リレーショナル・データベースの質問言語としての多層論理式と検索手続への変換アルゴリズム†

宇田川 佳久†† 大須賀 節雄††

本論文は多層論理式によるリレーショナル・データベースに対する質問の記述方法と，その論理式を検索手順に変換するアルゴリズムについて論じたものである．多層論理式は構造をもった対象に関する記述ができるように many sorted logic を拡張した形式言語である．多層論理式のデータベース質問言語への応用に関しては，その機能の一部である要素を値とする変数だけでなく集合や冪集合を値とする変数も陽に記述することができることを利用する．データベース質問言語としての多層論理式の特徴を以下に示す．(1)従来の論理型の質問言語のように特殊な記号や手続的な要素を導入することなく集合演算，さらにそれらのネストを含む質問を表現できる．(2)多層論理式によって計算機内に記憶されているリレーションから仮想的なリレーションを定義することができ質問に対する推論処理が行えるので，計算機内に記憶されているリレーションに対する質問と仮想的なリレーションに対する質問を同じ形式で表現できる．(3)多層論理式は自然言語型や例題型の質問言語の目的言語になりうる．

1. はじめに

リレーショナル・データベースに対する non-programmer のための質問言語は述語論理型，射像型，例題型，関係代数型，自然言語型に大別される．このうち述語論理型は(1)質問を非手続的に正確に記述できる，(2)自然言語型や例題型のような直観的に理解しやすい質問言語の目的言語となりうる，(3)論理式によってデータベースの視野 (view) を定義することができるので計算機内に記憶されているリレーション(ベース・リレーション)に対する質問だけでなく視野を通しての質問も行える，といった non-programmer 向け言語として適した特徴を備えている．

一方，通常の一階述語論理ではある変数に関して集約された集合に対する演算を含むような記述を行うことはむずかしい．そのために述語論理型の言語では DEDUCE, DEDUCE 2 (Chang, 1978)[3]や LQL (古川，1978)[12] のように特別な記号や手続的な要素を組み込むとか ALPHA (Codd, 1971)[4] のように必要な集合演算を関数として作り付けにするなどによって記述力の拡張が行われてきた．しかし，このような拡張は記述力に関する問題の十分な解決にならないだけでなく，論理型言語が備えていた上記の特徴さえも失いかねない．

ここで提案する多層論理は構造をもった対象に関する記述を行うことができるように many sorted logic (MSL) を拡張したものである．MSL では要素と領域の関係を陽に表現できるのに対し，多層論理では要素と領域の関係のみならず要素の集合や冪集合などと領域との関係も陽に表現することができる．以下ではこの機能を利用する．なお，多層論理における推論アルゴリズムは MSL の推論アルゴリズムをわずかに修正するだけで実現できるものである．

本稿は多層論理式によるリレーショナル・データベースへの質問の記述法と，それらの質問をデータベース検索手順に変換するアルゴリズムについて論じたものである．データベース質問言語としての多層論理式の特徴を以下に示す．

(1) 従来の論理型の質問言語のように特殊な記号や手続的な要素を導入することなく集合演算のネストを含む問合せを表現できる．

(2) 多層論理式によってベース・リレーションから仮想的なリレーション（視野）を定義することができ質問に対する推論処理が行えるので，ベース・リレーションに対する質問と仮想的なリレーションに対する質問を同じ形式で表現できる．

(3) 多層論理式は自然言語型や例題型の質問言語の目的言語になりうる．

2章では多層論理式について概論し，データベースへの質問の記述方法を例を用いて述べる．通常，仮想的なリレーションを導入した場合，それを含む質問を

† The Multi-layer Logic as a Relational Database Query Language and Its Reduction Algorithm to Relational Procedures by YOSHIHISA UDAGAWA and SETSUO OHSUGA (Institute of Interdisciplinary Research Faculty of Engineering, University of Tokyo).

†† 東京大学工学部境界領域研究施設

直接データベース検索手続に変換することはできない．そのために仮想的なリレーションをベース・リレーションによって表現しなおす必要がある．多層論理ではこれは推論処理の一部である．多層論理式の一般的な推論処理については別に述べることとし，ここでは仮想的なリレーションの処理に必要な部分のみを3章で述べる．4章では仮想的なリレーションを含まない多層論理式をデータベース検索手続に変換する方法について論ずる．5章ではまとめとして今後の研究課題について述べる．

2. 多層論理式とリレーショナル・データベースへの質問の記述

2.1 多層論理式について

many sorted logic (MSL) は記述対象が属する集合を複数用意し，すべての変数または定数はこれらのいずれか一つを領域とする述語論理である．MSL では要素と領域の関係が明示されているので，領域をリレーションの属性と関連づけることによってデータベースに対する質問を通常の1階述語論理よりも自然に記述することができた[7]．しかし，MSL では集合演算を含む質問を特別な記法を導入することなく記述することは困難である．このためわれわれは MSL を拡張した多層論理を提案した．多層論理式の基本的な考え方は構造をもった対象を論理式内に含めることができるように MSL を拡張することである．このような構造表現の一つとして MSL の領域の空集合を除く任意の部分集合のクラスの上で定義される変数を含む構造が含まれる．この場合，これらの変数の領域を明示する必要があるが，これは元の領域の空集合を除く任意の部分集合を要素とする集合，すなわち元の集合の冪集合から空集合を除いたものとすればよい．本稿ではSを空でない任意の集合とするとき$*S$によってSの冪集合から空集合を除いたものを表すことにする．たとえば S' を自然数のある有限集合とし $G(x, y)$ をxの値がy以上であるとき真，そうでないとき偽である述語とする．このとき，値が 50 以上の S' の部分集合Xに対しある集合演算 SO を行いたいとする．これは多層論理式によって $(^{\exists}X/*S')(^{\forall}x/X)G(x,50)\&SO(X)$ と表現される．& 記号より左側の論理式は S' の部分集合でそのすべての値が $G(x,50)$ を真とするXの存在を定義している．こうして定義されたXを含む集合演算を & 記号に続いて記述することができる．

