

# DISEÑO Y OPTIMIZACIÓN. NORMALIZACIÓN

Bases de datos (GIITIN)

Grado en Ingeniería Informática en Tecnologías de la Información

Elena Montañés Roces

# Motivación

- **Objetivo:** Obtener diseños de BD adecuados
  - Evitar repetición de información (eliminar redundancias)
  - Evitar anomalías de consistencia y pérdida de información
  - Evitar código adicional
- **¿Cómo?:**
  - Mediante dependencias funcionales
  - Descomposiciones adecuadas de las relaciones
  - Diferentes grados de redundancia (Formas Normales)

# Dependencias funcionales

- Una dependencia funcional  $\alpha \rightarrow \beta$  se cumple si para todo par de tuplas cuyos valores coinciden en  $\alpha$  entonces también coinciden en  $\beta$

Cod.	Nombre	cred.	escuela
1	Inform.	130	EPI
2	Derecho	100	FD
3	Teleco.	125	EPI
4	Inform.	130	EII

titulación

- $\text{nombre} \rightarrow \text{créditos}$
- $\text{nombre} \not\rightarrow \text{escuela}$

# Cierre de un conjunto de DF

- En general, cuando se dice que se cumplen ciertas dependencias puede que implícitamente se estén cumpliendo algunas más

Cod.	Nombre	cred.	escuela
1	Inform.	130	EPI
2	Derecho	100	FD
3	Teleco.	125	EPI
4	Telemática	130	EPI

## titulación

- Se cumple  $F=\{\text{nombre} \rightarrow \text{créditos} ; \text{créditos} \rightarrow \text{escuela}\}$ . Implícitamente también se cumple  $\text{nombre} \rightarrow \text{escuela}$
- Se obtienen todas con los axiomas de Armstrong, pero el algoritmo es de orden exponencial

# Recubrimiento canónico

- Conjunto de dependencias mínimo, sin redundancias

Cod.	Nombre	cred.	escuela
1	Inform.	130	EPI
2	Derecho	100	FD
3	Teleco.	125	EPI
4	Telemática	130	EPI

titulación

- Se cumple  $F=\{\text{nombre} \rightarrow \text{créditos}; \text{créditos} \rightarrow \text{escuela}; \text{nombre} \rightarrow \text{escuela}\}$ , pero el recubrimiento canónico es  $F=\{\text{nombre} \rightarrow \text{créditos}; \text{créditos} \rightarrow \text{escuela}\}$

# Descomposición

- Las relaciones  $R_1, \dots, R_n$  es una **descomposición** de la relación  $R$  si  $R = R_1 \cup \dots \cup R_n$ 
  - La unión de sus atributos es el total
  - En general  $R_i \cap R_j \neq \emptyset$  (estarán unidas por claves ajenas)
- Siempre se cumple  $r \subseteq r_1 \bowtie \dots \bowtie r_n$ , pero en general  $r \neq r_1 \bowtie \dots \bowtie r_n$
- Descomposición **equivalente**
  - Sin pérdida de información (se cumple  $r = r_1 \bowtie \dots \bowtie r_n$ )
  - Conservando las dependencias (se mantengan las relaciones entre los atributos), es decir,
    - el cierre de  $F$  (dependencias que se cumplen en  $R$ ) y de  $F_1 \cup \dots \cup F_n$  (la unión de las dependencias que se cumplen en cada  $R_i$ ) coinciden

# D. con pérdida de información (I)

r(R)

Titul.	escuela	alumno	Asig.	créditos
Industrial	FD	José	IN	9
Telema.	EI	Juan	MP	7,5
Inform.	EI	Juan	BD	4,5
Economia	FE	Gema	ES	6

$$r_2 = \Pi_{R2}(r)$$

Titul.	escuela	alumno
Industrial	FD	José
Telema.	EI	Juan
Inform.	EI	Juan
Economia	FE	Gema

alumno	Asig.	créditos
José	IN	9
Juan	MP	7,5
Juan	BD	4,5
Gema	ES	6

# D. con pérdida de información (II)

Titul.	escuela	alumno
Industrial	FD	José
Telema.	EI	Juan
Inform.	EI	Juan
Economia	FE	Gema

