

分散環境における包摂関係の統合と問い合わせ処理

7K-4

黒田 崇[†] 萬上 裕[‡] 横田 一正[†] 上林 彌彦[†]京都大学工学研究科[†] 京都大学工学部[‡]

1 はじめに

インターネットをはじめとする広域ネットワークの爆発的な広がりによって、広域ネットワーク上に分散する多種多様な情報源を融合し有効に利用するための研究がなされてきた。ネットワークを介して利用可能な分散環境上で独立に管理される情報源が持つ情報を統合して利用することができれば、目的に応じたより多くの情報を入手することが可能となる。しかし、分散環境下の情報源の統合には異種性や情報源間の矛盾などの問題が存在する。そこで本稿ではメディアエータシステムを用いて分散環境上の情報源が持つ情報を統合する際に、新たに情報源/メディアエータ記述のために演繹オブジェクト指向データベース言語 *Quizote* を拡張した知識表現言語 *QUIK* (*Quizote in Kyoto*) を提案し、更に *QUIK* を用いて包摂関係を統合すること及びその利用について考察する。

2 メディアエータシステム

データベースやファイルシステム、HTML 文書など分散した異種情報源が持つ情報を統合するひとつの手法としてメディアエータシステムがある。メディアエータシステムは情報源、ラッパ、メディアエータから構成される。情報源が持つ情報についての共通モデルを用いてラッパ内で表す。ラッパの出力する共通モデルを用いて、メディアエータに記された規則により分散情報源に対する統合ビューを与える。

メディアエータシステムを用いた分散情報統合のためのシステムに *TSIMMIS*^[1] がある。本稿では、再帰関係を含む一般的規則と共通モデルをメディアエータで扱う、更に情報源の包摂関係をラッパの記述で考慮するなど *TSIMMIS* システムを拡張することを考えている。

3 情報源 / メディアエータ記述言語 *QUIK*

WWW 上のファイルのような一定のスキーマを持たない情報源も分散環境上の情報源として統合対象とする場合、共通モデルには高い柔軟性が要求される。例えば、ファイル上の情報は表の形式で記述されていても、データによっては欄の数が増えたり追加されたりする。共通モデルにはこのようなスキーマを持たない半構造化された情

報を表現できる必要がある。また、高度な知識表現を可能にするには、単に複合オブジェクトを表現できるだけでなく、規則や包摂関係を表現できた方がよい。このような要件を考慮した言語 *QUIK*^[3] について考える。

3.1 *QUIK* プログラム

QUIK プログラム P は 4 つの構成要素からなる。構成要素はそれぞれデータディクショナリ / ディレクトリ (DD/D) D 、情報源関係の集合 I 、包摂関係の集合 S 、規則の集合 R からなっている。つまり、

$$P = (D, I, S, R)$$

の四つ組で表される。それぞれの情報源とメディアエータに対し、ひとつの *QUIK* プログラムが定義されるので、メディアエータは図1のように階層構造となりうる。この階層構造は情報源の探索範囲に対応している。

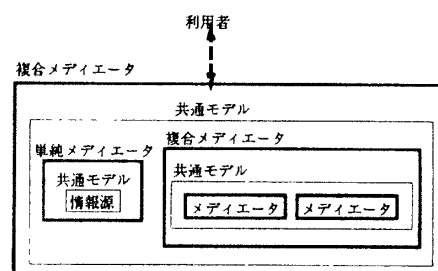


図 1: メディアエータの階層関係

DD/D は情報源の一種のメタデータであり、論理的には下位のメディアエータを含むひとつのメディアエータ内で定義されているデータを保持している。 DD/D D は情報源識別子、オブジェクト識別子、属性名、オブジェクトの同値関係から以下のように構成されている。

$$D = (ID, \{ID_i, \{O_{ij}, \{L_{ijk}\}\}\}, \{E_l\})$$

ID はメディアエータに関する自己識別子、 ID_i は自己を含めた所在情報を含む情報源識別子、 O_{ij} は情報源 ID_i に含まれるオブジェクト識別子、 L_{ijk} はオブジェクトが持つ属性名である。 E_l はオブジェクト識別子の同値関係を表しており、束の自動生成あるいはオブジェクト同定の結果から得られる同値関係を格納している。ここで、 D はオブジェクト識別子、属性名からも参照可能である。

情報源関係は、ある情報源がどの情報源の規則を参照するか、または継承するかを指定する。つまり I には二つ

Global Subsumption Relation in Distributed Environments and its Query Processing
Takashi KURODA, Yutaka BANJOU, Kazumasa YOKOTA and Yahiko KAMBAYASHI
Faculty of Engineering, Kyoto University

の関係、参照関係と継承関係がある。一般的にルール型言語では単なる外部参照か、継承かで意味が異なる。よってここでは2種類の情報源関係を導入した。以下にその例を示す。

【例】 s1: $p(X) \leftarrow r(X, Y), p(Y)$
 $p(a)$
 $r(c, a)$
 $r(d, b)$
 s2: $p(b)$

