Guía para control 2 CC42A, Otoño 2009 *

Diego Díaz Espinoza

23 de mayo de 2009

1. Resumen Materia

1.1. Pautas informales para el diseño

1. Semántica de los atributos : La semántica de los atributos involucrados en una relación debe ser explícita, clara y concisa.

2. Reducción de los valores redundantes en las tuplas : Minimizar el espacio utilizado en relaciones agrupando correctamente estas y evitar redundancias en las tuplas. Además hay que considerar las posibles anomalías de inserción, eliminación y actualización, a saber:

   - Inserción : inserción de tuplas de las que se desconoce valores (inserción de valores nulos).
   - Eliminación : eliminación de ciertas tuplas dejan a la relación sin atributos utilizados en otras tablas. Mejor es separar las tablas, caso típico: comunas, departamentos, etc.
   - Actualización : Mejor separar tablas para que una actualización no se propague en muchas tuplas.

3. Reducción de los valores nulos en las tuplas : Evitar los valores nulos en la tablas, puesto que semánticamente es dificil saber cuál es su origen. Un valor nulo puede representar: atributo no aplica a cierta tupla, se desconoce el valor del atributo para esa tupla ó el valor existe pero aún no se ingresa.

4. Prohibición de tuplas espurias : verificar que no existan tuplas erróneas o epurias cuando se utiliza reunión por igualdad sobre atributos que son claves primarias o foráneas.

1.2. Dependencias funcionales

1.2.1. Definición

Sea R una relación y X e Y subconjuntos de atributos de R, entonces una dependencia funciona de X sobr Y se anota $X \rightarrow Y$ siempre y cuando $\forall t_1, t_2$ (t_1 y t_2 tuplas de R) tal que $t_1[X] = t_2[X]$ se tiene que $t_1[Y] = t_2[Y]$. Dicho en lenguaje coloquial: los valores de Y dependen funcionalmente de los valores de X ó los valores de X determinan unívocamente los valores de Y.

*No reemplaza la lectura del texto y las cátedras.

1.2.2. Cerradura

Sea F el conjunto de dependencias funcionales definidos sobre una esquema de relación R, entonces, se denota por F^+ al conjunto de todas las dependencias funcionales que sobre R. Estas son todas aquellas dependencias funcionales que se cumplen en TODA extensión del esquema de relación R (notar que -generalmente- no basta con analizar una extensión de R para conocer todas las dependencias funcionales sobre R).

1.2.3. Reglas de inferencia

1. $Y \subseteq X \Rightarrow X \to Y$
2. $\{X \to Y\} \models XZ \to YZ$
3. $\{X \to Y, Y \to Z\} \models X \to Z$
4. $\{X \to YZ\} \models X \to Y$
5. $\{X \to Y, X \to Z\} \models X \to YZ$
6. $\{X \to Y, WY \to Z\} \models XW \to Z$

Se dice que 1 a 3 son completas y correctas. Completas: La cerradura de F sobre R (F^+) se puede obtener por el uso repetido de 1 a 3 (hasta que no se puedan obtener más relaciones). Correctas: Cualquier inferencia que hagamos apartir de 1,2 y/o 3 se aplica para toda extensión del esquema de relación R.

Para un algoritmo para generar la cerradura de F sobre R revisar Elmasri capítulo 12.

1.2.4. Cerradura de X (X^+) bajo F

Se dice que X^+ es la cerradura de X bajo F (donde X, X^+ son conjuntos de atributos y F es el conjunto de dependencias funcionales sobre R) si y sólo si todos los atributos de X^+ se pueden determinar funcionalmente de X.

1.2.5. Equivalencia funcional

Se dice que un conjunto de dependencias funcionales E está cubierto por F si: toda dependencia funcional de E está también en F o puede generarse a partir de F. Formalmente: toda dependencia funcional de E está también en F^+.

1.2.6. Conjunto minimal de df (dependencias funacionales

Se dice que un conjunto E de df es minimal si cumple con:

1. Para toda df en E el miembro derecho de la df tiene un solo atributo.
2. No se puede quitar ninguna df de E y seguir teniendo un conjunto equivalente -funcionalmente- a E.
3. No se puede reemplazar ninguna df $X \to A$ por $Y \to A$ con Y subconjunto propio de X y sequir teniendo un conjunto de df equivalente funcionalmente a E.

