

仮説生成を用いた分散情報源の統合

7K-3

横田 一正[†] 萬上 裕[‡] 黒田 崇[†] 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

ネットワーク上のデータベースは非常な勢いで増加しつつあり、分散環境における情報源の統合の研究が盛んに行われている。その一手法として、単にデータベースだけでなく、ファイルシステム、WWWページなど様々な分散異種情報源に対し、統合ビューを定義することによって統一アクセスを与えるメディエータシステムがある。

我々は統合ビュー定義のために必要なデータモデルとして *QUIXOTE*^[1] を拡張した新たな知識表現言語 *QUIK* を提案し、高度な知識表現の扱いを可能にする。さらにメディエータに仮説生成機能を持たせ、それとデータディクショナリ/ディレクトリを用いることによって、階層化されたメディエータに対する問合せが失敗したときあらかじめ統合規則で指定された以外の情報源へ問合せ、新たな回答を得る手法を提案する。

2 メディエータシステム

メディエータシステムは図1のように構成される。図1における情報源として、RDBやOODB、KB、SGMLやHTMLなどで記述された半構造化情報源であるファイルシステムなどをとる。ラッパーは各情報源のオブジェクトを共通モデルに変換し、メディエータがラッパーから出力される共通モデルオブジェクトを規則により融合し統合ビューを定義する。QUIKでは以下の点に特徴がある。

単純メディエータ

メディエータの階層構造において、情報源の記述を統一して扱うため、ラッパーの付加された情報源も単純メディエータとして扱う（図1を参照）。

複数ラッパーの定義

一つの情報源を異なる目的で利用するため、ラッパーを二つ以上定義できる（図2を参照）。この情報源の多目的利用はメディエータレベルでも可能であるが、情報源のセキュリティの問題から、ラッパーでも行えた方がよい。

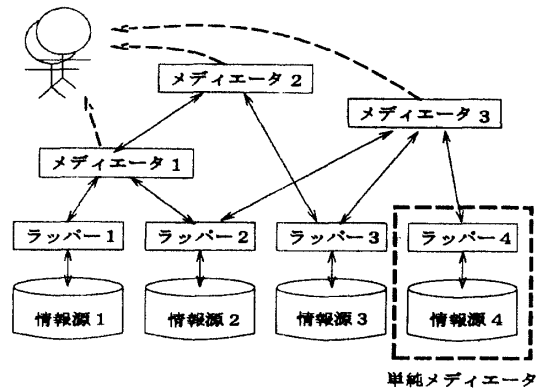


図1: メディエータ アーキテクチャ

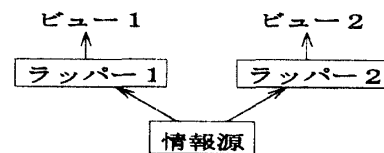


図2: 複数ラッパー

3 分散環境での問合せ処理の要件

分散環境の統合に必要な要素を挙げ、説明する。

3.1 メディエータの階層化

メディエータは利用者の目的に応じて構築される。新たなメディエータ構築時に、すでに存在するメディエータを利用可能な時はそれを利用し、階層化させることが有効である。

3.2 データディクショナリ/ディレクトリ

データディクショナリ/ディレクトリ（以後DD/Dと記述する）は一種の情報源のメタデータであり、一つのメディエータ（論理的には下位のメディエータを含む）内で定義されている情報源識別子、オブジェクト識別子、その属性名を保存する。DD/Dは単純メディエータではラッパーの構築時に作られ、メディエータ作成時にはそれが参照する下位のメディエータから情報を吸い上げる。これにより、メディエータは自分より下位の全てのデータを参照可能となる。図3はメディエータM3が参照可能な範囲（これをスコープと呼ぶ）を表している。図3においてM3のDD/DはM1、M2、M3で定義されているデータを保持している。また、DD/Dはオブジェクト識別子、属性名からも検索可能であるとする。

Integration of Distributed Information Sources Using Hypothesis Generation

Kazumasa YOKOTA[†], Yutaka BANJOU[‡],
 Takasi KURODA[†], Yahiko KAMBAYASHI[†]

Department of Information Science, Kyoto Univ.[†],
 Faculty of Engineering, Kyoto Univ.[‡]