という規則が個々の情報源にあると仮定する。そのとき $?-P(X)$ という質問をした場合を考える。規則を統合せずに質問を評価した場合には解は $s1: X = a, c, s2: X = b$ となる。一方 $s1, s2$ の規則を統合した後に解を求めると $X = a, b, c, d$ となり結果は異なる。これは p の終了条件が規則の統合により増えたことによる。規則を継承して統合する場合、規則の中で指定される情報源識別子は無視される。このように前者のような質問評価の方法を参照関係、後者の場合を継承関係という。規則での情報源の外部参照は、ここで規定する参照関係に従っていなければならない。これは包摂関係の一貫性を保つためである。

S は D の同値関係を反映した概念の束である。すなわちこの情報源を形づくる構成要素であり、これらの要素の間には包摂関係 " \sqsubseteq " がつけられる。 $a \sqsubseteq b$ は a が b より詳細化されていることを表し、直感的には "is-a 関係" に相当する概念間の半順序関係である。半順序関係から束への変換手続きはよく知られている [2] ので S は一般性を失うことなく束であると仮定できる。

3.2 包摂関係の統合と利用

分散情報源を統合する場合、個々の情報源が持つ包摂関係から新たな統合包摂関係を生成し、メディアータはその結果を用いて個々の情報源が持つ情報間に制約を与えることがある。

メディアータ P がメディアータ P_1, P_2, \dots, P_n が持つ情報を統合しているとき、それぞれの包摂関係を L, L_1, L_2, \dots, L_n とすると、 P における統合包摂関係は $L \cup L_1 \cup L_2 \cup \dots \cup L_n$ から得られる。

一般的に束 L_1 と L_2 の合併 $L_1 \cup L_2$ は必ずしも束とはならない。この束の生成アルゴリズムはよく知られているが、実際には $\top \cong \perp$ が得られる危険性がある。しかし QUIK では、束の合併はメディアータ作成時の支援環境と位置づけ、メディアータ作成者との対話的構築によって合併が可能となることとする。つまり、自動統合の結果得られた束と同値関係を利用してメディアータ作成時に選択的な束の統合機能を持たせる。この束の合併の結果生じる同値関係が新たに DD/D に貯えられる。束 L_1 と L_2 から生成できる束を $\overline{L_1 \cup L_2}$ と書くことにする。 $I = (D, I, L, R), I_i = (D_i, I_i, L_i, R_i) (1 \leq i \leq n)$ (簡略化のために包摂関係定義とそれから生成される束を同一視

している)、そして

$$L' = \overline{L \cup L_1 \cup L_2 \cup \dots \cup L_n}$$

のとき、 L' を P, P_1, P_2, \dots, P_n に共通の包摂関係として定義し、各情報源を以下のように考える。

$$P = (D, I, L', R), P_i = (D_i, I_i, L', R_i) (1 \leq i \leq n)$$

このことにより、関係するすべての情報源の包摂関係は、統合を行うメディアータにより定義されることとなり、そのメディアータ内で一貫性をもった大域的な包摂関係がつけられたことになる (図2参照)。つまりメディアータで生成されたこの大域的包摂関係は関係する情報源に輸出され、各情報源での制約解消や属性継承は、対応するメディアータから呼び出されたときにはこの制約解消に従うことになる。大域的包摂関係は元の情報源に対する拡張されたビューと考えることもでき、問い合わせに対する拡張となっている。

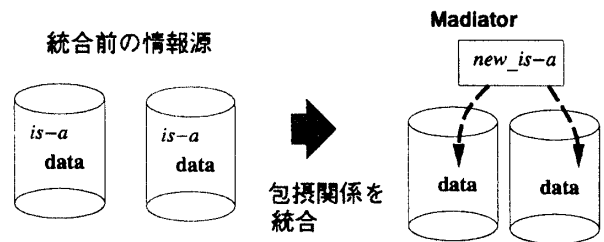


図 2: 包摂関係の統合と利用

4 おわりに

本稿では演繹オブジェクト指向データベース言語 *Quixote* を分散環境に対応させた知識表現言語 *QUIK* を提案し、*QUIK* のメディアータアーキテクチャへの利用を考えた。さらに包摂関係の統合について述べ、その利用について議論した。

今後の課題として、*QUIK* 処理系の本格的実装とそれに基づいた適用可能性の評価があげられる。また、包摂関係以外の制約に関する扱いについても議論する必要がある。応用として地理情報システムへの適用を考えている。

参考文献

- [1] Y.Papakonstantinou, S.Abiteboul and H.Garcia-Molina. "Object Fusion in Mediator System", *Proc. VLDB conf.*,1996.
- [2] Hassan Ait-Kaci, Robert Boyer, Patrik Lincoln and Roger Nasr, "The Efficient Implementation of Object Inheritance", *MCC Tech. Report AI-102-87*, 1987.
- [3] Kazumasa Yokota, "メディアータシステムからエージェントシステムへ", 科研重点領域「高度データベース」成果報告会, Jan, 1997.