

リレーショナルデータベースシステムの動向

Trends of Relational Database Systems

リレーショナルデータベースは、その使いやすさ、柔軟性から、最近のユーザーニーズに合致したデータベースとして急速に脚光を浴び始めている。リレーショナルDBMSの開発も米国を中心に行なわれており、最近幾つかの商用システムが出始めている。

本論文では、まずリレーショナルデータベースの基礎概念について説明し、次に、従来のデータベースシステムと比較して、リレーショナルデータベースシステムの長所と短所を明らかにする。リレーショナルDBMSのユーザーニーズを分析し、リレーショナルDBMSの動向に触れた後、現在の技術課題と将来展望について述べる。

福嶋慎一* *Shin'ichi Fukushima*

武藤英男** *Hideo Mutô*

附田克晃*** *Katsuaki Tsukuda*

1 緒言

リレーショナルデータベースは、1970年にE.F.Coddが論文¹⁾で発表したリレーショナルデータモデルの考えに基づくデータベースである。このモデルでは、データベースを表の集まりとして考えることができ、データベースに対する操作が表に対する操作として表現できる。表のもつ理解しやすさ、使いやすさ、柔軟性から、最近のユーザーニーズに合致したデータベースとして、リレーショナルデータベースが急速に脚光を浴び始めている。リレーショナルDBMS(データベース管理システム)の開発も米国を中心に行なわれており、幾つかの商用システムが出始めている。

本論文では、まずリレーショナルデータベースの基礎となる考え方、理論を説明し、従来の階層形やネットワーク形のデータベースとの比較で長所、短所を明らかにする。次に、リレーショナルDBMSに対するユーザーニーズの分析と製品動向の紹介を行なう。そして、簡潔で柔軟性の高いユーザーインターフェースを提供する反面として、リレーショナルDBMSが解決しなければならない技術問題について述べる。

2 リレーショナルデータベースの基礎概念

データモデルとは、現実世界のデータをどのように整理し関連づけてデータベースにするか、というモデル化の手段を与えるものであり、ユーザーから見えるデータの論理的構造とデータに対する操作方法を規定するものである。

リレーショナルモデルでは、図1に示すように、データベース中のデータは表として表現できる。この表は、数学の集合論でいう関係(Relation)を具体的に表現したものに相当し、次の規則を満たすものである。

- (1) 表及び欄には、名前が付けられている。
- (2) 一つの表には、同じ行がない。
- (3) 一つの表の中での行の出現順序や欄の出現順序は、自由である。
- (4) 一つの枠(フィールド)には、一つの値しか入れられない。

従来のファイルの概念で言えば、表がファイルに、行がレコードに、欄がレコード中のデータ項目又はフィールドにそれぞれ対応する。

リレーショナルモデルでのデータ操作は、表に対する操作であり、関係演算と呼ばれる選択(Selection)、射影(Projec-

従業員表

従業員番号	プロジェクトコード	作業従事率	部門コード	部門名
2150	P1	40	14	データベース部
2150	P3	60	14	データベース部
2200	P2	20	23	言語プロセッサ部
2200	P3	50	23	言語プロセッサ部
2200	P4	30	23	言語プロセッサ部

表の規則

- (1) 表及び欄には名前が付けられている。
- (2) 同じ行がない。
- (3) 行や欄の順序は自由である。
- (4) 一つの枠(フィールド)には、一つの値しか入れられない。

図1 表形式のデータ構造 リレーショナルデータベースは表の集まりである。

tion)、結合(Join)の3操作が基本となる。「選択」は、表からある条件を満足する行を取り出す。「射影」は、表の不要な欄を切り落とし、必要な欄だけを残す。「結合」は、二つの表を共通の欄をもとにしてはり合わせて一つの表にまとめる。射影、結合の操作例を図2に示す。この図の射影の例では、在庫表から部品番号と数量の欄を切り落とし、部品名の欄だけから成る表を得る。結合の例では、部品名表と部品数量表に共通の部品番号の欄をもとに、部品番号の一致する行を二つの表から取り出して、部品番号、部品名、数量の欄から成る一つの表を作り出す。これらの操作に共通する特徴は、表というレコードの集合に作用して得られる結果が再び表になることである。これらの表操作を表現するための言語として、関係代数、関係論理、写像形、図形形など種々の言語が提案されている。

データベースを構成する表をどのように設計するかの指針として、表の正規化という概念がある。正規化とは、欄の間の対応関係に注目して、一つの表に含まれる別々の対応関係をそれぞれ別の表に含まれるように、元の表を分解していく

