

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第4201831号
(P4201831)

(45) 発行日 平成20年12月24日 (2008. 12. 24)

(24) 登録日 平成20年10月17日 (2008. 10. 17)

(51) Int. Cl.

F I

G 0 6 F 17/30 (2006. 01)

G 0 6 F 17/30 3 4 O Z

請求項の数 9 (全 18 頁)

(21) 出願番号	特願平8-510815	(73) 特許権者	501359984
(86) (22) 出願日	平成7年9月21日 (1995. 9. 21)		クリクテック、インターナショナル、アク
(65) 公表番号	特表平10-505930		チボラグ
(43) 公表日	平成10年6月9日 (1998. 6. 9)		Q L I K T E C H I N T E R N A T I O
(86) 国際出願番号	PCT/SE1995/001071		N A L A B
(87) 国際公開番号	W01996/009589		スウェーデン国エス - 2 2 3 7 0、ルンド
(87) 国際公開日	平成8年3月28日 (1996. 3. 28)		、フォルスクニングスビン、イデオン (番
審査請求日	平成14年7月12日 (2002. 7. 12)		地なし)
審判番号	不服2005-24192 (P2005-24192/J1)	(74) 代理人	100064285
審判請求日	平成17年12月15日 (2005. 12. 15)		弁理士 佐藤 一雄
(31) 優先権主張番号	9403175-4	(74) 代理人	100077609
(32) 優先日	平成6年9月21日 (1994. 9. 21)		弁理士 玉真 正美
(33) 優先権主張国	スウェーデン (SE)	(74) 代理人	100088889
			弁理士 橘谷 英俊

最終頁に続く

(54) 【発明の名称】 データベースから情報を抽出するための方法及び装置

(57) 【特許請求の範囲】

【請求項 1】

データベースから情報を抽出するための方法であって、そのデータベースは複数のデータレコードを有し、それぞれのデータレコードは、少なくとも2つのデータ要素を有し、それぞれのデータ要素はデータ要素の型式とデータ要素の値とにより定義される、方法であって、

ー前記データベースの前記データレコードをコンピュータの主メモリに読み込むステップと；

ーそれぞれのデータ要素の型式のそれぞれ異なるデータ要素の値に対して、異なるバイナリ・コードを割り当てて、前記データレコードをバイナリ・コード化された状態で格納するステップと；

ー少なくともひとつのデータ要素の値がユーザにより選択されたことを検出するステップと；

ー前記ユーザによる選択の対象となるすべてのデータ要素の値を得るために、前記バイナリ・コード化されたデータレコードを検査するステップと；

ー前記ユーザによる選択の対象となるすべてのデータ要素の値を、状態セットに記録するステップと；

を備えるとともに、

前記ユーザによる選択によって除外されていないすべてのデータ要素の値が最初は状態セットに含まれており、

10

20

前記データレコードを検査するステップは、
データ要素の値がすべて前記状態セットに属しているようなすべてのデータレコードを見つけるステップと、

前記データレコードを見つけるステップにおいて見つけれたいずれのデータレコードにも現れないデータ要素の値を前記状態セットから除外するステップと、
を備え、

前記データベースの前記データレコードは2またはそれ以上の表で構成されており、前記データレコードを検査するステップでは、前記表の前記データレコードは、前記状態セットの要素が変化しなくなるまで繰り返して検査されるとともに、

前の選択の対象になったデータ要素の値のうちの少なくともいずれかがユーザにより新たに選択されたことをさらに検出するたびに、前記バイナリ・コード化されたデータレコードを検査するステップと、前記選択の対象となるすべてのデータ要素の値を状態セットに記録するステップとを繰り返すことを特徴とする方法。

【請求項2】

少なくとも前記選択の対象となる前記データ要素の値を人間が読みとれる型式でユーザに提示するステップをさらに備えたことを特徴とする請求項1記載の方法。

【請求項3】

前記状態セットは、それぞれのデータ要素の型式についての状態ベクトルを備え、
それぞれの状態ベクトルは、対応するデータ要素の型式のそれぞれのデータ要素の値について1つの要素を有し、

前記データレコードを検査するステップは、前記選択の対象となるデータ要素の値に対応する前記状態ベクトルのそれぞれの要素を第1の値に設定し、前記選択の対象となるデータ要素の値に含まれていないデータ要素の値に対応する状態ベクトルのそれぞれの要素を第2の値に設定することを特徴とする請求項1または2に記載の方法。

【請求項4】

それぞれのデータ要素の型式について頻度ベクトルを供給するステップをさらに備え、
それぞれの頻度ベクトルは、対応するデータ要素の型式のそれぞれのデータ要素の値について1つの要素を有しており、

前記データレコードを検査するステップは、すべてのデータ要素の値が前記状態セットに属しているようなデータレコードのデータ要素の値に対応する頻度ベクトルの要素を、1だけ増加させるステップを備えることを特徴とする請求項1～3のいずれかに記載の方法。

【請求項5】

すべてのデータレコードの検査が終了した時にゼロの頻度カウントを有する頻度ベクトルの要素に対応するそれぞれのデータ要素の値を、前記状態セットから除外するステップをさらに備えたことを特徴とする請求項4記載の方法。

【請求項6】

第1のデータ要素の型式のデータ要素の値が論理ANDによって関係づけられていることを検出するステップを備え、

前記少なくともひとつのデータ要素の値がユーザにより選択されたことを検出する前記ステップは、前記第1のデータ要素の型式のn個のデータ要素の値が選択されたことを検出し、

前記データレコードを検査するステップは、頻度ベクトル要素の値としてnを有する第2のデータ要素の型式のデータ要素の値を得るステップを備えたことを特徴とする請求項4または5に記載の方法。

