

Auxiliar 5/5/2008: Dependencias funcionales

Qué es una dependencia funcional

¿Cuál es la idea tras las dependencias funcionales?

Una dependencia funcional es una relación entre atributos, en el “cual uno es determinado por el otro.” Lo que se quiere decir, es que existe una “función” en el “mundo” que asocia un dato a otro. Por ejemplo,

$$Y = f(X)$$

representa una dependencia funcional, donde Y es determinado por X . En particular, el grafo de la función es

$$(x, f(x))$$

y la función se expresa como

$$f : X \rightarrow Y$$

De manera similar a las funciones, anotamos una dependencia funcional como $A \rightarrow B$, donde el atributo A determina al atributo B . Habrá una función en el mundo que permite obtener B en función de A , pero no siempre la sabemos analíticamente. Pero bien podremos saber que existe, y que implícitamente $A \rightarrow B \Rightarrow B = f(A)$, con $f : A \rightarrow B$ la función implícita.

La definición formal de una dependencia funcional

Sean X, Y dos conjuntos de atributos de un esquema R . Entonces, $X \rightarrow Y$ en R se define como

$$\forall s, t \in r(R), s[X] = t[X] \Rightarrow s[Y] = t[Y]$$

donde la igualdad de atributos de tuplas se define de la manera usual, o sea, $s[X] = t[X] \Leftrightarrow (\forall i) s[X_i] = t[X_i]$ (igualdad atributo a atributo).

Axiomas de Armstrong

No son los originales, pero son simples y equivalentes:

1. Dependencia trivial/reflexividad: $X \rightarrow Y \Rightarrow X \rightarrow y \subseteq Y$
2. Aumentación: $X \rightarrow Y \Rightarrow XZ \rightarrow YZ$
3. Transitividad: $X \rightarrow Y \wedge Y \rightarrow Z \Rightarrow X \rightarrow Z$

Estos axiomas son sencillos de probar usando la definición de dependencia funcional.

1 Llaves candidato

Problema. Sea $R(A, B, C, D, E)$ y sea $F = \{A \rightarrow BC, CD \rightarrow E, B \rightarrow D, E \rightarrow A\}$. Encuentre las llaves candidato.

Solucion. Computaremos $A+$, $B+$, etc. paso a paso y trataremos de obtener todas las llaves candidato, esto es, las llaves minimales (irreducibles). Entonces:

$A+$: Partamos con $A \rightarrow BC \in F$:

- (1) $A \rightarrow BC \Rightarrow A \cup A \rightarrow BC \cup A \Rightarrow A \rightarrow ABC$ ($A \rightarrow BC \in F \oplus$ aumentación)
- (2) $A \rightarrow ABC \wedge B \rightarrow D \Rightarrow A \rightarrow ABC \wedge ABC \cup B \rightarrow ABC \cup D$ (aumentación)
 $\Rightarrow A \rightarrow ABC \wedge ABC \rightarrow ABCD \Rightarrow A \rightarrow ABCD$ (transitividad)
- (3) $A \rightarrow ABCD \wedge CD \rightarrow E \Rightarrow A \rightarrow ABCD \wedge ABCD \cup CD \rightarrow ABCD \cup E$ (aumentación)
 $\Rightarrow A \rightarrow ABCD \wedge ABCD \rightarrow ABCDE \Rightarrow A \rightarrow ABCDE$ (transitividad)
- (4) Como $A \rightarrow ABCDE$, A determina toda la relación R (podríamos anotar $A \rightarrow R$). Luego A es llave. Y es llave minimal al ser irreducible, luego es llave candidato.

(Nota) Llave candidato = Llave candidata = Clave candidato = Clave candidata = Llave alterna = Clave alterna ... Hay muchas notaciones alternativas para el mismo concepto.

