Inférence, couvertures de dépendances

UE fondement de bases de données - TD4

1. Soit I l'ensemble de DI suivant sur le schéma de bases de données $\mathbf{D} = \{R, S, T\}$ avec $R = ABCD, S = EFGH$ et $T = IJKL$.
   – $R[ABC] \subseteq S[EFG]$
   – $R[F] \subseteq T[J]$
   – $S[FE] \subseteq T[LK]$
   En utilisant les règles de Casanova et al., calculez I^+.
   Tant pis si on ne les donne pas toutes... l'idée est qu'ils apréhendent la complexité de ce problème.
   Je vous rappelle les règles de casanova et al :
   Soit I un ensemble de DI sur un schéma de base de données $\mathbf{R}$. Les règles d'inférence de DI suivantes sont appelées *système d'inférence de Casanova et al.* :
   – Reflexivité : $I \vdash R[X] \subseteq R[X]$.
   – Projection et permutation :
     si $I \vdash R[A_1...A_n] \subseteq S[B_1...B_n]$ alors $I \vdash R[A_{\sigma 1}...A_{\sigma m}] \subseteq S[B_{\sigma 1}...B_{\sigma m}]$ pour chaque séquence $\sigma 1...\sigma m$ d'entier distincts dans $\{1...n\}$
   – transitivité : si $I \vdash R[X] \subseteq S[Y]$ et $I \vdash S[Y] \subseteq T[Z]$ alors $I \vdash R[X] \subseteq T[Z]$
   en plus de celles déjà écrites :
   toutes les projections :
   $R[A] \subseteq S[E]; R[B] \subseteq S[F]; R[C] \subseteq S[G]; S[F] \subseteq T[L]; S[E] \subseteq T[K]; R[AB] \subseteq S[EF]; R[AC] \subseteq S[EG]$; $R[BC] \subseteq S[FG]$
   + toutes les permutations de ces DI
   par exemple $R[ACB] \subseteq S[EGF]$ etc...
   + les transitives à partir de toutes celles trouvées
   la on doit trouver : $R[B] \subseteq T[J]; R[AB] \subseteq T[KL]$
   Il me semble qu'il y en a 34 soit 37 avec celles de départ, à partir seulement de 3 DI...
2. Démontrez que la règle d'inférence suivante pour les DF et les DI est juste.
   "(Collection) Si $F \cup I \vdash \{R[UV] \subseteq S[XY]; R[UW] \subseteq S[XZ]; X \longrightarrow Y\}$ alors $F \cup I \vdash R[UVW] \subseteq S[XYZ]$"
   Soit $d = \{r, s\}$ une base de données telle que $d \models F \cup I$ Soit un tuple $t \in r$. On doit montrer qu'il existe un tuple $t' \in s$ tel que $t'[XYZ] = t[UVW]$.
   Puisque $d \models R[UV] \subseteq S[XY]$ et $d \models R[UW] \subseteq S[XZ]$ alors $\exists t'_1, t'_2 \in s$ tels que $t'_1[XY] = t[UV]$ et $t'_2[XZ] = t[UW]$. Remarquer que l'on a alors $t'_1[X] = t'_2[X]$. Or, $d \models X \longrightarrow Y$, donc par définition $t'_1[Y] = t'_2[Y](= t[V])$. Donc $t'_2[XYZ] = t[UVW]$, et t'_2 est le tuple t' qu'on recherchait.
   On a donc $d \models R[UVW] \subseteq S[XYZ]$. Puisque celà est vrai quelquesoient d et t, on a $F \cup I \vdash R[UVW] \subseteq S[XYZ]$.
3. Soit la relation de la table 1.
   (a) Écrire la liste de toutes les DF que vous pouvez trouver. Essayez d'être méthodiques pour ne pas en oublier. Soit F l'ensemble des DF trouvées.
   Utiliser une approche par niveau sur les parties gauches. Pour la fermeture, on détecte les valeurs identiques à gauche, et on élimine à droite tous ceux qui n'ont pas de valeurs identiques sur les mêmes tuples.

partie gauche	fermeture (donc partie droite maximale)
A	A
B	BDE
C	E
D	BE
E	E

No. Tuple	A empno	B depno	C annee	D depnom	E dir
1	1	1	85	Biochimie	5
2	1	5	94	Admission	12
3	2	2	92	Informatique	2
4	3	2	98	Informatique	2
5	4	3	98	Géophysique	2
6	5	1	75	Biochimie	5
7	6	5	88	Admission	12

Tab. 1 – Affectation des employés à un département

On passe au niveau 2.

partie gauche	fermeture (donc partie droite maximale)
AB	ABCDE
AC	ABCDE
AD	ABCDE
AE	ABCDE
BC	ABCDE
BD	BDE
BE	BDE
CD	ABCDE
CE	CE
DE	B

Au niveau 3, on ne considère que les parties gauches dont aucun sous-ensemble n'est une clé ; sinon, on ne peut rien rajouter à droite.

partie gauche	fermeture (donc partie droite maximale)
BDE	BDE

On ne peut donc pas aller plus haut.

On peut alors construire les règles de la façon suivante :

$$X \longrightarrow (X^+ - X) - \bigcup_{Y \subseteq X} Y^+$$

ce qui donne :

$B \longrightarrow DE$
$C \longrightarrow E$
$D \longrightarrow BE$
$AB \longrightarrow C$
$AC \longrightarrow BD$
$AD \longrightarrow C$
$AE \longrightarrow BCD$
$BC \longrightarrow A$
$CD \longrightarrow A$

(b) Calculez une couverture minimum pour F.

Algo donné en cours :

Require: F un ens de DF
Ensure: G une couverture minimum de F
  $G := \emptyset$;
  for all $X \longrightarrow Y \in F$ **do**

$G := G \cup \{X \longrightarrow X^+\}$;
end for
for all $X \longrightarrow X^+ \in G$ **do**
if $G - \{X \longrightarrow X^+\} \vdash X^+$ **then**
$G := G - \{X \longrightarrow X^+\}$;
end if
end for
Return G.

Notez que le résultat n'est pas unique, ca dépend dans quel ordre on considère les DF dans la deuxième boucle. En revanche le nombre de DF est minimal pour représenter F.