

## ARES におけるデータベース操作機能の拡張†

平川 正人<sup>\*\*</sup> 清水 達男<sup>\*\*</sup> 市川 忠男<sup>\*\*</sup>

一般ユーザを対象としたデータベース検索環境では、検索容易性に優れたユーザインタフェース機能が重要な意味をもつ。本論文では、条件にある程度の幅をもった問合せを受理する情報検索システム ARES の検索のメカニズムを明らかにする。ARES は関係モデルに基づくデータベースシステムであり、コンベンショナル・リレーションとセマンティック・リレーションという二つの型のリレーションをもっている。セマンティック・リレーションは要素間のセマンティックな関連を表現しており、ユーザの要求を反映させるためのセマンティック規定支援ツールを介して導かれる。検索条件にあいまい性が付与された問合せは、従来の関係代数演算の系列に変換されたのちに実行される。さらに、データは問合せ条件との一致度に従って順序づけられ、ユーザの希望する個数まで出力されるようになっている。これによって、一般の関係データベースシステムとの整合性をそこなうことなく、柔軟で処理効率の高い検索機能が提供できるようになった。また、国産乗用車データベースによる実験システムが現在ミニコンピュータ上で稼働中であり、これを通してシステムの有効性を確認した。

### 1. はじめに

筆者らは、データ間のセマンティックな相互関連度を評価する新しいスキーマを連想メモリの形態の上で ARES (A Memory, Capable of Associating Stored Information through Relevancy Estimation) として提案した<sup>1)</sup>。このスキーマでは、符号間距離という概念を導入してデータ間の相互関連度を評価する。これによって入力データに対してセマンティックな関連のあるデータを引きだすことができる。これはまた、データベース環境にこのスキーマを利用した場合、従来の断定的な問合せ条件だけでなく、ある程度の幅をもった条件解釈が実現されること、したがってより柔軟な問合せが可能となることを示している<sup>2)</sup>。筆者らは、自動車を対象とした物件案内データベースを構築し、これを通してシステムの有効性を確認した<sup>3)</sup>。なお、ここではユーザインタフェースとして QBE (Query-by-Example)<sup>4)</sup> を採用し、その上で“…に近い”という条件指定を行うための比較演算子“≐”を新たに導入した。さらにこの比較演算子を用いて記述された条件に対する許容度、すなわち許容される条件の幅は、感覚的な尺度でグラフィック・ターミナルを介してシステムにインタラクティブに与えられる。

本論文では、効率のよい新しい関連度評価法を示し、これに基づく検索のメカニズムを明らかにする。以後

ARES というときは、たんなる関連スキーマではなくて、関連度評価機能を取り入れた情報検索システム (Associative Information Retrieval System) 全体を指すものとする。

ARES は関係モデル<sup>5)</sup>に基づくデータベースシステムであり、コンベンショナル・リレーションとセマンティック・リレーションという二つの型のリレーションをもっている。セマンティック・リレーションは定義域 (domain) 内の個々の要素間のセマンティックな関連を表現しており、それはユーザの要求を反映させるための支援ツールを通して決定される。このセマンティック・リレーションをコンベンショナル・リレーションから独立して管理することによって、セマンティックの定義をユーザに解放することができ、かつ、システムの追従も容易となる。

検索条件にあいまい性が付与された場合には、検索条件が記述された属性 (attribute) の定義域に対応したセマンティック・リレーションに対する選択 (selection) 操作に続いて、コンベンショナル・リレーションとの結合 (join) 操作が行われる。このようにして、検索条件にあいまい性が付与された問合せもまた従来の関係代数演算の領域に持ち込まれ、確立された従来の技法で実行可能となる。

さらに、検索条件を満たす組 (tuple) が選定されると、これらは条件に対する一致度の度合いによって内部的に順序づけられる。そしてユーザによって希望出力数の上限が指定されると、その範囲内で問合せ条件との一致度の強いものから順に出力される。

† Augmentation of Database Manipulation Facility in ARES by MASAHIRO HIRAKAWA, TATSUO SHIMIZU and TADA O ICHIKAWA (Faculty of Engineering, Hiroshima University).  
\*\* 広島大学工学部

