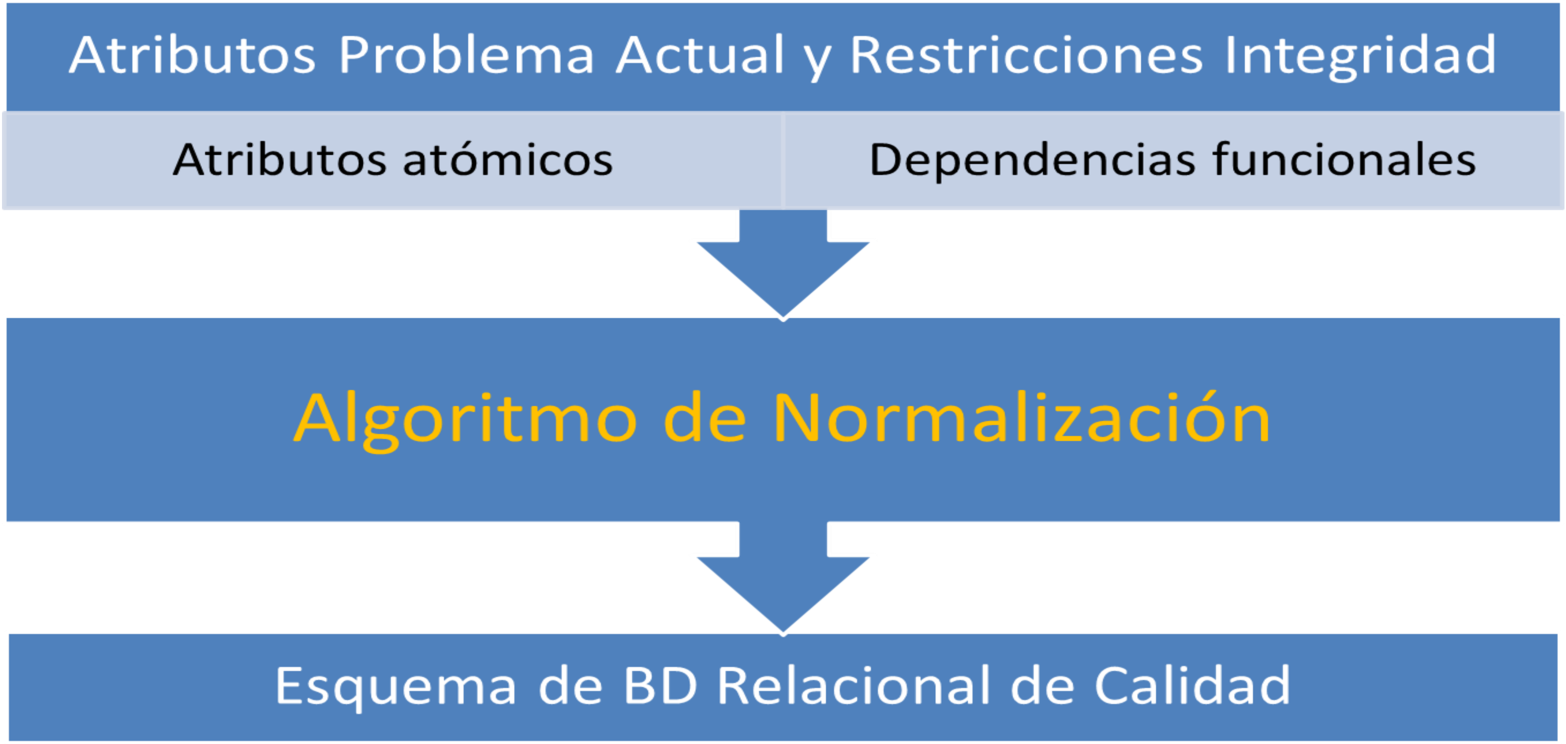
Capítulo 5

Normalización

Normalización

- **Meta**: aprender a hacer **diseños de calidad** de esquemas de BD relacionales.
- **Solución**: algoritmos de normalización.



