

具体的な講義の項目

- イン트로ダクション
- データベースのモデル
- 関係データベースの操作(関係代数、**関係論理**)
- **正規化**
- 問い合わせ言語
- トランザクション処理
- データベースの内部構造
- データベースの問い合わせ処理
- その他(オブジェクト指向DB、アクティブDB等々)

2020/5/11

データベース (©横田)

68

関係論理 (Relational Calculus)

- 集合の表現方法:
 - 外延表現 (Extensional Expression)
 - $\{ 4, 5, 6, 7 \}$
 - 内包表現 (Intensional Expression)
 - $\{ X \mid X \text{ は整数、} X > 3、X < 8 \}$
- 関係自体は？
- 関係論理
 - 検索の結果得られる関係の内容についての内包表現
 - 条件は論理で記述
 - vs. 関係代数: 結果を得るための操作の記述

2020/5/11

データベース (©横田)

69

関係論理(続き)

- 基底節のみを対象とした一階述語論理
 - 再帰を含まなければ VIEW (後述)
 - 再帰を含むと演繹データベース (後述)
- 扱う変数によって 2 種類に分られる
 - タプル関係論理 (Tuple Relational Calculus)
 - タプル変数 (Tuple Variable)を使う
 - 領域関係論理 (Domain Relational Calculus)
 - 領域変数 (Domain Variable)を使う

2020/5/11

データベース (©横田)

70

関係論理の定義(1)

- 原子論理式 (Atomic Formula)
 1. R を関係名、 t をタプル変数とすると、 $R(t)$ は原子論理式 ($t \in R$ と同義)
 2. t_1, t_2 をタプル変数、 A_1, A_2 を属性名、 θ を算術比較演算子 ($=, \neq, <, \leq, >, \geq$) とすると、
$$t_1[A_1] \theta t_2[A_2]$$
は原子論理式
 3. t をタプル変数、 A を属性名、 c を定数、 θ を算術比較演算子 とすると、
$$t[A] \theta c$$
は原子論理式

2020/5/11

データベース (©横田)

71

関係論理の定義(2)

□ タプル関係論理式 (Tuple Relational Calculus)

1. 原子論理式はタプル関係論理式
2. p_1, p_2 をタプル関係論理式とすると、
 $p_1 \vee p_2, p_1 \wedge p_2, \neg p_1$
はタプル関係論理式
3. p をタプル関係論理式とすると、
 $(\forall t)p$ および $(\exists t)p$
はタプル関係論理式
4. 以上の規則以外で作られる論理式はタプル関係論理式でない。
 - 領域関係論理式(Domain Relational Calculus)は領域変数で記述 ($\langle d_1, d_2, \dots, d_n \rangle \in R$)

2020/5/11

データベース (©横田)

72

関係代数と関係論理の等価性(1)

□ Union:

$$\blacksquare R_a \cup R_b = \{ t \mid R_a(t) \vee R_b(t) \}$$

□ Intersection:

$$\blacksquare R_a \cap R_b = \{ t \mid R_a(t) \wedge R_b(t) \}$$

□ Difference:

$$\blacksquare R_a - R_b = \{ t \mid R_a(t) \wedge \neg R_b(t) \}$$

□ Cartesian Product:

$$\blacksquare R_a \times R_b = \{ (t_a, t_b) \mid R_a(t_a) \wedge R_b(t_b) \}$$

2020/5/11

データベース (©横田)

73

関係代数と関係論理の等価性(2)

□ Projection:

$$\blacksquare \pi_{X_1, X_2, \dots, X_n}(R) = \{ t[X_1, X_2, \dots, X_n] \mid R(t) \}$$

□ Selection:

$$\blacksquare \sigma_{X \theta a}(R) = \{ t \mid R(t) \wedge t[X] \theta a \}$$

□ Restriction:

$$\blacksquare \sigma_{X \theta Y}(R) = \{ t \mid R(t) \wedge t[X] \theta t[Y] \}$$

