数理解析研究所講究録
第591巻 1986年 74-83


# ネットワークデータベースにおける巡航操作で解ける質問のクラス

九州大学工学部　古川哲也（Tetsuya Furikawa）
九州大学工学部　上林彌彦（Yahiko Kambayashi）

## １．まえがき

ネットワークデータベースシステムは、効率が良くある程度の自由度の高さを有しており、実用的データベースとして広く普及している。ネットワークモデルでの質問処理法を明確にすることは重要であり、本稿では巡航操作に基づく質問処理の複雑さについて検討している。

関係モデルにおける質問処理は、関係間の演算として集合的に行なわれ、演算は新しい関係を生成する。中間結果の関係ももとの関係と同じように扱うことができるが、それには最終的な解とはならないものも含まれる。一方ネットワークモデルにおける質問処理は、リンクをたどる巡航操作によりレコード間の対応を得、質問の解となる値の組を１つずつ求めることに基づいている。即ち巡航操作では直接解を得る必要があり、巡航中に現在求めている組が解となるかどうかが分かっていなければならない。巡航操作のみで質問の解が得られなければ、その結果に対しさらに親言語を用いた処理を必要とする。従って、結果が解となるように巡航を行なう方がよいが、巡航操作の能力によってはそれのみで解けない質問も存在するため、巡航操作の能力とそれにより解ける質問のクラスを明確にしておかなければならない。質問処理に関しては、質問のクラスを限定した処理法[1][2]が知られているが、巡航操作の能力自体は明らかにされていない。

本稿では、ネットワークデータベースに対する質問をデータベースの構造から分類し、巡航操作の能力の違いによる解ける質問のクラスを明確にすることを試みる。さらにそのクラスの質問が関係モデルにおける質問のクラスとどのように対応しているかについて議論する。

## ２．ネットワークモデルにおける単純質問

ネットワークモデルは、同じデータ項目（本稿では関係モデルとの対応のため属性と呼ぶ）からなるレコードの集合であるレコード型と、２つのレコード型間

のレコードの１対多の対応を表わす親子集合型の集合によって表現される。属性集合Xで構成されるレコード型RをR(X)で、親レコード型が$R_o$，子レコード型が$R_m$である親子集合型Sを$S<R_o,R_m>$で表わす。$S<R_o,R_m>$ ではRoの各レコードに対し$R_m$の任意個のレコードが対応づけられており、それぞれを親子集合という。

ネットワークモデルの構造はバックマン線図と呼ばれる有向グラフＢ(V,A) で表わされる。Vは各レコード型に対応する節点集合，Aは各親子集合型に対応し、親レコード型に対応する節点から子レコード型に対応する節点に向かう有向枝の集合である。値が定まればレコード型R(X)の対応するレコードがただ１つ定まるような最小の属性集合Kを、R(X)のキーと呼ぶ。

ネットワークモデルにおける単純質問を、関係代数演算を用いて定義する。次に関係代数の主要な演算を示す。組tの属性集合Xの値をt[X]で表わし、$\theta$を比較演算子＝,＞,≧,≦,＜,≠ のいずれかとする。

選択：関係R(U)と属性$A_1,A_2(\in U)$，定数'c'について

$$1:\ R[A_1\,\theta\,A_2]=\{t \mid t\in R,\ t[A_1]\,\theta\,t[A_2]\}$$

$$2:\ R[A_1\,\theta\,'c']=\{t \mid t\in R,\ t[A_1]\,\theta\,'c'\}$$

のいずれか。文献[3] ではさらに条件の論理和も許しており（論理積は複数の選択演算で表わされる）、文献[4]では２のみを選択演算としている。本稿では両方を含めて選択演算とし、１を制約演算，２を狭義の選択演算として区別する。

射影：関係R(U)の属性集合$X(\subseteq U)$について

$$R[X]=\{t[X] \mid t\in R\}$$

結合：関係$R_1(U_1),R_2(U_2)$と属性$A_{1i},A_{2i}(A_{1i}\in U_1,A_{2i}\in U_2(1\leqq i\leqq n))$につて

$$R_1 \bowtie R_2=R(U_1U_2),\ c=(R_1.A_{11}\,\theta_1\,R_2.A_{21})\wedge\cdots\wedge(R_1.A_{1n}\,\theta_n\,R_2.A_{2n})$$

$$R=\{t \mid t[U_1]\in R_1,\ t[U_2]\in R_2,\ t[A_{1i}]\,\theta_i\,t[A_{2i}](1\leqq i\leqq n)\}$$

