

データ模型の分類学

データベース・モデル研究委員会中間報告

千葉恭弘

(株)日本ユニバック総合研究所

序

1970年に発表されたE. F. Coddによる関係形式模型を一つの契機として、データベース・システムの模型化が非常に流行し、現在までに多数の模型が提唱されてきた。データベース・システムの模型化には、現実世界の知識の抽象化としてのデータ模型と、その上で定義されるデータベース・システムの行動模型の2つの側面が必要である。前者は現実世界における情報の持つ意味を明確に把握し、データベースにおける表現形式へと翻訳する過程を含み、後者はこれらのデータに対する論理的なアクセスの過程を含む。最近の傾向では、データ模型の中に意味論的分析を強く加え、より良い抽象化の道具を構成しようとする動きが活発に見られる。

多くのデータ模型では用いられている用語の多様性や特異性のために、相互の比較・検討が充分行なわれてきたとは言い難い。1976年にトロント大学のCSRG (Computer Systems Research Group) では、これらのデータ模型の相互の位置づけを明確にする目的で「データ模型の分類学」(A Taxonomy of Data Models) [1] を発表した。本稿ではこの論文を紹介し、幾つかのデータ模型の特徴と相互の関連性を検討してみよう。

1. データ模型の概略

CSRGの分類学の対象となっているデータ模型は次に掲げる15個である。

- | | |
|---|--------------------------------|
| ① 関係形式模型 (Relational Data Model) | Codd [2] [3] |
| ② DBTG / CODASYL 模型 | CODASYL [4] |
| ③ 階層模型 (Hierarchical Model) | IMS, System 2000 |
| ④ DIAM II 模型 | Senko [5] |
| ⑤ 情報論理模型 (Infological Model) | Langefors [6], Sundgren [7] |
| ⑥ 2項論理結合模型 (Binary Logical Association Model) | Bracchi et-al. [8] |
| ⑦ 事象・関係性模型 (Entity-Relationship Model) | Chen [9] |
| ⑧ データ意味論 (Data Semantics) | Abrial [10] |
| ⑨ 函数的データ模型 (Functional Data Model) | Nijssen [11], Kerschberg [12] |
| ⑩ 意味ネットワーク模型 (Semantic Network Model) | Roussopoulos & Mylopoulos [13] |
| ⑪ 拡張集合論模型 (Extended Set Theory Model) | Childs [14] |
| ⑫ 情報管理概念 (Information Management Concepts) | Durchholz & Richter [15] |
| ⑬ 情報代数模型 (Information Algebra Model) | 小林 [16], CODASYL [17] |
| ⑭ データ空間模型 (Data Space Model) | Bachman [18] |
| ⑮ Lindgreen 模型 | Lindgreen [19] |

各模型の概略を簡単に記述してみよう。

1.1 関係形式模型

関係形式模型はデータ項目間の関係性を数学的な几項関係として表現する形式

的模型である。複数の定義域 (domain) S_1, S_2, \dots, S_n 上に定義された関係 R とは n個組 (n-tuple)で構成される1つの集合である。換言すれば関係 R は複数の定義域の直積 $S_1 \times S_2 \times \dots \times S_n$ の1つの部分集合であると考えられる。集合 S_i は関係 R の i 番目の定義域と呼ばれ、 R の次数 (degree)は n であるという。

関係 R を表として見た時、その表の各々の列は属性 (attribute)と呼ばれ、関係の定義域に対応している。同じ定義域を持った複数の属性がある場合、これらを識別するために役目名 (role name)を付加し、1つの関係内では属性名をユニークとする。関係 R の鍵 (key) K とは次のような方向に独立性性質を充す R の属性の部分集合である。

- 一意識別性 ... K の値は R の各々の組を一意に識別する。
- 非冗長性 ... K の中のどんな属性も、性質 a をそこなうことなく削除することはできない。