□ Join:

$$\blacksquare R_a \bowtie_{X_a \theta X_b} R_b \\ = \{ (t_a, t_b) \mid R_a(t_a) \wedge R_b(t_b) \wedge t_a[X_a] \theta t_b[X_b] \}$$

2020/5/11

データベース (©横田)

74

関係論理での書き方の例

- 秋葉原支店で、販売価格が100,000円以下で、在庫が10以上である商品の商品番号と販売価格と在庫のデータが欲しい場合の関係論理式

$\{ t1[\text{商品番号}, \text{販売価格}, \text{在庫}] \mid$
秋葉原支店($t1$) $\wedge t1[\text{販売価格}] \leq 100,000 \wedge$
 $t1[\text{在庫}] \geq 10 \}$

キャラクタ入力

$\{ t1[\text{商品番号}, \text{販売価格}, \text{在庫}] \mid$
秋葉原支店($t1$) and $t1[\text{販売価格}] \leq 100,000$ and
 $t1[\text{在庫}] \geq 10 \}$

2019/4/8

データベース (©横田)

75

ちょっと考えてみよう(3-1)

□ 自然言語による問い合わせ

- 秋葉原支店か日本橋支店で、販売価格が100,000円以上の商品の商品番号と販売価格のデータが欲しい場合、キャラクタ入力での関係論理で表現してみよう。

もう少し複雑な問い合わせの例

□ 自然言語による問い合わせ

- 商品分類がComputerに属するGELLの商品で、秋葉原支店か日本橋支店で販売価格が100,000円以下で売られている商品の商品名と商品番号を関係論理式で
 - 製造元関係(商品番号、商品名、製造元)
 - 商品分類関係(商品分類、商品名)
 - 秋葉原支店関係(商品番号、販売価格、在庫)
 - 日本橋支店関係(商品番号、販売価格、在庫)

関係論理での問い合わせ例

- $\{ t1[\text{商品名、商品番号}] \mid$
製造元関係($t1$) $\wedge t1[\text{製造元}] = \text{"GELL"} \wedge$
($\exists t2$) (商品分類関係($t2$)
 $\wedge t1[\text{商品名}] = t2[\text{商品名}]$
 $\wedge t2[\text{商品分類}] = \text{"Computer"}$
 $\wedge ((\exists t3)(\text{秋葉原支店関係}(t3)$
 $\wedge t1[\text{商品番号}] = t3[\text{商品番号}]$
 $\wedge t3[\text{販売価格}] \leq 100,000)$
 $\vee (\exists t4)(\text{日本橋支店関係}(t4)$
 $\wedge t1[\text{商品番号}] = t4[\text{商品番号}]$
 $\wedge t4[\text{販売価格}] \leq 100,000))) \}$

関係代数と関係論理の等価性(3)

□ Division:

- $R_a[X] \div R_b[Y]$
 $= \{ t_a[X_c] \mid R_a(t_a) \wedge (\forall t_b)(R_b(t_b) \Rightarrow$
 $(\exists t_i)(R_a(t_i) \wedge t_i[X_c] = t_a[X_c] \wedge$
 $t_i[X] = t_b[Y])) \}$
- すなわち
 $R_a[X] \div R_b[Y]$
 $= \{ t_a[X_c] \mid R_a(t_a) \wedge (\forall t_b)(\neg R_b(t_b) \vee$
 $(\exists t_i)(R_a(t_i) \wedge t_i[X_c] = t_a[X_c] \wedge$
 $t_i[X] = t_b[Y])) \}$

関係代数と関係論理の等価性(3')

□ Division:

- $R_a[X] \div R_b[Y] = \{t[X_c] \mid R_a(t) \wedge R_b[Y] \subseteq \text{Iset}(R_a, X_c, t[X_c])\}$
 - ここで、 $\text{Iset}(R, X, x) = \{t[X_c] \mid R(t) \wedge t[X] = x\}$
 - つまり、 $\text{Iset}(R_a, X_c, t[X_c]) = \{t[X] \mid R_a(t) \wedge t[X_c] = t[X_c]\}$
 - $R_a(t)$ に含まれるタプル t の中で、補属性の値 $t[X_c]$ と同じ値を持つ R_a 中のタプルの対象属性の集合 Iset に対して R_b が部分集合となるようなタプルの補属性を集めたもの
 - $R_s(\text{販売実績関係})[\text{販売員}] \div R_p(\text{販売員関係})[\text{販売員}]$ の場合:
 - t が (田中, DeskTop) の時は、 $\text{Iset}(\text{販売実績}, \text{商品名}, \text{DeskTop})$ が {田中, 佐藤} となるので、 R_p つまり {田中, 鈴木, 佐藤} はその部分集合にならないが、 t が (田中, LapTop) なら $\text{Iset}(\text{販売実績}, \text{商品名}, \text{LapTop})$ は {田中, 鈴木, 佐藤} で R_p の部分集合なので、「LapTop」は結果に含まれることになる。

2020/5/11

データベース (©横田)

80

ちょっと考えてみよう(3-2)

□ $R_a[X] \div R_b[Y]$ の余りを表す関係論理式は？

□ キャラクタ入力

- $(\exists t_b) R_b(t_b) \rightarrow E(tb)Rb(tb)$
- $\neg (\forall t_i) R_a(t_i) \rightarrow \text{not}(A(ti))Ra(ti)$

2020/5/11

データベース (©横田)

81

正規化 (Normalization)

□ 第一正規形 (First Normal Form: 1NF)

- 提案者: E. F. Codd [1972]
- 関係の属性の値が全て原子的(Atomic)
 - それ以上細かくならない
 - 関係代数、関係論理を適用するための条件
- vs. 非第一正規形 (Non-First Normal Form: NF²)
 - 属性の値が原子的でない: 集合を属性の値として持つ

2020/5/11

データベース (©横田)

82

非第一正規形の例:

販売地域	販売員	国籍	製造元	商品名	商品番号
千代田区	田中	米国	GELL	Desktop	GDC10
		日本	Fepson	Laptop	FLC33
			Hanon	Laser Pt.	HLP05
新宿区	鈴木	米国	GELL	Laptop	GLC20
		日本	Tenao	LCDisplay	TLCD70
渋谷区	佐藤	日本	Hanon	Laser Pt.	HLP05

2020/5/11

データベース (©横田)

83

第一正規形に変形:

販売地域	販売員	国籍	製造元	商品名	商品番号
千代田区	田中	米国	GELL	Desktop	GDC10
千代田区	田中	日本	Fepson	Laptop	FLC33
千代田区	田中	日本	Hanon	Laser Pt.	HLP05
新宿区	鈴木	米国	GELL	Laptop	GLC20
新宿区	鈴木	日本	Tenao	LCDisplay	TLCD70
渋谷区	佐藤	日本	Hanon	Laser Pt.	HLP05

正規化の動機

□ 更新時異常 (Update Anomaly)

- 一つの事項を変更するのに、複数の部分の変更を要求する
 - 一時的にデータベースの状態の一貫性がなくなる。
- 削除によって必要以上の情報が消失
- 関係を分解することによって対応
 - どのように分解するかの方針: 第2正規形以降の正規形

□ 情報無損失分解 (Information Lossless Decomposition)

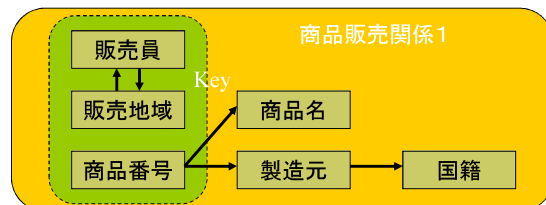
- 関係を射影によって分解後、自然結合によって元の関係を復元可能

関数従属性 (Functional Dependency: FD)