alumno	Asig.	créditos
José	IN	9
Juan	MP	7,5
Juan	BD	4,5
Gema	ES	6

$$r_1 = \Pi_{R1}(r)$$

$$r_2 = \Pi_{R2}(r)$$

Titul.	escuela	alumno	Asig.	créditos
Industrial	FD	José	IN	9
Telema.	EI	Juan	MP	7,5
Telema.	EI	Juan	BD	4,5
Inform.	EI	Juan	MP	7,5
Inform.	EI	Juan	BD	4,5
Economia	FE	Gema	ES	6

$$r_1 \bowtie \dots \bowtie r_n$$

# D. con pérdida de información (III)

Titul.	escuela	alumno	Asig.	créditos
Industrial	FD	José	IN	9
Telema.	EI	Juan	MP	7,5
Telema.	EI	Juan	BD	4,5
Inform.	EI	Juan	MP	7,5
Inform.	EI	Juan	BD	4,5
Economia	FE	Gema	ES	6

$r_1 \bowtie \dots \bowtie r_n$   
tuplas espurias

Titul.	escuela	alumno	Asig.	créditos
Industrial	FD	José	IN	9
Telema.	EI	Juan	MP	7,5
Inform.	EI	Juan	BD	4,5
Economia	FE	Gema	ES	6

# D. sin pérdida de información (I)

Titul.	escuela
Industrial	FD
Telema.	EI
Inform.	EI
Economia	FE

Asig.	créditos	Titul.
IN	9	Industrial
MP	7,5	Telem.
BD	4,5	Inform.
ES	6	Economía

$$r_1 = \Pi_{R1}(r)$$

$$r_2 = \Pi_{R2}(r)$$

alumno	Asig.
José	IN
Juan	MP
Juan	BD
Gema	ES

$$r_3 = \Pi_{R3}(r)$$

$$r = r_1 \bowtie \dots \bowtie r_n$$

# Comprobación de la d. sin pérdida

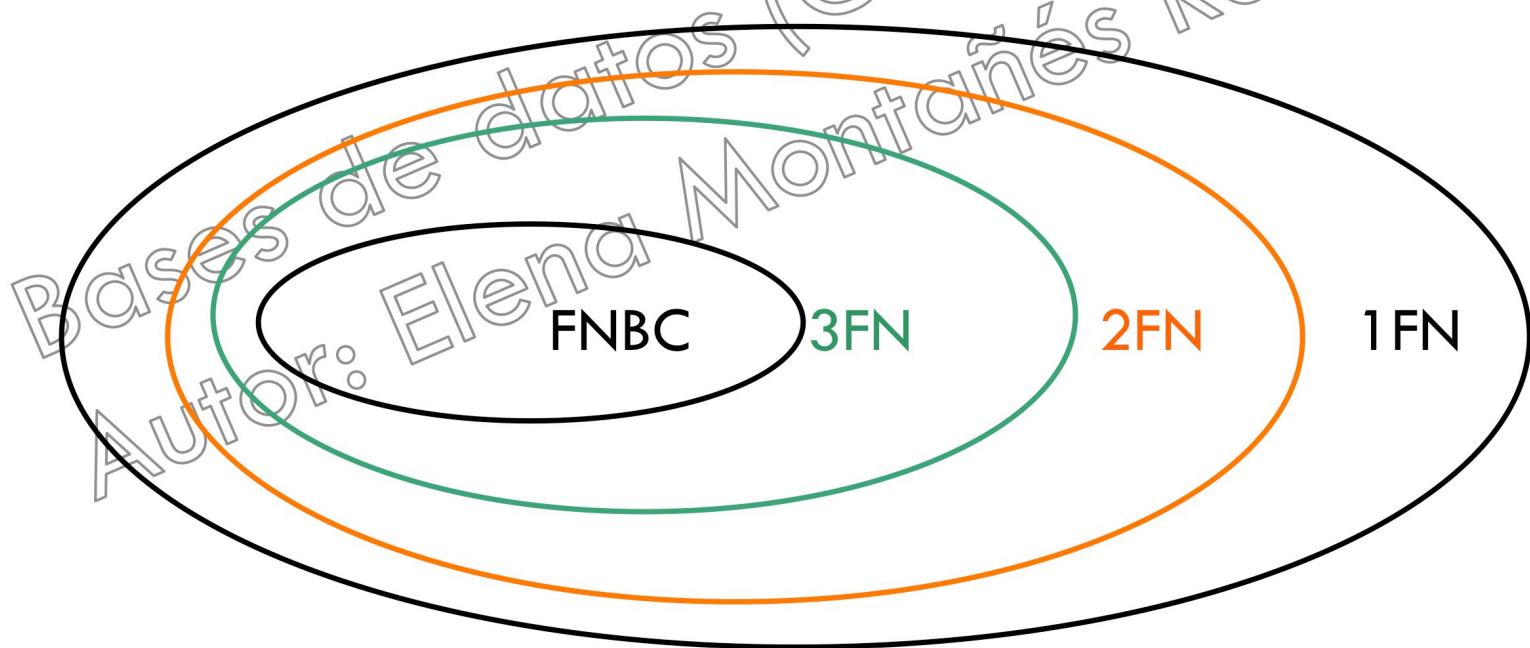
- Las **tuplas** en la base de datos **cambian** (se insertan, borran y modifican datos)
- Las dependencias funcionales se **mantienen**
- Por tanto, se la **descomposición sin pérdida** se comprueba a partir de las dependencias funcionales
  - Si  $n=2$  mediante el 1º Principio de Rissanen
  - Si  $n>2$  mediante un algoritmo