関係形式モデルではデータへのアクセスは関係代数 (relational algebra)もしくは関係計算 (relational calculus)によって指示される。前者は幾つかの関係演算で構成され、引数として関係をとり、他の関係を結果として形成する。後者では求めたい関係の記述が1階述語計算で定式化された述語でもって与えられる。

1.2 DBTG / CODASYL 模型

DBTGのデータ模型における2つの基本的概念はレコード型 (record type)と親子組 (DBTG set)である。レコード型とは具体的なレコードの成員 (occurrence)の集まりを示す統体的な概念である。各々のレコード型は幾つかのデータ項目 (data item)で構成される。データ項目は個々のレコード型の成員に具体的なデータ項目値 (data item value)を持っている。レコード型は集用 (aggregate)、ベクトル (vector)及び繰り返レグループ (repeating group)を持つことができる。DBTGの親子組はレコード型間の結合として機能する。1つの親子組は1つの親レコード型と1つ(もしくは複数)の子レコード型を結びつける。

すなわち親子組の概念は親レコード型と子レコード型のレコード成員間に存在する関係性の集合である。これらの関係づけには次の2つの制約が課せられる。

- 子レコード型のレコード成員はたかだか1個の親レコード成員と結ばれる。
- 1つの親子組では親レコードと子レコードの型は同じであってはならない。

データ構造図 (data structure diagram)はDBTG模型のスキーマを表現する有向グラフである。各々の有向枝は1つの親子組に対応している。枝の方向は親レコード型から子レコード型へと向う。ここで1つの枝が子レコード型のレコード成員から親レコード型のレコード成員への関数を意味している。

1.3 階層模型

階層型データ模型の2つの基本概念は、レコード型と親子関係 (parent-child relationship)である。レコード型はセグメント型 (segment type)とも呼ばれ、具体的なセグメント成員を持つ。1つの親子関係はDBTGの親子組と同様に、2つのレコード型を結びつけ、2つのレコード型成員間の関係性の集まりを表現する。1つの親子関係は子レコード型の成員から親レコード型の成員への関数を表現しており、かつこの関数は全関数 (total function)である。すなわち各々の子レコード型成員は親レコード型のただ1つの成員と関係づけられ、この関係づけがなされない場合は存在できない。階層データ模型には次のような親子関係に対する制約が課せられている。

- どんな2つのレコード型の間にもたかだか1つの親子関係しか存在しない。

後、始点と終点の組が一意にそれを識別し、陽に名前付ける必要がない。
 b. レコード型の向て定義されるすべての親子関係は1つのデータ構造図を構成し、これは木(tree)となる(階層定義木 hierarchical definition tree)。

1.4 DIAM-II 模型

DIAM-II 模型は ①エンド・ユーザ・レベル ②情報レベル ③ストリング・レベル ④コード化レベル ⑤物理装置レベル の5つのレベルを対象とする。ここではANSI/SPARCの概念レベルに対応する②の情報レベルを検討する。現実世界における概念は、事象(entity)、属性(property)、事実(fact)、事象集合(entity set)などの用語で構成されているが、情報レベルではこれに対し事物の名前が扱われ、用いられる用語は属性名(attribute name)、属性値(attribute value)、識別名(identifier name)、識別値(identifier value)、事実表現(fact representation)などである。

1つの事実表現とは属性名の対であり、連想対(association pair)とも呼ばれ、2つの識別子を結びつける。連想対は対称的である。すなわち1つの属性名は1つの識別子を主語として用い、他の識別子をその値として用いる。一方もう1つの属性名は正にこれと反対のことを行う。

連想の対称性をを用いることは2つの目的に貢献する。第1にどの対象(object)が記述されるもの(described)で、どの対象が記述しているもの(describing)であるというような偏向を排除することである。第2に連想の両着がデータベースからの検索に有効に作用していることである。

1.5 情報論理模型

この模型に従うと、1つの情報システムは対象システムと呼ばれる現実の切片に関する情報を受け入れ、備蓄し、生成するように設計される。対象システムは対象(object)、属性(property)及び関係(relation)で構成されている。対象は属性を有し、時間軸の1点で、もしくは時間を通じて相互に関係づけられる。

