LIF10 – FONDEMENTS DES BASES DE DONNÉES
TD4 – fermeture d'un ensemble d'attributs

Licence informatique – Automne 2013–2014

Exercice 1 : dépendances d'inclusion

Soit $I = \{R[ABC] \subseteq S[EFG], S[F] \subseteq T[J], S[FE] \subseteq T[LK]\}$ l'ensemble de dépendances d'inclusion sur le schéma de bases de données $\mathbf{D} = \{R, S, T\}$ avec $R = ABCD$, $S = EFGH$ et $T = IJKL$.

1. Prouver $I \models R[AB] \subseteq T[KL]$ à l'aide du système de Casanova donné en figure 1, en précisant σ quand la règle de permutation et projection est utilisée.
2. Enumérer les différentes DI que l'on peut produire par application de la règle de projection et permutation sur $R[ABC] \subseteq S[EFG]$. Les dénombrer dans le cas général.
3. Prouver la propriété suivante d'intéraction entre DFs et DIs

$$\{R[XY] \subseteq S[TU], S : T \to U\} \models R : X \to Y$$

4. Même question pour la propriété

$$\{R[XY] \subseteq S[TU], R[XZ] \subseteq S[TV], S : T \to U\} \models R[XYZ] \subseteq S[TUV]$$

$$\frac{}{R[X] \subseteq R[X]}\ \gamma_R \text{ (réflexivité)} \qquad \frac{R[X] \subseteq S[Y] \qquad S[Y] \subseteq T[Z]}{R[X] \subseteq T[Z]}\ \gamma_T \text{ (transitivité)}$$

$$\frac{R[A_1 \ldots A_n] \subseteq S[B_1 \ldots B_n]}{S[A_{\sigma(1)} \ldots A_{\sigma(k)}] \subseteq S[B_{\sigma(1)} \ldots B_{\sigma(k)}]}\ \gamma_P \text{ (permutation \& projection)}$$

avec σ une permutation d'un sous-ensemble de $\{1 \ldots n\}$

FIGURE 1 – Axiomatisation de Casanova pour les DIs

	A	B	C	D	E
t_1	1	1	85	Biochimie	5
t_2	1	5	94	Admission	12
t_3	2	2	92	Informatique	2
t_4	3	2	98	Informatique	2
t_5	4	3	98	Géophysique	2
t_6	5	1	75	Biochimie	5
t_7	6	5	88	Admission	12

TABLE 1 – Affectation des employés à un département

Exercice 2 : dépendances fonctionnelles satisfaites par une instance

Soit la relation de la table 1 dans laquelle ont indique pour chaque employé sont numéro, le numéro de son département, l'année du d'entrée dans le département, le nom du département en question et le numéro du responsable.

1. Écrire la liste de toutes les DF que vous pouvez trouver. Proposer une méthode pour ne pas en oublier.
2. Indiquer à votre avis quelles sont les DFs précédentes qui n'ont pas d'interprétation naturelle vis-à-vis des données.
3. Écrire la liste de toutes les DI « internes » que vous pouvez trouver. Commenter.

