

# ARESにおける関係演算の機能的拡張

平川正人 清水達男 市川忠男  
(広島大学工学部)

## 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。これにより、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出すことができる。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に対してある程度の幅をもった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ーフェイスとしてQBE (Query-By-Example)<sup>(4)</sup>を採用し、その上で“…に近い”という条件指定をおこなうための比較演算子“≍”が新たに導入されている。さらに“≍”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、感覚的にグラフィック・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あきらかにする。以後ARESというときは、単なる関連スキーマではなくて、関連度評価の機能を取り入れた情報検索システム (As-

sociative Information Retrieval System) 全体をさすものとする。

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2つの型のリレーションを持っている。セマンティック・リレーションは定義域 (domain) の個々の要素間の関連を表現しており、これはユーザの要求を反映させるための設計ツールを通して決定される。

検索条件にあいまい性が付加された場合には、まず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・リレーションに対して選択 (selection) をおこなう。次に、この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 (join) をおこなう。このようにして、検索条件にあいまい性が付加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むことになり、確立された従来の技法で実行可能となる。さらに、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り、セマンティック・リレーションの意義をユーザに解放することができ、かつ、定義されたリレーションに対するシステムの追従も容易となる。

また、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関連度の大小に従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との関連度の強いも

のから順に出力される。

以下では、セマンティック・スキーマとその設計ツールを説明したのちに、そのスキーマに基づいたARIESの検索処理のメカニズムを詳説する。

## 2 データ・セマンティックスキーマとその設計ツール

### 2.1 セマンティックスキーマについて

E. F. Coddによつて提案されたリレーショナル・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 $n$ 項関係として定義している。しかしこの定義域という概念は、個々の要素間のセマンティックな関連を陽に記述したものではない。筆者らは、セマンティックな関連を積極的に取り入れるために、要素間の関連の様子にしたがつて定義域を次の3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からなる定義域

### 2.2 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

コンベンショナル・リレーションは、現実世界の事象を表現するもので、従来のリレーションに対応する。リレーション $CR$ は、 $n$ 個の定義域 $D_i$  ( $1 \leq i \leq n$ )の直積 $D_1 \times D_2 \times \dots \times D_n$ の部分集合として定義される。

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の度合いを規定するものであり、要素間に関連づけがなされる定義域(タイプⅡ, タイプⅢ)ごとに定義される。定義域 $D_i$ に対応するセマンティック・リレーション $SR_{D_i}$ は、 $D_i \times D_i \times D_i$  ( $D_i$ は、要素として類似度の値をもつ定義域)の部分

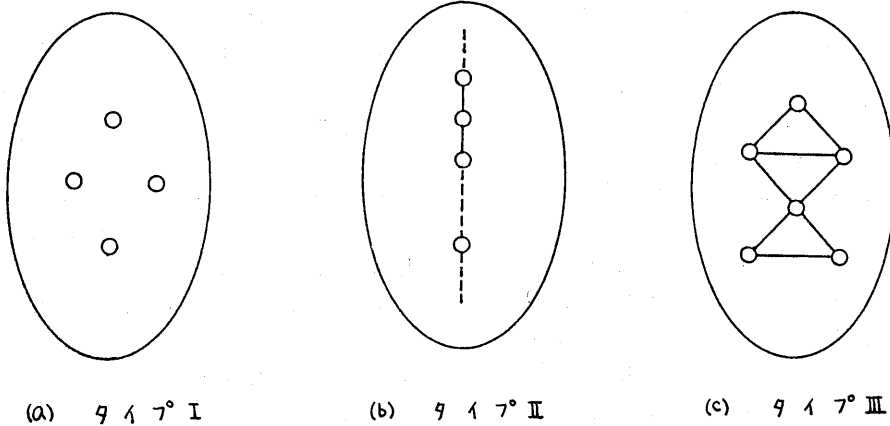


図1. 定義域の分類

分集合となり、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) からなる。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程小さい値をとる。

