Capítulo 5

Dependencias Funcionales

 Meta: aprender a hacer diseños de calidad de esquemas de BD relacionales, idealmente de manera automática.

• Repaso:

- Un esquema relacional es un conjunto de esquemas de relaciones.
- Un esquema de relación es una lista de nombres de atributos.

- Para tener un diseño de calidad: ¿Qué cosas pedir?
 - Evitar problemas de redundancia de información.
 - Evitar problemas de comprensibilidad.
 - Evitar problemas de incompletitud.
 - Restricciones de integridad incompletas.
 - Relaciones entre atributos no contempladas por esquemas de BD.
 - Evitar problemas de ineficiencia.
 - Chequeo ineficiente de restricciones de integridad.
 - Consultas ineficientes por tener un esquema inadecuado de BD.

Enfoque ya visto para lograr esta meta

- Hacer buen diseño de esquema-ER.
- Mapear esquema-ER a esquema relacional.
- Dificultades:
 - Es un trabajo manual
 - Corregir esquemas-ER con problemas de calidad.
 - Tomar decisiones de diseño:
 - Considerar alternativas suficientes de diseño
 - Elegir entre las alternativas la mejor basándose en criterios de calidad.

 Problema ¿Cómo evitar tener que diseñar diagrama ER, tener que decidir entre alternativas de diseño ER, tener que corregir problemas de calidad en diseños ER?

Solución:

- Identificar K el conjunto de todos los atributos (atómicos) del problema actual.
- Luego a partir de K y para el problema actual definir un conjunto de restricciones de integridad I,
 - que servirán de guía para hacer un buen diseño.
- Se aplica un algoritmo llamado de normalización que:
 - Calcula un esquema de BD relacional a partir de K y de I.

Atributos Problema Actual y Restricciones Integridad

Atributos atómicos

Dependencias funcionales

Algoritmo de Normalización

Esquema de BD Relacional de Calidad

- Situación: para aplicar un algoritmo de normalización necesitamos conocer primero los atributos del problema actual.
- Dado el problema actual, el **esquema universal** consta de todos los atributos atómicos del mismo.

- **Ejemplo**: sistema de bibliotecas
 - Se tiene un sistema de bibliotecas de una ciudad, el sistema está formado por bibliotecas de las que se proveen su nombre, y su domicilio formado por calle y número; las bibliotecas tienen libros de los que se almacena su ISBN, su título y sus autores; las bibliotecas llevan un número de inventario para distinguir entre las distintas copias de libros; además las bibliotecas tienen socios para los que se almacena nombre, DNI y posición (la misma puede ser egresado, docente, estudiante, etc.); a los socios se les prestan ejemplares de libros.
 - ¿Cuál es el esquema universal para este problema?

- **Ejemplo**: sistema de bibliotecas
 - Se tiene un sistema de bibliotecas de una ciudad, el sistema está formado por bibliotecas de las que se proveen su nombre, y su domicilio formado por calle y número; las bibliotecas tienen libros de los que se almacena su ISBN, su título y sus autores; las bibliotecas llevan un número de inventario para distinguir entre las distintas copias de libros; además las bibliotecas tienen socios para los que se almacena nombre, DNI y posición (la misma puede ser egresado, docente, estudiante, etc.); a los socios se les prestan ejemplares de libros.
 - SmaBibliotecas = (nombre, DNI, posición, numInv, nombreBib, calle, numero, ISBN, título, autores)
 - ¿Cuál es el significado de una tupla de una relación para el esquema universal?

- **Ejemplo**: sistema de bibliotecas
 - Se tiene un sistema de bibliotecas de una ciudad, el sistema está formado por bibliotecas de las que se proveen su nombre, y su domicilio formado por calle y número; las bibliotecas tienen libros de los que se almacena su ISBN, su título y sus autores; las bibliotecas llevan un número de inventario para distinguir entre las distintas copias de libros; además las bibliotecas tienen socios para los que se almacena nombre, DNI y posición (la misma puede ser egresado, docente, estudiante, etc.); a los socios se les prestan ejemplares de libros.
 - SmaBibliotecas = (nombre, DNI, posición, numInv, nombreBib, calle, numero, ISBN, título, autores)
 - Significado: Una tupla de información para una relación de ese esquema para un sistema de bibliotecas consiste de: datos sobre una persona (correspondientes a nombre, DNI, posición) datos sobre un libro (correspondientes a numInv, ISBN, título, autores), datos sobre una biblioteca (correspondientes a calle, número, nomBib), tal que esa persona es socia de esa biblioteca, y se le ha prestado ese libro que es de esa biblioteca.

