

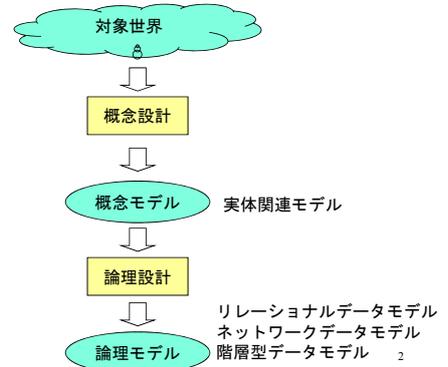
データベース概説(第3回)

2008.05.12

宇陀則彦

1

データモデリングの過程



データモデルの役割と代表的なモデル

- 概念モデル(実世界のモデル化ツール)
 - 実体関連モデル
- 論理モデル(データベースでのデータ表現)
 - リレーショナルデータモデル
 - ネットワークデータモデル
 - 階層データモデル

3

リレーショナルデータモデル

直観的には表 (テーブル)

支店

支店名	所在地	代表電話番号
みなと	横浜	045-123-4567
なにわ	大阪	06-555-6666
おえど	東京	03-3444-8888
はかた	福岡	092-333-7777

4

リレーショナルデータモデル

数学に基づく形式的定義

定義

- リレーションスキーマ $R(A_1, \dots, A_n)$ が与えられたとき、 $D_1 \times \dots \times D_n$ の有限部分集合 $r \subseteq D_1 \times \dots \times D_n$ をリレーションと呼ぶ。

5

集合

集合 S を

$$S = \{ a, b, c \}$$

としたとき、 a, b, c を集合 S の要素という。

(要素の重複 $S = \{ a, a, b, c \}$ は認めない!)

a, b, c が S に属することを

$$a \in S, b \in S, c \in S$$

と書く。

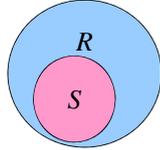
6

集合

集合 S が集合 R に含まれることを

$$S \subset R$$

と書く



7

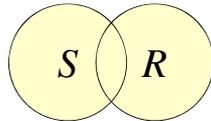
集合の演算

- 和集合演算 $S \cup R$
- 差集合演算 $S - R$
- 共通集合演算 $S \cap R$

8

和集合

$$S \cup R$$



$$S = \{a, b, c\}$$
$$R = \{b, c, d\}$$

$$S \cup R = \{a, b, c, d\}$$

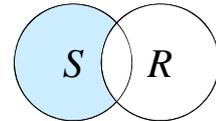
重複は除く

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

9

差集合

$$S - R$$



$$S = \{a, b, c\}$$
$$R = \{b, c, d\}$$

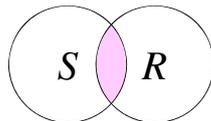
$$S - R = \{a\}$$

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

10

共通集合

$$S \cap R$$



$$S = \{a, b, c\}$$
$$R = \{b, c, d\}$$

$$S \cap R = \{b, c\}$$

$$S \cap R = \{t \mid t \in S \wedge t \in R\}$$

11

空集合

空集合 \varnothing : 何も入っていない集合

- $\{a, b\} \cap \{c, d\} = \varnothing$
- $\varnothing \subset A$ (A は任意の集合)
 \varnothing はどんな集合にも含まれると定義する。

12

組(順序ある組)

集合S、Rの要素を
(Sの要素、Rの要素)
という順で並べたもの

例

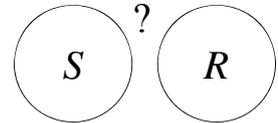
$S=\{0,1\}$, $R=\{a,b,c\}$
(0,a) (0,b) (1,c) など

