

# RIQSⅡV2 統計データベース機能

## — 多次元・多階層・要約データの管理 —

1E-4 小林敏彦<sup>1</sup>、小林次朗<sup>1</sup>、伊藤 進<sup>1</sup>、神辺武司<sup>1</sup>、 鴨井 功<sup>2</sup>

<sup>1</sup>日本電気ソフトウェア㈱

<sup>2</sup>日本電気㈱

### 1 はじめに

RIQSⅡV2 (Relational Information Query System II Version 2) は日本電気㈱の汎用コンピュータACOS上で動作するリレーショナルデータベース管理システムである。多次元、多階層、要約データを管理する統計データベース機能をリレーショナルデータベース上でやっている。統計データベース機能の概要と関連技術について報告する。

### 2 統計データベースの概要と適用分野

#### 2.1 要約データ

参考文献[1]によると、統計データベースとは「管理されるデータがサマリデータであるようなデータベース」として定義されている。ここでサマリデータとは「記述の対象が一般のDBのように個々の事物ではなく、それらを類別した集団についての要約したデータ」である。具体的には、東京に住む太郎や花子の収入を合計した東京の収入等がこれに該当する。

#### 2.2 多次元データ

データベースに格納されているデータは、その記述対象によって幾つかの軸の座標上に位置付けることができる。たとえば図1のようなデータは地域軸、

地域軸	時間軸	分類軸	統計軸	
国名	年	産業	就業者	収入
日本	1980	農業	100	50
日本	1980	工業	200	90
日本	1980	商業	300	150
中国	1980	農業	400	200
⋮	⋮	⋮	⋮	⋮
イタリ	1980	農業	200	300
⋮	⋮	⋮	⋮	⋮
日本	1985	農業	200	90
⋮	⋮	⋮	⋮	⋮

図1 多次元データの例

時間軸、分類軸の各座標に対応して(就業者、収入)の値が存在する。これらのデータをおある2つの軸について展開し、図2のように展開することができる。

農業・国別・年別・収入

年\国	日本	中国	イタリ	フランス	⋮
1980	100	400	200	300	⋮
1985	200	500	300	400	⋮
1990	300	600	400	500	⋮
⋮	⋮	⋮	⋮	⋮	⋮

図2 多次元データのマトリックス表現

#### 2.3 多階層データ

現実世界のデータは、日本の上にアジア、アジアの上に世界のように多階層の意味関係を構成していることが多い。これらの多階層意味関係を一つの軸について図3のように展開し、簡潔にマトリックスで表現することができる。

1990年・産業別・国別・就業者数

国\産業	農業	工業	商業
世界	55100	54200	33300
アジア	5100	4200	3300
日本	100	200	300
⋮	⋮	⋮	⋮
ヨーロッパ	4500	6400	5300
イタリ	500	400	300
⋮	⋮	⋮	⋮

図3 多階層データの表現

#### 2.4 統計データベースの適用分野

上記のような3つの性質を持っているデータを以後統計データと呼ぶことにする。統計データベースの適用範囲は、国や地方自治体で収集している社会の様々なデータを要約したもの、民間企業の経営上の基礎資料の要約結果などを対象とする。

RIQSⅡV2 Statistics Database Function  
— Multi-dimension, Multi-level, Summary Data Management —

Toshihiko KOBAYASHI<sup>1</sup>, Jiro KOBAYASHI<sup>1</sup>, Susumu ITO<sup>1</sup>, Takeshi KANBE<sup>1</sup>, Isao KAMOI<sup>2</sup>

<sup>1</sup>NEC Software, Ltd

<sup>2</sup>NEC Corporation

3 機能

R I Q S I I V 2 統計データベース機能の構成は図4のようになっている。統計データに関する知識定義、素データから統計データの生成、統計データの検索・分析までを一貫して支援する。

