関係データベースにおける全称記号を含む質問の一処理方式について

古川康一 （電子技術総合研究所）

1. はじめに

大規模データベースの質問応答を，人間にとってより使い易い，自然なものとするための研究が，自然言語の使用をも含めて，近来数多く行われている。[4),5),9),13)] これらは，質問応答機能の高度化と考えられるが，これを支える重要な技術は，データベースに関する知識の表現と，その利用である。データベースに関する知識としては，その内容に関する知識と構造に関する知識に大別されるが，前者は質問の意味の理解に役立ち，後者は答を取り出すためのデータベース・アクセス・プログラムの作成あるいは最適化に役立つ。

Coddは，関係モデルを導入して，データベースの構造を関係としてとらえることを提案した。彼は，さらに，関係モデルに対する質問言語として，関係論理および関係代数を導入した。関係論理は存在および全称限定子を許す非手続き的言語で関係代数は，集合演算を基にした関数的言語である。

関係データベースに対して推論能力を付加する試みがいくつかなされてきた。[2),4),7),8),11),14)] データベースの研究分野では，これはユーザに独自のデータの把握の仕方を許す user's view あるいは外部スキーマという概念でとらえられている。[6)] Emden[4)] は，関係の中の各組を事実の言明と考え，推論能力を付加した関係データベースを，PROLOGのプログラムとして与えた。これは関係データベースとHorn論理の結合と考えられる。Reiterはfunctionを許さない制限された述語論理の証明システムの中に，関係代数を埋め込む方式を開発した。[11)] さらにClosed World Assumptionの下で働く効率の良い質問処理アルゴリズムを与えた。[12)]

全称限定された質問の処理については，Coddが関係の除算に変換するアルゴリズムを与え，Palermo[10)] がその改善を行っている。Reiterは，その考えを一般化して，上述のfunctionを許さない述語論理の下での扱いを可能とした。[11)]

ここでは，Closed World Assumption の下で，全称限定された質問に含まれる含意の意味をそのまま解釈することによって，より効率の良い処理アルゴリズムを導入する。それは，SEQUEL 2[1)] などの手続き的言語に見られるGROUPBY 演算に相当する，関係の切断という操作を基にしている。

また，切断面の要素の数え上げによって，数量限定子の処理が可能であり，それが全称限定子の拡張としてとらえることができることを示す。

従来，推論能力を持った関係データベースでは，変数を含まない言明，すなわち外延データのみが関係ファイルに対応づけられ，変数を含む一般的な命題，すなわち内包データは，データベースの外部スキーマや言葉の定義などを表わすものとして，関係ファイルの外に置かれる知識ベースと対応づけられてきた。しかしながら，階層的データベースによって推移的な性質を表わすことができるように，同一の形式を持つ複数の内包データを形式化された関係ファイルによって，表現することが可能である。本論文では，データベースの一部を直積分解を用いて再構成することによって，内包データの一部を関係ファイル化することができ，それによって，全称限定された質問が効率よく処理し得ることを示す。

2. Closed World Assumption の下での質問処理

2.1. 諸定義

ここでは，Reiter [1978a] に従って，用語および記号の定義，ならびに Closed World Assumption における質問処理の概要について述べる。

定義　質問の形は，つぎの通りである。

$\langle x_1/\tau_1, \cdots, x_n/\tau_n \mid (q_1 y_1/\theta_1)\cdots(q_m y_m/\theta_m)\, W(x_1, \cdots, x_n, y_1, \cdots, y_m)\rangle$

あるいは $\langle \boldsymbol{x}/\boldsymbol{\tau} \mid (\boldsymbol{q}\boldsymbol{y}/\boldsymbol{\theta})\, W(\boldsymbol{x}, \boldsymbol{y})\rangle$．ここで q_i は∃または∀で，W は $x_1, \cdots, x_n, y_1, \cdots, y_m$ を自由変数とする論理式である。また，τ_i，θ_j はタイプであり，x_i，y_j のドメインを与える。（ドメイン自身は $|\tau_i|$，$|\theta_j|$ などで表わす）

定義　データベース（DB）は，関数記号を含まない節の集合である。

定義　$Q = \langle \boldsymbol{x}/\boldsymbol{\tau} \mid (\boldsymbol{q}\boldsymbol{y}/\boldsymbol{\theta})\, W(\boldsymbol{x}, \boldsymbol{y})\rangle$ とし，DB をデータベースとする。定数の n 組の集合 $\{\boldsymbol{c}^{(1)}, \cdots, \boldsymbol{c}^{(r)}\}$ が答である必要十分条件は，

1. $\boldsymbol{c}^{(i)} \in |\boldsymbol{\tau}|$, $i = 1, \cdots, r$ および
2. $DB \vdash \bigvee_{i \le r} (\boldsymbol{q}\boldsymbol{y}/\boldsymbol{\theta})\, W(\boldsymbol{c}^{(i)}, \boldsymbol{y})$．ここで $|\boldsymbol{\tau}|$ は，$|\tau_1| \times \cdots \times |\tau_n|$ である。

ここでの答 $\{\boldsymbol{c}^{(1)}, \cdots, \boldsymbol{c}^{(r)}\}$ は，「$\boldsymbol{c}^{(1)}$ か，あるいは $\boldsymbol{c}^{(2)}$ か… あるいは $\boldsymbol{c}^{(r)}$ のいずれかが $(\boldsymbol{q}\boldsymbol{y}/\boldsymbol{\theta})\, W(\boldsymbol{x}, \boldsymbol{y})$ を満たすが，どれとはいえない」ことを表わしているので，不確定な答（indefinite answer）である。また $\{\boldsymbol{c}^{(1)}, \cdots, \boldsymbol{c}^{(r)}\}$ が答ならば，任意の $\boldsymbol{c}$ に対して $\{\boldsymbol{c}^{(1)}, \cdots, \boldsymbol{c}^{(r)}, \boldsymbol{c}\}$ も答となるので，最小性の概念が必要となる。

