Capítulo 5

Dependencias Funcionales

Dependencias Funcionales

- Hay que definir las restricciones de integridad para el *conjunto de tablas legales.*
  - **Tablas legales** son las tablas con las que la empresa quiere poder trabajar.
  - Son tablas donde las tuplas tienen un cierto significado y cumplen con ciertas propiedades obligatorias.
- Las **dependencias funcionales (DF)** requieren que para las tablas legales,
  - el valor de un cierto conjunto de atributos determine unívocamente el valor de otro conjunto de atributos.

Dependencias Funcionales

- **Formalización**:
  - Sea R un esquema relacional

$$\alpha \subseteq R \ \ y \ \beta \subseteq R$$

  - La dependencia funcional

$$\alpha \rightarrow \beta$$

se cumple en R si y solo si para todas las relaciones legales r(R), cada vez que dos tuplas t_1 y t_2 de r coinciden en los atributos α, también coinciden en los atributos β. Formalmente:

$$t_1[\alpha] = t_2[\alpha] \ \Rightarrow \ t_1[\beta] = t_2[\beta]$$

Dependencias Funcionales

- **Problema**: Dado *R* esquema universal. ¿Cómo encontrar las dependencias funcionales (DF) de un problema del mundo real (con atributos en *R*)?
- **Solución**: considerar un proceso iterativo. En cada paso de iteración se tiene *F* conjunto de DF y se tienen los siguientes pasos:
  1. Proponer **DF candidata** *f*
     - Una **DF candidata** puede **ser o no ser válida**;
     - en caso de ser válida puede **convenir o no agregarla a *F***.
     - Para encontrar DF candidata usaremos **reglas heurísticas**.
  2. Evaluar si *f* es **válida**:
     - *f* es **válida** si tiene sentido para el problema del mundo real.
     - Para esto usaremos **reglas de descarte de DF** y
     - **razones positivas** para tener una DF que no cumplió ninguna regla de descarte.
  3. Si *f* es válida, ver si **vale la pena agregarla** a *F*.
     - Veremos que no siempre vale la pena agregar una DF candidata valida a *F*.

Derivando Dependencias Funcionales

- No hace falta listar todas las DF de un problema del mundo real.
- **Ejemplo**: persona = (DNI, nombre, posición)
  - DNI -> nombre, posición
  - Resulta bastante obvio que : DNI-> nombre y DNI -> posición.
  - Estas últimas se pueden derivar de la primera.

Derivando Dependencias Funcionales

- **Ejemplo**: Sea R = (A, B, C, G, H, I),
  - F = { A→ B, A→ C, B → H }
  - Resulta bastante obvio que también se cumplen:

    A→ H y A → B, C.

  - Estas últimas se pueden derivar o deducir de las de F. Luego no hace falta listarlas.

Derivando Dependencias Funcionales

- **Conclusión 1:** Si *F* conjunto de DF, hay otras DF fuera de *F* que también se cumplen y que se pueden derivar de las de *F*.
- **Conclusión 2:** no necesito dar todas las DF que se cumplen en el problema del mundo real sino un subconjunto de ellas lo menor posible tal que todas las demás DF se puedan derivar de ese subconjunto.
- Recordar que las dependencias funcionales son restricciones de integridad y las restricciones de integridad necesitan ser chequeadas cada vez que cambia la base de datos. Y esto tiene su costo computacional.
  – Por lo tanto, cuantas menos dependencias funcionales necesiten ser chequeadas, mejor.

Derivando Dependencias Funcionales

- Necesitamos formalizar qué significa derivar. Para esto consideramos un conjunto de reglas de inferencia.
- Las reglas consideradas se llaman **axiomas de Armstrong:**
  - if $\beta \subseteq \alpha$, then $\alpha \rightarrow \beta$ **(reflexividad)**
  - if $\alpha \rightarrow \beta$, then $\gamma\,\alpha \rightarrow \gamma\,\beta$ **(aumentatividad)**
  - if $\alpha \rightarrow \beta$, and $\beta \rightarrow \gamma$, then $\alpha \rightarrow \gamma$ **(transitividad)**

Derivando Dependencias Funcionales

- **Ejemplo**: queremos deducir $AC \rightarrow D$ a partir de las DF $\{A \rightarrow B; CB \rightarrow D\}$ usando las reglas de Armstrong.
  - **Bosquejar una deducción.**

Derivando Dependencias Funcionales

- **Ejemplo**: queremos deducir $AC \rightarrow D$ a partir de las DF $\{A \rightarrow B; CB \rightarrow D\}$ usando las reglas de Armstrong.

  1) A->B
  2) CB ->D
  3) AC -> CB (aumentatividad a 1))
  4) AC -> D (transitividad a 3) y 2))

Derivando Dependencias Funcionales

- **Generalizando**: Dado un esquema relacional R, una DF f con atributos en R **se deduce** de un conjunto de DFs F con atributos en R *si:*
  - existe una lista de DFs $f_1,\ldots, f_n$ tales que $f_n = f$ y para todo $1 \leq i \leq n$:
  1. $f_i \in F$ *o*
  2. f_i se obtiene por aplicar la regla de reflexividad o
  3. f_i se obtiene por aplicar aumentatividad o transitividad a pasos anteriores en la lista.
- **Notación**: Usaremos $F \vdash f$ para decir que f se deduce de F.