上記の考え方はさらに一般化され要素，要素の集合のみならず要素の冪集合などを値とする変数を論理式内に陽に含めることができる．$x_0, x_1, \cdots, x_n$ $(n\geqq0)$ がそれぞれ $S, *S, \cdots, *_n\cdots*_1S$ 上の変数であるとき $(Q_nx_n/*_n\cdots*_1S)(Q_{n-1}x_{n-1}/x_n)\cdots(Q_0x_0/x_1)$ は多層論理式の限量子である．ただし $Q_i(0\leqq i\leqq n)$ は量記号$\forall$または$\exists$のいずれかである．$n=0$ のとき限量子は (Q_0x_0/S) であり MSL の限量子と等しい．$*_n\cdots*_1S$ をレベル $n+1$ の集合とよぶ．ただし，レベル0の集合とは記述の対象となる要素を表し，レベル0の集合の要素（レベル0の要素）とはその要素自身を表すものとする．

通常の述語論理と同様に多層論理でも記述の対象とする問題に応じて解釈の定まっている述語を導入する．このような述語には解釈がリレーションによって与えられているモデル述語[7]と解釈がアルゴリズムによって与えられる述語定数がある．$A_i(1\leqq i\leqq n)$ を空でない集合とし x_i を A_i 上の変数とする．述語 $R(x_1, \cdots, x_n)$ がリレーション $R'(A_1, \cdots, A_n)$ のモデル述語であるとは，n組 $\langle t_1, \cdots, t_n\rangle$ が R' の要素であるとき $R(t_1, \cdots, t_n)$ が真，そうでないとき偽となる述語である．すなわちリレーションを内包的に表現したものがそのリレーションに対するモデル述語である．とくに区別する必要がない限り，リレーションとモデル述語とを同義に用いる．

つぎに以下の説明で必要となる関数を示す．

avg(X): X はレベル $m(m\geqq1)$ の集合でそのベースとなっているレベル0の要素は四則演算が定義されているものとする．このとき avg(X) は X の平均値を求める．なお $m\geqq2$ のとき avg(X) はレベル1の各集合に対する平均値を要素とするレベル $m-1$ の集合である．

count(X): X はレベル $m(m\geqq1)$ の集合であるとする．このとき count(X) は集合 X の要素数を求める．なお $m\geqq2$ のとき count(X) はレベル1の各集合に対する要素数を要素とするレベル $m-1$ の集合である．

sub(X, Y): X, Y はともにレベル $m(m\geqq0)$ の集合で，それらのベースとなっているレベル0の各要素は減算が定義されており，かつXとYのレベル0の要素の対応関係がついているものとする．このとき sub(X, Y) は X のレベル0の要素とそれに対応するYのレベル0の要素との差を求め，それをベースとするレベルmの集合である．

たとえば，

avg (count ({{4, 8}, {3, 1, 6}})

=avg ({count ({4, 8}), count ({3, 1, 6})})

=avg ({2, 3})=2.5

また

sub ({5, 7}, {−2, 4})

={sub (5, −2), sub (7, 4)}

={7, 3}

である．

つぎに以下の説明で必要となる述語定数を示す．

GE(X, Y), LT(X, Y) と EQ(X, Y): X, Y はともにレベル $m(m\geqq 0)$ の集合で，それらのベースとなっているレベル0の要素は順序付けられており X と Y の要素間の対応関係もわかっているものとする．このとき GE(X, Y), LT(X, Y), EQ(X, Y) は X のすべてのレベル0の要素 x に対応する Y のレベル0の要素 y に対し $x\geqq y$, $x<y$, $x=y$ が成り立つとき真，他のとき偽である．

LET(X, Y): X がレベル $m(m\geqq 0)$ の集合を値とする変数で Y がレベル m の集合とする．このとき X, Y のレベル0の要素が代入可能であり Y を X に代入したとき真，他のとき偽である．

SSET(X, Y): X, Y はともにレベル $m(m\geqq 1)$ の集合である．このとき SSET(X, Y) は $X\subseteq Y$ であるとき真，他のとき偽である．

たとえば X を整数のレベル1の集合を値とする変数であるとするとき GE ({8, 3}, {7, 3}), LET (X, {6, 4}), SSET ({b}, {a, b}) はいずれも真である．

2.2 多層論理式による質問の記述

本節では多層論理式によるリレーショナル・データベースに対する質問とすでに定義されているリレーションから仮想的なリレーションを定義する方法を例を用いながら述べる．**図1**に以下の説明で用いるデータベースを示した．リレーション LYP, LUD はそれぞれある年度 (YEAR：西暦) における土地 (LAND) と価格 (PRIC：万円/m²) の関係，および土地の用途 (USAG) と都心からの距離 (DIST：km) の関係を表している．以下に，このデータベースに対する質問とそれを多層論理式によって表現したものを示す．なお，"^" の付いた変数は答えとして求められている変数であることを示す．"#" に続いて書かれた数字列は数値を表し，' ' で囲まれた文字列は文字定数を表す．

LYP	LAND	YEAR	PRIC
	A1	1977	50
	A1	1979	55
	A1	1981	58
	A2	1979	80
	A2	1981	86
	B1	1979	63
	C1	1979	22
	C1	1981	25
	D1	1981	40
	D2	1981	36

(a)

LUD	LAND	USAG	DIST
	A1	a	40
	A2	a	25
	B1	b	30
	B1	c	30
	C1	c	10
	D1	d	20
	D2	d	30

(b)

図1 リレーショナル・データベースの例
Fig. 1 An example relational database.