* 日立製作所ソフトウェア工場 ** 日立製作所システム開発研究所 *** 日立ソフトウェアエンジニアリング株式会社

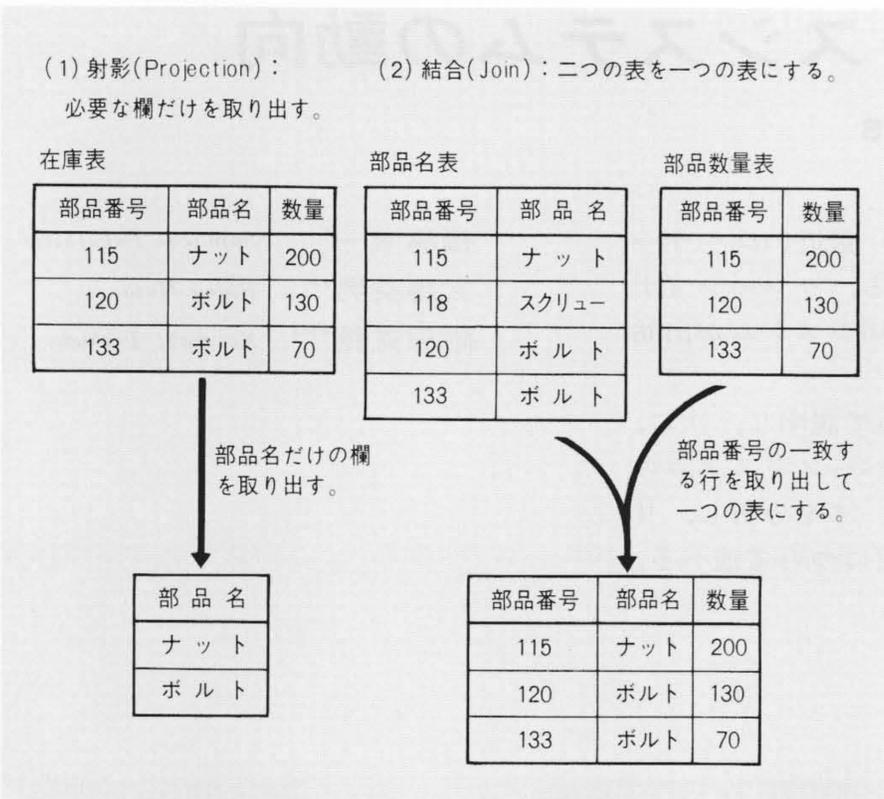


図2 表に対する操作 射影, 結合などの操作により, 基本となる表から必要な表を作り出す。

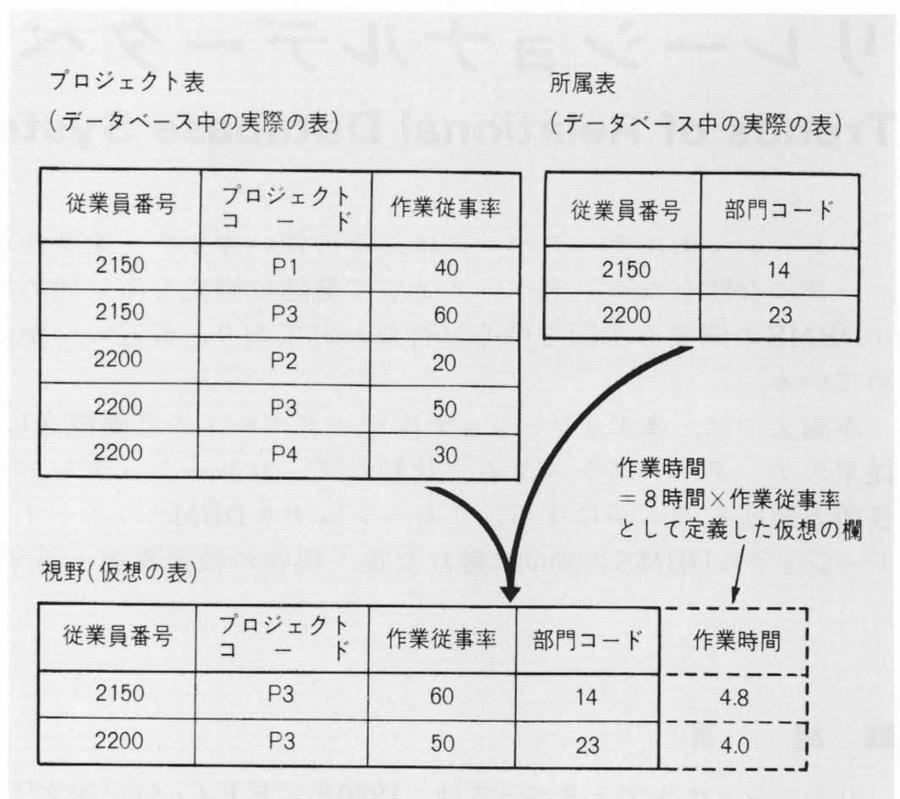


図4 視野(仮想の表) データベース中のプロジェクト表, 所属表からプロジェクトP3に関するデータを, 一つの表にまとめた視野を定義する。

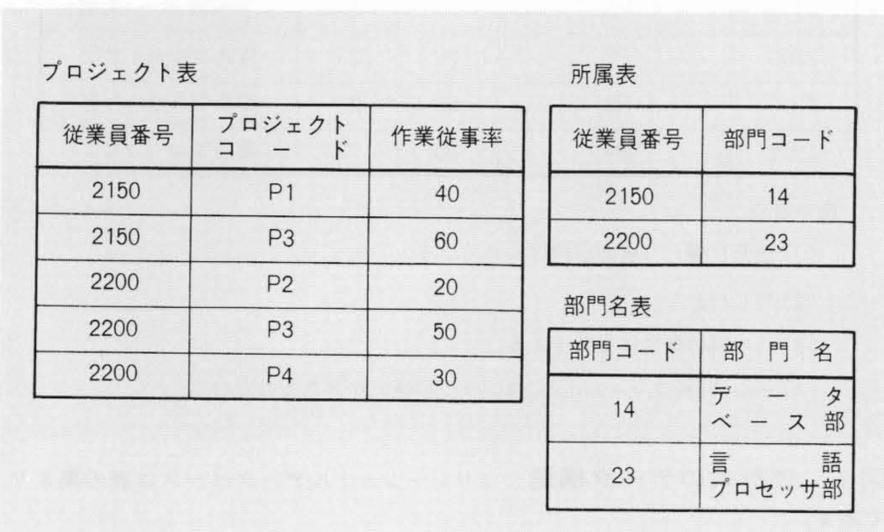


図3 正規化された表 図1の従業員表を正規化して, プロジェクト表, 所属表, 部門名表の三つの単純な表に分解する。

ことである。例として図1の従業員表を考えてみる。この従業員表は例えば従業員番号2150の人は部門コード14のデータベース部に所属し, プロジェクトP1とP3の作業にそれぞれ40%と60%の割合で従事していることを示している。この従業員表には同じ従業員番号をもつ行が存在し, その従業員の所属が変わった場合, その従業員番号に対応する部門コード(及びこの場合部門名)をすべて更新する必要がある。もしその一部の更新を忘れると, 従業員番号と部門コードとの対応に矛盾が生ずる危険性がある。