Dépendances Fonctionnelles et Normalisation

Jef Wijsen

UMONS

7 novembre 2023

Exemple

Bouquet	Fleur	Couleur	Nombre	PrixFleur
Valentin	tulipe	jaune	5	100
Valentin	rose	rouge	10	150
Belge	tulipe	noir	3	100
Belge	tulipe	jaune	4	100
Belge	tulipe	rouge	6	100

$$\text{Fleur}, \text{Couleur} \rightarrow \text{PrixFleur}$$

$$\text{Fleur} \rightarrow \text{PrixFleur}$$

$$\text{Bouquet}, \text{Fleur}, \text{Couleur} \rightarrow \text{Nombre}$$

Accidentellement, la table satisfait Nombre $\rightarrow$ Bouquet.

Respect des dépendances fonctionnelles (informel)

Soit U un assemble d'attributs qui contient $A_1, \ldots, A_k, B$ (ainsi que zéro, un, ou plusieurs autres attributs).

La dépendance fonctionnelle $A_1, \ldots, A_k \rightarrow B$ (=syntaxe) exprime (=sémantique) la contrainte suivante :

> pour chaque relation R sur U, la projection
> $P := \pi_{\{A_1,\ldots,A_k,B\}}(R)$ doit respecter `P UNIQUE(`$A_1, \ldots, A_k$`)`.

Par exemple, Fleur $\rightarrow$ PrixFleur exprime que dans la projection sur $\{\text{Fleur}, \text{PrixFleur}\}$, aucune espèce de fleur ne peut apparaître plus d'une fois.

(Le syllabus utilise $\mathcal{A}$ au lieu de U.)

Décomposition

<u>Bouquet</u>	<u>Fleur</u>	<u>Couleur</u>	Nombre
Valentin	tulipe	jaune	5
Valentin	rose	rouge	10
Belge	tulipe	noir	3
Belge	tulipe	jaune	4
Belge	tulipe	rouge	6

$\text{Bouquet}, \text{Fleur}, \text{Couleur} \rightarrow \text{Nombre}$

<u>Fleur</u>	PrixFleur
tulipe	100
rose	150

$\text{Fleur} \rightarrow \text{PrixFleur}$

Où est $\text{Fleur}, \text{Couleur} \rightarrow \text{PrixFleur}$?

Conséquence logique

Soit $U = ABCDE$.

Affirmation : Si une relation respecte $AB \rightarrow C$ et $BC \rightarrow D$, elle respectera automatiquement $AB \rightarrow D$.

Pourquoi ?

Raisonnement par l'absurde. Soit R une relation qui respecte $AB \rightarrow C$ et $BC \rightarrow D$, mais qui ne respecte pas $AB \rightarrow D$. Cette relation doit contenir deux tuples t_1, t_2 comme suit, avec $d_1 \neq d_2$:

R	A	B	C	D	E	
	a	b	c_1	d_1	e_1	(t_1)
	a	b	c_2	d_2	e_2	(t_2)
			$\vdots$			

Puisque R respecte $AB \rightarrow C$, on a $c_1 = c_2$. Donc,

$$t_1[BC] = t_2[BC] = \{B : b, C : c_1\}.$$

Alors, puisque R respecte aussi $BC \rightarrow D$, on a $d_1 = d_2$, une contradiction.

Notation

$R \models AB \to C$

- Lire : R respecte (ou satisfait) $AB \to C$.
- Sémantique : il n'existe pas deux tuples t_1, t_2 dans R tels que $t_1[AB] = t_2[AB]$ et $t_1[C] \neq t_2[C]$.

$\{AB \to C, BC \to D\} \models AB \to D$

- Lire : $AB \to D$ est une conséquence logique de $\{AB \to C, BC \to D\}$.
- Sémantique : toute relation qui respecte à la fois $AB \to C$ et $BC \to D$, respectera automatiquement $AB \to D$.

Exercice

Soit

$$\Sigma = \{BC \to A, AC \to B, AE \to C, D \to BE, B \to DE, C \to E, AF \to G\}.$$

1. Prouvez :
   1. $\Sigma \models AD \to C$
   2. $\Sigma \models AB \to C$
   3. $\Sigma \models AE \to BD$
   4. $\Sigma \models AC \to D$
   5. $\Sigma \models CD \to A$
2. Calculez l'ensemble

$$\{N \in ABCDEFG \mid \Sigma \models AE \to N\}.$$

Solution

Soit

$$\Sigma = \{BC \to A, AC \to B, AE \to C, D \to BE, B \to DE, C \to E, AF \to G\}.$$

Prouvons que $\Sigma \models AE \to BD$. Raisonnement par l'absurde. Soit R une relation qui respecte toute DF de Σ, mais qui ne respecte pas $AE \to BD$. Cette relation doit contenir deux tuples t_1, t_2 comme suit, avec $b_1 \neq b_2$ ou $d_1 \neq d_2$:

$$\begin{array}{c|ccccccc}
R & A & B & C & D & E & F & G \\
\hline
 & a & b_1 & c_1 & d_1 & e & f_1 & g_1 & (t_1) \\
 & a & b_2 & c_2 & d_2 & e & f_2 & g_2 & (t_2) \\
 & & & & \vdots & & & &
\end{array}$$

À compléter pendant le cours.