- Ya que el esquema universal es un esquema de BD relacional,
 - ¿conviene usarlo como diseño de una BD relacional?
 - Respuesta: no, porque el esquema universal usualmente tiene problemas de calidad, como redundancia de información.

- Sea el siguiente esquema universal:
 - SmaAutomotor = (DNI, nombre, marca, modelo, patente, numSeguro, compañíaSeguro, direcciónCS)

• Problemas de diseño:

- Redundancia de información: direcciónCS aparece repetido para cada coche asegurado por esa compañía de seguros; nombre aparece repetido por cada coche que tiene esa persona.
- Manejo de valores nulos: si una persona no tiene coche, aparecen los demás campos en nulo; si una compañía de seguro no tiene coches asegurados, aparecen los demás campos en nulo.
 - Antes de decidir si agregamos valores nulos, tenemos que hacer consultas.
 - P.ej. si a una persona le sacamos coche asegurado, hay que consultar si tiene otros coches asegurados, sino hay que poner nulos.

- Problema: ¿Cómo se pueden eliminar los problemas citados?
- Solución: descomponer el esquema universal para eliminar los problemas citados.

• Ejemplo:

- O Persona = (DNI, nombre)
- O Auto = (marca, modelo, patente)
- CompañíaAseguradora = (compañíaSeguro, direcciónCS)
- Seguro = (patente, compañíaSeguro, numSeguro)
- O TieneAuto = (patente, DNI)

Problemas de usar esquema universal

- La teoría de normalización estudia cómo descomponer esquemas universales para eliminar los problemas citados.
 - O Implicaciones de usar normalización:
 - no hace falta ser bueno en modelado o diseño;
 - pero hay que ser bueno encontrando las restricciones de integridad necesarias (i.e. dependencias funcionales) y el esquema universal.

Descomposiciones

- Motivación: Un algoritmo de normalización va descomponiendo un esquema universal
 - o hasta obtener un esquema de BD relacional de calidad.
- Sea *R* un esquema de relación. Un conjunto de esquemas de relación {*R*1,...,*Rn*} es una descomposición de *R* sii

$$R = R1 \cup ... \cup Rn$$
.

- Sea el siguiente esquema relacional:
 - SmaBibliotecas = (nombre, DNI, posición, numInv, nombreBib, calle, numero, ISBN, título, autores)
- Podemos identificar la siguientes redundancias de información:
 - el mismo usuario saca varios libros: se repite nombre y posición varias veces
 - o para cada persona que es socia de biblioteca se repite calle y número
 - o el mismo libro se saca varias veces prestado y se repite titulo y autores

- Situaciones que obligan a trabajar con valores nulos:
 - si tengo una persona solamente (no es socia, no le prestaron libros, etc.) entonces los campos van ser nulos.
 - lo mismo si tengo libro solamente, los demas campos van a ser nulos.

• Lo malo de trabajar con valores nulos:

- Si borro una tupla de tabla del esquema para eliminar una asociación entre entidades, hay que tener cuidado de no eliminar entidades que deben permanecer.
- Esto obliga a hacer consulta extra y posible modificación de tupla agregando valores nulos.

- Considero:
- Biblioteca = (<u>nombreBib</u>, calle, número)
- SmaBibliotecas2=(DNI, numInv, nombreBib, ISBN, título, autores)
- Así elimino la segunda redundancia de información.

- Problema: ¿Cómo arreglar problemas de redundancia de información?
- Idea: puedo ir descomponiendo gradualmente el esquema universal de acuerdo a conceptos que sirven para sacar atributos que ocasionan redundancia de información afuera.
- Consideremos el esquema anterior:

```
SmaBibliotecas = (nombre, DNI, posición, numInv,
nombreBib, calle, numero, ISBN, título, autores)
```

- Persona = (nombre, <u>DNI</u>, posición)
- SmaBibliotecas1 = (DNI, numInv, nombreBib, calle, numero, ISBN, titulo, autores)
- Saqué de SmaBibliotecas: nombre y posición eliminando la primera redundancia de información.