[質問 1] 用途が b の土地とその土地の都心からの距離を求めよ．

Q 1: ($^\exists l$^/LAND)($^\exists d$^/DIST)[LUD(l, 'b'/USAG, d)].

[質問 2] 都心からの距離が 30 km 以内で用途が c 以外の土地を求めよ．

Q 2: ($^\exists l$^/LAND)($^\exists u$/USAG)($^\exists d$/DIST)[LUD{l, u, d)& LT(d, #30/DIST)&∼EQ(u, 'c'/USAG)].

[質問 3] 用途が a，都心からの距離が 40 km 以内で 1981 年度における価格が 60 万円/m² 以下の土地を求めよ．

Q 3: ($^\exists l$^/LAND) ($^\exists p$/PRIC) ($^\exists d$/DIST) [LYP(l, #1981/YEAR, p)& LUD(l, 'a'/USAG, d)& LT(p, #60/PRIC)& LT(d, #40/DIST)].

[質問 4] すべての調査年度について調査されている土地を求めよ．

Q 4: ($^\exists l$^/LAND)($^\forall y$/YEAR)($^\exists p$/PRIC)[LYP(l, y, p)].

Q 4 の属性 PRIC は質問 4 の中には現れていない．このように質問の中に現れない変数は ∃ 限量され，一連の限量子の最後に現れる．このような変数は最終的には射影演算が行われて消去される．本稿では記述を簡単にするためにこのような質問に直接関係のない属性に対して "@" を用いた略記法を用いる．たとえば Q 4 は ($^\exists l$^/LAND) ($^\forall y$/YEAR) [LYP (l, y, @)] のように表現する．

[質問 5] すべての調査年度について調査されている土地とその用途と都心からの距離を求む．

Q 5: ($^\exists l$^/LAND) ($^\exists u$^/USAG) ($^\exists d$^/DIST ($^\forall y$/YEAR) [LYP(l, y, @) & LUD(l, u, d)].

[質問 6] 2年以上にわたって調査されている土地を求む．

Q 6: ($^\exists l$^/LAND) ($^\exists Y$/∗YEAR) ($^\forall y$/Y) [LYP(l, y, @)& GE(count(Y), #2/INTG)].

ここで INTG は整数の集合を表す．これはシステムに作り付けにされた集合である．

［質問 7］ 1979 年度に何か所で土地価格の調査が行われたか.

Q 7: $(^{\exists}L/{*}\text{LAND})\ (^{\exists}n^{\wedge}/\text{INTG})\ (^{\forall}l/L)\ [\text{LYP}(l, \#1979/\text{YEAR}, \text{ⓐ})\ \&\ \text{LET}(n, \text{count}(L))]$.

［質問 8］ すべての調査年度にわたって調査されている土地は何か所あるか.

Q 8: $(^{\exists}L/{*}\text{LAND})/^{\exists}n^{\wedge}/\text{INTG})(^{\forall}l/L)(^{\forall}y/\text{YEAR})\ [\text{LYP}(l, y, \text{ⓐ})\ \&\ \text{LET}(n, \text{count}(L))]$.

［質問 9］ 2年以上にわたって調査されている土地は何か所あるか.

Q 9: $(^{\exists}L/{*}\text{LAND})\ (^{\exists}n^{\wedge}/\text{INTG})\ (^{\forall}l/\text{L})\ (^{\exists}Y/{*}\text{YEAR})(^{\exists}y/Y)\ [\text{LYP}(l, y, \text{ⓐ})\ \&\ \text{GE}(\text{count}(Y), \#2/\text{INTG})\ \&\ \text{LET}(n, \text{count}(L))]$.

Q 9は, LAND の部分集合 L でそのすべての要素 l に対して要素数が2以上である YEAR の部分集合 Y が対応しているものを求め, その要素数 n を求めることを表現している. 図1のデータベースの場合 $L=\{\text{A}1, \text{A}2, \text{C}1\}$ であるから $n=3$ となる.

［質問 10］ 1981 年度に調査されたか, 用途が b である土地を求む.

Q10: $(^{\exists}l^{\wedge}/\text{LAND})\ [\text{LYP}(l, \#1981/\text{YEAR}, \text{ⓐ}) \cup \text{LUD}(l, {}'b'/\text{USAG}, \text{ⓐ})]$.

［質問 11］ 用途が b または d の土地で 1981 年度に調査された土地は何か所か.

Q11: $(^{\exists}L/{*}\text{LAND})\ (^{\exists}n^{\wedge}/\text{INTG})\ (^{\forall}l/L)\ [\text{LYP}(l, \#1981/\text{YEAR}, \text{ⓐ})\ \&\ (\text{LUD}(l, {}'b'/\text{USAG}, \text{ⓐ}) \cup \text{LUD}(l, {}'d'/\text{USAG}, \text{ⓐ}))\ \&\ \text{LET}(n, \text{count}(L))]$.

［質問 12］ 用途 c の土地のうち, この用途の土地が調査対象となったすべての年にわたって調査されたものを除いた土地を求む.

Q12: $(^{\exists}Z/{*}\text{YEAR})\ (^{\forall}z/Z)\ (^{\exists}m/\text{LAND})\ (^{\exists}L/{*}\text{LAND})(^{\forall}l^{\wedge}/L)(^{\exists}Y/{*}\text{YEAR})\ (^{\exists}y/Y)[\text{LYP}(m, z, \text{ⓐ})\ \&\ \text{LUD}(m, {}'c'/\text{USAG}, \text{ⓐ})\ \&\ \text{LYP}(l, y, \text{ⓐ})\ \&\ \text{LUD}(l, {}'c'/\text{USAG}, \text{ⓐ})\ \&{\sim}\text{SSET}(Z, Y)]$.

［質問 13］ 用途 c の土地のうち, 全調査年度にわたって調査されたものを除いた土地を求む.