以下では、ARES におけるセマンティック・スキーマとデータ・セマンティックスの規定を支援するツールについて述べたのちに、そのスキーマに基づいた検索のメカニズムを詳説する。

## 2. データ・セマンティックスとその規定支援ツール

### 2.1 データ・セマンティックス

E. F. Codd によって提案された関係モデル<sup>5)</sup>は、現実世界の事象を  $n$  個の定義域の直積の部分集合という  $n$  項関係で定義している。しかしながらこの定義域という概念では、個々の要素間のセマンティックな関連は記述されていない。ARES では、このセマンティックな関連を積極的に取り入れることによって、より強力なデータベース環境を提供する。

要素間の関連の様子にしたがって定義域を次の三つに分類する。すなわち、

- タイプ I: 要素が互いに分離して、要素間には何ら関連が存在しない要素からなる定義域。
- タイプ II: 数値のように、要素間に一次元的な関連が存在する要素からなる定義域。
- タイプ III: 要素間に多次元的な関連構造をもつ要素

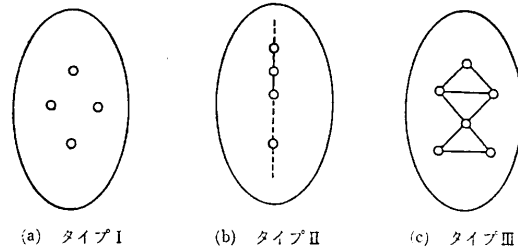


図 1 定義域の分類  
Fig. 1 Classification of domains.

からなる定義域。これらを図示したものが図 1 である。

### 2.2 セマンティック・スキーマ

データ・セマンティックスを記述するために、ARES ではセマンティック・リレーションと呼ぶ特別なリレーションを定義する。これにともない、以下では従来のリレーションをコンベンショナル・リレーションと呼んでセマンティック・リレーションと区別する。

セマンティック・リレーションは定義域内の要素間の関連情報を表現するもので、要素間に関連づけがなされる定義域 (タイプ II, タイプ III) ごとに存在する。ある定義域  $D_i$  に対応するセマンティック・リレーシ