# Comprobación de la conservación de dependencias

- Comprobar directamente que el cierre de  $F$  y de  $F_1 \cup \dots \cup F_n$  coinciden, siendo  $F_i = F^+ \cap R_i$ 
  - Si no se conservan, en general ocurre que en el cierre de  $F_1 \cup \dots \cup F_n$  faltan dependencias
- Mediante un algoritmo

# Formas Normales

- Descomposición mejor
  - ▣ Elimina redundancias
- Distintos niveles de normalización



# Primera Forma Normal (1FN)

- R que cumple F está en 1FN si todos los atributos de R pertenecen a un dominio simple
- Ejemplo:

Titul.	escuela
Industrial	FD
Telema., Inform.	EI
Economia	FE

✗ No está en 1FN

Titul.	escuela
Industrial	FD
Telema.	EI
Inform.	EI
Economia	FE

✓ Sí está en 1FN

# Segunda Forma Normal (2FN)

- R que cumple F está en **2FN** si:
  - Está en 1FN
  - Cada atributo no principal (no forma parte de la clave candidata) depende de manera completa de cada clave candidata
- Ejemplo:
  - $R=(\text{titulación}, \text{escuela}, \text{alumno}, \text{asig}, \text{créditos})$  y  
 $F=\{\text{alumno}, \text{asig} \rightarrow \text{titulación}, \text{escuela}, \text{créditos};$   
 $\text{asig} \rightarrow \text{créditos};$   
 $\text{titulación} \rightarrow \text{escuela}\}$
  - Todos los atributos son simples → Está en 1FN
  - Claves candidatas: Sólo  $\{\text{alumno}, \text{asig}\}$
  - Atributos no principales: titulación, escuela, créditos
  - Créditos no depende de manera completa de la clave → No está en 2FN

# Propiedades de la 2FN

- Cualquier R **cuyos atributos sean todos principales** está en 2FN
  - $R=(\text{alumno}, \text{asig})$  y  $F=\emptyset$ . La CC es  $\{\text{alumno}, \text{asig}\}$ . Todos son principales, luego R está en 2FN
- Cualquier R **cuyas claves candidatas sean todos simples** está en 2FN
  - $R=(\text{asig}, \text{créditos}, \text{titulación}, \text{escuela})$  y  $F=\{\text{asig} \rightarrow \text{créditos}, \text{titulación}; \text{titulación} \rightarrow \text{escuela}\}$ . La CC es  $\{\text{asig}\}$  que es simple. Todos los atributos no principales dependen de manera completa de la única clave candidata, luego R está en 2FN
- Cualquier R **formada por dos atributos** está en 2FN
  - $R=(\text{alumno}, \text{asig})$  y  $F=\{\text{asig} \rightarrow \text{alumno}\}$ . La CC es  $\{\text{asig}\}$  y alumno depende de manera completa de la única clave candidata, luego R está en 2FN
  - $R=(\text{alumno}, \text{asig})$  y  $F=\{\text{alumno} \rightarrow \text{asig}\}$ . La CC es  $\{\text{alumno}\}$  y asig depende de manera completa de la única clave candidata, luego R está en 2FN