Normalización

- Si proveo inputs deficientes, la calidad de la solución calculada se va a ver perjudicada.
- **Aun cuando no aplico algoritmos de normalización necesito tener los inputs.**
  1. Debo considerar los atributos del problema en el diseño.
     - Sino el cliente va a quedar descontento.
  2. Las DF son restricciones de integridad que necesitan ser capturadas para mantener la integridad de la BD.
- **Conclusión**: El aplicar normalización es un bonus por tener los inputs adecuados que **son obligatorios**.

Normalización

- **Repaso**: si R esquema con redundancia de información en un conjunto de atributos β y la DF α -> β (α y β son disjuntos) se cumple en R:
  - Si α no determina todos los atributos de R,
  - la dependencia α -> β puede ser usada para eliminar redundancia de información por medio de la descomposición de R:
  - Para eliminar la redundancia de información se saca β de R y se crea un esquema con los atributos de α U β.
  - Al hacer esto desaparece la redundancia de información para los atributos de β.

Normalización

- Si hay 2 tuplas distintas de R que coinciden en α y vale α -> β y α no determina todos los atributos de R,
  - entonces para esas tuplas se van a repetir los valores de β; o sea, tenemos redundancia de información para los atributos de β.
- Sea R, F (conjunto de DF). Que α no determina todos los atributos de R es lo mismo que decir:
  - que α -> R no se deduce de F
  - que α no es superclave de R

Normalización

- **Problema**: Queremos caracterizar un esquema *R* que no tiene redundancia de información proveniente de dependencias funcionales por medio de una propiedad.
  - Un esquema que cumple esa propiedad diremos que está en la forma normal Boyce-Codd (FNBC).

Forma normal de Boyce Codd

- **Definición**: Un esquema R está en **forma normal de Boyce-Codd (FNBC)** con respecto a un conjunto F de DFs si para todas las DFs en F^+ de la forma $\alpha \to \beta$, donde $\alpha \subseteq R$ y $\beta \subseteq R$, al menos una de las siguientes propiedades se cumple:
  - $\alpha \to \beta$ es trivial (i.e., $\beta \subseteq \alpha$)
  - α es una superclave de R (i.e. $\alpha \to R \in F^+$).
- **Definición**: Sea R esquema universal, F conjunto de DFs. Una descomposición $\{R_1,\dots,R_n\}$ de R está en Forma normal de Boyce-Codd (FNBC) con respecto a F si y solo si cada R_i está en FNBC con respecto a F.

Forma normal de Boyce Codd

- ***¿Cómo comprobar que un esquema R con respecto a F no está en FNBC?***
- Una DF de F^+ que no cumple la condición de FNBC se llama **violación** o **DF testigo**.
  - Es una DF $\alpha \to \beta$ no trivial en F^+ tal que $\alpha \to R \notin F^+$
- Para probar que *R* no está en FNBC con respecto a *F* basta con encontrar una DF testigo en F^+.
  - A veces (pero no siempre) la DF testigo está en *F*.

Forma normal de Boyce Codd

- **Ejemplo**: Sea $R = (A, B, C)$ esquema con DFs:
  $F = \{A \rightarrow B, B \rightarrow C\}$.
  - R no está en FNBC. **¿Por qué?**
  - $\{A\}$ es clave candidata de R
  - $B \rightarrow C$ es testigo:
    - A no está en B+ = {B, C},
    - B no es superclave de R
  - Sea la descomposición de R: $R_1 = (A, B)$, $R_2 = (B, C)$
  - Esta descomposición está en FNBC

Forma normal de *Boyce* Codd

- **Ejemplo**: Sea el esquema relacional *R* = (A, C, D) con DFs: *F* = {A→B, B→C}.
- **¿Está *R* en FNBC?**
- A->B: no se puede usar porque B no está en R. Cuidado con este tipo de error común!
- A->C está en F+ (por transitividad de las DF de F)
- A no superclave de R porque D no está en A+
- Luego R no está en FNBC.

