

CODASYL DML に対する非手続的グラフ問合せ言語の設計と実現†

滝 沢 誠†† 野 口 正 一†††

本論文では、CODASYL モデルを提供する CODASYL DBS を概念、論理、仮想、物理の4階層に分け、論理層と仮想層の論理構造と、層間写像とについて考察する。論理層の論理データ構造は、集合と二つの集合間の部分関数とで表現され、論理操作言語は述語論理形式の CODASYL 問合せ言語 CQL とこれと等価な CODASYL 問合せグラフ (CQG) とで定義される。CQG は論理操作演算を見やすい形式で表せるので利用者言語として優れている。仮想層では、論理データ構造に新たに順序関係が導入されて、仮想データ構造が定まる。この上の仮想操作演算 VOP が従来の COBOL DML に対応したものとして与えられる。本論文では、この意味を仮想的な CODASYL 機械 (VCM) として、多テープチューリング機械を用いて表し、DML の意味を明らかにする。概念層の任意の CQG を VCM 上の演算、すなわち COBOL DML プログラムで表せることを示す。さらに、各 CQG に対して、中間ファイルを不要とし、ファイル操作として順次書き込みだけで済む DML が存在することを明らかにし、その構成方法を示す。本論文は、従来の CODASYL DBS 上に非手続的な利用者インタフェースを設計するための論理的基礎を与えている。

1. はじめに

CODASYL モデル^{1),2)}を提供するデータベースシステム (DBS) を CODASYL DBS とする。CODASYL DBS は、COBOL DML²⁾等の手続的操作言語を用いて、高性能性が求められる大型の定型業務に広く用いられてきていた。しかし、最近、一般利用者がデータベースを非定型的に利用するための非手続的インタフェースが求められてきている。このような非手続的インタフェースを設計するためには、まず従来の CODASYL モデルの論理的な構造を明確にする必要がある。しかし、現在まで、CODASYL モデルに対する十分な論理的考察はほとんどなされていない。わずかに、DML の意味について、文献 13) が状態遷移により、14) がデータ抽象化の公理的方法により意味記述を試みている程度である。また、述語論理形式の非手続的操作演算を手続的 DML に変換する問題は、限定された形式の検索演算について文献 5)~7), 12) が、更新演算については文献 8) が論じている。

本論文では、①CODASYL モデルの論理的階層構造を明確に与え、②このもとで、DML の意味を明らかにし、③自由な形式の非手続的検索演算を満足する DML プログラムの構成方法を与える。②として、従

来の COBOL DML²⁾の意味を多テープチューリング機械の動作として与えることで、DML の明確な意味を与える。③として、まず、従来の CODASYL モデルを抽象化して得られる集合と部分関数とで与えられるデータ構造に、述語論理形式の検索言語 CODASYL 問合せ言語 CQL と、そのグラフ表現 CQG (CODASYL 問合せグラフ) とを与える。次に、CQG の条件を満足する DML プログラムが存在することと、目標集合を格納するための出力ファイル数を最小化する DML プログラムを構成できることを示す。本論文の方法は、文献 5)~7), 12) に対して、①完備した検索演算を備えた CQG を利用者に提供でき、②CQG の条件を満足する目標集合を、一つの物理的な出力ファイルだけを用いた DML プログラムで求められる特徴をもつ。

以上を検討するうえで、従来の CODASYL モデルとして、文献 15) 等も考えられるが、その論理的に本質の問題を検討するには、文献 1)と 2)で十分と考える。

2章では、従来の CODASYL モデルの論理階層を明確にする。3章と4章ではおのおの、論理階層のなかの CQL を提供する概念層と DML を提供する仮想層とを定式化する。5章では、CQLとDMLとの関係を論じ、6章では、LC モデル⁸⁾と DML との対応を示す。

2. 従来の CODASYL モデルの論理階層

従来の CODASYL モデル^{1),2)}は、物理的側面と論理的側面の議論が不十分であり、専門家向けの COB-

† Design and Implementation of Non-procedural Graph Query Language for CODASYL DML's by MAKOTO TAKIZAWA (Japan Information Processing Development Center) and SHOICHI NOGUCHI (Research Institute of Electrical Communication, Tohoku University).

†† (財)日本情報処理開発協会

††† 東北大学電気通信研究所

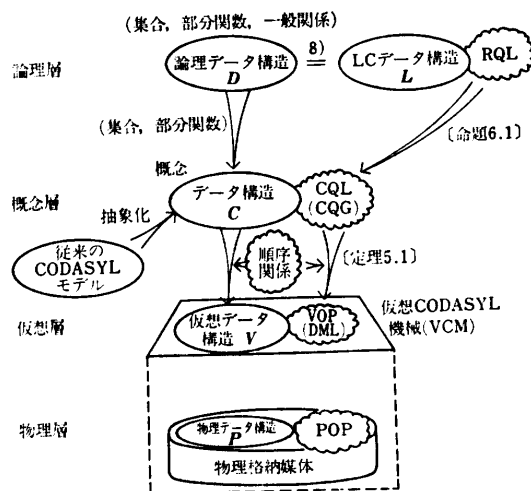


図1 CODASYL DBS の階層構造
Fig. 1 Hierarchy of the CODASYL DBS.

OL DML²⁾等の手続的操作言語を提供している。しかし、一般利用者には、論理的なデータ構造を非手続的に操作することが望ましい。このためには、物理と論理的側面を分離し、DMLの操作層と非手続的な言語の操作層との関係を明らかにする必要がある。

本論文では、CODASYLモデルは、論理、概念、仮想、物理の4層の階層構造でモデル化できることを示す(図1)。まず、従来のCODASYLデータ構造¹⁾の論理構造を抽象化すると、集合(レコード(R)型)と集合間の部分関数(セット(S)型)とで与えられるデータ構造が導ける。このデータ構造上に、述語論理形式の非手続的操作言語を定めることができる。このデータ構造と言語で定まるモデルを、概念CODASYLモデルとする。このモデルは、論理的データ構造と非手続的操作演算を利用者に提供するもので、3章で定式化する。さらに、概念CODASYLデータ構造を一般化すると集合間の一般関係(リンク(L)型)が定まる*。集合、部分関数に加えてこの一般関係が定められたデータ構造を、概念CODASYLデータ構造上の論理CODASYLデータ構造とする。このデータ構造は、文献8)で定義された関係型のローカル概念(LC)データ構造⁹⁾との間に一对一の対応関係が与えられ、詳細は文献8)で示されている。

