

リレーショナルモデル概説、関係代数

直積の部分集合を関係と呼ぶ。

A, B なる属性で定義される関係を $R(A, B)$ と書く。たとえば

$$R = \{(a_1, b_1), (a_2, b_1), (a_3, b_2)\} \quad (\text{この例では属性は } A, B \text{ の } 2 \text{ つしか想定してない})$$

一般的な関係 R は

$$R = \{(a_1, a_2, \dots, a_n) \mid a_1 \in A_1, a_2 \in A_2, \dots, a_n \in A_n, r(a_1, a_2, \dots, a_n)\}$$

と書ける。 $r(a_1, a_2, \dots, a_n)$ は述語といい、 a_1, a_2, \dots, a_n の組み合わせにより定まる条件である。 A_1, A_2, \dots を R の属性値集合、または定義域という。

関係モデルは表の形で表す（関係表）ことが多い。

列の中の項目の順番、行の順番等、他のモデルでは問題にするが、リレーショナルモデルでは一切問題にしない。

関係代数

和

$$R \cup S = \{x \mid x \in R \vee x \in S\}$$

在学生 ∪ 卒業生

積

$$R \times S = \{x \mid x \in R \wedge x \in S\}$$

在学生 × 3年生

差

$$R - S = \{x \mid x \in R \wedge x \notin S\}$$

注意 例に上げる、表形式の説明図では各欄に自由な文字を入れられるように感じられるが、データベースのデータは計算機が自動的に処理するのが目的のため、勝手なデータを入れる訳にはいかず、定められたものの中から選ぶのが基本と覚えておいてほしい。

学生表

学生番号	氏名	出身地
1011	新井 一	九州
1015	井上太郎	大阪
1123	宇野三太	名古屋
1238	大村花子	和歌山
1350	加藤雄一	神奈川

データの正規化

受講表

学生番号	授業科目番号	受講年度	成績
1011	201	1989	70
	203	1990	65
1015	202	1991	90
	203	1992	50
1123	202	1992	80
1238	203	1991	65
1350	201	1993	81

このような表は RDB では作らない。1つのマス目には1つだけ入れるようにする（第1正規形）。また、縦横整然としたマス目であることを要する（上記のような2つのマス目が統合したものは不可）。参考 正規化

成績表

学生番号	授業科目番号	受講年度	成績
1011	201	1989	70
1011	203	1990	65
1015	202	1991	90
1015	203	1992	50
1123	202	1992	80
1238	203	1991	65
1350	201	1993	81

授業科目表

授業科目番号	科目名	教官名
201	電子工学	夏目素石
202	文学	唯野耽二
203	英語	Peter Smith

ただし、ここでは科目名と教官名の組に対して一意となるように授業科目番号を振るものとしている。

射影

ある関係（表）から、いくつかの属性のみを取り出して関係を作り出すこと。この際重複するデータは1つだけにする。

学生番号	出身地	受講年度
1011	九州	1989
1015	大阪	1990
1123	名古屋	1991
1238	和歌山	1992
1350	神奈川	1993

学生表から学生番号と
出身地に着目した射影

受講表から受講年度に着目した射影

(射影を厳密に定義すると)

元の $R(A_1, A_2, \dots, A_n)$ から属性 $A_{P1}, A_{P2}, \dots, A_{Pk}$ だけを取り出し、その属性よりなる関係を作る。

$$R(A_{P1}, A_{P2}, \dots, A_{Pk}) = \{ (r[A_{P1}], r[A_{P2}], \dots, r[A_{Pk}]) \mid r \in R \}$$

$r[A_{P1}]$ はあるタプル r の属性 A_{P1} の値。

選択 (制限)

