Base de Datos. 1C2009.
Clase práctica de Normalización (Continuación)
Docente: Alejandro Eidelsztein

DEPENDENCIAS MULTIVALUADAS

EJEMPLO:

Sea el siguiente esquema:

PROFESIONAL(NOMBRE, TITULO, IDIOMA)

Donde almacenamos la información correspondiente a un grupo de profesionales con los títulos que cada uno posee y los idiomas que cada uno domina.

Sea la siguiente instancia r1:

NOMBRE	TITULO	IDIOMA
hugo	físico	inglés
hugo	matemático	francés
hugo	físico	francés
hugo	matemático	inglés
maría	médica	alemán
maría	médica	italiano
luis	abogado	portugués
luis	abogado	inglés
luis	matemático	portugués
luis	matemático	inglés

Cuál es la clave de este esquema?

Aunque PROFESIONAL no tiene otras dependencias funcionales que las triviales (y por lo tanto está en FNBC) observamos que hay redundancia de información: Tenemos que repetir el título por cada idioma que el profesional sabe y de igual forma tenemos que repetir el idioma por cada título que tiene.
Se presentan anomalías de inserción: Por ejemplo si Hugo aprende un nuevo idioma debemos agregar dos tuplas, lo mismo si obtiene un nuevo título. Si queremos agregar un nuevo profesional seguramente deberemos agregar más de una tupla.

Esto es así porque (NOMBRE-TITULO) y (NOMBRE-IDIOMA) son independientes entre si.

Vemos que este esquema se puede decomponer en **(PROF-TIT)** y **(PROF-IDIOMA)**:

r1.1:

NOMBRE	TITULO
hugo	físico
hugo	matemático
maría	médica
luis	abogado
luis	matemático

r1.2:

NOMBRE	IDIOMA
hugo	inglés
hugo	francés
maría	alemán
maría	italiano
luis	portugués
luis	inglés

y que esta descomposición es SPI dado que r1.1 |X| r1.2 = r1
Si en cambio ahora consideramos esta otra instancia r2 (donde eliminamos algunas tuplas), vemos que la descomposición propuesta no es SPI dado que r2.1 |X| r2.2 es un superconjunto de r2.

r2:

NOMBRE	TITULO	IDIOMA
hugo	físico	inglés
hugo	matemático	francés
hugo	físico	francés
maría	médica	alemán
maría	médica	italiano
luis	abogado	portugués
luis	abogado	inglés
luis	matemático	portugués

r2.1:

NOMBRE	TITULO
hugo	físico
hugo	matemático
maría	médica
luis	abogado
luis	matemático

r2.2:

NOMBRE	IDIOMA
hugo	inglés
hugo	francés
maría	alemán
maría	italiano
luis	portugués
luis	inglés

O sea que para poder descomponer PROFESIONAL en forma SPI se debe cumplir *que ciertas tuplas deben existir obligatoriamente*

Definiremos un nuevo tipo de dependencia que llamaremos *dependencia multivaluada (DMV)*

En nuestro caso diremos que se cumple la DMV: NOMBRE ->-> TITULO

(Decimos que NOMBRE multidetermina TITULO o que TITULO es multideterminado por NOMBRE)

También se verifica: NOMBRE ->-> IDIOMA

En forma general diremos que dado:

	X	Y	Z	(donde Z =R–XY)
t1				
t2				
t3				
t4				
…				

La DMV X->->Y se cumple en R, si para cualquier instancia r de R, se cumple que si existen 2 tuplas t1 y t2 en r para las cuales t1[X] = t2[X], luego deben existir otras 2 tuplas t3 y t4 en r tal que:

t1[X] = t2[X] = t3[X] = t4[X]

t3[Y] = t1[Y] y t3[Z] = t2[Z]

t4[Y] = t2[Y] y t4[Z] = t1[Z]

	X	Y	Z	(Z =R–XY)
t1	x1	y1	z2	
t2	x1	y2	z1	
t3	x1	y1	z1	
t4	x1	y2	z2	

Dadas t1 y t2 la DMV asegura que t3 debe existir
La existencia de t4 se desprende del carácter simétrico de la definición

Con esta definición las DFs son un caso particular de las DMVs.

EJEMPLO:

Sea R(A,B,C, D) sujeto a la DMV: A->->B

¿Si sabemos que tenemos las tuplas t1=(a b1 c1 d1), t2=(a b2 c2 d2), t3=(a b3 c3 d1) están en r, qué otras tuplas deben estar en r?