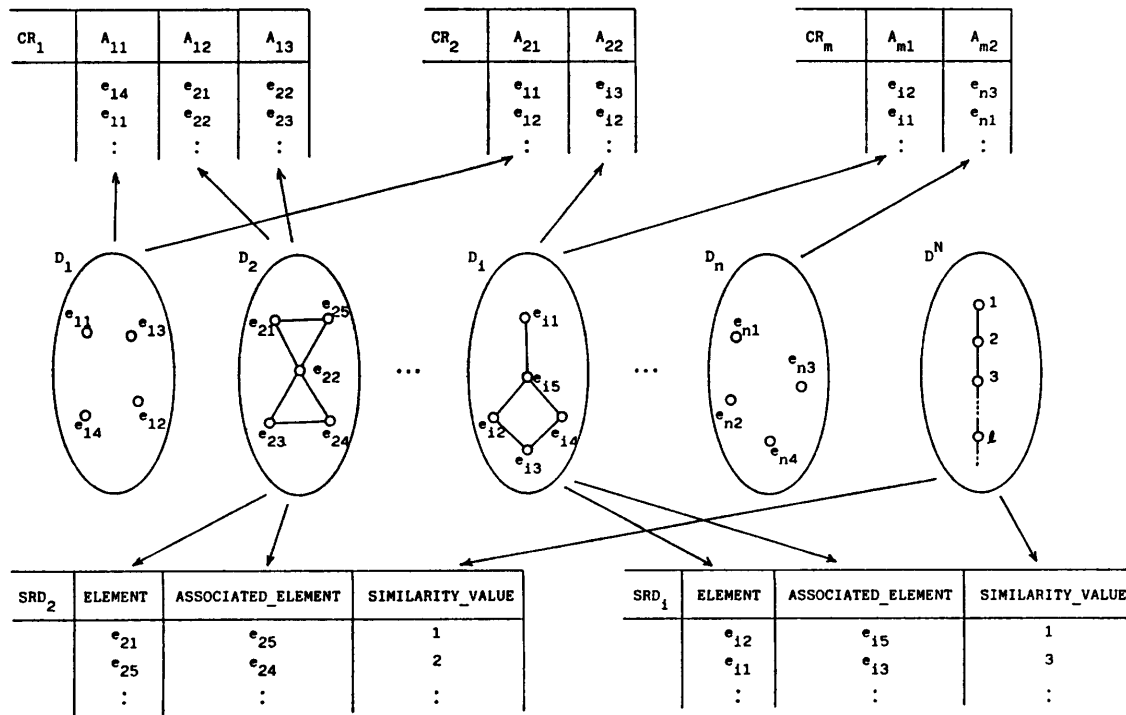


図 2 コンベンショナル・リレーションとセマンティック・リレーションの関連  
Fig. 2 Relationships between conventional and semantic relations.

ョン  $SRD_i$  は次のように定義される。すなわち、

$$SRD_i \subseteq D_i \times D_i \times D^N$$

ここで  $D^N$  は類似度の値を要素としてもつ定義域である。セマンティック・リレーションの各属性値は要素 (ELEMENT), それに関連する要素 (ASSOCIATED-ELEMENT), およびその要素間の類似度 (SIMILARITY-VALUE) を表している。ただし処理の内部機構から、この類似度は要素間の関連の度合いが強いものほど小さい値をとる。

現実世界の事象はコンベンショナル・リレーション CR およびセマンティック・リレーション SRD によって表現される。定義域を介して結びつけられたこの二つの型のリレーションの様子を図 2 に示す。

### 2.3 データ・セマンティックスの規定支援ツール

システムの振舞いを保証するためには、セマンティック・リレーションにユーザの感覚を十分に反映したデータ・セマンティックスを取り込む必要がある。ARES では、そのためにデータ・セマンティックスの規定を支援するツールが用意されている<sup>6)</sup>。このツールにおける手続きの流れを図 3 に示す。

セマンティック・リレーションは、要素間の関連構造を写しとったセマンティック・グラフを介して導かれる。このセマンティック・グラフはグラフィック・ターミナルに表示され、セマンティック・リレーションにおける記述の論理的無矛盾性の感覚的な検証、な

らびに修正・変更を容易にする。

ARES では、セマンティック・グラフを導くための手法として、直接規定法および間接規定法の 2 種類

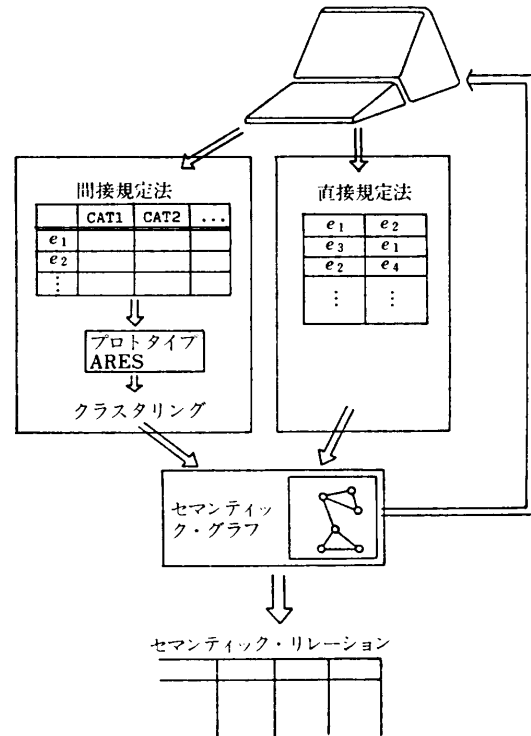


図 3 セマンティック・リレーションの設計プロセス  
Fig. 3 Design process for semantic relation.