関係から、ある属性の値が定められた条件を満たすタプルのみを取り出すこと。

例 受講表から 成績 80

学生番号	授業科目 番号	受講年度	成績
1015	202	1991	90
1123	202	1992	80
1350	201	1993	81

厳密に表すと

$$R[A_P = C] = \{ r \mid r \in R \wedge r[A_P] = C \}$$

A_P はある属性値 $C = \{ =, \neq, <, >, \leq, \geq \}$ C は定数または $r[A_{Pq}]$

上の例では $A_P = \text{成績}$ 、 $C = 80$

受講表と授業科目表の直積

学 生 番 号	授 業 科 目 番 号	受 講 年 度	成 績	授 業 科 目 番 号	科 目 名	教 官 名
1011	201	1989	70	201	電子工学	夏目素石
1011	201	1989	70	202	文学	唯野耽二
1011	201	1989	70	203	英語	Peter Smith
1011	203	1990	65	201	電子工学	夏目素石
1011	203	1990	65	202	文学	唯野耽二
1011	203	1990	65	203	英語	Peter Smith
1015	202	1991	90	201	電子工学	夏目素石
1015	202	1991	90	202	文学	唯野耽二
1015	202	1991	90	203	英語	Peter Smith
1015	203	1992	50	201	電子工学	夏目素石
1015	203	1992	50	202	文学	唯野耽二
1015	203	1992	50	203	英語	Peter Smith
1123	202	1992	80	201	電子工学	夏目素石
1123	202	1992	80	202	文学	唯野耽二
1123	202	1992	80	203	英語	Peter Smith
1238	203	1991	65	201	電子工学	夏目素石
1238	203	1991	65	202	文学	唯野耽二
1238	203	1991	65	203	英語	Peter Smith
1350	201	1993	81	201	電子工学	夏目素石
1350	201	1993	81	202	文学	唯野耽二
1350	201	1993	81	203	英語	Peter Smith

自然結合 (Natural Join)

2つの関係の直積を作り、着目する属性値の一致するものだけを残し、属性の重複を避けるようにしたもの。

受講表と授業科目表を授業科目番号に着目して自然結合して できたもの

学生番号	授業科目番号	受講年度	成績	科目名	教官名
1011	201	1989	70	電子工学	夏目素石
1011	203	1990	65	英語	Peter Smith
1015	202	1991	90	文学	唯野耽二
1015	203	1992	50	英語	Peter Smith
1123	202	1992	80	文学	唯野耽二
1238	203	1991	65	英語	Peter Smith
1350	201	1993	81	電子工学	夏目素石

一般に関係 R の属性 A と関係 S の属性 B に関する結合は

$$R[A \bowtie B]S = \{(r, s) \mid r \in R \wedge s \in S \wedge (r[A] = s[B])\}$$

が等号= であるとき等結合という。この場合結果に同じ内容の属性 (A と B) が表れるので一方を省略することが多い。これを自然結合という。

関係データベースでは (ユーザに見えるように) 存在するのは表だけ、これに反し他の DB では RDB での演算に相当する情報もデータに埋め込まれている。

RDB において、元々の表に演算を施して、さまざまな表を作ることができる。こうして得られたものと、元々の、ものを区別するため、前者を実表 (base table)、後者をビュー (view) またはビュー表という。

データの正規化

帳票等のデータ形式をそのままデータベース化すると冗長性が多い、その冗長性を省いた構成にするのと等価なのが正規化である。正規化は：

- (1) データ構造を単純にする
- (2) 挿入や更新に対し、関係の再構成をできるだけ少なくするという効果がある。

第1正規形 リレーショナルデータベースでは、1つの欄に1つの情報しか入れられない。もし、それに相当するデータを作りたいときは、他の欄をすべて繰返して2つのタプルを作る。このようにして得られるものを第1正規形と呼ぶ。

従属 ある属性値が定まると別の属性値も定まるとき、後の属性は前の属性に従属する（関数的に従属する、機能従属する）という。

候補キー プライマリキー、またはその候補となるキー

第2正規形 ある関係が第1正規形の条件を満たしているとき、候補キー以外の属性（非主要属性）に候補キーに従属するものがあるとき、これを分離したものを第2正規形という。間接的に他の属性に従属しているとき推移従属という。

第3正規形 ある関係が第2正規形の条件を満たしているとき、すべての非主要属性が推移従属しないとき第3正規形という。

（第3正規形は第2正規形が候補キーについてだけの機能従属性に限定していたものを候補キー以外にも拡張したもの）

参考書

滝沢 誠 データベースシステム入門技術解説 ソフトリサーチセンター

増永良文 リレーショナルデータベースの基礎 オーム社

魚田勝臣、小碓輝雄 データベース 日科技連

（飯沢篤志、白田由香利 データベース面白講座、共立出版）

小碓暉雄：データベース入門、啓学出版、1700円