Discussion

- Apparemment, il existe un algorithme pour le problème suivant :

  INPUT : Un ensemble Σ de DFs sur un ensemble U d'attributs, une DF $X \to Y$ sur U.

  QUESTION : Est-ce que $X \to Y$ est une conséquence logique de Σ ?

- Notez qu'une DF peut être exprimée en calcul relationnel. Par exemple, pour $U = ABC$:

$A \to B$	$\varphi_{A\to B} := \forall x \forall y_1 \forall y_2 \forall z_1 \forall z_2\, ((R(x, y_1, z_1) \wedge R(x, y_2, z_2)) \to y_1 = y_2)$
$B \to C$	$\varphi_{B\to C} := \forall x_1 \forall x_2 \forall y \forall z_1 \forall z_2\, ((R(x_1, y, z_1) \wedge R(x_2, y, z_2)) \to z_1 = z_2)$
$A \to C$	$\varphi_{A\to C} := \forall x \forall y_1 \forall y_2 \forall z_1 \forall z_2\, ((R(x, y_1, z_1) \wedge R(x, y_2, z_2)) \to z_1 = z_2)$

En plus, pour $\Sigma = \{A \to B, B \to C\}$, soit $\varphi_\Sigma := \varphi_{A\to B} \wedge \varphi_{B\to C}$.

Vérifiez qu'en logique des prédicats, $\varphi_{A\to C}$ est bien une conséquence logique de la formule φ_Σ.

Exercice

Soit Σ un ensemble de DFs sur l'ensemble U d'attributs.
Prouvez que les axiomes d'Armstrong sont corrects :

	Si $X \rightarrow Y$ est dans Σ, alors $\Sigma \models X \rightarrow Y$.
Réflexivité :	Si $Y \subseteq X \subseteq U$, alors $\Sigma \models X \rightarrow Y$.
Augmentation :	Si $\Sigma \models X \rightarrow Y$ et $Z \subseteq U$, alors $\Sigma \models X \cup Z \rightarrow Y \cup Z$.
Transitivité :	Si $\Sigma \models X \rightarrow Y$ et $\Sigma \models Y \rightarrow Z$, alors $\Sigma \models X \rightarrow Z$.

Dépendances fonctionnelles

Definition 1 (Dépendance fonctionnelle)

Une dépendance fonctionnelle (DF) sur $\mathcal{A}$ est une expression $X \to Y$ avec $X, Y \subseteq \mathcal{A}$.

Une relation R sur $\mathcal{A}$ satisfait (ou respecte) la DF $X \to Y$, dénoté par $R \models X \to Y$, si pour tout $s, t \in R$, si $s[X] = t[X]$, alors $s[Y] = t[Y]$.

Soit Σ un ensemble de DF sur $\mathcal{A}$.

- Une relation R sur $\mathcal{A}$ satisfait Σ, dénoté par $R \models \Sigma$, si R satisfait chaque DF dans Σ.
- Une DF $X \to Y$ sur $\mathcal{A}$ est une conséquence logique de Σ, dénoté par $\Sigma \models X \to Y$, si pour toute relation R sur $\mathcal{A}$, si $R \models \Sigma$, alors $R \models X \to Y$.

Dépendances fonctionnelles

Definition 1 (Dépendance fonctionnelle)

Une dépendance fonctionnelle (DF) sur $\mathcal{A}$ est une expression $X \to Y$ avec $X, Y \subseteq \mathcal{A}$.

Une relation R sur $\mathcal{A}$ satisfait (ou respecte) la DF $X \to Y$, dénoté par $R \models X \to Y$, si pour tout $s, t \in R$, si $s[X] = t[X]$, alors $s[Y] = t[Y]$.

Soit Σ un ensemble de DF sur $\mathcal{A}$.

- Une relation R sur $\mathcal{A}$ satisfait Σ, dénoté par $R \models \Sigma$, si R satisfait chaque DF dans Σ.
- Une DF $X \to Y$ sur $\mathcal{A}$ est une conséquence logique de Σ, dénoté par $\Sigma \models X \to Y$, si pour toute relation R sur $\mathcal{A}$, si $R \models \Sigma$, alors $R \models X \to Y$.

Dépendances fonctionnelles

Definition 1 (Dépendance fonctionnelle)

Une dépendance fonctionnelle (DF) sur $\mathcal{A}$ est une expression $X \to Y$ avec $X, Y \subseteq \mathcal{A}$.

Une relation R sur $\mathcal{A}$ satisfait (ou respecte) la DF $X \to Y$, dénoté par $R \models X \to Y$, si pour tout $s, t \in R$, si $s[X] = t[X]$, alors $s[Y] = t[Y]$.

Soit Σ un ensemble de DF sur $\mathcal{A}$.