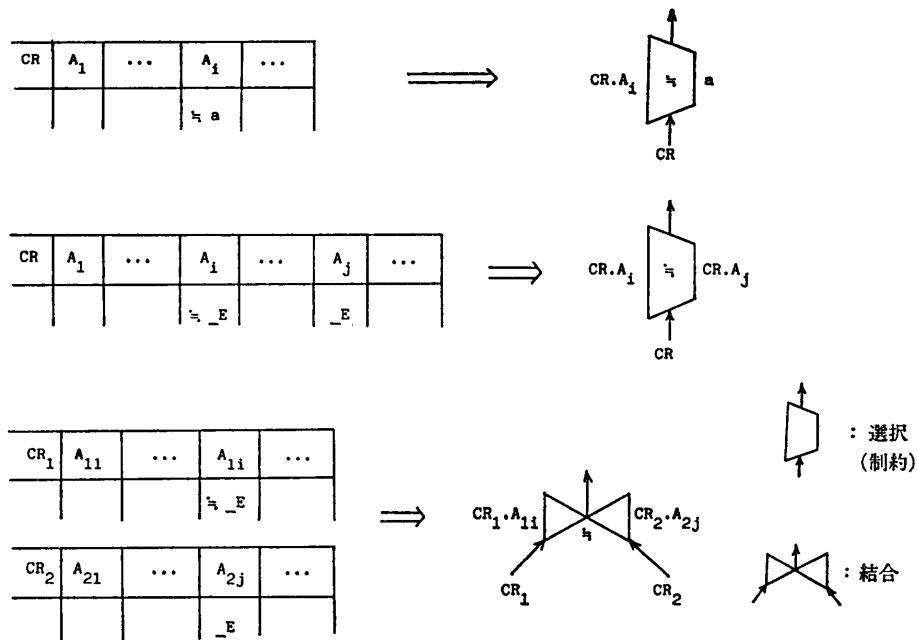


図 4 あいまい性のある条件の拡張関係代数への変換  
Fig. 4 Translation of ambiguous conditions into extended relational algebra.

が用意された。システムはこのいずれかの手法によってセマンティック・グラフを求めたのち、セマンティック・リレーションを構築する。

直接規定法は、関連の強い要素対を直接規定することによってセマンティック・グラフを求める手法である。間接規定法では、まず、個々の要素の特徴をいくつかのカテゴリを用いて多角的に評価し、カテゴリ別に観察された部分的特徴から各要素の全体的な特徴を表現する。次に初期の関連度評価スキーマ (プロトタイプ ARES) に従って要素間のつながりを評価し、関連の強い要素からなるクラスタを見つけ出す。クラスタリングの階層化とそのくり返し実行によって、セマンティック・グラフが導かれる。

定義域の要素数が多い場合には、関連の強い要素対を拾いだしてそれを直接的に記述することは困難である。また、主観が全体の関連度表現に影響を及ぼす度合いも強くなる。したがって、直接規定法の適用は、要素数が比較的少ない場合に限られる。

### 3. セマンティック・スキーマに基づく検索機構

#### 3.1 条件にあいまい性を含む問合せの変換

ARES におけるユーザインタフェースは QBE を基本としているが、筆者らはその上に “ $\approx$ ” という比較演算子を新たに導入した。これは、“...ぐらい”、“...に近い” という条件指定を意味するもので、ユーザの概念に沿ったより自然な問合せの記述を可能にする。

ユーザによって記述された問合せは、実行に先だって関係代数演算の系列に変換される。とくに、条件にあいまい性を含んだ問合せは、適用可能な比較演算子として “ $\approx$ ” が許された選択 (selection), 制約 (restriction), および結合 (join) の各演算によって表現される。このために従来の関係代数演算記法<sup>7)</sup>を若干拡張する。これを図 4 に示す。

比較演算子 “ $\approx$ ” が指定された各関係代数演算では、セマンティック・スキーマに基づいた関連度評価が行われる。次節では、これがどのようにして実行されるかについて述べる。

なお、関係代数レベルでの比較演算子

“ $\approx$ ” の導入は選択, 制約, 結合に限られる。その他の関係代数演算にあいまい性を含んだ条件解釈を許すとセマンティクスに矛盾が生じる。

#### 3.2 拡張された関係代数演算の実行

ARES では、比較演算子 “ $\approx$ ” が記述された選択, 制約, 結合の各関係代数演算の実行のために独自の演算を新たに定義していない。これらの演算も、すべて通常の関係代数の演算系列に吸収される。これにより、通常の関係データベースシステムへの移植が容易となる。

