

Tema 5: Normalización en Bases de Datos

Andrés Cordon Franco
e-mail: `acordon@us.es`

Bases de Datos 2007/08

Ciencias de la Computación e IA (<http://www.cs.us.es/>)
Universidad de Sevilla

1 Introducción

2 Dependencias funcionales

- Definición
- Dependencia funcional elemental
- Dependencia funcional transitiva. Campos equivalentes

3 Formas normales

- Primera Forma Normal
- Segunda Forma Normal. Campos principales
- Tercera Forma Normal

4 Bibliografía

Introducción (I)

<i>AUTOR</i>	<i>PAÍS</i>	<i>COD-LIBRO</i>	<i>TÍTULO</i>	<i>EDITORIAL</i>
Date, C.	USA.	01	DB	Ad
Date, C.	USA	02	SQL(I)	Ad
Gardarin	Chile	03	ModeloER	Verlag
Gardarin	Chile	04	SQL(II)	ACM
Kim,W.	China	04	SQL(II)	ACM

Problemas:

- 1 Redundancia de datos
- 2 Anomalías de modificación de datos
- 3 Anomalías de inserción de datos
- 4 Anomalías de borrados de datos

Introducción (II)

Definición

Formas normales (FN): conjunto de restricciones sobre tablas relacionales que evitan los problemas de redundancia de datos y de anomalías de modificación, inserción y borrado de datos.

$$1FN \Leftarrow 2FN \Leftarrow 3FN \Leftarrow FNBC \Leftarrow 4FN \Leftarrow 5FN$$

Nota: En este curso sólo estudiaremos hasta la tercera forma normal, 3FN.

Dependencias funcionales

Definición

Sean R una tabla relacional y X_1, \dots, X_k e Y campos de R .

Diremos que Y **depende funcionalmente** de X_1, \dots, X_k si al fijar un valor de los campos X_1, \dots, X_k , se fija también el valor que puede tomar el campo Y en la tabla R .

Lo escribiremos: $X \rightarrow Y$.

$X \equiv$ determinante o implicante

$Y \equiv$ campo implicado

Nota: $X_1, \dots, X_k \rightarrow Y_1, \dots, Y_n$ significa que cada campo Y_i tiene dependencia funcional respecto del conjunto X_1, \dots, X_k .

Dependencias funcionales. Ejemplo

ESCRIBE(autor,país,código,título,editorial,páginas)

$autor \rightarrow pais$

$codigo \rightarrow titulo$ $codigo \rightarrow editorial$ $codigo \rightarrow paginas$

$codigo \rightarrow titulo, editorial, paginas$

$cod, autor \rightarrow editorial, pais$

Nota: No debe confundirse con el concepto de *campo derivado*. Es claro que si Y es un campo derivado de X (por ejemplo, *edad* de *fechanacimiento*), entonces $X \rightarrow Y$.

Sin embargo, al revés no es cierto. Para que Y tenga dependencia funcional de X , no es necesario que se pueda deducir el valor *concreto* del campo Y a partir del campo X .

Dependencia funcional plena (I)

Definición

*Diremos que Y tiene dependencia funcional **plena** o **completa** del conjunto de campos X_1, \dots, X_k si depende funcionalmente de dicho conjunto, pero **NO** depende de ningún subconjunto **propio** de X_1, \dots, X_k .*

Dependencia funcional plena (II)

Ejemplo: EVAL(alumno, asignatura, dpto, nota)

(●) El campo *nota* tiene dependencia **plena** de *alumno*, *asignatura*:

$alumno, asignatura \rightarrow nota$

$alumno \nrightarrow nota$

$asignatura \nrightarrow nota$

(●) El campo *dpto* **NO** tiene dependencia **plena** de *alumno*, *asignatura*:

$alumno, asignatura \rightarrow dpto$

$asignatura \rightarrow nota$

Dependencia funcional trivial

Definición

*Diremos que el campo Y tiene dependencia funcional **trivial** del conjunto de campos X_1, \dots, X_k si Y es uno de los campos X_1, \dots, X_k .*