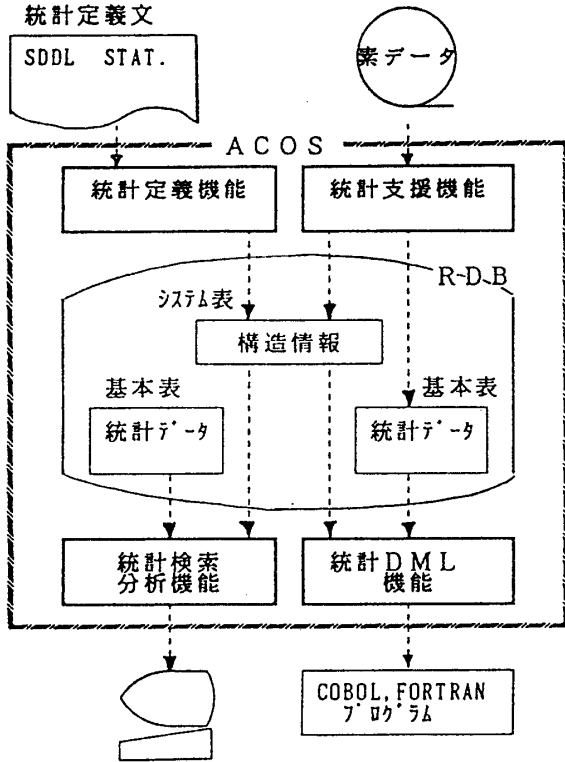


図4 統計データベース機能の構成

4 関連技術

4.1 統計データの構造の定義 (統計定義機能)

統計データはN次元空間の各座標に存在し、値を持つと考える。統計定義機能で記述された項目は、軸ごとに一意の番号が付られ、各項目の上下関係(図5)、各最下位項目に対する選択条件等が構造情報としてR I Q S I I V 2 のシステム表に格納される。

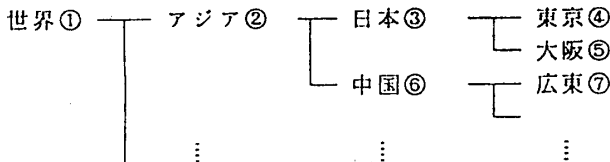


図5 地域軸の項目の相互関係と番号

4.2 統計データの生成 (統計支援機能)

統計定義機能で記述した統計データの構造をもとに、個々の事物を記述している素データから多次元・多階層・要約データを生成する。統計データの生成処理は図6のようになる。

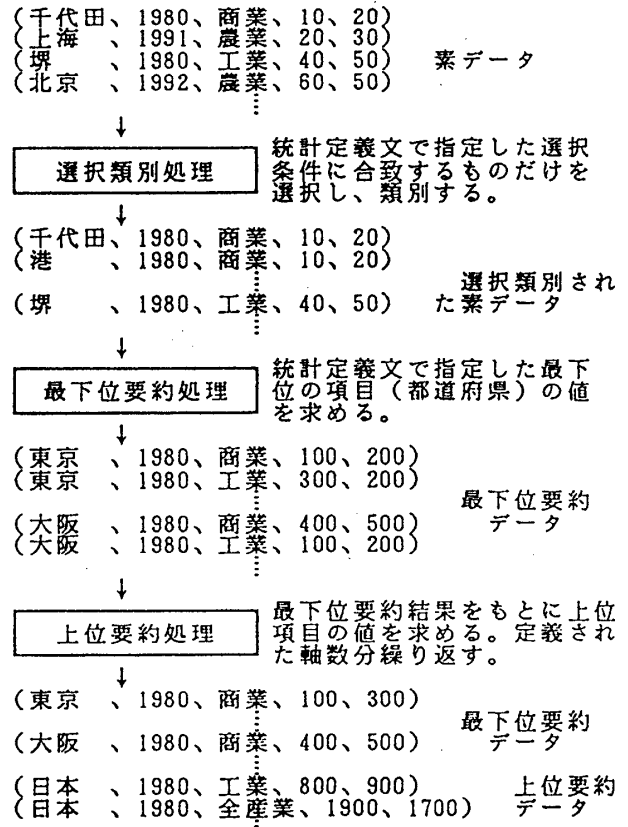


図6 統計データの生成処理概要

4.3 統計データの分析 (統計検索分析機能)

統計データには先頭にN次元空間上の座標を表すフィールド(固有値)が付加され、R I Q S I I V 2 の基本表に格納されている。N次元座標から固有値を生成する関数を用意しており、任意の座標の統計データを瞬時に取得できる。

f (東京, 1990, 工業) = 6 9 8 7 インデックス	固有値	就業者	収入
	6986	100	50
	6987	100	200

図7 統計データの格納形式

統計データの構造を生かした時系列分析、対照表示、タイトル変更などの豊富な分析機能が利用できる。分析結果は容易にグラフ化できる。

5 おわりに

多次元・多階層・要約データを管理するR I Q S I I V 2 統計データベース機能について概要と関連技術について報告した。

参考文献

[1] 「統計データベースの設計と開発」  
穂鷹良介・佐藤英人 オーム社