定義　Q の答 A が最小であるのは，A の真部分集合で Q の答となるものがないときで，かつそのときに限る。

定義　Q の最小な答が唯ひとつの n 組より成るとき，A は Q の確定した答（definite answer）である。

定義　$\|Q\|_{OWA}$ は，Q の最小な答の集合である（OWA は，Open World Assumption を表わす）。

つぎに Closed World Assumption について述べる。それは，与えられた節集合から論理的に導くことのできない事実は，その否定が真であるとする仮定である。一般のデータベースでは，このことが成り立つことが多い。たとえば，列車の時刻表を考えると，余程の例外的な場合を除いて，時刻表に載っていない列車が運行されることはない。逆に言えば，データベースの利用者は，必要なすべての情報が常にデータベース中にあることを期待しており，そうでない場合には，データベースを利用する価値が薄れてくる。

定義　$\overline{EDB}$ は論理的に導けない言明の集合 $\{\overline{P\boldsymbol{c}} \mid P$：述語記号，$\boldsymbol{c}$ 定数ベクトル，$DB \nvdash P\boldsymbol{c}\}$ である（EDB は Extensional Data Base を表わす）。

定義　$\{\boldsymbol{c}^{(1)}, \cdots, \boldsymbol{c}^{(r)}\}$ が $\langle \boldsymbol{x}/\boldsymbol{\tau} \mid (\boldsymbol{q}\boldsymbol{y}/\boldsymbol{\theta})\, W(\boldsymbol{x}, \boldsymbol{y})\rangle$ の CWA での答である必要十分条件は，

1. $\boldsymbol{c}^{(i)} \in |\boldsymbol{\tau}|$, $i = 1, \cdots, r$ および
2. $DB \cup \overline{EDB} \vdash \bigvee_{i \le r} (\boldsymbol{q}\boldsymbol{y}/\boldsymbol{\theta})\, W(\boldsymbol{c}^{(i)}, \boldsymbol{y})$ である。

定義　$\|Q\|_{CWA}$ は Q の最小な CWA での答の集合である。

すべての最小な CWA での答は，確定していることが知られている。

OWA の下で矛盾のないデータベースは，必ずしも CWA の下で矛盾がないと

は限らない。たとえば，$|\tau| = \{A, B\}$，$DB = \{$犯人$(A) \vee$ 犯人$(B)\}$ を考えると，$\overline{EDB} = \{\overline{犯人(A)}, \overline{犯人(B)}\}$となり，$DB \cup \overline{EDB}$ は矛盾する。DBが矛盾のないHorn節集合の場合には，CWAの下でも矛盾がないことが知られている。

$\overline{EDB}$ はDBから論理的に導かれない事実全体であるので，一般にはDBと全く関係のない事柄を含んでいてもよく，それはいくらでも大きい集合になってしまう。ゆえに$\overline{EDB}$を使う推論は役に立たない。ところが，CWAの下で矛盾のないデータベースでは，$Q = \langle x/\tau \mid (\exists y/\theta)\, W(x, y) \rangle$において$W$が素式のとき，$\|Q\|_{CWA} = \|Q\|_{OWA}$となることが知られている。すなわち，素式のみで表わされる質問は$\overline{EDB}$を使わずにCWAでの答を求めることができる。

Wが素式以外のとき，Wを分解して，いくつかの素式の質問と，それらの間の集合演算に変換することができる。以下にその変換のための諸規則を列挙する。

<u>規則1.</u>（AND, OR の分解）

1. $\|\langle x/\tau \mid W_1 \wedge W_2 \rangle\|_{CWA} = \|\langle x/\tau \mid W_1 \rangle\|_{CWA} \cap \|\langle x/\tau \mid W_2 \rangle\|_{CWA}$
2. $\|\langle x/\tau \mid W_1 \vee W_2 \rangle\|_{CWA} = \|\langle x/\tau \mid W_1 \rangle\|_{CWA} \cup \|\langle x/\tau \mid W_2 \rangle\|_{CWA}$

<u>規則2.</u>（否定の除去）

1. $\|\langle x/\tau \mid \overline{W} \rangle\|_{CWA} = |\tau| - \|\langle x/\tau \mid W \rangle\|_{CWA}$
2. $\|\langle x/\tau \mid W_1 \wedge \overline{W_2} \rangle\|_{CWA} = \|\langle x/\tau \mid W_1 \rangle\|_{CWA} - \|\langle x/\tau \mid W_2 \rangle\|_{CWA}$

ただし，W, W_1, W_2は限定作用素を含まず，$DB \cup \overline{EDB}$が矛盾がないとする。

<u>規則3.</u>（存在記号の OR への分配）

$\langle x/\tau \mid (\exists y/\theta)(W_1 \vee W_2) \rangle = \langle x/\tau \mid (\exists y/\theta)\, W_1 \vee (\exists y/\theta)\, W_2 \rangle$

<u>規則4.</u>（存在記号の除去）

1. $\|\langle x/\tau \mid (\exists y/\theta)\, \overline{W} \rangle\|_{CWA} = \Pi_y(|\tau| \times |\theta| - \|\langle x/\tau, y/\theta \mid W \rangle\|_{CWA})$
2. $\|\langle x/\tau \mid (\exists y/\theta)[W_1 \wedge W_2] \rangle\| = \Pi_y(\|\langle x/\tau, y/\theta \mid W_1 \rangle\|_{CWA} \cap \|\langle x/\tau, y/\theta \mid W_2 \rangle\|_{CWA}$

