

データ・セマンティクスの回帰表現モデル

石田 喬也 (三菱電機)

1. まえがき

現実世界の情報を、その意味を極力正しくデータの形に表現して扱うことを意図するデータ・セマンティクスの問題は、データ・ベースにと、本質的に不可避なテーマである⁽¹⁾。しかし極の二最近までこのテーマが余り注目されることになかったのは、データ・ベースを実用性あるツールと見なされることか、さしあつりの最大の関心事であつた点のせいであると思われる。CODASYL DBTGレポート⁽²⁾がその問題にひとつの結論を与えて後、その方向への批判も合せて、改めば理想論の議論が活発化したと言える。その旗頭となつてゐるのがリレーショナル・モデルである。しかしそのオリジナル論文⁽³⁾でE. F. Coddが自ら述べてゐるように、彼がこのモデルを提案した動機はあくまでデータ独立性にある。より正規形⁽⁴⁾もその観点から導出されたもので、関数従属性の概念もその目的から取り入れられたにすぎない。勿論データ独立性の追求は、データ・セマンティクスを回避してあり得ず、関数依存性概念は明らかに大きな成果であり、リレーショナル・モデルがデータ・セマンティクスの問題を喚起したことは事実であると思われる。

データ・セマンティクスの立場から、リレーショナル・モデルに關し多くの疑問が投げかけられており、かつそれら問題点を克服すると標榜するいくつかの新たなモデルが提案されてゐる。^{(5),(6),(7),(8)} その中でもM. E. SENKOは当初のDIAM-Iモデル提案⁽⁹⁾のときから、リレーショナル・モデルとは独立にデータ・セマンティクスの問題を深く追求してあり、いわばこの分野の先駆者と言える。またJ. R. ABRIALは⁽¹⁰⁾ PLANNER流の発想を取り入れて、データ・セマンティクスの自己記述的な、明快なフォーマイズをおこなつた点での功績が大きい。

データ・セマンティクス記述の基本はENTITY概念にある。具象的であれ、概念的であれ、現実世界で互に区別して識別され得るすべてのもののひとつひとつがENTITYであると呼ばれる。ENTITYに關する情報の記述をデータ・ベース記述の中核に据えることの必要性を最初に主張したのは、GUIDE/SHAREのDBMS要望レポート⁽¹¹⁾である。ANSI/SPARCレポート⁽¹²⁾における、CONCEPTUAL SCHEMA思想は明らかにこの主張を継承したものである。つまり、ENTITYこそすべての記述の基盤であるが、従来のデータ・セマンティクス議論において最も問題にされるべきなのは、この基盤そのものに一貫性が欠けてゐることにある。ENTITYの定義を純粋に貫くならば、属性、値、関係、タイプ、カテゴリ等、ENTITY以外の余命な概念を一切必要としないはずだからである。そうであるならば、そのモデルは単にデータ・ベースにとどまらず、環境世界をも含めて統一的に記述できるグローバル性を持つべきである。

筆者らは以前⁽¹³⁾、FILE と VOLUME というふたつの概念を独自に定義し、それによつて計算機システムを含め、情報処理システム一般を統一的に記述することを試みた。ここでは、ENTITY というより根元的概念を使つてそのモデルを再構成し、データ・セマンティクス記述の核モデルを形成することを試みる。

2. 核モデルの条件とその意義

いかなるデータ・セマンティクスをも、あるモデルの回帰適用によつて表現できるとすれば、それはデータ・セマンティクスの核モデルであると呼ぶことにする。核モデルである条件は以下の通り。

- (1) すべてのデータ・ベース・モデル (CODASYL, リレーショナル, DIAM, 等) を表現可能である。
- (2) データ・ベースだけでなく、DBMS 自体、さらにはそれととりまく外界自体も表現可能である。
- (3) DBMS がいくつかの階層から構成される (下と之は、ANSI/SPARC レポートの DBMS では、EXTERNAL, CONCEPTUAL ならびに INTERNAL の3階層から成る) 場合、どの階層をも共通的に表現可能である。
- (4) その表現できる世界の広がりに限界がなく、表現する対象がいかに大きくとも、いかに小さくとも共通的に表現可能である。
- (5) 人間の情報処理方式や思考方式に則している。