- 関係 R が与えられたとき、 R 中の任意のタプルについて、 R の属性 X の各属性値が、 R の他の属性 Y のただ1つの属性値と正確に対応する場合に、 R の属性 Y は属性 X に関数従属である。

$$(X \rightarrow Y) \Leftrightarrow (\forall t, t' \in R, t[X] = t'[X] \Rightarrow t[Y] = t'[Y])$$

- 前出の例の関数従属性



第2正規形 (Second Normal Form: 2NF)

□ 提案者

- E. F. Codd [1972]

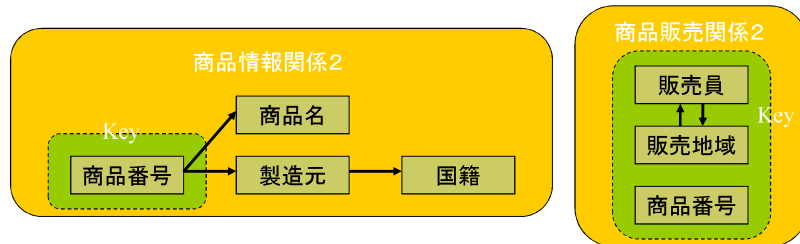
□ 定義:

- 1NF の関係 R において、どのキー候補にも属さない属性 A は、いかなるキー候補 K の真部分集合にも関数従属でない。

□ 前出の 1NF の関係の例を 2NF に

第2正規形

□ 例における関数従属性



注: ここでは {販売員、商品番号} および {販売地域、商品番号} はキー候補とする。

第2正規形

商品情報関係2

国籍	製造元	商品名	商品番号
米国	GELL	Desktop	GDC10
日本	Fepson	Laptop	FLC33
日本	Hanon	Laser Pt.	HLP05
米国	GELL	Laptop	GLC20
日本	Tenao	LCDisplay	TLCD70

商品販売関係2

販売地域	販売員	商品番号
千代田区	田中	GDC10
千代田区	田中	FLC33
千代田区	田中	HLP05
新宿区	鈴木	GLC20
新宿区	鈴木	TLCD70
渋谷区	佐藤	HLP05

ちょっと考えてみよう(3-3)

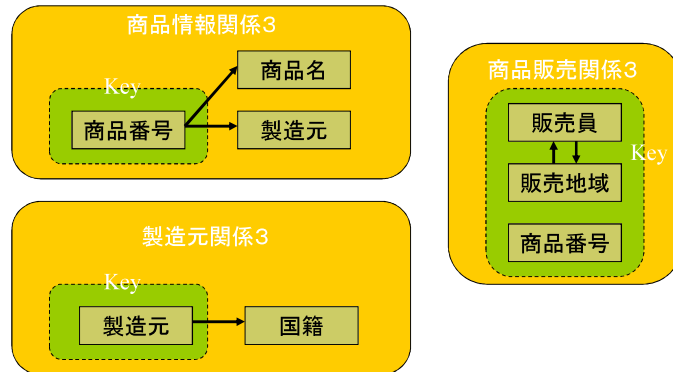
- 前出の第1正規形から第2正規形にすることで解決された更新時異常は何でしょう

第3正規形 (Third Normal Form: 3NF)

- 提案者
 - E. F. Codd [1972]
- 定義:
 - 2NF の関係 R において、どのキー候補にも属さない属性 A は、基本キーに対して非推移的に従属する
- 前出の 2NF の関係の例を 3NF に

第3正規形

□ 例における関数従属性



2020/5/11

データベース (©横田)

92

第3正規形

商品情報関係3			商品販売関係3		
製造元	商品名	商品番号	販売地域	販売員	商品番号
GELL	Desktop	GDC10	千代田区	田中	GDC10
Fepson	Laptop	FLC33	千代田区	田中	FLC33
Hanon	Laser Pt.	HLP05	千代田区	田中	HLP05
GELL	Laptop	GLC20	新宿区	鈴木	GLC20
Tenao	LCDisplay	TLCD70	新宿区	鈴木	TLCD70
			渋谷区	佐藤	HLP05

