Normalización. Conceptos adicionales

Definición formal de normalización

Descomposición sin pérdida de información ni de semántica de la relación universal (o de una colección de relaciones equivalentes a la misma) en una colección de relaciones en la que las anomalías de actualización (inserción, borrado y modificación) no existan o sean mínimas.

Descriptor

Cualquier subconjunto de atributos de una relación. Formalmente:

- Dada la relación R (A_1, A_2,..., A_n), Se dice que X es un descriptor de R,
  Si X $\subseteq$ (A_1, A_2,..., A_n)

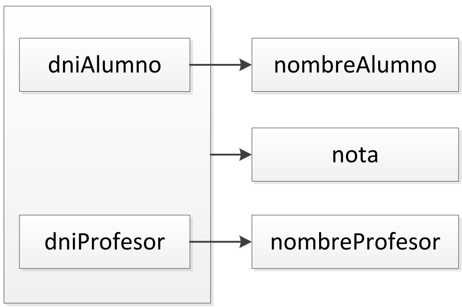
Grafo de dependencias funcionales

Es una representación de las dependencias funcionales elementales del conjunto de atributos de una relación.

Supongamos una relación:

R(A={dniProfesor, dniAlumno, nombreProfesor, nombreAlumno, nota},
DF={dniProfesor→nombreProfesor; dniAlumno→nombreAlumno;
dniProfesor, dniAlumno→nota})

Se representaría como:



Atributo extraño

Dada una dependencia funcional X → Y, siendo X el determinante, decíamos que la dependencia funcional era plena o completa si X → Y, y no existe ningún subconjunto de X, X1, tal que X1 → Y.

Si la dependencia funcional no es completa, será porque existe ese subconjunto X1 que determina a Y. Luego habrá otro subconjunto complementario de campos que “sobra” en la implicación. Por ejemplo:

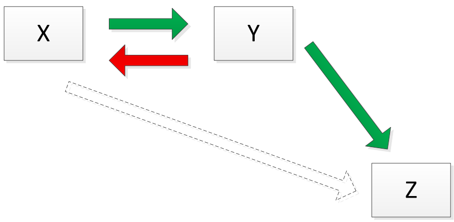
Alumno, asignatura → edad

Si X = {alumno, asignatura}, Y = {edad}, en este caso vemos que X1 = {alumno} ya verifica que X1 → Y (alumno → edad), y sobraría X2 = {asignatura}. Este sería el concepto de atributo extraño, asignatura en este caso.

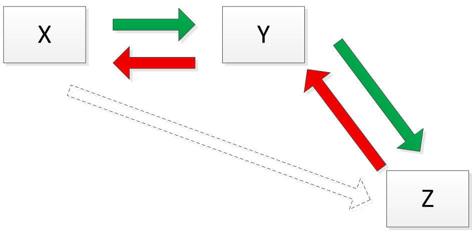
Un **atributo extraño** o **ajeno** en una dependencia funcional (DF) es un atributo del determinante de la DF que provocan que ésta no sea plena.

Dependencia transitiva estricta

Decíamos que Z tenía una dependencia funcional transitiva con respecto a X a través de Y si: X → Y, Y → Z y Y /→X. (X, Y, Z no necesariamente disjuntos)



Pues bien, esta dependencia será estricta si además no Z / → Y.



Por ejemplo, dada la relación: CURSOPROG (codcurso, codprograma, coddepto), en donde se tiene para cada curso su código, el programa que lo incluye y el departamento del que depende el programa (suponemos que un curso se imparte en un único programa y que un programa lo prepara un único departamento) se tendrán las siguientes DF:

codcurso → codprograma
codprograma → coddepto

Como en un programa se imparten varios cursos:

codprograma /→ codcurso

Se cumple la **DF transitiva**:

codcurso → coddepto

Además, **es estricta** porque en un departamento se tienen varios programas:

coddepto /→ codprograma

Superclave y clave de una relación

Dada una relación R(A), se denomina **superclave** de la relación a un subconjunto K tal que:

$K \subseteq A,\ K \neq \phi,\ K \rightarrow A$

Es decir, un subconjunto de atributos no vacío que determina a todos los demás.

Si además no existe ningún subconjunto K' de K, tal que: K' → A, entonces decimos que K es una **clave** de R.

Como vimos, en la relación R(T) se denominan ***atributos principales*** a aquellos que son ***elementos de alguna clave*** de R(T). A los demás atributos se les denomina ***no principales***.

Consecuencia lógica

A partir del conocimiento de ciertas DF se puede inferir la existencia de otras que no se encontraban en el conjunto inicial:

- Dado un esquema de relación (una relación del DL, una tabla relacional): R (A, DF) con sus atributos y dependencias funcionales, es posible deducir de DF nuevas dependencias funcionales que sean una consecuencia lógica del conjunto de partida.
- Las nuevas dependencias **f** son consecuencia lógica de DF (vienen implicadas por DF) y se representan como: **DF |= f**.

Cierre de un conjunto de dependencias

El **cierre de un conjunto de dependencias funcionales** DF (que se denota DF^+) es el conjunto de todas las dependencias que son consecuencia lógica de DF:

$DF^+ = \{ X \rightarrow Y \mid DF \models X \rightarrow Y \}$

DF será siempre un subconjunto del cierre ($DF \subseteq DF^+$). Por lo tanto, las notaciones R(A,DF) y R(A,DF^+) definen el mismo esquema de relación (en el sentido de que la segunda proviene de la primera, se puede obtener de ella).

Estas definiciones no permiten el cálculo del cierre, siendo necesarias unas **reglas de derivación** que faciliten la implicación lógica de dependencias.