13

直積

S, Rの要素から作られる「順序のある組」全ての集合

$S \times R$



$S=\{0,1\}$

$R=\{a,b,c\}$

$S \times R = \{(0,a), (0,b), (0,c), (1,a), (1,b), (1,c)\}$

•直積の要素を組(タプル)という

14

直積

練習

• $\{a, b, c\} \times \{1, 2\} \times \{A, B\}$ の直積を求めよ。

15

直積

練習

• $\{a, b, c\} \times \{1, 2\} \times \{A, B\}$ の直積を求めよ。

$\{(a, 1, A), (a, 1, B), (a, 2, A), (a, 2, B), (b, 1, A), (b, 1, B), (b, 2, A), (b, 2, B), (c, 1, A), (c, 1, B), (c, 2, A), (c, 2, B)\}$

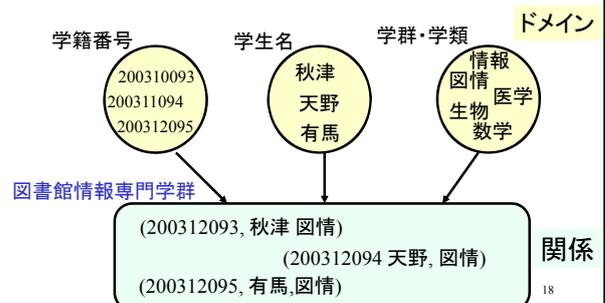
16

直積から関係(リレーション)へ

- ドメイン(定義域): 集合
 - $D_1=\{x \mid x \text{は学籍番号}\}$
 - $D_2=\{y \mid y \text{は学生名}\}$
 - $D_3=\{z \mid z \text{は学群・学類}\}$
- 直積: $D_1 \times D_2 \times D_3$
- 「図書館情報専門学群の学生」というリレーションは?
- 「弓道部」というリレーションは?

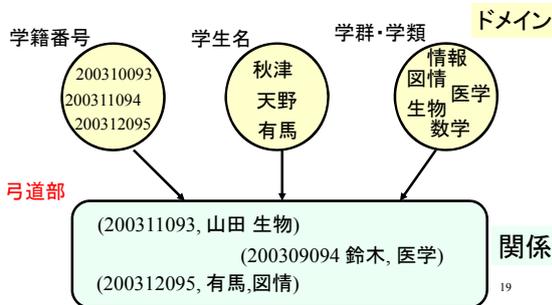
17

直積から関係(リレーション)へ



18

直積から関係(リレーション)へ



19

関係(リレーション)

図書館情報専門学群

(200312093, 秋津 図情)
(200312094 天野, 図情)
(200312095, 有馬, 図情)

リレーション名:
図書館情報専門学群

学籍番号	氏名	学群	属性名
200312093	秋津	図情	
200312094	天野	図情	
200312095	有馬	図情	

20

関係(リレーション)

弓道部

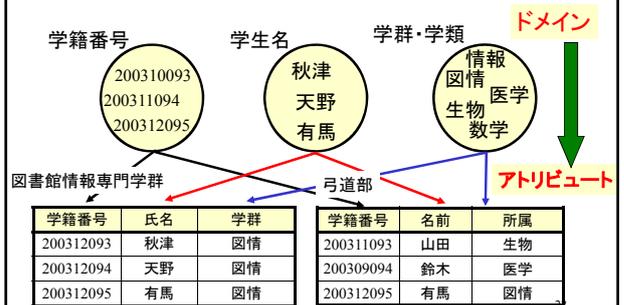
(200311093, 山田 生物)
(200309094 鈴木, 医学)
(200312095, 有馬, 図情)

リレーション名:
弓道部

学籍番号	名前	所属	属性名
200311093	山田	生物	
200309094	鈴木	医学	
200312095	有馬	図情	

21

関係(リレーション)



ドメイン、アトリビュート、スキーマ

- ドメイン(定義域)の集合の要素は、「シンプルな値」であること。
- アトリビュート(属性)の名前は、そのリレーション(関係)を適切に表現した(現実世界の意味を反映した)名前にしてよい。
- リレーションスキーマRは、アトリビュートの組 $(A_1, A_2, A_3, \dots, A_n)$ によって記述される。