	A	B	C	D	
t1	a	b1	c1	d1	
t2	a	b2	c2	d2	
t3	a	b3	c3	d1	
t4	**a**	**b1**	**c2**	**d2**	**(por t1 y t2)**
t5	**a**	**b2**	**c1**	**d1**	**(por t1 y t2)**
t6	**a**	**b1**	**c3**	**d1**	**(por t1 y t3)**
t7	**a**	**b3**	**c1**	**d1**	**(por t1 y t3)**
t8	**a**	**b3**	**c2**	**d2**	**(por t2 y t3)**
t9	**a**	**b2**	**c3**	**d1**	**(por t2 y t3)**

REGLAS DE INFERENCIA PARA DFs y DMVs:

3 reglas de Armstrong para DFs ya dadas:

1. Reflexividad: Si Y $\subseteq$ X entonces X-- >Y
2. Aumento: Para cualquier W, si X-->Y entonces XW-->WY
3. Transitividad: Si X-->Y e Y-->Z entonces X-->Z

3 reglas adicionales para FDs ya dadas:

4. Unión: X-->Y y X-->Z entonces X-->YZ
5. Pseudotransitividad: Para cualquier W, X-->Y e YW-->Z entones XW--> Z
6. Descomposición: X-->YZ entonces X-->Y y X-->Z

4 reglas específicas para DMVs:

7. Complementación: Si X->->Y entonces X->->R-XY
8. Reflexividad: Si Y $\subseteq$ X entonces X->->Y
9. Aumento: Para cualquier W, si V$\subseteq$ W, X->->Y entonces XW->->YV
10. Transitividad: Si X->->Y e Y->->Z entonces X->->Z-Y

3 reglas adicionales para DMVs:

11. Unión: Si X->->Y y X->->Z entonces X->->YZ
12. Pseudotransitividad: Si X->->Y y YW->->Z entonces XW->->Z-WY
13. Descomposición: Si X->->Y y X->->Z entonces X->->(Y$\cap$Z)
X->->(Y- Z)
X->->(Z-Y)

3 reglas que conectan DFs y DMVs:

14. Conversión: Si X-->Y entonces X->->Y
15. Interacción: Si X->->Y y existe un W tal que W$\cap$Y=ϕ, W-->Z y Z $\subseteq$Y entonces X-->Z
16. Pseudotransitividad mixta: Si X->->Y y XY-->Z entonces X-->Z-Y

EJEMPLO:

Dados:

R(A,B,C,D,E) y
D= {A->->BC, DE->->C}

Derivar AE->->BD

1. A->->BC (dada)
2. A->->DE (complementación de 1)
3. DE->->C (dada)
4. A->->C (transitividad de 2 y 3)
5. AE->->C (aumento de 4 con W=E y V=ϕ)
6. AE->->BD (complemento de 5)

EJEMPLO:

Demostrar la siguiente regla:

Si X->->Y entonces X->->(Y-X)

1. X->->Y (dada)
2. X-->X (reflexividad de DFs)
3. X->->X (conversión)
4. X->->(Y-X) (descomposición entre 1 y 3)

EJEMPLO:

Sean:
R(A,B,C,D,E) y
D= {A->->BC, DE->->C}

Aplicando las reglas de inferencia vemos que:

D |= {A->->DE, A->->C, AD->->BE}

La siguiente instancia r cumple con todas estas DMVs:

A	B	C	D	E
a	b	c	d	e
a’	b’	c’	d	e
a’	b’	c	d	e
a	b	c’	d	e
a”	b’	c’	d’	e

EJEMPLO:

Sean:
R(A,B,C,D,E) y
D= {A->->BC, D-->C}

Aplicando las reglas de inferencia vemos que:

D |= {A-->C}

La siguiente instancia r cumple con estas DFs y DMVs:

A	B	C	D	E
a	b	c’	d	e
a	b’	c’	d’	e’
a	b’	c’	d	e
a	b	c’	d’	e’

DMVs TRIVIALES:

Dados R y $X,Y \subseteq R$,

X->->Y es trivial si

$R= XY$ o $Y \subseteq X$

DESCOMPOSICION BINARIA SPI CON FDs y DMVs:

Dados R, ρ= (R1, R2) y **D** (conjunto de DFs y DMVs)
decimos que ρ es SPI sii

1) La dependencia multivaluada (R1 $\cap$ R2) ->- > (R1 – R2) está en **D+**
o
2) La dependencia multivaluada (R1 $\cap$ R2) ->- > (R2 – R1) está en **D+**

4FN (CUARTA FORMA NORMAL):

R está em 4FN com respecto a un conjunto **D** de DFs y DMVs si para toda DMV no trivial de la forma X->->Y, X es una superclave de R.

De la definción se desprende que:

Si R está en 4FN, también está en FNBC
Si R está en 4FN, las dependencias no triviales que valen son funcionales

ALGORITMO PARA DESCOMPONER EN 4FN SPI:

Es análogo al que usamos para descomponer en FNBC, o sea:

Si la dependencia no trivial X->->Y viola 4FN entonces hay que descomponer en

XY y R-Y

y así sucesivamente…

Si hay una dependencia multivaluada X ->-> Y que viola 4FN se parte R en R1 y R2 de la siguiente forma:

```
         R
        /  \      X ->- >Y
       /    \
      /      \
R1= XY      R2= R-Y
```

EJEMPLO:

Sean R(A,B,C,D,E)
D= {A-->BC, C->->DE}

Vemos que R no está en 4FN porque la clave es ADE y C->->DE viola la 4FN.

Entonces descomponemos en:

R1= ABC   R2= CDE

Esta descomposición está en 4FN con respecto a **D** aunque **D**|= A->->B, A->->C que son no triviales y valen en R1 pero A es una clave para R1.

```
             ABCDE
              /   \        C->->DE
             /     \
            /       \
      R2= CDE      R1= ABC
```