【請求項7】

得られた前記第2のデータ要素の型式のデータ要素の値を、前記第1のデータ要素の型式のn個のデータ要素の値の前記選択によって除外される前記第1のデータ要素の型式のデータ要素の値を得るための選択として利用することを特徴とする請求項6記載の方法。

【請求項8】

前記コンピュータの主メモリに読み込まれた前記複数のデータレコードは複数のリストの形態によって前記コンピュータのディスプレイ上に提示され、
それぞれのリストは異なるデータ要素の型式のデータ要素の値を有し、
それぞれのデータ要素の値はそれぞれのリストに一度だけ現れ、
前記選択の対象となるデータ要素の値は第1の方法によって示され、
前記選択の対象とならないデータ要素の値は第2の方法によって示されることを特徴とする請求項1～7のいずれかに記載の方法。

【請求項9】

データベースから情報を抽出するための装置であって、そのデータベースは複数のデータレコードを有し、それぞれのデータレコードは、少なくとも2つのデータ要素を有し、それぞれのデータ要素はデータ要素の型式とデータ要素の値とにより定義される、装置であって、

－前記データベースの前記データレコードをコンピュータの主メモリに読み込む手段と；
－それぞれのデータ要素の型式についてそれぞれ異なるデータ要素の値に対して異なるバイナリ・コードを割り当てて、前記データレコードをバイナリ・コード化された状態で格納する手段と；

－少なくともひとつのデータ要素の値がユーザにより選択されたことを検出する手段と；
－前記ユーザによる選択の対象となるすべてのデータ要素の値を得るために、前記バイナリ・コード化されたデータレコードを検査する手段と；

－前記ユーザによる選択の対象となるすべてのデータ要素の値を、状態セットに記録する手段と；

を備えるとともに、

前記ユーザによる選択によって除外されていないすべてのデータ要素の値が最初は状態セットに含まれており、

前記データレコードを検査する手段は、

データ要素の値がすべて前記状態セットに属しているようなすべてのデータレコードを見つける手段と、

前記見つける手段において見つけられたいずれのデータレコードにも現れないデータ要素の値を前記状態セットから除外する手段と、

を備え、

前記データベースの前記データレコードは2またはそれ以上の表で構成されており、前記データレコードを検査する手段では、前記表の前記データレコードは、前記状態セットの要素が変化しなくなるまで繰り返して検査されるとともに、

前の選択の対象になったデータ要素の値のうちの少なくともいずれかがユーザにより新たに選択されたことをさらに検出するたびに、前記データレコードを検査する手段による、前記バイナリ・コード化されたデータレコードの検査と、前記データ要素の値を前記状態セットに記録する手段による、その選択の対象となるすべてのデータ要素の値の記録とを繰り返し行う手段をさらに備えることを特徴とする装置。

【発明の詳細な説明】

技術分野

本発明は、複数のデータレコード（data records）を備えたデータベースから情報を抽出するための方法及び装置に関し、ここで、それぞれのデータレコードは少なくとも2つのデータ要素を備え、それぞれのデータ要素は、データ要素の形式と、データ要素の値とにより定義される。

本発明の背景

データベースに関して抽出するするための周知な方法は、SQL（Structured Query Language）を用いるものである。データベースから情報を抽出するために、そのデータベースに対して、選択命令文（SELECT statements）が質問される。基本的には、その選択命令文に対する回答は、表（table）を構成しているデータレコードの数である。

SQLのひとつの欠点は、SQLにおいてはデータベースの表構造について知っていることが前提とされていることである。より具体的には、ユーザは、情報を表にしている選択命令文において、検索される形式で命令をすることが必要とされる。例えば、データベースが衣服に関するものであり、情報が2つの表に格納されている場合を想定すると、それらの表のうちの第1の表は衣服の形やそのサイズを格納し、第2の表は衣服の形と価格とを格納することができる。所定のサイズと所定の価格とを有するすべての形の衣服を抽出するためには、2つの表の名前と、衣服の形、価格及びサイズを含むテーブルのフィールドの名前と、所定の価格と、所定サイズとを特定した複雑な選択命令文を用いなければならなかった。

さらに、検索されるべき情報が、データ要素に索引の付けられていない(not indexed)カラム(column)にある場合には、そのカラムはリニア(linearly)に検索しなければならず、そのカラムがたくさんのデータ要素値を有する場合には、応答時間が長くなる。これは、特に、そのデータベースが商用のものであり、ユーザがそのデータベースに接続している時間に対して料金を支払うような場合に、深刻な欠点となる。

本発明の要約

従って、本発明の目的は、前述したような欠点を軽減し、データ・ベースから情報を抽出する新規な方法であって、ユーザがそのデータ・ベースの表構造に関してなんら前もって知識を有することが必要とされずにデータ・ベースから迅速に情報を得ることができる方法を提供することである。

この目的は、請求項1と請求項13とに列挙されている特徴を有する方法及び装置によりそれぞれ達成される。それらの望ましい実施の形態は、従属項に列挙されている。

本発明によれば、データをオフ・ライン(off-line)で処理するために、データ・ベースのデータレコードは、コンピュータの主メモリに読み込まれる。データ・ベースは2次メモリに格納されるか、またはコンピュータがモデムによって接続されているデータ・ベースに遠隔的に格納されても良い。例えば、このように主メモリに読み込まれるデータ・ベースは、もっと大きいデータ・ベース或いは2またはそれ以上の複数のデータ・ベースの組み合わせ、のうちの選択された一部でもよい。

さらに、それぞれのデータ要素の型式のそれぞれの異なるデータ要素の値にはバイナリ・コードが割り当てられ、データレコードはバイナリ・コード化された形式で格納される。バイナリ・コード化することにより、きわめて迅速なサーチを表について実行することができる。さらに、冗長な情報を除去することが可能となり、その結果としてデータ量を減少することができる。