Forma normal de Boyce Codd

- **Proposición**: Para comprobar si *R*, *F* está en FNBC, donde los atributos de *F* están contenidos en *R*, basta con chequear la condición de FNBC para las DF de *F*.
  – Hay que ver que ninguna DF es testigo.
- **Consecuencia**: para descomponer el esquema universal basta con buscar testigos en F.
- Si los atributos de F no están contenidos en R necesitamos otra manera de probar que un esquema está en FNBC.

Forma normal de Boyce Codd

- **Ejemplo**: Sea $R = (A, B, C)$ esquema con DFs:

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

  - **¿Está *R* en FNBC?**

- Atributos(F) contenidos en R.
- Por proposición anterior basta con chequear dependencias del F.
- A, B y C son superclaves (justificarlo).
- Entonces ningún miembro de F es testigo.
- Por lo tanto R está en FNBC.

Comprobación de Forma normal de Boyce Codd

- **Situación:** lo intentamos y no encontramos una DF testigo.
  - En ese caso intentar probar que tenemos un esquema en FNBC.
  - Chequear todos los $\alpha \rightarrow \beta$ de F+ con atributos en R es demasiado costoso.
- **Comprobación de FNBC**: Sea R_U universal, con DFs F y sea R_i que forma parte de descomposición de R_U; para probar que R_i está en FNBC se puede hacer la siguiente comprobación:
  - $\forall\, \alpha \subseteq R_i : \alpha^+ \cap (R_i - \alpha) = \phi \lor R_i \subseteq \alpha^+$
- La primera condición del V significa: todas las DF con lado izquierdo α son triviales.

Comprobación de Forma normal de Boyce Codd

- **Prueba**: Supongamos que R_i está en FNBC y $\neg R_i \subseteq \alpha^+$:
  - toda $\alpha \rightarrow \beta$ en F^+ con atributos en R_i es trivial.
  - Esto equivale a $\beta \cap (R_i - \alpha) = \phi$
  - Luego: $\alpha^+ \cap (R_i - \alpha) = \phi$ (tomo $\beta = \alpha^+$)

Comprobación de Forma normal de Boyce Codd

- **Ejercicio**: Sea *F* dado por:
  1. nomBib → calle, numero
  2. calle, numero → nomBib
  3. ISBN → título, editorial, autores, edición
  4. nomBib, numInv → ISBN
- Sea la descomposición:
  - BibLibs = (nomBib, numInv, ISBN)
  - Biblioteca = (nomBib, calle, número)
  - Libro = (ISBN, título, editorial, autores, edición)
- Comprobar que *Biblioteca*, *Libro* están en FNBC.

Forma normal de Boyce Codd

- La comprobación de FNBC anterior va a ser usada por el algoritmo de normalización.
- **Situación:** la comprobación de FNBC falla para un α.
  - Eso nos permite definir una DF testigo.
  - ***¿Cuál es una dependencia testigo?***
  - Considerar la negación de la comprobación de FNBC

Forma normal de Boyce Codd

- **Observación**: Si $\alpha \subseteq R_i$ viola la condición:

$$\forall \alpha \subseteq R_i : \alpha^+ \cap (R_i - \alpha) = \phi \vee R_i \subseteq \alpha^+$$

  entonces la siguiente DF es testigo:

  – $\alpha \rightarrow \alpha^+ \cap (R_i - \alpha)$.

- Usamos esta dependencia para descomponer Ri.
- Ahora estamos en condiciones para presentar el algoritmo de normalización.