$B+$: Usaremos una costumbre de notación algo diferente de ahora en adelante. Si tenemos $X \rightarrow Y$, anotaremos inmediatamente $X \rightarrow XY$ (esto es válido por aumentación: $X \rightarrow Y \Rightarrow X \cup X \rightarrow Y \cup X \Rightarrow X \rightarrow XY$). Y si tenemos $X \rightarrow XY$ y $U \rightarrow Z$ ($U \subseteq XY$), entonces anotaremos inmediatamente $X \rightarrow XYZ$ (aumentación: $U \rightarrow Z \Rightarrow U \cup XY \rightarrow Z \cup XY \Rightarrow XY \rightarrow XYZ \oplus$ transitividad: $X \rightarrow XY \wedge XY \rightarrow XYZ \Rightarrow X \rightarrow XYZ$). Observe que si una dependencia funcional es incluida en el determinado (Y , si pensamos en $X \rightarrow Y$), nunca más es necesario volver a incluirla ya que no añadirá información.

- (1) $B \rightarrow BD$ (tercera regla en F)
- (2) Terminamos. Ninguna dependencia funcional nos sirve.

$C+$: $C \rightarrow C$ y concluimos. Nada más por hacer.

$D+$: $D \rightarrow D$, y concluimos.

$E+$: Tenemos una dependencia en F que nos sirve:

- (1) $E \rightarrow AE$ ($E \rightarrow A \in F$)
- (2) $E \rightarrow AE \rightarrow ABCDE$ (pues $A \rightarrow ABCDE$)
- (3) Concluimos que E es llave candidato.

$BC+$: No revisamos combinaciones de atributos de la forma AX pues A es llave y AX no sería llave minimal ($A \neq X$).

- (1) $BC \rightarrow BCD$ ($B \rightarrow D \in F$)
- (2) $BC \rightarrow BCD \rightarrow BCDE$ ($CD \rightarrow E \in F$)
- (3) $BC \rightarrow BCD \rightarrow BCDE \rightarrow ABCDE$ (por $E \rightarrow A \in F$ o porque $E \rightarrow R$ al ser llave candidato)
- (4) Luego BC es llave candidato.

$BD+$: Proseguimos igual que antes.

- (1) $BD \rightarrow BD$ ($B \rightarrow D \in F$ o dependencia trivial).
- (2) Terminamos.

$CD+$: Como siempre.

- (1) $CD \rightarrow CDE$ ($CD \rightarrow E \in F$).
- (2) $CD \rightarrow CDE \rightarrow ABCDE$ (E es llave candidato).
- (3) Luego CD es llave candidato.

Terminamos: No es posible crear grupos más grandes de atributos sin crear superllaves. Entonces, A , E , BC y CD son llaves candidato.

2 Equivalencia de conjuntos de dependencias

Problema. Sea la relación $R(A, B, C, D)$ y los siguientes conjuntos de dependencias funcionales:

$$F = \{AC \rightarrow D, BCD \rightarrow AC, B \rightarrow C\}$$

$$G = \{AB \rightarrow D, BD \rightarrow A, B \rightarrow BC\}$$

$$H = \{AB \rightarrow D, BD \rightarrow A, B \rightarrow BC, C \rightarrow BC\}$$

Con esto pruebe que:

1. $G+ \subset F+$
2. $F+ \subset H+$
3. ¿Qué se puede decir de H y G ?

Solución.

(PARTE 1) Primero probemos que $G+ \subset F+$. Básicamente hay que probar que las dependencias de G son generadas a partir de las dependencias de F , y que al revés esto no es cierto (pues tenemos $\subset$ no $\subseteq$).

Probemos que F produce $AB \rightarrow D \in G$. $AB+$ en F es: $AB \rightarrow ABC (B \rightarrow C \in F) \rightarrow ABCD = R(AC \rightarrow D \in F)$. Como $AB \rightarrow ABCD \Rightarrow AB \rightarrow D$. Pero prosigamos... si por casualidad $AB+$ es distinto en F y G , $F+ \neq G+$. (Si logramos $G+ \subseteq F+$ y $G+ \neq F+$, entonces tenemos $G+ \subset F+$.) $AB+$ no puede aumentar en F porque es llave: $AB \rightarrow ABCD = R$. Ahora, en $G+$: $AB \rightarrow ABC (B \rightarrow C) \rightarrow ABCD = R(AB \rightarrow D)$. Luego, $AB+$ es igual en $F+$ y $G+$.