次に、概念CODASYLデータ構造の集合と関係の元に、ある規準によって定まる順序関係を与える。こ

れにより、このデータ構造の操作として、与えられた順序関係に基づいて各演算が逐次行われることになる。この演算が従来のCOBOL DML²⁾に対応している。このデータ構造と操作演算で定まるモデルを、概念CODASYLモデルの一つ論理的下位の仮想CODASYLモデルとし、4章で定式化する。

次に、仮想CODASYLモデルを物理格納装置上に実現するモデルとして、物理CODASYLモデルが定義される。

以上の論理階層を設けることによって、従来のCODASYLモデルの論理構造が明確となり、これをもとにしてCODASYL DBS上の非手続的利用者インタフェース設計の論理的基礎とすることができる。

3. 概念 CODASYL モデル

本章では、概念CODASYLモデルのデータ構造と操作演算とを定義する。

3.1 概念データ構造

概念データ構造Cは、時間不変な論理構造を与える概念スキーマCと、これに実際のデータを与えた(時間可変な)概念データベースCとから成る。Cはレコード(R)型とセット(S)型とから成る。R型Rは項目集合 $Q_R = \{@R, t_1, \dots, t_m\}$ ($m \geq 0$)とインテグリティ条件 Γ_R とから成る($R = (Q_R, \Gamma_R)$)。各項目 $t \in Q_R$ のとりうる値の集合(定義域)を $\text{dom}(t)$ とする。 $@R$ はdbキーで、 $\text{dom}(@R)$ はRの識別子集合である。各 $t_i \in Q_R$ はデータ項目である。 Γ_R は、 Q_R がとりうる値の組(レコード(R)実現値)を定める述語で、この組の集合をレコード実現値(RO)集合 $R = \{o \mid o \in \text{dom}(\bigcap_{i=1}^m Q_R) \wedge \Gamma_R(t)\}$ (ここで、 $\text{dom}(Q_R) = \text{dom}(@R) \times \text{dom}_{i=1}^m (t_i)$)とする。各 $o \in R$ の意味のある部分項目集合 $I(\subseteq Q_R)$ の値を $I(o)$ とする。各 $o \in R$ のdbキー $@R$ の値は、Rのみならず他のRO集合内でも一意であるので、 $@R$ はRの主キーと考えられ、 $@R(o)$ を $\delta(o)$ と表す。また利用者にはdbキー情報が知られないので、R型を $R(t_1, \dots, t_m)(\Gamma_R \text{は省略})$ とも表す。各Rについて、次のR述語 $P_R: \text{dom}(Q_R) \rightarrow \{\text{true}, \text{false}\}$ を定める。

$$P_R(o) = \text{true if } o \in R, \text{ false otherwise} \quad (1)$$

Cでは、二つのR型間の関係として、セット(S)型が定義される。すなわち、部分関数 $f_s: \text{dom}(Q_{R_1}) \rightarrow \text{dom}(Q_{R_2})$ が定義されているとき、S型 $S = (R_1, R_2, \Gamma_s)$ と表す。 R_1 と R_2 は、おのおの f_s の値域と定義域を与えるR型であり、親と子R型とする。 Γ_s は

* 概念CODASYLデータ構造のなかで、複数のR型 $R_1, \dots, R_n$ と、共通の子RをもつS型 $S_1, \dots, S_m$ が存在するとき、新たにRと $S_1, \dots, S_m$ をまとめて、 $R_1, \dots, R_n$ 間の一般関係を与えるリンク(L)型を定義する。

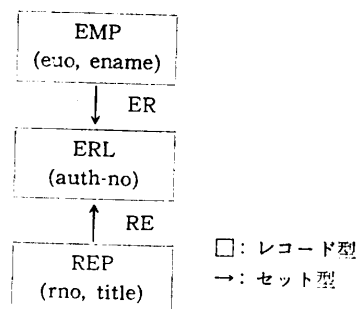


図2 論理スキーマ例
Fig. 2 Logical CODASYL schema.

Q_{R_1} と Q_{R_2} の関係より定まる述語で、これにより S に実際に与えられる R 実現値対 (セット (S) 実現値) が定まる。そしてこの集合をセット実現値 (SO) 集合 $S = \{u | u \in \times_{i=1}^2 \text{dom}(Q_{R_i}) \wedge \Gamma_S(u)\}$ とする。各 $u = (o_1, o_2) \in S$ で、 $o_1 \in R_1, o_2 \in R_2$ でなければならず、おのおのを親子 R 実現値、また o_1 と o_2 は親子関係にあるという。 S は f_S により与えられるので、各 $o_1 \in R_1$ は $h(\geq 0)$ 個の子 $o_2 \in R_2$ を、また各 $o_2 \in R_2$ はたかだか一つの親 $o_1 \in R_1$ をもてる。 Γ_S の一つとして、あるデータ項目 $t \in Q_{R_i}$ について、各 $o_1 \in R_1$ のすべての子 $o_2 \in R_2$ が互いに異なった t 値をもたねばならない条件がある。この条件を満たす t を S の DNA (duplicates are not allowed) 項目という。 $@R_i$ は R_i の主キーであり、識別子であるので、各 $(o_1, o_2) \in S$ を db キーの対 $(\delta(o_1), \delta(o_2))$ によって表すこともある。本論文では、以下 $S = \{u | u = (\delta(o_1), \delta(o_2)) \wedge u' = (o_1, o_2) \in \times_{i=1}^2 \text{dom}(Q_{R_i}) \wedge \Gamma_S(u')\}$ とする。 S は次の S 述語 $\sigma_S: \times_{i=1}^2 \text{dom}(Q_{R_i}) \rightarrow \{\text{true}, \text{false}\}$ と同じ意味をもつ。

$$\begin{aligned} \sigma_S(o_1, o_2) = & \text{true if } (\delta(o_1), \delta(o_2)) \in S, \\ & \text{false otherwise} \end{aligned} \quad (2)$$