- Une relation R sur $\mathcal{A}$ satisfait Σ, dénoté par $R \models \Sigma$, si R satisfait chaque DF dans Σ.
- Une DF $X \to Y$ sur $\mathcal{A}$ est une conséquence logique de Σ, dénoté par $\Sigma \models X \to Y$, si pour toute relation R sur $\mathcal{A}$, si $R \models \Sigma$, alors $R \models X \to Y$.

Équivalence

Definition 2 (Équivalence)

Soit Σ_1, Σ_2 deux ensembles de DFs sur $\mathcal{A}$.

- On écrira $\Sigma_1 \models \Sigma_2$ si pour toute DF $X \to Y$ dans Σ_2, on a $\Sigma_1 \models X \to Y$.
- On dira que Σ_1 et Σ_2 sont équivalents, dénoté par $\Sigma_1 \equiv \Sigma_2$, si $\Sigma_1 \models \Sigma_2$ et $\Sigma_2 \models \Sigma_1$.

Example 3

$\{A \to B, B \to C\} \equiv \{A \to ABC, B \to BC, AB \to C\}$

Fermeture d'un ensemble d'attributs

$$X^{*,\Sigma} := \{A \in \mathcal{A} \mid \Sigma \models X \to A\}$$

Il est facile de voir que

$$\Sigma \models X \to Y \iff Y \subseteq X^{*,\Sigma}.$$

Algorithme pour calculer la fermeture

Algorithme 1

Entrée : un ensemble Σ *de DFs, un ensemble* X *d'attributs*
Sortie : $X^{*,\Sigma}$

$\texttt{NonUtilisé} := \Sigma$
$\texttt{Fermeture} := X$
répéter aussi longtemps que $\texttt{NonUtilisé}$ *change*
 si $W \to Z \in \texttt{NonUtilisé}$ *et* $W \subseteq \texttt{Fermeture}$
 alors i. $\texttt{NonUtilisé} := \texttt{NonUtilisé} \setminus \{W \to Z\}$
 ii. $\texttt{Fermeture} := \texttt{Fermeture} \cup Z$
retourner $\texttt{Fermeture}$

L'Algorithme 1 calcule correctement $X^{*,\Sigma}$

À la fin,

- $\boxed{\texttt{Fermeture} \subseteq X^{*,\Sigma}}$ Facile à prouver.
- $\boxed{X^{*,\Sigma} \subseteq \texttt{Fermeture}}$
  Par contraposition, supposons $A \notin \texttt{Fermeture}$.
  Il faut montrer $A \notin X^{*,\Sigma}$, i.e., $\Sigma \not\models X \to A$.
  Peut-on construire une relation qui satisfait Σ, mais pas $X \to A$?

attributs de Fermeture			autres attributs		
A_1	$\dots$	A_ℓ	$A_{\ell+1}$	$\dots$	A_n
0	$\cdots$	0	0	$\cdots$	0
0	$\cdots$	0	1	$\cdots$	1

Cette relation satisfait Σ (à prouver) mais ne satisfait pas $X \to A$ (parce que $X \subseteq \texttt{Fermeture}$ après l'initialisation et, par hypothèse, $A \notin \texttt{Fermeture}$).

L'Algorithme 1 calcule correctement $X^{*,\Sigma}$

À la fin,

- $\boxed{\texttt{Fermeture} \subseteq X^{*,\Sigma}}$ Facile à prouver.
- $\boxed{X^{*,\Sigma} \subseteq \texttt{Fermeture}}$
  Par contraposition, supposons $A \notin \texttt{Fermeture}$.
  Il faut montrer $A \notin X^{*,\Sigma}$, i.e., $\Sigma \not\models X \to A$.
  Peut-on construire une relation qui satisfait Σ, mais pas $X \to A$?

$$\begin{array}{|cccccc}
\multicolumn{3}{c}{\overbrace{\hspace{4cm}}^{\text{attributs de }\texttt{Fermeture}}} & \multicolumn{3}{c}{\overbrace{\hspace{4cm}}^{\text{autres attributs}}} \\
A_1 & \dots & A_\ell & A_{\ell+1} & \dots & A_n \\
\hline
0 & \cdots & 0 & 0 & \cdots & 0 \\
0 & \cdots & 0 & 1 & \cdots & 1
\end{array}.$$

Cette relation satisfait Σ (à prouver) mais ne satisfait pas $X \to A$ (parce que $X \subseteq \texttt{Fermeture}$ après l'initialisation et, par hypothèse, $A \notin \texttt{Fermeture}$).

Theorème de Heath

Proposition 1 (Theorème de Heath)

Soit R une relation sur $\mathcal{A}$ telle que $R \models X \to Y$ et $Z = \mathcal{A} \setminus XY$.
Alors, $R = \pi_{XY}(R) \bowtie \pi_{XZ}(R)$.

Théorème de Heath : preuve pour 3 attributs

Hypothèse : R est une relation sur ABC qui respecte $B \to C$.

Prouvez : $R = (\pi_{AB}(R)) \bowtie (\pi_{BC}(R))$.

Preuve de $R \subseteq \pi_{AB}(R) \bowtie \pi_{BC}(R)$. Exercice.