1.3. Formas Normales

1.3.1. Recordatorio

1. Superclave : Conjunto de atributos S de la relación R tal que no existe ningún par de tuplas t_1, t_2 distintas tal que $t_1[S] = t_2[S]$
2. Clave : Superclave minimal (todos los atributos son necesarios y suficientes para hacerlo superclave).
3. Clave candidata : toda clave es clave candidata.
4. Clave primaria : Clave candidata escogida arbitrariamente. Todo esquema de relación R tiene una y sólo una clave primaria.
5. Atributo primo : Un atributo del esquema de relación R es primo si y sólo si es miembro de cualquier clave de R.
6. Atributo no primo : Un atributo del esquema de relación R tal que no es miembro de ninguna clave de R.
7. Dependencia funcional total : una df $X \to Y$ es total si cualquier eliminación de un atributo A de X hace que la df deje de ser válida.
8. Dependencia funcional parcial : una df $X \to Y$ es parcial si puedo eliminar un atributo A de X y la df sigue siendo válida.
9. Dependencia funcional transitiva : una df $X \to Y$ es transitiva si existe un conjunto de atributos Z que no sea un subconjunto de cualquier clave de R tal que cumple con $X \to Z$ y $Z \to Y$
10. Cobertura mínima de G para F : Es el conjunto de df minimal equivalente a F.

1.3.2. Primera Forma Normal

Prohíbe relaciones dentro de relaciones. Prohíbe atributos multivaluados. Sólo permite atributos únicos e indivisibles (atómicos) para cada uno de los atributos.

1.3.3. Segunda Forma Normal

Si y sólo si todo atributo no primo A en R depende funcionalmente de manera total de la clave primaria de R. Formalmente: Ningún atributo no primo A en R depende parcialmente de cualquier clave de R.

1.3.4. Tercera Forma Normal

Si y sólo si:

1. Está en 2FN.
2. Ningún atributo no primo de R depende transitivamente de la clave primaria.

De manera formal se tiene que para toda df $X \to A$ en R se cumple una de las dos siguientes condiciones:

- X es superclave.
- A es atributo primo de R.

1.3.5. Forma Normal de Boyce-Codd

De manera formal se tiene que para toda df $X \to A$ en R se cumple:

- X es superclave.

2. Ejercicios Resueltos Dependencias Funcionales:

Los siguientes ejercicios resueltos son un complemento a las soluciones de la guía "Guía Dependencias Funcionales" y "Normalización en 3FN y FNBC" de Mauricio Monsalve del semestre primavera 2007. Los puntos 1 y 2 se dejan al alumno ya que son altamente recomendable que los piense y resuelva solo. Los items que no tienen solución es porque la solución está en la misma guía.

1.
2. 2.7 Por definición: 1)$t_1[A] = t_2[A] \Rightarrow t_1[BC] = t_2[BC] \Rightarrow t_1[B] = t_2[B] \wedge t_1[C] = t_2[C]$ Además: 2) $t_1[C] = t_2[C] \Rightarrow t_1[D] = t_2[D]$. Luego por 1) y 2) $t_1[A] = t_2[A] \Rightarrow t_1[D] = t_2[D]$ y $t_1[A] = t_2[A] \Rightarrow t_1[C] = t_2[C]$ y por 1) $t_1[A] = t_2[A] \Rightarrow t_1[B] = t_2[B]$. Así tenemos que A "genera" $\{B, C, D\}$. Además no hay ninguna dependencia funcional para E por lo que no podemos generar E a partir de ninguna df. Por lo mismo la clave es $\{AE\}$.
3. [1]Para todos estos ejercicios hay dos opciones: calcular la cerradura del conjunto de df usando el algoritmo del libro y buscar si aparece por ahí la df particular que piden; ó, hacer una inferencia apartir de las df dadas.

   3.2 Usando 4) sobre $A \to BC$ se tiene que: $A \to C$ y $A \to B$ luego se desprende que $A \to E$ (usando 3) y $C \to E$). Además usando $A \to B$ y $B \to D$ se tiene por 3) que $A \to D$. Luego usando que $A \to D$ y $CD \to F$ junto con 6) se infiere que $AC \to F$. Ahora como $A \to C$ entonces $A \to AC$ (por 1, ya que $A \to AA \to AC \to F$ con lo que $A \to F$. Finalmente usando que $A \to E$ y $A \to F$ más la regla 5) se tiene que $A \to EF \to G$ con lo que $A \to G$.