比較演算子に “ $\approx$ ” が想定された関係代数演算は、基本的には、一つの自然結合 (natural join) の追加によって評価される。この結合は、問合せが記述されたコンベンショナル・リレーション  $CR$  と、その条件づけられた属性  $A_i$  の定義域  $D_i$  に対応するセマンティッ

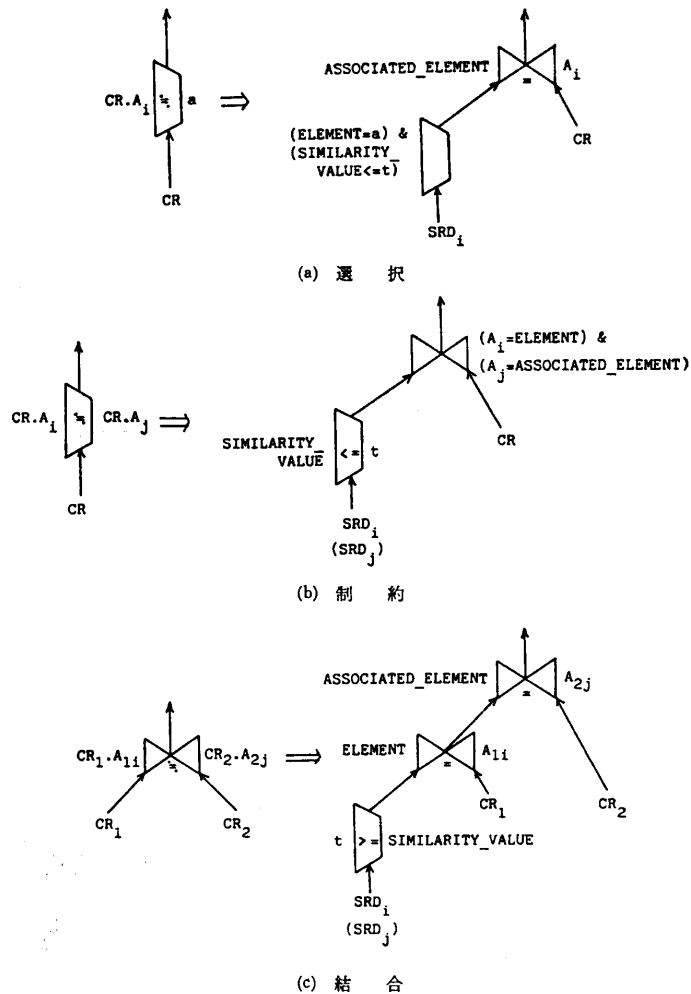


図 5 拡張関係代数の変換

Fig. 5 Translation of extended relational algebra operations.

ク・リレーション  $SRD_i$  との間で行われる。選択、制約、および結合の比較演算子に“ $\equiv$ ”を持ち込んだ関係代数演算が、それぞれどのように通常の関係代数演算の命令列に変換されるかを図5に示す。ここで、条件に対してユーザが感覚的に指定するあいまい性許容度を  $\epsilon$  とする。

なお、条件にあいまい性を含んだ問合せの関係代数演算への変換操作の理解を助けるために、付録Iにその具体例を示す。

### 3.3 出力数制御のメカニズム

条件にあいまい性を含んだ問合せが記述された場合、検索されるデータ数をあらかじめ予想することは一般に困難である。したがって、ARES では、あるしきい値を設定して出力表示するデータ数を制御する機能を付加した。

出力数の制御機能を実現するために、総合類似度 (TOTAL-SIMILARITY) と呼ぶ属性を新たに定義する。これは、システム内部で取り扱われる仮想的な属性としてコンベンショナル・リレーションに付加されるもので、問合せ条件と各組 (tuple) との関連の度合いを表現する。総合類似度の値は各属性値と問合せ条件との類似度 (セマンティック・リレーションの類似度の値) の総和として、各組ごとに個別に算出される。ここでもまた前述の類似度の場合と同様に、総合類似度の値が小さい組ほど、問合せ条件に近いものとなる。