Probemos que F produce $BD \rightarrow A \in G$. $BD+$ en F es: $BD \rightarrow BCD (B \rightarrow C) \rightarrow ABCD = R(BCD \rightarrow A)$. Así que $BD \rightarrow A \in G$ también es producido por F (está en $F+$). Ahora, $BD+$ en G es: $BD \rightarrow ABD (BD \rightarrow A) \rightarrow ABCD = R(B \rightarrow BC)$. Luego, $BD+$ es igual en $F+$ y $G+$.

Probemos que F produce $B \rightarrow BC \in G$. Trivial, $B \rightarrow C \in F$: $B \rightarrow BC$. También es fácil ver que $B+$ es idéntico para $F+$ y $G+$.

Probemos que G no produce F . Ya probamos que $G \subseteq F+$, luego $G+ \subseteq F+$ (esta afirmación se puede probar rápidamente por contradicción). Lo que nos queda es ver que no todas las reglas de F se pueden obtener a partir de G . Para eso, veamos $AC \rightarrow D \in F$ y veamos si $AC \rightarrow D \in G+$; rápidamente vemos que en $G+$, $AC+$ sólo contempla dependencias triviales. Luego, $G+ \neq F+$.

Conclusión. Al ver que $G+ \subseteq F+$ y $G+ \neq F+$, tenemos que $G+ \subset F+$. $\square$

(PARTE 2) Ahora veamos que $F+ \subset H+$. Procederemos igual que antes, viendo que $F+ \subseteq H+$ y que $F+ \neq H+$.

Problemas que F no produce H . Comparemos $C+$. En el caso de $F+$: $C \rightarrow C$ (trivial). En el caso de $H+$: $C \rightarrow BC\dots$. No es necesario ir más lejos para ver que $C+$ es distinto en $F+$ y $H+$, luego $F+ \neq H+$. Debido a esto, sólo queda probar que $F+ \subseteq H+$.

Probemos que H produce $AC \rightarrow D \in F$. Veamos $AC+ \in H+$: $AC \rightarrow ABC(C \rightarrow BC) \rightarrow ABCD = R(AB \rightarrow D)$. En particular, hemos producido $AC \rightarrow D \in F$.

Probemos que H produce $BCD \rightarrow AC \in F$. Veamos $BCD+ \in H+$: $BCD \rightarrow ABCD = R(BD \rightarrow A)$. En particular, hemos producido $BCD \rightarrow AC \in F$.

Probemos que H produce $B \rightarrow C \in F$. Trivial, dado que $B \rightarrow BC \in H$ es una regla equivalente a $B \rightarrow C \in F$ ($B \rightarrow B$ es trivial).

Conclusión. Dado que $F+ \neq H+$ y $F+ \subseteq H+$, $F+ \subset H+$. $\square$

(PARTE 3) Sencillamente, de $G+ \subset F+$ y $F+ \subset H+$ tenemos $G+ \subset H+$ por la propiedad de transitividad de $\subset$. $\square$

3 Introducción a la normalización

Como manera de dejar ejercicios más intensivos (demostraciones menos obvias) para la siguiente clase auxiliar, se dio una breve y simple introducción a la normalización.

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

Solución. Se había visto en clases que:

Forma Normal de Boyce-Codd (FNBC). Para toda dependencia $X \rightarrow Y$ válida en un esquema R (X, Y son grupos de atributos de R), $X \rightarrow Y$ es una dependencia trivial o bien X es una superllave.

Tercera Forma Normal (3FN). Para toda dependencia $X \rightarrow Y$ válida en un esquema R (X, Y son grupos de atributos de R), $X \rightarrow Y$ es una dependencia trivial, X es una superllave o Y es atributo primo (es parte de alguna llave candidato).