以上の RO 集合と SO 集合の集合を C の概念データベース C とする。図2は職員と論文の情報を表す C の例である。箱は R 型、矢印は親から子 R 型への S 型を表す。

3.2 概念操作演算

概念操作演算としては、検索と更新演算がある。更新演算の基本的問題は、文献8)で論じているので、本論文では検索演算について述べる。

3.2.1 CODASYL 問合せ言語 (CQL)

概念データベース C 上の検索演算は、①目標スキーム設定と、②検索条件設定との二つの手順によって表される。検索条件 Ψ は、次の3種の基本式、(a) r 。

$\alpha\theta v$, (b) $r.\alpha\theta r'.\beta$, (c) $\sigma_S(r, r')$ から成る論理式で表される。ここで、 r はある RO 集合 R 内の R 実現値をとる O 変数、 $r.\alpha$ は r のとる R 実現値の項目 α の値とする。 θ は比較演算子、 v は定数である。 r' は RO 集合 R' の O 変数で、 β は R' の項目である。(c)は(2)式で与えられる S 述語である。このとき、 Ψ は次のように正規化される。

$$\Psi = \sigma \wedge \rho \wedge \pi \quad (3)$$

ここで、 σ は S 述語 σ_S の積、 ρ は各 O 変数 r ごとの条件式 ρ_r の積で、 ρ_r は $r.\alpha\theta v$ の積・和の形で表現される。 π は O 変数 r と r' 間の条件式 $\pi_{rr'}$ の積で、 $\pi_{rr'}$ は $r.\alpha\theta r'.\beta$ で表現される。 Ψ 内の O 変数を $r_1, \dots, r_l (l \geq 1)$ とすると、 $\Psi: \times_{i=1}^l \text{dom}(Q_{R_i}) \rightarrow \{\text{true}, \text{false}\}$ である (r_i は RO 集合 R_i の O 変数とする)。

目標スキームを $\tilde{A}_G = \{a_1, \dots, a_k\} (k \geq 1)$ とし、 a_i を目標項目という。いま $R_1, \dots, R_l$ を C 内の R 型とすると、 $\tilde{A}_G$ は $\tilde{\lambda}(Q_{R_1}, \dots, Q_{R_l}) = \tilde{A}_G$ なる関係 $\tilde{\lambda}$ で定義される*。 $\tilde{\lambda}$ から誘導される関数 $\lambda: \times_{i=1}^l \text{dom}(Q_{R_i}) \rightarrow \text{dom}(\tilde{A}_G)$ を目標関数とする。

検索演算は、 (λ, Ψ) と表せて、次の意味をもつ。

$$G = \{t | t = \lambda(o) \wedge \Psi(o) \wedge o = (o_1, \dots, o_l) \wedge \bigwedge_{i=1}^l \mathcal{P}_{R_i}(o_i)\} \quad (4)$$

以上をもとに、検索演算は CODASYL 問合せ言語 CQL で次のように表せる。

$$\text{var } (r_1, R_1) \dots (r_l, R_l); \quad (5)$$

$$\text{get into } G(A) \text{ where } \Psi; \quad (6)$$

(5)式は、(1)式の $\mathcal{P}_{R_i}(r_i)$ を表し、 RO 集合 R_i に O 変数 r_i を定める。 A は目標関数 λ によって定まるもので、 $\{\beta_1, \dots, \beta_k\}$ で与えられる。 β_i は、目標項目 a_i の値を与える関数 $\varepsilon_i: \times_{j=1}^l \text{dom}(Q_{R_j}) \rightarrow \text{dom}(a_i)$ である。 β_i は、 $a_i = \varepsilon_i$ と与えられる。

各 ε_i 内の項目が出力項目である。 Ψ は(3)式で与えられる条件式である。また $A = (\beta_1, \dots, \beta_k)$ を目標リストという。

[定義 3.1] 検索演算の CQL 表現を CODASYL 問合せと定義する。

3.2.2 CODASYL 問合せグラフ (CQG)

CODASYL 問合せのグラフ表現としての CODASYL 問合せグラフ (CQG) Q は次のようにして与えられる。

[CQG の構成方法]

* $\tilde{\lambda}$ は目標項目として、項目集合 Q_{R_i} からどのような項目を選ぶかを与える関係である。

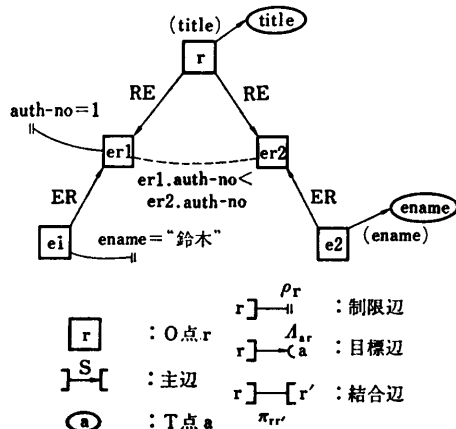


図 3 CODASYL 問合せグラフ (CQG)
Fig. 3 CODASYL Query Graph (CQG).

① CODASYL 問合せの各 O 変数 r に対して O 点 r を設ける (以下 O を省略し点 r とする)。② Ψ の各 $\sigma_s(r, r')$ に対して、この値が真ならば点 r から r' に有向な主辺 S を設ける。③ 各 $\pi_{rr'}$ に対して、点 r と r' 間に結合辺 $\pi_{rr'}$ を設ける。④ 各 ρ_r に対して、点 r に接する目標辺 ρ_r を設ける。⑤ A 内の各 $a = \varepsilon$ に対して、T 点 a を設け、⑥ ε 内の T 点に対応する点 r から T 点 a に有向な目標辺 A_{ra} (ε 内の r の出力項目集合) を設ける。

【定義 3.2】 CQG Q で、目標辺に接続する点を出力点とする。Q が主辺について連結なときのみ Q を連結とする。連結な Q 内の任意の 2 点 r と r' 間の路とは、接続する主辺とその上の点を逐次たどることによって得られる r と r' 間の点と主辺の系列である。