データ・ベースの中の少なくとも一つのデータ要素が選択されたことが認識された時、すなわち、例えばユーザがそのデータ要素を「クリック」した時には、その選択の対象とするすべてのデータ要素の値を見つけるためにバイナリ・コード化されたデータレコードが検査され、そして、これらのデータ要素の値は状態セット(status set)に記録される。すなわち、本発明による方法では、1またはそれ以上のデータ要素の値の選択の対象がどのデータレコードであるかではなく、その選択の対象はどのデータ要素の値であるかが記録され、それによって選択の結果がきわめて明瞭かつコンパクトな方法でユーザに提示される。

この方法の結果として得られたデータ要素の値はユーザに直接、提示してもよく、または提示する前にさらに処理されても良く、または例えば別のプログラムに対する入力として利用するように内部的に利用しても良い。

結果の提示は、その選択の対象とされるすべてのデータ要素の値のリストの形式をとることもできるし、または、選択の対象とされる値が第1の方法で示され、その選択により除外された値は第2の方法或いはその他の適当な形式により示されるようにした、すべてのデータ要素の値のリストとすることもできる。

データ・ベースからデータを抽出するための本発明の方法は、きわめて速い。応答時間は、データ・ベースのデータの量に線形に依存する。データ要素の選択の対象とされるデー

10

20

30

40

50

タレコードのすべてのデータ要素の値が示されるので、ユーザは次になしうるあらゆる選択を容易に把握することができ、このことによってデータ・ベースからの情報について継続的にサーチすることが容易になる。

【図面の簡単な説明】

本発明の実施の形態は、補遺 A の表および図面を参照しつつ、以下に例示によって説明され、ここで、

図 1 は、本発明による方法の一般的な概略を提供するダイアグラム図であり、

図 2 は、本発明による方法の O R モードにおける表の検索を例示したダイアグラム図であり、

図 3 は、本発明による方法の A N D モードにおける表の検索を例示したダイアグラム図である。

10

発明の実施の形態

図 1 及び図 2、及び、本発明の特定の実施例に関する補遺 A の表 1 - 3 9 を参照しつつ、以下に、本発明の基本的なモードと O R モードとについて説明する。表 1 及び表 2 は、様々な衣服に関する情報を含んだデータベースの一例を表すものである。表 1 は、データ要素の型式として、「衣服」、「色」、「サイズ」及び「科目番号 (A r t N o)」を含んでいる。表 2 は、データ要素の型式として、「科目番号」と「価格」とを含んでいる。これらの表から分かるように、「衣服」の型式で定義されているデータ要素の型式には、データの値「シャツ」と、データの値「ジーンズ」とがある。同様に、「色」の型式のデータ要素には、データの値として「青」と「赤」とがあり、また、「サイズ」の型式のデータ要素にはデータの値として「3 2」、「3 5」、「3 9」或いは「4 0」があり、また、「科目番号」の型式のデータ要素にはデータの値として「1 0 0 1」、「1 0 0 3」或いは「2 1 0 0」があり、また、「価格」の型式のデータ要素にはデータの値として「1 8 0」と「2 2 0」とがある。

20

通常は、これらのデータ要素の値は、アスキー・コードによる文字列 (A S C I I - c o d e d s t r i n g s) の型式で格納されている。

本発明による方法は、コンピュータ・プログラムにより実行される。本発明による方法の第 1 のステップ (図 1 の枠 1 0 1 を参照のこと。) においては、例えば、データベースのすべての表、すなわちこの場合には表 1 及び表 2 を選択するような選択命令文を用いることによって、プログラムは、データベースのすべてのデータレコードを読み込む。

30

次に、入力されたそれぞれの表について、次のステップ (枠 1 0 2 を参照のこと。) が実行される。第 1 に、その表のカラム名、すなわちデータ要素の型式が連続的に読み込まれる。新しいデータ要素の型式が現れる度に、データ構造は、そのデータのために例示 (i n s t a n t i a t e d) される。次に、すべてのデータレコードをバイナリ型式で含むために、内部表構造 (i n t e r n a l t a b l e s t r u c t u r e) が例示され、これにより、データレコードは、連続的に読み出され、バイナリ・コード化される。読み出されたそれぞれのデータ要素の値について、対応するデータ要素の型式のデータ構造がチェックされ、その値が前に割り当てられたバイナリ・コードであるか否が確認される。そうである場合には、バイナリ・コードは、上述した表構造の中の適当な場所に挿入される。そうでない場合には、データ構造に挿入される前に、新しいバイナリ・コード、望ましくは、昇順で次のコードが割り当てられ、そのデータ要素の値は、データ構造に追加される。換言すれば、それぞれのデータ要素の型式に関して、異なるデータ要素の値ごとに異なるバイナリ・コードが割り当てられる。

40

表 3 - 7 は、表 1 及び表 2 を構成しているデータベースの中の異なるデータ要素の値に対して割り当てられたバイナリ・コードを表す。表 8 及び表 9 (枠 1 0 4 を参照のこと。) は、表 1 及び表 2 のすべてのデータレコードをそれぞれバイナリ型式で格納している。すなわち、表 8 及び表 9 は、図 1 の表 1 及び表 2 を非常に圧縮したものとされており、表中のデータを極めて速くリニア検索することができる。読み出しを容易にするために、表 3 - 9 のバイナリ・コードは、それぞれ 1 0 進数によって表現されている。

入力された表のすべてのデータレコードが読み込まれて、バイナリ・コード化されたら、

50

前述した内部表構造 (`internal table structure`) のバイナリレコードがソート (`sort`) され、重複したデータレコードはソートされた表から取り除かれる。

前述したような、バイナリ化されソートされた表を作るステップは、適当なアルゴリズムによって、 $N^2 \log N$ 回行うことができ、ここで、 N は表の中のデータの数である。