# Tercera Forma Normal (3FN)

- R que cumple F está en **3FN** si cada  $\alpha \rightarrow \beta \in F$  cumple al menos una de las sgtes. condiciones:
  - $\beta \subseteq \alpha$
  - $\alpha$  contiene una clave de R
  - Cada atributo de  $\beta - \alpha$  pertenece a alguna clave candidata de R
- Ejemplo:
  - $R = (\text{titulación}, \text{escuela}, \text{alumno}, \text{asig}, \text{créditos})$  y  
 $F = \{\text{alumno}, \text{asig} \rightarrow \text{titulación}, \text{escuela}, \text{créditos}; \text{titulación} \rightarrow \text{escuela}\}$
  - El antecedente de  $\text{alumno}, \text{asig} \rightarrow \text{titulación}, \text{escuela}, \text{créditos}$  contiene una clave de R → Cumple la 2<sup>a</sup> condición
  - En  $\text{titulación} \rightarrow \text{escuela}$ 
    - $\text{escuela}$  no está contenida en  $\text{titulación}$ ,
    - ni su antecedente ( $\text{titulación}$ ) contiene una clave de R
    - ni cada atributo del consecuente ( $\text{escuela}$ ) está contenido en ninguna clave candidata → No cumple ninguna condición,
    - Luego, esta dependencia impide que R esté en 3FN

# Propiedades de la 3FN

- Cualquier R **cuyos atributos sean todos principales** está en 3FN
  - Todas las dependencias cumplen la 3<sup>a</sup> condición
- Cualquier R **formada por dos atributos** está en 3FN
  - Si  $F=\emptyset$ . Todos los atributos son principales.
  - Si  $F\neq\emptyset$ . Todas las dependencias cumplen la 2<sup>a</sup> condición

# Forma Normal de Boyce-Codd (FNBC)

- R que cumple F está en **FNBC** si cada  $\alpha \rightarrow \beta \in F$  cumple al menos una de las sgtes. condiciones:
  - $\beta \subseteq \alpha$
  - $\alpha$  contiene una clave de R
- Ejemplo:
  - $R = (\text{titulación}, \text{escuela}, \text{alumno}, \text{asig}, \text{créditos})$  y  
 $F = \{\text{asig} \rightarrow \text{créditos}, \text{titulación}; \text{titulación} \rightarrow \text{escuela}\}$
  - Ninguna dependencia de F el consecuente está en el antecedente y ni asig ni titulación contienen una clave de R
  - Luego, ambas dependencias hacen que R no está en FNBC

# Propiedades de la FNBC

- Cualquier R formada por dos atributos está en FNBC
  - Si  $F = \emptyset$ . Todas las dependencias son triviales
  - Si  $F \neq \emptyset$ . Todas las dependencias cumplen la 2<sup>a</sup> condición

Bases de datos (GIITIN)  
Autor: Elena Montañés Roces

# Anomalías

- $R=(\text{codPf}, \text{nombrePf}, \text{categPf}, \text{codPy}, \text{cuantiaPy})$  y  
 $F=\{\text{codPf} \rightarrow \text{nombrePf}, \text{categPf}; \text{codPy} \rightarrow \text{cuantiaPy}\}$ 
  - La única clave candidata es  $\{\text{codPf}, \text{codPy}\}$  que será clave primaria
  - **Inserción:** Siempre que se añada un profesor hemos de asociarle un proyecto para que ningún atributo de la clave valga null. Idem para un proyecto.
  - **Actualización:** Si un profesor cambia de categoría debemos cambiar el valor del atributo tantas veces como proyectos tenga asignados.
  - **Borrado:** Si se borra un proyecto en el que solo haya participado un profesor entonces desaparecen también los datos del profesor.