Derivando Dependencias Funcionales

- Usando solo los axiomas de Armstrong muchas veces las derivaciones se tornan un poco largas.
- Con algunos axiomas adicionales, muchas derivaciones se tornan más cortas y sencillas. Estos son:
  - If $\alpha \rightarrow \beta$ holds *and* $\alpha \rightarrow \gamma$ holds, then $\alpha \rightarrow \beta\gamma$ holds (**unión**)
  - If $\alpha \rightarrow \beta\gamma$ holds, then $\alpha \rightarrow \beta$ holds and $\alpha \rightarrow \gamma$ holds (**decomposición**)
  - If $\alpha \rightarrow \beta$ holds *and* $\gamma\beta \rightarrow \delta$ holds, then $\alpha\gamma \rightarrow \delta$ holds (**pseudotransitividad**)

Las reglas anteriores se pueden inferir a partir de los axiomas de Armstrong.

Cierre de conjunto de DF

- Sea *R* esquema relacional y *F* conjunto de DF con atributos en *R*;
  - el **cierre de *F*** se denota con *F*+ y
  - son todas las dependencias funcionales que se deducen de *F*.
  - Formalmente:

$$F+ = \{\alpha \rightarrow \beta \mid \alpha \subseteq R,\ \beta \subseteq R,\ F \vdash \alpha \rightarrow \beta\}$$

- **Situación**: tenemos *R* esquema universal para un problema del mundo real y queremos hallar un conjunto adecuado de DF para el mismo.
- Comenzaremos con un problema más sencillo:
  - **Problema**: hemos calculado un conjunto de DF *F* con atributos en *R* y tenemos una DF candidata *f* (con atributos en *R*). ¿Vale la pena agregar *f* a *F*?
  - Después veremos algunas maneras de encontrar DF candidatas.

Cierre de conjunto de DF

- **Problema**: tenemos *F* conjunto de DF para problema de mundo real y queremos saber si vale la pena agregar DF *f* a *F*.
- **Idea 1**: calcular *F*+ y ver si *f* está en *F*+
  - Recordar que *F*+ = $\{\alpha \rightarrow \beta \mid \alpha \subseteq R,\ \beta \subseteq R,\ F \vdash \alpha \rightarrow \beta\}$
  - Si *f* está en *F*+: entonces *f* no se agrega a *F* porque *f* se deduce de *F*.
  - en caso contrario: hay que agregar *f* a *F*.
- **¿Es esta idea viable?**

Cierre de conjunto de DF

- **¿Es viable calcular *F+* si tenemos muchos atributos en el problema del mundo real?**

- En la práctica hay cientos de atributos en esquema universal y no es viable calcular *F+*.

- La razón es que *F+* es demasiado grande:

  – Para α hay $2^{|\alpha|}$ dependencias triviales.

  – Además, si $\alpha \rightarrow \beta$ en *F* entonces hay $2^{|R|}$ maneras de aplicar aumentatividad a $\alpha \rightarrow \beta$ (*R* esquema universal).

- Como ven, *F+* es demasiado grande como para poder calcularlo.

Cierre de conjunto de DF

- **Idea 2**: intentar deducir *f* de *F* y si lo logramos: no se agrega *f* a F.
- ¿Y si no lo logramos?
  - puede que no sepamos como derivar *f* de *F* o
  - que *f* no sea derivable de *F*.
  - Necesitamos saber que *f* no es deducible de *F*, pero no sabemos hacer ese tipo de pruebas.

Cierre de un conjunto de atributos

- **Idea 3**: Quiero saber si α -> β es deducible de *F*. Si calculamos solo las dependencias de *F*+ con lado izquierdo α,
  - este conjunto es mucho más chico que *F*+.
  - **Por lo tanto**: para responder si $F \vdash \alpha \rightarrow \beta$, podríamos contestar

    ¿ $\alpha \rightarrow \beta \in \{\alpha \rightarrow \varphi \mid F \vdash \alpha \rightarrow \varphi\}$?

- Pero se puede hacer mejor aún, calculando solo una pequeña fracción de las dependencias de $\{\alpha \rightarrow \varphi \mid F \vdash \alpha \rightarrow \varphi\}$.
- **Idea 4**: Calcular solo

$$\{\alpha \rightarrow A \mid F \vdash \alpha \rightarrow A\}$$

  - Aquí hay menos cantidad de dependencias que $|R|$.
  - ¿Por qué esta idea funciona?

