

LIF10 – FONDEMENTS DES BASES DE DONNÉES

TD6 – minimisation, normalisation

Licence informatique – Automne 2013–2014

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

Exercice 2 : couverture minimale

On rappelle qu'une couverture minimum des d'un ensemble de DFs possède un nombre minimal de DF. L'algorithme 1 prend en entrée une couverture minimum, et donne en sortie une couverture canonique avec un nombre réduit d'attributs en parties gauche et droite. Soit Σ l'ensemble de DFs $\Sigma = \{A \rightarrow B; A \rightarrow C; D \rightarrow E; C \rightarrow D; B \rightarrow C; BC \rightarrow A\}$

1. Donner une couverture minimum de Σ en calculant $\Sigma' = \{X \rightarrow X^+ \mid X \rightarrow Y \in \Sigma\}$ puis en supprimant 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 1.

Exercice 3 : modélisation

On souhaite modéliser des données sur des films. On analyse les données de la façon suivante.

Chaque film est identifié par un unique numéro d'exploitation. Un film est également déterminé par son titre et sa date de sortie. Un film peut être une reprise d'un ou plusieurs autres films avec lesquels il possède un taux de ressemblance. Chaque film a un réalisateur et un producteur, et implique un ou plusieurs studios et un ensemble d'acteurs. Pour chaque acteur, on souhaite stocker le nombre de minutes qu'il joue dans un film ainsi que son cachet pour le film.

Notons qu'un réalisateur peut-être producteur ou acteur dans un autre film, et inversement. Un réalisateur/producteur/acteur possède un identifiant unique, un nom, un prénom et une adresse. La connaissance du nom, du prénom et de l'adresse permet d'identifier uniquement une personne. Un studio est identifié par un numéro et possède un nom unique et une adresse.

1. Faire l'inventaire des attributs (relation universelle) et des contraintes (DF, DI, DMV) du problème.
2. Proposer une décomposition sans pertes de la relation universelle en précisant les clés primaires et étrangères.

Algorithme 1: réduction des attributs d'un ensemble minimum de DFs

Data : Un ensemble *minimum* de DF F sur R .

Result : F avec un nombre minimal d'attributs

```
1  $Min := F$ 
  /* Réduction des parties gauches */
2 for  $X \rightarrow Y \in Min$  do
3    $W := X$ 
4   for  $A \in X$  do
5     if  $Min \models (W - A) \rightarrow X$  then  $W := W - \{A\}$ 
6    $Min := (Min - \{X \rightarrow Y\}) \cup \{W \rightarrow Y\}$ 
  /* Réduction des parties droites */
7 for  $X \rightarrow Y \in Min$  do
8    $W := Y$ 
9   for  $A \in Y$  do
10     $G := (Min - \{X \rightarrow Y\}) \cup \{X \rightarrow (W - A)\}$ 
11    if  $G \models X \rightarrow Y$  then  $W := W - \{A\}$ 
12   $Min := (Min - \{X \rightarrow Y\}) \cup \{X \rightarrow W\}$ 
13 return  $Min$ 
```

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 $\{B \rightarrow ABCDE; D \rightarrow DE; C \rightarrow CDE; A \rightarrow ABCDE\}$
2. On obtient $\{B \rightarrow A; D \rightarrow E; C \rightarrow D; A \rightarrow BC\}$

Solution de l'exercice 3

1. On obtient l'ensemble de DFs suivant :
 - $IDFilm \rightarrow Titre, Date$
 - $Titre, Date \rightarrow IDFilm$
 - $IDFilm, IDReprise \rightarrow Ressemblance$
 - $IDFilm \rightarrow IDProd, IDReal$
 - $IDFilm \twoheadrightarrow IDStudio$ ou $IDFilm \twoheadrightarrow IDActeur$
 - $IDFilm, IDActeur \rightarrow Duree, Cachet$
 - $IDPersonne \rightarrow NomPersonne, Prenom, AdressePersonne$
 - $IDStudio \rightarrow NomStudio, AdresseStudio$
 - $IDProd \subseteq IDPersonne$
 - $IDReal \subseteq IDPersonne$
 - $IDActeur \subseteq IDPersonne$
 - $IDReprise \subseteq IDFilm$
2. On va décomposer la relation universelle ainsi :
 - $R_1 = \{IDFilm, Titre, Date, \#IDProd, \#IDReal\}$, on choisit par exemple $IDFilm$ en clé primaire et on ajoute une contrainte d'unicité sur $Titre, Date$ comme alternative.
 - $R_2 = \{IDFilm, \#IDActeur, Duree, Cachet\}$
 - $R_3 = \{IDFilm, \#IDStudio\}$
 - $R_4 = \{IDFilm, \#IDReprise, Ressemblance\}$
 - $R_5 = \{IDStudio, NomStudio, AdresseStudio\}$
 - $R_6 = \{IDPersonne, NomPersonne, Prenom, AdressePersonne\}$