Ejemplo: Las siguientes dependencias funcionales son triviales:

autor, cod \rightarrow autor

alumno, asignatura, nota \rightarrow nota

libro \rightarrow libro

Dependencia funcional elemental (I)

Definición

*El campo Y tiene dependencia funcional **elemental** del conjunto de campos X_1, \dots, X_k si dicha dependencia es:*

- *completa, y*
- *no trivial*

Nota: En una dependencia funcional elemental $X_1, \dots, X_k \rightarrow Y$, el campo implicado Y siempre ha de ser **unitario**.

Dependencia funcional elemental (II)

Ejemplo: EVAL(alumno,edad,asignatura,dpto,nota)

Un conjunto de dependencias funcionales **elementales** que describen la tabla EVAL es:

alumno, asignatura \rightarrow *nota*

alumno \rightarrow *edad*

asignatura \rightarrow *dpto*

Nota: Para la normalización de una tabla relacional, solamente se tienen en cuenta las dependencias funcionales **elementales**.

Dependencia funcional transitiva (I)

Definición

*El campo Z tiene dependencia funcional **transitiva** respecto de los campos X_1, \dots, X_k a través de los campos Y_1, \dots, Y_n si se cumple que:*

- $X_1, \dots, X_k \rightarrow Y_1, \dots, Y_n$, y
- $Y_1, \dots, Y_n \rightarrow Z$, pero
- $Y_1, \dots, Y_n \not\rightarrow X_1, \dots, X_k$

Dependencia funcional transitiva (II)

Ejemplo 1: LIBRO(isbn,páginas,autor,país)

pais depende transitivamente de *isbn* a través de *autor*, pues:

$isbn \rightarrow autor$, $autor \rightarrow pais$, y
 $autor \not\rightarrow isbn$

Ejemplo 2: CORREO(login,NIF,nombre,apellidos,edad,país)

pais NO depende de manera transitiva de *NIF* a través de *login*,
puesto que:

$NIF \rightarrow login$, $login \rightarrow pais$, pero
 $login \rightarrow NIF$

Nota: *NIF* y *login* son campos **equivalentes**

Campos equivalentes (I)

Definición

Sean R una tabla relacional y $X_1, \dots, X_k, Y_1, \dots, Y_n$ campos de R . Diremos que X_1, \dots, X_k e Y_1, \dots, Y_n son conjuntos de campos **equivalentes** si:

- X_1, \dots, X_k depende funcionalmente de Y_1, \dots, Y_n , e
- Y_1, \dots, Y_n depende funcionalmente de X_1, \dots, X_k , e

Lo escribiremos

$$X_1, \dots, X_k \leftrightarrow Y_1, \dots, Y_n$$

Nota: Habitualmente $k = n = 1$ y hablaremos de campos equivalentes de una tabla.

Campos equivalentes (II)

VENTA(código,número,vendedor,artículo,color,precio)

$\text{codigo} \leftrightarrow \text{numero}, \text{vendedor}$

$\text{codigo} \rightarrow \text{articulo}$

$\text{articulo} \rightarrow \text{color}, \text{precio}$

(●) *articulo* NO depende transitivamente de *numero, vendedor*:

$\text{numero}, \text{vendedor} \rightarrow \text{codigo}, \text{codigo} \rightarrow \text{articulo}, \text{ pero}$

$\text{numero}, \text{vendedor} \leftrightarrow \text{codigo}$

(●) *precio* SÍ depende transitivamente de *numero, vendedor*:

$\text{numero}, \text{vendedor} \rightarrow \text{articulo}, \text{articulo} \rightarrow \text{precio}, \text{ pero}$

$\text{articulo} \nrightarrow \text{numero}, \text{vendedor} (\star)$

(★ Se admiten devoluciones de artículos)

Primera Forma Normal: 1FN

Definición (Codd, 1970)

*Una tabla relacional R está en primera forma normal (**1FN**) si NO contiene campos multivaluados.*

Nota: Por la propia definición del modelo de datos relacional, NO se admiten campos multivaluados. En consecuencia, TODAS las relaciones del modelo de datos relacional están automáticamente en 1FN.

