Fundamentos de bases de datos Tarea 5 Dependencias y Normalización

Díaz Gómez Silvia Eugenio Aceves Narciso Isaac Quiroz Castañeda Edgar

26 de Abril del 2019

1. Preguntas de repaso

- a. ¿Qué es una dependencia funcional y cómo se define? Es una relación entre un conjunto de atributos X y otro Y, denotada como X → Y, donde a cada posible valor de Y se le asocia un único valor de Y. También se les llama propiedades semánticas. Por otra parte, se crean a partir de las 'reglas del negocio' y características concidas de las relaciones (semántica) ó reglas de inferencia.
- b. ¿Para qué sirve el concepto de **dependencia** en la normalización? Sirve para especificar restricciones entre los atributos de la relación, y determinar si son legales bajo esas restricciones. Además es en base a estas dependencias funcionales que se define el concepto de llave y nos permite identificar tablas que potencialmente podrían introducir anomalías (redundancia o pérdida de integridad) que es de donde vienen las formas normales que se y la normalización de relaciones.
- c. Sea A la llave de R(A, B, C). Indica **todas** las dependencias funcionales que implica **A**. $\{A \to A, A \to B, A \to C, A \to AB, A \to AC, A \to BC, A \to ABC\}$
- d. ¿Qué es una forma normal? ¿Cuál es el objetivo de normalizar un modelo de datos?

 Una relación está en forma normal si cumple una serie de restricciones respecto a su esquema.

 Cada regla aumenta el grado de normalización.

El propósito de esto es eliminar redundancia y garantizar "JOIN" sin pérdida, todo esto sin perder dependencias funcionales sen el proceso.

- e. ¿En qué casos es preferible lograr 3NF en vez de BCNF?
 - La **BCNF** elimina toda la redundancia de la base de datos, pero se pueden perder dependencias funcionales en el proceso. En un caso real, si se pierden las dependencias funcionales entonces la base de datos ya no estaría modelando fielmente la realidad del caso de uso, por lo que no es aceptable perderlas.
 - La 3NF no elimina la redundancia completamente, pero mantiene todas las dependencias funcionales.

Entonces, si al usar **BCNF** se pierden dependencias funcionales, es necesarios usar una forma de normalización menos estricta, entiéndase **3NF**.

- 2. Proporciona algunos ejemplos que demuestren que las siguientes reglas no son válidas:
 - a) Si $\mathbf{A} \to \mathbf{B}$, entonces $\mathbf{B} \to \mathbf{A}$

En un sistema de registro de vehiculos y conductores (particulares), si hay un auto, es necesario que le asigne un propietario (quien pagará multas y se hará responsable) y entonces a todo auto 'A' le asigno una persona con licencia 'B'. No obstante, no todas las personas con licencia estan obligadas a tener un auto propio y entonces la segunda dependencia funcional realmente no existe/se cumple.

b) Si $AB \rightarrow C$, entonces $A \rightarrow C y B \rightarrow C$

Podemos tomar 'A' como el dinero que tiene una persona, 'B' como el número de lugares disponibles en sus cines asociados (digamos, los más cercanos de la franquicia Mexipolis) y 'C' un atributo que nos indica si la persona ha visto o no la película "DC:Los Vendedores, Jugada Final" porque queremos saber si podemos hablar de spoilers con la persona.

La primera regla tiene sentido (El que tengas dinero suficiente para un boleto y que haya lugares en el cine pueden determinar si vas a ver la película/ya la viste) pero ni $A \to C$ se cumple, ya que el cine podría estar

lleno y no importaría que tienes dinero, ni $B \to C$, ya que puede haber lugares pero sin dinero para boletos no puedes ir al cine. Por lo tanto ninguna de las dos puede determinar realmente si ya viste la pelicula (a ambas les falta información para eso) y no existen esas dependencias funcionales.

c) Si $\mathbf{A} \rightarrow \mathbf{C}$, entonces $\mathbf{A} \rightarrow \mathbf{C}$

Digamos que 'A' es el identificador de un equipo de tareas en FBD y 'C' es la calificación de una tarea. Entonces podemos tener una tabla con (1,5T1),(1,6T2),(2,10T2) que representa que el equipo 1 ha entregado dos tareas con 5 y 6 de calificación mientras que el equipo 2 sólo entregó la tarea 2 pero sacó 10 en ella.

