Lif10 – Fondements des bases de données
TD5 – Relation d'Armstrong, couvertures et dépendances d'inclusion

Licence informatique – Automne 2014–2015

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

Algorithme 1: calcul d'une relation d'Armstrong

```
Data : R un schéma de relation, F un ensemble de DF sur R.
Result : Une relation d'Armstrong r pour F.
for A ∈ R do
    t[A] := 0
r := {t}
i := 1
for X ∈ Cl(F) \ R do
    for A ∈ R do
        if A ∈ X then
            t[A] := 0
        else
            t[A] := i
    r := r ∪ {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 3 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 $X \to Y \in F$ **do**
  $G := G \cup \{X \to X^+\}$;
for $X \to X^+ \in G$ **do**
  if $G - \{X \to X^+\} \vdash X^+$ **then**
    $G := G - \{X \to X^+\}$;
return G;

Algorithme 3: 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

$Min := F$
`// Réduction des parties gauches`
for $X \to Y \in Min$ **do**
  $W := X$
  for $A \in X$ **do**
    if $Min \models (W \setminus \{A\}) \to X$ **then** $W := W \setminus \{A\}$
  $Min := (Min \setminus \{X \to Y\}) \cup \{W \to Y\}$
`// Réduction des parties droites`
for $X \to Y \in Min$ **do**
  $W := Y$
  for $A \in Y$ **do**
    $G := (Min \setminus \{X \to Y\}) \cup \{X \to (W \setminus \{A\})\}$
    if $G \models X \to Y$ **then** $W := W \setminus \{A\}$
  $Min := (Min - \{X \to Y\}) \cup \{X \to W\}$
return Min

Exercice 2 : couverture minimale (†)

1. (†) Soient $F_1 = \{A \to C, AC \to D, E \to AD, E \to F\}$ et $F_2 = \{A \to CD, E \to AF\}$. 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 $F = \{A \to BC, D \to E, C \to D\}$ est minimal.

Exercice 3 : Couvertures non redondantes

Soit l'ensemble F de DF suivant le schéma $R = ABCDEFG$:

$D \to A$	$D \to C$	$D \to E$
$D \to F$	$CE \to G$	$AG \to F$
$ADG \to B$	$BG \to ADE$	$BF \to DG$

1. Démontrer que F n'est ni optimum, ni minimum, ni non-redondante.
2. Calculer une couverture canonique de F (couv. minimum + réduction, algorithme 3).

Exercice 4 : 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

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

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

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

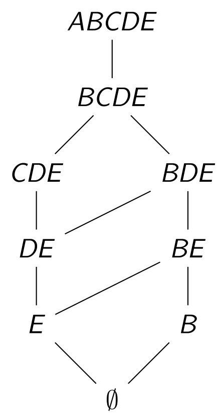
avec σ une permutation d'un sous-ensemble de $\{1 \ldots n\}$

FIGURE 1 – Axiomatisation de Casanova pour les DIs

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. Nous savons que Couv. Optimun $\Rightarrow$ Couv. minimum et couv. minimum $\Rightarrow$ couv. non redondante. Il suffit donc d'utiliser la contraposée et de montrer que F est redondante. Pour cela, nous pouvons exhiber la DF $D \to A$ qui peut être retrouvée à l'aide des autres DFs de F.
2. Calculons d'abord la couverture minimum. Pour cela, il faut reécrire les règles sous la forme $X \to X^+$, soit l'ensemble Σ suivant :
   $D \to R$
   $CE \to CEG$
   $AG \to AGF$
   $ADG \to R$
   $BG \to R$
   $BF \to R$
   Notons que la DF $D \to R$ n'apparait qu'une seule fois et pas quatre car nous adoptons une vision ensembliste (les 3 autres occurences supprimées).
   Il faut maintenant supprimer les DFs de Σ inutiles. Ici, il n'y en a qu'une seule : $ADF \to R$ ($D \to R$)
   Il faut donc réduire l'ensemble suivant :
   $D \to ABCDEFG$
   $CE \to CEG$
   $AG \to AGF$
   $BG \to ABCDEFG$
   $BF \to ABCDEFG$

   Attention la réduction va dépendre de l'ordre dans lequel on traite les DFs. Ici, que des parties droites à réduire :
   – $BF \to ABCDEFG$ simplifié en $BF \to D$ puisque $D \to R$.
   – Idem pour $BG \to ABCDEFG$: $BG \to D$.
   – $AG \to AGF$, il faut supprimer A et G qui apparaissent déjà en partie gauches.
   – Idem pour $CE \to CEG$.
   – $D \to ABCDEFG$: on peut supprimer le D (apparait à gauche), le G (car $CE \to G$), le F ($AG \to F$) : $D \to ABCE$
   Ce qui donne le résultat final (un parmi d'autre) :
   $D \to ABCE$
   $CE \to G$
   $AG \to F$
   $BG \to ABCDEFG$
   $BF \to D$

Solution de l'exercice 4

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 $0 \le k \le n$, 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 $t_2[TUV] = t[XYZ]$.