Q13: $(^{\exists}L/{*}\text{LAND})\ (^{\forall}l^{\wedge}/L)\ (^{\exists}Y/{*}\text{YEAR})(^{\exists}y/Y)\ (^{\exists}Z/{*}\text{YEAR})\ (^{\forall}z/Z)\ [\text{LYP}(l, y, \text{ⓐ})\ \&\ \text{LUD}(l, {}'c'/\text{USAG}, \text{ⓐ})\ \&\ \text{LYP}(\text{ⓐ}, z, \text{ⓐ})\ \&{\sim}\text{SSET}(Z, Y)]$.

［質問 14］ 1981 年度における用途 a の土地およびその価格と, その年に調査された全体の土地価格の平均との差を求む.

Q14 $(^{\exists}l^{\wedge}/\text{LAND})\ (^{\exists}p^{\wedge}/\text{PRIC})\ (^{\exists}Q/{*}\text{PRIC})(^{\exists}d^{\wedge}/\text{PRIC})\ (^{\forall}q/Q)\ [\text{LYP}(l, \#1981/\text{YEAR}, p)\ \&\ \text{LUD}(l, {}'a'/\text{USAG}, \text{ⓐ})\ \&\ \text{LYP}(\text{ⓐ}, \#1981/\text{YEAR}, q)\ \&\ \text{LET}(d, \text{sub}(p, \text{avg}(Q)))]$.

［質問 15］ 1981 年度における用途 a の土地について, その平均価格と個々の土地価格との差を求む.

Q15: $(^{\exists}L/{*}\text{LAND})(^{\exists}P/{*}\text{PRIC})(^{\forall}l^{\wedge}/L)(^{\exists}p^{\wedge}/P)\ (^{\exists}d^{\wedge}/\text{PRIC})\ [\text{LYP}(l, \#1981/\text{YEAR}, p)\ \&\ \text{LUD}(l, {}'a'/\text{USAG}, \text{ⓐ})\ \&\ \text{LET}(d, \text{sub}(\text{avg}(P), p))]$.

［質問 16］ 1981 年度における用途別の土地価格の平均を求めよ.

Q16: $(^{\forall}u^{\wedge}/\text{USAG})\ (^{\exists}L/{*}\text{LAND})(^{\exists}P/\text{PRIC})(^{\exists}q^{\wedge}/\text{PRIC})\ (^{\forall}l/L)\ (^{\exists}p/P)\ [\text{LYP}(l, \#1981/\text{YEAR}, p)\ \&\ \text{LUD}(l, u, \text{ⓐ})\ \&\ \text{LET}(q, \text{avg}(P))]$.

［質問 17］ 1981 年度における用途別の土地価格の平均の平均を求めよ.

Q17: $(^{\exists}PP/{*}{*}\text{PRIC})\ (^{\exists}q^{\wedge}/\text{PRIC})\ (^{\forall}u/\text{USAG})\ (^{\exists}L/{*}\text{LAND})\ (^{\exists}P/PP)\ (^{\forall}l/L)\ (^{\exists}p/P)\ [\text{LYP}(l, \#1981/\text{YEAR}, p)\ \&\ \text{LUD}(l, u, \text{ⓐ})\ \&\ \text{LET}(q, \text{avg}(\text{avg}(PP)))]$.

質問1～17を他の質問言語と比較するため ALPHA, DEDUCE, QUERY BY EXAMPLE (Q. B. E), SEQUEL, SQUARE による表現をまとめたものを**表1**に示した. この表において'Y', '?', 'N' はそれぞれ表現できる, 表現できるかどうか判断できない, 表現できないを表す記号である. この表から, 多層論理式が述語論理型に属する既存の質問言語 ALPHA, DEDUCE にまさる記述力を備えていることがわかる. なお, 表1の作成に当たっては参考文献 1)～5), 9) を参考にした.

Q1～Q17 のなかには, その質問を表現するためには必要でない属性を含んでいるものがある (Q4～Q17). この現象は質問を表現するために必要となるリ

表 1 既存の言語によって本文中の質問が表現できるかどうかを示した表. 'Y', '?', 'N' はそれぞれ表現できる, 不明, 表現できないことを示す.

Table 1 This table indicates whether queries listed in this paper could be expressed in given languages. 'Y', '?' and 'N' indicate expressible, un-decidable and not expressible, respectively.

言語	質問																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
ALHPA	Y	Y	Y	Y	Y	Y	Y	Y	?	N	Y	Y	Y	N	N	N	N
DEDUCE	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	?	?
Q. B. E.	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	?	Y	N	N	Y	N
SEQUEL	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	?
SQUARE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

レーションの数が増大するほど頻繁に生ずる傾向にある．この表記上の問題を解決するために，質問の記述に必要な属性だけを含む仮想的なリレーションを定義し，これに対して質問を行えるようにする必要がある．次に示した多層論理式はリレーションLYP, LUDからLAND, YEAR, PRIC, USAGを属性とするリレーションYPUを定義しているものである．

K. $(^{\forall}l/\mathrm{LAND})(^{\forall}y/\mathrm{YEAR})(^{\forall}p/\mathrm{PRIC})(^{\forall}u/\mathrm{USAG})(^{\exists}d/\mathrm{DIST})\ [(\mathrm{LYP}\ (l,y,p)\ \&\ \mathrm{LUD}\ (l,u,d))\rightarrow \mathrm{YPU}\ (l,y,p,u)]$.

この論理式はリレーションLYPとLUDを属性LANDで結合し，LUDの属性DISTで射影したものがリレーションYPUであると読むことができる．仮想的なリレーションYPUを用いて質問17を表現したものを次に示す．

Q' 17 $(^{\exists}PP/\mathrm{PRIC})(^{\exists}q\hat{\ }/\mathrm{PRIC})(^{\forall}u/\mathrm{USAG})(^{\exists}L/{*}\mathrm{LAND})(^{\exists}P/PP)(^{\forall}l/L)(^{\exists}p/P)\ [\mathrm{YPU}\ (l,\#1981/\mathrm{YEAR},p,u)\ \&\ \mathrm{LET}(q,\ \mathrm{avg}\ (\mathrm{avg}(PP)))]$.