以下の定義は、ネットワーク構造として与えられたバックマン線図Ｂとその実現値に対して行なう。

〔定義〕　バックマン線図Ｂの連結な部分グラフで表わされるネットワーク構造をＢの<u>部分ネットワーク構造</u>という。 □

〔定義〕　バックマン線図Ｂで表わされるネットワーク構造に含まれるレコード型を$R_1,R_2,\cdots,R_n$とする。レコード集合 $t=\{r_1,r_2,\cdots,r_n\}$ （$r_i$は$R_i$のレコード($1\leqq i\leqq n$)）でＢに親子集合型$<R_i,R_j>$ があれば、$r_i$は$r_j$の親レコードであるとき、tをＢにおける<u>レコードの組</u>と呼ぶ。 □

〔定義〕　バックマン線図Ｂで表わされるネットワーク構造に含まれるレコード型を構成する属性集合の和集合を$U=\{R.A \mid$ レコード型Rの属性A$\}$とする。Ｂの関係はU上の関係uであり、u(Ｂ)で表わす。u(Ｂ)の各組はＢにおけるレコードのすべての組と１対１に対応し、対応する組とレコードの属性の値は等しい。□

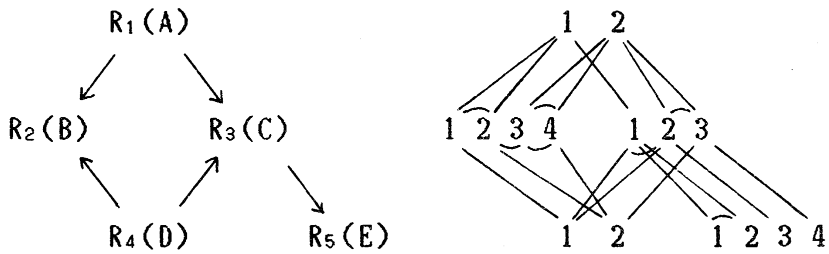
〔定義〕　質問Ｑの解がある部分ネットワーク構造Ｂの関係u(Ｂ)に対する選択及び射影演算で求められるとき、その質問を単純質問という。また、ＢをＱの質問スキーマ$B_Q$，$u(B_Q)$をＱの（質問スキーマの）関係という。□

バックマン線図Ｂ，質問Ｑに対し、Ｂの冗長性によりＢにおけるＱの質問スキーマは一意ではない。単純質問は、質問スキーマの定義により木質問と巡回質問に分類される。

〔定義〕　バックマン線図Ｂで表わされるネットワークスキーマにおいて非巡回質問スキーマが存在する単純質問を木質問，それ以外の単純質問を巡回質問と呼ぶ[5]。巡回性はバックマン線図を無向としたときのものである。□

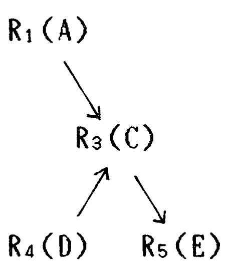
以下の議論でも閉路，巡回性はバックマン線図を無向とした上で考える。

（例１）　ネットワーク構造全体を図１(a)に示されるバックマン線図Ｂ，そのレコードの対応を図１(b)とする。属性集合ADE の対応を求める必要がある質問




(a)　　(b)

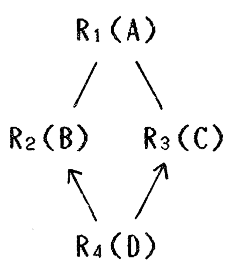
図１ ネットワーク構造とレコードの対応




(a)

| A | C | D | E |
|---|---|---|---|
| 1 | 1 | 1 | 1 |
| 1 | 1 | 1 | 2 |
| 2 | 2 | 1 | 3 |
| 2 | 3 | 2 | 4 |

(b)




(c)

| A | B | C | D |
|---|---|---|---|
| 1 | 1 | 1 | 1 |
| 2 | 3 | 3 | 2 |
| 2 | 4 | 3 | 2 |

(d)

図２ 質問スキーマとその関係

$Q_1$の質問スキーマ$B_{Q1}$は図２(a)，その関係u($B_{Q1}$)は図２(b)となる。また、属性集合ABCDの対応を求める必要がある質問$Q_2$の質問スキーマ$B_{Q2}$は図２(c)，その関係u($B_{Q2}$)は図２(d)であり、この質問は巡回質問である。□