ここで $\Pi_y = \Pi_{y_1} \Pi_{y_2} \cdots \Pi_{y_m}$で，$\Pi_{y_i}$は y_i に関する射影（projection）である。

<u>2.2. Join に基づく存在記号の除去</u>

規則4の2. では，$Q = \langle x/\tau \mid (\exists y/\theta)\, W_1 \wedge W_2 \rangle$ のとき，先ずW_1とW_2についてそれらを満すすべての(x, y)の組を求め，つぎにそれらの共通集合を求め，最後にyについての射影を行っている。ここで，W_1およびW_2には，変数yはすべて現われるが，変数xはすべて現われるとは限らない。$x = x_1 \cup x_2$で，W_1にはx_1のみ現われ，W_2にはx_2のみ現われるものとしよう（$x_1 \cap x_2 \neq \emptyset$でもかまわない）。$|\tau_i^c| = \Pi_{x_i^c}(|\tau|)$，$Q_i^c = \langle x_i/\tau, y/\theta \mid W_i \rangle$，$i = 1, 2$，とすると，$Q_i = \langle x/\tau, y/\theta \mid W_i \rangle$，$i = 1, 2$，のCWAの下での答の集合$\|Q_i\|_{CWA}$は，直積 $|\tau_i^c| \times \|\langle x_i/\tau, y/\theta \mid W_i \rangle\|_{CWA}$に等しいが，この計算は，両集合の要素数の積に比例する計算時間および記憶容量を必要とする。

ここで，このような直積を求めずに，$\|Q_i'\|_{CWA} = \|\langle x_i/\tau, y/\theta \mid W_i \rangle\|_{CWA}$から$\|Q\|_{CWA}$を求めることを考える。それは，Join 演算$*$を用いることにより達成される。

<u>定理1</u> $\|\langle x/\tau \mid (\exists y/\theta)(W_1 \wedge W_2) \rangle\|_{CWA}$

$= \Pi_y(\|\langle x_1/\tau, y/\theta \mid W_1 \rangle\|_{CWA} * \|\langle x_2/\tau, y/\theta \mid W_2 \rangle\|_{CWA})$

ここで$x_1 \cup x_2 = x$ で，x_iはxのうちW_iに現われるものとする。$*$は，変数yおよび$x_1 \cap x_2$ について行うものとする。

本論文では，諸定理の証明は，Appendix でまとめて記述する。

3. 全称記号の扱い

3.1. 全称記号の扱いを含んだ変換規則系

Codd の関係論理では，質問に現われる全称記号は関係代数の除算によって除去される。Reiterは，その方法を不確定な答を許す場合に拡張した[11]。CWA の下では，答は確定するので，そこではReiterの拡張された除算は，本質的にはもとの除算と同じである。ここでは，Reiter の記述に従って除算Δを定義しよう。

定義: $Q = \langle x/\tau,\ z/\psi \mid (\exists y/\theta)\ W(x, y, z) \rangle$ とする。$\|Q\|_{CWA}$ の z による商 $\Delta_z\|Q\|_{CWA}$ は，以下の性質を満たすn組 C の集合である：

$C \in \Delta_z\|Q\|_{CWA} \leftrightarrow$ すべての $a \in |\psi|$ に対して (c, a) は Q の CWA での答である。

Reiterは，CWAでの質問の変形規則に全称記号の扱いを含ませていないが，それを含ませることによって，規則の体系はよりすっきりする。

第1に，否定の除去を行う規則において，W, W1, W2, が限定作用素を含んではいけないという制約を除くことができる。もし $W = (\exists y/\theta) W_1$ ならば，$\overline{W} = (\forall y/\theta)\ \overline{W}_1$ となるので，

$$\| \langle x/\tau \mid (\forall y/\theta)\ \overline{W}_1 \rangle \|_{CWA} = |\tau| - \| \langle x/\tau \mid (\exists y/\theta)\ W_1 \rangle \|_{CWA}$$

となる。

第2に，規則3に双対な規則が存在する。それは，次の通りである。

規則3'.（全称記号のANDへの分配）

$$\langle x/\tau \mid (\forall y/\theta)(W_1 \wedge W_2) \rangle = \langle x/\tau \mid (\forall y/\theta) W_1 \wedge (\forall y/\theta) W_2 \rangle$$

第3に，すでにのべたように除算による全称記号の除去の規則を追加する。

規則5.（全称記号の除去）

$$\| \langle x/\tau \mid (\forall y/\theta) W(x, y) \rangle \|_{CWA} = \Delta_y \| \langle x/\tau,\ y/\theta \mid W(x, y) \rangle \|_{CWA}$$

ここで $\Delta_y = \Delta_{y_1} \cdot \cdots \cdot \Delta_{y_m}$ である。

変換規則系1〜5は，冗長である。それは，$(\forall y/\theta) W = \overline{(\exists y/\theta)\ \overline{W}}$ により，全称記号を存在記号と否定によって表わすことができるからである。このことは全称記号を除去する第2の方法としてつぎのような規則があることを示している。

規則6.（全称記号の除去－2）

$$\| \langle x/\tau \mid (\forall y/\theta) W(x, y) \rangle \|_{CWA} = |\tau| - \Pi_y (|\tau| \times |\theta| - \| \langle x/\tau,\ y/\theta \mid W(x, y) \rangle \|_{CWA})$$