本発明による方法の次のステップにおいては、それぞれのデータ要素の型式に対して、異なるデータ要素の値のそれぞれについてひとつの位置 (`position`) を有する頻度ベクトルが生成される。さらに、それぞれのデータ要素の型式について、それぞれ異なるデータ要素の値についてひとつの位置を有する2つのブール文字列 (`Boolean strings`) が生成される。それぞれのデータ要素の値に対する第1のブール文字列は、ユーザがした選択を表すために用いられる。以下の説明では、この第1のブール文字列は、ユーザ選択ベクトル (`user selection vector`) と称する。第2のブール文字列は、データ要素の値の状態を表すために用いられる。以下の説明では、この第2のブール文字列は、状態ベクトル (`status vector`) と称する。

以上の説明に関連して、状態ベクトルは、厳密な意味では、ベクトルでなくとも良いことも説明する必要がある。すなわち、それは、特定の状態を有するようすべてのデータ要素の値のリスト、例えば、ユーザの選択によって含まれまたは除外されるようなもの、または、データ要素の値の状態を記録するための他のいかなる適当な状態のセット、であっても良い。しかし、状態ベクトルを前述のように利用することによって、状態のセットを簡単に表現することができる。

上述したステップは、一度だけ実行される。そして、このステップが完了すると、バイナリ化された表において、繰り返し検索が実行されることもある。

本発明による方法の次のステップ (枠103を参照のこと。) では、データベースの内容が、コンピュータのディスプレイ上でユーザに示され、この時に、望ましくは、表10-14のような型式がとられ、ここで、それぞれのデータ要素の型式についてのデータ要素の値は、それぞれ別の縦書きのリストに示されている。それぞれのリストは、そのデータ要素の型式についての見出しを有する。互いに異なるそれぞれの値は、このリストにおいては、一度しか現れない。もちろん、この表示モードは、本発明による方法以外の方法について用いることもできる。

これで、ユーザがディスプレイ上のリストの中の1またはそれ以上のデータ要素の値を選択することにより、データベースの中の情報を調べることができるようになる。その選択は、例えば、選択する1またはそれ以上のデータ要素の値をクリックするか触ることにより、することができる。または、そのような選択は、キーボードによって入力することもできる。

ユーザにより選択がされたことが検知されると、次のステップが実行される (枠105-108を参照のこと)。第1に、データ要素の値が選択されたリスト (またはデータ要素の型式) に対応するユーザ選択ベクトルの中の、このデータ要素の値に対応する位置に、1 (または、真 (`TRUE`)) が挿入される。このユーザ選択ベクトルのその他のすべての位置と、その他のユーザ選択ベクトルのその他のすべての位置には、0 (または、偽 (`FALSE`)) が挿入される。これらのステップの結果は、表15-19に示されている。

第2に、データ要素の値が選択されたリストに対応するユーザ選択ベクトルが、対応する状態ベクトル (`status vector`) にコピーされる。いずれのデータ要素の値も選択されなかったリストに対応する、残りの状態ベクトルは、そのリストのすべての可能なデータ要素の値が選択されたものとして取り扱われる。何故ならば何の選択もないということは、このデータ要素の型式に関しては何の制限もないということを意味するからである、従って、それら全ての位置に1が挿入される。

たとえば、ユーザが値「シャツ」を選択した場合を想定する。すると、表20に示したように、リスト「衣服」に関する状態ベクトルの第1の位置には1が挿入され、第2の位置には0が挿入される。その他のすべてのブール文字列は表21-24に示したように1に

10

20

30

40

50

よって満たされる。

これに関連して、ユーザ選択ベクトルは、ORモードにおける方法を実行することに関しては義務的ではないことを説明する必要がある。ユーザの選択は、独立したベクトルまたはリストに記録することが可能であり、ユーザの選択は、状態ベクトルに直接エンコードできる。しかし、ユーザ選択ベクトルは、ANDモード（枠109を参照のこと。）においては用いられる。

次のステップでは、プログラムは引き続いて、バイナリ化された表のデータレコードの検索（*traverse*）をする（図1の枠110-111及び図2を参照のこと。）。第1に頻度ベクトルがクリアされる（枠202）。それぞれのデータレコードのそれぞれの値について、対応する状態ベクトルの対応する値が1であるかをチェックする（枠203）。もしデータレコードのすべての値についてそうであるならば、プログラムは「ヒット」または「合致（*match*）」であると判断する（枠204）。そして、対応する頻度ベクトルの対応する値が1だけ増加される（枠205）。

表の中のすべてのデータレコードが調べられたら（枠206）、対応する頻度ベクトルの頻度カウントがゼロであるような状態ベクトルのすべての位置をクリアする（0に設定することにより、対応する状態ベクトルが更新（*update*）される（枠207及び208）。このようにして、頻度ベクトルは、その選択に含まれないこれらのデータ要素の値を探すために利用することができる。

データベースにもっと表がある場合には、これらの表について前述のステップが繰り返される。すべての表が検査されたら、前述したステップはすべての表についてもう一度繰り返される。この手続きは、状態ベクトルがもはや変化しなくなるまで続けられる（枠111）。通常は、それぞれの表を1回だけ、または2回検索することが必要とされる。しかし、データベースがひとつの表のみを有する場合は、この表はもちろん1回だけ検査されることが必要とされる。

表25-29は、データ要素の値として「シャツ」が選択された後に表8と9が一度だけ検索された時の、頻度ベクトルの内容を示している。状態ベクトルの要素のうちで、対応する頻度ベクトル要素の値がゼロよりも大きいものは「1」を有し、その他のすべての要素は、「0」を有する。