図2にARESによって提供される2つの型のリレーション、すなわちコンベンショナル・リレーションとセマンティック・リレーションとの定義域を介しての結びつきを示す。

### 2.3 セマンティック・リレーションの設計ツール

各々の定義域について要素間のセマンティックな関連をセマンティック・

リレーションとして記述する場合、ユーザの感覚を十分に反映し得るような支援ツールを用意する必要がある。そこで、セマンティック・リレーションを規定し、かつその論理的な無矛盾性の感覚的な検証を容易にするためにセマンティック・グラフを導入する。なお、セマンティック・グラフ上の節点は各要素を、枝は要素間の直接的な関連を示すものとする。

ARESでは、このセマンティック・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・グラフを求めた後、最短経路アルゴリズムを用いて2節点間の距離を計算する。距離は節点間の枝の数に対応する。これによって、互いに関連をもつ要素対とその間に規定される距離が求めら

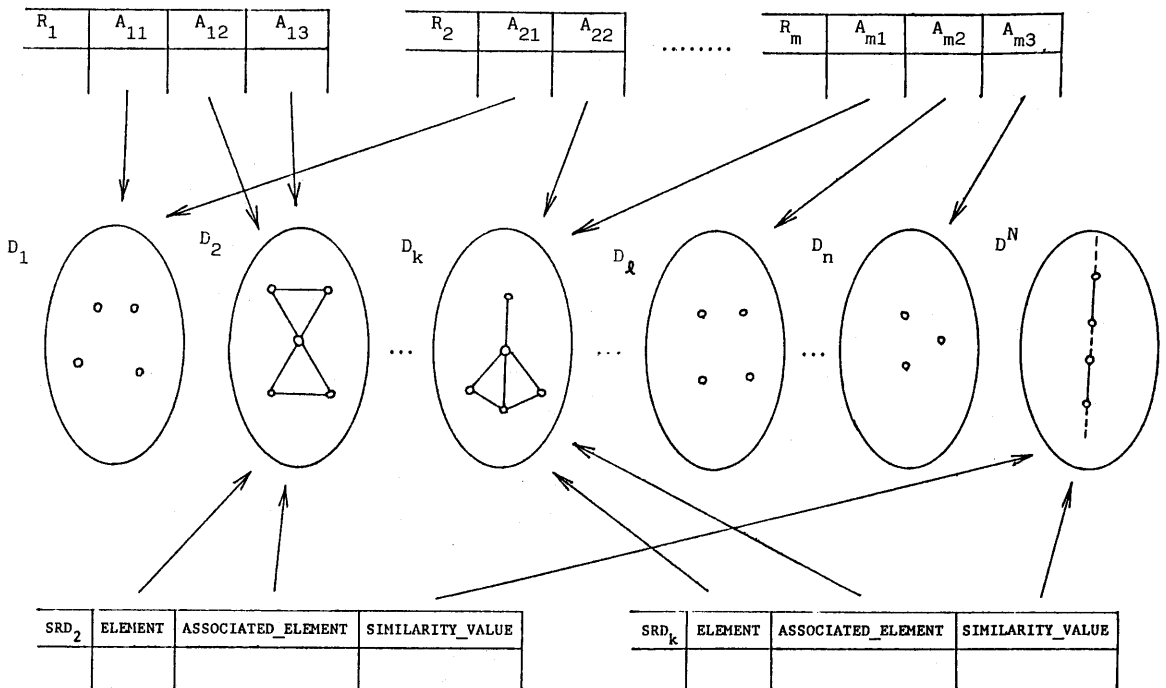


図2. コンベンショナル・リレーションとセマンティック・リレーションの関連

れ、セマンティック・リレーションが導かれる。

直接規定法は、関連の強い要素対をユーザに直接記述させるものである。セマンティック・グラフは、記述された要素対に対応する節点間を線で結ぶことにより、導かれる。

間接規定法は、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。次に初期の符号間距離に基づく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、関連の強い要素のクラス9を見つけ出す。クラス9リングの

階増化とこの繰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、関連の強い要素対だけを拾い出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。従って、直接規定法の適用は、要素数が比較的小さい場合に限られる。

セマンティック・リレーションの設計プロセスを図3に示す。セマンティック・グラフをグラフィック・ターミナル上に表示することによって、セマンティック・リレーションの修正と変更が、さらには論理的な無矛盾性の検証が容易になる。