国籍	製造元
米国	GELL
日本	Fepson
日本	Hanon
日本	Tenao

製造元関係3

2020/5/11

データベース (©横田)

93

Boyce/Codd 正規形 (Boyce Codd Normal Form: BCNF)

□ 提案者

- Boyce and Codd [1974]

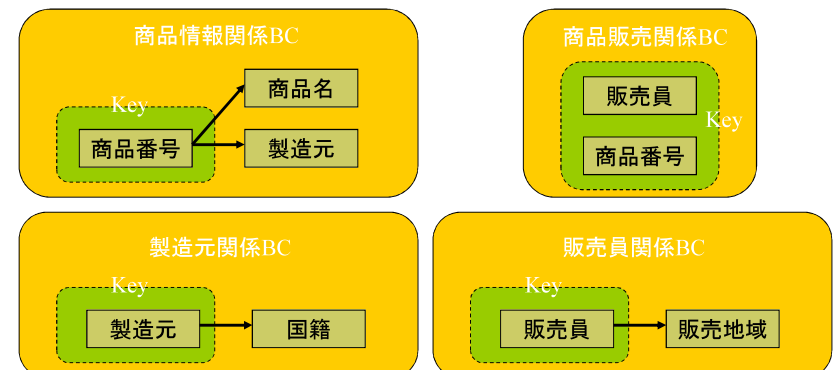
□ 定義:

- 1NF の関係 R において、いかなる属性集合 X に対しても、 X に含まれない属性のうちどれかが X に関数従属であるときは、 R の全ての属性は、 X に関数従属となる。

□ 前出例を BCNF に分解

Boyce/Codd 正規形

□ 例における関数従属性



2020/5/11

データベース (©横田)

94

2020/5/11

データベース (©横田)

95

Boyce/Codd正規形

商品情報関係BC

製造元	商品名	商品番号
GELL	Desktop	GDC10
Fepson	Laptop	FLC33
Hanon	Laser Pt.	HLP05
GELL	Laptop	GLC20
Tenao	LCDisplay	TLCD70

商品販売関係BC

販売員	商品番号
田中	GDC10
田中	FLC33
田中	HLP05
鈴木	GLC20
鈴木	TLCD70
佐藤	HLP05

製造元関係BC

国籍	製造元
米国	GELL
日本	Fepson
日本	Hanon
日本	Tenao

販売員関係BC

販売地域	販売員
千代田区	田中
新宿区	鈴木
渋谷区	佐藤

2020/5/11

データベース (©横田)

96

ちょっと考えてみよう(3-4)

講義名	開講期	使用講義室	担当教員	教員連絡先	単位数
-----	-----	-------	------	-------	-----

- 上記のようなスキーマの関係「講義」があったとする。また、それぞれの属性の間には、{講義名、開講期}→使用講義室、{講義名、開講期}→担当教員、担当教員→教員連絡先、講義名→単位数という関数従属性があるとする。
 - 情報無損失分解によって、2NFではあるが、3NFではないような関係の集合になるか。関係名を適当につけて、それらの関係のスキーマを示せ。
 - この分解によりどのような問題点が改善されたか述べよ。
 - BCNFにするには、どのような関係の集合にすればよいか。関係名を適当につけて、それらの関係のスキーマを示せ。

2020/5/11

データベース (©横田)

97

関数従属性の限界

□ 例:

- 「新宿区」の販売員は「鈴木」と「吉田」で、「新宿区」の重点商品は「Laptop」と「Inkjet Pt.」。「渋谷区」の販売員は、「佐藤」で、「渋谷区」の重点商品は、「Laptop」、「Laser Pt.」、「LCDisplay」。

販売地域	販売員	重点商品
新宿区	[鈴木、吉田]	[Laptop、Inkjet Pt.]
渋谷区	佐藤	[Laptop、Laser Pt.、LCDisplay]

第一正規形ではない

2020/5/11

データベース (©横田)

