

データベースにおける意味表現とその述語論理による管理

5F-4

片山 紀生
東京大学大学院

高須 淳宏
学術情報センター研究開発部

1はじめに

データベースはデータを蓄積する機能と同時にそれらの意味を表現する機能を持っている。例えばRDBやOODBでは、リレーションやオブジェクトといったデータ構造のどの位置に格納されるかによってデータの意味が決まる。一般にデータが持っている意味は、意味ネットワークのようなネットワーク構造によって表現されるため、RDBやOODBのようなデータ構造による意味表現はデータがある程度画一的な場合には有効だが、多種多様なデータを取り扱う場合には適していない。そのため、構造化文書を取り扱うフルテキストデータベースなどデータの種類や構造に多様性がある場合、ネットワーク構造の意味表現を直接管理することが必要になる。

そこで本稿ではネットワーク構造の意味表現を直接的に管理する手法として述語論理に基づく方法を提案する。

2意象モデルによるデータの意味表現

我々はネットワーク構造を持つ意味表現を管理するためのデータモデルとして、意象モデル[1, 2]を提案している。意象モデルは、以下の四つの構成要素から成る。

意象(S-object): 意象の役割はデータベースにおける意味的な対象を一意に表現することにある。意象は意味表現のネットワーク構造においてノードになる。

物象(P-data): 物象は抽象データ型に基づいてカプセル化された物理データである。それぞれの物象はその意味を表現するための意象をひとつだけ持ち、物象と意象との間に1対1の対を作る。

意関(S-relation): 意関の役割は意象相互に存在する意味的な関連を一意に表現することにある。例えば「Xの著者がYである」という意味的な関連は、下のような述語によって表現する。

Is-Author-Of [Person = X, Document = Y]

項の値として意象を与えることによって意象間のリンクを定義する。例えば「(J.Smith)が(Paper A)の著者である」というリンクは下のように記述する(図1)。

Is-Author-Of [Person = (J.Smith),
Document = (Paper A)]

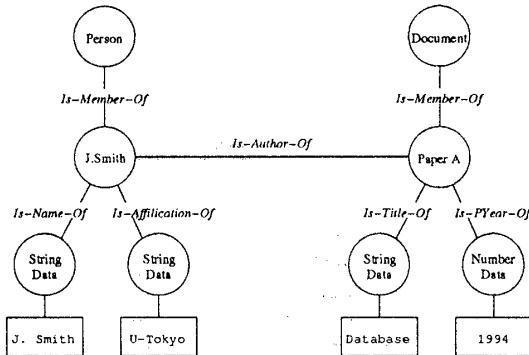


図1：著者と文書の間の意味的な関連

意約(S-constraint): 意味表現の一貫性を維持する役割を持つのが意約である。意約は次の三種類に分類される。矛盾と等価の例を図2に示す。

矛盾: $\leftarrow \Gamma$

導出: $R \leftarrow \Gamma$

等価: $\Gamma_1 \leftrightarrow \Gamma_2$

ここで、 R は意関、 $\Gamma, \Gamma_1, \Gamma_2$ は意関の状態である。意関の状態とは「意関にどのようなリンクが定義されているか」を表し、述語論理に基づいて記述する。すなわち、述語、変数、定数、論理否定(\neg)、論理積(\wedge)を用いて記述する。述語は、該当するリンクが存在するとき真、存在しないとき偽と解釈する。

導出については、演繹データベース[3]で用いられている層化プログラム(Stratified Program)[4, 5]の概念を用いる。層化プログラムでは \leftarrow を論理的な含意ではなく新たな述語の導出と解釈し、述語間に層状の導出関係があるという条件を設ける。すると、ボディに論理否定を含んでいても推論が単調推論になり不動点が存在する。

3計算手続きによって定義される述語

3.1機能的意関

物理データの間には、下のIs-Sum-OfやIs-Substring-Ofなど計算手続きによって定義される関連が存在する。

Is-Sum-Of [Sum = 3, Value1 = 1, Value2 = 2]
Is-Substring-Of [Part = "BC", Whole = "ABC"]

このような関連も意味的な関連の一種としてみなせるため、意象モデルでは意関のひとつとして取り扱い、機能的意関と呼ぶ。機能的意関は計算によってリンクを生成するため、項に値を入れない限りリンクを特定することはできない。

A Semantic Representation in a Database and Its Management with Predicate Logic

Norio Katayama¹, Atsuhiro Takasu², Jun Adachi²

¹University of Tokyo

²National Center for Science Information Systems

矛盾: Is-Author-Of の「Person」項に入る意象は、人間でなければならない。

$\leftarrow \text{Is-Author-Of} [\text{Person} = X, \text{Document} = Y] \wedge \neg \text{Is-Member-Of} [\text{Set} = \{\text{Person}\}, \text{Member} = X].$