基本状態(e-situation)とは、対象システムにおける知識を表現し、ある時点での識別された対象(もしくは対象のn個組)を保有している基本属性(もしくは関係)である。基本状態の2つのタイプは次の3つ組で表現できる。

$\langle 0, p, t \rangle$ | $0, 0_1, \dots, 0_n$: 対象 p : 属性
 $\langle \langle 0_1, \dots, 0_n \rangle, r, t \rangle$ | r : 関係 t : 時間

基本状態も属性も関係もすべて対象であると考えられる。

一方、人間が1つの対象システムについて認識し、思考する際に用いるものは参照(reference)であり、これは概念的・精神的な身象である。参照は多くの方法で結合され参照表現式(reference expression)に構成される。1つの基本状態を参照する参照表現式は基本メッセージ(e-message)と呼ばれ、情報システムにおける情報の基本単位である。属性型及び関係型の基本メッセージは次の3つ組で表現される。

$\langle p(\text{object}), p(\text{property}), p(\text{time}) \rangle$ | p : 参照
 $\langle \langle p(\text{obj-1}), \dots, p(\text{obj-n}) \rangle, p(\text{obj-relation}), p(\text{time}) \rangle$

次に対象は集合に類別され、値は属性として構成される。これに対応して基本状態を基本状態型に分類する。参照のレベルでは基本メッセージが基本メッセージ型に分類され、これは基本概念(elementary concept)と呼ばれる。

現実世界では、ある基本状態が他の原因及び結果であることがある。これは情報システムでは基本メッセージ間に先件(precedence)関係が存在することによって反映

されている。ここで基本プロセス (elementary process)とは先件関係にある基本メッセージからある基本メッセージを生成する手順として定義される。

情報システムの基本メッセージはデータベースに格納されている。情報論理データベースの主要なサブ・システムは核 (nucleus), スキーマ (schema) 及びフィルター (filter) で構成される。核とは物理的に存在するファイルの中のデータによって陽に表現されるような基本メッセージの集合である。スキーマは基本プロセスを表現し、核の中から基本メッセージを生成する権限を有する。フィルターはデータベース及び利用者を保護し、意味のないメッセージの生成を予防する。

1.6 2項論理結合模型

Bracchiの提唱したこの模型はANSI/SPARCの概念レベルで用いられることを想定している。論理結合の概念は企業管理者 (enterprise administrator) によって抱かれる論理的結び付きを表現している。数学的には1つの論理結合は7つの関係であり、概念の集合向に存在する結び付きを定義している。この模型では概念 (concept) が事象 (entity) を構成するのではない点に特徴がある。事象とは外部スキーマ (external schema) で考えられるものである。

この模型では論理結合はすべて2項に制限されており、論理結合の定義域である概念の集合はすべて基本定義域 (すなわち simple) である。n項関係を表現するために仮想の概念集合 (dummy set) が考えられ、仮想概念集合の内容は他の基本概念集合でもって表現される。

論理結合ではその関数性は概念レベルにおいては無視している。

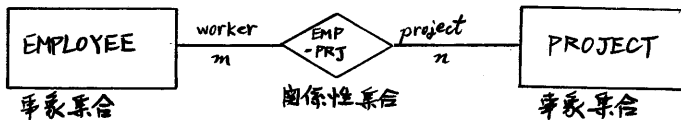
1.7 事象・関係性模型

この模型は現実世界で明確に識別される事物を示す事象の集合で構成される。関係性 (relationship)とは事象間の7つの連合 (association)であり、関係性集合 (relationship set) Rとは事象集合に属するn個の事象間に定義される数学的な関係である。7つの関係性の中の事象の役目 (role)はその関係性の中でその事象が機能する関数である。事象の関係性には2つの型がある。7つは弱事象関係 (weak entity relation)と呼ばれ親子の階層関係を表現する。もう7つは独立の事象間の関係を示すもので正規事象関係 (regular entity relation)と呼ばれる。