98

第一正規形に直すと

販売地域	販売員	重点商品
新宿区	鈴木	Laptop
新宿区	鈴木	Inkjet Pt.
新宿区	吉田	Laptop
新宿区	吉田	Inkjet Pt.
渋谷区	佐藤	Laptop
渋谷区	佐藤	Laser Pt.
渋谷区	佐藤	LCDisplay

- キー候補は、{販売地域、販売員、重点商品}
 - よって左の関係は BCNF
- 「渋谷区」の担当に「伊藤」加わると、3タプルを同時に入れる必要が生じる。(更新時異常)

2020/5/11

データベース (©横田)

99

多値従属性 (Multivalued Dependency: MVD)

- 提案者: R. Fagin [1977]
- 属性 A, B, C を関係 R の属性集合の任意の部分集合とする。 R のある (A -値、 C -値) の対に対応する B の値の集合が A の値にのみ依存して、 C の値と独立である場合、かつそのときに限り B は A に多値従属しているといい、次のように表す
 - $A \twoheadrightarrow B \mid C$ (B が (A) に多値従属しているという)
- ここで $\mid C$ は省略可能、また同時に逆も成り立つ
 - $A \twoheadrightarrow C \mid B$
- $B \subseteq A$ または $C = \varnothing$ の時、常に $A \twoheadrightarrow B$ が成り立つので、自明の多値従属性という

多値従属性

- 関数従属性をより一般化したもの
 - MVD: 値から集合への従属性
 - 集合の要素が一つなら FD
- 前出の例:
 - 自明でない多値従属性:
 - 販売地域 $\twoheadrightarrow$ 販売員 $\mid$ 重点商品
 - 販売地域 $\twoheadrightarrow$ 重点商品 $\mid$ 販売員
 - 販売店が決まれば、販売員の集合が決まるし、重点商品の集合も決まる。

第4正規形 (Fourth Normal Form: 4NF)

- 提案者: R. Fagin [1977]
- 多値従属性を使って、正規形を定義
- 定義
 - 自明でない任意の多値従属性 $A \twoheadrightarrow B$ を満たすような関係 R の属性の部分集合 A と B があるとき、 R の全ての属性が A に関数従属である
- 前出の例は、候補キーからの関数従属でない多値従属を含むので、4NF でない。
 - (販売地域 $\twoheadrightarrow$ 販売員) と (販売地域 $\twoheadrightarrow$ 重点商品) のどちらか一方が関数従属 (つまり対象が集合でない) なら第4正規形

第4正規形 (Fourth Normal Form: 4NF)

- 定義
 - 関係 R において、ある自明でない多値従属性 $X \twoheadrightarrow Y$ が成り立つときはいつでも関係 R の全ての属性 A が X に関数従属である (オリジナル)
 - (a) 属性 X を持つ 1NF の関係 R が 4NF であるとは、関係 R 中の各自明でない多値従属性 $U \twoheadrightarrow V$ に対して、関係 R 中で X は U に関数従属である
 - (b) 1NF の関係 R のキー従属性の集合を Δ としたとき、関係 R 中の各多値従属性 σ は Δ から導かれる。すなわち、全ての多値従属性はキーの結果となる。
 - (c) 属性 X を持つ 1NF の関係 R 中の各多値従属性 σ に対して、 $K \rightarrow X$ から σ が導かれるようなキー従属性 $K \rightarrow X$ が関係 R 中に存在する

第4正規形(続き)

- $A \twoheadrightarrow B \mid C$ であるとき、 $\{A, B\}$ と $\{A, C\}$ に情報無損失分解できる
 - 分解された関係は、第4正規形

販売員関係4		重点商品関係4	
販売地域	販売員	販売地域	重点商品
新宿区	鈴木	新宿区	Laptop
新宿区	吉田	新宿区	Inkjet Pt.
渋谷区	佐藤	渋谷区	Laptop
		渋谷区	Laser Pt.
		渋谷区	LCDisplay

n-分解可能 (n-decomposable)

