Pauta C1-P2

Enunciado Considere el esquema de relación $R(A, B, C, D, E, F, G, H)$ y el conjunto de dependencias funcionales $F = \{AB \to C, A \to DE, B \to FC, F \to GH\}$.

Considere las descomposiciones:

(a) $R_1(B, C), R_2(A, D, E), R_3(B, F), R_4(F, G, H)$

(b) $R'_1(A, B, C, D, E), R'_2(B, F, G, H)$

Determine para (a) y (b):

1. Si la descomposició preserva o no dependencias funcionales.
2. Si tiene o no pédida de informació.
3. ¿Cuál descomposición es preferible? Argumente.
4. Encuentre una descomposición de (R, F) en 3FN.
5. Encuentre una descomposición de (R, F) en FNBC. ¿Qué desventaja podría tener?

1. Preservación de dependencias Tenemos que ver que cada dependencia de la forma $X \to Y$ está contenida en una relación (i.e. $XY = \pi_{XY} R_k$, para algún R_k).

En rigor, tenemos que comprobar que $X \to Y$ **se preserve incluso a través de transitividad**. Por lo anterior, nosotros trataremos de no preocuparnos de las dependencias que se vayan a deducir; nos preocuparemos sólo de las dependencias con las que contamos. **También, si $Y = Y_1 Y_2$, e Y_1, Y_2 son atributos, deberiamos examinar $X \to Y_1$ y $X \to Y_2$ por separado** cuando no encontremos $X \to Y$ en una primera inspección.

Para (a), vemos que $AB \to C$ no se preserva, PERO se deduce de $B \to FC$. Y $B \to FC$ se preserva a través de R_1 ($B \to C$) y R_3 ($B \to F$). En cuanto a $A \to DE$, vemos que se preserva en R_2. Y $F \to GH$ se preserva en R_4.

Para (b), en R_1' vemos que se preservan $AB \to C$, $B \to C$ y $A \to DE$. Y en R_2' se preservan $B \to F$ y $F \to GH$.

En conclusión, **ambas descomposiciones preservan las dependencias funcionales**.

2. Pérdida de información En términos generales, una descomposición sin pérdida de información es aquella que no crea ni destruya tuplas respecto a la relación original.

En términos de álgebra relacional, una descomposición no tiene pérdida cuando la original se reconstruye a través de inner joins o joins naturales.

Basándonos en la observación anterior, una descomposición sin pérdidas se puede analizar desde las dependencias funcionales preservadas y los atributos comunes entre las relaciones. **Debemos ser capaces de determinar las relaciones a través de los atributos comunes y las dependencias que se cumplen.**

Lo anterior permite deducir el algoritmo visto en clases (la matriz). También nos permite ver que, si no se preserva ninguna llave candidato, no podemos reconstruir sin pérdidas; **una llave candidato sólo se puede deducir a través de otra llave candidato (alternas), sino no se puede reconstruir**.

Al ejercicio: veamos que AB es llave. $A \to DE, B \to CFGH \Rightarrow AB \to R$. Si la llave no se preserva, nos ahorramos la tabla.

Para (a), la llave no se preserva: AB no se pueden encontrar juntos en una relación. Luego, **no preserva información**.

Para (b), la llave se preserva. Y el atributo común entre R_1' y R_2' es B. Luego, R_1' accede y determina R_2', determinando todos los atributos originales de R (esto es equivalente a completar una fila de la tabla). Luego, **preserva la información**.

3. ¿Cuál descomposición es preferible? La segunda descomposición es preferible, porque preserva dependencias e información. En otras palabras, minimizamos el riesgo de ingresar información inválida (las depedencias actúan como restricciones, idealmente) y de recuperar tuplas espurias (que no existían).

La primera descomposición falla en lo de de recuperar tuplas. Podemos recuperar tuplas espurias... De hecho, R_2 es disjunto del resto de las relaciones

(!), luego nos enfrentamos a un potencial generador de redundancia al hacer la reunión y tratar de recuperar R.

4. Descomponer en 3FN. Esto es bastante trivial... Buscamos el conjunto covertor mínimo. Bueno, es fácil ver que $AB \to C$ está de sobra ante $B \to FC$. Luego, nuestro F se reduce a

$$F = A \to DE, B \to CFGH$$

y la llave es AB. Luego

$$R_1(A, B), R_2(A, D, E) R_3(B, C, F, G, H)$$

cumple 3FN.

Validez: en R_1 la llave es AB. La llave de R_2 es A, y la única dependencia válida ahí es $A \to DE$. La llave de R_3 es B, y la única dependencia válida es $B \to CFGH$. Luego, esta descomposición está en 3FN.

Ojo que la descomposición presentada no es única. Por ejemplo,

$$R_1(A, B), R_2(A, D, E), R_3(B, C, F), R_4(B, G), R_5(B, H)$$

tambíen está en 3FN.

5. Descomponer en FNBC. En este problema, cualquier descomposición en 3FN estará en FNBC. Esto es así ya que no hay manera de que una dependencia funcional haya determinado atributos primos. Luego, toda dependencia aplicable partía de una (super)llave.

Recordatorio: 3FN y FNBC. Las definiciones son:

(3FN) Para toda $X \to Y$ válida en R, si
- X es superllave, o
- Y es atributo primo,

diremos que R está en tercera forma normal.

(FNBC) Para toda $X \to Y$ válida en R, si
- X es superllave,

diremos que R está en forma normal de Boyce-Codd.