La primera representa una DMV (Dependencia Multivaluada). Si se cumpliera lo segundo (\rightarrow) , estaríamos obligando a que a un equipo siempre se le asigne la misma claifiación/tarea, i.e. (1,10T2),(1,7T3) no sería válido porque a cada valor de 'A' le debe corresponder un único valor en 'C', pero esto no es lo que se planteaba la primera de las reglas originalmente.

- 3. Para cada uno de los esquemas que se muestran a continuación:
 - a) R(A,B,C,D,E) con $F=\{AB \rightarrow CD, E \rightarrow C, D \rightarrow B \}$
 - b) R(A,B,C,D,E) con $F={AB \rightarrow C, DE \rightarrow C, B \rightarrow D}$
 - Especifica de ser posible dos **DF** no triviales que se puedan derivar de las dependencias funcionales dadas. Usando la cerradura de atributos.
 - a) $\{AD\}$ += $\{ADBC\}$ y $\{AE\}$ += $\{AEC\}$ a partir de estas cerraduras obtenemos las siguientes dependencias: $AD \rightarrow BC$ y $AE \rightarrow C$
 - b) $\{AB\}+=\{ABCD\}$ y $\{BE\}+=\{EBDC\}$ a partir de estas cerraduras obtenemos las siguientes dependencias: $AB \to CD$ y $BE \to DC$
 - Indica una llave candidata para R

Usamos la cerradura para encontrar una llave candidata

- a) $\{AB\}+=\{ABCD\}, \{E\}+=\{EC\}, \{D\}+=\{DB\}, la cerradura de AB es la que contiene mas elementos de la relación por lo tanto una llave para R sería :$ **ABE**
- b) $\{AB\}+=\{ABCD\}, \{DE\}+=\{E\}, \{B\}+=\{BD\}, la cerradura de AB es la que contiene mas elementos de la relación por lo tanto una llave para R sería :$ **ABE**
- Especifica todas las violaciones a la BCNF
 - a) Las tres dependencias son violaciones a BCNF porque no aparece del lado izquierdo de las DF la llave que es ABE.
 - b) Las tres dependencias son violaciones a BCNF porque no aparece del lado izquierdo de las DF la llave que es ABE.
- Normaliza de acuerdo a BCNF, asegúrate de indicar cuáles son las relaciones resultantes con sus respectivas dependencias funcionales.
 - a) Como todas las DF son violaciones, tomamos a AB \rightarrow CD y su cerradura es {AB}+= {ABCD}. Así que definimos dos nuevas relaciones S y T,

$$S(A,B,C,D)$$
 con $\{AB \to CD, D \to B\}$

T(A,B,E) con ABE \rightarrow ABE y perdemos $E \rightarrow C$

En S la llave es AB por lo tanto D \rightarrow B es una violación.

Ahora para S tomamos $D \to B$ y calculamos la cerradura para $\{D\} + = \{DB\}$

Definimos otras dos nuevas relaciones

 $U(D,B) \text{ con } D \to B$

V(D,A,C) con DAC \rightarrow DAC

En esta nueva partición se pierde AB \rightarrow CD.

Observamos que en U y V ya no se tienen violaciones, por lo tanto el esquema en BCNF para R es:

 $T(A,B,E) \text{ con } ABE \rightarrow ABE$

 $U(D,B) con D \rightarrow B$

V(D,A,C) con $DAC \rightarrow DAC$

b) Como todas las DF son violaciones, tomamos a $B \to D$ y su cerradura es $\{B\} += \{BD\}$. Así que definimos dos nuevas relaciones S y T,

 $S(B,D) con \{B \rightarrow D\}$

T(B,A,C,E) con $AB \rightarrow C$ y perdemos $DE \rightarrow C$

En S, B es llave por lo tanto S ya esta en BCNF.

