

6D-7

コレクションオブジェクトを支援する
オブジェクトマネージャの問合せ処理

高橋 賢一 石川 佳治 植村 俊亮

奈良先端科学技術大学院大学 情報科学研究科

1 はじめに

オブジェクト指向データベース (OODB) などの複合オブジェクトを対象としたデータベースでは、集合、マルチセット、リストなどのさまざまなコレクションオブジェクトを統一的に、かつ効率的に扱いたいという要求がある。近年、データベースプログラミング言語の研究分野において、このような要求をもとにコレクション型 (collection type)、あるいはバルクデータ型 (bulk data type) などと呼ばれる、コレクションオブジェクトのための言語の構成手法に関する議論が行われている。

このようなコレクション型の枠組みを OODB などについてこれまで研究されてきた問合せ処理方式と統合し、問合せの最適化の形式化や実行処理をの効率化を行なおうとする研究も進められている^{3), 7)}。

一方、石川らはこれまで、テキスト検索の分野で一般的な索引手法であるシグネチャファイル (signature file)²⁾ の手法を、包含関係 (\supseteq) などによる集合データの検索に利用することを提案し、シグネチャファイルの物理的構成法や性能の評価などを行ってきた⁴⁾。また、集合値シグネチャファイルの概念を一般化し、コレクションオブジェクトに対する索引として活用することを、5) において提案した。5) では、Fegarar が提案したコレクション型の枠組みに基づいて、集合シグネチャファイルのファイル構造や、集合シグネチャファイルを用いた問合せ処理の記述を試み、コレクションオブジェクトの問合せ処理に対する一般化について検討した。

本研究では、5) の研究をもとに、コレクションオブジェクトに関する問合せ処理を効率的に支援するオブジェクトマネージャの設計と実装を行う。本稿では特に、オブジェクトマネージャにおける問合せ処理の部分に焦点を当て、主に問合せ処理の流れについて述べる。

2 オブジェクトマネージャの概要

本研究で対象とするオブジェクトマネージャは索引機構の一つとして、集合シグネチャファイル (set-based signature file)⁴⁾ を実装していることを特徴の一つとし、集合の包含関係 (\supseteq) などの集合値検索条件⁴⁾ を効率的に支援する。

また、集合以外のコレクション型 (マルチセット、リストなど) に関する問合せについても、集合シグネチャ

ファイルを用いることで効率的な問合せ処理が可能なのが存在することから⁵⁾、そのような問合せに対し積極的に集合シグネチャファイルを活用することを試みる。

問合せ処理を表現するための枠組みとして、Fegararらにより提案されたモノイドに基づくコレクション型の枠組み³⁾ を利用する。Fegararらは、数学的なモノイドの概念に基づいて、さまざまなコレクション型を統一的に扱い、分割統治処理に適したコレクション型の構成手法を提案した。この枠組みでは、上位のモジュールから与えられたモノイド論理 (monoid calculus) の問合せが段階的に変形され、最終的にモノイド代数 (monoid algebra) の canonical form に帰着され、これが問合せ実行プラン (QEP) となる。彼らは、これらの言語を基にしてオブジェクト指向データベースの問合せ処理のフォーマルな記述を行なった³⁾。

3 問合せ処理の概要

集合値検索の例として、ODMG-93 Standard¹⁾ の OQL 言語に基づく、以下の問合せを考える：

```
select distinct d.name
from d in Department
where d.orders is_superset? set("a", "b")
```

この問合せは、商品 "a"、"b" の両方とも注文した部門の名前を求める問合せであり、集合値検索の一種である。このようなユーザレベルの問合せはオブジェクトマネージャの上位のモジュールにおいて変換され、

```
set{d.name | d ← D, d.orders ⊇ Q}
```

というモノイド論理の論理式としてオブジェクトマネージャに渡される。ただし、D はクラス Department、Q は集合 set("a", "b") を表すものとする。

このようなモノイド論理式に基づく問合せを受けると、オブジェクトマネージャはこの問合せが集合値検索や結合処理など、その処理のために特定のアルゴリズムが存在するような問合せであるかどうかを調べる。このためにパターンマッチが行われ、この例の場合には集合値検索であると判定される。

次いで、Department クラスの orders 集合値属性に集合シグネチャファイル S が存在するかどうか調べられ、存在するならば集合値索引を利用した問合せとして、集合シグネチャファイルを明示的に用いたモノイド論理の等価な問合せ

```

set{d.name |
  d ← FDR(ssig-scan(S, Q, has-subset),
           orders, Q, has-subset)}