このような核モデルがあるとすれば、データ・ベースあるいはそれをひとつの要素として持つ計算機システムは核モデルによつて共通的に表現される。それらは、やはりより大核モデルで表現される、それらととりまく人間を含めたより大きな情報システムの、断絶がない単なるひとつの構成要素として位置づけられる。

データ・ベースが、核モデルの回帰構造で表現されるならば、データ・ベースのすべての管理やデータ・ベースに対するアクセスは、いわば核処理マシンと呼ばれるものの回帰的適用によつて処理することができるとする。したがつて DBMS の構成は単純化され、またデータ・ベース専用プロセッサの考へ方に結びつけることも可能になるであろう。

一方、人間の概念認識の仕方を思い起すならば、内包 (INTENSION) と外延 (EXTENSION) の回帰がその基本となっていることを知る。内包とは概念の意味であり、外延は内包によつて示される適用条件が適用される対象のクラスである。したがつて、内包と外延の一对を核モデルの核部分に対応せしめるならば、それは人間の思考方式に則したものであるであろう。⁽¹⁴⁾

また人間のものごとの認識はすべて相対的であることを留意する必要がある。ものごとを絶対的にユニークに識別することの要請は、人間の情報処理方式に則さない。核モデルはその回帰性によつて、データ・ベース中のすべてのデータを相対的に位置づけることを可能ならしめる。そのためにこの核モデルのユーザは、ある時点で常にひとつの文脈 (CONTEXT) が与えられる。内包と外延の一对を核とするモデルでは、文脈は特定の回帰階層の内包に対応する。

3. ENTITY 概念の3レベルとその回帰性

ENTITY は3種類のレベルで認識されることが出来る。ひとつは ENTITY の実体そのものであり、これを R-ENTITY (REAL ENTITY) と名付ける。次のレベルはそれが抽象化された一点として認識されるものであり、これを A-ENTITY (ABSTRACTED ENTITY) と名付ける。次のレベルは A-ENTITY が具体化表現されるものであり、これを M-ENTITY (MATERIALIZED ENTITY) と名付けることにする。

R-ENTITY のひとつひとつは、物理空間内あるいは概念空間内に、あるひらがりを占めるものとして客観的に存在する。通常 ENTITY と呼ばれるものは、この R-ENTITY に対応する。

R-ENTITY の、そのような空間に占めるひらがりがいかに大きくとも、あるいはいかに小さくとも、我々はそれらのおのおのを他とは異なる点としての存在として識別することが出来る。かくしてそれぞれの R-ENTITY を、我々は一点で代表せしめることをおこなう。その各点が A-ENTITY である。

A-ENTITY は、それがひとつの R-ENTITY から抽象されて存在せしめられると、その ENTITY が R-ENTITY としての存在ではもはや消滅した後にも、存在し続けるものであることに注意する必要がある。

A-ENTITY は抽象空間内の一点であり、他の A-ENTITY から異なる一点として区別される以上のほにもものでもない。ここで、A-ENTITY のひとつの点集合を、特に FILE と名付けることにする。

A-ENTITY の点のひとつの集まりである FILE を、一種のざる状のいれもので可くいとるといふ、観念上の操作を施すことを想定してみる。その結果、そのいれものの中には FILE の中のいくつかの A-ENTITY のみが可くいとられ、他の A-ENTITY はざるの目から落ちこぼれることになるであろう。そのとき、可くいとられた A-ENTITY は、そのいれものに通して具体化されたと考える。このような観念上のいれものを、特に VOLUME と名付けることにする。

VOLUME によつて可くいとられた、可かわる VOLUME の中に具体化された A-ENTITY が M-ENTITY であり、FILE の中の A-ENTITY とは区別される。