3. 推論処理

Q' 17のように仮想的なリレーションを含む質問は，データベース，関数および述語定数を用いて直接評価することはできない．そのために仮想的なリレーションを，それを定義している論理式を用いて展開し，仮想的なリレーションを含まない質問に変換しなければならない．この処理を推論処理とよぶ．推論処理は基本的には質問内の仮想的なリレーションを表す述語を，それを含意する述語を含む論理式によって置き換えることである．多層論理式における推論処理の詳細は別稿に譲り，ここではデータベースの質問処理に関して必要となる部分を例を用いて述べる．

Pが述語記号で，$m\geqq k\geqq 0$，$n\geqq k\geqq 0$であるとするとき $(Q'_n y_n/{*}_n\cdots{*}_1 S)\cdots(Q'_k y_k/y_{k+1})\cdots(Q'_0 y_0/y_1)P(y_k)$ が $(Q_m x_m/{*}_m\cdots{*}_1 S)\cdots(Q_k x_k/x_{k+1})\cdots(Q_0 x_0/x_1)P(x_k)$ を含意するための条件を**表2**に示した．Q' 17の述語YPUは仮想的なリレーションを表しているので推論処理の対象となる．量記号を含めたQ' 17内の述語YPUは，F: $(^{\forall}u/\mathrm{USAG})(^{\forall}l/L)(^{\exists}p/P)$ YPU

表2 含意の条件
Table 2 Implication condition.

Q'_k	Q_k	Condition
∀	∀	$y_{k+1}\supseteq x_{k+1}$
∀	∃	$y_{k+1}\wedge x_{k+1}\neq\phi$
∃	∃	$y_{k+1}\subseteq x_{k+1}$

$(l,\ \#1981/\mathrm{YEAR},\ p,\ u)$であり，K内の述語YPUは，F': $(^{\forall}l'/\mathrm{LAND})(^{\forall}y'/\mathrm{YEAR})(^{\forall}p'/\mathrm{PRIC})(^{\forall}u'/\mathrm{USAG})$ YPU (l',y',p',u')である．変数xの領域をDom(x)で表すときF, F'における変数の領域はDom$(l')\supseteq$Dom(l), Dom$(p')\wedge$Dom$(p)\neq\phi$, Dom$(u')\supseteq$Dom(u)なる関係を満たし，変数yに対しては定数1981が対応しているのでF'はFを含意する．したがってQ' 17のYPUをKによって置き換える．このときKに含まれQ' 17に含まれない変数dが導入される．変数dの限量子内での位置はQ' 17の変数pよりも後であることがKの限量子から判明する．この結果Q' 17は次に示すQ'' 17に変換される．

Q'' 17 $(^{\exists}PP/{*}{*}\mathrm{PRIC})(^{\exists}q\hat{\ }/\mathrm{PRIC})(^{\forall}u/\mathrm{USAG})(^{\exists}L/{*}\mathrm{LAND})(^{\exists}P/PP)(^{\forall}l/L)(^{\exists}p/P)(^{\exists}d/\mathrm{DIST})\ [\mathrm{LYP}(l,\ \#1981/\mathrm{YEAR},\ p)\ \&\ \mathrm{LUD}(l,u,d)\ \&\ \mathrm{LET}(q,\ \mathrm{avg}\ (\mathrm{avg}(PP)))]$.

4. 多層論理式からデータベース検索手続きへの変換

4.1 多層論理式のグラフによる表現

モデル述語内の変数の値はその述語を評価する（対応するリレーションをアクセスする）ことによって定まるのに対し，述語定数を評価するためにはその引数の値が定まらなければならない．このように述語とその引数の依存関係は述語によって異なるが，これまで用いてきた多層論理式の記法ではこの依存関係は陽に表現されない．そこで多層論理式を以下に示す**CAVグラフ**(Constant-Atom-Variable relationship graph)を用いて表現し，これに基づいて検索手続きへの変換アルゴリズムを説明する．

CAVグラフは変数，述語，定数または関数を表すノードと，ノードの関係を表すアークから成る．v, wがともに変数を表すノードであるとき$v\leftarrow w$はvがwに先行することを表し，$v\leftarrow\!\!\leftarrow w$は$w$の領域は$v$であることを表す．$b$が述語または関数を表すノードであるとき$v\leftarrow b$は$b$を評価することによって$v$の値が定まることを示し，$v\rightarrow b$は$b$を評価するために$v$の値が定まっていなければならないことを示す．また，cが定数または関数を表すノードでaが述語を表すノードであるとき$c\leftarrow a$はcがaに含まれていることを示す．**図2**はQ'' 17をCAVグラフを用いて表現したものである．

4.2 関係演算および集約演算の定義

ここでは関係演算および集約演算の定義を与える．

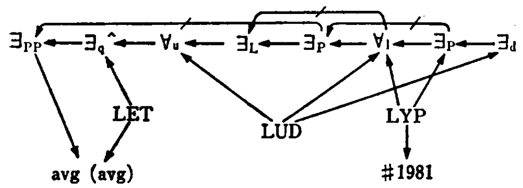


図 2 Q'' 17 を CAV グラフによって表現したもの
Fig. 2 CAV-graph representation of Q'' 17.

