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

Licence informatique – Automne 2014–2015

Les questions marquées du symbole (†) sont à préparer pour la séance

Exercice 1 : 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 d'entrée de l'employé 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.

	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 et logique propositionnelle ([1])

Dans cet exercice on va s'intéresser au lien entre les dépendances fonctionnelles $X \rightarrow Y$ et les formules de la logique propositionnelle classique de la forme $X \Rightarrow Y$ pour montrer que la correspondance entre les notations n'est pas simplement superficielle.

Soit Σ un ensemble de DFs sur un schéma $\mathbf{R}$. A chaque attribut $A \in \mathbf{R}$ on associe une *variable propositionnelle* $\underline{A} \in \mathcal{P}$. On étend cette transformation aux DFs en faisant correspondre à une DF $A_1 \dots A_p \rightarrow B_1 \dots B_q \in \Sigma$ la formule de logique propositionnelle $\underline{A}_1 \wedge \dots \wedge \underline{A}_p \Rightarrow \underline{B}_1 \wedge \dots \wedge \underline{B}_q$ sur l'ensemble de propositions $\mathcal{P}$. Enfin, on note $\underline{\Sigma}$ l'ensemble de formules logiques correspondant aux DFs de Σ en étendant la transformation à un ensmeble de dépendances. Soit $\alpha = A_1 \dots A_p \rightarrow B_1 \dots B_q$ une DF. On va montrer que les trois propositions suivantes sont équivalentes :

$$\Sigma \models \alpha \tag{1}$$
$$\Sigma \models_2 \alpha \tag{2}$$
$$\underline{\Sigma} \models_P \underline{\alpha} \tag{3}$$

Dans ces formules, $\Sigma \models \alpha$ désigne l'implication logique entre un ensemble de DFs et une DF, $\Sigma \models_2 \alpha$ désigne l'implication logique de DFs limitée *aux relations ne contenant que deux tuples*, autrement dit $\Sigma \models_2 \alpha \equiv \forall r.(|r| = 2 \wedge r \models \Sigma) \Rightarrow r \models \alpha$. Enfin, $\underline{\Sigma} \models_P \underline{\alpha}$ désigne l'implication logique classique, c'est-à-dire, que pour toute valuation des variables propositionnelles $\nu : \mathcal{P} \rightarrow \{0, 1\}$ telle que $\nu \models \underline{\Sigma}$ on a aussi $\nu \models \underline{\alpha}$. On donne le lemme suivant :

Lemme 1. *Soient $\nu : \mathcal{P} \rightarrow \{0, 1\}$ une valuation des variables propositionnelles, $r = \{t_1, t_2\}$ la relation à deux tuples tels que $t_1[A] = 1$ et $t_2[A] = \nu(\underline{A})$ pour tout $A \in \mathbf{R}$ et $\alpha = A_1 \dots A_p \rightarrow B_1 \dots B_q$ une dépendance arbitraire. Alors on a $\nu \models_P \underline{\alpha}$ si et seulement si $r \models \alpha$.*

1. Montrer que la proposition (1) est équivalente à la proposition (2). Pour le sens (2) $\Rightarrow$ (1) on procédera par l'absurde en montrant que supposer $\Sigma \models_2 \alpha$ et $\Sigma \not\models \alpha$ conduit à une contradiction.
2. Montrer le lemme 1 en construisant une relation à deux tuples « à la Armstrong » à partir d'une interprétation ν donnée et vice-versa.
3. Montrer que la proposition (3) est équivalente à la proposition (2) en utilisant le lemme 1. Procéder par l'absurde pour les deux implications.
4. Conclure.

Corrections

Solution de l'exercice 1

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.