これらの表の検査に必要とされる時間は、これらの表に含まれているデータの量に線形に比例している。

次のステップでは、選択の結果がユーザに示される（枠112と103）。この選択に含まれるデータ要素の値、すなわちその選択に合致するデータレコードのデータ要素の値、は第1の方法、例えば第1の色、により示され、その選択から除外されたデータ要素の値は、異なる方法、例えば第2の色により示される。これらの値は状態ベクトルから得られる。さらに、選択された値、すなわちユーザ選択ベクトルから得られた値、は望ましくはさらに異なる方法により示される。表30-34は、値として「シャツ」が選択されたような前述の場合において、その結果がユーザにどう示されるかを示している。表に示されているように、サイズ32と35及び科目番号2100が除外されている、何故ならばサイズ32または35、或いは科目番号2100のシャツは存在しないからである。

ここで、ユーザがさらに採りうるデータ要素の値のうちのいずれか、例えばサイズの値「40」、を選択した場合は、頻度ベクトルはクリアされ、前述のステップが繰り返されて、どのデータレコードが選択「シャツ」と選択「40」の両方に合致するかを調べる。この結果は表35-39に示されている。ここで、ユーザがサイズ40のシャツを希望する場合は、価格が180の青のシャツとなることが明らかになる。

前述の例では、異なるデータ要素の型式の間で、想定された論理的AND操作が存在することが示されている。すなわち、（衣服＝シャツ）AND（サイズ）＝（40）AND（色）＝（青）AND（科目番号）＝（1001）AND（価格）＝（180）である。しかし、いずれかのデータ要素の値と、同じデータ要素の型式とは、通常は論理OR演算によって関連づけられる。このことは、第2の選択の前に、サイズは「39」と「40」のいずれかであったという事実によって示されている。

10

20

30

40

50

もし、ユーザが除外されている値を選択した場合は、少なくとも、その新たな選択と矛盾するそれまでのすべての選択は、その状態をそれに応じて変化させる。また、ユーザが一度に複数のデータ要素の値を選択することも可能である。

望ましくは、本発明による方法はAND機能、すなわち論理AND演算によって関連づけられる同一のデータ要素の型式のうちの第1のデータ要素の値と第2のデータ要素の値とを選択すること、を許容する。これは、SQLによって実行すると、時間がかかり、しかも煩雑な操作である。本発明による方法によれば、ユーザは単に、ANDカテゴリーにある所望のデータ要素の型式を特定すればよい。

ANDの原理は、表40-46を参照しつつ説明される。表40-44に示されているように、4つの異なる国において7種類の異なる調査が実施された場合を想定する（枠301）。スウェーデンを選択した結果は、表43と44に示されている。調査1003、1005及び1006は除外されている。何故ならばこれらのいずれも調査もスウェーデンでは実施されていないからである。イタリアも除外されている。何故ならばスウェーデンとイタリアの両方で実施された調査はないからである。

データ要素の型式が、ANDのカテゴリーにあると特定された場合は、プログラムはまず、すべてのデータ要素の型式がORのカテゴリーにあった前述の場合と同様に、バイナリ化された表を検索し、その結果として、調査1001、1002、1004及び1007を見つける。スウェーデンを選択したことによって除外された国を見つけるためには、最初の検索の間に見つけられた前述の調査を入力（選択された値）として、バイナリ化された表をもう一度検索する。この第2の検索の結果として、イタリアについてヒットは無く、すなわち、除外される。

ユーザが次のステップで、フィンランドを選択する場合は、その結果は表45と46に示されているようになり、すなわち調査1001及び1007のみが対象となりうる。バイナリ化された表の最初の検索が、スウェーデンとフィンランドとを選択された値として、前述のように実行される。複数のヒットのうちのいずれがAND条件を満足するかを決定するために、プログラムは頻度ベクトルをチェックする。頻度カウントが選択された国の数、ここでは2、に等しいような調査のみが、このAND条件を満足する。より一般的には、第1のデータ要素の型式がANDカテゴリーにあるものと特定され、この型式のn個のデータ要素の値が選択されている場合は、頻度カウントがnに等しいような第2の型式のデータ要素の値のみが対象となりうる。

本発明による方法のANDモードの望ましい実施の形態について、図3を参照しつつ以下により詳細に説明する。

以下では、表40に対応するバイナリ化された表がすでに生成されており、ユーザ選択ベクトルと状態ベクトルと頻度ベクトルとが例示されていると想定する（枠301）。また、説明の容易のために、表の右側カラムはANDカテゴリーにあるものと特定されたものであると想定する。

ANDとして特定されたデータ型式の1または複数のデータ要素の値をユーザが選択したことが検知されると、そのユーザの選択は、ユーザ選択ベクトルにエンコードされる。そして、状態ベクトルに初期値が割り当てられる。ORモードにおける場合とは異なり、状態ベクトルのすべての要素は「1」に設定される、何故ならば、ANDリストの中の1または複数の値を選択したことは、同一のリストの中の別の値を除外したことを意味するわけではないからである。しかし、ANDモードにおいて最初にバイナリ化された表の検索をする時に、その表の左側カラムに対応する状態ベクトルは、そのバイナリ化された表の右側カラムに対応するユーザ選択ベクトルと共に、入力として用いられる。

バイナリ化された表を最初に検索する前に、その表の左側カラムに対応する頻度ベクトルはクリアされる（枠302）。そして、その表のそれぞれのデータレコードについて（枠303）、次のステップが実行される。もし、表の右側カラムのデータ要素の値に対応するユーザ選択ベクトルの要素が設定されている場合、左側カラムのデータ要素の値に対応する頻度ベクトルの要素が1だけ増加される（枠304）。

すべてのデータレコードの検索が終了すると（枠305）、次のステップが実行される。

頻度ベクトル要素がANDカテゴリーのリストの中でユーザによりなされた選択の数、すなわち右側のユーザ選択ベクトルにおける1の総数、に等しくない場合は、表の左側カラムに対応する頻度ベクトルのそれぞれの要素について、左側の状態ベクトルの対応する要素がクリアされる（枠306）。

