Lif10 – Fondements des bases de données

TD5 – relations d'Armstrong, couverture minimale, introduction à la normalisation

Licence informatique – Automne 2013–2014

Algorithme 1: calcul d'une relation d'Armstrong

Data : R un schéma de relation, Fun ensemble de DF sur R.
Result : Une relation d'Armstrong r pour F.

for $A \in R$ **do**
  $t[A] := 0$
$r := \{t\}$
$i := 1$
for $X \in Cl(F) \setminus R$ **do**
  for $A \in R$ **do**
    if $A \in X$ **then**
      $t[A] := 0$
    else
      $t[A] := i$
  $r := r \cup \{t\}$
  $i := i + 1$
return r

Exercice 1 : relation d'Armstrong

Soit un schéma de relation R et un ensemble F de DF sur ce schéma. Une *relation d'Armstrong* pour F est une relation r sur R vérifiant *exactement* F^+. Soit le schéma $R = ABCDE$ et l'ensemble de DF $F = \{A \to BC, D \to E, C \to D\}$ pour la suite.

1. Calculer l'ensemble des fermés de F, défini par $Cl(F) = \{X^+ \mid X \subseteq R\}$. Pour cela, utilisez une exploration systématique des sous-ensembles de R en commençant par ceux de taille 1, puis 2 . . . Penser à ne pas visiter les sur-ensembles des clés pour limiter le nombre de calculs.
2. Utiliser l'algorithme 1 pour calculer une relation d'Armstrong r.
3. Supposons que l'on supprime le tuple correspondant au fermé BE de la relation r obtenue à la question précédente. Il y a-t-il des DFs nouvellement vérifiées qui ne l'étaient pas ?
4. Même question que précédemment avec le fermé DE.
5. Quand $F = \emptyset$, combien il y a-t-il de tuples dans la relation d'Armstrong ? Même question quand F est l'ensemble de toutes les DFs possibles sur R ?

Algorithme 2: calcul d'une couverture minimum

Data : F un ensemble de DF
Result : G une couverture minimum de F

$G := \emptyset$
for $X \rightarrow Y \in F$ **do**
    $G := G \cup \{X \rightarrow X^+\}$;
for $X \rightarrow X^+ \in G$ **do**
    if $G - \{X \rightarrow X^+\} \vdash X^+$ **then**
        $G := G - \{X \rightarrow X^+\}$;
return G;

Exercice 2 : couveture minimale

1. Soient $F_1 = \{A \rightarrow C, AC \rightarrow D, E \rightarrow AD, E \rightarrow F\}$ et $F_2 = \{A \rightarrow CD, E \rightarrow AF\}$. Déterminer si chaque DF de F_2 peut être inférée par F_1 et réciproquement.
2. En utilisant l'algorithme 2 déterminer une couverture minimale de F_1. Comment comparer la couverture ainsi obtenue à F_2 ?
3. En utilisant l'algorithme 2 déterminer si $F = \{A \rightarrow BC, D \rightarrow E, C \rightarrow D\}$ est minimal.

Exercice 3 : modélisation

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

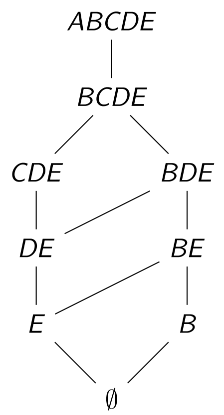
> *Une recette est identifiée par son numéro. Elle a un nom et un type particulier (soupe, entrée, dessert, ...). Elle utilise un ou plusieurs ingrédients (carottes, viande de boeuf, poivre, ...). Un ingrédient est identifié par son numéro et a un nom. Pour chaque ingrédient dans une recette, on précisera 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 $NumR, NumI \rightarrow Qte$ 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 $R_1 = (\underline{NumR}, NomR, TypeR)$, $R_2 = (\underline{NumR, NumI}, Qte)$, $R_3 = (\underline{NumI}, NomI)$, $R_4 = (\underline{NumR, Ustensile})$.

Corrections

Solution de l'exercice 1

1. On trouve les fermés suivants : $ABCDE, BCDE, B, CDE, DE, E, BDE, BE$. Muni de l'inclusion, on peut représenter graphiquement cet ensemble ordonné par son diagramme de Hasse :



2. On trouvera, en les considérant dans l'ordre où les fermés sont donnés :

fermé	A	B	C	D	E
$ABCDE$	0	0	0	0	0
$BCDE$	1	0	0	0	0
B	2	0	2	2	2
CDE	3	3	0	0	0
DE	4	4	4	0	0
E	5	5	5	5	0
BDE	6	0	6	0	0
BE	7	0	7	7	0

3. La relation à deux tuples composée de $\langle 7, 0, 7, 7, 0\rangle$ associé à BE et de $\langle 0, 0, 0, 0, 0\rangle$ associé à $ABCDE$ falsifie les DFs de la forme $BE \to X$ avec $X \in ACD$. On peut voir que parmi celles-ci c'est la DF $BE \to D$ qui n'a plus de contre exemple si on supprime $\langle 7, 0, 7, 7, 0\rangle$. Les DFs $BE \to A$, $BE \to C$ restent quant à elles encore violées grâce à $\langle 6, 0, 6, 0, 0\rangle$ associé à BDE.
4. Dans ce cas là, il n'y a aucun contre-exemple qui est perdu. En effet le tuple de DE permet d'invalider $DE \to A$, $DE \to B$ et $DE \to C$ mais chacune de ces DFs trouve encore un contre-exemple soit via le tuple associé à CDE soit via celui de BDE. En fait, la méthode utilisée pour générer une relation d'Armstrong produit des relations avec un nombre maximal de tuples. Pour avoir un nombre minimal il ne faut pas considérer la famille de Moore $\{ABCDE, BCDE, B, CDE, DE, E, BDE, BE\}$ mais seulement son sous-ensemble $\{ABCDE, BCDE, B, CDE, BDE, BE\}$ qui l'engendre par intersection (théorème 4.1 de *On the Structure of Armstrong Relations for Functional Dependencies*).
5. Si $F = \emptyset$ alors chaque $X \subseteq R$ est fermé et il y a ainsi $2^{|R|}$ tuples dans la relation d'Armstrong. A l'inverse, si F contient toutes les DFs possibles, alors il n'y a qu'un seul fermé $ABCDE$ et la relation obtenue ne contient qu'un seul tuple.

Solution de l'exercice 2

1. Pour la première question, on calcule $A^+ = ACD$ via F_1 et donc $F_1 \models A \to CD$. De même $E^+ = ACDEF$ et $F_1 \models E \to AF$. Dans l'autre sens c'est également le cas, et ainsi on a bien $F_1^+ = F_2^+$ et les deux ensembles sont équivalents.
2. Après la première boucle, on obtient la couverture $\{A \to ACD, AC \to ACD, E \to ACDEF\}$. Après la deuxième boucle, on supprime $AC \to ACD$ et on obtient $\{A \to ACD, E \to ACDEF\}$. On remarque que ces DFs sont minimales à gauche car composée d'un singleton mais qu'on peut les simplifier à droite pour obtenir $A \to CD$ et $E \to AF$, soit exactement F_2.
3. On utilise l'algorithme on obtient $F' = \{A \to ABCDE; D \to DE, C \to CDE\}$ qui a le même nombre de DFs et l'ensemble est bien minimal.

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.