【命題 3.1】 任意の CODASYL 問合せと CODASYL 問合せグラフ (CQG) とは一対一対応する。

【証明】 CQG の構成方法より明らかである。

命題 3.1 よりすべての問合せは CQG で表現され、この表現法は利用者インタフェースとして非常に見やすい形になっている。

【例 3.1】 図 2 に対する検索演算「鈴木を第一著者とする共著論文名とその共著者を求めよ」は、CQL で (7) のように書ける。

```
var (e1, EMP) (e2, EMP) (er1, ERL) (er2, ERL) (r,
REP); get into G (r. title, e2. ename) where e1.
ename = "鈴木" and ER (e1, er1) and RE (r, er1)
and RE (r, er2) and ER (e2, er2) and er1. auth-no
= 1 and er2. auth-no > er1. auth-no; (7)
```

(7) に対応した CQG を図 3 に示す。箱は点、点間の矢と点線はおのおの、主辺と結合辺を、また点に接する接地記号は制限辺を表す。楕円は T 点を、点から T

点への有向辺は目標辺を表す。

4. CODASYL DBS の仮想層

本章では、概念層と物理層の中間に位置づけられ、従来の DML²⁾ の操作層に対応した仮想層の定義を行う。

4.1 仮想データ構造

概念データ構造 $C = (C, C)$ の仮想層での表現を仮想データ構造 $V = (V, V)$ とする。V と V をおのおの、仮想スキーマと仮想データベースとする。C 内の各 R 型 $R = (Q_R, \Gamma_R)$ に対して、新たに仮想 R(VR) 型 $X = (Q_X, \Gamma_X, <_X)$ が定義できる。ここで $Q_X = Q_R, \Gamma_X = \Gamma_R$ である。 $<_X$ は db キーによる昇順の順序関係である。以上のことから、X に実際にデータを与えた仮想 RO (VRO) 集合 X は、RO 集合 R を $<_X$ で全順序づけたものである。各 R 実現値 $o \in R$ に対応した $v \in X$ を、 o の仮想 R(VR) 実現値とする。

各 S 型 $S = (R_1, R_2, \Gamma_S)$ に対して、仮想 S(VS) 型 $Y = (X_1, X_2, \Gamma_Y, <_Y)$ が定義される。 X_i は R_i の VR 型で ($i=1, 2$), $\Gamma_Y = \Gamma_S, <_Y$ は順序関係である。 X_1 と X_2 をおのおの、Y の親と子 VR 型とする。各 R 実現値 $o_1 \in R_1$ の VR 実現値 $v_1 \in X_1$ に対して、 $m_Y(v_1)$ を S の子 $o_2 \in R_2$ の VR 実現値 $v_2 \in X_2$ の集合とする。Y に実際にデータを与えた仮想 SO (VSO) 集合を、 $Y = \{(v_1, m_Y(v_1)) | v_1 \in X_1 \wedge m_Y(v_1) \neq \emptyset\}$ とする。各 $(v_1, m_Y(v_1)) \in Y$ を仮想 S(VS) 実現値とする。 $m_Y(v_1)$ は $<_Y$ で全順序づけられている。 $<_Y$ としては、①あるデータ項目 $t \in Q_{X_1}$ についての昇順と②降順、③任意の順との 3 種がある。①と②の t を Y の整列項目とする。

VRO 集合と VSO 集合の集合を仮想データベース V とする。

【性質 4.1】 仮想データ構造 V から、順序関係を除くと概念データ構造 C になる。

4.2 仮想 CODASYL 機械による仮想操作演算の表現

従来の COBOL DML²⁾ の論理的意味を明らかにするために、多テープチューリング機械として、仮想 CODASYL 機械 (VCM) を考える。VCM は有限制御 $\mathcal{G}$ と任意の有限本のテープ集合 $\mathcal{G}$ とから成る (図 4)。各テープ T は任意の有限個の単位セルから成る。各単位セルは、識別子とデータをおのおの記憶する ID セルと D セルから成る。 $\mathcal{G}$ は T ごとにヘッド H_T をもち、時刻ごとに一つの単位セルを見ることができる。 $\mathcal{G}$ は仮想データベース V を記憶する DB テープ、仮想

表 1 VCM 演算 (VOP) と COBOL DML
Table 1 VCM Operations and COBOL DML's.

