

5. Dependencias multivaluadas.

5.1 Definiciones básicas .

Una dependencia multivaluada es una sentencia que se escribe : $X \twoheadrightarrow Y$ y cuyo significado intuitivo es el siguiente : A cada valor de X se le asocia un conjunto de valores de Y independiente del contexto (si X e Y son subconjuntos de T , el contexto es $Z = T - (X \cup Y)$) .

Como vimos en la Introducción , esto quiere decir que la dependencia multivaluada asigna a cada valor de $\text{Dom} (X)$ un valor de $P (\text{Dom} (Y))$, y esta asignación no varía con el contexto .

Es muy importante no confundir la existencia de una dependencia multivaluada (en adelante , mvd) con el hecho de que entre los dominios de X e Y pueda establecerse una correspondencia 1 : n , lo que es un error bastante frecuente .

Veremos ahora tres definiciones (por supuesto , equivalentes) de mvd . Cada una tiene su mayor o menor utilidad según el punto de vista bajo el que las mvd's sean consideradas .

Definición usando la notación de dependencia generalizada :

X	Y	Z
x	y	z
x	y'	z'
x	y	z'
x	y'	z

$X \twoheadrightarrow Y$ sii la existencia de las tuplas $\langle x \ y \ z \rangle$, $\langle x \ y' \ z' \rangle$, implica la existencia de estas otras : $\langle x \ y \ z' \rangle$, $\langle x \ y' \ z \rangle$.

En notación de dependencia generalizada , a las dos primeras se les llama “tuplas hipótesis “ y a las dos últimas “ tuplas conclusión “ .

La consistencia requiere de la aparición de las tuplas conclusión en cualquier instancia válida del esquema en el que la mvd tiene lugar , dada la aparición de las tuplas hipótesis .

En un contexto deductivo , es fácil formular la mvd como una regla , y no es necesario que las conclusiones acompañen de modo explícito a las hipótesis ,siendo suficiente que se puedan deducir de las mismas .

Definición usando el criterio de proyección-join :

En el esquema $R(T , L)$ tiene lugar la mvd $X \twoheadrightarrow Y$, sii para toda instancia r del mismo se cumple :

$$\Pi_{XY}(r) \bowtie \Pi_{XZ}(r) = r$$

Esta definición no es más que enunciar de otra forma el Teorema de Delobel y Casey que veremos enseguida .

Su utilidad es la aplicación al diseño del conocimiento previo de las mvd's , pero no resulta adecuada para la extracción de las mismas .

Definición usando el criterio de selección-proyección :

Cuando veamos la estructura inferencial de las mvd's , justificaremos que , en toda mvd , puede suponerse vacía la intersección de implicante e implicado .

Sea $Z=T-(X \cup Y)$.

La dependencia $X \rightarrow \rightarrow Y$ tiene lugar en $R (T)$ sii , para toda instancia r del esquema :

$\Pi_Y(\sigma_{X=x} (r)) = \Pi_Y(\sigma_{X=x \wedge Z=z} (r))$, siendo x , z valores genéricos de los dominios de X y Z , respectivamente .

Sea la relación :

$R (\text{ Pintor , Cuadro , Museo })$

Con las dependencias funcionales :

$\text{Cuadro} \rightarrow \text{Pintor}$

$\text{Cuadro} \rightarrow \text{Museo}$

Dado que un pintor determinado tiene asociado el conjunto bien definido de sus cuadros , podría pensarse en si tiene o no lugar la mvd $\text{Pintor} \rightarrow \rightarrow \text{Cuadro}$.

Designemos por comodidad a los atributos por sus iniciales . De acuerdo con la definición , la mvd tendrá lugar sii se cumple :

$\Pi_C(\sigma_{P=Goya} (R)) = \Pi_C(\sigma_{P=Goya \wedge M=Prado} (R))$

Y esto para cualquier par de valores (p , m) de los dominios de P y M .

Claramente se ve que no se cumple la igualdad con carácter general . La primera proyección nos da los cuadros de Goya , mientras que la segunda nos da el subconjunto de los mismos perteneciente a los fondos del Prado .

Esta definición es la que usaremos para extraer las mvd's de un cierto esquema .

5.2 Axiomas para mvd's .

La estructura inferencial completa para mvd's fue enunciada por Beeri , Fagin y Howard , y puede consultarse en el libro de Ullmann citado en la Bibliografía básica .

No usaremos en este curso todos los Axiomas , ya que , siendo nuestro objetivo el diseño , y siendo las mvd's propiedades relativas al contexto ,

“aparecen y desaparecen “ en el proceso de normalización , por lo que no analizaremos todas las existentes en la relación universal $R (T , L)$, sino que será suficiente con ir detectando alguna de las posibles mvd's que vulneren la definición de Cuarta Forma Normal .