事象や関係性に関する情報は属性と値の対で表現される。属性は事象集合もしくは関係性集合から値集合 (もしくは値集合の直積) への関数として考えられる。

概念的情報構造は事象や関係性を参照するのに対し、情報構造のレベルでは事象と関係性は値で表現されている。このレベルでは事象と関係性は基本鍵 (primary key) の値でユニークに識別される。

この模型では次のような事象関係性図を用いて構造を表現する。



1.8 データ意味論

Abriaの提唱したデータ意味論模型では2項関係を基本要素として用いる。この2項関係は2つの対象 (object) のカテゴリ (category) 間の逆想として定義される。この関係は2つのアクセス関数で特定化される。これは定義域カテゴリと値域カテゴリの対応濃度 (cardinality) を示す最小・最大指示パラメータを持つている。 $(rel(A, B) \dots \alpha_{fn1}: A \rightarrow \mathcal{P}(B), \alpha_{fn2}: B \rightarrow \mathcal{P}(A))$

データベースは現実世界を映す模型であり、模型の状態 (state) はその知識を表現しているものである。知識は次の4つのクラスに分類される。① 基本事実 (elementary fact) ② 単純規則 (simple rule) ③ 複合規則 (complex rule) ④ 演繹規則 (deduced rule)。模型における対象 (object) は抽象的对象 (abstract object) と具体的対象 (concrete object) があり、7つの対象は他の対象との関係性で記述される。模型の行動にとって重要な対象には プログラム, プロセス, 文脈 (context), 及び ユーザ の概念があり、これらの意味は相互の関係性で形式的に記述される。この模型ではデータの意味記述はすべて2項関係とそのアクセス関数で表現され形式的な自己記述性 (self-descriptive) を目指している。データへの論理的アクセスはプログラムによってアクセス関数上で作用する基本演算 (elementary operation) を用いて実行される。また意味の拡張やデータに対する問い合せは1階述語計算と等価な手順 (procedure) を用いてなされる。

1.9 関数的データ模型

関数的模型の重要な構成要素は回路を含む名前付けされた有向マルチグラフである。グラフの節は 事象集合 (entity set) もしくは 抽象集合 (abstract set) を表わす。事象集合は現実世界の事象を模型化しており、抽象集合は事象集合間の結合を表現している。グラフの枝は 全関数 (total function) を意味している。7つの集合はその名前とその上で定義された関数によって記述され、この関数の集まりを 関数規定 (functional specification) と呼ぶ。7つの集合 S の鍵 (key) とは、 S からその関数の値域集合に対する写像が1対1になるような S の関数の集まりである。鍵が7つの関数で構成される時は 単純鍵 (simple key)、複数の関数で構成されている時は 複合鍵 (composite key) と呼ぶ。

関数模型に対する論理的アクセスは高水準の宣言的言語によって行なわれる。問い合せの指示はグラフ上を巡航 (navigate) することによってなされる。

1.10 意味ネットワーク模型

意味ネットワークは枝と節の両者が名前付けられた有向グラフである。特に枝の名前は関連した意味的性質を表わし、推論に有効的に用いられる。

この模型では知識を表現する節に4つの型がある。概念 (concept) は現実世界の本質的な定数やパラメータに対応し、物理的もしくは抽象的な対象を表現する。事象 (event) は現実世界での行動 (action) を表現し、ケース・グラム (case-grammar) で記述される。これはその事象が関係する幾つかの節を枝で結んだもので、各枝にはその役目を示す役目名 (role-name) が指示される。この役目名には 主格 agent (a), 目的 affected (aff), 関連 topic (t), 手段 instrument (i), 結果 result (r), 始点 source (s), 終点 destination (d), 対象 object (o) などがある。

性質 (characteristics) は状態を表現したり、概念や事象や他の性質を修飾するものに用いられる。7つの性質は定義域の要素を値域の値に対応させる2項関係である。性質から定義域の要素に向う枝には 'Ch' のラベルが、値に向う枝には 'V' のラベルが付される。性質には対応の関数性により多対多, 多対1, 1対多, 1対1の型がある。値節 (value-node) は性質の値域として機能する。