前述のステップの結果は、ユーザ選択に含まれている、表の左側カラムのデータ要素の値である。右側カラムのデータ要素の値の状態は、この表の最初の検索によって影響を受けない。右側カラムのどのデータ要素の値がそのユーザの選択によって含まれるのかを決定するために、その表の左側カラムに対応する更新された状態ベクトルとその表の右側カラムに対応する状態ベクトルを入力として、バイナリ化された表の第2の検索が実行される。

10

まず、右側カラムの頻度ベクトルがクリアされる（枠307）。次に、その表のそれぞれのデータレコードについて（枠308）、次のステップが実行される。もし、表の左側カラムのデータ要素の値に対応するユーザ選択ベクトルの要素が設定されている場合、右側カラムのデータ要素の値に対応する頻度ベクトルの要素が1だけ増加される（枠309）。

すべてのデータレコードの検査が終了すると（枠310）、右側の状態ベクトルの要素であって、対応する左側の頻度ベクトルの要素がゼロであるような、これらの要素はクリアされる。そして、「1」であるような状態ベクトルの要素は、ユーザの選択に含まれているデータ要素の値を示し、その結果は、表43及び44または45及び46に示されるようになる。

20

以上の説明では、カラムが2つの表を参照しつつ、ANDモードについて説明した。しかし、複数カラムの表から2カラムの表を生成することによって、複数カラムの表の中の1またはそれ以上のカラムに関連してANDを実行することも可能である。ANDが1つのカラムについて実行される場合には、2カラムの表を生成するために、その複数カラムの表から、このカラムとANDカラムとが抽出される。そして、ANDが2またはそれ以上のカラムについて実行される場合は、これらのカラムは合成されて新しい一つのカラムにされ、2カラムの表を生成するためにANDカラムと共に用いられる。

時に、データベースが表に除外のみを有し、その他のすべての関係は、通常の法則に従っている場合がある。このことを反映するために、余分なデータレコードを表の最後に付け加えて、その中の1またはそれ以上、しかしすべてではない、のデータ要素の値をアスタリスク（*）とすることができる。このアスタリスクは、そのデータ要素の値が特定されておらず、または開放されていることを示す。

30

「ワイルド・カード」レコードまたはデフォルトレコードと称される、このデータレコードについての情報は、別に格納される。この情報に基づいて、プログラムは、その表がワイルド・カードのデータレコードを含むことと、その表がそれに従って取り扱われるべきものであることを、認識する。

ここで、表47及び48に示した例について考えてみる。表47は、自動車のメーカーとその国とを示している。表48は、それぞれの自動車メーカーのシリンダ容積を示している。表48は表47よりも多くのメーカーを示しており、表47に示されていないのは、日本のメーカーであることが分かる。コンピュータの主メモリに読み込まれた表に対して、この情報

40

を含ませるために、表47にワイルド・カードのデータレコードを追加することができ、そのメーカーの値はアスタリスクとして、その国の値は日本とする。プログラムがワイルド・カードを含む表をバイナリ化する時には、ワイルド・カードのカラムと同一の値を有するデータレコードを互いに隣り合わせになるように、その表をソートする。

次のステップでは、プログラムはその表の中のデータレコードの数と同数の位置を有するベクトルを生成する。そして、そのベクトルの最初の位置には、0が挿入される。その表が検索される時に、ワイルド・カードのカラムの値が前のデータレコードのこれらのカラムの値と同一である場合は、検査されるデータレコードに対応するベクトルの要素は、ベクトルの前の値と同一の値に設定される、そして、同一でない場合には、同一の値に1を

50

加算した値に設定される。バイナリ化された表の検索が終了した時、ユニークな (unique) ヒットの数にカウントされ、すなわち、前述のベクトルにおいて異なる値を有するヒットのみがカウントの対象とされる。ヒットの数それぞれのワイルド・カードのカラムの中の選択された値の数の積に等しい場合は、この場合のみ、そのワイルド・カードのデータレコードは除外されたものである。

例えば、表 48 においてシリンダ容積として「1. 4」という値が選択された場合を考えると、メーカとして「フォード」と「サブ」が含まれる事となる。これは、また、表 47 において 2 つのユニークなヒットを生ずる事となり、何故ならばワイルド・カードのカラムの値が異なるからである。ここで、ワイルド・カードのカラムにおいて 2 つの値が選択され、2 つのヒットがあるので、そのワイルド・カードのデータレコードは除外されている。

10

もし、シリンダ容積の値として「2. 0」が代わりに選択されると、自動車メーカ「ボルボ」と「マツダ」が含まれ、表 47 に対する入力として選択される。しかし、ワイルド・カードのカラムのこれらの選択された値は、1 よりも多いヒット (ボルボ・スウェーデン) を生じさせない。その結果として、ワイルド・カードのデータレコードは依然として選択が可能である。

本発明によれば、現存しているデータ・ベースの応用可能性を広げることができる。例えば、多くの企業が多数のデータ・ベースを有し、それらのデータ・ベースは、ネットワーク上の別々のコンピュータの上で走るさまざまなデータ・ベース管理システムに組み込まれている。しかし、個々のユーザはそれらのさまざまなデータ・ベースに含まれている情報のうちのほんの一部分のみを必要とするに過ぎない場合がしばしば生ずる。本発明によれば、ユーザは、特定のプログラミングをなんら必要とせずに、さまざまなデータ・ベースから関係のあるデータレコードの表を引き出して、全く新規で均一な、例えば企画などを創り出すことができる。