- ¿Cómo eliminar la redundancia de información: para cada persona que es socia de biblioteca se repite calle y número?
- SmaBibliotecas1 = (DNI, numInv, nombreBib, calle, numero, ISBN, titulo, autores)
- Sea el esquema:
- Biblioteca = (<u>nombreBib</u>, calle, número)
- SmaBibliotecas2=(DNI, numInv, nombreBib, ISBN, título, autores)
- Saqué de SmaBibliotecas1: calle y número.

- ¿Cómo eliminar la redundancia de información: el mismo libro se saca varias veces prestado y se repite titulo y autores?
- SmaBibliotecas2=(DNI, numInv, nombreBib, ISBN, título, autores)
- Sea el esquema:
- Libro = (**ISBN**, título, autores)
- SmaBibliotecas3 = (DNI, numInv, nombreBib, ISBN)
- Saqué de SmaBibliotecas2: título y autores.

- Pero sigue el problema de los nulos:
 - Persona en biblioteca sin préstamo de libro: nulo en numlnv.
 - o Libro en biblioteca pero no hubo préstamo, DNI nulo.
- Eliminar estos problemas introduciendo otros conceptos con sus respectivos esquemas.
- Divido SmaBibliotecas3 = (DNI, numInv, nombreBib, ISBN) en:
- Socio = (<u>DNI, nombreBib</u>)
- Préstamo = (<u>DNI, nombreBib, numInv</u>)
- CopiaLibro = (<u>nombreBib</u>, <u>numInv</u>, ISBN)
- Así se resuelven los problemas anteriores.
- Pero hay que tener cuidado que al hacer esto no se pierdan relaciones entre atributos del problema.

Observaciones

- Receta: arreglar problemas de diseño introduciendo esquemas para conceptos varios que dependen de conocimiento del dominio y que son una descomposición del esquema universal
- Recordar que modelado ER consideraba la introducción de conceptos varios y por lo tanto si se hace bien se obtendrá un buen diseño relacional luego de pasar a tablas.
- La diferencia de modelado ER con lo que hicimos aquí es que hoy analizamos problemas de diseño y los fuimos eliminando uno a uno con conceptos que son introducidos.

Observaciones

- Meta: Queremos que la descomposición salga automáticamente y no que dependa de un conocedor del dominio que hace uso inteligente del mismo.
 - Esto se torna muy trabajoso si el esquema universal tiene cientos de atributos, cosa que suele pasar en la vida real.

- Ejemplo: SmaBibliotecas = (nombre, DNI, posición, numlnv, nombreBib, calle, numero, ISBN, título, autores)
- Tiene atributos redundantes nombre y posición
- El valor de DNI determina unívocamente los valores de nombre y posición y lo indicamos de la siguiente forma: DNI -> nombre, posición.
- Usamos DNI -> nombre, posición para descomponer SmaBibliotecas en:
 - SmaBibliotecas1 = (DNI, numInv, nombreBib, calle, numero, ISBN, titulo, autores)
 - (<u>DNI</u>, nombre posición)

- SmaBibliotecas1 = (DNI, numInv, nombreBib, calle, numero, ISBN, titulo, autores)
- Tiene atributos redundantes calle y número.
- nombreBib -> calle, número
- Usamos nombreBib -> calle, número para descomponer SmaBibliotecas1 en:
 - (nombreBib, calle, número)
 - SmaBibliotecas2=(DNI, numInv, nombreBib, ISBN, título, autores)

- SmaBibliotecas2=(DNI, numInv, nombreBib, ISBN, título, autores)
- Tiene atributos redundantes título y autores.
- ISBN -> título, autores
- Usamos ISBN -> título, autores para descomponer SmaBibliotecas2:
 - (ISBN, título, autores)
 - SmaBibliotecas3 = (DNI, numInv, nombreBib, ISBN)

Observaciones

- Fijarse que hemos seguido el siguiente procedimiento:
 - o los atributos de la derecha de la DF se sacan del esquema que descompongo y
 - o agrego un esquema para los atributos de toda la DF.
- Usualmente puedo ir descomponiendo el esquema universal para eliminar mucha de la redundancia de información en el mismo por medio del uso de dependencias funcionales.
- Más aun, si tenemos las DF del problema, esto es automatizable (debido a la forma repetitiva de descomponer esquemas).