Preuve de $\pi_{AB}(R) \bowtie \pi_{BC}(R) \subseteq R$.

Soit $\{A : a, B : b, C : c\}$ un tuple dans $\pi_{AB}(R) \bowtie \pi_{BC}(R)$.

$\implies \{A : a, B : b\} \in \pi_{AB}(R)$ et $\{B : b, C : c\} \in \pi_{BC}(R)$.

$\implies \exists c' \exists a' : R$ contient $\{A : a, B : b, C : c'\}$ et $\{A : a', B : b, C : c\}$.

Puisque $R \models B \to C$, on a $c = c'$.

Donc, R contient $\{A : a, B : b, C : c\}$.

Dépendance de jointure : Exemple 1

R	A	B	C
	a	b	c'
	a'	b	c

$\pi_{AB}(R)$	A	B
	a	b
	a'	b

$\pi_{BC}(R)$	B	C
	b	c'
	b	c

$\pi_{AB}(R) \bowtie \pi_{BC}(R)$	A	B	C
	a	b	c'
	a'	b	c
	a'	b	c'
	a	b	c

On dira que R ne respecte pas $\bowtie [AB, BC]$.

Dépendance de jointure : Exemple 2

R	A	B	C
	1	a	△
	1	b	□
	2	b	△

$\pi_{AB}(R)$	A	B
	1	a
	1	b
	2	b

$\pi_{AC}(R)$	A	C
	1	△
	1	□
	2	△

$\pi_{BC}(R)$	B	C
	a	△
	b	□
	b	△

$\pi_{AB}(R) \bowtie \pi_{AC}(R) \bowtie \pi_{BC}(R)$	A	B	C
	1	a	△
	1	b	□
	2	b	△
	1	b	△

On dira que R ne respecte pas $\bowtie [AB, AC, BC]$.

Dépendance de jointure : Exemple 3

S	A	B	C
	1	a	$\triangle$
	1	b	$\bigcirc$
	1	b	$\triangle$
	2	b	$\triangle$

$\pi_{AB}(S)$	A	B
	1	a
	1	b
	2	b

$\pi_{AC}(S)$	A	C
	1	$\triangle$
	1	$\bigcirc$
	2	$\triangle$

$\pi_{BC}(S)$	B	C
	a	$\triangle$
	b	$\bigcirc$
	b	$\triangle$

$\pi_{AB}(S) \bowtie \pi_{AC}(S) \bowtie \pi_{BC}(S)$	A	B	C
	1	a	$\triangle$
	1	b	$\bigcirc$
	1	b	$\triangle$
	2	b	$\triangle$

On dira que S respecte (ou satisfait) $\bowtie [AB, AC, BC]$.

Dépendance de jointure

Definition 4 (DJ)

Une dépendance de jointure (DJ) sur $\mathcal{A}$ est une expression

$$\bowtie [X_1, X_2, \ldots, X_\ell]$$

avec $\ell \geq 1$ et $\mathcal{A} = \bigcup_{i=1}^{\ell} X_i$.

Une relation R sur $\mathcal{A}$ satisfait cette DJ si

$$R = \pi_{X_1}(R) \bowtie \pi_{X_2}(R) \bowtie \cdots \bowtie \pi_{X_\ell}(R).$$

Dépendance de jointure : Exemple

B Bouquet	Fleur	Couleur
Campagnard	bleuet	bleu
Campagnard	tournesol	jaune
Belge	tulipe	rouge
Belge	tulipe	jaune
Belge	rose	rouge
Belge	rose	jaune

Cette table ne respecte pas

$$\bowtie [\{\text{Bouquet}, \text{Fleur}\}, \{\text{Bouquet}, \text{Couleur}\}],$$

mais respecte bien [1]

$$\bowtie [\{\text{Bouquet}, \text{Fleur}\}, \{\text{Bouquet}, \text{Couleur}\}, \{\text{Fleur}, \text{Couleur}\}].$$

1. Noter : Si l'on “décompose” et “ré-compose” avec 3 composants , la combinaison (Campagnard, bleuet, jaune) n'apparaîtra pas à cause du composant $\{\text{Fleur}, \text{Couleur}\}$.

Dépendance de jointure : Exemple

B	Bouquet	Fleur	Couleur
	Belge	tulipe	jaune
	Belge	rose	rouge
	Valentin	tulipe	rouge

Cette table ne respecte pas

$$\bowtie [\{\text{Bouquet}, \text{Fleur}\}, \{\text{Bouquet}, \text{Couleur}\}, \{\text{Fleur}, \text{Couleur}\}],$$

parce qu'elle ne contient pas la combinaison (Belge, tulipe, rouge).

Dépendance multivaluée : Exemple

Une dépendance multivaluée est une dépendance de jointure avec exactement 2 composants.