# Anomalías cuando no está en 2FN

- $R=(\text{codPf}, \text{nombrePf}, \text{categPf}, \text{codPy}, \text{cuantiaPy})$  y  
 $F=\{\text{codPf} \rightarrow \text{nombrePf}, \text{categPf}; \text{codPy} \rightarrow \text{cuantiaPy}\}$ 
  - Clave primaria:  $\{\text{codPf}, \text{codPy}\}$

codPf	nombrePf	categPf	codPy	cuantíaPy
1	Roberto	TU	7	1000€
1	Roberto	TU	15	2000€
2	Patricia	PCD	15	2000€

- Se repite nombre y categoría del profesor por cada proyecto
- Se repite cuantía del proyecto por cada profesor

# Anomalías cuando no está en 3FN

- $R=(\text{codPf}, \text{nombrePf}, \text{categPf}, \text{codPy}, \text{cuantiaPy})$  y  
 $F=\{\text{codPf} \rightarrow \text{nombrePf}, \text{categPf}; \text{codPy} \rightarrow \text{cuantiaPy}, \text{codPf}\}$ 
  - Clave primaria:  $\{\text{codPy}\}$

codPf	nombrePf	categPf	codPy	cuantíaPy
1	Roberto	TU	7	1000€
1	Roberto	TU	15	2000€
1	Roberto	TU	20	3000€

- Se repite nombre y categoría del profesor por cada proyecto

# Anomalías cuando no está en FNBC

- $R=(\text{codPf}, \text{codPy}, \text{codBecario})$  y  $F=\{\text{codPf}, \text{codPy} \rightarrow \text{codBecario}, \text{codBecario} \rightarrow \text{codPy}\}$ 
  - Clave primaria:  $\{\text{codPf}, \text{codPy}\}$  ó  $\{\text{codPf}, \text{codBecario}\}$

codPf	codPy	codBecario
1	7	4
1	15	5
2	15	5
	20	3

- Se repite el proyecto al que está asignado un becario por cada profesor
- Como  $\text{codPf}$  siempre forma parte de la clave siempre habría que asignarle un Pf a un proyecto cuando se le asigna un becario

# Dependencias Multivaloradas

- Representación tabular de  $\alpha \rightarrow\!\!\!\rightarrow \beta$ :

	$\alpha$	$\beta$	$R-\alpha-\beta$
$t_1$	$\alpha$	$\beta_1$	$\mu_1$
$t_2$	$\alpha$	$\beta_2$	$\mu_2$
$t_3$	$\alpha$	$\beta_1$	$\mu_2$
$t_4$	$\alpha$	$\beta_2$	$\mu_1$

- La DM  $\alpha \rightarrow\!\!\!\rightarrow \beta$  quiere decir que la relación entre  $\alpha$  y  $\beta$  es independiente de la relación entre  $\alpha$  y  $R-\beta$ .

# Ejemplo

alumno	asignatura	hermano
Ignacio	SII	Rafael
Ignacio	SII	Marta
Ignacio	AT	Rafael
Ignacio	AT	Marta
Rafael	SII	Marta
Rafael	SII	Ignacio
Roberto	AT	

- Única clave candidata {alumno, asignatura, herman@}
- Dependencias funcionales triviales, está en FNBC
- Sin embargo, se repiten los hermanos que tiene un alumno por cada asignatura en la que está matriculado y viceversa.

# Cuarta Forma Normal (4FN)

- Una relación R que cumple F está en **Cuarta Forma Normal (4FN)** si cada una de las dependencias multivaloradas  $\alpha \rightarrow\!\!\!\rightarrow \beta$  de F cumple, al menos, una de las siguientes condiciones:
  - $\beta \subseteq \alpha$  ó  $\alpha \cup \beta = R$
  - $\alpha$  contiene una clave de R

# Ejemplo

- En el ejemplo anterior,
  - se cumplen las dependencias multivaloradas alumno→→asignatura y alumno→→hermano
  - se cumple que alumno no contiene una clave puesto que la clave es {alumno, asignatura, herman@}
  - Por tanto la relación no está en 4FN.

alumno	asignatura
Ignacio	SII
Ignacio	AT
Luis	SII
Roberto	AT

alumno	hermano
Ignacio	Rafael
Ignacio	Marta
Luis	Carlos
Luis	Paula

- Se ha conseguido guardar la información en 16 casillas frente a 21 originalmente.

# Dependencias jerárquicas

- Alguna proyección de R contiene alguna dependencia multivalorada

asignatura	profesor	alumno	nota
SII	Elena	Alfonso	8
SII	Jorge	Alfonso	7
SII	Elena	Nerea	9
SII	Jorge	Nerea	8
AT	Oscar	Antonio	9

- Se cumple la dependencia jerárquica asignatura  $\rightarrow \rightarrow$  profesor | alumno