- ☐ 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,
 - o el valor de un cierto conjunto de atributos determine unívocamente el valor de otro conjunto de atributos.

□ 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:

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

- ¿Las DF solo pueden ser usadas para eliminar redundancia de información?
- Ahora veremos con un ejemplo que la respuesta es no.

• **Ejemplo**: Sea el esquema universal de un banco.

```
Préstamo = (numSucursal, ciudad, activo, numCliente, numPréstamo, importe)
```

- o Lo descomponemos en:
 - SucursalCliente = (numSucursal, ciudad, activo, numCliente)
 - ClientePréstamo = (numCliente, numPréstamo, importe)
- Evaluación de la descomposición: Si tenemos un cliente con varios préstamos en distintas sucursales:
 - no se puede decir el préstamo que pertenece a cada sucursal en la descomposición que tenemos.
 - Luego en la descomposición que tenemos se perdió información con relación al esquema universal.

- Aquí se pierde relación entre atributos del problema al descomponer. Esto es algo grave y hay que evitarlo.
- El problema del ejemplo anterior se resuelve así:
- numSucursal -> ciudad, activo
- la uso para descomponer préstamo en:
- Sucursal = (numSucursal, ciudad, activo)
- prest = (numSucursal, numCliente, numPrestamo, importe)
- Nuevamente resolvimos el problema usando una DF para descomponer el universal.

- Recapitulando: usualmente que tengo un esquema con redundancia de información,
 - hay una DF con atributos a la derecha que caracterizan la información redundante y con atributos a la izquierda que no determinan todos los atributos del esquema.

Consequencias:

- Las DF ayudan a identificar redundancia de información.
- Se puede usar una DF para descomponer el esquema con redundancia de información.

Derivando Dependencias Funcionales

- Problema: ¿Cómo encontrar las dependencias funcionales (DF) de un problema del mundo real?
- 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.
 - o Estas últimas se pueden derivar de la primera.

Derivando Dependencias Funcionales

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

$$A \rightarrow H$$
 y $A \rightarrow B$, C.

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

- Necesitamos formalizar qué significa derivar. Para esto consideramos un conjunto de reglas de inferencia.
- Las reglas consideradas se llaman axiomas de Armstrong:

```
\circ if \beta \subseteq \alpha, then \alpha \to \beta (reflexividad)
```

- o if $\alpha \to \beta$, then $\gamma \alpha \to \gamma \beta$ (aumentatividad)
- o if $\alpha \to \beta$, and $\beta \to \gamma$, then $\alpha \to \gamma$ (transitividad)

- **Ejemplo**: queremos deducir $AC \rightarrow D$ a partir de las DF $\{A \rightarrow B; CB \rightarrow D\}$ usando las reglas de Armstromg.
 - Bosquejar una deducción.
- 1) A->B
- 2) CB ->D
- 3) AC -> CB (aumentatividad a 1))
- 4) AC -> D (transitividad a 3) y 2))

- Usar solo los axiomas de Armstrong es un poco pesado para hacer derivaciones.
 Y con algunos axiomas adicionales, muchas derivaciones se tornan más cortas y sencillas.
- Podemos simplificar más las derivaciones a partir de *F*, usando las siguientes reglas adicionales:
 - o If $\alpha \to \beta$ holds and $\alpha \to \gamma$ holds, then $\alpha \to \beta \gamma$ holds (union)
 - o If $\alpha \to \beta \gamma$ holds, then $\alpha \to \beta$ holds and $\alpha \to \gamma$ holds (decomposición)
 - \circ If $\alpha \to \beta$ holds and $\gamma \beta \to \delta$ holds, then $\alpha \gamma \to \delta$ holds (pseudotransitividad)

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

- 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 $f_1,...,f_n$ tales que $f_n = f$ y para todo $1 \le i \le n$:
 - 1. $f_i \in F$ ó
 - 2. f_i se obtiene por aplicar la regla de reflexividad ó
 - 3. f_i se obtiene por aplicar aumentatividad ó transitivi-dad a pasos anteriores en la lista.
- Notación: Usaremos F⊢ f para decir que f se deduce de F.