システムは、これらの準備のもとに、最終的に条件を満足した組に対して、総合類似度の値に従って昇順にこれらの並べかえを行う。その結果、順番に希望出力数  $\delta$  まで出力することによって、出力数の制御が行われる。このため、単に出力データ数の上限  $\delta$  を変更する場合には、改めてデータベース全体を検索しなおす必要はなくなる。なお、この出力データ数の制御の過程を具体例を用いて付録IIに示す。

ユーザは検索条件、検索条件に対するあいまい性許容度  $\epsilon$ 、ならびに出力データ数の上限  $\delta$  をインタラクティブに与えることによって問合せを行う。これにより、現実的なデータベース使用環境を保証した上で、ユーザの感覚に沿った問合せが可能となる。

## 4. 結 論

本論文では、効率のよい新しい関連度評価法を示し、これに基づく検索のメカニズムを明らかにした。データベースはコンベンショナル・リレーションとセ

マンティック・リレーションから構築される。要素間の関連度情報は、セマンティック・リレーションとしてコンベンショナル・リレーションとは独立に記述・保持される。このことによって、データ・セマンティックの修正・変更に対する操作性が高められた。

セマンティック・リレーションは、ユーザの感覚を直接反映できるような設計ツールを介して求められる。このツールでは、要素間の関連構造を最初にセマンティック・グラフとして表現し、これに基づいてセマンティック・リレーションが導かれる。また、グラフィック・ターミナル上にこのセマンティック・グラフを表示することによって、類似度表現の論理的な無矛盾性を確認することもできる。

次に、筆者らは、条件にあいまい性を含む問合せがセマンティック・リレーションとコンベンショナル・リレーションとの結合操作で評価できることを示した。これにより、一般の関係データベース環境において現実性を失うことなく、条件にあいまい性を含む問合せ機能を実現することができるようになった。

さらに、ARES では出力数の制御機能が付加されており、条件にあいまい性を伴った問合せを受理する能力とあわせて、より使いやすいデータベース環境が用意された。これらすべての問合せ操作はグラフィック・ターミナルを介して視覚的に行われる。

なお、自動車の物件案内を対象とした実験システムがミニコンピュータ HP-1000 上で稼働中である。そのプログラムは、フォートラン約 10,000 ステップで記述されている。

**謝辞** 本研究の遂行過程において有益な討論をいただいた上林憲行博士、システム開発に協力された蔵重衆治、福原勝彦、および宮迫良己氏に感謝する。

## 参 考 文 献

- 1) Ichikawa, T., Sakamura, K. and Aiso, H.: ARES—A Memory, Capable of Associating Stored Information through Relevancy Estimation, *Proc. AFIPS NCC*, Vol. 46, pp. 947-954 (1977).
- 2) Ichikawa, T., Kikuno, T., Kamibayashi, N. and Hirakawa, M.: ARES in Relational Database Environments, *Proc. COMPSAC '80*, pp. 557-561 (1980).
- 3) 平川正人, 上林憲行, 市川忠男: あいまいな問合せを受理する情報検索システム, *信学論 (D)*, Vol. J65-D, No. 1, pp. 64-71 (1982).
- 4) Zloof, M. M.: Query-by-Example: A Data

CAR	CAR_NAME	BODY_COLOR	PRICE
	P. SUNNY	P. RED	P. <= 200

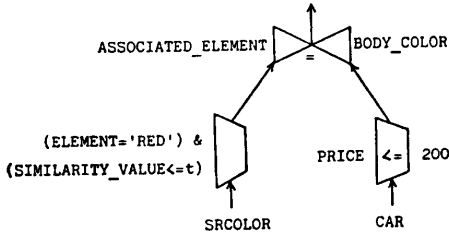
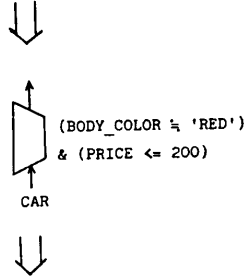


図 A.1 問合せ 1 の変換手順  
Fig. A.1 Total flow of translation—Query 1.