A-ENTITY は R-ENTITY に対応して唯一であるが、A-ENTITY に対応する M-ENTITY は、それが可くいとられる VOLUME の数だけ存在する。

図-1 にこれら3種類の ENTITY 間の関係を観念的に図示する。

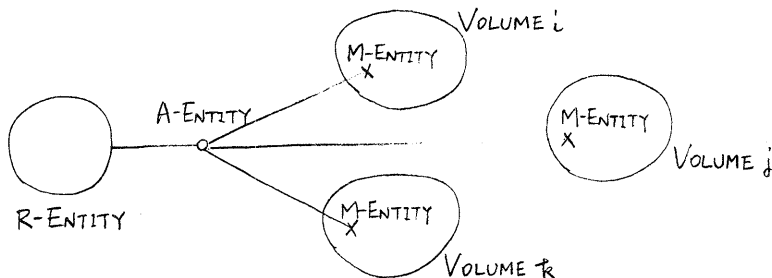


図-1 ENTITY 表現の3レベル

VOLUMEは対象世界 (UNIVERSE OF DISCOURSE) に唯一の最高位VOLUMEを例外として、常に特定唯一の直接上位 (IMMEDIATE SUPERIOR) のVOLUMEの中に包含されて存在する。対象世界のFILE中のA-ENTITYは、最高位VOLUMEによってすべてつくられる、すなわちすべて具体化されることが出来る。

一般にひとつのVOLUMEの中に具体化されるA-ENTITYは、そのVOLUMEの直接下位 (IMMEDIATE INFERIOR) のいくつかのVOLUMEのどれかひとつの中に具体化される。このようにして、VOLUMEによるA-ENTITYの論理的分割が徐々に進められる。この分割の仕方は、VOLUMEの選び方によって幾通りもあることに注意する必要がある。

VOLUMEもそれ自体、他から区別して認識され得るひとつの存在、すなわちひとつのENTITYである。対象世界のVOLUME階層の中のVとつVOLUME V_i に着目すれば、 V_i は V_i であるENTITYのR-ENTITYである。一方 V_i の直接上位VOLUMEを V_{i-1} とすれば、 V_i は同時にVとつENTITYのA-ENTITY e_i の、VOLUME V_{i-1} の中に具体化されたM-ENTITYである、とみなすことができる (図-2参照)。

この関係を、 V_i なるENTITYが、 V_{i-1} なるENTITYに e_i なる限定化が施された結果として作られている、と理解することになる。そしてこのことを次の式で表現する。

$$V_i = e_i \cdot V_{i-1}$$

一般的には次の式で表現される。

$$V_i = \prod_{j=1}^k e_{i-j+1} \cdot V_{i-k}$$

$$V_i = \prod_{j=1}^i e_{i-j+1} \cdot V_0$$

ただし、

V_{i-k} ; V_i の上位直接上位VOLUME

V_0 ; 対象世界の最高位VOLUME

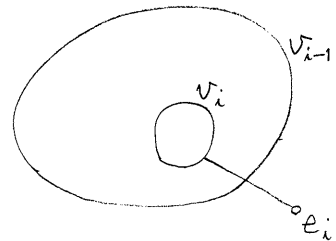
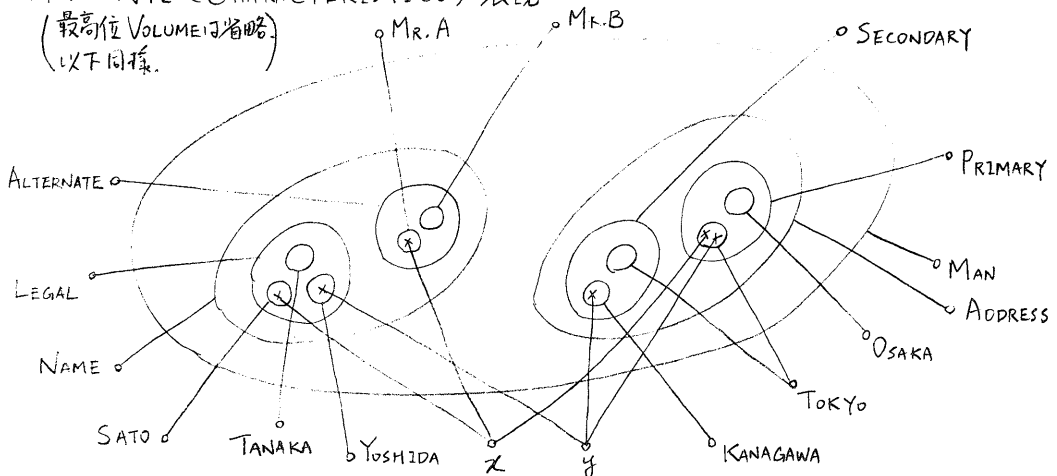