B	Bouquet	Fleur	Couleur
	Belge	tulipe	rouge
	Belge	tulipe	jaune
	Belge	rose	rouge
	Belge	rose	jaune
	Valentin	rose	rouge
	Valentin	rose	jaune

Cette table respecte

$$\bowtie [\{\text{Bouquet}, \text{Fleur}\}, \{\text{Bouquet}, \text{Couleur}\}],$$

Trois notations possibles : $\bowtie [\{\text{Bouquet}, \text{Fleur}\}, \{\text{Bouquet}, \text{Couleur}\}]$

Bouquet $\twoheadrightarrow$ Fleur

Bouquet $\twoheadrightarrow$ Couleur

Dépendance multivaluée

Definition 5 (DMV)

Une dépendance multivaluée (DMV) sur $\mathcal{A}$ est une expression

$$X \twoheadrightarrow Y$$

avec $X, Y \subseteq \mathcal{A}$.

Une relation R sur $\mathcal{A}$ satisfait cette DMV si $R \models \bowtie [XY, XZ]$ avec $Z = \mathcal{A} \setminus XY$.
Une DMV est donc un cas spécial d'une DJ.

Dans l'autre direction, on peut vérifier qu'une DJ $\bowtie [X_1, X_2]$ avec deux composants est équivalente à $X_1 \cap X_2 \twoheadrightarrow X_1$ (ce qui est équivalent à $X_1 \cap X_2 \twoheadrightarrow X_1 \setminus X_2$).

Prouvez que les cinq expressions suivantes sont toutes équivalentes :

$$\bowtie [AB, AC],\ A \twoheadrightarrow B,\ A \twoheadrightarrow C,\ A \twoheadrightarrow AB,\ A \twoheadrightarrow AC.$$

Interaction

Voici deux façons alternatives pour exprimer le Théorème de Heath :

- $\{X \rightarrow Y\} \models X \twoheadrightarrow Y$, ou
- $\{X \rightarrow Y\} \models \bowtie [XY, XZ]$ avec $Z = \mathcal{A} \setminus XY$.

Proposition 2

Soit $\mathcal{A}$ un ensemble d'attributs, et soit $\{X, Y, Z\}$ une partition de $\mathcal{A}$.
Soit Σ un ensemble de DF sur $\mathcal{A}$.
Alors $\Sigma \models \bowtie [XY, XZ]$ si et seulement si $\Sigma \models X \rightarrow Y$ ou $\Sigma \models X \rightarrow Z$.

Décomposition en BCNF

Bouquet	Fleur	Couleur	Nombre	PrixFleur
Valentin	tulipe	jaune	5	100
Valentin	rose	rouge	10	150
Belge	tulipe	noir	3	100
Belge	tulipe	jaune	4	100
Belge	tulipe	rouge	6	100

$$\text{Fleur}, \text{Couleur} \rightarrow \text{PrixFleur}$$
$$\text{Fleur} \rightarrow \text{PrixFleur}$$
$$\text{Bouquet}, \text{Fleur}, \text{Couleur} \rightarrow \text{Nombre}$$

<u>Bouquet</u>	<u>Fleur</u>	<u>Couleur</u>	Nombre
Valentin	tulipe	jaune	5
Valentin	rose	rouge	10
Belge	tulipe	noir	3
Belge	tulipe	jaune	4
Belge	tulipe	rouge	6

$$\text{Bouquet}, \text{Fleur}, \text{Couleur} \rightarrow \text{Nombre}$$

<u>Fleur</u>	PrixFleur
tulipe	100
rose	150

$$\text{Fleur} \rightarrow \text{PrixFleur}$$

Où est $\text{Fleur}, \text{Couleur} \rightarrow \text{PrixFleur}$?

Boyce-Codd Normal Form

Definition 6 (BCNF)

Un schéma-DF est un couple $(\mathcal{A}, \Sigma)$ avec $\mathcal{A}$ un schéma-de-relation et Σ un ensemble de DFs sur $\mathcal{A}$.

Un schéma-DF $(\mathcal{A}, \Sigma)$ est en BCNF (Boyce-Codd Normal Form) si pour toute DF $X \to Y \in \Sigma$ telle que $Y \not\subseteq X$, on a $\Sigma \models X \to \mathcal{A}$.

Definition 7

Une décomposition du schéma-DF $(\mathcal{A}, \Sigma)$ est un ensemble

$$\{(\mathcal{A}_1, \Sigma_1), \ldots, (\mathcal{A}_\ell, \Sigma_\ell)\}$$

tel que

1. $\mathcal{A} = \bigcup_{i=1}^{\ell} \mathcal{A}_i$;
2. $\Sigma \models \bowtie [\mathcal{A}_1, \ldots, \mathcal{A}_\ell]$ (préservation du contenu) ;
3. pour $i \in \{1, \ldots, \ell\}$, $\Sigma_i \equiv \{X \rightarrow Y \mid XY \subseteq \mathcal{A}_i, \Sigma \models X \rightarrow Y\}$.

On dira que cette décomposition préserve les DFs si $\Sigma \equiv \bigcup_{i=1}^{\ell} \Sigma_i$.
Chaque élément $(\mathcal{A}_i, \Sigma_i)$ est appelé un composant de la décomposition.

Exemple

Une décomposition de

$$(ABCD, \overbrace{\{A \rightarrow B, B \rightarrow C, AB \rightarrow C, ABC \rightarrow D\}}^{\Sigma})$$

est

$$\{(ABD, \{A \rightarrow B, A \rightarrow D\}), \quad (BC, \{B \rightarrow C\})\}$$

Pourquoi ?