Campos principales y no principales (I)

Para estudiar si una tabla R está en 2FN o 3FN se parte de:

- 1 Conjunto de las **dependencias elementales** de R ,
- 2 conjunto de todas las **claves candidatas** para R ,
- 3 conjunto de los **campos equivalentes** de R ,
- 4 conjunto de los **campos principales** de R , y
- 5 conjunto de los **campos no principales** de R .

Definición

*Un campo de una tabla relacional R es **principal** si pertenece a alguna de las claves candidatas de R .*

Campos principales y no principales (II)

VENTA(código,número,vendedor,artículo,color,precio)

- Dependencias elementales:

codigo → *numero*, *codigo* → *vendedor*, *codigo* → *articulo*

numero,vendedor → *codigo*

articulo → *color*, *articulo* → *precio*

- Claves candidatas:

- 1 *codigo*

- 2 (*numero,vendedor*)

- Campos equivalentes: *codigo* ↔ *numero,vendedor*

- Campos principales: *codigo,numero,vendedor*

- Campos no principales: *articulo,color,precio*

Segunda Forma Normal: 2FN

Definición (Codd, 1970)

*Una tabla relacional R está en **segunda forma normal (2FN)** si:*

- *R está en 1FN, y*
- *cada campo **no principal** de R tiene dependencia funcional plena respecto de cada una de las claves candidatas de R .*

Nota: Si R tiene k claves candidatas y m campos no principales, entonces para asegurar que R está en 2FN son necesarias $m \cdot k$ comprobaciones.

Segunda Forma Normal. Ejemplo (I)

EVAL(alumno,edad,asignatura,dpto,nota)

- Dependencias funcionales elementales:

alumno → *edad*

asignatura → *dpto*

alumno, asignatura → *nota*

- Claves candidatas: (*alumno, asignatura*)
- Campos equivalentes: no hay
- Campos principales: *alumno, asignatura*
- Campos no principales: *edad, dpto, nota*

Número de comprobaciones: $3 \cdot 1 = 3$

Segunda Forma Normal. Ejemplo (II)

EVAL(alumno,edad,asignatura,dpto,nota)

Comprobaciones:

- 1 ¿Es *alumno, asignatura* \rightarrow *edad* completa? NO
- 2 ¿Es *alumno, asignatura* \rightarrow *dpto* completa? NO
- 3 ¿Es *alumno, asignatura* \rightarrow *nota* completa? SÍ

EVAL **no está en 2FN** porque el campo no principal *edad* no tiene dependencia plena respecto de la clave *alumno, asignatura*.

Segunda Forma Normal. Ejemplo (III)

EVAL(*alumno*,*edad*,*asignatura*,*dpto*,*nota*)

(●) ¿Cómo conseguir un conjunto de tablas en 2FN equivalente?

R1(*alumno*,*edad*)

PK=*alumno*

R2(*asignatura*,*dpto*)

PK=*asignatura*

R3(*alumno*,*asignatura*,*nota*)

PK=(*alumno*,*asignatura*)

FK=*alumno* (\rightsquigarrow R1) FK=*asignatura* (\rightsquigarrow R2)

Segunda Forma Normal. Propiedades

Si el implicante de una dependencia funcional es unitario, entonces la dependencia es automáticamente completa. Por tanto:

Propiedad 1: Si todas las claves candidatas de una tabla relacional R son *unitarias*, R está automáticamente en 2FN.

Ejemplo: ALUMNO(código,nif,nombre,edad,dirección)

Claves candidatas: (1) *codigo*, (2) *nif*

Puesto que ambas claves candidatas son unitarias, la tabla ALUMNO está en 2FN.