このような不都合を取り除くために, 従業員表に含まれている幾つかの事実関係, すなわち従業員の所属に関する事実, 従業員とプロジェクトとの従事関係に関する事実, 及び部門に関する事実を別々の表に分離して, 図3に示す三つの表に分解する。このように正規化によって, 表に冗長な情報が含まれないようにして, 表の更新時などに起こるデータの不整合の問題を防止することができる。

正規化された表を正規形の表と呼ぶ。正規形は, 正規化の程度により, 第1, 第2, 第3正規形へと進んでいく。第1正規形の表は, 図1に示す規則だけを満足する単純な表であ

る。図3の表はすべて第3正規形であり, 結合操作により元の表を復元できる。

データベース中に実際に存在する表のほかに, 視野(View)と呼ばれる仮想の表という概念がある。これは, データベース中に存在する表の論理的な見方を定義するものであり, 実在する表の一部分の行又は欄だけを見ることができたり, 二つ以上の実在する表を組み合わせるとして定義したりすることができる。また, 実在する表にはない仮想の欄を視野として定義できる。図4に示す視野の例では, データベース中に実際に存在するプロジェクト表, 所属表からプロジェクトP3に関するデータを一つの表として見ることができるとして定義している。この例では, 作業時間の欄を仮想の欄として付加している。

視野を通してデータベース中の表を見せることにより, その表を利用するユーザーに都合のよい表を提供することができる。そのユーザーに許されるデータだけを見せるように制御することができ, データの安全保護を図ることができる。