Ahora en T la llave sigue siendo ABE por lo tanto AB \rightarrow C es violación, calculamos la cerradura para $\{AB\}+=\{ABC\}$

Definimos otras dos nuevas relaciones

 $U(A,B,C) \text{ con } AB \to C$

 $V(A,B,E) \text{ con ABE} \rightarrow ABE$

Observamos que en U y V ya no se tienen violaciones, por lo tanto el esquema en BCNF para R es:

 $S(B,D) con B \rightarrow D$

 $U(A,B,C) \text{ con } AB \rightarrow C$

 $V(A,B,E) \text{ con } ABE \rightarrow ABE$

- 4. Para cada una de las siguientes relaciones con su respectivo conjunto de dependencias funcionales:
 - a) R(A,B,C,D,E,F) con $F = \{B \rightarrow D, B \rightarrow E, D \rightarrow F, AB \rightarrow C\}$
 - b) R(A,B,C,D,E) con $F = \{A \rightarrow BC, B \rightarrow D, CD \rightarrow E, E \rightarrow A\}$
 - Indica todas las violaciones a la 3NF

Que no aparezca una llave candidata en el lazo izquierdo de las DF o que no aparezca a la derecha. Calculamos la cerradura:

- a) {B}+={BDE}, {D}+={DF}, {AB}+={ABCDE} una llave para R es **AB** Las dependencias que violan la 3NF son B \rightarrow D, B \rightarrow E, D \rightarrow F
- b) {A}+={ABCDE}, {B}+={B},{CD}+={CDEAB}, {E}+={EABCD} una llave para R es $\bf A$, $\bf E$ o $\bf DC$ B \rightarrow D viola la 3NF
- Normaliza de acuerdo a la 3NF

Para normalizar en 3NF se deben buscar superfluos por la izquierda y por la derecha.

a) • Superfluos por la izquierda:

En este caso la dependencia que tiene mas de un atributo por la izquierda es la DF que contiene a la llave del lado izquierdo por lo tanto no es necesario verificarlo.

• Superflos por la derecha:

Haciendo uso de la propiedad de la unión tenemos que F queda como F= $\{B \to DE, D \to F, AB \to C\}$ Tomamos la dependencia que se violacion a 3NF y tenga mas de un atributo a la derecha, $B \to DE$ y buscamos elementos superfluos:

¿D es superfluo? B \rightarrow E

obtenemos un nuevo conjunto de dependencias funcionales $F' = \{B \to E, D \to F, AB \to C\}$ y calculamos la cerradura para B

 $\{B\}+=\{BE\}$ como D no aparece por lo tanto D no es superfluo.

 $E = \text{E} \cdot \text{E}$

obtenemos un nuevo conjunto de dependencias funcionales $F' = \{B \to D, D \to F, AB \to C\}$ y calculamos la cerradura para B

{B}+={BDF} como E no aparece por lo tanto E no es superfluo.

Así que obtenemos que $F_{min} = \{B \to DE, D \to F, AB \to C\}$ a partir de este conjunto creamos una relación por cada DF

 $R_1(B,D,E) \text{ con } B \to DE$

 $R_2(D,F) \text{ con } D \to F$

 $R_3(A,B,C) \text{ con } AB \to C$

Como la llave esta contenida en la relación R₃ por lo tanto esta es la normalización para R en 3NF.

- b) Superfluos por la izquierda:
 - En este caso la dependencia que tiene mas de un atributo por la izquierda es $CD \to E$ y verificamos si algun atributo es superfluo.

 - {D}+={D}, E no aparece en la cerradura de D, por lo tanto C no es superfluo.
 - ¿D es superfluo? $C \to E$
 - $\{C\}+=\{C\}$, E no aparece en la cerradura de C, por lo tanto D no es superfluo.
 - Esta parte no era necesario verificarlo porque CD es llave candidata.
 - Superflos por la derecha:
 - Tomamos la dependencia que tenga mas de un atributo a la derecha, $A \to BC$ y buscamos elementos superfluos:
 - ¿B es superfluo? A \rightarrow C
 - obtenemos un nuevo conjunto de dependencias funcionales $F' = \{A \to C, B \to D, CD \to E, E \to A\}$ y calculamos la cerradura para A
 - $\{A\}+=\{AC\}$ como B no aparece por lo tanto B no es superfluo.
 - ¿C es superfluo? $A \rightarrow B$
 - obtenemos un nuevo conjunto de dependencias funcionales $F' = \{A \rightarrow B, B \rightarrow D, CD \rightarrow E, E \rightarrow A\}$ y calculamos la cerradura para A
 - $\{A\}+=\{ABD\}$ como C no aparece por lo tanto C no es superfluo.
 - Así que obtenemos que $\mathbf{F}_{min} = \{ A \to BC, B \to D, CD \to E, E \to A \}$ a partir de este conjunto creamos una relación por cada DF
 - $R_1(A,B,C) \text{ con } A \to BC$
 - $R_2(B,D) \text{ con } B \to D$
 - $R_3(C,D,E) \text{ con } CD \to E$
 - $R_4(E,A) \text{ con } E \to A$