規則6において，x の全ドメイン $|\tau|$ から Π_y 以下を差し引く代りに，初めから $\| \langle x/\tau \mid (\exists y/\theta) W(x, y) \rangle \|_{CWA} = \Pi_y \| \langle x/\tau,\ y/\theta \mid W(x, y) \rangle \|_{CWA}$ の範囲を考えれば十分である。このようにして，つぎの規則を得る。

規則6'（全称記号の除去－3）

$$\| \langle x/\tau \mid (\forall y/\theta) W(x, y) \rangle \|_{CWA} = \Pi_y R - \Pi_y (\Pi_y R \times |\theta| - R)$$

ここで $R = \| \langle x/\tau,\ y/\theta \mid W(x, y) \rangle \|_{CWA}$ である。

3.2. 切断に基づく全称記号の除去

全称記号によって限定された質問 $\langle x/\tau \mid (\forall y/\theta) W(x, y) \rangle$ は，実は含意をその中に含んでいる。すなわち，$\langle x/\tau \mid (\forall y)(y \in |\theta| \rightarrow W(x, y)) \rangle$ に等しい。この含意形の質問をそのまま解釈すると，$|\theta|$ の中の全ての y に対して W が真であるような x を求めればよいことになる。このために，集合 $R = \| \langle x/\tau,\ y/\theta \mid W(x, y) \rangle \|_{CWA}$ の x による切断 $\rho_x(R)$ およびその切断面 $\rho_x(R)(c)$ をつぎのように定義する。

定義　$R = \| \langle x/\tau,\ y/\theta \mid W(x, y) \rangle \|_{CWA}$ とするとき，R の x による切断 $\rho_x(R)$ は $\Pi_y R$ から $2^{\pi x R}$ への関数で，$C \in \Pi_y R$ に対して，

$$\rho_x(R)(C) = \{ d \mid (c,\ d) \in \| \langle x/\tau,\ y/\theta \mid W \rangle \|_{CWA} \}$$

となる。ここで $\rho_x(R)(\mathbb{C})$ を R の（x に関する）$\mathbb{C}$ における切断面という。

この切断は，SEQUEL 2[1] における GROUPBY の演算に相当している。すなわち，関係 R を x の値によってグループ分けしたときの，$x = u$ に対する y の値の集合が切断面 $\rho_x(R)(u)$ になっている。

切断面に対してつぎの補助定理が成り立つ。

補助定理 1. 関係 $R = \| \langle x/\mathbb{T},\ y/\theta \mid \mathbb{W}(x, y) \rangle \|_{CWA}$ の x に関する $\mathbb{C}$ における切断面を $\rho_x(R)(\mathbb{C})$ とすると，$d \in \rho_x(R)(\mathbb{C})$ であるための必要十分条件は，$DB \cup \overline{EDB} \vdash \mathbb{W}(\mathbb{C}, d)$ となることである。

補助定理 2. 論理式 $\mathbb{W}$ が x を含まなければ関係 R の x に関する切断 $\rho_x(R)$ は，定数 $\| \langle y/\theta \mid \mathbb{W}(y) \rangle \|_{CWA}$ を与える定数関数となる。

関係の切断を用いて全称記号を除くことができる。

定理 2. $DB \cup \overline{EDB}$ が矛盾がなければ，次の式が成り立つ：

$$\| \langle x/\mathbb{T} \mid (\forall y/\theta)\ \mathbb{W}(x, y) \rangle \|_{CWA} = \{ \mathbb{C} \mid |\theta| \subset \rho_x(\| \langle x/\mathbb{T},\ y/\theta \mid \mathbb{W} \rangle \|_{CWA})(\mathbb{C}) \}.$$

この定理は，切断と包含テストによって除算が実現できることを示しているが，さらに $\mathbb{W}$ 自身が含意命題の場合には，切断を用いて質問を $\mathbb{W}$ の成分の質問に分解することが可能となる。

定理 3. $DB \cup \overline{EDB}$ が矛盾がなければ，含意質問 $Q = \langle x/\mathbb{T} \mid (\forall y/\theta)(\mathbb{W}_1 \to \mathbb{W}_2) \rangle$ の答は，次の式で与えられる：

$$\| Q \|_{CWA} = (|\mathbb{T}| - \| \langle x/\mathbb{T} \mid (\exists y/\theta)\ \mathbb{W}_1 \rangle \|) \cup \{ \mathbb{C} \mid \emptyset \neq \rho_x(R_1)(\mathbb{C}) \subset \rho_x(R_2)(\mathbb{C}) \}.$$

ここで $R_i = \| \langle x/\mathbb{T},\ y/\theta \mid \mathbb{W}_i \rangle \|$，$i = 1, 2$，である。

この定理の式の第1項は，$\mathbb{W}_1$ を満たす y が存在しない場合に対応している。たとえば，"PEN を販売していない部門" を求めるとき，$\mathbb{W}_1$ を "部門 x が y を販売している" とし，$\mathbb{W}_2$ を "y はペンでない" とすれば，質問は $Q = \langle x/部門 \mid (\forall y/品目)(\mathbb{W}_1(x, y) \to \mathbb{W}_2(y)) \rangle$ と表わせるが，全く品物を販売していない部門，たとえば人事部門や企画部門などが，この第1項の答に相当する。

全称記号により限定された含意命題を，除算とその他の変換規則で処理すると，

$$\| \langle x/\mathbb{T} \mid (\forall y/\theta)(\mathbb{W}_1 \to \mathbb{W}_2) \rangle \|_{CWA} = \Delta y \{ (|\mathbb{T}| \times |\theta| - \| \langle x/\mathbb{T},\ y/\theta \mid \mathbb{W}_1 \rangle \|_{CWA}) \cup \| \langle x/\mathbb{T},\ y/\theta \mid \mathbb{W}_2 \rangle \|_{CWA} \}$$