Estas reglas de derivación, se conocen como **Axiomas de Armstrong**, y forman un conjunto completo y correcto de axiomas.

Reglas de Derivación

Dado un conjunto DF de dependencias funcionales, se dice que **f se deriva de DF**, lo que se representa por DF |— f, ***si f se puede obtener por aplicación sucesiva de los axiomas de Armstrong a DF (o a un subconjunto de DF)***, es decir, si existe una secuencia de dependencias f1, f2, … fn tal que fn = f, donde cada fi es bien un elemento de DF o ha sido derivada a partir de las dependencias precedentes aplicando las reglas de derivación.

Aunque son conceptos distintos, se cumple siempre que si una dependencia f es una consecuencia lógica de un conjunto de dependencias, también será posible derivarla de dicho conjunto aplicando los axiomas de Armstrong, y viceversa; es decir:

∀ f | DF |– f se_implica_que DF |= f (propiedad de corrección), y

∀ f | DF |= f se_implica_que DF |– f (propiedad de plenitud)

Axiomas de Armstrong

Básicos

- A1: Reflexividad:

  X → X
  Si Y ⊆ X, X → Y (X → Y es una DF trivial)

- A2: Aumentatividad

  Si X → Y y Z ⊆ W, entonces XW → YZ

- A3: Transitividad

  Si X → Y e Y → Z, entonces X → Z

Derivados

- D1: Proyectividad

  Si X → Y, entonces X → Y' si Y' ⊂ Y

- D2: Unión o aditividad

  Si X → Y y X → Z, entonces X → YZ

- D3: Pseudotransitividad

  Si X → Y e YW → Z, entonces XW → Z

Ejercicio:

Dado el esquema de relación: R(A, B, C, D, E; {A → B, C → D, D → E}).

Demostrar que AC → ABCDE

Inconveniente

Ver biblio: http://alarcos.inf-cr.uclm.es/doc/bda/doc/teo/ant/BDa-t7a.pdf, diap. 10.

Cierre de un Descriptor

Dado un esquema de relación R(A, DF), el cierre transitivo de un descriptor X de R respecto al conjunto de dependencias DF, denotado como X^{+}_{DF} es el subconjunto de los atributos de A tales que $X \rightarrow X^{+}_{DF} \in DF^{+}$, siendo X^{+}_{DF} máximo en el sentido de que la adición de cualquier atributo vulneraría la condición anterior.

Algoritmo de Ullman

Entrada:

Un conjunto de dependencias DF y de atributos, R(A,DF).

Un descriptor X subconjunto de A

Salida:

X^{+}, cierre de X respecto a DF.

Proceso:

1) $X^{+} = X$

2) Repetir hasta que no se añadan más atributos a X^{+}:

Para cada dependencia $Y \rightarrow Z \in DF$

si $(Y \in X)$ y $\neg(Z \in X+)$ entonces

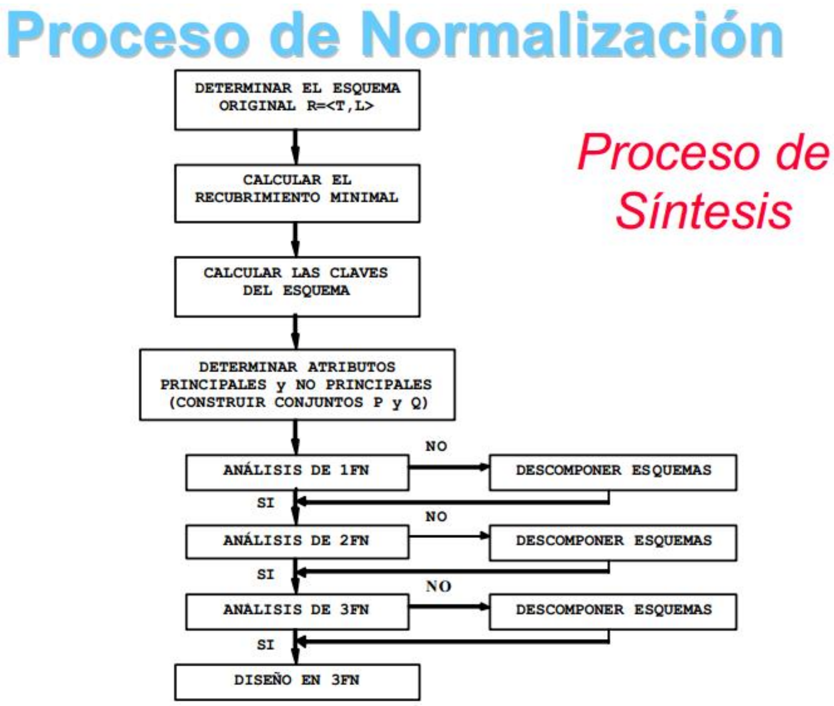
$X^{+} = X^{+} \cup Z$

Ejercicio:

Dada la relación R ({CE, NE, P, G, CP, C}, DF), con DF={CE → NE, NE → CE, P → CE, G → P, (CP, P) → G, CE → C, P → C}, hallar el cierre del descriptor (CP,P).

Ejercicios adicional

Determinar el nivel de normalización de la relación R=<T,L>, tal que T={A, B, C, D, E, F} y L={E→C; AC → D; D → A, A → BF; CD → E}, sabiendo que L es un RMNR y que las claves son AC, AE, CD y DE.



(Ver detalle en:

http://www-oei.eui.upm.es/Asignaturas/BD/BD/docbd/tema/normalizacion.pdf)