Sistemas de Bases de Datos I - CI3311

Normalización - I

Profesores: Marlene Goncalves y José Tomás Cadenas

Introducción

- Técnica que busca garantizar que un esquema relacional cumpla con ciertas propiedades deseables, antes de ser implantado
- Normalización lidia con el problema de determinar redundancias

Anomalías de Operaciones

- Problemas derivados de efectuar operaciones sobre una BD con redundancia
- Depende del tipo de operación que la origina (inserción, actualización o eliminación)
- Se busca que los diseños relacionales no produzcan este tipo de anomalías

Normalización

- Verifica si un esquema dado esta libre de anomalías de operaciones
- Ofrece técnicas para transformar un esquema erróneo en un esquema correcto, bajo la óptica de las anomalías de operaciones

Dependencia Funcional (DF)

- Un conjunto de atributos X *implica funcionalmente* a un conjunto de atributos Y si el valor de X determina un valor único para Y
- Son obtenidas de restricciones del mundo real sobre los atributos, en base a la semántica, no pueden determinarse observando una instancia de la relación

Definición DF

Sea $R(A_1,...,A_n)$ una relación y $X,Y \subseteq \{A_1,...,A_n\}$ subconjuntos de atributos de R. Se dice que X **implica funcionalmente a** Y o que Y **depende funcionalmente** de X (se denotará $X \rightarrow Y$) y que **$X \rightarrow Y$ es una dependencia funcional de R**, se denotará como $df(X \rightarrow Y, R)$, si se cumple que:

$$(\forall r \forall t_1 \forall t_2 \mid r(R) \wedge tupla(t_1,r) \wedge tupla(t_2,r) \wedge t_1 \neq t_2 \wedge t_1[X] = t_2[X] \rightarrow t_1[Y] = t_2[Y])$$

Dependencia Funcional (DF)

- Toda superclave de R implica funcionalmente a cualquiera de los subconjuntos de los atributos de R
- También se cumple para cualquier clave (primaria o candidata) de R (recordar que una clave es una superclave minimal)

Definición (antecedente y consecuente)

Sea $R(A_1,...,A_n)$ una relación y X e Y subconjuntos de los atributos de R. Si $f_1 = X \rightarrow Y$ es una dependencia funcional de R - $df(f_1,R)$-, entonces X se dice **conjunto antecedente** de la dependencia funcional f_1 o **lado izquierdo** de f_1- lo denotaremos como antecedente(f_1,X)- e Y se dice **conjunto consecuente** de la dependencia funcional F_1 o lado derecho de f_1 - lo denotaremos como consecuente(f_1,Y)

Inferencia de DFs

- Existen muchas dependencias funcionales que pueden ser deducidas a partir de dependencias dadas
- En 1974, Armstrong definió un conjunto minimal de reglas de inferencia que permiten obtener nuevas dependencias usando el álgebra de dependencias funcionales (Axiomas de Armstrong)

Definición (Conjunto de DFs)

Sea $R(A_1,\ldots,A_n)$ una relación y F un conjunto de dependencias funcionales. Se dice que F_R es un conjunto de dependencias funcionales de R - se denotará como cdf(F,R)- si se cumple que:

$$(\forall f | pertenece(f,F_R) \rightarrow df(f,R))$$

Definición (Axiomas de Armstrong)

Sea $R(A_1,...,A_n)$ una relación y X,Y,Z subconjuntos de los atributos de R. Las siguientes reglas permiten obtener dependencias funcionales de R.

- **Regla de Reflexividad**:

  $Y \subseteq X \rightarrow df(X \rightarrow Y,R)$

- **Regla de Aumentatividad**:

  $df(X \rightarrow Y,R) \rightarrow df(XZ \rightarrow YZ,R)$

- **Regla de Transitividad**:

  $df(X \rightarrow Y,R) \wedge df(Y \rightarrow Z,R) \rightarrow df(X \rightarrow Z,R)$

Definición (Reglas derivadas de Armstrong)

Sea $R(A_1,...,A_n)$ una relación y X,Y,Z,W subconjuntos de los atributos de R. Las siguientes reglas derivadas de los axiomas de Armstrong permiten obtener dependencias funcionales de R:

- **Regla de Unión**:

  $df(X \rightarrow Y,R) \wedge df(X \rightarrow Z,R) \rightarrow df(X \rightarrow YZ,R)$

- **Regla de Seudotransitividad:**

  $df(X \rightarrow Y,R) \wedge df(WY \rightarrow Z,R) \rightarrow df(WX \rightarrow Z,R)$.