## 3．巡航操作による単純質問の処理

本節では、単純質問を処理するための巡航操作の機能について議論する。ネットワークモデルには次の現在指示子(currency indicator:ci) があり、巡航操作はこれらのデータベース中の移動である。

1．レコード型の現在指示子(rci)

2．親子集合型の現在指示子(sci)

3．実行単位の現在指示子(uci)

1，2はそれぞれ各レコード型，親子集合型ごとにあり１つのレコードを指す。3は実行単位に１つありrci,sciのうちのどちらかと同じレコードを指す。

〔定義〕　現在指示子の移動によりレコードの対応を求める操作を基本巡航操作という。

rci：レコード型のレコードを順に動く。sci がそのレコード型のレコードを指せばそれに一致する。

sci：親→子　uciが指す親レコード型のレコードの最初の子レコードを指す。

子→親　uciが指す子レコード型のレコードの親レコードを指す。

子→子　uciが指す子レコード型のレコードの同一親子集合内での次の子レコードを指す。

uci：rci,sciのうち、最も新しく設定したものに一致する。

(rci,sciの任意のものにも一致できる)

これら現在指示子のうち、uciが指すレコードを読むことができる。□

まず、単純質問Qに対し、Qの関係u($B_Q$)が基本巡航操作により求められるかどうかについて検討する。そのために、ネットワーク構造の巡航順序を表わす巡航列を定義する。

〔定義〕　ネットワーク構造B(V,A)の巡航列は次の条件を満たす列$R_0$,($S_1$,$R_1$),($S_2$,$R_2$),…,($S_n$,$R_n$)である。ここで、$R_i$($0\leqq i\leqq n$),$S_i$($1\leqq i\leqq n$)に対応する節点,有向枝をそれぞれ$v_i$,$a_i$とする。

(1) $v_i\in V(0\leqq i\leqq n)$, $a_i\in A(1\leqq i\leqq n)$

(2) $\bigcup_{i=0}^{n} v_i=V$, $\bigcup_{i=1}^{n} a_i=A$

(3) $a_i \neq a_j$ $(i \neq j)$

(4) $a_i$は$v_i$と$v_i'$を結ぶ有向枝（方向はどちらでもよい）であり、

$v_i' \in \bigcup_{j=0}^{i-1} v_j$ □

木質問の質問スキーマの巡航列は文献[1]での木質問の処理のための巡航順序と同じとなる。[1]ではコストを最小にするための巡航順序の決定法を示している。

**［補題１］** 木質問の関係は、基本巡航操作のみで求めることができる。□

**（証明）** 任意の木となる部分ネットワーク構造の関係が求められることを、節点数nの帰納法で示す。n＝1のとき明らか。n＝k のとき、部分ネットワーク構造Ｂの関係が基本巡航操作のみで求められたとする。これに木となるようにレコード型Rを追加したものをＢ’，Ｂ’でのRの親（木構造での親で親子集合型の親とは限らない）とRを結ぶ親子集合型Sとすると、Ｂ’の巡航列はＢの巡航列に(S,R)を加えたものとすることができる。Ｂのレコードの対応を表している各sci 集合に対し、親子集合型Sのsci を、この対応に対するRのレコードを順に指すようにする。これによりＢ’の対応がすべて得られる。□

巡回質問に対しては基本巡航操作ではその関係を得ることはできない。これは１つのレコードの組を求めるときに２度以上検索されるレコード型が必ず存在しそのレコードは一致しなければならないためである。従って、巡回質問を解くにはレコードの値の比較ができる必要となる。次にその機能を持つ巡航操作では巡回質問の関係を求めることができることを示す。

**〔定義〕** sci の移動のとき、他のciが指すレコードの値と比較して与えられた条件を満足する次のレコードを求めることができる機能を基本巡航操作に付加したものを<u>比較可能巡航操作</u>という。□

比較可能巡航操作では巡航列の順序に従って巡航を行なえば巡回質問の関係を求めることができる。その際、２度目以降に検索されるレコード型では、先に得たレコードと同じものを求めなければならない。