Cierre de un conjunto de atributos

- Quiero saber si $\alpha \rightarrow \beta$ es deducible de *F*. La idea anterior funciona porque si se cumple:

$$\beta \subseteq \{A \in R \mid F \vdash \alpha \rightarrow A\}$$

  - Entonces por regla de unión aplicada finitas veces a todos los atributos de β sale $F \vdash \alpha \rightarrow \beta$.

- Llegamos así a un conjunto conocido como **cierre de un conjunto de atributos.**
  - es el conjunto a la derecha de la inclusión en esta página.

Cierre de un conjunto de atributos

- Sea R el <u>esquema universal</u>, F conjunto de DF del problema del mundo real (con atributos en el universal), sea $\boldsymbol{\alpha} \subseteq R$.
  - El **cierre de $\boldsymbol{\alpha}$ bajo F** (denotado por α_F^+) se define:

$$\alpha^+_F = \{A \in R : F \vdash \alpha \rightarrow A\}.$$

Cierre de un conjunto de atributos

- **Proposición:** $F \vdash \alpha \rightarrow \alpha^{+}_{F}$
- **Prueba**: sale aplicando unión finitas veces .
- **Proposición**: $F \vdash \alpha \rightarrow \beta$ si y solo si $\beta \subseteq \alpha^{+}_{F}$
- **Problema**: tenemos *F* conjunto de DF para problema de mundo real y queremos saber si vale la pena agregar DF α -> β a *F*.
- **Solución**: chequear si se cumple $F \vdash \alpha \rightarrow \beta$ usando la proposición anterior.
  - Si $\beta \subseteq \alpha^{+}_{F}$: la respuesta es sí; por lo tanto, no agregar α -> β a *F*.
  - Sino: α -> β no se deduce de *F*, por lo tanto, agregamos α -> β a *F*.

Cierre de un conjunto de atributos

- Solo está faltando un algoritmo para computar α^+_F
- Como $\alpha^+_F = \{A \in R : F \vdash \alpha \rightarrow A\}$.
  - Una idea seria para cada $A \in R$ intentar deducir $\alpha \rightarrow A$ *de F.*
    - En ese caso, solo terminaremos calculando un subconjunto de α^+_F ,
      - Para las deducciones que logremos completar.
      - En algunos casos podemos no encontrar una deducción aun existiendo una, o que esta no exista.
  - Ahora veremos que hay otra forma eficiente de calcular α^+_F.

Cierre de un conjunto de atributos

- Algoritmo para computar α^+_F (la clausura de α bajo *F*)

  result := α;
  while (changes to *result*) **do**
      for each $\beta \rightarrow \gamma$ **in** *F* **do**
      begin
          if $\beta \subseteq$ *result* **then** *result* := *result* $\cup \gamma$
      end

  – Luego de cada asignación a la variable *result* se cumple *result* $\subseteq \alpha^+$.
  – Por lo tanto, cuando el algoritmo termina se cumple

  $$result \subseteq \alpha^+ .$$

  – La prueba de que en este momento $\alpha^+ \subseteq$ *result* escapa al alcance de la materia.

Cierre de un conjunto de atributos

- **Ejercicio**: Dados $R = (A, B, C, G, H, I)$,

  $F = \{ A \rightarrow B, A \rightarrow C, CG \rightarrow H, CG \rightarrow I, B \rightarrow H$

  - Calcular A^+_F y $(AG)^+_F$
  - Aplicar el algoritmo anterior.

Superclaves y Claves Candidatas

- **Definición**: Sea $K \subseteq R$, R esquema de relación; K es una **superclave** de R si
  - los valores para K son suficientes para identificar una tupla única en cada posible relación $r(R)$.
- En otras palabras para todo par de tuplas t1 y t2: Si t1[K] = t2[K] entonces t1 = t2
- Pero esto es lo mismo que: para todo par de tuplas t1 y t2: si t1[K] = t2[K] entonces t1[R] = t2[R]
- Esto es lo mismo que decir que K→R es válida.
- Si tenemos *F* las dependencias funcionales del problema; entonces K→ R tiene que poder derivarse a partir de *F*. Entonces podemos hacer la siguiente definición:
- **Definición**: Sea *R* esquema relacional y *F* es conjunto de DF del problema, entonces α es **superclave** de *R* si y solo si $F \vdash \alpha \rightarrow R$.

Superclaves y Claves Candidatas

- $\boldsymbol{\alpha}$ **clave candidata** de *R* si y solo si:
  - $\boldsymbol{\alpha}$ superclave de *R*
  - para todo A en $\boldsymbol{\alpha}$: $\boldsymbol{\alpha}$ - {A} no es superclave de *R*
- **Para chequear que α superclave de *R:***
  - computar α^{+}, y chequear si α^{+} contiene todos los atributos de *R*.

Superclaves y Claves Candidatas

- **Ejercicio**: Dados $R = (A, B, C, G, H, I)$,

  $F = \{ A \rightarrow B, A \rightarrow C, CG \rightarrow H, CG \rightarrow I, B \rightarrow H \}$.

  1. Probar que de F no se deduce $A \rightarrow I$
  2. Probar que AG es clave candidata.