Axiomas :

- Toda dependencia funcional es un caso particular de mvd .(1)
- Axiomas de Armstrong para dependencias funcionales .(2)
- Siendo $Z=T-(X \cup Y)$, $\{ X \rightarrow \rightarrow Y \} \Rightarrow X \rightarrow \rightarrow Z$ (3)
- $\{ X \rightarrow \rightarrow Y , Y \rightarrow \rightarrow Z \} \Rightarrow X \rightarrow \rightarrow (Z - Y)$ (4)
- $\{ X \rightarrow \rightarrow Y \} \Rightarrow X \rightarrow \rightarrow (Y - X)$ (5)

Con este subconjunto de axiomas tendremos suficiente para nuestros objetivos de diseño .

Mencionar por último que una mvd es “trivial” si la unión de implicante e implicado es T .

Supongamos la siguiente relación :

Alumno	Asignatura	Actividad
J.Fernández	Inglés	Deportes
J.Fernández	Historia	Teatro
J.Fernández	Literatura	Deportes
J.Fernández	Inglés	Teatro
J.Fernández	Historia	Deportes
J.Fernández	Literatura	Teatro

En ella se reflejan las asignaturas en las que están inscritos los alumnos de un colegio y las diferentes actividades opcionales que han elegido .

El alumno J.Fernández cursa tres asignaturas y ha elegido dos actividades . Para que esta información sea correctamente reflejada en la Base de Datos , necesito seis tuplas (2×3) . Y ello para que la respuesta a cualquier interrogación que relacione **asignatura** con **actividad** sea completa .

Por lo demás , es fácil ver que la relación es 3FN , siendo su clave el conjunto de los tres atributos (no hay dependencias funcionales) .

Analicemos ahora la situación desde el punto de vista de las dependencias multivaluadas .

Por comodidad , llamaremos A al alumno , S a la asignatura y C a la actividad . El análisis de las posibles mvd's es sencillo .

Comenzaremos por establecer si $A \rightarrow \rightarrow S$, aplicando el criterio de selección-proyección .

$\Pi_S (\sigma_{A=a} (R))$ = conjunto de asignaturas que cursa el alumno “a” .

$\Pi_S (\sigma_{A=a \wedge C=c} (R))$ =conjunto de asignaturas que cursa el alumno “a” que ha elegido la actividad “c” .

El segundo acceso está “sobreespecificado”, y ambos conjuntos son idénticos. La dependencia analizada es cierta, y por tanto lo es también $A \rightarrow \rightarrow C$, como puede comprobarse aplicando el axioma (3).

Analicemos las restantes posibilidades :

$C \rightarrow \rightarrow A$ ¿?

$\Pi_A(\sigma_{C=c}(R))$ = conjunto de alumnos que han elegido la actividad “c”.

$\Pi_A(\sigma_{C=c \wedge S=s}(R))$ = conjunto de alumnos que han elegido la actividad “c” y están inscritos en la asignatura “s”.

En general, el segundo conjunto está incluido en el primero, así que la mvd no tiene lugar y, como consecuencia, tampoco $C \rightarrow \rightarrow S$ tiene lugar.

De modo análogo se llega a que no hay mvd's no triviales implicadas por S.

No analizamos las posibles mvd's implicadas por dos atributos, ya que, en este contexto, son triviales (una mvd es trivial si la unión de implicante e implicado es T).

La clave del esquema R (sigue definiéndose como siempre, en relación a las dependencias funcionales), es ASC.

La **Cuarta Forma Normal (4FN)**, fue definida por R.Fagin y es como sigue :

Un esquema 1FN es 4FN sii toda mvd no trivial está implicada por una clave.

En nuestro ejemplo, hay dos mvd's no triviales implicadas por A, que no es clave. No es 4FN y, como hemos visto, presenta un alto grado de redundancia.

5.3 Teorema de Delobel y Casey .

Sea R(T) un esquema y sean $R_1(T_1)$ y $R_2(T_2)$ sus proyecciones sobre T_1 y T_2 , respectivamente.

La descomposición es LJ sii $T_1 \cap T_2 \rightarrow \rightarrow T_i - T_j$, $(i,j)=(1,2)$ ó $(2,1)$.