知識はシナリオ (scenario) として整えられる。シナリオは事象や性質や因果関係を示す数学的述語の集まりで構成され、データベースの意味的整合性を判定する型盤 (template) として機能する。

データベースに格納された関係に対する論理的アクセスには意味論的関係代数が用いられ、自然言語をこの模型の高級内い合せ言語とする方向を目指している。

1.11 拡張集合論模型

Childsの提唱した集合論的データ構造論(STDS... set theoretic data structure)では概念レベルから機械表現レベルまでを統一して形式化することが考えられている。この模型の基本的な対象は複合体(complex, quasi-set), 集合及びn個組である。2つの対象間に存在する基本的な関係はi要素関係(i-membership)である。xがyのi番目の要素である時 $x \in_i y$ と書く。複合体は整数と要素の対で構成されている。集合は1要素関係だけで成っている。1つのn個組では、1からnまでただ1つのi要素を持っている。要素の位置を添字で表現すると、複合体 $a = \{b^1, c^1, d^3\}$ は $b \in_1 a, c \in_1 a, d \in_3 a$ を意味している。3つ組 $\langle a, b, c \rangle$ は $\{a^1, b^2, c^3\}$ と表現でき、集合 $\{a, b, c\}$ は複合体構造の表現では $\{a^1, b^1, c^1\}$ となる。

拡張集合論における演算は通常の集合演算の一般化である。共通部分を求める演算は次のようになる。

$$C = A \cap B \stackrel{def}{\Rightarrow} \forall x \forall i (x \in_i C) \Leftrightarrow \forall x \forall i (x \in_i A \ \& \ x \in_i B)$$

組を集合から独立に定義した結果、集合演算が組に対しても意味を持つ。

$$\langle a, b, c \rangle \cup \langle a, b, d, e \rangle = \{a^1, b^2, c^3, d^3, e^4\}$$

$$\langle a, b \rangle \cap \langle c, b, e \rangle = \{b^2\}$$

Childsはほとんどの情報処理アプリケーションで充分と思われる拡張集合演算の系を提示している。

1.12 情報管理概念

情報管理概念(IMC)はユーザとDBMSの間の対象領域を模型化するものである。両者で交換される情報の基本的な単位は構造体(construct)と呼ばれる。分解不可能な最小の基本構造体は値(value)である。幾つかの構造体を結合して作られた構造体は集団(aggregate)と呼ばれる。集団は構成要素(Constituent)から構成されているという。模型に対し情報構造の導入と選択機能を与えるために構造体の構成要素に名前付けがなされる。名前付けがされた構成要素を持つ集団を有資格(nomination), 名づけられない構成要素を持つ集団を集まり(collection)と呼ぶ。教育的には有資格とは名前から構造体への関数であり、集まりは有限集合である。ある集合は個々の要素を判別する属性から集合自体を規定する属性を分離した1つの抽象化であることが出来る。この抽象化を表現した構造体の集合を型(type)と呼ぶ。型は型記述言語(TDL)で定義される。型の1つの要素をその型の成員(instance)と呼ぶ。TDLで与えられた型記述をスキーマと呼ぶ。各々の型は名前を割り当てられ、これを型表示(type designation)という。データベースの中には種々の構造体が格納されている。すべての構造体を構成要素として含む総括的な集団を考えると、これを枠組構造体(framework construct)と呼ぼう。情報処理の基本的な操作はこの枠組構造体の中で要素を位置づけ、枠組構造体自体を変更する要求を出すことである。

名前は有資格の選択に用いられる。有資格のネストしている構造体では各レベルの名前を指示する。これを識別子(identifier)という。集まりの中の構成要素は識別子を持たないので、属性で指示する。このような識別子や属性の1つの列は最終的に1つの構造体に辿りつくので、この列は点(spot)を定義しているという。以上のような概念に基づき、DBMSの基本的な関数が定義される。これらには構造体の入力、点の生成、点の名前の割当て、点の名前の削除などが含まれている。