- 提案者: Aho, Beeri, Ullman [1977]
- n 個の射影に情報無損失分解できることを指す。
- ある関係を二つの射影には情報無損失分解できないが、三つの射影なら情報無損失分解可能である場合がある。
- 例:

販売員	商品名	販売先
鈴木	Laptop	東工大
吉田	Laptop	東工大
吉田	Inkjet Pt.	東工大
吉田	Laptop	早稲田

n-分解可能(続き)

射影1		射影2		射影3	
販売員	商品名	商品名	販売先	販売員	販売先
鈴木	Laptop	Laptop	東工大	鈴木	東工大
吉田	Laptop	Inkjet Pt.	東工大	吉田	東工大
吉田	Inkjet Pt.	Laptop	早稲田	吉田	早稲田

販売員	商品名	販売先
鈴木	Laptop	東工大
鈴木	Laptop	早稲田
吉田	Laptop	東工大
吉田	Inkjet Pt.	東工大
吉田	Laptop	早稲田

射影1と射影2を自然結合する

元の関係になかったタプルが現れる

この結果に射影3を結合すると、元の関係に戻る

結合従属性 (Join Dependency: JD)

- 提案者: R. Fagin [1979]
- R を関係とし、 $A, B, \dots, Z$ を R の属性集合の任意の部分集合とする。このとき、 R が $A, B, \dots, Z$ 上の射影の結合と等しいとき、その時に限り、 R は結合従属性 $* (A, B, \dots, Z)$ を満たす。
- 結合従属性と多値従属性の関係
 - 関係 R が、多値従属性 $A \twoheadrightarrow B \mid C$ を満たすときかつその時に限り $* (AB, AC)$ を満たす。
- 結合従属性は、多値従属性を一般化したもの

第5正規形 (Fifth Normal Form: 5NF)

- 提案者: R. Fagin [1979]
- 定義
 - 関係 R の全ての結合従属性が R の候補キーによって与えられる
- 射影結合正規形(PJNF)とも呼ばれる。
- 前出の商品販売先関係は 5NF ではない
 - 3分解可能で、その3分解可能性は {販売員、商品名、販売先}の組み合わせが候補キーであることに意味されないため
- 結合従属性は、多値従属性を一般化したものであるため、全ての 5NF は 4NF である。

2020/5/11

データベース (©横田)

108

従属性の関係

- 従属性の一般化
 - 関数従属性を一般化したものが多値従属性
 - 多値従属性を一般化したものが結合従属性
- 正規形の定義の類似性
 - BCNF、4NF、5NF の定義の言い換え
 - BCNF: 関係 R の全ての関数従属性が R の候補キーによって与えられるときかつその時に限り R は BCNF である。
 - 4NF: 関係 R の全ての多値従属性が R の候補キーによって与えられるときかつその時に限り R は 4NF である。
 - 5NF: 関係 R の全ての結合従属性が R の候補キーによって与えられるときかつその時に限り R は 5NF である。

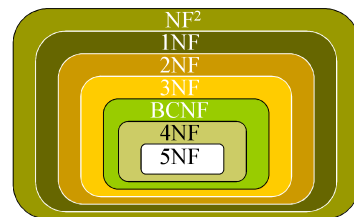
2020/5/11

データベース (©横田)

109

正規形間の関係

- 1NF、2NF、3NF、BCNF、4NF、5NF は包含関係にある
 - 例:
 - 3NF なら必ず 2NF
 - 理由:
 - 2NF の定義は、1NF を前提
 - 3NF の定義は、2NF を前提
 - BCNF なら、キー候補の部分集合からの従属性はなく、推移的でもない
 - 4NFの多値従属性は BCNF の関数従属性を一般化したもの
 - 5NFの結合従属性は 4NF の多値従属性を一般化したもの



2020/5/11

データベース (©横田)

110

本日の講義の最後に

- 初めてのオンライン講義でこなれていませんが、コメントがあったら書き込んで下さい(オプション)

2020/5/7

データベース (©横田)

111