1. $ABCD = ABD \cup BC$
2. Le théorème de Heath nous dit que $\{B \rightarrow C\} \models \bowtie [ABD, BC]$.
3. Noter que $\Sigma \equiv \{A \rightarrow B, B \rightarrow C, A \rightarrow D\}$.

Noter que cette décomposition préserve les DFs.

Exercice : Quels des trois schémas sont en BCNF ?

Exemple

Prof	Cours	Heure
Albert Einstein	Physique I	lundi, 8H15
Albert Einstein	Physique I	mercredi, 10H15
Albert Einstein	Mécanique quantique	lundi, 10H15
Marie Curie	Chimie II	lundi, 8H15

$$\text{Cours} \rightarrow \text{Prof}$$

$$\text{Prof}, \text{Heure} \rightarrow \text{Cours}$$

Décomposition en BCNF

Proposition 3

Chaque schéma-DF a une décomposition en BCNF.

Soit

$$\begin{aligned} \mathcal{A} &= \{\text{Prof}, \text{Cours}, \text{Heure}\}, \\ \Sigma &= \{\text{Cours} \rightarrow \text{Prof}, \{\text{Prof}, \text{Heure}\} \rightarrow \text{Cours}\}. \end{aligned}$$

Ce schéma-DF n'est pas en BCNF.
Manifestement, puisque la DF $\{\text{Prof}, \text{Heure}\} \rightarrow \text{Cours}$ contient les trois attributs du schéma-DF, il n'existe pas de décomposition en BCNF qui préserve les DFs.

3NF

Definition 8 (3NF)

Un sous-ensemble K de $\mathcal{A}$ est appelée une clé d'un schéma-DF $(\mathcal{A}, \Sigma)$ si K est un ensemble minimal (par rapport à $\subseteq$) tel que pour tout $A \in \mathcal{A}$, $\Sigma \models K \to A$. Un attribut A est appelé premier s'il appartient à une ou plusieurs clés.

Un schéma-DF $(\mathcal{A}, \Sigma)$ est en 3NF si pour toute DF $X \to Y \in \Sigma$ telle que $Y \not\subseteq X$, soit $\Sigma \models X \to \mathcal{A}$, soit chaque attribut de $Y \setminus X$ est premier (soit les deux).

Proposition 4

Chaque schéma-DF a une décomposition en 3NF qui préserve les DFs.

3NF : Définition (un peu) plus pratique

Une DF $X \to \{B\}$ avec B un attribut tel que $B \notin X$ est appelée singulière.

On peut supposer sans perte de généralité que Σ ne contient que des DFs singulières.

Le schéma-DF $(\mathcal{A}, \Sigma)$ est alors en 3NF si pour toute DF $X \to \{B\} \in \Sigma$, soit $\Sigma \models X \to \mathcal{A}$, soit B est premier (soit les deux).

Intuition derrière la définition de 3NF

Pour le schéma

$$\begin{aligned}\mathcal{A} &= \{\text{Prof}, \text{Cours}, \text{Heure}\},\\ \Sigma &= \{\text{Cours} \rightarrow \text{Prof}, \{\text{Prof}, \text{Heure}\} \rightarrow \text{Cours}\},\end{aligned}$$

les clés sont $\{\text{Cours}, \text{Heure}\}$ et $\{\text{Prof}, \text{Heure}\}$.

Le schéma est bien en 3NF. Notez en particulier que pour Cours $\rightarrow$ Prof, on “utilise” que Prof est premier.

Contrairement à BCNF, 3NF ne nous impose pas de décomposer en $\{\text{Cours}, \text{Prof}\}$ et $\{\text{Cours}, \text{Heure}\}$. Cette décomposition est problématique car elle “perd” $\{\text{Prof}, \text{Heure}\} \rightarrow \text{Cours}$.

Décomposition en 3NF

Definition 9 (Schéma-DF)

Un schéma-DF $(\mathcal{A}, \Sigma)$ est irréductible si les trois conditions suivantes sont respectées :

1. Chaque DF dans Σ est singulière.
2. Pour chaque DF $X \to A \in \Sigma$ et $B \in X$,
$$(\Sigma \setminus \{X \to A\}) \cup \{(X \setminus B) \to A\} \not\equiv \Sigma.$$
3. Pour chaque DF $X \to A \in \Sigma$, $(\Sigma \setminus \{X \to A\}) \not\equiv \Sigma$.

Proposition 5

Soit $(\mathcal{A}, \Sigma)$ un schéma-DF irréductible avec $\Sigma = \{X_1 \to A_1, \ldots, X_\ell \to A_\ell\}$. Soit K une clé pour $(\mathcal{A}, \Sigma)$. L'ensemble suivant est une décomposition de $(\mathcal{A}, \Sigma)$ en 3NF qui préserve les DFs :
$$\{(K, \Sigma_0), (X_1 A_1, \Sigma_1), \ldots, (X_\ell A_\ell, \Sigma_\ell)\},$$
avec $\Sigma_0, \Sigma_1, \ldots, \Sigma_\ell$ comme spécifiés dans la Définition 7.