1.73 情報代数模型

情報代数は CODASYL によって提唱され小林によって拡張された。この模型では現実の世界は属性 (property) を持った事象 (entity) によって構成されている。事象の各々の属性は値 (value) を持つ。小林の拡張では事象間には種々な関係性 (relationship) が定義される。この状態は情報空間 (information space) で表現される。情報空間はすべての属性に対応した値集合の直積である。事象間の関係性は情報空間上で定義された組で表現される。すべての n 項関係は基本的な 2 項関係に分解される。2 項関係そのものも事象と見なすことができる。

情報空間の任意の部分集合を領域 (area) と呼ぶ。重なり合わない 2 項関係の集合を構造 (structure) と呼ぶ。従って構造は有向グラフと考えることができる。構造は枝のパターンによって線型、木、ネットワークの 3 つの型に分類される。

情報処理の模型では 3 つの言語レベルが設定されている。中間レベルでは集合演算が定義される。これには通常の集合演算の他に、導入された構造を処理する新しい演算が提唱されている。これらの演算は点関数 (FOP... function of point), 線関数 (FOL... function of line), 領域関数 (FOA... function of area) を基礎とする。線は情報空間上の点の順序集合である。低レベル言語ではこれらの中間レベル操作の詳細化 (refinement) として要素アクセス (element by element) のレベルが定義され、上位レベル言語では中間レベルのより抽象化 (abstraction) として問題向き記述のレベルが定義されている。

1.74 データ空間模型

Bachman のデータ空間模型では、データベースは 3 次元空間上の点の集合として考えられる。3 つ組は事象 (entity), 属性 (attribute) 及び測度 (measure) と呼ばれる。各々の事象には事象軸上の値であるユニークな事象番号が割り当てられている。事象軸に垂直な平面は 1 つの事象に関するすべてのデータを表現しているのが同一事象面 (iso-entity plane) と呼ばれる。同様に同一属性面 (iso-attribute plane), 同一測度面 (iso-measure plane) が考えられる。

このようなデータベースに対して次のような操作が定義できる。

- ・ 同一平面上の点を選択
- ・ 同一線上の点を選択
- ・ 2 線で定義される面を構成
- ・ 2 面で定義される線を構成
- ・ 線に垂直な同一面を構成 etc.

1.75 Lindgreen 模型

この模型ではデータベースは何らかの外部システムの構図であると考ええる。システムは相互に関係する事象 (entity) で構成され、各事象は属性 (property) を持ち、各々の属性は値 (value) を持つ。情報の基本的な要素は $\langle e, p, v \rangle$ で表わされる 3 つ組である。このような 3 つ組の集まりを情報集合 (information set) と呼ぶ。同じ属性要素を持つ情報集合を同型 (isotypic) 情報集合 と呼び、同じ事象要素を持つ情報集合を同義 (isonyms) 情報集合 と呼ぶ。等一情報集合 (HIS... homogeneous information set) とは、同じ属性の集合を持った同質情報集合に分割できるような情報集合である。

論理的アクセスは 1 つの単項演算と 6 つの 2 項演算で指示される。これには、抽出 (extraction), 合併 (conglomeration), 分離 (separation), 演繹 (derivation), 探査 (detection), 表出 (exposition), 包含 (inclusion) がある。

2. データ模型の分類

前述した 15 個の模型に対し分類を試みる。CSRQ のアプローチでは、検討

項目を①構造的比較 ②論理アクセスの方法 ③概念の比較 ④意味的比較の4つに分けて行っている。

2.1 構造的比較

模型の構造を検討する際に各々の模型をグラフ理論的側面、集合論的側面、数学的構造の側面にわたって考察している。グラフ理論的検討では節の構造を分類し、構造を持つ節と構造を持たない節の関係が考察されている。また各模型において節や枝に何を比定しているかが検討され、枝の関数性が重要な点として分析されている。集合論的検討では集合を構成する要素である組の次数と、要素がそれ自身集合であるか否か(集合のネスト)の判別が考えられている。数学的構造の検討では各々の模型で受容される対象間の関係性を考察し、関係定義を2項、3項および2項+3項のタイプに分類している。構造的比較の結果はAPPENDIX表1, 表2, 表3に掲げられている。