**［補題２］** 単純質問の関係は比較可能巡航操作のみで求めることができる。□

**（証明）** 木質問のとき明らか。任意の巡回質問Ｑに対しその質問スキーマの巡航列が次のようにして作れる。質問スキーマ$B_Q$の全域木をMin($B_Q$)とする。Min($B_Q$)の巡航列に次の(S,R)を加える。即ち、$B_Q$のMin($B_Q$) に含まれない各枝について、その枝に対応する親子集合型Sとその枝の一方の端点（どちらでもよい）に対応するレコード型R である。この巡航列で得たレコードの組では、組中のレ

コードr$_i$,r$_j$（r$_i$,r$_j$のレコード型をそれぞれR$_i$,R$_j$とする）に対し、R$_i$がR$_j$ の親レコード型であればr$_i$はr$_j$の親レコードである。 □

**（例２）** 例１の質問Ｑ$_1$ で、巡航を始めるレコード型をR$_1$とすると、巡航列は$R_1$,$(S_{12},R_2)$,$(S_{24},R_4)$,$(S_{52},R_2)$ 又は$R_1$,$(S_{12},R_2)$,$(S_{52},R_5)$,$(S_{24},R_4)$（$S_{ij}$は親子集合型<R$_i$,R$_j$>を表わす）とすることができる。質問Ｑ$_2$に対しては、$R_1$,$(S_{12},R_2)$,$(S_{13},R_3)$,$(S_{35},R_5)$,$(S_{52},R_5)$等の巡航列を作ることができる。 □

狭義の選択演算は現在指示子の移動で選択条件を満たすレコードだけを求めることにより，射影演算は射影される属性のみを出力することにより実現できる。制約演算は巡航中に他のレコード型のレコードと比較しなければならない場合があり、基本巡航操作では実現できない。しかし比較可能巡航操作では、巡回質問の関係を求めるときと同様にして実現できる。

**［定理１］** 単純質問が基本巡航操作で解くことができるための必要十分条件はその単純質問がレコード型間にまたがる制約演算を含まない木質問であることである。 □

**［定理２］** 単純質問は、比較可能巡航操作のみで解くことができる。 □

## ４．単純質問の拡張

本節では単純質問を拡張した結合質問を定義し、さらに結合質問のどのクラスが巡航操作で解くことができるかについて議論する。

**〔定義〕** バックマン線図Ｂの部分ネットワーク構造の集合をＪ＝{Ｂ$_1$,Ｂ$_2$,…,Ｂ$_n$}，Ｂ$_i$の関係をu(Ｂ$_i$)($1 \leq i \leq n$)とする。u(Ｂ$_i$）の直積をＪの関係u(Ｊ)といい、質問Ｑの解がu(Ｊ）の選択及び射影演算で求められるＪが存在するとき、その質問を<u>結合質問</u>という。また，ＪをＱの質問スキーマＪ$_Q$，u(Ｊ$_Q$）をＱの（質問スキーマの）関係という。

$V_i$,$A_i$をそれぞれＢ$_i$の節点集合，枝集合としたとき、$V_i \cap V_j = \phi$ ($1 \leq i \leq n, 1 \leq j \leq n, i \neq j$)となる質問スキーマＪ$_Q$＝{Ｂ$_1$,Ｂ$_2$,…,Ｂ$_n$} が存在する結合質問を<u>節点結合質問</u>，$V_i \cap V_j = \phi$，$A_i \cap A_j = \phi$ ($1 \leq i \leq n, 1 \leq j \leq n, i \neq j$)となる結合質問を<u>離散結合質問</u>と呼ぶ。

さらにすべてのＢ$_i$ が木となる質問スキーマが存在するものを木質問，存在しないものを巡回質問に分類する。 □

結合質問Ｑの関係が巡航操作のみで求めることができるには、次の条件を満たす質問スキーマが存在しなければならない。

［条件１］　質問スキーマ$J_Q=\{B_1,B_2,\cdots,B_n\}$で、$B_i(1\leqq i\leqq n)$ の各組について$u(B_j)(i<j\leqq n)$の直積を求めたあと$B_i$の次の組を求めることができる。 □

$B_i$に対する各ciの状態が各$B_j(i<j\leqq n)$ の巡航の後も保存されていればよいが、$B_i$と$B_j$で枝（親子集合型）に重複があれば$B_j$の巡航中にそのsciが変わってしまう。また、巡航を始めるレコード型のrciも$B_j$の巡航中に変わる場合がある。条件１を満足する質問スキーマは補題３で特徴付けられる。