Décomposition en 3NF : Exemple

Considérez

$$(ABCD, \overbrace{\{A \rightarrow B, B \rightarrow C, AB \rightarrow C, ABC \rightarrow D\}}^{\Sigma}).$$

Une version irréductible de Σ est :

$$\{A \rightarrow B, A \rightarrow D, B \rightarrow C\}.$$

Une clé est $\{A\}$.
Proposition 5 nous donne la décomposition :

$$\{(A, \emptyset), \quad (AB, \{A \rightarrow B\}), \quad (AD, \{A \rightarrow D\}), \quad (BC, \{B \rightarrow C\})\}.$$

Notez qu'une autre décomposition en 3NF est :

$$\{(ABD, \{A \rightarrow B, A \rightarrow D\}), \quad (BC, \{B \rightarrow C\})\}$$

Décomposition en 3NF : Exemple

Pour le schéma

$$\begin{array}{rcl} \mathcal{A} & = & \{\text{Bouquet}, \text{Fleur}, \text{Prix}\}, \\ \Sigma & = & \{\text{Fleur} \rightarrow \text{Prix}\}, \end{array}$$

la seule clé est $\{\text{Bouquet}, \text{Fleur}\}$.

$(\mathcal{A}, \Sigma)$ est irréductible.
La décomposition en 3NF de la Proposition 5 est :

$$\left\{ \begin{array}{l} (\{\text{Bouquet}, \text{Fleur}\}, \emptyset), \\ (\{\text{Fleur}, \text{Prix}\}, \{\text{Fleur} \rightarrow \text{Prix}\}) \end{array} \right\}$$

5NF : Exemple

Un fleuriste affirme : "dans chacun de mes bouquets, chaque fleur du bouquet doit toujours être disponible en chaque couleur du bouquet, pourvu que la fleur existe dans la couleur". En d'autres termes :

$$\bowtie [\{\text{Bouquet}, \text{Fleur}\}, \{\text{Bouquet}, \text{Couleur}\}, \{\text{Fleur}, \text{Couleur}\}].$$

B	Bouquet	Fleur	Couleur
	Campagnard	bleuet	bleu
	Campagnard	tournesol	jaune
	Belge	tulipe	rouge
	Belge	tulipe	jaune

5NF nous impose de décomposer en trois composants :

Bouquet	Fleur
Campagnard	bleuet
Campagnard	tournesol
Belge	tulipe

Bouquet	Couleur
Campagnard	bleu
Campagnard	jaune
Belge	rouge
Belge	jaune

Fleur	Couleur
bleuet	bleu
tournesol	jaune
tulipe	rouge
tulipe	jaune

5NF

Definition 10

Un schéma-DF-DJ est un couple $(\mathcal{A}, \Xi)$ avec $\mathcal{A}$ un ensemble d'attributs et Ξ un ensemble qui contient des DFs et DJs sur $\mathcal{A}$.

Un schéma-DF-DJ $(\mathcal{A}, \Xi)$ est en 5NF s'il existe un schéma-DF [a] $(\mathcal{A}, \Sigma)$ en BCNF tel que $\Sigma \equiv \Xi$.

a. Donc, tout élément de Σ est une DF.

Intuition derrière la définition de 5NF

Pays	PSurface	Capital	CPopulation
Belgique	30.689	Bruxelles	176.543
Italie	301.338	Rome	2.860.009

$$\Xi = \left\{ \begin{array}{l} \text{Pays} \rightarrow \text{PSurface}, \text{Capital}, \text{CPopulation} \\ \text{Capital} \rightarrow \text{Pays} \\ \bowtie [\{\text{Pays}, \text{Capital}\}, \{\text{Pays}, \text{PSurface}\}, \{\text{Capital}, \text{CPopulation}\}] \end{array} \right\}$$

$$\Sigma = \left\{ \begin{array}{l} \text{Pays} \rightarrow \text{PSurface}, \text{Capital}, \text{CPopulation} \\ \text{Capital} \rightarrow \text{Pays} \end{array} \right\}$$

Puisque Σ est en BCNF et $\Sigma \equiv \Xi$, aucune décomposition ne s'impose : le schéma est bien en 5NF.

Exercice

Soit

$$\Sigma = \left\{ \begin{array}{l} \text{Pays} \rightarrow \text{PSurface}, \text{Capital}, \text{CPopulation} \\ \text{Capital} \rightarrow \text{Pays} \end{array} \right\}.$$

Prouvez :

$$\Sigma \models \bowtie [\{\text{Pays}, \text{Capital}\}, \{\text{Pays}, \text{PSurface}\}, \{\text{Capital}, \text{CPopulation}\}].$$

Esquisse de preuve. Essayons d'avoir un tuple t *"de trop"* dans la jointure des projections :

Pays	PSurface	Capital	CPopulation
Belgique	s_1	Bruxelles	u_1
Belgique	30.689	c_2	u_2
p_3	s_3	Bruxelles	176.543

La jointure des projections contient $t =$(Belgique, 30.689, Bruxelles, 176.543). Mais Σ impose $s_1 = 30.689$, $c_2 =$ Bruxelles, $p_3 =$ Belgique et finalement $u_1 = 176.543$. Donc, t n'est pas *"de trop"*.