VOP	意 味	COBOL DML
(1) <u>first</u> (X);	H _x を T _x の先頭に移動 (H)	<u>find first</u> X <u>within</u> A.*
(2) <u>last</u> (X);	H _x を T _x の最後に移動 (T)	<u>find last</u> X <u>within</u> A.
(3) <u>next</u> (X);	H _x を一つ右に移動 (R)	<u>find next</u> X <u>within</u> A.
(4) <u>prior</u> (X);	H _x を一つ左に移動 (L)	<u>find prior</u> X <u>within</u> A.
(5) <u>CALC</u> (X);	β_x 内の t の値を D にもつ単位セルに H _x を移動 (AC)	<u>find any</u> X.
(6) <u>first</u> (Y); Y=(X ₁ , X ₂) とする	H _{x1} 下の単位セルを親とする先頭の子に H _r および H _{x2} を動かす	<u>find first</u> X ₁ <u>within</u> Y.
(7) <u>last</u> (Y);	H _{x1} 下の単位セルを親とする最後の子に H _r と H _{x2} を動かす	<u>find last</u> X ₁ <u>within</u> Y.
(8) <u>next</u> (Y);	H _r を一つ右へ動かす (H _r =R). H _{x2} も H _r 下の R 実現値に動かす	<u>find next</u> X ₁ <u>within</u> Y.
(9) <u>prior</u> (Y);	H _r を一つ左に動かす (H _r =L). H _{x2} も H _r 下の R 実現値に動かす	<u>find prior</u> X ₁ <u>within</u> Y.
(10) <u>owner</u> (Y);	H _r 下の単位セルの親に H _{x1} を動かす	<u>find owner</u> <u>within</u> Y.
(11) <u>get</u> (X);	m _x =RD	<u>get</u> X.
(12) <u>if</u> $\sigma=F$. <u>go to</u> L;	(σ)=F ならば, ID=L の単位セルに H _p を動かす	<u>if</u> $\sigma=F$, <u>go to</u> L.
(13) <u>open</u> (T _i);	HT _i を TT _i の先頭に移動 (H)	<u>open</u> T _i .
(14) <u>close</u> (T _i);	HT _i を TT _i の最後に移動 (T)	<u>close</u> T _i .
(15) <u>write</u> (T _i);	HT _i の単位セルに β_{T_i} を書き出す (AP)	<u>write</u> T _i .
(16) <u>accept</u> (X, v);	v に (δ_x) を記憶する	<u>accept v from</u> X <u>current</u> .
(17) <u>accept</u> (Y, v);	v に (δ_y) を記憶する	<u>accept v from</u> Y <u>current</u> .
(18) <u>find</u> (X, v);	H _x を, (v) を ID (db キー) としてもつ単位セルに動かす	<u>find</u> X; <u>db-key is</u> v.
(19) <u>stop</u> ;	VCM の停止	<u>stop</u> .
(20) <u>store</u> (X);	(β_x) を L _x を満たすように T _x に加える (AP)	<u>store</u> X.
(21) <u>erase</u> (X);	H _x 下の VR 実現値を含むすべての単位セルをテープから消す (DL)	<u>erase</u> X.
(22) <u>modify</u> (X);	H _x のいる単位セルの D を (β_x) とする (RP)	<u>modify</u> X.
(23) <u>connect</u> (Y);	H _{x1} , H _{x2} 下の単位セルを親子として, T _r に L _r を保つように加える	<u>connect</u> X ₁ <u>to</u> Y.
(24) <u>disconnect</u> (Y);	DL を行う (H _r のいる単位セルの消去)	<u>disconnect</u> X ₁ <u>from</u> Y.
(25) <u>modify</u> (Y);	H _{x2} 下の単位セルの親を H _{x1} 下の単位セルにする	<u>modify</u> X ₁ <u>only</u> Y.
(26) <u>go to</u> L;	ID=L の単位セルに H _p を動かす	<u>go to</u> L.
(27) $p(v_1, \dots, v_e)$	$v_1, \dots, v_e$ 内の値について述語 P を評価する	$p(v_1, \dots, v_e)$
(28) $v_i \leftarrow v_j$;	(v_j) を v_i に記憶する	<u>move</u> v_j <u>to</u> v_i .
(29) $A(v_1, \dots, v_e)$	$v_1, \dots, v_e$ 内の値について関数 A を評価する	$A(v_1, \dots, v_e)$
(30) <u>read</u> (T _i)	HT _i のいる D の情報を β_{T_i} に読み出す	<u>read</u> T _i

* A は, X の属するエリア [CODAS 73] である.

以下で X と Y=(X₁, X₂) はおのおの VR 型と VS 型とする.

応し、その動作は表1で表される。

以上より、仮想 CODASYL データベース V 上の COBOL DML の意味が多テープチューリング機械の動作として正確に記述され、CODASYL DBSアーキテクチャ設計の基礎を与えることができる。これにより、CQG (3.2.2項) の動作が明確に与えられ、以下に述べるシステム評価の考察の基礎を与える。

5. CODASYL 問合せの VOP 表現

概念データ構造 C 上の CODASYL 問合せグラフ (CQG) の条件を満たす操作演算が、仮想 CODASYL データ構造上の仮想操作演算 VOP (すなわち、COBOL DML) としてどのように表現されるかについて論じる。

5.1 CQG の VOP 表現可能性

[定義 5.1] VOP (DML) の系列を VOP(DML) プログラムとする。

[定義 5.2] CQG Q の目標リスト A と検索条件式 Ψ とを満たす目標集合 G を導出する VOP (および DML) プログラム P が存在するとき、 P は Q の条件を満たすという。

[定理 5.1] 任意の CQG Q の条件を満足する VOP (および DML) プログラム P が存在する。

[証明] 次の二つの基本操作ができれば定理は明らかである。① Q の点 r_i, r_j とその主辺 S に与えられている条件をすべてチェックし、それを満たす目標集合 G_s を選ぶ操作と、② これらの G_s を1本または複数本のテープに出力し、その後与えられた順序で出力する操作。①の操作は、各点 $r_i, r_j, (r_i \text{ から } r_j)$ の主辺 S の条件 $\rho_i, \rho_j, \sigma_s, \pi_{ij}$ を満足する R 実現値対 ($o_i \in R_i, o_j \in R_j$) に対応した VR 実現値対 ($v_i \in X_i, v_j \in X_j$) (X_i と X_j はおのおの、 R_i と R_j に対応した VRO 集合) を見いだすことで行える。それは次のようにして求められる。(a) X_i を表1の VOP(1)と(2)を用いて ρ_i を満たす v_i を見つけ、(b) その子のなかから(6)と(8)を用いて操作し、 ρ_j と π_{ij} を満たす v_j を見つけ、(c) O テープ To_s に (v_i, v_j) を(15)で書く。よって、①を VOP で行える。また②を VOP で行えるのは明らかである。

本定理は、CQG の条件と VOP との対応を考えるうえの基本として必要である。

5.2 最適化—出力テープ数の最小化

CQG の条件を満たす VOP プログラムは、定理5.1で示した以外にも存在する。そのなかで、ある目標値

を最適にするプログラムを求める問題がある。目標値として、① O テープ数の最小化と、② 応答時間の最小化との二点⁸⁾がある。定理5.1の方法では、検索のために主辺数+1本の O テープが必要になる。 O テープは、通常の物理的ファイルに対応するので、テープ数の増加はファイル空間と操作量の増加をもたらす、好ましくない。以下の考察では、 O テープ数の最小化問題を考える。②については、文献9)で論じる。

5.2.1 巡航木

[定義 5.3] VOP (および DML) プログラム P が次の条件を満たすとき、 P は巡航的であるという。① 1本の O テープ To のみを用い、② To の open と close は1回のみで、他に To 上の write のみを用い、③ close したとき、 To 内に目標集合 G が定められた順序に従って書き出されている (冗長な組を含んでもよい)。