Como la llave esta contenida en la relación R₃ por lo tanto esta es la normalización para R en 3NF.

5. Sea el esquema:

$$R(A,B,C,D,E,F)$$
 con $F=\{BD \rightarrow E, CD \rightarrow A, E \rightarrow C, B \rightarrow D\}$

- ¿Qué puedes decir de A+ v F+?
 - $\mathbf{A} += \{A\} \text{ y } \mathbf{F} += \{F\} \text{ No alcanzan ningun otro attributo.}$
- Calcula B+, ¿qué puedes decir de esta cerradura?
 - {B}+={BDECA} Casí contiene todos los atributos de la relación por lo tanto podemos agregarle el atributo F y sería una llave para la relación R.
- Obtén todas las llaves candidatas.
 - $\{BD\} + = \{BDECA\}, \{CD\} + = \{CDA\}, \{E\} + = \{EC\}, \{B\} + = \{BDECA\}$
 - Las llaves candidatas son: BF. Una llave debe ser mínima, y como BF es llave ninguna conjunto de tres o más atributos puede ser llave en este esquema. Luego, todas los de conjuntos de un sólo atributo no son llaves, pues ninguna dependencia alcanza a F, por lo que para que alcance a todos, debe tener explícitamente a F.
 - Luego, FA, FC y FD sólo tiene dependencias triviales, pues no forman parte de ninguna dependencia.
 - Y $\{EF\}+=EFC$, por lo que no es llave.
 - Entonces la única llave es BF.
- ¿R cumple con BCNF? ¿Cumple con 3NF? (en caso contrario normaliza)
 - R no cumple BCNF ni con 3NF.
 - La llave candidata es BF.
 - Normalización con BCNF
 - Todas las dependencias son violaciones a BCNF, tomamos BD \rightarrow E, así tenemos la partición como:
 - S(B,D,E) con $BD \to E$, (ya no tiene ninguna violación) T(B,D,A,C,F) con $\{B \to D, CD \to A\}$
 - y E \rightarrow C se pierde. en T las DF son violaciones por lo tanto hacemos una nueva partición,
 - U(B,D) con $B \to D$, (va no tiene ninguna violación)

V(B,A,C,F) con BACF \rightarrow BACF y perdemos CD \rightarrow A.

Nuestro esquema normalizado con BCNF queda de la siguiente manera:

 $S(B,D,E) \text{ con } BD \to E,$

 $U(B,D) \text{ con } B \to D,$

V(B,A,C,F) con BACF \rightarrow BACF

• Normalización con 3NF

Todas las dependecias son violaciones a 3NF

o Superfluos por la izquierda

Tomamos a BD \rightarrow E,

B? D \rightarrow E,

 $\{D\}+=\{D\}$, E no aparece en la cerradura de D, por lo tanto B no es superfluo. ¿D? B \rightarrow E,

{B}+={BDECA}, E aparece en la cerradura de B, por lo tanto D es superfluo.

Entonces $\mathbf{F}_{min} = \{B \to E, CD \to A, E \to C, B \to D\}$ y por la propiedad de la unión $\mathbf{F}_{min} = \{B \to DE, CD \to A, E \to C\}$

o Superfluos por la derecha:

Tomamos la dependencia que tenga mas de un atributo a la derecha, en este caso es $B \to DE$ y buscamos elementos superfluos:

¿D es superfluo? B \rightarrow E

obtenemos un nuevo conjunto de dependencias funcionales $F' = \{B \to E, CD \to A, E \to C\}$ y calculamos la cerradura para B

 $\{B\}+=\{BEC\}$ como D no aparece por lo tanto D no es superfluo.