以上述べてきた基礎概念の詳細については, 他の文献^{2,3)}を参照されたい。

3 リレーショナルデータベースシステムの特徴

リレーショナルデータベースシステムは, 従来の階層形及びネットワーク形のデータベースシステムと比較して, 次に述べる長所と短所をもつ^{4)~6)}。

3.1 リレーショナルデータベースシステムの長所

(1) 理解しやすい。

データ構造は表形式であり, データ操作は表の操作である。他のデータベースシステムのように, 固有のデータ構造やデータ表現方法を習得する必要がない。表の概念や操作は広く理解されており, 基礎となる概念がすぐ理解できる。

(2) データ独立性が高い。

データのもつ情報はデータの値として明示的に表現され, データ操作はデータの値だけを使って行なう。従来のシステムで見られるポインタやリンケージパスなどのデータベース

の内部構造に依存した補助手段は、ユーザーには見えない。ユーザーから見えるデータ構造とそれを実現する構造とを区別している。このため、ユーザーはデータベースの内部構造や処理方式の変更に影響されず、高いデータ独立性を実現できる。

(3) データ操作が強力で使いやすい。

従来のように1レコードずつ処理するのではなく、レコードの集合である表を一括して処理できて強力である。また、ユーザーは結果として何を得心かだけを記述すればよく、その結果をどのようにして得心かはDBMS側の責任となる。

(4) データ定義がデータ操作と一貫している。

データの定義を表の定義、表に対する操作として考えることができ、データ定義と操作を一貫して扱える。これにより、データベース管理者、プログラム作成者、エンドユーザーが、表という共通の概念を通して会話できるようになる。

(5) 理論的裏付けがある。

リレーショナルデータベースは理論から始まり、確立した理論的基礎がある。今後の理論的研究の成果も受けることができる。

3.2 リレーショナルデータベースシステムの短所

(1) 大容量データベースでの実用上の性能が十分でない。

リレーショナルデータベースシステムでは、制限の非常に少ない表の操作を許すために、現在の技術では内部的に多くの索引を利用する。データ検索では有利な索引の多用が、データ更新を多く含む操作では、索引の内容の更新のために性能を低下させることになる。性能の問題は、リレーショナルモデルそのものの問題というよりは、それを実現するDBMSの問題である。

(2) データの意味の表現力が不足している。

構造形やネットワーク形のデータベースでは、データ取り出しの道筋があらかじめ定義されているので、データ間の関係を理解するのは早い。これに対してリレーショナル形では、表の間の関係は、ユーザーがデータの値で自由に関係づけられる代わりに、データベースの内容だけで理解するには時間を必要とする。

以上述べてきたリレーショナルデータベースシステムの長所と短所を表1にまとめる。

4 リレーショナルDBMSに対するユーザーニーズ

ハードウェアの価格性能比の向上とコンピュータの普及によるコンピュータ適用可能業務の拡大を背景として、リレーショナルDBMSに対して図5に要約される次のようなユーザーの要求がある。

(1) 非定型業務を簡単に処理したい。

突発的に発生する非定型の業務を、プログラムを作成せずすぐに処理したい。

(2) 自由に強力なデータ検索を行ないたい。

データベースを、任意の情報を条件として検索したい。

(3) 簡単に、高水準のユーザーインタフェースがほしい。

最小限の学習でエンドユーザーにも使えて、端末からもプログラム中でも同様に使えるユーザーインタフェースがほしい。

(4) データベースを簡単に作成したい。

小容量のデータを含め、種々のデータをデータベース化して、豊富なデータ操作機能を利用したい。データベースの物理的構造や参照時のことを考えずに作成したい。

(5) 変動するデータベース環境に対応できる柔軟性がほしい。

フィールドの追加などのデータ構造の変更を、必要なとき

表1 リレーショナルデータベースシステムの長所と短所 長所は主にリレーショナルデータモデルのもつ長所であり、短所は主に実現方式の問題である。

区分	項目	説明
長所	理解しやすい。	表の概念や操作は広く知られている。
	データ独立性が高い。	データの値だけを意識してデータを操作する。
	データ操作が強力で使いやすい。	集合操作ができ、何を得心かだけを記述する。
	データ定義と操作が一貫している。	表の定義、表の操作としてデータを定義できる。
	理論的裏付けがある。	理論から始まり、理論的基礎が確立している。
短所	大容量データベースでの実用上の性能が十分でない。	更新の多い表操作で性能が十分でない。
	意味の表現力が不足している。	表の間の関係がすぐには理解できない。