Corrections

Solution de l'exercice 1

1. Avec $\sigma_1(1) = 1, \sigma_1(2) = 2$ et $\sigma_2(1) = 2, \sigma_2(2) = 1$
$$\dfrac{\dfrac{R[ABC] \subseteq S[EFG]}{R[AB] \subseteq S[EF]}\ \gamma_P\ (\sigma_1) \qquad \dfrac{S[FE] \subseteq T[LK]}{S[EF] \subseteq T[LK]}\ \gamma_P\ (\sigma_2)}{R[AB] \subseteq T[LK]}$$
2. Dans le cas général de la règle de permutation et projection, pour chaque sous-ensemble de $\{1..n\}$ de cardinalité k on peut choisir n'importe quelle permutation possible, autrement dit c'est le nombre d'arrangement de k parmi n, soit $A_n^k = n!/(n-k)!$. On peut choisir toutes les valeurs de $0 \leq k \leq n$, on a donc ainsi $\sum_{i=0}^{n} A_n^i$ possibilités. Dans le cas de $R[ABC] \subseteq S[EFG]$, en énumérant la partir gauche cela fait $\{\emptyset\} \cup \{A, B, C\} \cup \{AB, AC, BA, BC, CA, CB\} \cup \{ABC, ACB, BAC, BCA, CAB, CBA\}$ soit 16 possibilités. Voir `http://oeis.org/A000522`.
3. Supposons $\{r, s\}$ une base de données telle que $\{r, s\} \models \{R[XY] \subseteq S[TU], S : T \to U\}$. Il faut montrer que $r \models X \to Y$ également. Pour cela, supposons $t_0, t_1 \in r$ avec $t_0[X] = t_1[X] = \overline{x}$, $t_0[Y] = \overline{y_0}$ et $t_1[Y] = \overline{y_1}$. Comme $\{r, s\} \models R[XY] \subseteq S[TU]$, ils existent $s_0, s_1 \in s$ avec $s_0[T] = s_1[T] = \overline{x}$, $s_0[U] = \overline{y_0}$ et $s_1[U] = \overline{y_1}$. Or, comme $s \models S : T \to U$ et $s_0[T] = s_1[T]$ on a $t_0[Y] = s_0[U] = \overline{y_0} = \overline{y_1} = s_1[U] = t_1[Y]$ et donc $r \models X \to Y$.
4. Il faut montrer que $\forall t_0 \in r. \exists t_1 \in s. t_0[XYZ] = t_1[TUV]$. Supposons donc un tuple $t \in r$. D'une part, comme $\{r, s\} \models R[XY] \subseteq S[TU]$, on un certain tuple $t_1 \in s$ avec $t[XY] = t_1[TU]$, et d'autre part on a un certain tuple $t_2 \in s$ avec $t[XZ] = t_2[TV]$. On illustre $\pi_{TUV}(s)$ ainsi :

	T	U	V
t_1	$\overline{x}$	$\overline{y}$	$\overline{v}$
t_2	$\overline{x}$	$\overline{u}$	$\overline{z}$

Or comme $t_1[T] = t_2[T] = t[X]$ et $s \models T \to U$, on a $t_1[U] = t_2[U] = t[U]$ et ainsi $t_2[TUV] = t[XYZ]$.

Solution de l'exercice 2

1. Utiliser une approche par niveaux croissants sur les parties gauches. Pour trouver quels sont les attributs qui dépendent de l'ensemble d'attributs X, on va regarder toutes les couples (s, t) de tuples tels que $s[X] = t[X]$ et on élimine tous les attributs qui n'ont pas de valeurs identiques pour au moins un couple. Par exemple, pour $X = A$, on a $t_1[A] = t_2[A]$ mais tous les autres attributs sont différents, on a donc seulement $A \to A$.
On commence par les attributs seuls, ensuite, on va regarder toutes les combinaisons possibles (sans ordre entre les attributs). Au niveau 3, on ne considère que les parties gauches dont aucun sous-ensemble n'est une clé, le seul à considérer est donc BDE. Il n'y a pas à aller plus loin car en ajoutant soit A soit C à BDE on obtient une clé.

partie gauche	fermeture
A	A
B	BDE
C	CE
D	DBE
E	E
AB	ABCDE
AC	ABCDE
AD	ABCDE
AE	ABCDE
BC	ABCDE
BD	BDE
BE	BDE
CD	ABCDE
CE	CE
DE	BDE
BDE	BDE

2. On a $B \to DE$ et $D \to BE$, autrement dit on peut identifier un département par son nom ou son numéro et connaître l'un ou l'autre permet de déterminer qui est le reponsable du département. Que AB et AD soient clés est donc logique. En revanche, les clés faisant intervenir C ou E en partie gauche sont douteuses. Typiquement $C \to E$ n'est pas sensée, c'est juste un effet de bord dû au fait qu'informatique et géophysique aient le même directeur.
3. $R[A] \subseteq R[A], R[B] \subseteq R[A], R[C] \subseteq R[C], R[D] \subseteq R[D], R[E] \subseteq R[E]$ pour les triviales, et $R[B] \subseteq R[A]$. Clairement $R[B] \subseteq R[A]$ est un effet de bord, mais on s'attendrait par contre à avoir $R[E] \subseteq R[A]$!