- **Regla de Descomposición**:

  $df(X \rightarrow YZ,R) \rightarrow df(X \rightarrow Y,R) \wedge df(X \rightarrow Z,R)$

Definición (Relación de Inferencia)

Sea $R(A_1,...,A_n)$ una relación y F un conjunto de dependencias funcionales de R. Se dice que la dependencia funcional $f = X \rightarrow Y$ puede inferirse a partir de F -se denotará como $F \models f$ - si f puede obtenerse a partir de F aplicando los axiomas de Armstrong o las reglas derivadas de los mismos o f forma parte de F.

Definición (Clausura de un conjunto de dependencias funcionales)

Sea $R(A_1,...,A_n)$ una relación y F un conjunto de dependencias funcionales de R. La clausura de F, se denotará como F^+, es el conjunto:

$$F^+ = \{ X \rightarrow Y \mid F \models X \rightarrow Y \}$$

Determinación de DFs

Para poder determinar las características de todas las DFs de la clausura de un conjunto de dependencias F debemos conocer todos los atributos que pueden obtenerse a cada uno de los antecedentes de las dependencias funcionales en F. A este conjunto de atributos los llamaremos la clausura del antecedente respecto al conjunto de dependencias funcionales. Siguiendo esta técnica se pueden expresar las reglas de formación de todas las dependencias funcionales de F^+

Definición
(Clausura de un conjunto de atributos)

Sea $R(A_1,...,A_n)$, F un conjunto de dependencias funcionales de R y X un subconjunto de los atributos de R. Se define la clausura de X respecto a F, se denota como $X^+_{F,}$ al siguiente conjunto:

$$X^+_F = \{\, Y \mid \text{atributo}(Y,R) \wedge F \models X \rightarrow Y \}$$

Algoritmo
(Clausura_Atributos_CDF)

Entrada:

X, un subconjunto de los atributos de una relación $R(A_1,\ldots,A_n)$

F, un conjunto de dependencias funcionales sobre $R(A_1,\ldots,A_n)$

Salida:

X^+_F, la clausura del conjunto de atributos X sobre el conjunto de dependencias funcionales F

Variables:

Nuevos, atributos agregados en la iteración i del algoritmo a X_i

Algoritmo
(Clausura_Atributos_CDF)

$X_0 \leftarrow X$

$i \leftarrow 0$

Repetir

$i \leftarrow i + 1$

$X_i \leftarrow X_{i-1}$

Nuevos $\leftarrow \varnothing$

Para cada $Y \rightarrow Z$ en F **Hacer**

Si $Y \subseteq X_i$ **Entonces** Nuevos $\leftarrow$ Nuevos $\cup$ Z

Fin

$X_i \leftarrow X_i \cup$ Nuevos

Hasta $(X_i = X_{i-1})$

Devolver X_i

Caracterización de la Clausura de un Conjunto de Dependencias Funcionales

Para poder caracterizar las DF de la clausura de un conjunto de dependencias F, basta con calcular la clausura de cada antecedente de las dependencias de F. Una vez realizado este proceso, si X es un antecedente de una DF f de F y X^{+}_{F} es su clausura respecto a F, puede decirse que en F^{+} existirán las siguientes dependencias funcionales:

- La DF original
- Las DFs cuyo antecedente es X y cuyo consecuente es cualquier subconjunto X^{+}_{F}.
- Las DFs resultantes de aumentar cualquier DF obtenida por el apartado anterior

Determinación de Claves

A través de un conjunto de dependencias funcionales de una relación R y el algoritmo de clausura de un conjunto de atributos respecto a un conjunto de dependencias funcionales pueden determinarse fácilmente todas las claves de la relación.

Definición (Clave)

Sea $R(A_1,...,A_n)$ una relación y F un conjunto de dependencias funcionales de R. Se dice que un subconjunto X de los atributos de R es clave de R si se cumple:

1. $F \models X \rightarrow A_1,...,A_n$ (Noción derivada de superclave)
2. $(\neg\exists Y \mid Y \subseteq X \wedge F \models Y \rightarrow A_1,...,A_n)$ (Minimalidad)

Determinación de Claves

1. Toda clave de la relación contiene al menos los atributos que no aparecen en ningún consecuente de las DFs del conjunto dado
2. Una vez encontrada una clave de la relación, intercambiar uno o más de los atributos de la clave que aparezcan como consecuentes de alguna dependencia por los que aparezcan como antecedentes en la misma. Esto generará una nueva clave para la relación.

Formas Normales

- *Normalización:* divide una relación en dos o más relaciones que garantizan la eliminación de anomalías a costa de crear redundancia **sólo** a nivel de claves
- *Forma Normal:* Condiciones utilizando las claves y las DFs de una relación para certificar si un esquema de relación está en una forma normal en particular
- Codd introdujo tres formas normales: Primera Forma Normal (1NF), Segunda Forma Normal (2FN) y Tercera Forma Normal (3NF).
- Boyce y Codd introdujeron una forma normal más restrictiva a la que llamaron Forma Normal de Boyce y Codd (BCNF)