Lif10 – Fondements des bases de données
TD6 – Formes normales, minimisation, décomposition

Licence informatique – Automne 2014–2015

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

Exercice 1 : normalisation (†)

Donner la forme normale la plus avancée des schémas de relation suivants munis de l'ensemble Σ de contraintes. Si ce n'est pas fait, normaliser la relation en la décomposant.

1. (†) $Employe(NomEmp, NomDept, NomChef)$
   avec $\Sigma = \{NomEmp \to NomDept; NomDept \to NomChef\}$;
2. (†) $Employe(NomEmp, NomEnfant, Salaire)$
   avec $\Sigma = \{NomEmp \to Salaire; NomEmp \twoheadrightarrow NomEnfant\}$;
3. (†) $Adresse(Rue, Ville, CodePostal)$
   avec $\Sigma = \{Rue, Ville \to CodePostal; CodePostal \to Ville\}$;

Exercice 2 : Couverture minimale (†)

Soit Σ l'ensemble de DFs $\Sigma = \{A \to B; A \to C; D \to E; C \to D; B \to C; BC \to A\}$

1. (†) Donner une couverture minimum de Σ en utilisant l'algorithme vu précédemment dont on rappelle le principe :
   1. calculer $\Sigma' = \{X \to X^+ \mid X \to Y \in \Sigma\}$
   2. supprimer les DFs $f \in \Sigma'$ redondantes, c'est-à-dire telles que $\Sigma' \setminus \{f\} \models f$
2. (†) Réduire la couverture minimum en appliquant l'algorithme de réduction des parties droites et gauches.

Exercice 3 : modélisation

On souhaite créer une base de données de recettes de cuisine, décrites comme suit :

> *Une recette est identifiée par son numéro. Elle a un nom et un type particulier (soupe, entrée, dessert, ...). Elle utilise un ou plusieurs ingrédients (carottes, viande de boeuf, poivre, ...). Un ingrédient est identifié par son numéro et a un nom. Pour chaque ingrédient dans une recette, on précisera sa quantité.*

1. Dresser l'inventaire des DF valides sur la *relation universelle* $R = \{NumR, NomR, TypeR, NumI, NomI, Qte\}$
2. Exhiber les problèmes de redondance à partir de cet exemple. La DF $NumR, NumI \to Qte$ engendre-t-elle un problème de redondance ?
3. Proposer de façon intuitive une décomposition sans pertes et sans redondance de R.
4. On veut maintenant ajouter à cette modélisation qu'à chaque recette correspond un ensemble d'ustensiles (attribut *Ustensile*). Cette information se traduit-elle par une nouvelle DF ?
5. Que penser de la décomposition $R_1 = (\underline{NumR}, NomR, TypeR)$, $R_2 = (\underline{NumR, NumI}, Qte)$, $R_3 = (\underline{NumI}, NomI)$, $R_4 = (\underline{NumR, Ustensile})$.

Exercice 4 : définitions de la troisième forme normale

L'objectif de cette exercice est de montrer que les deux définitions **3FN-1** et **3FN-2** ci-après de la troisième forme normale sont équivalentes. On considère les paires $\langle R, F \rangle$ formées d'un schéma de relation R et d'un ensemble de F de dépendances fonctionnelles sur R. On donne les définitions suivantes :

Directe Une DF $X \to Y$ est *directe ssi* $\nexists Z.X \to Z \land Z \not\to X \land Z \to Y$.

Premier Un attribut $A \in R$ est *premier* s'il appartient à *au moins* une clé minimale de R.

2FN $\langle R, F \rangle$ est en **2FN** *ssi* il n'existe pas de DF non-triviale $X \to A \in F^+$ avec A non-premier et X sous-ensemble *propre*[1] d'une clé minimale.

3FN-1 $\langle R, F \rangle$ est en **3FN** *ssi* $\langle R, F \rangle$ est en **2FN** et toutes les DFs non-triviales $X \to A \in F^+$ avec A non-premier sont *directes*.

3FN-2 $\langle R, F \rangle$ est en **3FN** *ssi* pour toute DF non-triviale $X \to A \in F^+$, soit X est (super)clé, soit A est premier.

1. Montrer que si la condition **2FN** n'est pas respectée, alors la **3FN-2** ne l'est pas non plus.
2. Montrer que s'il existe une DF non-triviale non-directe $X \to A \in F^+$ avec A non-premier, alors la condition **3FN-2** n'est pas respectée.
3. Conclure des deux questions précédentes que la définition **3FN-2** implique la définition **3FN-1**.
4. Montrer que s'il existe une DF $X \to A$ avec A non-premier et X qui n'est pas (super)clé, alors la condition **3FN-1** n'est pas satisfaite.
5. Conclure sur l'équivalence des définitions **3FN-1** et **3FN-2**.