すなわち、VCM が、 P に基づいて、DB テープを動作させながら、 To に順編成ファイルに対応した結果を順次書いていく場合である。

[定義 5.4] CQG Q が主辺について連結で、閉路をもたず、ある点を根とするとき、 Q を木と定義する。根と葉間の路を街道とし、各主辺 S に接続する2点 r_i と r_j のなかで、根に近い点 r_i を上点、他の r_j を下点とする。下点 r_j が S 型 S の子の O 変数を表すとき M 型、親のとき O 型とする。この主辺 S について、 R 実現値 $o_i \in R_i, o_j \in R_j$ が親子または子親関係にあるときのみ真となる主述語 $\beta_s(o_i, o_j)$ を定義する。

[定義 5.5] 木 Q が次の条件を満たすとき、 Q を巡航木という。① 各主辺で、上点に対し下点が順序づけられ、② 全出力点は Q の最右の街道内にあり、③ 任意の結合辺は同一街道内にある。ここで、葉に最も近い出力点と根との間の路を幹とする。各点 r について、 r から根までの路内のすべての点 r' と r 間の条件式 $\pi_{rr'}$ の積を上方結合式 Π_r とし、下方結合式 Δ_r を Π_r 以外の r の結合式の積とする。根 r_0 から r_i までの路を $r_0, S_1, \dots, S_i, r_i$ とすると、 r_i の上方条件式 ϕ_i を次のように定義する。 o_j を O 変数 r_j がとる R 実現値とすると ($j=0, 1, \dots, i$)

$$\begin{aligned} \phi_i(o_0, \dots, o_i) \\ \triangleq \begin{cases} \text{true} & \text{if } \rho_i(o_i) \wedge \beta_{S_i}(o_{i-1}, o_i) \wedge \Pi_i(o_0, \dots, o_i) \\ \text{false} & \text{otherwise} \end{cases} \quad (8) \end{aligned}$$

図3の CQG は根を r とする木であるが、結合辺が同一街道内がないので巡航木ではない。しかし、根を

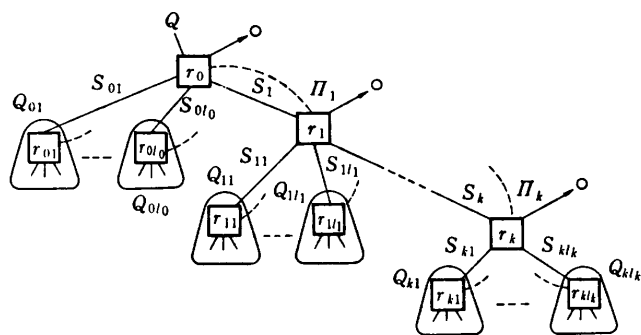


图 6 巡航木
Fig. 6 Navigational tree.

図6の巡航木Qを考える。Qの根を r_0 、幹を $r_0, S_1, r_1, \dots, S_k, r_k$ とする。各点 r_i は主辺 $S_{i,j}$ で連結する幹外の下点 $r_{i,j}$ ($j=1, \dots, l_i$ (≥ 0))をもつ。 $Q_{i,j}$ を $r_{i,j}$ を根とする部分木とする。幹点 r_k のとりR実現値を o_h ($h=0, 1, \dots, i$)とする。 $o_0, \dots, o_i$ に対して、 $Q_{i,j}$ 内の全点で(1)式の上方条件式 ϕ が真となるときのみ真となる $\phi_{i,j}$ を $Q_{i,j}$ の部分木述語とする。

$$g_Q = \{t \mid t = A(o) \wedge o = (o_0, o_1, \dots, o_k) \wedge \bigwedge_{i=0}^k \mathcal{P}R_i(o_i)$$

$$\bigwedge_{i=0}^k \phi_i(o_0, \dots, o_i) \quad \bigwedge_{i=0}^k \bigwedge_{j=1}^{l_i} \Phi_{ij}(o_0, \dots, o_i) \quad (9)$$

【性質 5.1】 (8) 式の論理演算を行う VOP が存在する.

【定理 5.2】 任意の巡航木 Q が与えられたとき、一本の O テープを用いて、 Q の条件を満たす目標集合を一定順序で求める 巡航的 VOP プログラムが存在する。

次に出力が巡航的であることを以下の帰納法で示

【定義 5.6】 木 Q が次の条件を満足するとき、 Q を準巡航木という。①出力点は最右の街道内か、または出力点から根までの全点が O 型の路内に存在し、②各主辺で上点に対し下点が順序づけられ、③各結合辺は同一路内にある。

[証明] 略.

5.2.2 CQG の巡航木表現

【定理 5.4】 任意の CQG Q と等価な巡航木 N が存在する.

【定理 5.5】 任意の CQG の条件を満足する巡航的 VOP (および DML) プログラムが存在する.

〔証明〕 定理 5.2 と 5.4 より明らかである.

5.2.3 巡航木の条件を満たす VOPプログラムの構成

巡航木 N から, 4.2 節で与えた VCM 上の巡航的 VOP プログラムの構成手順 VOPG を以下に与える。

〔VOPG 手順〕 N の根を r_0 , 各主辺 S の上点を r' , 下点を r とする。まず, r_0 に対して表 2 (a) より VOP を P テープ T_P に書く。次に, 点 r' まで VOP が構成されているとする。 r' から与えられた順序に従って下点 r を求め, 表 2 (b)~(e) より条件に従って VOP を求め T_P に書く。同様のことを, 以下の全点について行う。すなわち, depth-first の木探索を行う。

表 2 内で代表的な (b) を説明する。主辺 S の上点を r' , 下点 r とする。いま, r' が表す VR 実現値 v' に対応した単位セルに $H_{X'}$ がいるとする。 r に H_X のいる単位セル内の VR 実現値の db キー, 目標項目値, 下方結合式 Δ_r 内の r の項目値のおのおのを, 必要に応じて記憶する。おのおのの記憶を $\gamma_r, \lambda_r, \alpha_r$ とする (おのおのは, VCM 内の作業用レジスタである)。以下に r での操作を示す。
 ① r' の v' の先頭の子を T_r より求める (②)。 ⑤ 存在すれば (⑤), T_x より D の情報を β_x に読み出し (⑥), 条件式 ρ_r, Π_r を調べる (⑦)。 ③ 条件を満たせば, β_x より必要な情報を α_r と λ_r に記憶する。 r の一つ下位の点について以下同様に行う。 ④ 条件を満たさねば H_r を一つ右に進め (④), ①~④を行う。 ⑥ 子がなくなったとき, v' の次の順序の VR 実現値 v'' を v' として ①~⑥を行う (⑥)。 r が最右の葉のときは, 全条件を満たせば, 幹内の各出力点 r_P の (λ_{r_P}) よ

表 2 CQG と VOP
Table 2 CQG and VOP.