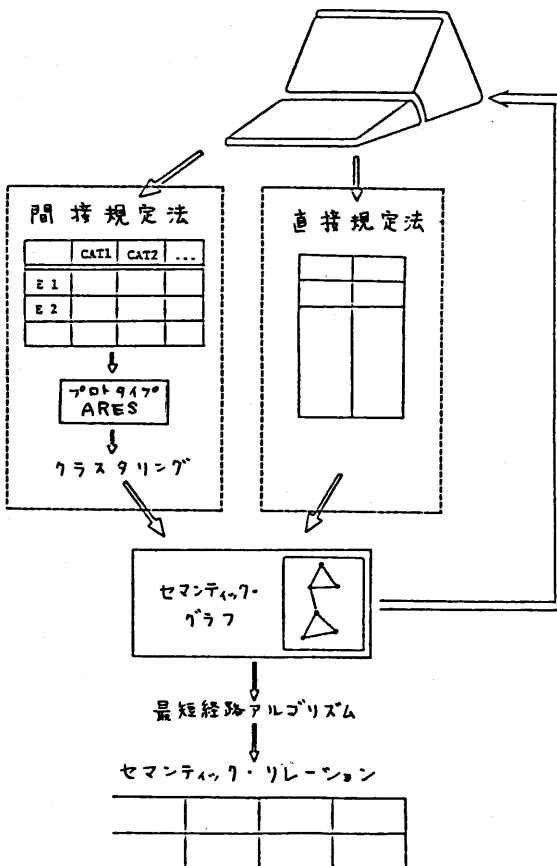


図3. セマンティック・リレーションの設計プロセス

### 3 セマンティック・スキーマに基づく検索の実行

#### 3.1 条件にあまい性を含む問合せの変換

ARESにおける問合せはQBEを基本としているが、筆者らはその上に“≒”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、システムに関連度評価機能の実行を要求する。

ユーザによって記述された問合せは、実行に先だてて関係代数演算の系列に変換されるが、ここで、選択(selection)、制約(restriction)、および結合(join)に適用可能な比較演算子として“≒”を許すためには従来の関係代数演算記法を若干拡張する必要が生じる。これを図4に示す。