図-2. VOLUMEの限定化

例1. 特性 (CHARACTERISTICS) 表現

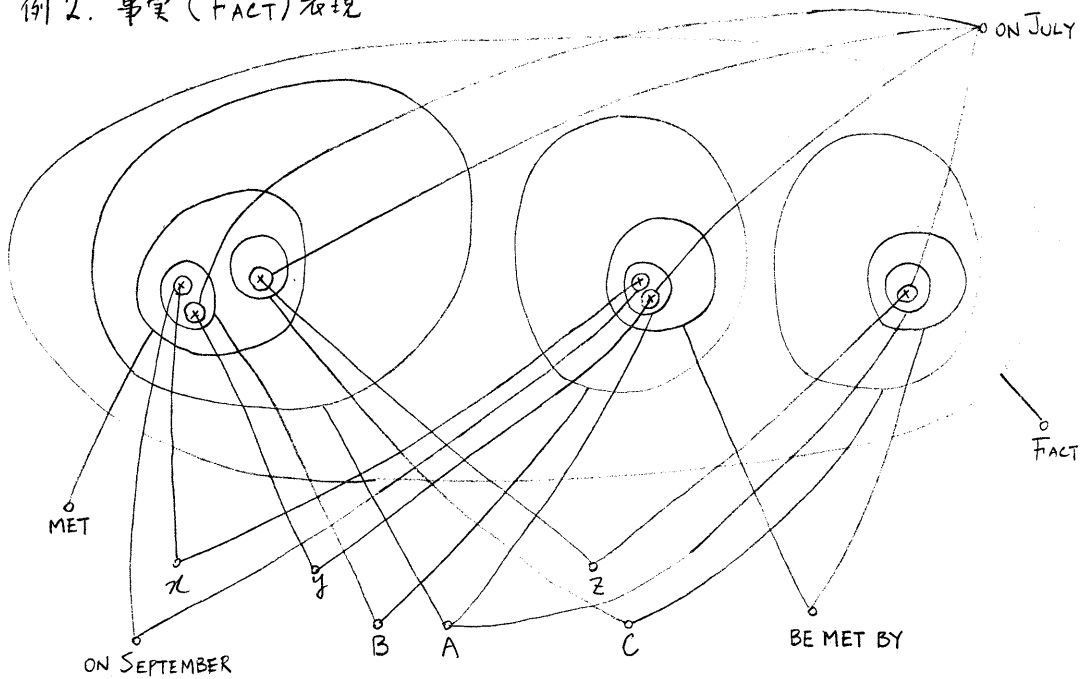
(最高位VOLUMEは省略)
以下同様。



ENTITY xは、LEGAL NAME=SATO, ALTERNATE NAME=MR.A, PRIMARY ADDRESS=TOKYO, のMANである。

ENTITY yは、LEGAL NAME=YOSHIDA, PRIMARY ADDRESS=TOKYO, SECONDARY ADDRESS=KANAGAWA, のMANである。

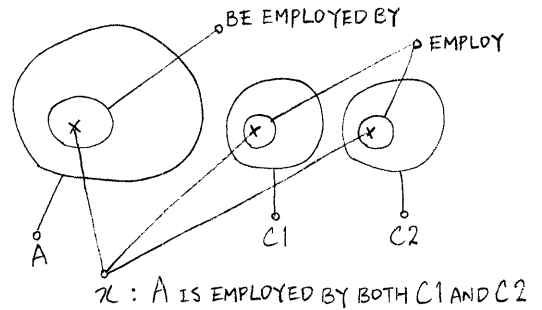
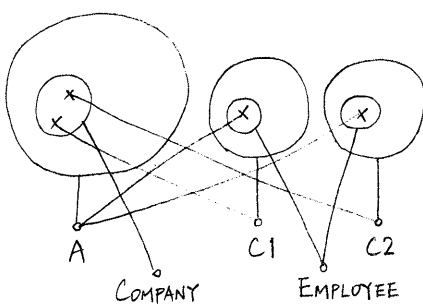
例2. 事実 (FACT) 表現



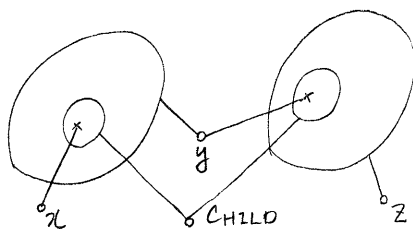
- x : A MET B ON JULY (B WAS MET BY A ON JULY)
 y : A MET B ON SEPTEMBER (B WAS MET BY A ON SEPTEMBER)
 z : A MET C ON JULY (C WAS MET BY A ON JULY)

例3. 関係 (RELATION) 表現

ENTITY 間の関係は一般に事実の形で表現可る方が望ましい。
 下図は左の関係表現より、右の事実表現の方が関係を顕在化するので良い。



例4. 関係推論 (RELATIONAL INFERENCE)⁽¹⁵⁾ 表現



$$\begin{aligned}
 y &= \text{CHILD} \cdot Z \\
 x &= \text{CHILD} \cdot y \\
 &\downarrow \\
 x &= \text{CHILD} \cdot \text{CHILD} \cdot Z \\
 &= \text{GRANDCHILD} \cdot Z