CQG	VOP
(a)	$\begin{aligned} & \text{open } (T_0); \quad \beta_0 \leftarrow \phi; \\ & \text{first } (X_0); \quad \text{go to } L_0; \\ & K_0: \llbracket \text{find } (X_0, r_0); \rrbracket \\ & M_0: \text{next } (X_0); \\ & L_0: \text{if } \sigma = F \quad \text{close } (T_0); \text{stop}; \\ & \quad \text{get } (X_0); \quad \text{if } \sim \rho_0(\beta_{X_0}) \text{ go to } M_0; \\ & \quad \text{accept } (X_0, r_0); \\ & \quad \alpha_0 \leftarrow \beta_{X_0}, \tau_0; \quad \lambda_0 \leftarrow \beta_{X_0}, \Delta_0; \\ & \quad [\dots]^* \quad \text{go to } M_0; \end{aligned}$
(b) 幹内の M 型点 (r)	$\begin{aligned} & S_r: \llbracket \text{find } (X_r', r_r'); \rrbracket \\ & \text{first } (Y); \quad \text{go to } L_r; \\ & K_r: \llbracket \text{find } (X_r, r_r); \rrbracket \\ & M_r: \text{next } (Y); \\ & L_r: \text{if } \sigma = F \quad \text{go to } K_l; \\ & \quad \text{get } (X_r); \quad \text{accept } (X_r, r_r); \\ & \quad \text{if } \sim (\rho_r(\beta_{X_r}) \wedge \Pi_r(\alpha_r, \dots, \alpha_r', \beta_{X_r})) \text{ go to } M_r; \\ & \quad \alpha_r \leftarrow \beta_{X_r}, \tau_r; \quad \lambda_r \leftarrow \beta_{X_r}, \Delta_r; \\ & \quad [\dots]^* \\ & \quad \text{go to } M_r; \end{aligned}$
(c) 幹内の O 型点 (r)	$\begin{aligned} & S_r: \llbracket \text{find } (X_r', r_r'); \rrbracket \\ & \text{owner } (Y); \\ & \text{if } \sigma = F \quad \text{go to } K_l; \\ & \text{get } (X_r); \quad \text{accept } (X_r, r_r); \\ & \text{if } \sim (\rho_r(\beta_{X_r}) \wedge \Pi_r(\alpha_r, \dots, \alpha_r', \beta_{X_r})) \text{ go to } K_l; \\ & \alpha_r \leftarrow \beta_{X_r}, \tau_r; \quad \lambda_r \leftarrow \beta_{X_r}, \Delta_r; \\ & [\dots]^* \\ & \text{go to } K_l; \end{aligned}$
(d) 幹外の M 型点 r	$\begin{aligned} & S_r: \llbracket \text{find } (X_r', r_r'); \rrbracket \\ & \text{first } (Y); \quad \text{go to } L_r; \\ & K_r: \llbracket \text{find } (X_r, r_r); \rrbracket \\ & M_r: \text{next } (Y); \\ & L_r: \text{if } \sigma = F \quad \text{go to } K_l; \\ & \quad \text{get } (X_r); \\ & \quad \text{accept } (X_r, r_r); \\ & \quad \text{if } \sim (\rho_r(\beta_{X_r}) \wedge \Pi_r(\alpha_r, \dots, \alpha_r', \beta_{X_r})) \text{ go to } M_r; \\ & \quad \alpha_r \leftarrow \beta_{X_r}, \tau_r; \\ & \quad [\dots]^* \\ & \quad \text{go to } K_l; \end{aligned}$
(e) 幹外の O 型点 r	$\begin{aligned} & S_r: \llbracket \text{find } (X_r', r_r'); \rrbracket \\ & \text{owner } (Y); \\ & \text{if } \sigma = F \quad \text{go to } K_l; \\ & \text{get } (X_r); \\ & \text{accept } (X_r, r_r); \\ & \text{if } \sim (\rho_r(\beta_{X_r}) \wedge \Pi_r(\alpha_r, \dots, \alpha_r', \beta_{X_r})) \text{ go to } K_l; \\ & \alpha_r \leftarrow \beta_{X_r}, \tau_r; \\ & [\dots]^* \\ & \text{go to } K_l; \end{aligned}$

ここで,

$[\dots]^* = [\beta_0 \leftarrow \Delta(\lambda_0, \lambda_1, \dots, \lambda_k);$

$\text{if } \beta_0 \neq \beta_0' \{ \text{write } (T_0); \beta_0 \leftarrow \beta_0'; \}]^*$

$l = r$ から根 r_0 までの街道内で, r に最も近い M 型点, なければ r_0

$k = r'$ を親とし, r の右隣りの点, なければ l

β_{X_r}, Δ_r = レジスタ β_{X_r} 内の目標項目集合 Δ_r の値

β_{X_r}, τ_r = レジスタ β_{X_r} 内の Δ_r の目項集合 τ_r の値

$[\dots]^*$ は, 点 r が最右の葉点のとき必要な VOP

り目標組を求め O テープに出力する (⑨)*. 表 2 の他も同様である.

【命題 5.1】 任意の巡航木に対して, これを満足する巡航的 VOP プログラムを, 表 2 と VOPG 手順より構成できる.

【証明】 任意の巡航木の各主辺 S と下点 r について考えると, r は幹内か幹外の点で, かつ M 型か O 型である. したがって, S と r は, 表 2 の (b) ~ (e) のどれかのタイプであり, 木の根は (a) のタイプである. VOPG により構成された VOP プログラムの出力が巡航的であることは, 定理 5.2 と表 2 の VOP の意味より明らかである. ■

【定理 5.6】 概念 CODASYL データ構造上の任意の CQL 問合せの条件を満たす巡航的 VOP プログラムを, 表 2 と VOPG 手順より構成できる.