次節では、比較演算子≒を含む問合せ表現をいかにして通常の関係代数演算に変換するかについて述べる。

なお、関係代数レベルでの比較演算

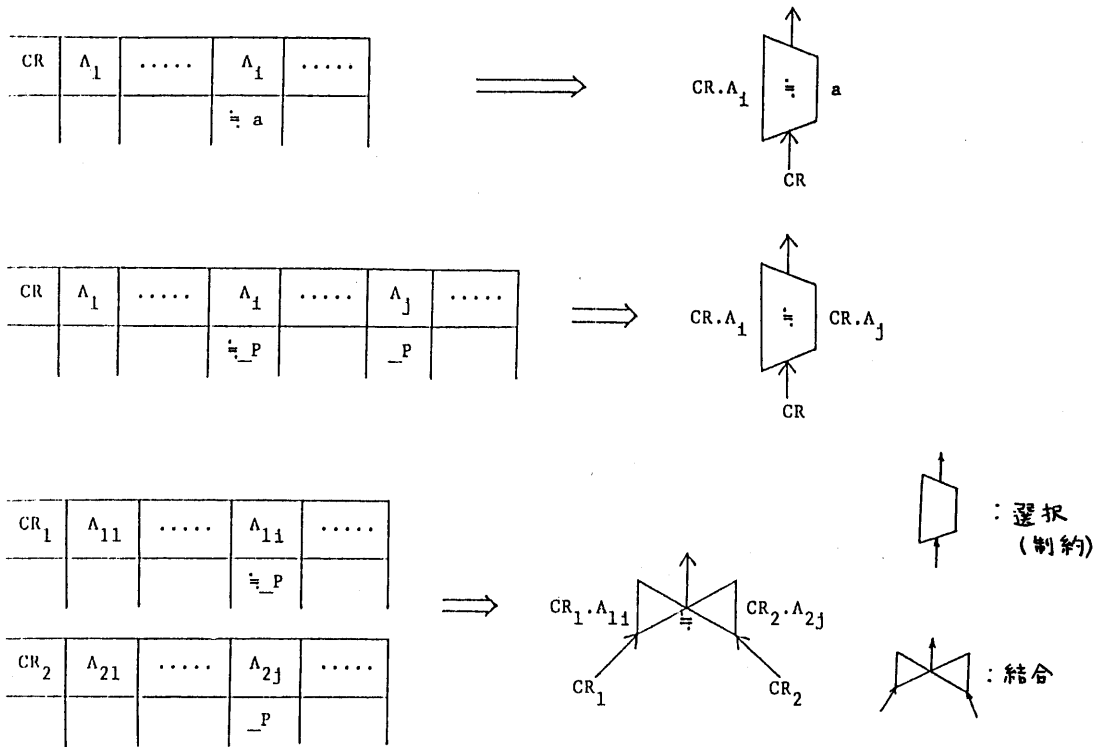


図4. あいまい性のある条件の拡張関係代数への変換<sup>(6)</sup>

子“ $\equiv$ ”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ックスの矛盾が生じる。

### 3.2 従来の関係代数演算への変換

3.1節で述べたように、選択、制約、結合の関係代数演算では仮想的に比較演算子“ $\equiv$ ”が許容される。しかしながら、ARESではその実行のために独自の演算を新たに定義すること避けている。比較演算子として意味的に“ $\equiv$ ”が許される関係代数演算も、すべて通常の演算系列に吸収される。これは、通常の関係データベースシステムとのマッチングを考慮したことによるものである。

比較演算子に“ $\equiv$ ”が想定される関

係代数演算は、基本的には、1つの自然結合(natural join)の追加によって評価される。その際、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、条件づけられた属性A<sub>i</sub>の定義域D<sub>i</sub>に対応するセマンティック・リレーションSR<sub>D<sub>i</sub></sub>との間で行なわれる。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したあいまい性許容度を $\alpha$ 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1にその具体例を示す。