20

さらに、そのユーザはデータ・ベースからのデータにローカル・データやコメントを結合することもできる。データの処理はローカルに行われるので、ローカル・データは処理されて中央のデータと同等の重要度を有するようになる。これによって、ローカル・データと中央のデータとの組み合わせを実現することができる。

さらに、従来のデータ・ベースにおいては、ビデオ、音声およびカラー写真などを含むマルチメディアは容易に扱えない場合がしばしば生じていた。また、それらは、通常のコンピュータ・ネットワークを介して伝送されることが多くはなかった。本発明によれば、いたるところに存在しているマルチメディアを接続することが可能になる、例えば、ローカルに接続された CD プレイヤーなどである。関係のあるマルチメディアのファイル名とフォーマットは、特定の値によって表に簡単に与えられる。その値が選択されると、対応するマルチメディアの実演が開始される。

30

補遺 A

表 1

衣 服	色	サ イ ズ	科目番号
シ ャ ツ	青	3 9	1 0 0 1
シ ャ ツ	青	4 0	1 0 0 1
シ ャ ツ	赤	3 9	1 0 0 3
ジ ー ン ズ	青	3 2	2 1 0 0
ジ ー ン ズ	青	3 5	2 1 0 0

表 2

科目番号	価 格
1 0 0 1	1 8 0
1 0 0 3	2 2 0
2 1 0 0	1 8 0

10

表 3

衣 服	
シ ャ ツ	0
ジ ー ン ズ	1

表 4

色	
青	0
赤	1

表 5

サ イ ズ	
3 9	0
4 0	1
3 2	2
3 5	3

20

表 6

科目番号	
1 0 0 1	0
1 0 0 3	1
2 1 0 0	2

表 7

価 格	
1 8 0	0
2 2 0	1

30

表 8

衣 服	色	サ イ ズ	科目番号
0	0	0	0
0	0	1	0
0	1	0	1
1	0	2	2
1	0	3	2

10

表 9

科目番号	価 格
0	0
1	1
2	0

20

表10

衣 服
シャツ
ジーンズ

表11

サ イ ズ
39
40
32
35

表12

色
青
赤

表13

価 格
180
220

表14

科目番号
1001
1003
2100

30

表15-29は衣服のデータ要素の型式において「シャツ」を選択した場合を示す。

表15

ユーザ選択ベクトル - 衣服

1	0
---	---

表20

初期状態ベクトル - 衣服

1	0
---	---

表25

頻度ベクトル - 衣服

3	0
---	---

表16

ユーザ選択ベクトル - サイズ

0	0	0	0
---	---	---	---

表21

初期状態ベクトル - サイズ

1	1	1	1
---	---	---	---

表26

頻度ベクトル - サイズ

2	1	0	0
---	---	---	---

表17

ユーザ選択ベクトル - 色

0	0
---	---

表22

初期状態ベクトル - 色

1	1
---	---

表27

頻度ベクトル - 色

2	1
---	---

表18

ユーザ選択ベクトル - 価格

0	0
---	---

表23

初期状態ベクトル - 価格

1	1
---	---

表28

頻度ベクトル - 価格

1	1
---	---

表19

ユーザ選択ベクトル - 科目番号

0	0	0
---	---	---

表24

初期状態ベクトル - 科目番号

1	1	1
---	---	---

表29

頻度ベクトル - 科目番号

1	1	0
---	---	---

表30から34は衣服のデータ要素の型式から「シャツ」を選択した後のユーザ・インターフェイスの一例を示す。

表30

衣 服
シャツ
ジーンズ

表31

サ イ ズ
32
35
39
40

表32

色
青
赤

表33

価 格
180
220

表34

科目番号
1001
1003
2100

カラー・コード
選択可
選択済
除 外

10

20

30

40

50

表 3 5 から 3 9 は衣服のデータ要素の型式から「シャツ」、サイズのデータ要素の型式から「4 0」を選択した後のユーザ・インターフェイスの一例を示す。

表 3 5

衣 服
シ ャ ツ
フ ェ ャ ッ

表 3 6

サ イ ズ
3 2
3 5
3 9
4 0

表 3 7

色
青
赤

表 3 8

価 格
1 8 0
2 2 0

表 3 9

科目番号
1 0 0 1
1 0 0 3
2 1 0 0

カラー・コード
選択可
選択済
除 外

表40

調 査	国
1001	スウェーデン
1001	フィンランド
1002	スウェーデン
1002	デンマーク
1003	デンマーク
1003	フィンランド
1004	スウェーデン
1005	フィンランド
1005	イタリア
1006	デンマーク
1006	イタリア
1007	スウェーデン
1007	フィンランド
1007	デンマーク