【証明】 命題 3.1 より, CQL 問合せと CQG とは一対一対応する. さらに, 定理 5.4 と命題 5.1 が成り立つので, 本定理は明らかである. ■

すなわち, CQL で表せる任意の検索演算から, この条件を満たす巡航的 VOP プログラムを, 表 2 と VOPG 手順とによって構成できる.

6. 非手続的インタフェース

文献 8) では, ローカル概念 (LC) モデルと LC 操作言語 RQL とを定め, LC と論理 CODASYL データ構造の一対一対応性を示した. データ構造の一対一対応性より, 次の命題が成り立つことは明らかである.

【命題 6.1】 任意の RQL 検索演算と CQL 検索演算 (および CQG) とは一対一対応する. ■

【定理 6.1】 任意の RQL 検索演算の条件を満たす巡航的 VOP (DML) プログラムが存在する.

【証明】 命題 6.1 と定理 5.5 より明らかである. ■

【定理 6.2】 任意の RQL 演算条件を満たす VOP (DML) プログラムが存在する.

【証明】 RQL 演算は (ρ, A, Ψ) と表される. 基本操作演算 ρ を VOP で表せることは文献 8) より明らかである. 目標リスト A と検索条件 Ψ を VOP で表せることは, 定理 6.1 より明らかなので, 定理は成り立つ. ■

よって, LC データ構造上の任意の RQL 演算の表す条件を COBOL DML プログラムで表せることがわ

かり, CODASYL DBS 上に非手続的インタフェース設計の基礎が明らかとなった.

7. おわりに

本論文では, CODASYL DBS を論理, 概念, 仮想, 物理の 4 層から成るとし, このなかで, 概念層と仮想層, および層間写像を論理的に明らかにした. 概念層では, 集合と部分関数から成る概念データ構造 C と, 述語論理形式の CODASYL 問合せ言語 (CQL) および等価な CODASYL 問合せグラフ (CQG) との定義を与えた. CQG は見やすい形になっていることから, 利用者インタフェースとして優れている. 仮想層では, C に順序関係を導入した仮想データ構造 V を与え, 従来の COBOL DML と一対一対応する仮想操作演算 (VOP) を実行できる仮想 CODASYL 機械 VCM を多テープチューリング機械で定義し, 従来の DML の意味を明らかにした. 層間写像として, すべての CQG を, 一本の出力テープとこれへの逐次書込みだけで目標集合を導出できる巡航的 VOP プログラムによって正しく表せることを論理的に明らかにし, その具体的方法を示した. 本方法は, 文献 5), 6) に対して, 完全な逐次出力で目標集合を導出でき, 文献 12) に対しては, さらに任意の結合辺と, 主辺について木型以外の検索演算を許している点で優れている.

CQG の条件を満たす VOP を考えるとき, 出力テープ数と操作量の最小化に加えて, 応答時間と出力量の最小化が目標となる. これについては文献 9) で論じる.

以上より, 本論文は, 従来の CODASYL DBS 上に非手続的利用者インタフェース設計の論理的基礎を与えている. また, 本論文で述べた概念に基づいて, CODASYL DBS 上の非手続的インタフェースシステムローカルデータベースプロセッサ (LDP-V2)^{3)-5), 7), 8)} が Jipdec の AIM (M-170F), ADBS (Acos-700) 上ですでに実働している.

謝辞 LDP-V2 の開発を共同して行った (財) 日本情報処理開発協会の横塚実氏に感謝する.

参 考 文 献

- 1) CODASYL DDL Committee: CODASYL Data Description Language, J. Dev. (1973).
- 2) Olle, T. W.: *The Codasyl Approach to Data Base Management*, John Wiley & Sons, New York (1978).
- 3) Takizawa, M., et al.: Resource Integration

* COBOL DML では, 同一 R 型に対応した複数の点が存在するときには, ヘッド位置を正しく保つために表 2 の [...] の VOP が必要である.

- and Data Sharing on Heterogeneous Resource Sharing System, Proc. ICCS '78, pp. 253-258 (1978).
- 4) Takizawa, M. et al.: The Four-Schema Concept as the Gross Architecture of Distributed Databases and Heterogeneity Problems, *JIP (IPSJ)*, Vol. 2, No. 3, pp. 134-142 (1979).
 - 5) Takizawa, M. et al.: Query Translation in Distributed Databases, Proc. IFIP '80, pp. 451-456 (1980).
 - 6) Takizawa, M.: Distribution Problems in Distributed Databases, *JIP (IPSJ)*, Vol. 5, No. 3, pp. 139-147 (1982).
 - 7) 滝沢, 横塚他: 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-V 1.5) の設計と実現, 情報処理学会論文誌, Vol. 23, No. 3, pp. 665-675 (1982).
 - 8) 滝沢, 野口: CODASYL データベースシステムに対する非手続的更新インタフェースの基本概念, 情報処理学会論文誌, Vol. 24, No. 6, pp. 857-866 (1983).
 - 9) 滝沢, 野口: CODASYL 問合せの最適化と評価について, 準備中.
 - 10) Held, G. et al.: INGRES—A Relational Database System, AFIPS Conf. Proc., pp. 409-416 (1976).
 - 11) Codd, E. F.: A Relational Model of Data for Large Shared Data Bank, *CACM*, Vol. 13, No. 6, pp. 337-387 (1980).
 - 12) Dayal, U. et al.: Query Optimization for CODASYL Database Systems, Proc. ACM SIGMOD, pp. 138-150 (1980).
 - 13) Biller, H. et al.: On the Semantics of Data Bases: The Semantics of Data Manipulation Language, Proc. IFIP TC-2, pp. 239-267 (1976).
 - 14) Gangopadhyay, D. et al.: Semantics of Network Data Manipulation Languages: An Object-Oriented Approach, Proc. the 8th VLDB, pp. 357-369 (1982).
 - 15) CODASYL DDL Committee: Report of the CODASYL Data Description Language Committee, *Inf. Syst.*, Vol. 3, No. 4, pp. 247-320 (1978).

(昭和 58 年 9 月 21 日受付)

(昭和 59 年 3 月 6 日採録)