Algoritmo de normalización en FNBC

- **Problema:** Sea *R*, *F* conjunto de DFs. ¿Cómo hallar una descomposicion de *R* que está en FNBC?
- **Solución**: Algoritmo de normalización en FNBC.
- *result* := {*R*};

  while (there is a schema R_i in *result* that is not in BCNF) **do**

  begin

  let $\alpha \rightarrow \beta$ DF testigo de R_i and $\alpha \cap \beta = \emptyset$;

  $result := (result - R_i) \cup (R_i - \beta) \cup (\alpha, \beta);$

  end

Algoritmo de normalización en FNBC

- Algunas aclaraciones sobre el algoritmo anterior si se **implementa automáticamente**:
  - Para buscar esquema que no está en FNBC se puede usar el algoritmo de comprobación de que esquema está en FNBC.

$$\forall \alpha \subseteq R_i : \alpha^+ \cap (R_i - \alpha) = \phi \vee R_i \subseteq \alpha^+$$

  - Ese algoritmo va a encontrar un α que no cumple la condición. Y a partir del mismo se puede obtener la DF testigo:

$$\alpha \rightarrow \alpha^+ \cap (R_i - \alpha) .$$

Algoritmo de normalización en FNBC

- **Ejercicio**: Aplicar el algoritmo de normaliza-ción en FNBC a:
  - R = (A, B, C, D, E, F)
  - $\mathcal{F}$ = {A → CB, E → FA}

Algoritmo de normalización en FNBC

- **Ejercicio**: Sea el esquema universal:

BibLibs = (nomBib, calle, número, numInv, ISBN, título, editorial, autores, edición)

Sea *F* dado por:

- nomBib → calle, número
- calle, número → nomBib
- ISBN → título, editorial, autores, edición
- nomBib, numInv → ISBN

Aplicar el algoritmo de normalización en FNBC.

Algoritmo de normalización en FNBC

- **¿nomBib -> calle, numero es testigo?**
  - Sí porque no es trivial y
  - nomBib+ = {nomBib, calle, numero} <> R
  - Luego nomBib no es superclave de R.
- **¿Cómo queda la descomposición de BibLibs?**
  - BibLibs2= (nomBib, numInv, ISBN, titulo, editoral, autores, edicion)
  - R1 = (nomBib, calle, numero)
- **¿ISBN → título, editorial, autores, edición es testigo?**
  - Sí porque no es trivial y
  - ISBN+ = {ISBN, titulo, editoral, edición, autores} que es menor que BibLib2
  - Luego ISBN no es superclave de BibLib2.
- **¿Cómo queda la descomposición de BibLibs2?**

Algoritmo de normalización en FNBC

  - BibLibs3 = (nomBib, numInv, ISBN)
  - R2 = (ISBN, titulo, editorial, edicion, autores)
- En un ejercicio anterior sale que R1 y R2 están en FNBC.
- Falta ver que BibLib3 está en FNBC.
- Usamos la comprobación: ▪ $\forall \alpha \subseteq R_i : \alpha^+ \cap (R_i - \alpha) = \emptyset \vee R_i \subseteq \alpha^+$
- {numInv, nomBib}+ = BibLibs ⊃ BibLibs3
- Luego {numInv, nomBib} superclave y no hace falta chequear superconjuntos.
- {ISBN, numInv} + = R2 U {numInv}, luego no contiene BibLibs3
- {ISBN, numInv}+ ∩ BibLibs3 - {ISBN, numInv} = R2 ∩ {nomBib} = Ø
- {ISBN, nomBib} + = R2 U R1, luego no contiene BibLibs3
- {ISBN, nomBib} + ∩ BibLibs3 - {ISBN, nombib} = (R2 U R1) ∩ {numInv} = Ø

Algoritmo de normalización en FNBC

- $nomBib^+$ = {nomBib, calle, numero}
- $nomBib^+ \cap BibLibs3 - nomBib$ = {nomBib, calle, numero} ∩ {numInv, ISBN} = ∅.
- $numinv^+$ = numinv
- $ISBN^+$ = R2   luego ISBN no es superclave de BibLib3
- $ISBN^+$ ∩ {nomBib, numInv} = ∅
- Hemos chequeado todos los casos, por lo tanto BibLibs3 está en FNBC.