Cierre de conjunto de DF

- El cierre de un conjunto de DF F (F+) son todas las dependencias que se deducen de F.
- 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+
- ¿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+.
- Una razón es que para α hay $2^{|\alpha|}$ dependencias triviales.
- Además, si $\alpha \to \beta$ en F entonces hay $2^{|R|}$ maneras de aplicar aumentatividad a $\alpha \to \beta$ (R esquema universal).
- Como ven F+ es demasiado grande como para poder calcularlo todo.

- 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.
- Solución (viable): Quiero deducir α -> β de F. Si calculamos las dependencias de F+ con lado izquierdo α, este conjunto es mucho más chico que F+ (porque hay 2ⁿ tales α).

• **Por lo tanto**: para responder si $F \models \alpha \rightarrow \beta$, bastaría contestar

$$\alpha \rightarrow \beta \in \{\alpha \rightarrow \phi \mid F \models \alpha \rightarrow \phi\}$$
 o mejor contestar:

$$\beta \in \{ \phi \mid F \models \alpha \rightarrow \phi \}$$
 o mejor contestar:

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

 Llegamos así a un conjunto conocido como cierre de un conjunto de atributos – es el conjunto a la derecha de la inclusión.

Observaciones

- \circ Fijarse que tengo una cantidad enorme de φ: basta tomar el φ más grande del segundo conjunto y por la regla de descomposición todos los subconjuntos de φ también pertenecen al conjunto. Vamos a tener más de $2^{|φ|}$ elementos en el segundo conjunto.
- El tercer paso es muy ingenioso porque solo tenemos un subconjunto de atributos de R esquema universal y esta sí que es una cantidad manejable.
- o Entre cada paso y el siguiente se puede ver que hay un si y sólo si.
- Sea R el <u>esquema universal</u>, F conjunto de DF del problema del mundo real (con atributos en el universal), sea α ⊆ R. El cierre de α bajo F (denotado por α _F⁺) se define:

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

- 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.
 - \circ Si β \underline{C} α^+_F : la respuesta es sí, por lo tanto no agregar $\alpha \to \beta$ a F.
 - \circ Sino: $\alpha \rightarrow \beta$ no se deduce de F, por lo tanto agregamos $\alpha \rightarrow \beta$ a F.
 - \circ Por lo tanto, necesitamos un algoritmo para computar $\alpha^+_{\ \ F}$

• Algoritmo para computar α^+_F (la clausura de α bajo F) $result := \alpha;$ while (changes to result) do for each $\beta \to \gamma$ in F do begin if $\beta \subseteq result$ then $result := result \cup \gamma$ end

- Luego de cada asignación a la variable $result \subseteq \alpha^+$.
- 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.

• Ejercicio: Dados R = (A, B, C, G, H, I), $F = \{A \rightarrow B, A \rightarrow C, CG \rightarrow H, CG \rightarrow I, B \rightarrow H \text{ Calcular } A^+_F \text{ y } (AG)^+_F \}$

- Sea R esquema relacional y F es conjunto de DF, entonces α
 superclave de R si y solo si α -> R está en F+.
- α clave candidata de R si y solo si:
 - α superclave de R
 - o para todo A en α : α {A} no es superclave de R
- Para chequear que α superclave de R
 - \circ computar α^{+} , y chequear si α^{+} contiene todos los atributos de R.

• 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.

Otros conceptos

• **Definición**: $\alpha \to \beta$ se cumple en r(R) si para todo t1 \neq t2 en r: t1[\alpha] = t2[\alpha] \Rightarrow t1[\beta] = t2[\beta].

• Ejercicio: Listar algunas DF que no se cumplen y algunas DF que se cumplen

en la siguiente tabla:

Α	В	С
a1	b1	c1
a2	a1	c1
a1	b2	c2

- No confundir las DF que se cumplen en una tabla con las DF de un problema del mundo real para el cual la tabla es legal.
 - El problema suele cumplir menos DF que la tabla.

Otros conceptos

- Una DF trivial se cumple en todas las tablas de un esquema.
 - Ejemplo: En SmaBibliotecas
 - nombreBib, calle \rightarrow nombreBib
 - $calle \rightarrow calle$
- Después veremos que las DF triviales juegan su papel en los algoritmos de normalización.
- Ejercicio (de la práctica): probar la siguiente
 - o Proposición: α → β es trivial si y solo si $\beta \subseteq \alpha$.