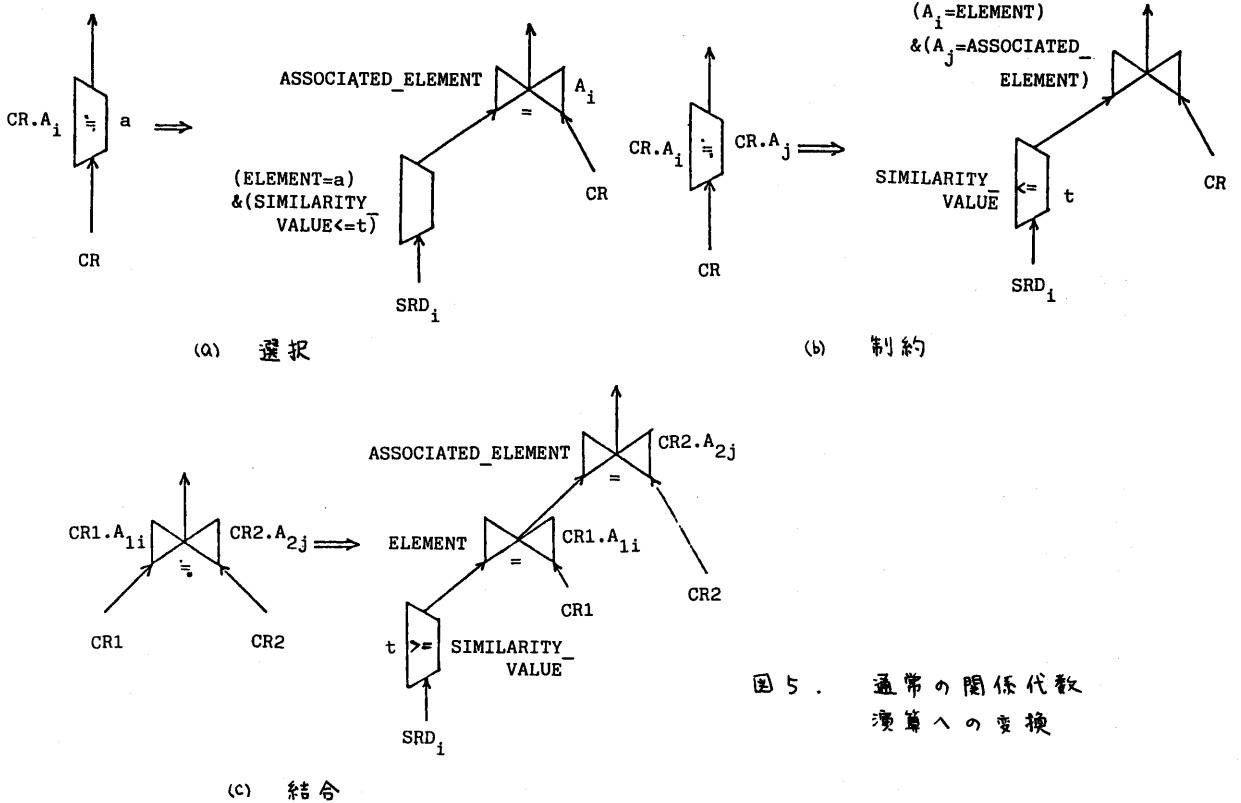


図5. 通常の関係代数演算への変換

### 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を可能にした場合、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規制できるようにしている。すなわち、ユーザは検索条件、検索条件に対するあいまい性許容度 $\alpha$ 、ならびに出力データ数の上限 $\beta$ をインタラクティブに与えることにより、問合せをおこなう。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。

さて、出力数を制限するために、総合類似度と呼ぶ属性を新たに定義する。これは、システム内部で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、問合せと関係内の組(tuple)との関連の度合いを表現する。総合類似度は、各組ごとに評価された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)の総和である。ここでもまた前述の類似度と同様に、総合類似度の値が小さい組程、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この総合類似度の値の順にソーティングをおこなう。値の小さい順に出力数 $\beta$ まで出力することによって、出力数の制御がおこなわれる。このため、単に出力データ数の上限 $\beta$ を変更する場合に、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プロセスを具体例を用いて付録2に示す。

## 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あきらかにした。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される。このようにリレーションを2つの型に区別し、データセマンティックスを従来のリレーションから独立させることにより、セマンティックスの修正・変更に対する操作性が高められた。また、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。セマンティック・リレーションの設計手順は、セマンティック・グラフの構築、ならびに最短経路の計算という2つのステップに分けられる。また、グラフ・フィック・ターミナル上にセマンティック・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、一般のリレーショナル・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の実現が可能となった。

さらに、ARESでは出力数制御機能が付加されており、条件に対するあいまい性の付与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。以上の問合せ操作はすべてグラフィック・ターミナル上で視覚的に行なわれる。

なお、自動車の物件検索を対象とした実験システムが、現在筆者らの所属する広島大学工学部情報システム研究

室のミニコンピュータ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昭和56年度本学情報システム研究室在籍者、蔵重衆治、福原勝彦、および宇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 Base Language," IBM Systems J., Vol. 16, No. 4, pp.845-853, 1977.
- (5) Codd, E. F., "A Relational Model of Data for Large Shared Data Banks," Comm. of ACM, Vol. 13, No. 6, pp.377-387, June 1970.
- (6) Smith, J. M., and Chang, P. Y., "Optimizing the Performance of a Relational Algebra Database Interface," Comm. of ACM, Vol. 18, No. 10, pp.568-579, Oct. 1975.