Hay que tener cuidado que $X \rightarrow XY$, si Y es atributo primo, no viola 3FN. Pero todo el grupo de atributos XY no es un grupo de atributos primos por sí mismo. En casos como $X \rightarrow XYZ$, es mejor desagregar, y estudiar $X \rightarrow X$, $X \rightarrow Y$ y $X \rightarrow Z$ por separado para ver si se cumple, o no, 3FN.

(Forma normal) Busquemos la llave. Primero, vemos que nada determina A , luego A es parte de toda llave. Asimismo, C es parte de la llave. Viendo $AC+$: $AC \rightarrow ABC(A \rightarrow B) \rightarrow ABCD(A \rightarrow D) \rightarrow ABCDE(C \rightarrow E)$. Luego, AC es llave candidato (la única posible). Como vemos, en $A \rightarrow B$ tenemos una dependencia no trivial, donde A no es superllave y donde B no es atributo primo. Luego, R no está ni en FNBC ni en 3FN. Por lo tanto, hay que normalizar.

(FNBC) En este caso, se trata de ir partiendo el esquema paso a paso, sacando dependencias y creando nuevas relaciones:

1. $\{ABCDE\} \Rightarrow R_1(\underline{A}, B), \{ACDE\}$ por usar $A \rightarrow B$.
2. $\{ACDE\} \Rightarrow R_2(\underline{A}, D), \{ACE\}$ por usar $A \rightarrow D$.
3. $\{ACE\} \Rightarrow R_3(\underline{C}, E), \{AC\}$ por usar $C \rightarrow E$.
4. $\{AC\} \Rightarrow R_4(\underline{A}, \underline{C})$, por lo que quedó.

La descomposición R_1, R_2, R_3, R_4 es el resultado de la normalización. Las descomposiciones R_1, R_2 y R_3 *preservan las dependencias*. Como en 4 se preserva la llave, y además preservamos las dependencias, podemos reconstruir la relación original sin crear tuplas espurias: basta con usar *join natural* ($\bowtie$ ó $*$). Por ende, $R_1, \dots, R_4$ *preservan la información*.

La información no se preserva cuando la relación original no se puede reconstruir si se pierden tuplas o aparecen tuplas espurias (que no existían originalmente). Si tenemos, $R(A, B, C)$ y se cumple $A \rightarrow B$, la descomposición $R'(\underline{A}, B), R''(\underline{C})$ no permite recuperar la información... Por ejemplo, si hablásemos sobre personas, y guardásemos la relación $R(A = \text{ciudad_natal}, B = \text{nacionalidad}, C = \text{nombre})$, habríamos separado la ciudad natal y la nacionalidad del individuo. Claramente, con una relación de nombres de individuos y con otra relación con las ciudades natales y nacionalidades, no podemos saber la nacionalidad de cada persona, cosa que podíamos hacer en la relación original. Pero, si tenemos $R'(\underline{A}, B), R''(\underline{A}, \underline{C})$, podemos recuperar la información original. (Este es un caso de ejemplo, el esquema es idealizado y erróneo por ser demasiado simple.) Si preservamos la llave, y preservamos las dependencias de F , podemos usar la llave y la transitividad de las dependencias para localizar únicamente cada atributo. Y evitamos perder información.

(3FN) Vemos que F es irreducible ($F+$ siempre cambia si eliminamos alguna dependencia). Luego, podemos utilizar F para generar la descomposición en 3FN:

1. $R_1(\underline{A}, B)$ para guardar $A \rightarrow B$.
2. $R_2(\underline{A}, D)$ para guardar $A \rightarrow D$. (En la clase vimos que $A \rightarrow B$ y $A \rightarrow D$ se podían mezclar en $A \rightarrow BD$, cuya forma era más conveniente.)
3. $R_3(\underline{C}, E)$ para guardar $C \rightarrow E$.
4. Vemos que ninguna relación ha guardado la llave original. Así que la guardamos en una nueva relación $R_4(\underline{A}, \underline{C})$.

Y hemos obtenido la misma descomposición anterior.