

関数型データベース管理システムの実現手法

4R-11

家富誠敏 舛谷和幸 富井尚志 有澤博

横浜国立大学 電子情報工学科

1 はじめに

マルチメディアデータベースでは数値・文字列、映像や音声、時には三次元映像などといった、様々なメディアのデータを扱うことが考えられる。

これを実現するものとして、現在、著者らが行なっているリアルワールドデータベース [1] の研究は、多様な見方や自由度の高い意味的つながりを表現できるという観点から、単位となる情報が小さく、かつ平坦な構造をもつ AIS モデルと呼ばれる関数型データモデルに基づいて行なわれている。

しかし現在に至るまで、AIS モデルを扱うデータベースシステムというものは存在せず、新しい専用のデータベースシステムの開発が望まれていた。

そこで今回、AIS モデルに基づくデータベースシステムとして AIS データベースシステムのプロトタイプを構築した。本稿では、AIS データベースシステムの概要について述べる。

2 AIS データベースシステム

AIS データベースシステムは AIS モデルで表現されるデータの管理を目的として考えられたデータベースシステムである。

このシステムには以下のような機能が要求される。

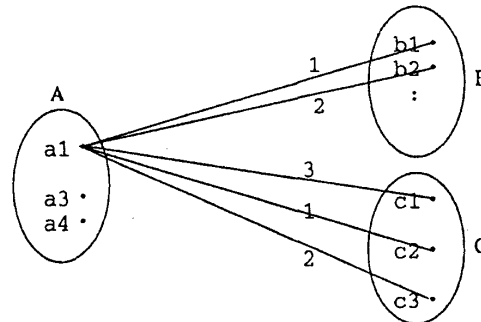
2.1 データベース関数の実現

AIS モデル [2] とは関数型データモデルの一種であり、現実世界を表現するのに、主体 (entity) と関数 (function) という概念を用いている。

主体は現実世界の「もの」を表し、数字や文字等の値 (value) も一つの主体である。共通な性質を持つ主体の集合は、主体型と呼ばれる。

関数は主体と主体の間の意味的なつながりを表し、定義域及び値域となる主体型ごとに区別され、名前

(関数名) が付けられている。関数名としては通常、定義域と値域の主体型名を並べたものを使用することが多い。



例) $AB(a1) = \langle b1, b2, b3, \dots \rangle$
 $AD(a1) = \langle d2, d3, d1 \rangle$

図 1: 主体と関数

主体に対して関数を適用すると、その主体から関数によって結び付いた主体列を得ることができる。

データベースへの検索は、関数を用いて主体同士の結びつきを得ることにより行なわれる。データベースの検索に用いられるこのような関数のことをデータベース関数 [3] と呼ぶ。

AIS データベースシステムではデータベース関数の実現を考慮した設計が必要とされる。

2.2 順序の管理

例えば動画画像データ等の扱いを考えると、これを構成する一枚一枚のフレームに順序があると再生等の扱いが便利である。このような時間軸を持つデータ等、現実世界の事象は順序を持つものが多く、これらの表現のために AIS モデルではデータベース関数の返り値である主体集合に順序が振られている。

それ故、順序についての情報を管理し、蓄積、検索、更新等の操作を効率良く行なうことがデータベースシステムに要求される。

3 データ構造

前節で述べたような立場に基づいて設計したのが AIS データベースシステム (プロトタイプ) である。

主体間を結ぶ関数を実現するために、AIS データベースでは以下のようなデータ構造を用いた。前項図 1 の関係を実際のデータ構造で表したのが図 2 である。

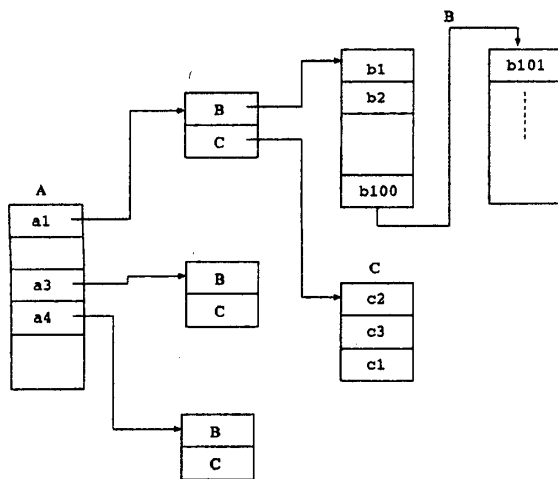


図 2: AIS データベースのデータ構造

- 関数を表現するには、引数を索引 (インデックス) として、対応する値を取り出せるような機構が必要である。ここでは物理データ配置による索引づけを行なった。すなわち領域を固定長に区切って、 n 番目に $a(n)$ というデータを配置することにより、任意のデータに瞬時にアクセスできる。
- データ構造において、ある引数に対応する全ての関数を一ヶ所に集めることで、任意の引数に対して複数の関数を適応する場合の最適化を図った。例えば $a1$ に対して、それにつながる B と C の主体列は同時に取り出すことが容易である。
- データベース関数の戻り値は順序を持つ主体列だが、この順序列への挿入・削除があった場合でも影響が全体に広がらないように、長大な値列は複数の小ブロックに分割して格納した。図 2 でいえば、 $b1$ と $b2$ の間に値を挿入する場

合、 $b1$ から $b100$ が入っている小ブロック内だけで値をずらせばよい。この場合にはブロック内が完全に埋まっていて新しい値が入らないので、 $b1$ から $b49$ 、 $b50$ から $b100$ 等、二つのブロックに分割する操作が必要となる。

このような構造を採用することにより、ある引数に対する任意の関数の戻り値は要素数によらず一定の計算量で求めることができる。

また引数ごとに独立した構造をしており、並列プロセッサを使えば更に高速なアクセスを行なうことが可能である。

問題点としては、物理データ配置を用いているため、削除等を行なった際、その要素の存在していた領域が空き領域となってしまうことが考えられるが、新しい要素を作る時、その領域を再利用することによって空間資源の無駄使いを極力防いでいる。

4 まとめ

AIS モデルに基づくデータベースシステムとして、AIS データベースシステムを設計、試作した。

これによって、AIS モデルで表現されたデータをそのまま格納することができ、かつ素早い検索および更新を可能とした。また将来に向けた並列化、最適化も念頭においた構成となっている。

参考文献

- [1] 有澤 博, “リアルワールド・データモデリングについての考察”, 電子情報通信学会技術研究報告, DE96-5, 1996.
- [2] H. Arisawa, T. Tomii, H. Yui, and H. Ishikawa, “Data Model and Architecture of Multimedia Database for Engineering Applications”, IEICE TRANS. INF. & SYST., vol E78-D, No.11, 1995.
- [3] 富井 尚志, 有澤 博, “マルチメディアデータベースにおける映像モデリングと操作言語”, 電子情報通信学会論文誌 Vol.J79-D-II No.4, 1996.