2.2 論理アクセス

データベースへの論理アクセスの検討では、アクセスの型を要素アクセス型(element-at-a-time)と集合アクセス型(set-at-a-time)に分けている。要素アクセス型は通常の巡航タイプが含まれる。集合アクセス型には代数型(algebra)、写像型(mapping)、及び計算型(calculus)がある。代数型には関係代数、情報代数、Lindgreen模型、拡張集合論模型が類別されている。写像型にはDIAM-II、2項論理結合、関数的模型、一般ネットワーク、関係模型(SEQUEL型)、情報代数(小林の構造処理)が含まれている。計算型には階層型(system 2000, TDMS)事象・関係性模型及び関係計算が含まれている。

2.3 概念の比較

各々の模型で用いられている概念の比較を行っている。ここでの基準として設定されているのは事象(entity)、属性(property)、値(value)および関係性(relationship)である。各々の模型でこれらの概念と等価であると考えられる用語を表4に整めてある。

2.4 意味的比較

意味をどのように定義し把握するかには種々な議論がある。CSRQのアプローチは言語学(linguistic)的な方法を採用している。意味に対する言語学的アプローチにPredication Analysisがある。これは文の意味的構造を扱うものである。叙述文の解析によってその叙述文の構造が理解される。ある模型で主文と関連叙述文をどのように表現しているかが理解できれば、その模型が意味的叙述構造を情報構造にどのように写像しているかが判断できる。各模型が表現している情報構造は種々な意味的な抽象化のレベルを体現している。これは表相意味から深層意味までをカバーする。従って意味スペクトル(semantic spectrum)を導入して各々のデータ模型を考察することが出来る。このような方法により、CSRQは表相意味の模型として事象・関係性模型、意味ネットワークを掲げ、深層意味の模型としてAbrial, DIAM-II, 階層, 2項論理結合模型を掲げている。

おわりに

本稿で抄訳したデータ模型の分類学は、データ模型の洪水の中で相互の位置づけを目指した点で多くの有益な指摘を含んでいるものと思われる。ここでとりあげられている模型は情報処理学会データベース・モデル研究委員会でも研究対象とされたものである。本稿を整める上で研究委員の方々に多大の援助を蒙り、未算ながら感謝の意を表させていただきます。

APPENDIX

Table 1: Graph Theoretic Models

	Model	Node	Node Element	Edge	Functionality
structured node elements	DBTG	Record Type	Record	(Co) Set Type	member→owner (against arrow)
	Hierarchical	Segment Type	Segment	Parent/Child Relationship	child→parent (against arrow)
	General Network	Record Type	Record	Link	none
	Information Algebra (Kobayashi)	Area	Point	Arc	none
unstructured node elements	DIAM-II	Identifier	Id. Value	Association Pair	none
	Binary Logical Associations	Set of Concepts	Concept	Logical Association	none
	Abrial's	Category	Object	Relation	none
	Functional	Set	Entity, Abstract Entity, Value	Function	domain→range (with arrow)
multiple node types	Entity-Relationship	Value Set Entity Set Relationship Set	Value Entity Relationship	Attribute Role	Attr:ES→VS Attr:RS→VS Role:RS→ES
	Semantic Network	Value Node Concept Node Event Node Characteristic Node	Value "Entity" Event Property	Characteristic Role Characteristic	ch:Char.node→Concept ch:Char.node→Event v:Char.node→Value