10

20

表41

調 査
1001
1002
1003
1004
1005
1006
1007

表42

国
スウェーデン
フィンランド
デンマーク
イタリア

30

40

表43と44は、ANDモードにおいて、国のデータ要素の型式から「スウェーデン」を選択した後のユーザ・インターフェイスの一例を示す。

表43

調 査
1001
1002
1003
1004
1005
1006
1007

表44

&国
スウェーデン
フィンランド
デンマーク
イタリア

カラー・コード
選択可
選択済
除 外

10

表45と46は、ANDモードにおいて、国の表から「スウェーデン」と「フィンランド」を選択するユーザ・インターフェイスの一例を示す。

表45

調 査
1001
1002
1003
1004
1005
1006
1007

表46

&国
スウェーデン
フィンランド
デンマーク
イタリア

カラー・コード
選択可
選択済
除 外

20

表47

メーカ	国
ボルボ	スウェーデン
サーブ	スウェーデン
フォード	アメリカ
フォード	ヨーロッパ
B M W	ジャーマニー
*	日 本

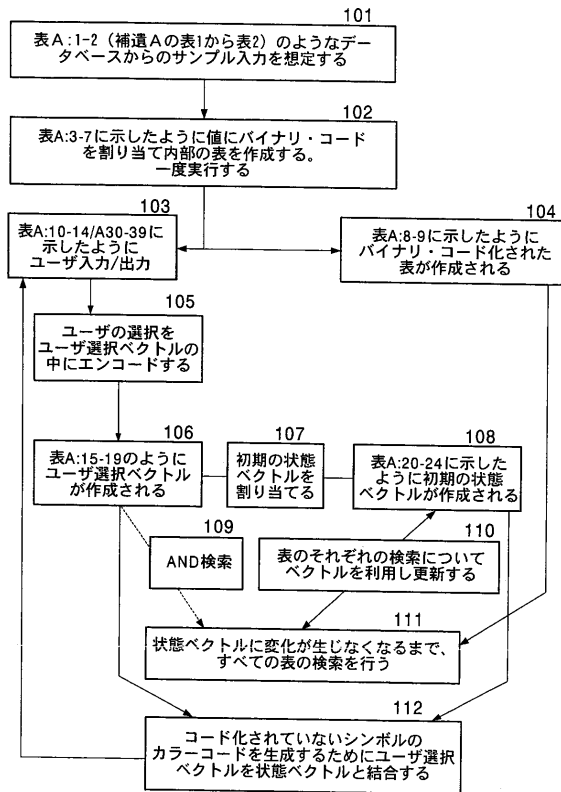
表48

シリンダ容量	メーカ
0.8	トヨタ
0.9	マツダ
1.0	トヨタ
1.0	ホンダ
1.4	フォード
1.4	サーブ
2.0	ボルボ
2.0	マツダ
2.8	B M W

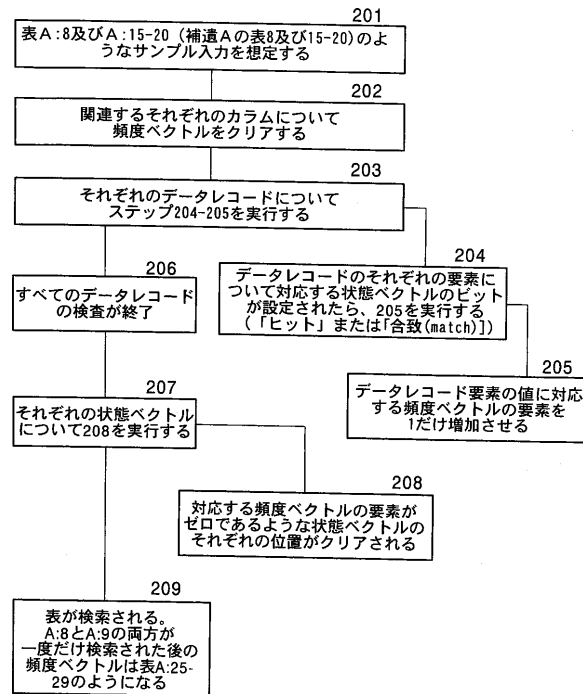
30

40

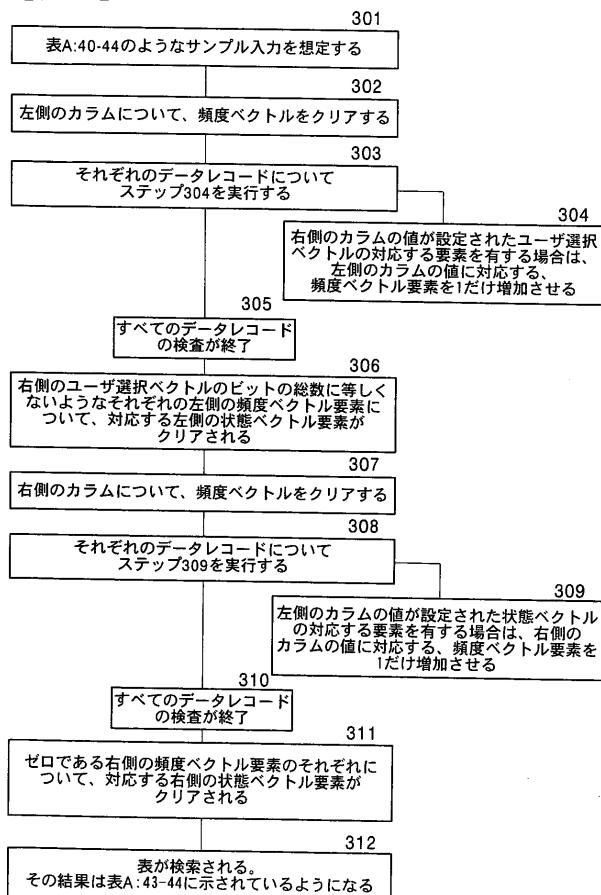
【図 1】



【図 2】



【図 3】



フロントページの続き

- (72)発明者 ゲストリウス, スタファン
スウェーデン国エスー2 4 5 6 3、ヤルプ、ドレシンベージェン、1 4
- (72)発明者 フランソン、ホーガン
スウェーデン国エスー2 2 4 7 2、ルンド、ゼボルグスリンガン、1 8
- (72)発明者 ブラレン, ベルグ
スウェーデン国エスー2 2 2 4 0、ルンド、モレベングスベージェン、7 1

合議体

審判長 長島 孝志

審判官 飯田 清司

審判官 長 由紀子

- (56)参考文献 特開平0 6－0 3 5 9 7 3 (J P, A)
特開平0 7－2 7 1 7 9 9 (J P, A)
特開平0 7－1 3 4 7 1 4 (J P, A)
特開平0 5－3 0 7 6 6 7 (J P, A)
特開平0 5－0 2 0 3 4 5 (J P, A)
特開平0 5－0 8 1 2 7 3 (J P, A)

- (58)調査した分野(Int.Cl., D B名)
G06F17/30