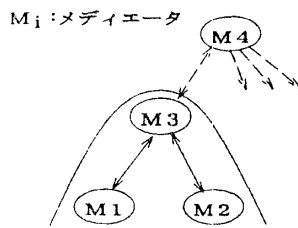


図3: メディエータのスコープ

3.3 メディエータでの仮説生成機能

仮説生成は部分情報を持ったデータベースでの問合せ処理で有効である^[1]。これを分散環境にまで拡張したとき、問合せに回答するために必要な情報を探索するのに利用できる。メディエータのスコープ内に存在するオブジェクトとその属性をDD/Dに保存することによって、生成された仮説に対し有効な情報を探索する。スコープ内で解が得られないときは、より上位のメディエータに対し仮説による代替問合せを行う。

【例1】 図3において各メディエータのデータが

M3: $p \leftarrow M1:q, M2:r$.
 M1: $q; r$.
 M2: $r \leftarrow s$.

のとき、メディエータ M3 に対する問合せ $?-p$ は仮説生成機能がない場合は失敗するが、DD/D により r が M1 にあることを知り、 $M1:r$ の代替サブゴールにより、解 YES が得られる。□

【例2】 図3において

M3: $p/[l = X] \leftarrow M1:q/[l = X], M2:r/[l = X]$.
 M1: $q/[l = a]$.
 M2: r .

のとき、メディエータ M3 に対する問合せ $?-p/[l = a]$ は仮説付き解 IF $M2:r.l = a$ THEN yes を生成する。仮説から代替問合せ $?-r/[l = a]$ を生成し、上位のメディエータ M4 に対して問合せを行う。□

4 QUIK での導出

仮説生成の概念と問合せ処理時の導出過程について説明する。仮説生成は QUIKOTE では導出列を

$$(G_0, A_0, \emptyset) \Rightarrow (G_1, A_1, C_1) \Rightarrow \dots \Rightarrow (\emptyset, A_n, C_n)$$

の3つ組のノードの系列として考えた^[1]。ここで、 G_i は未解消オブジェクト項の集合、 A_i は未解消制約の集合、 C_i は解消済制約の集合である。オブジェクトのゴールが充足されさえすれば、充足されない制約 A_n のみを仮説として生成し、解のひとつとして (\emptyset, A_n, C_n) を生成してき

た。これはオブジェクトの存在は仮説としては生成しないが、その属性は仮説とすることに対応していた。

いま、問合せ $?-G$ に対する失敗した導出列の最終ノードが $N_i = (G_i, A_i, C_i)$ であるとする。QUIK における導出には、最終ノード (G_n, A_n, C_n) の種類に応じて以下の種類がある。

i) $(\emptyset, \emptyset, C_n)$: 成功

ii) (\emptyset, A_n, C_n) :

(1) DD/D を用い A_n を他の情報源で探索

(2) 結果によって i か ii を繰り返す

(3) A_n が解消できなければ仮説付き解として成功

iii) (G_n, A_n, C_n) :

(1) DD/D を用い G_n を他の情報源で探索

(2) $G_n = \emptyset$ となれば i か ii の処理

(3) G_n が解消できなければ $(G_{n-1}, A_{n-1}, C_{n-1})$ に対して iii を繰り返す

(4) $n = 0$ となれば失敗

ただし iii) の効率が問題となる。つまり、仮説付き解の生成が爆発する危険性がある。したがって実際には何らかの制限を設ける必要があるだろう。

5 おわりに

本稿では知識表現言語 QUIK での仮説生成機能とそれを用いた情報源探索機構について検討した。これにより、動的に変化する統合環境において、部分情報を持つ情報源を複数個利用してより柔軟に情報を得ることが可能になると考えられる。

今後の課題として、導出アルゴリズムの最適化の検討、分散した等価な情報が非同期に更新されたときの一貫性の保持の問題、QUIK メディエータシステムの実装、そして運用、評価がある。応用として、分散地理情報システムへの適用を考えている。

謝辞

種々の御助言および御協力を頂きました上林研究室の皆様へ感謝致します。なお、本研究は文部省科学研究費(基盤研究(A)(2)試験)によるものである。

参考文献

- [1] K.Yokota, T.Nishioka, H.Tsuda, S.Tojo, "Query Processing for Partial Information Databases in QUIKOTE", 6th IEEE International Conference on TAI, pp.359-365, Nov., 1994.
- [2] 横田一正, "メディエータシステムからエージェントシステムへ-エージェント型データベースの研究-", 科研重点領域研究「高度データベース」成果報告会, 1997.