 \end{aligned}$$

4. 回帰表現モデルに基づく情報処理システム

このような回帰表現モデルは、単にデー々・ベースのセマンティクスの表現の世界へのみ適用が限定されるものではない。前述した意味の核モデルのひとつであることとすることができる。

デー々・ベース・モデルと呼ばれるものは、ほんらかの形でデー々のセマンティクスを表現すること意図しており、その表現の仕方は多様であるが、これらの表現はいずれも、A-ENTITYのVOLUME中への具体化、として抽象することができるであろう。

回帰表現モデルにおいては、DBMSとそれをとり囲む外界は同質(HOMOGENEOUS)である。ひとつのDBMS全体はひとつのVOLUMEであり、その外界はそのVOLUMEを包含するより下きなひとつのVOLUMEである。現実世界全体をひとつのVOLUMEとみれば、DBMSや人間そのものを含めて現実界に多数存在する情報処理システムのそれぞれが、そのVOLUMEの下位VOLUMEのひとつとして位置づけられる。

一方ひとつのDBMSに視点を移すと、ANSI/SPARCレポートに従うものであれば、その中ではデー々はINTERNAL MODEL空間、CONCEPTUAL MODEL空間、さらにEXTERNAL MODEL空間という全く異質の三種の空間に同時に位置を占めていることになる。このことはしかし、デー々に対応するA-ENTITYが具体化されるVOLUMEが、この空間の教だけあるケースと、回帰表現モデル内では本質的に区別されない。

このようにして、回帰表現モデルはDBMSの外なる世界も内なる世界も一貫して同質化することによって、情報受け渡しの壁を取り除く。

VOLUMEであるENTITYのR-ENTITYは、物理的のいふことであるが概念的のいふことであるかのいすれかである。現実世界の物理空間の任意の部分は、前者のタイプのひとつのVOLUMEとすることができる。また現実世界で認識される任意の概念の内包が、後者のタイプのひとつのVOLUMEとすることができる。