¿E es superfluo? B \rightarrow D

obtenemos un nuevo conjunto de dependencias funcionales $F' = \{B \to D, CD \to A, E \to C\}$ y calculamos la cerradura para B

{B}+={BD} como E no aparece por lo tanto E no es superfluo.

Así que obtenemos que $\mathbf{F}_{min} = \{ B \to DE, CD \to A, E \to C \}$ a partir de este conjunto creamos una relación por cada DF

 $R_1(B,D,E) \text{ con } B \to DE$

 $R_2(C,D,A)$ con $CD \to A$

 $R_3(E,C) \text{ con } E \to C$

Y como la llave no esta en ninguna relación creamos una nueva relación que la contenga $R_4(B,F)$ con $BF \to BF$.

Por lo tanto R₁, R₂, R₃ y R₄ es la normalización en 3NF.

- Se ha decidido dividir \mathbf{R} en las siguientes relaciones $\mathbf{S}(\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D},\mathbf{F})$ y $\mathbf{T}(\mathbf{C},\mathbf{E})$, ¿se puede recuperar la información de \mathbf{R} ? No podemos recuperar toda la información de \mathbf{R} porque se pierde la dependencia $\mathbf{B}\mathbf{D} \to \mathbf{E}$
- 6. Para cada uno de los esquemas, con su respectivo conjunto de dependencias multivaluadas, resuelve los siguientes puntos:
 - a) R(A,B,C,D) con $DMV = AB \rightarrow C, B \rightarrow D$
 - b) R(A,B,C,D,E) con $DMV = A \rightarrow B$, $AB \rightarrow C$, $A \rightarrow D$, $AB \rightarrow E$
 - Encuentra todas las violaciones a la 4NF
 - a) R(A,B,C,D) con $DMV = AB \rightarrow C, B \rightarrow D$

Consideremos las posibles llaves calculando las cerraduras.

$${B}+={BD}$$

Por lo que una llave puede ser **ABC**.

Como ninguna de las dependencias tiene a esta llave en su parte izquierda, entonces toda **DMV** son violaciones a la cuarta forma normal.

b) R(A,B,C,D,E) con $DMV = A \rightarrow B$, $AB \rightarrow C$, $A \rightarrow D$, $AB \rightarrow E$

Consideremos las posibles llaves calculando las cerraduras.

$${AB} + = {ABCDE}$$
$${A} + = {AD}$$

Por lo que una llave es AB.

Por lo que las violaciones son $A \rightarrow B y A \rightarrow D$

- Normaliza de acuerdo a la 4NF
 - a) R(A,B,C,D) con $DMV = AB \rightarrow C, B \rightarrow D$

Para intentar preservar las dependencias multivaluadas, primero empezemos seleccionando una violación que sea dependencia funcional, que es este caso sólo puede ser $\mathbf{B} \to \mathbf{D}$. Entonces, \mathbf{R} se parte en dos nuevas tablas $R_1(B,D)$ y $R_2(A,B,C)$.

En R_1 se preserva únicamente la dependencia funcional $\mathbf{B} \to \mathbf{D}$, que como incluye a todos los atributos de la relación es trivial, por lo que R_1 ya está en 4NF. En R_2 , sólo se preserva una dependencia multivaluada, pero esta incluye a todos los elementos de R_2 , por lo que no es una violación a la **4NF**. Por lo que R_2 también ya está normalizada.

Ahora, hay que revisar si la llave está contenida en alguna de las relaciones. Efectivamente, la llave está en R_2 , por lo que no es necesario agregar ninguna relación.

Entonces R_1 y R_2 son la normalización en $\mathbf{4NF}$ de \mathbf{R} .

b) R(A,B,C,D,E) con $DMV = A \rightarrow B$, $AB \rightarrow C$, $A \rightarrow D$, $AB \rightarrow E$

Para intentar preservar las dependencias multivaluadas, primero empezamos seleccionando una violación que sea dependencia funcional, que es este caso sólo puede ser $A \to D$.

Entonces, **R** se parte en dos nuevas tablas $R_1(A, D)$ y $R_2(A, B, C, E)$.

En R_1 se preserva únicamente la dependencia funcional $\mathbf{A} \to \mathbf{D}$, que como incluye a todos los atributos de la relación es trivial, por lo que R_1 ya está en 4NF.