属性 $A_1, \cdots, A_n (n \geqq 1)$ 上の関係 R を $R(A_1, \cdots, A_n)$ で，その要素を $\langle r_1, \cdots r_n \rangle$ で表す．また $\langle r_1, \cdots, r_n \rangle$ から $r_i (1 \leqq i \leqq n)$ を除いた $n-1$ 組 $\langle r_1, \cdots r_{i-1}, r_{i+1}, \cdots, r_n \rangle$ を $\overline{\langle r_i \rangle}$ で表す．

［定義］ c を定数とする．関係 $R(A_1, \cdots, A_n)$ の属性 A_i の c に関する制約 $\boldsymbol{R}_{A_i=c}(R)$ とは次の式で定義される集合である．

$$\boldsymbol{R}_{A_i=c}(R) = \{\overline{\langle r_i \rangle} \mid \langle r_1, \cdots, r_n \rangle \in R \text{ and } r_i = c\}.$$

［定義］ 関係 $R(A_1, \cdots, A_n)$ の属性 A_i に関する射影 $\boldsymbol{P}_{A_i}(R)$ は次の式で定義される．

$$\boldsymbol{P}_{A_i}(R) = \begin{cases} \{\overline{\langle r_i \rangle} \mid \langle r_1, \cdots, r_n \rangle \in R\} & (n > 1) \\ T & (n = 1) \end{cases}$$

T (top) は関係演算は正常に行われるが，結果として属性のないリレーションが得られたことを意味する．

［定義］ 関係 $R_1(A_1, \cdots, A_n)$, $R_2(B_1, \cdots, B_m)$ の属性 $A_i (1 \leqq i \leqq n)$ と $B_j (1 \leqq j \leqq m)$ の自然結合 $(R_1)_{A_i} \boldsymbol{N}_{B_j}(R_2)$ とは次の集合を満たす集合である．ただし，A_i と B_j は同じ定義域をもつ属性でなければならない．

$$(R_1)_{A_i} \boldsymbol{N}_{B_j}(R_2) = \{\langle r_1, \cdots, r_i, \cdots, r_n, s_1, \cdots, s_{j-1}, s_{j+1}, \cdots, s_m \rangle \mid \langle r_1, \cdots, r_n \rangle \in R_1 \text{ and } \langle s_1, \cdots, s_m \rangle \in R_2 \text{ and } r_i = s_j\}.$$

［定義］ 関係 $R(A_1, \cdots, A_n)$ において属性 A_i の値 $\langle x_i \rangle$ による属性 $A_j (i \neq j)$ の像集合 $g_R(A_j, A_i, \langle x_i \rangle)$ は次の式で定義される．

$$g_R(A_j, A_i, \langle x_i \rangle) = \{\langle r_j \rangle \mid \langle r_1, \cdots, r_n \rangle \in R \text{ and}$$

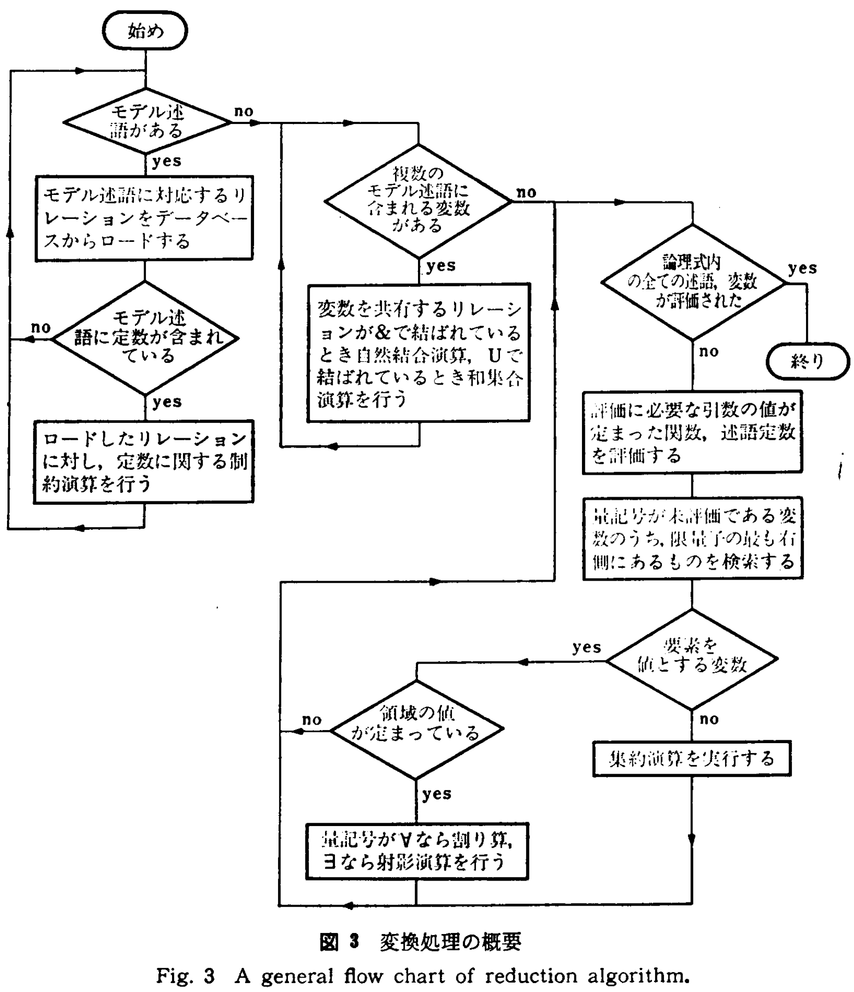


図 3 変換処理の概要
Fig. 3 A general flow chart of reduction algorithm.

$\langle r_i\rangle=\langle x_i\rangle\}$.