Base Language, *IBM Syst. J.*, Vol. 16, No. 4, pp. 324-343 (1977).

5) Codd, E.F.: A Relational Model of Data for Large Shared Data Banks, *Comm. ACM*, Vol. 13, No. 6, pp. 377-387 (1970).

6) 清水達男, 平川正人, 市川忠男, 藤永幸富: ARES におけるデータ・セマンティックスの規定支援ツール, 電子通信学会オートマトンと言語研究会資料, AL 82-60 (1982).

7) Smith, J.M. and Chang, P.Y.: Optimizing the Performance of a Relational Algebra Database Interface, *Comm. ACM*, Vol. 18, No. 10, pp. 568-579 (1975).

付録 I (問合せの変換手順)

次の三つの問合せが通常の関係代数演算の系列に変換される様子を図 A.1 から図 A.3 に示す。

問合せ 1: 価格が 200 万円以下で, ボディ・カラーが赤に似ている自動車を選べ。

CAR	CAR_NAME	BODY_COLOR	SEAT_COLOR
	XXX	P. RED	
	P. SUNNY		P. RED

CAR	CAR_NAME	BODY_COLOR	SEAT_COLOR
	P. SUNNY	P. RED	P. RED

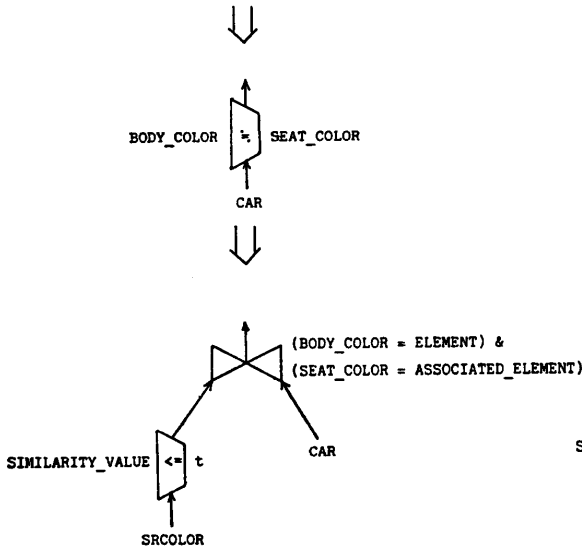


図 A.2 問合せ 2 の変換手順  
Fig. A.2 Total flow of translation—Query 2.

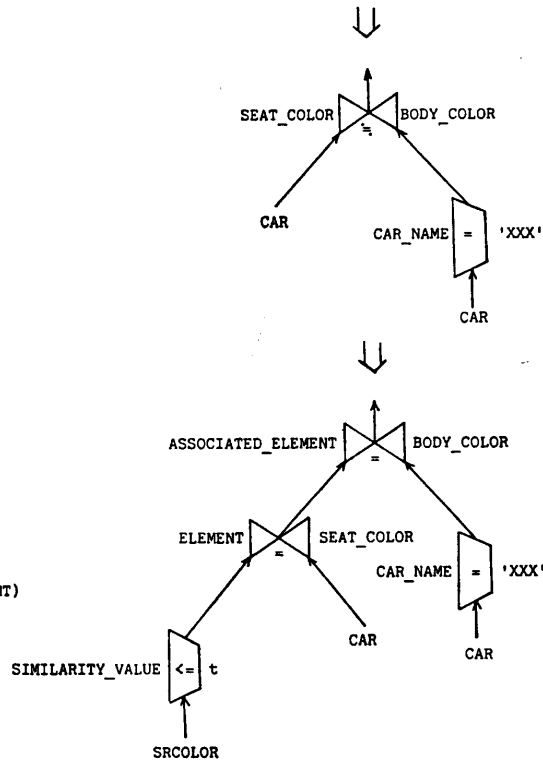


図 A.3 問合せ 3 の変換手順  
Fig. A.3 Total flow of translation—Query 3.

表 A.1 サンプル・データベース  
Table A.1 Sample databases.