［補題３］　質問スキーマ$J_Q=\{B_1,B_2,\cdots,B_n\}$ が条件１を満たすのは次のいずれかの場合である。

(1) Qが離散結合質問である。

(2) Qが節点結合質問であり、$B_i(1\leqq i<n)$中に$v\notin V_j(i<j\leqq n)$ となる節点が少なくとも１つ存在する。

(3) $B_i$の巡航列 $R_0,(S_1,R_1),(S_2,R_2),\cdots,(S_{ni},R_{ni})$で、$R_0,\cdots,R_m$のレコードを定めると$R_{m+1},\cdots,R_{ni}$のレコードが一意に定まり、$A_i\cap A_j\subsetneqq\bigcup_{k=1}^{m}a_k$ $(i<j\leqq n)$（$a_k$は$S_k$に対応する枝）である。 □

（証明）　(1) は明らか。Qが節点結合質問のとき、rciを用いるのは$B_i$で巡航を始めるレコード型$R_0$のみである。それに対応する節点を$v\notin V_j(i<j\leqq n)$とすれば$B_i$の巡航で$R_0$のrci は変わらない。枝（親子集合型）が重複していても(3)のときは$R_{m+1},\cdots,R_{ni}$の次のレコードの組は存在しないので$S_k(m<k\leqq ni)$のsciは保存しなくてもよい。 □

（例３）　質問$Q_3$を$u(B_{Q1})$と$u(B_{Q2})$ の直積に対する選択，射影演算で解を求める結合質問で、$B_{Q1},B_{Q2}$ の順序で巡航を行なうとする。$B_{Q1}$と$B_{Q2}$には枝（親子集合型）の重複があり、$B_{Q1}$の巡航列を $R_1,(S_{12},R_2),(S_{52},R_5),(S_{24},R_4)$とすると、$S_{12},S_{25}$のsciの内容は$B_{Q2}$の巡航中に変わってしまう。$B_{Q1}$の巡航列を$R_2,(S_{24},R_4),\cdots$とすると、$R_2,R_4$のレコードの組に対する$R_1,R_5$のレコードは一意に定まるので、$S_{12},S_{25}$のsci が変わっても問題ない。従って質問$Q_3$は比較可能巡航操作で解くことができる。 □

rciの移動がsci に独立であれば(2)である必要はないが、同じ１つの節点のみからなる複数の部分ネットワーク構造が存在すると条件１を満たさない。選択演算を考慮すると(3) の条件を満たさなくても巡航操作で解ける場合がある。巡航順序をうまく決めることにより、レコード型のキーによる選択条件などから(3)のときと同様に重複する枝のsci を保存する必要がなくなるためで、このとき条

件１は次のようになる。

［条件２］　質問スキーマ $J_Q=\{B_1, B_2, \cdots, B_n\}$ で、$B_i(1 \leqq i \leqq n)$の選択条件を満たす各組について$u(B_j)(i<j \leqq n)$の直積を求めたあと$B_i$の次の選択条件を満たす組を求めることができる。□

［定理３］　基本巡航操作のみで解ける結合質問のクラスは、条件２を満たす質問スキーマが存在するレコード型間にまたがる制約演算を含まない結合木質問である。□

［定理４］　結合質問が比較可能巡航操作のみで解けるための必要十分条件は、条件２を満たす質問スキーマが存在することである。□

任意の結合質問を解くには現在指示子の状態を復元できる必要がある。無限個の変数を用いることができれば、その変数にレコード型のキー値を記憶しておくことによって復元できる。状態を復元するには、その部分ネットワーク構造の巡航を初めからやりなおし、各レコード型のレコードのキー値が記憶したキー値に一致するまで巡航を進めればよい。

〔定義〕　変数に検索したレコードの値を記憶しておくことができ、sci の移動のとき変数の値と比較して与えられた条件を満足する次のレコードを求めることができる機能を基本巡航操作に付加したものを変数付巡航操作という。□

変数付巡航操作は、以前に検索したレコードの値と比較してsci を移動できるので、比較可能巡航操作の機能も持つ。

［定理５］　変数付巡航操作は、結合質問を解くことができる。□

## ５．関係代数との比較

ネットワークモデルに対応する関係モデル上での関係代数で記述される質問のクラスを、ネットワークモデルにおける巡航操作で解ける質問のクラスと比較する。関係代数は、２節で示した関係演算に和，差，共通集合演算を加えたものである。このうち関係演算で記述される質問（選択・射影・結合質問）や結合をすべての共通属性に対する等号による結合に限った質問（自然結合質問）のクラスが注目される[6]。関係モデルとネットワークモデルの対応についてはいくつかの研究がある[5][7][8]が、本稿では簡単のため次のように定義する。