1. X est un sous-ensemble *propre* de Y noté $X \subsetneq Y$ *ssi* $X \subseteq Y \land X \neq Y$

Corrections

Solution de l'exercice 1

1. en 2FN : on peut normaliser en $R_1 = (NomEmp, NomDept)$ et $R_2 = (NomDept, NomChef)$ qui est en FNBC
2. en 1FN : on peut normaliser en $R_1 = (NomEmp, salaire)$ et $R_2 = (NomEmp, NomEnfant)$ qui est en 4FN et donc en FNBC.
3. en 3FN : on ne peut pas faire mieux sans perte de dépendances.

Solution de l'exercice 2

1. On obtient la couverture minimum $\{B \to ABCDE; D \to DE; C \to CDE; A \to ABCDE\}$
2. On obtient $\{B \to A; D \to E; C \to D; A \to BC\}$

Solution de l'exercice 3

1. On a les DFs suivantes $F = \{NumR \to NomR, TypeR; NumI \to NomI; NumR, NumI \to Qte\}$:
   - *Une recette est identifiée par son numéro. Elle a un nom et un type particulier* donne $NumR \to NomR, TypeR$,
   - *Un ingrédient est identifié par son numéro et a un nom* donne $NumI \to NomI$,
   - *Pour chaque ingrédient dans une recette, on précisera sa quantité* donne $NumI \to Qte$.
2. Un problème de redondance a lieu lorsqu'une DF prends deux fois la même valeur à gauche et à droite, autrement dit, si on a une DF $X \to Y$ et deux tuples t_1 et t_2 *différents* tels que $t_1[XY] = t_2[XY]$. La dernière DF ne pose pas de problème, puisque la partie gauche est une clé.
3. On va décomposer selon les parties gauches des DFs obtenues $R_1 = (\underline{NumR}, NomR, TypeR)$, $R_2 = (\underline{NumR, NumI}, Qte)$, $R_3 = (\underline{NumI}, NomI)$.
4. Non, il n'y a pas de nouvelle DF non-triviale.
5. Cette décomposition engendre une perte de jointure... Il suffit pour cela d'exhiber la relation suivante sur R :

No. Tuple	NumR	NomR	TypeR	NumI	NomI	Qte	Ustensile
1	1	Couscous	Plat principal	1	Pois Chiches	0,5kg	Couscoussière
2	1	Couscous	Plat principal	2	Agneau	1kg	Plat

On peut vérifier quette relation ne viole aucune DF, elle est donc valide. On construit alors les relations r_1, r_4 en prenant les projections correspondantes. Ensuite, on fait la jointure naturelle entre ces relations. On obtient :

No. Tuple	NumR	NomR	TypeR	NumI	NomI	Qte	Ustensile
1	1	Couscous	Plat principal	1	Pois Chiches	0,5kg	Couscoussière
3	1	Couscous	Plat principal	2	Agneau	1kg	Couscoussière
4	1	Couscous	Plat principal	1	Pois Chiches	0,5kg	Plat
2	1	Couscous	Plat principal	2	Agneau	1kg	Plat

Cette relation est différente de la première, il y a donc perte de jointure et la décomposition n'est pas correcte. On devrait préciser que soit l'ustensile dépend de l'ingrédient *et* de la recette ou alors introduire une MVD pour préciser l'indépendance entre ustensiles et ingrédients dans une recette.

Solution de l'exercice 4

1. $\langle ABC, \{AB \to C, B \to C\}\rangle$ n'est pas en 2FN. $\langle ABC, \{A \to B, B \to C\}\rangle$ est en 2FN mais pas en 3FN.
2. Supposons $\langle R, F\rangle$ viole la 2FN, on a donc une dépendance $X \to A$ avec $X \subsetneq K$ où K est une clé minimale et l'attribut A non-premier. Comme K est minimale, X ne peut pas être clé et donc super clé et la condition 3FN-2 n'est pas respectée.
3. Supposons que $\langle R, F\rangle$ satisfait une dépendance non-directe $X \to Z \to A$ avec $Z \not\to X$ avec A non-premier. Comme $Z \not\to X$, Z n'est pas clé et viole la condition 3FN-2.

4. On a montré par la contraposée que la 3FN-2 implique la 2FN et la 3FN-2 implique l'absence de dépendances non directe, c'est-à-dire la définition de la 3FN-1.
5. Supposons qu'il existe une dépendance $X \to A$ avec A non-premier et X qui n'est pas (super)clé. Soit K une clé minimale :
   – $X \subseteq K$ et donc $X \subsetneq K$ car X n'est pas super clé. On a une violation de la 2FN sur la dépendance $X \to A$
   – $X \not\subseteq K$ et $X \not\to K$ car sinon X serait clé. On a donc une dépendance transitive $K \to X \to A$ qui viole la condition 3FN-1.
6. La dernière question a montré par la contraposée que la 3FN-1 implique la 3FN-2 et d'après la question 4 on a l'équivalence recherchée.