Exercice

Pour le schéma $ABCDE$, démontrez

$$\{A \to BD, AB \to C, AD \to E, D \to C, DE \to A\} \not\models \bowtie [AB, AD, BE, CD].$$

4NF : Exemple

Un fleuriste tordu affirme : “dans chacun de mes bouquets, chaque fleur du bouquet doit toujours être disponible en chaque couleur du bouquet”.[2]
En d'autres termes :

$$\bowtie [\{\text{Bouquet}, \text{Fleur}\}, \{\text{Bouquet}, \text{Couleur}\}].$$

B

Bouquet	Fleur	Couleur
Belge	tulipe	rouge
Belge	tulipe	jaune
Belge	rose	rouge
Belge	rose	jaune
Valentin	rose	rouge
Valentin	rose	jaune

4NF nous impose de décomposer en deux composants :

Bouquet	Fleur
Belge	tulipe
Belge	rose
Valentin	rose

Bouquet	Couleur
Belge	rouge
Belge	jaune
Valentin	rouge
Valentin	jaune

2. Notez que cette contrainte est vraisemblablement trop contraignante : comment inclure un tournesol et un bleuet dans un même bouquet ?

4NF

Definition 11

Un schéma-DF-DMV est un couple $(\mathcal{A}, \Xi)$ avec $\mathcal{A}$ un ensemble d'attributs et Ξ un ensemble qui contient des DFs et DMVs sur $\mathcal{A}$.

Un schéma-DF-DMV $(\mathcal{A}, \Xi)$ est en 4NF s'il existe un schéma-DF [a] $(\mathcal{A}, \Sigma)$ en BCNF tel que $\Sigma \equiv \Xi$.

a. Donc, tout élément de Σ est une DF.

Clé versus clé primaire

Clé $=$

- Notion définie par rapport à un schéma-DF $(\mathcal{A}, \Sigma)$.
- Un ensemble $K \subseteq \mathcal{A}$ tel que
  1. $\Sigma \models K \to \mathcal{A}$, et
  2. *Condition de minimalité :* pour tout ensemble $K' \subsetneq K$, $\Sigma \not\models K' \to \mathcal{A}$.

Clé primaire $=$ Une contrainte `R PRIMARY KEY(`K`)` qui est respectée s'il n'y a pas deux tuples distincts ayant la même valeur pour (tous les attributs de) K. Pas de condition théorique de minimalité.

Quelle est la contrepartie théorique de la notion de clé étrangère ?

Dépendances d'inclusion

$\boxed{\text{Syntaxe}}$ Soit $\mathcal{R}$ un schéma-de-base-de données.
Une dépendance d'inclusion (DI) sur $\mathcal{R}$ est une expression

$$R[A_1, \ldots, A_k] \subseteq S[B_1, \ldots, B_k]$$

où

- $k \geq 1$;
- R et S sont des noms de relation dans $\mathcal{R}$ (il se peut que $R = S$) ;
- $A_1, \ldots, A_k$ sont des attributs distincts dans $\mathit{sorte}(R)$;
- $B_1, \ldots, B_k$ sont des attributs distincts dans $\mathit{sorte}(S)$.

$\boxed{\text{Sémantique}}$ Une base de données $\mathcal{I}$ sur le schéma $\mathcal{R}$ satisfait cette DI si pour tout $r \in R^{\mathcal{I}}$, il existe $s \in S^{\mathcal{I}}$ tel que pour tout $i \in \{1, \ldots, k\}$, $r(A_i) = s(B_i)$.

Dépendances d'inclusion : Exemple

Soit $\mathcal{I}$ la base de données contenant les relations suivantes :

R	A	B
	1	2
	2	3
	3	2

S	C	D
	2	1
	2	3

Cette base de données satisfait les DIs suivantes :

$$\begin{aligned} R[B] &\subseteq R[A] \\ S[C, D] &\subseteq R[B, A] \end{aligned}$$

Conséquence logique

Soit $\mathcal{R} = \{R\}$ un schéma-de-base-de-données avec $sorte(R) = \{A, B\}$.

Soit $\Sigma = \{A \rightarrow B, R[A] \subseteq R[B]\}$.

Argumentez que $\Sigma \models R[B] \subseteq R[A]$, c'est-à-dire, qu'il n'existe pas de relation finie [3] qui respecte Σ mais qui ne respecte pas $R[B] \subseteq R[A]$.

Proposition 6

Il n'existe pas d'algorithme pour le problème suivant :

Entrée : Une DF σ et un ensemble Σ contenant des DFs et des DIs.

Question : Est-ce que $\Sigma \models \sigma$?

Rappel : Ce problème est "facile" si Σ ne contient que des DFs.

3. En théorie des bases de données, on impose toujours que toute relation soit finie.