導出: Is-GrandFather-Of は、Is-Father-Of から導出される。

$\begin{aligned} &\text{Is-GrandFather-Of} [\text{GrandFather} = X, \text{GrandChild} = Z] \\ &\quad \leftarrow \text{Is-Father-Of} [\text{Father} = X, \text{Child} = Y] \wedge \text{Is-Father-Of} [\text{Father} = Y, \text{Child} = Z]. \end{aligned}$

図 2: 意約 (S-constraint) の例

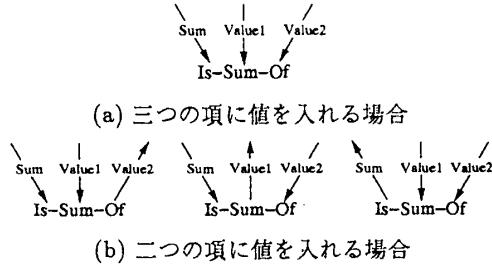


図 3: Is-Sum-Of 意関の評価法の表現

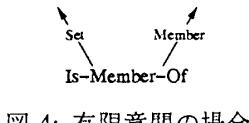


図 4: 有限意関の場合

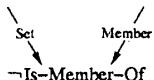


図 5: 否定を伴う場合

3.2 意関の状態の評価方法

意約では述語と論理記号を用いて意関の状態を記述する。これを評価するには意関に定義されているリンクを調べていけばよいが、機能的意関は直接評価することができないため、どの意関から評価すればよいのか評価手順を見つけ出す必要がある。

意関の評価法の表現

例えば、下のような加算の関係を表す機能的意関の場合、 X, Y, Z の三つの変数に値を入れて関連が成り立つかどうか判定させるか、どれか二つの変数に値を入れて残りの変数の値を特定させるかによって述語を評価することができる。しかし、ひとつの変数に値を入れて残りの変数の値を特定させることはできない。

$\text{Is-Sum-Of} [\text{Sum} = Z, \text{Value1} = X, \text{Value2} = Y]$

そこで、意関ごとにどのような評価が可能なのか、図 3 のようなグラフによって表現することにする(これを評価グラフと呼ぶ)。矢印が項への値の入れ方を示している。

定義されているリンクの数が有限な意関の場合には、すべての項の値を特定することができるので、図 4 のようにすべて外向きの矢印になる。

また、否定 (\neg) を伴う意関の場合には、すべての項に値が与えられないとリンクの有無を判定できないので、図 5 のようにすべて内向きの矢印になる。

評価手順の探索

評価手順の探索は、次のようにして行う。

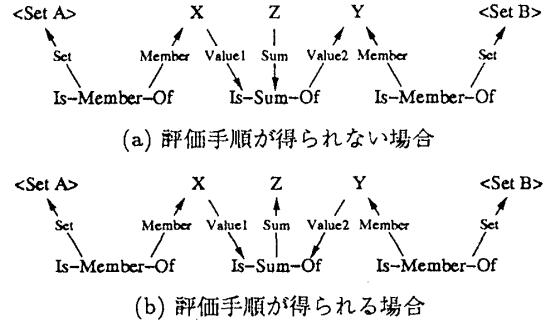


図 6: 評価手順の探索

- (1) 無限意関の評価グラフのあらゆる組合せについて、以下の処理を繰り返す。
- (2) 各意関の評価グラフを変数を介して連結する。
- (3) 定数および有限意関から出発し、矢印を辿っていくことによって特定できる変数を見つけていく。
- (4) すべての変数を特定できれば、その経路が評価手順を与える。

例えば下の記述の場合は図 6 のようなグラフが得られる。

$\begin{aligned} &\text{Is-Member-Of} [\text{Set} = \{\text{Set A}\}, \text{Member} = X] \wedge \\ &\text{Is-Member-Of} [\text{Set} = \{\text{Set B}\}, \text{Member} = Y] \wedge \\ &\text{Is-Sum-Of} [\text{Sum} = Z, \text{Value1} = X, \text{Value2} = Y] \end{aligned}$

図 6(a) の場合、変数 Z を特定することができない。図 6(b) の場合には、すべての変数を特定できるので、このグラフが評価手順を与える。

4 おわりに

現在、本手法の妥当性を検証するためにプロトタイプシステムを開発しており、今後、実装上の問題点などについて検討を進めていく考えである。

参考文献

- [1] 片山, 高須, 安達, 「データベースにおける意味的情報の構造的表現」, 情報処理学会 データベースシステム研究会, 93-DBS-94, 94-8.
- [2] 片山, 高須, 安達, 「全文データベースにおける巡回的検索および宣言的検索の実現法」, 1993 年 電子情報通信学会秋季大会 D-52.
- [3] 有澤 博 (編), 特集「演繹データベース」, 情報処理, 31, 2, (1990).
- [4] 勝野 裕文, 「演繹データベースの形式的意味論」, in [3], (1990), 198-205.
- [5] Ullman, J., *Principles of Database and Knowledge-base Systems*, Computer Science Press, Vol. I/II (1988).