23

リレーションアルデータモデル

数学に基づく形式的定義

定義

- リレーションスキーマ $R(A_1, \dots, A_n)$ が与えられたとき、 $D_1 \times \dots \times D_n$ の有限部分集合 $r \subseteq D_1 \times \dots \times D_n$ をリレーションと呼ぶ。

注: E-RモデルのRelationshipとリレーションアルデータモデルのRelationは異なる概念

24

リレーショナルデータモデル

直観的には表 (テーブル)

支店

支店名	所在地	電話番号
みなと	横浜	045-123-4567
なにわ	大阪	06-555-6666
おえど	東京	03-3444-8888
はかた	福岡	092-333-7777

25

リレーショナルデータモデル

直観的には表 (テーブル)

リレーション(relation)

支店

アトリビュート(attribute)

支店名	所在地	電話番号
みなと	横浜	045-123-4567
なにわ	大阪	06-555-6666
おえど	東京	03-3444-8888
はかた	福岡	092-333-7777

タプル
(tuple)

26

リレーショナルデータモデル

直観的には表 (テーブル)

テーブル(table)

支店

カラム(Column)

ロー
(row)

支店名	所在地	電話番号
みなと	横浜	045-123-4567
なにわ	大阪	06-555-6666
おえど	東京	03-3444-8888
はかた	福岡	092-333-7777

27

リレーションの次数と濃度

- カラム(column)の数を
次数 (degree)
- タプルの数を濃度
(cardinality)

支店名	所在地	電話番号
みなと	横浜	045-123-4567
なにわ	大阪	06-555-6666
おえど	東京	03-3444-8888
はかた	福岡	092-333-7777

濃度4

28

リレーションにおける注意点

- タプルの順序は問わない
 - 集合の要素に順序はない。
- タプルの重複禁止
 - 集合の要素に重複は許されない。
- 値はそれ以上分解できないこと
 - ドメインの直積であってはならない。
 - 姓と名の直積(山田, 太郎)はだめ
 - 値が集合(べき集合)であってはならない。
 - {弓道部, コンピュータクラブ}はだめ

29

例

1. **タプルの順序は無視**⇒下の2つの表は同じモノとみなす。

社員番号	社員名	所属支店
100250	夏目漱石	みなと
100583	正宗白鳥	おえど
101137	小宮豊隆	みなと

社員番号	社員名	所属支店
100583	正宗白鳥	おえど
100250	夏目漱石	みなと
101137	小宮豊隆	みなと

2. **同じタプル(全ての属性値が同じ)が複数あることは禁止**

右の表では2番目と
3番目が全く同じなので
禁止される。

社員番号	社員名	所属支店
100583	正宗白鳥	おえど
100250	夏目漱石	みなと
100250	夏目漱石	みなと

30

3. 属性の定義域の要素は「単純な値」であること。

ある支店が代表電話番号が二つ持っている、例えば

支店名	所在地	代表電話番号
みなと	横浜	045-123-4567, 045-123-4568
おえど	東京	03-3444-8888

とやりたくなるが、これは代表電話番号の属性値が二つの値の組になっているのでだめ。

31

リレーションにおける注意点

- リレーションスキーマとリレーションは異なる。
 - リレーションスキーマRは、時間経過に伴う値の変化(状態変化)とは独立した不変の枠組み
 - リレーションRは、リレーションスキーマの実現値であるインスタンス(オカレンス)の集合

例「支店(支店名、所在地、代表電話番号)をリレーションスキーマとするリレーション」を単に

リレーション 支店(支店名、所在地、代表電話番号)ということがある。(誤解を生じない限りにおいて)

32

メタデータの意味の違い(重要かも)

- データベースの世界では、スキーマ定義の記述情報を「メタデータ」という。
- 図書館情報学またはWebの世界における「メタデータ」は、データベースの世界からみれば、実体としての「データ(インスタンス、オカレンス)」である。
- メタデータデータベース？
メタメタデータ？