となる。この式では，$|\mathbb{T}| \times |\theta|$ の計算が含まれるが，この計算量は，時間，記憶量とも $|\mathbb{T}|$ と $|\theta|$ の要素数の積に比例するので，好ましくない。一方，定理 3 による計算では，直積は現われない。切断を求めるアルゴリズムは，*Sorting* を基にした方法や *Hashing* を基にした方法が考えられる。ハッシングを基にしたアルゴリズムによれば，$|R_1|$ および $|R_2|$ に比例した計算時間および記憶量のみを必要とするので，定理 3 による方法の方がずっと能率がよい。

実際の質問では，$\mathbb{W}_1$ は x を含まないことが多い。その場合には，$\rho_x(R_1)$ は定数関数となり，定理 3 はつぎのように簡単化される。

補題 3.1. $DB \cup \overline{EDB}$ が矛盾がなく，質問 $Q = \langle x/\mathbb{T} \mid (\forall y/\theta)(\mathbb{W}_1 \to \mathbb{W}_2) \rangle$ において $\mathbb{W}_1$ が x を含まないとき，$\rho_x(R_1)(\mathbb{C}) = S_1$ とすると，

$$\| Q \|_{CWA} = \begin{cases} \{ \mathbb{C} \mid S_1 \subset \rho_x(R_2)(\mathbb{C}) \} & \textit{if } S_1 \neq \emptyset \\ |\mathbb{T}| & \textit{otherwise.} \end{cases}$$

この式を定理 2 の式と較べてみることにより，$\| Q \|_{CWA}$ が除算によって求められることがわかる。さらに除算自身を，集合の包含判定を行わずに，要素の集合に対する所属判定および集合の要素の数え上げによって，能率よく行うことができる。

補題 3.2. 補題 3.1 と同じ条件の下で，

$\|Q\|_{CWA} = \{ c \mid |S_1| = |S_1 \cap g_x(R_2)(c)| \}$.

これは等式 $(A \subset B) \equiv (|A| = |A \cap B|)$ を用いて容易に証明できる。ここで，$|A|$ は集合Aの要素の数である。$g_x(R_2)(c) \cap S_1$ は，R_2 の切断を求める際に切断面の要素として S_1 に属しているもののみをとればよい。

この要素を数え上げる方法は，"2つ以上の"，"丁度3つの"，"半数以上の"などの数量限定子の処理にも応用できることが容易に分かる。すなわち，これらの限定子は，全称限定子の拡張と考えられて処理される。たとえば，"すべての階で売られている品物"は，階の総数をnとすると，"n個（以上）の階で売られている品物"と同じわけである。

$g_x(R_2)$が定数関数となる場合にも，全称限定子および数量限定子の効率の良い処理方法が考えられるが，ここでは詳細を略す。

3.3. 質問処理の例

全称記号によって限定された質問の処理例を見るために，図1のような関係データベースを考えよう。この例は palermo[1974] から採ったものである。関係SUPPLY(SU# PA# PJ#)は，SU# の納入業者が，部品PA#をプロジェクトPJ#に納入したことを示している。SUPPLIER（SU# SULOC SUNAME）は，納入業者 SU# の場所SULOCと名前 SUNAMEの情報を持っている。PROJECT(PJ# PJLOC PJNAME)は プロジェクト PJ#の場所PJLOCと名前PJNAMEを表わしている。PART(PA# PATYPE)は，部品PA#の型PATYPEを示している。

SUPPLY	SU#	PA#	PJ#
	211	31	971
	325	32	971
	211	33	970
	211	31	972
	237	31	970
	237	32	970
	237	33	970
	237	32	971
	237	31	971
	237	31	972

SUPPLIER	SU#	SULOC	SUNAME
	211	NY	AA
	325	SF	XX
	237	LA	YY

PROJECT	PJ#	PJLOC	PJNAME
	970	POK	A
	971	SJ	X
	972	SJ	Y

PART	PA#	PATYPE
	31	A
	32	A
	33	B

図1．関係データベースの例

このデータベースに対して次のような質問を考えよう。

質問　San Jose にあるプロジェクトの少くともいずれか1つに，型Aのすべての部品を少くとも1つづつ納めている納入業者の名前と場所を求めよ。

$$
\begin{aligned}
Q = \langle\, x/SUNAME,\ y/SULOC \mid\ & (\exists s/SU\#)(SUPPLIER(s, y, x) \\
& \wedge (\exists j/PJ\#)(PROJECT(ST, j) \\
& \wedge (\forall p/PA\#)(PART(p, A) \rightarrow SUPPLY(s, p, j)))) \,\rangle
\end{aligned}
$$

$\|Q\|_{CWA}$を求める手順を以下に示す。

1. $Q_1 = < x/SUNAME,\ y/SULOC,\ s/SU\# \mid SUPPLIER(s, y, x) >$

$Q_2 = < s/SU\# \mid (\exists j/PJ\#)(PROJECT(SJ, j) \wedge (\forall p/PA\#)(PART(p, A) \rightarrow SUPPLY(s, p, j))) >$

とすると，定理1より

$\| Q \|_{CWA} = \Pi_s(\| Q_1 \|_{CWA} * \| Q_2 \|_{CWA})$

となる。ここで Join * は s について行う。

2. $Q_{21} = < j/PJ\# \mid PROJECT(SJ, j) >$

$Q_{22} = < s/SU\#,\ j/PR\# \mid (\forall p/PA\#)(PART(p, A) \rightarrow SUPPLY(s, p, j)) >$

