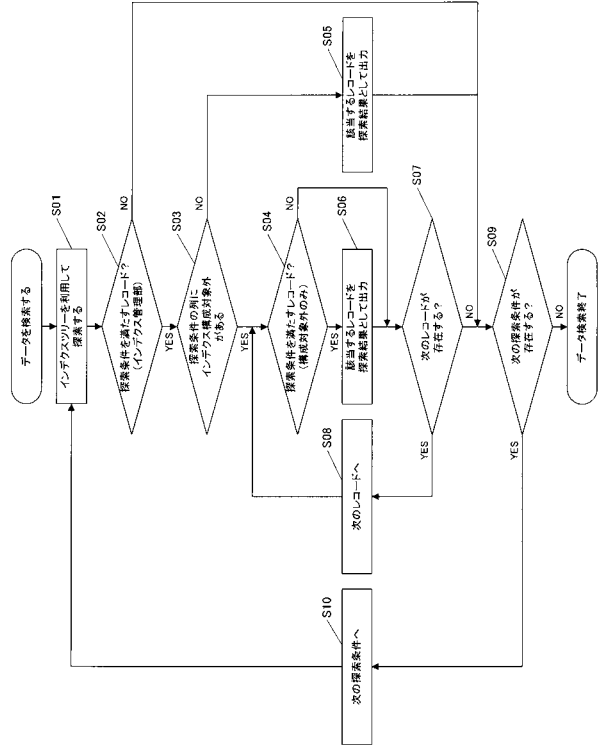
T1	列1	列2	列3
T11	あ	A	2
T12	い	E	3
T13	う	B	5
T14	お	C	7
	お	D	7
	き	G	9
	う	B	8
	お	C	7



【 図 5 】



【 図 6 】



【 図 7 】