［定義］ 関係 $R(A_1,\cdots,A_n)$ の属性 A_i による属性 $A_j(i\neq j)$ の集約 ${}_{A_j}\boldsymbol{G}_{A_i}(R)$ とは次の式を満たすマルチ・セット（重複を許す要素のあつまり）である．

$${}_{A_j}\boldsymbol{G}_{A_i}(R)=g_R(A_j, A_i, \langle r_i\rangle)$$

［定義］ 関係 $R(A_1,\cdots,A_n)$ の属性 A_i による割り算 $\boldsymbol{D}_{A_i}(R)$ は次の式で定義される．

$$\boldsymbol{D}_{A_i}(R)=\begin{cases}\{\overline{\langle r_i\rangle}\mid\langle r_1,\cdots,r_n\rangle\in R \text{ and } P_{\overline{A_i}} \\ (R)\subseteq g_R(A_i,\ \overline{A_i},\ \overline{\langle r_i\rangle}\} & (n>1)\\ T & (n=1)\end{cases}$$

ここで定義した集約演算は Codd (1971)[4] によって定義された像集合から直接定義されるものである．多層論理では集合を値とする変数を陽に使用しているので，その値を決定するために集約演算を使う必要がある．

4.3 検索手続への変換処理

3章で仮想的なリレーションを含む質問が推論処理によって仮想的なリレーションを含まない表現に変換されることを述べた．ここでは仮想的なリレーションを含まない質問をデータベース検索手続に変換するアルゴリズムを示す．

述語論理式をデータベース検索手続に変換するアルゴリズムは Codd (1971)[4], Reiter (1978)[8] によって理論づけられている．しかし，彼らが対象とした論理式は一階述語の範囲であり，集合を値とする変数を陽に含む多層論理式に対し彼らが示したアルゴリズムをそのまま適用することはできない．集合を値とする変数は，それを規定する条件がすべて評価されるまで値を決定することはできないからである．

図3に多層論理式をデータベース検索手続に変換するアルゴリズムの概要を示した．これは Reiter (1979)[8], Chang (1978)[3] によって論じられた変換アルゴリズムに集合の値を決定する処理を付け加えたものである．つぎに，このアルゴリズムを Q'' 17 に適用した例を示そう．以下ではノード $\boldsymbol{v}$ が評価されたことをⓥで示す．まず，Q'' 17 内のモデル述語をすべて評価する．モデル述語の引数に定数が含まれるときは，その定数に関する制約演算を行う．この処理によって図2に示した論理式は**図4**のようになる．変数 $\boldsymbol{l}$ は述語 LUD と LYP に含まれるので2回評価される．したがって LUD と LYP の $\boldsymbol{l}$ に関する結合演算が行われる．これによって $\boldsymbol{l}$ の◎は○に変わる．ここまでの処理によって図4は**図5**(a)になり，手続 $(\mathrm{LUD})_{\mathrm{LAND}}\ \mathbf{N}_{\mathrm{LAND}}\ (\mathbf{R}_{\mathrm{YEAR=1981}}(\mathrm{LYP}))$ が得られる．

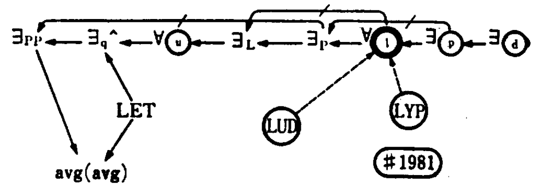
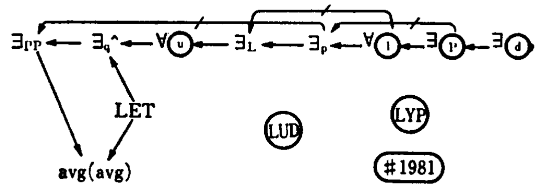


図 4 モデル述語の評価が終わったときの Q'' 17
Fig. 4 Q'' 17 after evaluation of model predicates.



(a)

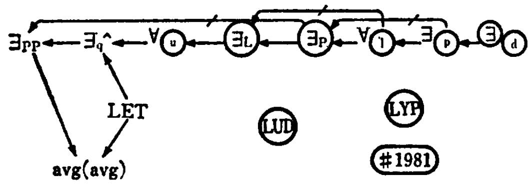
LAND	PRIC	USAG	DIST
A 1	58	a	40
A 2	86	a	25
C 1	25	c	10
D 1	40	d	20
D 2	36	d	30

(b)

図 5 (a) 結合演算が終わったときの Q'',
(b) 実行結果
Fig. 5 (a) Q'' 17 after join operation,
(b) result of this operation.

この手続の実行によって図5(b)のリレーションが得られる．

図3のアルゴリズムによれば，続いて述語定数 LET を評価しようとするが変数 PP の値が定まっていないのでこの時点で LET を評価することはできない．そこで量記号に対する処理が行われる（ただし，変数 q はモデル述語の引数になっていないのでこの処理の対象にはならない）．変数 d の量記号は∃であり領域は図5(b)のリレーションの属性 DIST であるから，d の量記号は DIST に関する射影演算に変換される．変数 p, l の領域 P, L の値はこの時点では決定していないので p, l の量記号に関する処理は行われない．変数 P, L に対しては，これに先行する変数 u の各値に対応する集合が定められる．また，変数 u の値とそれに対応する p, l の値から成るリレーションが u の値の数だけ作られる．これまでの処理によって手続 ${}_{\mathrm{PRIC}}\mathbf{G}_{\mathrm{USAG}}\ (\mathbf{P}_{\mathrm{DIST}}\ ((\mathrm{LUD})_{\mathrm{LAND}}\ \mathbf{N}_{\mathrm{LAND}}\ (\mathbf{R}_{\mathrm{YEAR=1981}}(\mathrm{LYP}))))$ と ${}_{\mathrm{LAND}}\mathbf{G}_{\mathrm{USAG}}\ (\mathbf{P}_{\mathrm{DIST}}((\mathrm{LUD})_{\mathrm{LAND}}\ \mathbf{N}_{\mathrm{LAND}}(\mathbf{R}_{\mathrm{YEAR=1981}}(\mathrm{LYP}))))$ が得られる．変数 P, L を領域とする変数 p, l に関する制約条件が論理式内にない（図5において p, l から述語を表すノードへ



(a)

LAND	PRIC	USAG
A 1	58	a
A 2	86	a

LAND	PRIC	USAG
C 1	25	c

LAND	PRIC	USAG
D 1	40	d
D 2	36	d

$L=\{A1,\ A2\}_a,\ \{C1\}_c,\ \{D1,\ D2\}_d$
$P=\{58,\ 86\}_a,\ \{25\}_c,\ \{40,\ 36\}_d$

(b)

図 6 (a) 集約演算の処理が終わったときの Q'' 17, (b) 実行結果. (b)で $\{\cdots\}_x$ は x で集約されたマルチ・セットを表す.