Tercera Forma Normal: 3FN

Definición (Cood, 1970)

*Una tabla relacional R está en **tercera forma normal (3FN)** si:*

- *R está en 2FN, y*
- *ningún campo no principal de R depende **transitivamente** de alguna clave candidata de R .*

Nota: Si R está en 2FN y tiene k claves candidatas y m campos no principales, entonces para asegurar que R también está en 3FN son necesarias $m \cdot k$ comprobaciones.

Tercera Forma Normal. Ejemplo (I)

LIBRO(código,isbn,título,autor,país)

- Dependencias funcionales elementales:

codigo \rightarrow *isbn*

isbn \rightarrow *codigo*, *isbn* \rightarrow *titulo*, *isbn* \rightarrow *autor*

autor \rightarrow *pais*

- Claves candidatas: (1) *codigo* (2) *isbn*
- Campos equivalentes: *codigo* \leftrightarrow *isbn*
- Campos principales: *codigo*, *isbn*
- Campos no principales: *titulo*, *autor*, *pais*

Puesto que todas las claves candidatas son unitarias, la tabla LIBRO está automáticamente en 2FN.

Tercera Forma Normal. Ejemplo (II)

LIBRO(código,isbn,título,autor,país)

¿Está la tabla LIBRO en 3FN? Núm. de comprobaciones: $3 \cdot 2 = 6$

- 1 *¿título* depende transitivamente de *código*? NO
- 2 *¿título* depende transitivamente de *isbn*? NO
- 3 *¿autor* depende transitivamente de *código*? NO
- 4 *¿autor* depende transitivamente de *isbn*? NO
- 5 *¿país* depende transitivamente de *código*? SÍ
- 6 *¿país* depende transitivamente de *isbn*? SÍ

El campo *país* tiene dependencia transitiva de *código* a través del campo *autor*. Luego, LIBRO **no está en 3FN**.

Tercera Forma Normal. Ejemplo (III)

LIBRO(código,isbn,título,autor,país)

(•) ¿Cómo conseguir un conjunto de tablas en 3FN equivalente?

R1(*autor, país*)

PK = *autor*

R2(*codigo, isbn, titulo, autor*)

PK = *codigo*

FK = *autor* (\rightsquigarrow R1)

Tercera Forma Normal. Propiedades (I)

Propiedad 2: Si R está en 2FN y posee un único campo no principal, entonces R también está en 3FN.

Ejemplo: SOCIO(*nif*,*código*,*edad*)

$nif \rightarrow codigo$, $nif \rightarrow edad$

$codigo \rightarrow nif$

- Claves candidatas: (1) *nif*, (2) *codigo*
- Campos equivalentes: $nif \leftrightarrow codigo$
- Campos principales: *nif*, *codigo*
- Campos no principales: *edad*

Puesto que todas las claves candidatas son unitarias, SOCIO está en 2FN. Puesto que tiene un único campo no principal(*edad*), SOCIO también está en 3FN.

Tercera Forma Normal. Propiedades (II)

Propiedad 3: Si todos los campos de R son principales, entonces R está automáticamente en 3FN.

Ejemplo: ESTUDIAR($nif, cod_alum, asignatura$)

$$nif \rightarrow cod_alum, \quad cod_alum \rightarrow nif$$

■ Claves candidatas:

1 $(nif, asignatura)$

2 $(cod_alum, asignatura)$

■ Campos equivalentes: $nif \leftrightarrow cod_alum$

■ Campos principales: $nif, cod_alum, asignatura$

■ Campos no principales: no hay

ESTUDIAR está automáticamente en 3FN porque sólo posee campos principales.

Bibliografía

- **Concepción y diseño de bases de datos**, Adoración de Miguel, Mario Piattini, RA-MA Editorial (1993).
- **Apuntes de Ficheros y Bases de Datos**, Mercedes Marqués, Universidad Jaume I en Castellón (2001).

<http://www3.uji.es/~mmarques/f47/apun/apun.html>