Table 2: Set Theoretic Models

Model	Tuple Size	Nesting
Relational	arbitrary	flat
Infological	arbitrary	flat
Information Algebra	arbitrary	flat
IMC	arbitrary	nested
Extended Set Theory	arbitrary	nested
Hierarchical Relations	arbitrary	nested
DIAM-I	triplet	flat
Data Space	triplet	flat
Lindgreen's	triplet	flat

Table 3: Mathematical Structure-Binary and/or N-ary

Binary	N-ary	Binary + N-ary
Data	Relational	DBTG
Lindgreen's	Information Algebra (CODASYL)	Information Algebra (Kobayashi)
DIAM-I	Infological	General Network
DIAM-II		Semantic Network
Binary Logical Association		Hierarchical
Abrial's		Hierarchical Relations
Functional		Information Management Concepts
		Extended Set Theory

Table 4: Conceptual Equivalents & Terminology

Model	Entity	Relationship	Type*	Property/Applicable to
Abrial-Data Semantics	Object	Relation	b	_____
DIAM I & II	Entity	Association	b	_____
Lindgreen's	Entity	_____	—	Property/Entity
Data Space	Entity	_____	—	Attribute/Entity
Information Algebra (CODASYL)	Entity	_____	—	Property/Entity
Information Algebra (Kobayashi)	Entity	Arc	b	Property/Entity
Benci et. al.	Object	Association	n	Property/Object, Association
Entity-Relationship	Entity	Relationship	n	Attribute/Entity, Relationship
Infological	Object	Relation	n	Property/Object
Schmid & Swenson	Indep. Object	Association	n	Characteristic/Indep. or Char. Object
Semantic Network	Concept	Event	n	Characteristic/Concept Event, Characteristic.
Functional	Entity	Abstract Entity	n	Function/Entity or Ab. Entity

* b denotes a binary relationship
n denotes an n-ary relationship

参考文献

- [1] L. Kerschberg, A. Klug, D. Tsichritzis, "A Taxonomy of Data Models," Univ. Toronto, CSRG-70, 1976
- [2] Codd, E.F., "A relational model of data for large shared data banks", CACM 13, 1970
- [3] Codd, E.F., "Relational completeness of data base sublanguages", Data Base Systems, 1971
- [4] CODASYL, "Data Base Task Report", ACM, New York, 1971
- [5] Senko, M.F., "Specification of Stored data structures and desired output Results in DIAM-II with FORAL", Conf. on V.L.D.B., 1975
- [6] Langefors, B., "Information Systems", Proc IFIP-74, North Holland, 1974
- [7] Sundgren, B., "Conceptual Foundation of the Infological Approach to Data Bases", Data Base Management, J.W. Klimbie ed., North Holland, 1974
- [8] Bracchi, G., Rolini, P., Pelagatti, G., "Binary Logical Associations in Data Modelling", Proc. IFIP-TC-2-WC, Black Forest, 1976
- [9] Chen, P., "The Entity-Relationship Model: Toward a Unified View of Data", TODS, 7, 1976
- [10] Abrial, J.R., "Data Semantics", Data Base Management, J.W. Klimbie ed., North Holland, 1974
- [11] Nijssen, G.M., "Set and CODASYL Set or Coset", Data Base Description, North Holland, 1975
- [12] Kerschberg, L & Pacheco, J.E.S., "A Functional Data Base Model", 1975
- [13] Roussopoulos, N., Mylopoulos, J., "Using Semantic Networks for Data Base Management", V.L.D.B., 1975
- [14] Childs, D.L., "Feasibility of a Set-Theoretic Data Structure", IFIP-68, 1968
- [15] Durchholz, R., Richter, G., "Information Management Concepts for Use with DBMS interface", IFIP-TC-2
- [16] Kobayashi, I., "Information and Information Processing structure", Information Systems 1, 1975
- [17] CODASYL, "An Information Algebra", CACM 5, 1962
- [18] Bachman, C.W., "Data Space mapped into three Dimensions", Infotech State of the Art Report 15, 1973
- [19] Lindgreen, P., "Basic Operations on Information as a Basis for Data Base Design", Proc. IFIP-74, 1974