ここであらためてR-ENTITYとは何かを問われる。R-ENTITYとは物理的にいる、概念的にいる、現実世界の中にあるのちかりを占めるものである。このことはすなわち、R-ENTITYが上記のいすれかのタイプのVOLUMEであることにすぎない。VOLUMEはまた前述のように、その直接上位VOLUMEの中に具体化されたひとつのM-ENTITYであるにすぎない。この帰結はR-ENTITYの概念を不用とする。すなわちENTITYに一見、実体なる側面があるが、それは実際には、現実世界というVOLUMEの中に具体化された、そのA-ENTITYのM-ENTITYである。

参考文献

- (1) 石田尚也: データベースのセマンティクスとインプリメンテーション, 情報処理学会データベース研究会資料, 74-10 (1975).
- (2) DATA BASE TASK GROUP REPORT TO THE CODASYL PROGRAMMING LANGUAGE COMMITTEE, ACM, NEW YORK (1971).
- (3) E. F. CODD: A RELATIONAL MODEL OF DATA FOR LARGE SHARED DATA BANKS, COMM. ACM, VOL. 13, No. 6, PP. 377~387 (1970).
- (4) E. F. CODD: FURTHER NORMALIZATION OF THE DATA BASE RELATIONAL MODEL, COURANT COMPUTER SCIENCE SYMPOSIA, VOL. 6, DATA BASE SYSTEMS, PRENTICE-HALL, NEW YORK (1971).
- (5) H. A. SCHMID ET AL.: ON THE SEMANTICS OF THE RELATIONAL DATA MODEL, PROCEEDINGS OF 1975 ACM SIGMOD INTERNATIONAL CONFERENCE ON THE MANAGEMENT OF DATA (1975).
- (6) M. E. SENKO: SPECIFICATION OF STORED DATA STRUCTURES AND DESIRED OUTPUT RESULTS IN DIAM-II WITH FORAL, INTERNATIONAL CONFERENCE ON VERY LARGE DATA BASES (1975).
- (7) N. ROUSSOPOULOS ET AL.: USING SEMANTIC NETWORKS FOR DATA BASE MANAGEMENT, INTERNATIONAL CONFERENCE ON VERY LARGE DATA BASES (1975).
- (8) P. P. CHEN: THE ENTITY-RELATIONSHIP MODEL TOWARD A UNIFIED VIEW OF DATA, ACM TRANSACTIONS ON DATA BASE SYSTEMS, VOL. 1, No. 1, PP. 9~36 (1976).
- (9) M. E. SENKO ET AL.: DATA STRUCTURES AND ACCESSING IN DATABASE SYSTEMS, IBM SYSTEMS JOURNAL, PP. 30~93 (1973).
- (10) J. R. ABRIAL: DATA SEMANTICS, DATA BASE MANAGEMENT, NORTH HOLLAND (1974).
- (11) THE JOINT GUIDE AND SHARE DATA BASE REQUIREMENTS GROUP: REQUIREMENTS FOR A DATA BASE MANAGEMENT SYSTEMS, GUIDE, CHICAGO (1970).
- (12) ANSI/X3/SPARC STUDY GROUP ON DATA BASE MANAGEMENT SYSTEMS INTERIM REPORT 75-02-08, ANSI DOC. No. 7514 TS01 (1975).
- (13) 坂, 石田, 飯川: 計算機システムの論理体系に関する一考察, 情報処理, VOL. 11, No. 5, PP. 260~268 (1970).
- (14) 石田尚也: データベースにおけるデータの意味表現, 昭和50年度情報処理学会第16回大会 (1975).
- (15) DATA BASE DIRECTIONS, PROCEEDINGS OF THE WORKSHOP OF THE NATIONAL BUREAU OF STANDARDS AND THE ASSOCIATION FOR COMPUTING MACHINERY, PP. 99 (1975).