```

に変形される⁵⁾。ただし、 $ssig\text{-}scan(S, Q, qtype)$ は集合シグネチャファイル S を集合 Q により走査し、問合せ条件を満たす可能性のある Department オブジェクトの OID の集合を返す処理を表す。 $qtype$ は集合値検索の種類を表す引数である。

$FDR(I, A, Q, qtype)$ は、フォルスドロップレゾリューション (false drop resolution) の処理を表している。シグネチャファイルではそのファイル構成上、与えられた問合せに対して得られた結果の中に検索条件を満たさないもの (フォルスドロップ) が存在する可能性がある (これはシグネチャ生成でハッシュ法などを用いるためである)。よって、シグネチャファイルの走査で得られた OID の集合を基に実際にオブジェクトを検索して条件を満たすかどうか検証する操作が必要となる。これがフォルスドロップレゾリューションであり、OID の集合 I をもとに、各オブジェクトを取りだし、そのオブジェクトの集合値属性 A が、与えられた集合 Q と実際に $qtype$ の検索条件を満たすかどうかチェックされる。

このモノイド論理式の間合せは、モノイド代数に直接的に変換され、変換ルールを用いて canonical form に変形される。最終的に得られる問合せプラン (QEP) は以下ようになる。(変換の詳細については、5) を参照のこと) :

```

hom[bag, set](hom[set, set](λd.{d.name})o
  (hom[set, set](λo.set{o | o.orders ⊇ Q})o
    (hom[set, set](λi.{ref(i)})o
      (λs.if has-subset(s.π1,
        make_ssig(Q, F, m, h))
        then {s.π2}
        else {})))) S

```

この QEP ではシグネチャファイル S にモノイド代数のオペレータ hom を階層的に適用している。

以上の問合せの流れを示したものが図 1 である。この図において、optimizer は入力された論理表現を上記述べたようなステップで QEP へと変換する。また、catalog manager はデータベース内の索引の有無や統計的情報を optimizer へ伝える。実行エンジンは QEP をデータベースに対して実行する機能を持つ。

4 まとめと今後の課題

本稿ではコレクション型の検索を支援するオブジェクトマネージャの間合せ処理の枠組みについて述べた。マルチセットなどの他のコレクション型に対する適用の例は 5) で述べられている。

本稿で述べた問合せ処理の最適化の為に Volcano Optimizer Generator⁶⁾ を実装に用いた問合せオブティマイザを設計中である。また、このオブジェクトマネージャの要求に適したシグネチャファイルの構成手法の検討も行なっていく。

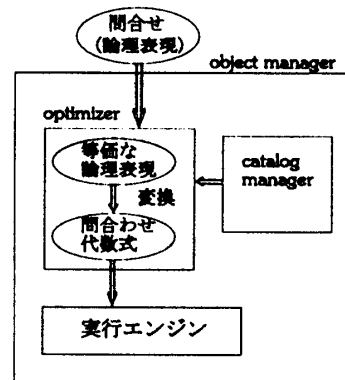


図 1: オブジェクトマネージャにおける問合せ処理

謝辞

本研究を遂行するにあたり貴重な御意見を頂いた筑波大学北川博之助教授、ならびに奈良先端科学技術大学院大学マルチメディア統合システム講座の皆様感謝いたします。

参考文献

- [1] R. Cattell ed.: *The Object Database Standard: ODMG-93*, Morgan Kaufmann, San Francisco, California, 1994, Release 1.1.
- [2] C. Faloutsos: "Signature Files", in W. B. Frakes and R. Baeza-Yates eds., *Information Retrieval - Data Structures and Algorithms*, pp. 44-65, Prentice-Hall, Englewood Cliffs, NJ, 1992.
- [3] L. Fegaras and D. Maier: "Towards an Effective Calculus for Object Query Languages", in *Proc. of ACM SIGMOD*, San Jose, California, May 1995.
- [4] 石川, 北川: "シグネチャファイルによる集合値検索のコスト評価", 情報処理学会論文誌, 36(2):385-395, 1995年2月.
- [5] 石川, 北川: "集合シグネチャによるコレクションオブジェクトの間合せ処理", 情報処理学会データベースシステム研究会研究報告, 95(65), 1995年7月 (95-DBS-104-2).
- [6] G. Graefe and W. J. McKenna: "The Volcano Optimizer Generator: Extensibility and Efficient Search", in *Proc. of IEEE 9th International Conference on Data Engineering*, pp. 209-218, Vienna, Austria, April 1993.
- [7] B. Vance: "An Abstract OODB Query Execution Language", in *Proc. of 4th Int'l. Workshop on Database Programming Languages*, pp.176-199, New York City, 1993