付録 1 (変換の手順)

図 A.1 から図 A.3 に 3 つの問合せの変換手順を示す。

問合せ 1 : 価格が 200 万円以下で  
ボデーカラーが赤に似  
た色の自動車を選べ。

問合せ 2 : ボデーカラーとシート  
カラーのよく似た自動  
車を選べ。

CAR	CAR_NAME	BODY_COLOR	PRICE
	P.	P. <sub>E</sub> = RED	P.<=200

CAR	CAR_NAME	BODY_COLOR	SEAT_COLOR
	P.	P. <sub>E</sub> = E	P. <sub>E</sub> = E

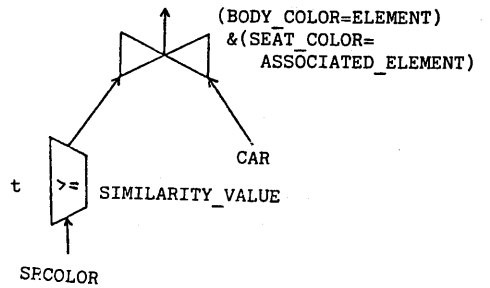
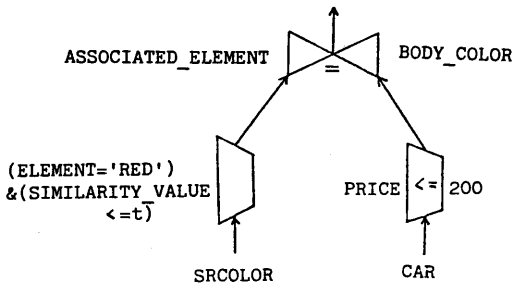
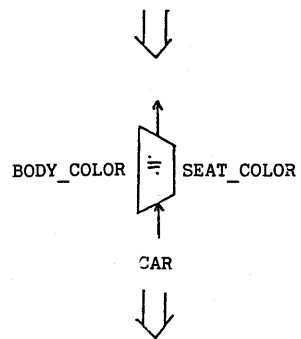
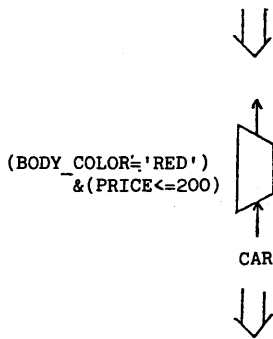


図 A.1 問合せ 1 の変換手順

図 A.2 問合せ 2 の変換手順



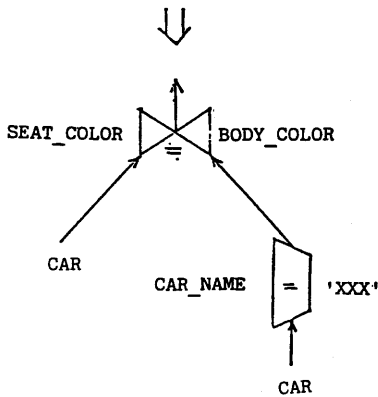
問合せ3: 車名がXXXのボディカラーとよく似たシートカラーの自動車を選べ。

付録2 (出力数制御)

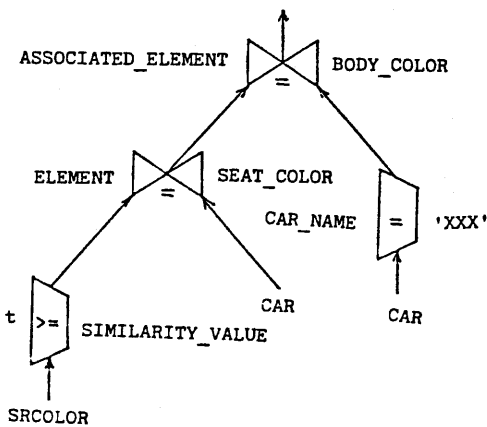
サンプル・データベースは、表A.1に示した3つのリレーションである。

CAR	CAR_NAME	BODY_COLOR	SEAT_COLOR
	XXX	P_E	
	P		P_E

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
	.	.	.
	.	.	.

図 A.3 問合せ3の変換手順

表 A.1 サンプル・データベース

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。  
あいまい性許容度はともに3とし、希望出力数は4とする。

表 A.2 にこの問合せの Q B E 形式による記述を示し、また、表 A.3 に出力制御のプロセスを示す。

CAR	CAR_NAME	BODY_COLOR	TYPE
	P.	P. = RED	P. = TYPE

表 A.2 | 問合せの Q B E 表現

CAR\* : CAR と S R T Y P E 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 S R C O L O R 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\*\* を総合類似度でソート

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

表 A.3 | 出力数制御のプロセス

出力:

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