Mientras que en R_2 se siguen manteniendo todas las demás dependencias multivaluadas.

Por esto, se sigue teniendo que AB sigue siendo llave y por lo tanto $A \rightarrow B$ sigue siendo violación de 4NF, y de hecho la única.

Entonces, tomándola como violación, R_2 se parte en dos nuevas tablas $R_3(A, B)$ y $R_4(A, C, E)$.

En R_3 se preserva únicamente A \rightarrow B, que como tiene a todos los atributos de la relación. no es violación de la **4NF**, por lo que R_3 ya está normalizada.

En R_4 no se preserva ninguna dependencia, por lo que las únicas presentes son las triviales, por lo que R_4 también ya está en 4NF.

Notemos que en este último paso se perdieron las dependencias funcionales $\mathbf{AB} \to \mathbf{E}$ y $\mathbf{AB} \to \mathbf{C}$. Por último, como la llave original está contenida en R_3 , no es necesario agregar ninguna relación.

Por lo que R_1 , R_3 y R_4 son la normalización en **4NF** de R.

7. Se tiene la siguiente relación:

R(idEnfermo, idCirujano, fechaCirugía, nombreEnfermo, direcciónEnfermo, nombreCirujano, nombreCirugía, medicinaSuministrada, efectosSecundarios)

• Expresa las siguientes restricciones en forma de **dependencias funcionales**:

A un enfermo sólo se le da una medicina después de la operación. Si existen efectos secundarios estos dependen sólo de la medicina suministrada. Sólo puede existir un efecto secundario por medicamento.

Sea $\mathbf{R}(A,B,C,D,E,F,G,H,I)$, cada letra el atributo que le corresponde de acuerdo a su posición y la definición de \mathbf{R} .

- a) $A, C, G \rightarrow H$
- b) $H \rightarrow I$
- Especifica otras dependencias funcionales o multivaluadas que deban satisfacerse en la relación R. Por cada una que definas, deberá aparecer un enunciado en español como en el inciso anterior. Cada cirugía tiene exactamente a un cirujano registrado como a cargo de esta. Un cirujano puede haber estado a cargo de varias cirujias distintas. Un enfermo puede haber tenido varias cirujias. En cada cirujía se trata a un solo enfermo.

- a) $C, G \rightarrow B, F$
- b) B,F \rightarrow C,G
- $c) A \rightarrow C,G$
- d) C,G \rightarrow A
- Normaliza utilizando el conjunto de dependencias establecido en los puntos anteriores.

Para normalizar usaré 4FN (DMV no dejan opción). Primero hay que buscar una llave candidata a partir de la cerradura. En este caso, $\{A,C,G,E,D\}$ funciona. Ahora buscamos violaciones (todas las demás reglas lo son) y tomamos $H \to I$, por ser no-DMV, con lo que nos quedan dos relaciones $R_1(H,I)$ y $R_2(A,B,C,D,E,F,G)$. R_1 ya esta en 4FN, así que seguirmos sobre $R_2...$

Podemos considerar $\{C,G,E,D\}$ como la llave, buscamos violaciones de preferencia no-DMV y encontramos $C,G\to B,F$ y ahora tenemos $R_3(B,C,G,F)$ y $R_4(A,E,D)$. R_4 ya esta listo, ahora sólo queda R_3 ...

Esta vez usamos la DMV B, FC, G y nos quedan $R_5(B,C,G,F)$ y $R_6(B,F)$ y ahora terminamos. Como un pequeño recap, las dependencias que son válidas en cada relación (tenemos 4) son:

- En R_1 , $H \to I$. Esta tabla tiene medicamentos y efectos secundarios.
- En R_4 , $A, E, D \rightarrow A, E, D$. Esta tabla tiene idEnfermo, nombreEnfermo y direcciónEnfermo
- \bullet En R_5 , $C, G \to B, F$. Esta tabla tiene fecha
Cirujía, nombreCirujía, idCirujano, nombreCirujano
- En R_6 , $B, F \to B, F$. Esta tabla tiene idCirujano, nombreCirujano.

Entonces terminamos con 4 tablas que podríamos llamar Enfermo, Cirujía y Medicamento/Efectos. Suena bien.