33

ここまでのまとめ

- E-Rモデルとリレーショナルデータモデルの位置づけの違いは？
- E-Rモデルでいう「関係」とリレーショナルデータモデルの「関係」の違いは？
- リレーショナルデータモデルの定義は？
 - 直積とは？
 - タプルとは？

34

ここまでのまとめ

- ドメインとアトリビュートの関係は？
- ドメインを定義するときの留意点は？
- アトリビュートを決めるときの留意点は？
- スキーマとインスタンスの違いは？
- 表形式の表現においては、
 - リレーションは何に相当するか？
 - タプルは何に相当するか？
 - アトリビュートは何に相当するか？

35

練習

- 2年次の標準履修科目というリレーションを設計しなさい。
 - ドメインとして何をどのように定義するか
 - アトリビュートをどのように定義するか
- (個人の)成績というリレーションを設計しなさい。
 - ドメインとして何をどのように定義するか
 - アトリビュートをどのように定義するか

数学的な定義を意識して書くこと

36

整合性制約(一貫性制約)

- 定義域制約(ドメイン制約)
 - タプルの値はドメインの値であること
(図書館情報専門学群の学生は筑波大学の学生)
 - ドメインの条件に従っていること
(学生の年齢が2歳とか500歳などでないこと)
(100点満点の成績は0から100までの値であること)

37

整合性制約(一貫性制約)

- キー制約
 - 候補キーを構成する属性値は唯一であること
 - 主キーを構成する属性値は唯一であり、かつ空値(null値)でないこと
- スーパーキー(super key)
- 候補キーまたは単にキー(candidate key)
- 主キー(primary key)

38

スーパーキー

- リレーションのタプルを一意に識別できる属性または属性の組合せをスーパーキーという。

39

リレーショナルデータモデル

支店

支店名	所在地	代表電話番号
みなと	横浜	045-123-4567
なにわ	大阪	06-555-6666
おえど	東京	03-3444-8888
はかた	福岡	092-333-7777

支店名はひとつしか存在しないものとする。 40

スーパーキー

- リレーションのタプルを一意に識別できる属性または属性の組合せをスーパーキーという。
 - (支店名、所在地、代表電話番号)
 - (支店名、所在地)
 - (支店名、代表電話番号)

41

候補キー

- スーパーキーのうち、極小(一意性を保つために除けないもの)の属性をいう。
 - (支店名、所在地、代表電話番号)
 - (支店名、所在地)
 - (支店名、代表電話番号)
- これらスーパーキーのうちで、
- (支店名)
 - (代表電話番号)
- が候補キーになりうる。(条件(状況)による)

42

リレーショナルデータモデル

支店

支店名	所在地	代表電話番号
みなと	横浜	045-123-4567
なにわ	大阪	06-555-6666
おえど	東京	03-3444-8888
はかた	福岡	092-333-7777

支店名はひとつしか存在しないものとする。

43

主キー

- 候補キーのうち、そのリレーションスキーマに最もふさわしい属性をいう。
 - (支店名)
 - (代表電話番号)
- これら候補キーのうち、
 - (支店名)が主キーとなる。
 - (空値となりえない。)
 - («支店»というリレーションに最適)

44

練習

- 学生(学籍番号、氏名、所属、住所)において、
 - スーパーキーはどれか？
 - 候補キーはどれか？
 - 主キーはどれにすればよいか？

45

整合性制約(一貫性制約)

- 参照整合性制約(参照一貫性制約)
 - あるリレーションの属性が別のリレーションの属性を参照する場合、その値が実在しなければならないという制約(同一のリレーションでも構わない。)
 - 履修(科目番号、科目名、学籍番号、成績)
 - 学生(学籍番号、氏名、所属)
 - 組織(学群、学類、専攻)
 - 履修における学籍番号は実在する番号でなければならない。
 - 外部キーを導入することで実現する。

46

外部キー

- 関係Rの属性Xが、関係Sの属性主キーYと一致するとき、XをRの外部キーという。
 - 履修(科目番号、科目名、学籍番号、成績)
 - 学生(学籍番号、氏名、所属)
 - 学籍番号は履修の外部キー

47