〔定義〕　ネットワークモデルの各レコード型に対し１つの関係，各親子集合型に対し親レコード型と子レコード型のキー属性からなる関係を作る。これにより２つのモデルの対応を定義する。□

**［補題４］** 関係モデルにおける自然結合質問は、結合される関係の属性集合の和集合に含まれるが結合されない関係がなければ、対応するネットワークモデルでは条件１を満たす質問スキーマが存在する結合質問となる。 □

**（証明）** 自然結合を表わす部分ネットワーク構造が存在しないのは、結合が情報損失となるときで、ネットワーク構造がその結合を表わす経路を持たないときのみである[9]。 このときは、結合される関係の属性集合の和集合に含まれるが結合されない関係が存在する。 □

普遍関係スキーム仮定の下での弱等価な質問[3] は、補題２中の関係は存在しないので、自然結合質問は対応するネットワークモデルでは条件１を満たす質問スキーマが存在する結合質問となる。従って自然結合質問はネットワークモデルでは比較可能巡航操作のみで解くことができる。

**［補題５］** 関係モデルにおける選択・射影・結合質問は、対応するネットワークモデルでは結合質問となる。 □

**（証明）** 結合演算は関係の直積に対する結合条件の制約演算で求めることができる。また、関係 $r_i = \pi_{ci}\ r_{i1} \otimes r_{i1} \otimes \cdots \otimes r_{ini}$（$i=1,2$, $\pi_{ci}$は条件がciである選択演算を，⊗は直積を表わす）に対し、

$$r_1 \otimes r_2 = \pi_{c1 \wedge c2}\ r_{11} \otimes \cdots \otimes r_{1n1} \otimes r_{21} \otimes \cdots \otimes r_{2n2}$$

となるので、すべての選択・射影・結合質問は関係の直積に対する選択，射影演算で表現できる。 □

共通集合演算による質問も関係の直積に対する制約で求めることができ、それを含む質問もネットワークモデルでは結合質問となる。和集合演算は各集合を独立に求めることによって実現でき、差集合演算は$r_1 \otimes r_2$で$r_1$の各組に対し$r_2$に制約条件を満たすものが存在しないときそれを解とすることで実現できる。

**［定理６］** 関係代数で記述される質問は、対応するネットワークモデルでは変数付巡航操作のみで解くことができる。 □

## ６．あとがき

ネットワークモデルにおける巡航操作による質問処理能力について検討した。比較可能巡航操作と同等の能力を持てば実用的な範囲で質問処理ができ、変数付巡航操作と同等の能力を持てば関係完備な質問を処理することができる。しかし質問の処理効率を考えると基本巡航操作で処理できる方がよく、データベース設

計時に考慮すべき問題である。また、処理コストを最小にする巡航順序の決定も残された問題である。

参考文献

[1] Dayal,U., Goodman,N., “Query Optimization for CODASYL Database Systems”, Proc. ACM SIGMOD Int. Conf. on Management of Data, pp.138-150, June 1982.
[2] Chen,H., Kuck,S.M., “Combining Relational and Network Retrieval Methods”, Proc. ACM SIGMOD Int. Conf. on Management of Data, pp.131-141, June 1984.
[3] Ullman,J.D., Principles of Database Systems, 2nd Edition, Computer Science Press, 1982.
[4] Maier,D., The Theorey of Relational Databases, Computer Science Press, 1983.
[5] 古川，上林，“ネットワークデータベースにおける木質問”，電子通信学会技術研究報告，AL84-62，1985年1月．
[6] 上林，“データベースの基礎理論（5）関係データベースにおける質問処理”，情報処理学会誌，第24巻，第3号，1983年3月．
[7] Dayal,U., Bernstein,P.A., “On the Updatability of Network View - Extending Relational View Theorey to the Network Model”, Information Systems, vol.7, no.1, pp.29-46, Jan. 1982.
[8] 滝沢，横塚，鈴木，“ＣＯＤＡＳＹＬデータベースシステムに対する関係インタフェースシステム（LDP-V1.5）の設計と実現”，情報処理学会論文誌，第23巻，第6号，1982年11月．
[9] Kambayashi,Y., Furukawa,T., “Semantic Constraints Expressed by Network Model”, Proc. Int. Conf. on Foundations of Data Organization, pp.201-206, May 1985.