CC42A – BASES DE DATOS
Profesores: Claudio Gutiérrez, Gonzalo Navarro
Auxiliar: Mauricio Monsalve

GUÍA DE EJERCICIOS:

- **Dependencias funcionales y multivaluadas**
- **Formas normales (1FN hasta 4FN)**
- **Normalización (3FN y FNBC)**

1. 4FN se define como lo siguiente: Para toda dependencia multivaluada $X \twoheadrightarrow Y$, o bien la dependencia es trivial, o bien X es superllave. Responda (*):

i. ¿Puede haber una dependencia multivaluada en 4FN que no sea dependencia funcional?
ii. Pruebe que 4FN está en FNBC.
iii. Pruebe que una relación en FNBC y sin dependencias multivaluadas explícitamente especificadas, necesariamente está en 4FN.

2. Dado R(A,B,C,D), pruebe lo siguiente (*):

i. Si $A \rightarrow C$ y $C \rightarrow D$ entonces $A \twoheadrightarrow BC$.
ii. Regla de condensación: Sean $X \twoheadrightarrow Y$ y $Z \subseteq Y$, y existe W tal que $W \subseteq R$ y $W \cap Y = \emptyset$ y $W \rightarrow Z$, entonces $X \rightarrow Z$.

3. Se dice que A depende transitivamente de B si: $B \rightarrow C$, $C \rightarrow A$ pero no ocurre que $C \rightarrow B$ ni que $A \subseteq C$. Un esquema está en 3FN con respecto a F (un conjunto de dependencias funcionales) si para todo atributo no primo A (o sea, A no está en llave candidato alguna) se tiene que no dependa transitivamente de una llave candidato de R. Probar la equivalencia con la definición dada en clases ($X \rightarrow Y$: es trivial, X es superllave o Y es atributo primo).

4. Sea R(A,B,C,D,E) y $F=\{A \rightarrow B, A \rightarrow D, C \rightarrow E\}$. Normalice en 3FN y en FNBC (*).

5. Muestre que no siempre es posible lograr una normalización en FNBC conservando todas las dependencias funcionales. ¿Qué mal podría traer esto?

6. Explique por qué es mejor lograr 4FN que FNBC, por qué es mejor lograr FNBC que 3FN, por qué es mejor lograr 3FN que 2FN y por qué es mejor lograr 2FN que tan solo 1FN. Hint: Sólo basta una respuesta.

7. Obtenga un conjunto minimal que represente el siguiente conjunto de dependencias funcionales para R(A,B,C,D,E,F,G):

$$F = \{A \rightarrow B,\ BC \rightarrow FG,\ D \rightarrow AC,\ BDF \rightarrow AG,\ BE \rightarrow F,\ C \rightarrow D,\ DF \rightarrow G\}$$

8. A continuación se presenta una serie de afirmaciones. Indique cuál es falsa y cuál verdadera, y demuestre:

i. Cualquier relación binaria está en 3FN.
ii. Cualquier relación binaria está en FNBC.
iii. Cualquier relación binaria está en 4FN.
iv. $R(A,B,C) = R1(A,B) * R2(B,C)$ si y sólo si $B \rightarrow C$.

9. En la publicación <<Dependency Structures of Data Base Relationships>>, Proc. IFIP Congress 1974, W. W. Armstrong formalizó por primera vez la teoría de las dependencias funcionales. Además presentó estos axiomas:

1. $A_1A_2 \ldots A_m \rightarrow A_i, \ \forall i \in \{1,2,3,\ldots,m\}$
2. $A_1A_2 \ldots A_m \rightarrow B_1B_2 \ldots B_r$ si y sólo si $A_1A_2 \ldots A_m \rightarrow B_i, \ \forall i \in \{1,2,\ldots,r\}$
3. Si $A_1A_2 \ldots A_m \rightarrow B_1B_2 \ldots B_r$ y $B_1B_2 \ldots B_r \rightarrow C_1C_2 \ldots C_p$, entonces se tiene que $A_1A_2 \ldots A_m \rightarrow C_1C_2 \ldots C_p$.

El ejercicio es probar que estos axiomas son equivalentes a los usados actualmente.

Solución a los ejercicios con (*):

1. - No, porque toda dependencia ocurre con respecto a la superllave, y una superllave determina funcionalmente cualquier atributo.
   - Basta ver que cualquier dependencia funcional no trivial cumple con ser de una superllave a otro atributo.
   - Esto ocurre porque toda dependencia funcional proviene de $X \rightarrow Y$, con X superllave, luego se deduce inmediatamente $X \twoheadrightarrow Y$ y $X \twoheadrightarrow R\text{-}X\text{-}Y$, pero en ambas dependencias multivaluadas el determinante es una superllave. Luego cualquier deducción de dependencias multivaluadas cumple con estar en 4FN.

2. - Bueno, veo que $A \rightarrow AC \rightarrow ACD$, entonces $A \rightarrow D$. Como $A \rightarrow D$ entonces $A \twoheadrightarrow D$ y por ende $A \twoheadrightarrow R\text{-}A\text{-}D$, o sea, $A \twoheadrightarrow BC$.
   - Usamos la definición de dependencia multivaluada. Entonces si tenemos t_1 y t_2 instancias de R tales que $t_1[X]=t_2[X]$, entonces se cumple que existe t_3 tal que $t_3[X]=t_1[X]$, $t_3[Y]=t_1[Y]$ y $t_3[R\text{-}X\text{-}Y]=t_2[R\text{-}X\text{-}Y]$. Sin embargo podemos modificar lo anterior sin miedo de fallo $t_3[R\text{-}Y]=t_2[R\text{-}Y]$, esto porque X no agrega información extra a las tuplas (ya se sabe el valor de X). Pero se tiene que W es disjunto de Y, luego W está en R-Y, se tiene que Z está en Y y también se tiene que $W \rightarrow Z$. Por lo tanto se tiene que $t_3[R\text{-}Y]=t_2[R\text{-}Y] \Rightarrow t_3[W]=t_2[W]$, $t_3[Y]=t_1[Y] \Rightarrow t_3[Z]=t_1[Z]$, pero $W \rightarrow Z$ por lo que $t_3[W]=t_2[W]$ y $t_3[Z]=t_1[Z] \Rightarrow t_2[Z]=t_3[Z]=t_1[Z] \Rightarrow X \rightarrow Z$.

3. Buscamos la llave. Como A y C no son determinados de forma no trivial por nadie, deben ser parte de toda llave candidato. Pero revisando, $AC \rightarrow ABCDE$, o sea, AC es llave candidato. Como ninguna dependencia funcional es del tipo trivial, superllave a algo o algo a atributo primo, entonces no está ni en FNBC ni en 3FN. Normalizamos. Primero se reduce $A \rightarrow B$ y $A \rightarrow D$ a $A \rightarrow BD$.
   FNBC: ABCDE ➔ ABD (fnbc) + ACE ➔ ABD (fnbc) + CE (fnbc) + AC (fnbc)
   3FN: ABD + CE + AC (más directo, y en este caso dio el mismo resultado que la descomposición de FNBC).