Fig. 6 (a) Q'' 17 after grouping operation, (b) result of this operation. $\{\cdots\}_x$ indicates multi-set group by x.

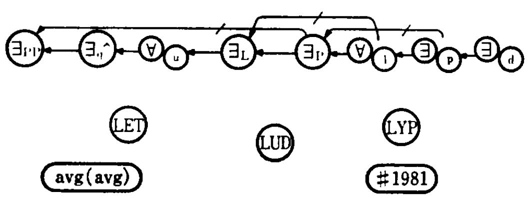


図 7 変換処理が終了したときの Q'' 17

Fig. 7 Q'' 17 after reduction completed.

のアークがない) ので上記の手続の実行結果がそれぞれPとLの値になる. これまでの処理で図 5 (a)は**図 6** (a)になり, 上記の手続の実行によって図 6 (b)に示した結果が得られる.

続いて, 変数pの量記号∃が射影に変数lとuの量記号∀が割り算に変換される. また, 変数PPの値は変数Pの値を要素とする集合, すなわち $\{\{58,\ 86\}_a,\ \{25\}_c,\ \{40,\ 36\}_d\}$ となる. PP の値が定まると関数 avg (avg (PP)) の値が求まるので LET 述語が評価され, 変数qの値が求まる. この一連の処理によって Q'' 17 内のすべての述語と変数およびリレーションの属性を領域とする変数の量記号が評価されたので変換処理は終了する (**図 7**). 変数 PP の値が決定した後の処理は, PP に対する集合演算と図 6 (b)に示したリレーションに対する処理が行われる. すなわち, avg({avg({58, 86}$_a$), avg({25}$_c$), avg({40, 36}$_d$)}) の値を求め変数qに代入することと, $\mathbf{D}_{USAG}$ ($\mathbf{D}_{LAND}$ ($\mathbf{P}_{PRIC}$ ($\mathbf{P}_{DIST}$ ((LUD)$_{LAND}$$\mathbf{N}_{LAND}$ ($\mathbf{R}_{YEAR=1981}$ (LYP)))))) の実行が行われる. 前者より $q=45$(万円/m^2) が得られ, 後者より関係演算が正常に終了したことを示すT (top) が得られる.

5. おわりに

本稿では多層論理式によるリレーショナル・データベースに対する質問の記述方法と推論アルゴリズム, ならびに検索手続への変換アルゴリズムについて論じた. これらのアルゴリズムは, いずれも many sorted logic におけるアルゴリズムを一部修正することによって実現できるものである. 現在 many sorted logic に基づいた知識システム SBDS-F 2[10] (KAUS[6] の推論処理およびデータベース・アクセスを行うシステム) の機能を修正することによってこれらのアルゴリズムを含むシステムを実現中である[11]. 本稿で提案した多層論理式はデータベース質問言語の内部表現形式であり, 現在グラフィックスを用いた対話型言語によって多層論理式を生成するアルゴリズムについて検討している. これらの研究については稿を改めて発表する予定である.

参考文献

1) Boyce, R. R. et al.: The SQUARE Data Sublanguage, *Comm. ACM*, Vol. 18, No. 11, pp. 621-628 (1975).
2) Chamberlin, D. D. et al.: SEQUEL 2: A Unified Approach to Data Definition, Manipulation and Control, *IBM J. Res. Dev.* Vol. 20, No 6, pp. 560-575 (1976).
3) Chang, C. L.: DEDUCE 2: Futher Investigation on Deduction in Relational Data Base, In Gallaire, H. and Minker, J. (ed.),: *Logic and Data Bases*, Plenum Press, New York, pp. 201-236 (1978).
4) Codd, E. F.: Relational Completeness of Data Base Sublanguages, In *Data Base Systems, Courant Computer Science Symposium 6*, Prentice Hall, New York, pp. 65-98 (1971).
5) Lacroix, M. and Pirotte, A.: Example Queries in Relational Languages, MBLE Technical Note N 107, Revised September (1977).
6) Ohsuga, S.: Perspectives on Computer Systems of the Next Generation—A Proposal for Knowledge-Base Systems, *JIP*, Vol. 3, No. 3, pp. 171-185 (1980).
7) Pirotte, A.: High Level Data Base Query Languages, In Gallaire, H. and Minker, J. (ed.), *Logic and Data Bases*, Plenum Press, New York pp. 409-436 (1978).

8) Reiter, R.: Deductive Question-answering on Relational Data Bases, In Gallaire, H. and Minker, J. (ed.), *Logic and Data Bases*, Plenum Press, New York pp. 149-177 (1978).
9) Zloof, M. M.: Query-by-Example: A Data Base Language, *IBM Syst. J.* Vol. 4, pp. 324-343 (1977).
10) 宇田川佳久，大須賀節雄：知識システムの設計問題への応用について，情報処理学会第20回全国大会予稿集，pp. 695-696 (1979).
11) 宇田川佳久，大須賀節雄：多層論理式からのデータベースアクセス手順の生成について，京都大学数理解析研究所講究録 423, pp. 40-60 (1981).
12) 古川康一：データベースの知的アクセスに関する研究，電総研研究報告，No. 804 (1979).

(昭和 56 年 5 月 29 日受付)
(昭和 57 年 5 月 19 日採録)