5.4 Normalización 4FN .

Dado R (T , L), una descomposición 4FN LJ se consigue en la forma siguiente :

- Calcular todas las claves.
- Analizar la existencia de posibles mvd's
- Si se detecta alguna no trivial ni implicada por una clave, $X \rightarrow \rightarrow Y$
- Proyectar R sobre $X \cup Y$ y sobre $T - Y$. De acuerdo con el Teorema de Delobel y Casey, la descomposición es LJ. El caso más general, obliga a estudiar desde el principio ambas proyecciones, calculando para

ambas la proyección de L , las claves y las mvd's que pudiesen surgir en los nuevos contextos (y que , con relación al esquema inicial , se denominan “embebidas”).

- Si en alguno de los subesquemas se detectan mvd's no triviales ni implicadas por alguna de las claves respectivas , el proceso continúa .

4FN \Rightarrow BCFN \Rightarrow 3FN

En nuestro ejemplo , proyectaríamos R sobre AS y AC , obteniendo R₁ y R₂ tales que $R=R_1 \bowtie R_2$ para cualquier instancia del esquema .

5.5 Ejemplo de diseño 4FN .

Consideremos la siguiente relación :

R

Asignatura	Prerrequisito	Alumno	Año
Ecuaciones diferenciales	Cálculo	J.Fernández	01
Ecuaciones diferenciales	Análisis matemático	J.Fernández	02
Ecuaciones diferenciales	Cálculo	E.Ruiz	00
Ecuaciones diferenciales	Análisis matemático	E.Ruiz	03
Ecuaciones diferenciales	Cálculo	S.García	01
Ecuaciones diferenciales	Análisis matemático	S.García	02

En la que , por cada asignatura que se imparte en una cierta Facultad , se indica cuáles son las que constituyen un prerrequisito para dicha asignatura , qué alumnos están matriculados en esa asignatura y en qué año aprobaron cada uno de los prerrequisitos de la misma .

Designaremos los atributos :

Asignatura (A)

Prerrequisito (B)

Alumno (C)

Año (D)

Es inmediato comprobar la existencia de una sola dependencia funcional no trivial :

$BC \rightarrow D$

La clave es ABC , de modo que R no es 2FN (D depende parcialmente de la clave) .

Para abordar el diseño 4FN , hay que analizar posibles mvd's .

Sólo interesan las que involucran un número de atributos menor o igual que 3 , ya que , por el axioma (5) , siempre podemos trabajar con dependencias que tengan disjuntos sus dos términos , y entonces toda dependencia de 4 atributos es trivial en este contexto (la unión de sus dos términos es T) .

Diseño:

1. Usando el axioma (1) , vemos en primer lugar que tiene lugar la mvd : $BC \rightarrow \rightarrow D$
2. Aplicando el Teorema de Delobel y Casey , proyectamos sobre ABC y BCD . El esquema $R_1(ABC)$ no tiene dependencias funcionales , por lo que su clave es ABC . En $R_2(BCD)$ se tiene que $BC \rightarrow D$ y la clave es BC. Ahora buscamos posibles mvd's .
3. En R_1 , analizamos si $A \rightarrow \rightarrow B$ (si es así , también se daría $A \rightarrow \rightarrow C$, por (3) . Empleamos el criterio de selección –proyección :
 $\Pi_B(\sigma_{A=a}(R))$ =conjunto de prerrequisitos de la asignatura “a”.
 $\Pi_B(\sigma_{A=a \wedge C=c}(R))$ =conjunto de prerrequisitos de la asignatura “a” en la que está matriculado el alumno “c” (acceso sobreespecificado).
 Es obvio que ambos conjuntos coinciden por lo que $A \rightarrow \rightarrow B$ y $A \rightarrow \rightarrow C$. Esto basta para ver que R_1 no es 4FN .
4. Aplicando de nuevo Delobel y Casey , proyectamos R_1 en $R_{11}(AB)$ y en $R_{12}(AC)$, ambas 4FN .
5. Analicemos ahora R_2 . Veamos , para empezar , si $B \rightarrow \rightarrow C$, por aplicación del criterio de selección-proyección : $\Pi_C(\sigma_{B=b}(R))$ =conjunto de alumnos que han aprobado el prerrequisito “b”. El conjunto de comparación es $\Pi_C(\sigma_{B=b \wedge D=d}(R))$ =conjunto de alumnos que han aprobado el prerrequisito “b” el año “d”. En general ambos conjuntos son diferentes, no teniendo lugar la mvd . Puede comprobarse que no hay mvd's no triviales en R_2 , que es por tanto 4FN (y , en consecuencia , 3FN y BCFN) .

En la práctica , cuando añadamos el R_0 (clave) a nuestro diseño según Ullmann , y cuando la clave tenga 3 ó más atributos , será frecuentemente necesario pasar R_0 a 4FN .

El proceso seguido , garantiza la preservación del carácter LJ del conjunto .