に、データベースを作り直さずに行なえるようにしたい。

(6) データベースの管理、運営を簡単にしたい。

個人レベルのデータベースは、個人で管理できるようにしたい。システムの円滑な運営のために、システムの管理情報を簡単に入手する手段がほしい。

これらの要求に対して、リレーショナルDBMSは、表形式のデータ構造、端末からの自由な表の操作、表の動的な追加、削除、変更、システム管理情報の表形式での表示という形で対応しようとしている。

5 リレーショナルDBMSの動向

1974年にIBM社でSystem R⁷⁾の研究が開始されて以来多くの実験システムが試作され、最近幾つかの商用リレーショナルDBMSが出始めている。また、リレーショナルモデルの長所を取り入れたDBMSも大形コンピュータからミニコンピュータまで多く見受けられるようになった。

表2に代表的なリレーショナルDBMSを示す。完全なリレーショナルDBMSとは、E.F.Codd⁸⁾によれば、データベースを表の集まりとしてユーザーに見せ、更に、彼の提唱した関係演算と同じ程度に強力なデータ操作言語を実現したDBMS

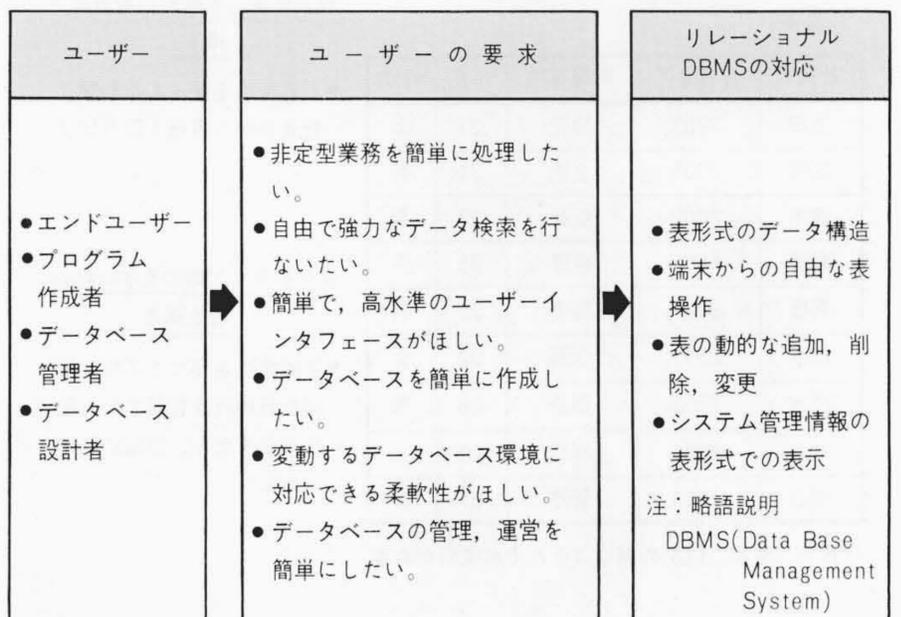


図5 リレーショナルDBMSに対するユーザーの要求 コンピュータ適用可能業務の拡大と既存DBMSに対する改善要求を背景に、多くの要求が出てきている。

表2 代表的なリレーショナルDBMS 完全なリレーショナルDBMSのほかに、リレーショナル的なDBMSが市場に出ている。

分類	DBMS名	提供者	主要な適用機種
完全なリレーショナルDBMS	SQL/DS	IBM	中形機
	ORACLE	Relational Software Inc.	ミニコンピュータ
	INGRES	Relational Technology, Inc.	ミニコンピュータ
その他	ADABAS	Software AG	中, 大形機
	Model 204	Computer Corp. of America	中, 大形機
	System/38 データベース管理機能	IBM	小形機

注：略語説明 SQL/DS(Structured Query Language/Data System)
INGRES(Interactive Graphics and Retrieval System)
ADABAS(Adaptable Data Base System)