とすると同じく，定理1より

$\| Q_2 \|_{CWA} = \Pi_j(\| Q_{21} \|_{CWA} * \| Q_{22} \|_{CWA})$

となる。ここで Join * は，j について行う。

3. $S_1 = \| < p/PA\# \mid PART(p, A) > \|_{CWA}$

$S_2 = \| < s/SU\#,\ j/PJ\#,\ p/PA\# \mid SUPPLY(s, p, j) > \|_{CWA}$

とする。$S_1 \neq \phi$ なので補題 3.2 より

$\| Q_{22} \|_{CWA} = \{ s_1, j_1 \mid |S_1| = |S_1 \cap \rho_{(s,j)}(S_2)(s_1, j_1)| \}$

実行の経過は図2のようになる。

1. $\| Q_1 \|_{CWA} =$

SU#	SULOC	SUNAME
211	NY	AA
325	SF	XX
237	LA	YY

--→ 1B. $\| Q \|_{CWA} = \Pi_s(\| Q_1 \|_{CWA} * \| Q_2 \|_{CWA})$

⇓

2. $\| Q_{21} \|_{CWA} =$

PJ#
971
972

--------→ 2B. $\| Q_2 \|_{CWA} = \Pi_j(\| Q_{21} \|_{CWA} * \| Q_{22} \|_{CWA})$

⇓

3. $S_1 =$

PA#
31
32

, $|S_1| = 2$

⇓

$S_2 =$ SUPPLYの表全体

s_1 SU#	j_1 PJ#	P#	$\rho_{(s,j)}(S_2)(s_1, j_1) \cap S_2$ 要素数
211	971	31	1
211	972	ϕ	0
211	970	31	1
325	971	32	1
237	970	31 32	2
237	971	32 31	2
237	972	31	1

$\| Q_{22} \|_{CWA} =$

SU#	PJ#
237	970
237	971

⇓

2B. $\| Q_2 \|_{CWA} =$

SU#
237

⟹ 1B. $\| Q \|_{CWA} =$

SULOC	SUNAME
LA	YY

図 2. 質問例の処理過程.

この処理過程は，より最適化することができる。第1に，ある条件の下で，JOIN $*$ と除算の演算順序を逆にすることができる。この例では，Q_{21} と Q_{22} のJOIN を行うかわりに，Q_{21} と S_2 のJOINを行って，その結果を S_1 で割ってもよい。

4. 関係データベースの直積分解

前節の例で，SU# 237 の納入業者が型Aのすべての部品をプロジェクト 971 に納入していることを，別の新たな表で表現することもできる。いま その表を SUPPLYALL と名づけよう。SUPPLYALL は，SU#, PTYPE, PR# の各欄から成る。この表には，納入業者 211 がプロジェクト 970 に型Bのすべての部品を納入している事実も同様におかれている。この表の各組は，述語 PART と SUPPLY との間につぎのような関係が成り立つ。

$$(\forall s/SU\#)(\forall j/PR\#)(\forall p/PA\#) \tag{4.1}$$
$$((\exists t/PTYPE)\ SUPPLYALL(s, t, j) \wedge PART(p, t) \rightarrow SUPPLY(s, p, j))$$

ここで前節の例の中の，Q_{22} を求める部分を考えてみよう。

$$Q_{22} = \langle s/S\#, j/PR\# \mid (\forall p/PA\#)(PART(P, A) \rightarrow SUPPLY(s, P, j)) \rangle$$

であるが，$SUPPLY(s, p, j)$ は，関係 SUPPLY からだけでなく，(4.1) 式より，関係 SUPPLYALL からも求められる。いま，$SUPPLY(s, p, j)$ が関係 SUPPLYALL と関係 SUPPLY のみから決まり，$PART(p, t)$ が関係 PART のみから決まるとすると，(4.1) より次式が成り立つ。

$$(\forall s/SU\#)(\forall p/PA\#)(\forall j/PJ\#) \tag{4.2}$$
$$SUPPLY(s, p, j) \equiv (\exists t/PATYPE)(SUPPLYALL(s, t, j) \wedge PART(p, t)) \vee SUPPLY(s, p, j)$$

このとき

$$Q_{22} = \langle s/SU\#, j/PJ\# \mid (\forall p/PA\#)(PART(p, A) \rightarrow$$
$$(\exists t/PATYPE)(SUPPLYALL(s, t, j) \wedge PART(p, t)) \vee SUPPLY(s, p, j)) \rangle$$

となるが，$\|Q_{22}\|_{CWA}$ の中には，関係 SUPPLYALL から得られる解 $\|Q_{221}\|_{CWA} = \|\langle s/SU\#, j/PJ\# \mid \underline{SUPPLYALL}(s, A, j) \rangle\|_{CWA}$ が含まれるべきである。このことを以下で示そう。

(4.3) 式は，補題 3.1 の形になっているので，除算によって解が求められる。また，$(\exists t/PATYPE)$ で始まる項で $t = A$ とすると，$\underline{SUPPLYALL}(s, A, j) \wedge \underline{PART}(P, A)$ となるが，この項の $\underline{SUPPLYALL}$ と $\underline{PART}$ は変数を共有していないので，この式を満足する p, s, j の組の集合 P_{psj} は，直積 $\|\langle P/PA\# \mid \underline{PART}(p, A) \rangle\|_{CWA} \times \|\langle s/SU\#, j/PJ\# \mid \underline{SUPPLYALL}(s, A, j) \rangle\|_{CWA}$ に等しくなる。この集合 R_{psj} は，上に述べた除算での被除数の一部になっており，それを除数 $\|\langle P/PA\# \mid \underline{PART}(P, A) \rangle\|_{CWA}$ で割れば，確かに $\|Q_{221}\|_{CWA}$ を商として得る。