Solution de l'exercice 2

1. Le sens (1) $\Rightarrow$ (2) est clair car si $\Sigma \models \alpha$ pour toute relation r, alors c'est aussi vrai pour le cas limité à $|r| = 2$. Pour la réciproque (2) $\Rightarrow$ (1), supposons le contraire, c'est-à-dire $\Sigma \models_2 \alpha$ et $\Sigma \not\models \alpha$, pour aboutir à une contradiction. Si $\Sigma \not\models \alpha$, alors il existe une relation r (de taille arbitraire) telle que $r \models \Sigma$ et $r \not\models \alpha$. Cela signifie qu'ils existent (au moins) deux tuples $t_1, t_2 \in r$ tels que $t_1[A_1 \dots A_p] = t_2[A_1 \dots A_p]$ et $t_1[B_1 \dots B_q] \neq t_2[B_1 \dots B_q]$. Il suffit donc de considérer la relation $r' = \{t_1, t_2\}$ qui a exactement deux tuples : elle vérifie Σ par hypothèse mais on vient de montrer qu'elle viole α, contradiction avec $\Sigma \models_2 \alpha$.
2. Pour la direction *si*. Supposons $r \models \alpha$, deux cas sont possibles :
   1. α est trivialement satisfaite car $t_1[A_i] \neq t_2[A_i]$ pour un certain i. Dès lors $\nu \not\models \underline{A}_1 \wedge \dots \wedge \underline{A}_p$ car $\nu(\underline{A}_i) = 0$ et ainsi $\nu \models \underline{A}_1 \wedge \dots \wedge \underline{A}_p \Rightarrow \underline{B}_1 \wedge \dots \wedge \underline{B}_q$ par inspection de la table de vérité de $\Rightarrow$.
   2. α est satisfaite car $t_1[A_i] = t_2[A_i]$ pour tout i et de même $t_1[B_j] = t_2[B_j]$. On a donc $\nu \models \underline{A}_1 \wedge \dots \wedge \underline{A}_p$ et également $\nu \models \underline{B}_1 \wedge \dots \wedge \underline{B}_q$. On conclut encore par inspection de la table de vérité de $\Rightarrow$.

   Pour la direction *seulement si*, on suppose que $\nu \models_P \underline{\alpha}$ et on considère encore les deux cas :
   1. soit $\nu \models_P \underline{\alpha}$ par vacuité car $\nu(\underline{A}_i) = 0$ pour un certain i et donc dans ce cas α est trivialement satisfaite car $t_1[A_i] \neq t_2[A_i]$,

2. soit on a $\nu(\underline{A_i}) = 1$ pour tout i et aussi $\nu(\underline{B_j}) = 1$ pour chaque j car $\nu \models_P \underline{\alpha}$. Dans ce cas aussi α est satisfaite car $\{t_1, t_2\}$ est le seul couple de tuple à considérer pour vérifier $r \models \alpha$.

3. Pour le sens (3) $\Rightarrow$ (2) on procède par l'absurde. On suppose donc $\Sigma \models_2 \alpha$ et $\underline{\Sigma} \not\models_P \underline{\alpha}$. Il existe donc une valuation ν telle que $\nu \models_P \underline{\Sigma}$ et $\nu \not\models_P \underline{\alpha}$. Par inspection de la table de vérité de $\Rightarrow$, cela signifie que $\nu(\underline{A_i}) = 1$ pour tout i mais qu'il existe (au moins) un index j tel que $\nu(\underline{B_j}) = 0$. A partir de cette valuation, construisons alors une relation $r = \{t_1, t_2\}$ telle que décrit dans le lemme 1. D'après le lemme on a $r \not\models \alpha$ et $r \models \Sigma$, contradiction.
   Pour le sens (2) $\Rightarrow$ (3), on a procéder de façon similaire en supposant $\underline{\Sigma} \models_P \underline{\alpha}$ et $\Sigma \not\models_2 \alpha$. On a donc au moins une relation à deux tuples $r = \{t_1, t_2\}$ qui vérifie Σ mais pas α. On construit alors la valuation ν telle que $\nu(\underline{A}) = 1$ si et seulement si $t_1[A] = t_2[A]$. On est dans le cas d'application du lemme 1 et d'après celui-ci on déduit que $\nu \not\models \underline{\alpha}$ et $\nu \models \underline{\Sigma}$, une contradiction.
4. D'après la question 1 on a (1) $\equiv$ (3), d'après la question 3 on a aussi (2) $\equiv$ (3), les trois propositions sont donc équivalentes.

Références

[1] R. Fagin. Functional dependencies in a relational database and propositional logic. *IBM J. Res. Dev.*, 21(6) :534–544, Nov. 1977.