である。その他に、4章で述べたリレーショナルDBMSに対するユーザーニーズを満たすものが市場に出ている。

6 技術課題と将来展望

リレーショナルDBMSの現在の主要な技術課題は性能向上である。このための方策として、データ検索手順の最適化とデータベースマシンが考えられている。

データ検索手順の最適化は、リレーショナルDBMS特有の問題である。3章で述べたように、リレーショナルデータベースのデータ検索では、ユーザーは何をほしいかを指定するだけである。このため、データベースからデータを取り出す手順はDBMS側で考えなければならない。このデータの取り出し手順は一般には複数存在する。最適化とは、考えられる取り出し手順から最も効率の良い手順を決定し実行する処理である。最適化の簡単な例を図6に示す。この図の社員表の社員名、年齢及び性別の欄にはそれぞれ索引が設けられているとする。「年齢順に社員名を得る」という検索要求に対しては、年齢索引を使って社員表の各行を取り出すことにより、年齢による並べ換え(ソート)処理を省略できる。また、「20歳

の男子の社員名を得る」という検索要求に対しては、まず年齢索引で20歳の社員の行を2行取り出して、次にその性別を調べるほうがよい。最初に性別索引で男性の社員を取り出してから年齢を調べても正しい検索結果が得られるが、この場合、社員表から5行取り出さなければならなくなる。

もう一つの性能向上策であるデータベースマシンとは、処理量の多いデータベース処理を、中央の処理装置から分離して別の専用ハードウェアに分担させようという趣旨のものであり、次のような形態のものが種々提案されている。

- (1) DBMS機能をホストコンピュータから分離し、別の専用コンピュータで実行するソフトウェアバックエンド方式。
- (2) ホストコンピュータに何らかのDBMS向きの機能をファームウェアなどで補強して、DBMSの性能向上を図る主計算機機能強化方式。
- (3) DBMSとデータベース全体を専用ハードウェア化するハードウェアバックエンド方式。

データベースマシンとして現在商品化されているものはごくわずかであるが、今後のデータベースマシン技術の発展とハードウェアの価格性能比の向上により、十分に現実性をもったものになると考えられる。

ユーザーの多様なニーズにこたえられる高い可能性をもつリレーショナルデータベースシステムは、ソフトウェアとハードウェア一体となったサポートにより、今後ますますデータ処理分野に浸透していくものと予想される。

7 結 言

以上本論文では、次のことについて紹介した。

- (1) リレーショナルデータベースの基礎となるデータモデル、正規化、正規形、視野の概念
- (2) 階層形やネットワーク形と比較してのリレーショナルデータベースシステムの長所と短所
- (3) コンピュータ適用可能業務の拡大を背景としたリレーショナルDBMSに対するユーザーニーズの分析
- (4) 現在、市場に出ている代表的なリレーショナルDBMS
- (5) データ検索手順の最適化とデータベースマシンを中心とした性能向上のための技術課題

リレーショナルDBMSに対するユーザーニーズは強く、日立製作所でもそのユーザーニーズにこたえるリレーショナルDBMSを現在開発中である。

参考文献

- 1) E.F.Codd: A Relational Model of Data for Large Shared Data Banks, Comm. ACM, 13, 6, 377~387 (1970)
- 2) C.J.Date: An Introduction to Database Systems, Third Edition, Addison-Wesley Publishing Company, Reading, Massachusetts (1981)
- 3) 植村: データベースシステムの基礎, オーム社 (1979)
- 4) G.Sandberg: A primer on relational database concepts, IBM Systems Journal, 20, 1, 23~40 (1981)
- 5) 植村: リレーショナル・データベースとは, コンピュートピア, 13, 通巻158号, 42~45 (1979)
- 6) 穂鷹: 商用リレーショナルDBMS, コンピュートピア, 13, 通巻158号, 48~55 (1979)
- 7) D.D.Chamberlin, et al.: A History and Evaluation of System R, Comm. ACM, 24, 10, 632~646 (1981)
- 8) E.F.Codd: Extending the Database Relational Model to Capture More Meaning, ACM Trans. Database Syst., 4, 4, 397~434 (1979)

社員名	社員番号	所属部門	年齢	性別
上田	7953	販売	21	女
佐藤	7305	企画	31	男
鈴木	8002	販売	21	男
高橋	6552	経理	35	男
高橋	8001	販売	20	男
田中	8201	企画	22	女
松本	7603	販売	28	男
村上	8051	経理	20	女
渡辺	7754	管理	27	女

検索要求: 年齢別に社員名を得る。

- 年齢索引を使って年齢順に社員表から各行を取り出す。

検索要求: 20歳の男子の社員名を得る。

- 年齢索引を使って年齢20歳の社員の行を社員表から取り出してから、性別の判定を行なう。

社員名、年齢、性別の欄にはそれぞれ索引がある。

図6 データ検索の最適化 検索の条件、索引の有無などに応じて、検索の最も効率の良い実行手順を決定する。