一般には，つぎの定理が成り立つ。

<u>定理4</u>　$\bar{W}_1(y) = \bar{W}_1'(y, e)$ で，y が $\bar{W}_1'(y, z)$ に対応する関係 R_1 の主キーなら，

$$Q = \langle x/\tau \mid (\forall y/\theta)(\bar{W}_1(y) \rightarrow ((\exists z/\psi)(\bar{W}_1'(y, z) \wedge \bar{W}_2(x, z)) \vee \bar{W}_3(x, y))) \rangle$$

の CWA での答は，次式によって求められる：

$$\|Q\|_{CWA} = \|\langle x/\tau \mid \bar{W}_2(x, e) \rangle\|_{CWA} \cup \|\langle x/\tau \mid (\forall y/\theta)(\bar{W}_1(y) \rightarrow \bar{W}_3(x, y)) \rangle\|_{CWA}$$

関係 SUPPLYALL は，関係 SUPPLY の一部分を直積の形に分解して得られたものと考えることができる。いま型Aの部品のみについて考えると

$\|\langle s/SU\#, j/PJ\# \mid SUPPLYALL(s, A, j) \rangle\|_{CWA} \times \|\langle P/PA\# \mid PART(P, A) \rangle\|_{CWA}$ は，型Aのすべての部品を納入している業者 s と納入先のプロジェクト j および部品番号 p を与えているので，もともとは関係 $SUPPLY(s, P, j)$ に置かれている情

報である。この関係SUPPLYALL(s, t, j)と，(4.1)式によって，もとの関係SUPPLYの部分が復元できる。SUPPLYALLは，他の関係と異なり，(4.1)式を見れば，明らかなように，内包データの一部分を構成している。つまり，SUPPLYALLの各組は，あるひとかたまりの情報を表わしている。たとえば，SUPPLYALL(s_0, t_0, j_0)が表わしているものは{s_0/SU#, p/PA#, j_0/PJ#|SUPPLY(s_0, j_0, p) ∧ PART(p, t_0)}である。

このように，関係データベースに直積分解を導入することは，内包データの一部を関係によって表現することに対応している。

5. おわりに

全称限定された質問は，実際の場面ではそれほど頻繁に現われるとは思われず，その割には処理が大変なので，多くのQAシステムでその機能が実現されていない。しかしながら，本論文で見たように，その処理手法は数量限定子の処理にも有効であり，それらの統一的な扱いは，システムを簡潔にするのに極めて役立つ。

直積分解によるデータ表現は，表現の簡約化の点で記憶効率の向上に役立つ。全称限定された質問の処理は，数式処理的な考えで処理できることを示したが，存在限定された質問の処理は，逆に直積を作ることが必要となる。あるいは，もっと高度な方法としては，関係上にある内包的な表現をそのまま出力として用いることも考えられる。

本論文で与えた処理方式を，すでに開発したLISP高水準データベースGEPHERDの上に実現することを計画している。

最後に，本研究の機会を与えていただいた石井治ソフトウェア部長，棟上昭男情報システム研究室室長，議論をしていただいた同研究室の諸氏に感謝致します。

Reference

1. Chamberlin, D. D. et. al. (1976) "SEQUEL 2: A Unified Approach to Data Definition, Manipulation and Control," IBM J. Res. Develop. Vol. 20, No. 6, 1976, 560-575.
2. Chang, C. L. (1978) "DEDUCE 2: Further Investigations of Deduction in Relational Data Bases," in Logic and Data Bases, (Ed. H. Gallaire and J. Minker), Plenum Press, 1978, 201-236.
3. Codd, E. F. (1972) "Relational Completeness of Data Base Sublanguages," in Data Base Systems, (Ed. R. Rustin), Prentice-Hall, 1972, 65-98.
4. Furukawa, K. (1977) "A Deductive Question Answering System on Relational Data Bases," Proc. Fifth IJCAI, 1977, 59-66.
5. Harris, L. R. (1977) "User Oriented Data Base Query with the ROBOT Natural Language Query System," Proc. Third VLDB, 1977, 303-311.
6. Interim Report ANCI/SPARC Study Group on Data Base Management Systems, "FDT"7, No. 2, ACM New York, 1975.
7. Kellog, C. et. al. (1978) "Deductive Planning and Pathfinding for Relational Data Bases," in Logic and Data Bases, 1978, 179-200.
8. Minker, J. (1978) "An Experimental Relational Data Base Systems on Logic," in Logic and Data Bases, 1978, 107-148.
9. Ohsuga, S. (1977) "Semantic Information Processing in Man-Machine Systems," Proc. 1977 IEEE Conf. on Decision and Control, 1977.
10. Palermo, F. P. (1974) "A Data Base Search Problem," in Information Systems, (Ed. J. T. Tou), Plenum Press, 1974.
11. Reiter, R. (1978a) "Deductive Question-Answering on Relational Data Bases," in Logic and Data Bases, 1978, 149-178.
12. Reiter, R. (1978b) "On Closed World Data Bases," in Logic and Data Bases, 1978, 55-76.
13. Sacerdoti, E. D. (1977) "Language Access to Distributed Data with Error Recovery," Proc. Fifth IJCAI, 1977, 196-202.
14. van Emden, M. H. (1977) "Computation and Deductive Information Retrieval," Proc. IFIP Working Conf. on Programming Concept, 1977.