CAR	CAR_NAME	BODY_COLOR	TYPE
	SPRINTER	BLUE	COUPE
	GEMINI	RED	COUPE
	CARINA	YELLOW	COUPE
	CIVIC	YELLOW	HATCHBACK
	FAMILIA	WHITE	HATCHBACK
	SKYLINE	RED	HATCHBACK
	BLUEBIRD	BLUE	SEDAN
	LANCER	BLACK	SEDAN
	LEONE	WINE_RED	SEDAN

SRTYPE	ELEMENT	ASSOCIATED_ELEMENT	SIMILARITY_VALUE
	COUPE	COUPE	0
	COUPE	HATCHBACK	1
	COUPE	LIFTBACK	1
	COUPE	HARDTOP	2
	COUPE	SEDAN	3
	COUPE	WAGON	4
	∴	∴	∴

SRCOLOR	ELEMENT	ASSOCIATED_ELEMENT	SIMILARITY_VALUE
	RED	RED	0
	RED	WINE_RED	1
	RED	ORANGE	1
	RED	YELLOW	2
	RED	GREEN	2
	RED	WHITE	3
	RED	BLUE	4
	RED	BLACK	5
	∴	∴	∴

表 A.2 問合せの QBE 表現  
Table A.2 An example of query (Query-by-Example).

CAR	CAR_NAME	BODY_COLOR	TYPE
	P. SUNNY	P. ≠RED	P. ≠COUPE

表 A.3 出力数制御の過程  
Table A.3 Process regulating number of outputs.  
CAR\*: CAR と SRTYPE の結合の結果

CAR_NAME	BODY_COLOR	TYPE	TOTAL_SIMILARITY
SPRINTER	BLUE	COUPE	0
GEMINI	RED	COUPE	0
CARINA	YELLOW	COUPE	0
CIVIC	YELLOW	HATCHBACK	1
FAMILIA	WHITE	HATCHBACK	1
SKYLINE	RED	HATCHBACK	1
BLUEBIRD	BLUE	SEDAN	3
LANCER	BLACK	SEDAN	3
LEONE	WINE_RED	SEDAN	3

CAR\*\*: CAR\* と SRCOLOR の結合の結果

CAR_NAME	BODY_COLOR	TYPE	TOTAL_SIMILARITY
GEMINI	RED	COUPE	0
CARINA	YELLOW	COUPE	2
CIVIC	YELLOW	HATCHBACK	3
FAMILIA	WHITE	HATCHBACK	4
SKYLINE	RED	HATCHBACK	1
LEONE	WINE_RED	SEDAN	4

CAR\*\*\*: CAR\*\* を総合類似度で並べかえた結果

CAR_NAME	BODY_COLOR	TYPE	TOTAL_SIMILARITY
GEMINI	RED	COUPE	0
SKYLINE	RED	HATCHBACK	1
CARINA	YELLOW	COUPE	2
CIVIC	YELLOW	HATCHBACK	3
FAMILIA	WHITE	HATCHBACK	4
LEONE	WINE_RED	SEDAN	4

出力:

CAR_NAME	BODY_COLOR	TYPE
GEMINI	RED	COUPE
SKYLINE	RED	HATCHBACK
CARINA	YELLOW	COUPE
CIVIC	YELLOW	HATCHBACK

問合せ 2: ボディ・カラーとシート・カラーのよく似た自動車を選べ.

問合せ 3: 車名が XXX のボディ・カラーとよく似たシート・カラーをもつ自動車を選べ.

#### 付録II (出力数の制御メカニズム)

サンプル・データベースとして, 表 A.1 に示す三つのリレーションを用いる.

問合せ: ボディ・カラーが赤に近く, 型がクーペによく似た自動車を選べ.

ここで, 検索条件に対するあいまい性許容度  $\epsilon$  はともに 3 とし, 希望出力数  $\delta$  を 4 とする.

QBE による問合せの記述例を表 A.2 に示す. また, 表 A.3 に出力数制御の過程を示す.

(昭和 57 年 11 月 17 日受付)

(昭和 58 年 1 月 17 日採録)