   3.5 Usando el hecho que $B \to CF$ se tiene que $B \to C$. Por otro lado dado que $G \to BI$ entonces $G \to B$. Además $B \to C \models B \to BC$ y por ende $BI \to BCI$. Luego uniendo esto último con $G \to BI$ y utilizando la regla de inferencia 3) se tiene que $G \to BCI$. Luego usando 2) se tiene que $AG \to ABCI$ y finalmente usando 4) se tiene que $AG \to BCI$.

[1]Notar que especifica sólo usando las reglas de Armstrong, esto es: las reglas de inferencia del 1) al 3), para esta solución se usarán indistintamente todas, ya que las reglas 4) a 6) pueden derivarse de las de Armstrong por completitud. Si no cree, demuéstrelo! :P

3. Ejercicios Resueltos Normalización

2.1 Primero busquemos las claves candidatas. Dado que $A \to D$, $C \to BD \models C \to B \wedge C \to D$ y $D \to ABC \models D \to A \wedge D \to B \wedge D \to C$ por lo que las claves candidatas son: $\{\{A\},\{C\},\{D\}$. Escojamos A como clave primaria. Ahora para que esté en 2FN tenemos que fijarnos en los atributos no primos de R, en este caso es sólo B, este atributo tiene que depender de manera total de la clave primaria de R (A). En este caso así es, dado que por reglas de inferencia $A \to B$. Luego la relación R está en 2FN. Veamos si está en 3FN. Vemos que la primera condición se cumple, puesto que para toda df en F de la forma $X \to Y$ se tiene que X es superclave, luego está en 3FN. Veamos si está en FNBC. Para esto se tiene que cumplir la misma condición en este caso, lo que nos dice que además está en FNBC.

2.4 Clave primaria A (puesto que A genera a todos los demás atributos). Primero busquemos una cobertura mínima G para F. Claramente es la dependencia funcional $A \to C$ puesto que de ella se pueden desprender todas las demás. Ahora aplicamos el algoritmo que aparece en Elmasri (capítulo 13, algoritmo 13.3 o 13.1). Con esto nos queda: $R_1(A,C), R_2(B,D)$. Este esquema conserva las dependencias funcionales y además es del tipo reunión sin pérdidas (verificar usando el algoritmo 13.2). Con esto lo tenemos en 3FN. Además si comprobamos nos daremos cuenta que además es del tipo FNBC. También podemos usar el algoritmo En este caso los pasos del algoritmo resulta en

- $R(A,B,C,D) \Rightarrow$ escogemos R y la df $B \to D$
- $R_1(A,B,C), R_2(B,D)$ escogemos R_1 y la df $C \to B$
- $R_1(A,B), R_2(B,D), R_3(B,C)$ fin, puesto que ya no queda ningún esquema que no esté en FNBC.

2.6 Primero apliquemos el algoritmo para dejarlo en 3FN. Primero necesitamos una cobertura mínima G para F. La cobertura mínima puede ser $\{\{AB \to D\},\{E \to C\}\}$. Ahora aplicando el algoritmo tendremos: $R_1(AB,D), R_2(E,C)$. Ahora apliquemos el algoritmo para FNBC, los pasos serían:

- $R(A,B,C,D,E)$ escogemos R y la df $D \to B$.
- $R_1(A,C,D,E), R_2(D,B)$. Fin.

2.8 [2]

a) Veamos las claves candidatas. $A \to B$ y $B \to D$ luego $A \to D$, clave candidata posible A. $A \to C$. De $A \to C$ y $A \to D$ se tiene que $A \to CD$ además $CD \to FG$ luego $A \to F$ y $A \to G$. También $G \to E$ luego claramente A es clave candidata. Pero tenemos que $E \to A$ y $A \to E$ luego las claves son $\{\{A\},\{E\}\}$.

b) No cumple 3FN ni FNBC.

c) R' está en FNBC y R'' no está ni en 3FN ni en FNBC.

[2] Notar que en la guía está resuelto el 2.8 pero hay un error porque en realidad está resuelto el 2.9 :P.