Appendix

ここでは、紙面の都合上，定理１および定理３の証明を行う。他の定理の証明も同様にしてできる。

定理1. $\|\langle x/\tau \mid (\exists y/\theta)(W_1 \wedge W_2)\rangle\|_{CWA}$

$$= \Pi_y(\|\langle x_1/\tau_1, y/\theta \mid W_1\rangle\|_{CWA} * \|x_2/\tau_2, y/\theta \mid W_2\rangle\|_{CWA})$$

ここで $x_1 \cup x_2 = x$ で，x_i は x のうち W_i に現われるものとする。Join $*$ は，変数 y および $x_1 \cap x_2$ について行うものとする。

証明. $Q=\langle x/\tau \mid (\exists y/\theta)(W_1\wedge W_2)\rangle$, $Q_i=\langle x_i/\tau_i, y/\theta \mid W_i\rangle$, $i=1,2$ とする。

(1). $c \in \|Q\|_{CWA}$ とする。c の中で x_1, x_2 に対応するものを，それぞれ c_1, c_2 とする。$DB\cup\overline{EDB} \vdash (\exists y/\theta)(W_1(c,y)\wedge W_2(c,y))$ より，$d\in|\theta|$ が存在して，$DB\cup\overline{EDB}\vdash W_1(c,d)\wedge W_2(c,d)$ となる（Reiter（1978b））。ところが $W_i(c,d)=W_i(c_i,d)$ であるので，$DB\cup\overline{EDB}\vdash W_i(c_i,d)$, $i=1,2$ となる。(c_1,d) と (c_2,d) の Join は (c,d) で，$\Pi_y(\{c,d\}) = c$ であるので，$c\in\Pi_y(\|Q_1\|_{CWA} * \|Q_2\|_{CWA})$ が成り立つ。

(2). $c\in\Pi_y(\|Q_1\|_{CWA}*\|Q_2\|_{CWA})$ とする。$d\in|\theta|$ が存在して $(c,d)\in\|Q_1\|_{CWA} * \|Q_2\|_{CWA}$ となる。ところが $\|Q_i\|$ は変数 x_i と y に対応するので $(c_i,d)\in\|Q_i\|_{CWA}$ である。ゆえに $DB\cup\overline{EDB}\vdash W_i(c_i,d)=W_i(c,d)$, $i=1,2$ である。したがって $DB\cup\overline{EDB}\vdash W_1(c,d)\wedge W_2(c,d)$ となり，$c\in\|Q\|$ が成り立つ。

(1)と(2)から，$\|Q\|_{CWA}=\Pi_y(\|Q_1\|_{CWA} * \|Q_2\|_{CWA})$ がいえる。

定理3. $DB\cup\overline{EDB}$ が矛盾がなければ，含意質問 $Q=\langle x/\tau \mid (\forall y/\theta)(W_1\to W_2)\rangle$ の答は，次の式で与えられる：

$$\|Q\|_{CWA} = (|\tau| - \|\langle x/\tau \mid (\exists y/\theta) W_1\rangle\|_{CWA}) \cup \{c \mid \phi \neq P_x(R_1)(c) \subset P_x(R_2)(c)\}.$$

ここで $R_i = \|\langle x/\tau, y/\theta \mid W_i\rangle\|_{CWA}$，$i=1,2$ である。

証明. 右辺の第1項を S_1 とし，第2項を S_2 とする。

(1) $c\in\|Q\|_{CWA}$ とする。

case1. $DB\cup\overline{EDB}\nvdash(\exists y/\theta)W_1(c,y)$ のとき。

$DB\cup\overline{EDB}\vdash\overline{(\exists y/\theta)W_1}(c,y)$ であるので，$c\in|\tau| - \|\langle x/\tau \mid (\exists y/\theta)W_1\rangle\| = S_1$.

case2. $DB\cup\overline{EDB}\vdash(\exists y/\theta)W_1(c,y)$ のとき.

仮定より，$d\in|\theta|$ が存在して $DB\cup\overline{EDB}\vdash W_1(c,d)$. ゆえに $P_x(R_1)(c)\neq\phi$（補助定理1より）. $DB\cup EDB\vdash(\forall y/\theta)(W_1(c,y)\to W_2(c,y))$ より，すべての $d\in P_x(R_1)(c)$ に対して $DB\cup\overline{EDB}\vdash W_2(c,d)$，すなわち $d\in P_x(R_2)(c)$，が成り立つ. ゆえに $\phi\neq P_x(R_1)(c)\subset P_x(R_2)(c)$ となり，$c\in S_2$ がいえる。

(2-1) $c\in S_1$ とする. $DB\cup\overline{EDB}\vdash\overline{(\exists y/\theta)W_1(c,y)} = (\forall y/\theta)\overline{W_1}(c,y)$ となるので，$DB\cup\overline{EDB}\vdash(\forall y/\theta)(\overline{W_1}(c,y)\vee W_2(c,y)) = (\forall y/\theta)(W_1(c,y)\to W_2(c,y))$ が成り立つ. ゆえに $c\in\|Q\|$ がいえる.

(2-2) $c\in S_2$ とする。$P_x(R_1)(c)\subset P_x(R_2)(c)$ が成り立つので，$DB\cup\overline{EDB}\vdash W_1(c,d)$ であるすべての $d\in|\theta|$ に対して，$DB\cup\overline{EDB}\vdash W_2(c,d)$ となる（補助定理1より）. したがって，すべての $d\in|\theta|$ に対して $DB\cup\overline{EDB}\vdash W_1(c,d)\to W_2(c,d)$ が成り立ち，$c\in\|Q\|_{CWA}$ がいえる。

(1), (2-1), (2-2) より，$\|Q\|_{CWA}=S_1\cup S_2$ がいえる。