LIF10 – FONDEMENTS DES BASES DE DONNÉES
TD3 – fermeture d'un ensemble d'attributs

Licence informatique – Automne 2014–2015

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

Algorithme 1: fermeture d'un ensemble d'attributs par des dépendances fonctionnelles

Data : Σ un ensemble de DF, X un ensemble d'attributs.
Result : X^+, la fermeture de X par Σ.

unused $:= \Sigma$
closure $:= X$
repeat
  closure' $:=$ *closure*;
  if $W \to Z \in$ *unused and* $W \subseteq$ *closure* **then**
    unused $:=$ *unused* $- \{W \to Z\}$;
    closure $:=$ *closure* $\cup Z$;
until *closure'* $=$ *closure*;
return *closure*;

Exercice 1 : inférence de dépendances (†)

Soit Σ l'ensemble des dépendances suivantes

$$BC \to A \qquad D \to BE$$
$$AC \to B \qquad B \to DE$$
$$AE \to C \qquad C \to E$$

1. (†) En utilisant les règles le système d'Armstrong augmenté des règles de composition et décomposition, prouver que les DF suivantes appartiennent à $\Sigma^+ = \{f \mid \Sigma \models f\}$:
   1. (†) $AD \to C$
   2. (†) $AB \to C$
   3. (†) $AE \to BD$
   4. $AC \to D$
   5. $CD \to A$
2. Même question en calculant la fermeture des parties gauches à partir de l'algorithme de fermeture.

Exercice 2 : correction de l'algorithme (†)

On rappelle qu'une preuve de $X \to Y$ à partir de Σ notée $\Sigma \vdash X \to Y$ est une *séquence* $\langle f_0, \dots, f_p \rangle$ de DFs telle que $f_p = X \to Y$ et $\forall i \in [0..p]$ on a soit $f_i \in \Sigma$, soit f_i est la conséquence de l'application d'une des règles d'Armstrong dont toutes les prémisses $f_0 \dots f_p$ apparaissent avant f_i dans la séquence. On donne le lemme

$$\Sigma \vdash X \to Y \text{ ssi } Y \subseteq X^\star$$

Cet exercice a pour objectif de montrer que l'algorithme donné calcule bien $X^\star = \{A \mid \Sigma \vdash X \to A\}$ à partir de X et Σ donnés. Pour cela, on considère la trace de l'algorithme définie comme une séquence $\langle X_0 \dots X_n \rangle$ de longueur maximale telle que $X_0 = X$, $X_i \subset X_{i+1}$ et $X_{i+1} = X_i \cup Z$ avec $W \to Z \in \Sigma$ et $W \subseteq X_i$. Notons que l'on fait abstraction du maintien de *unused* et *closure'* par soucis de simplification.

1. (†) Montrer par récurrence sur la longueur de la trace que $X_n \subseteq X^\star$. Pour l'étape de récurrence $X_i \subseteq X^\star$, pour $A \in X_{i+1}$ on considèrera les cas $A \in X_i$ ou $A \notin X_i$ en on utilisera le lemme.
2. Conclure que $X_n \subseteq \{A \mid \Sigma \models X \to A\}$.
3. Prouver par induction que $X \subseteq Y$ implique $X_n \subseteq Y_n$.
4. Prouver par induction sur la longueur d'une preuve de $\Sigma \vdash X \to Y$ que $X^\star \subseteq X_n$. Pour l'induction on procède cas par cas pour chaque règle d'Armstrong (utiliser la proposition précédente).
5. Conclure sur la correction de l'algorithme.

Exercice 3 : fermeture algébrique

En algèbre abstraite on appelle fermeture une application $\phi : \wp(E) \to \wp(E)$ qui soit :
- *extensive* $X \subseteq \phi(X)$
- *croissante* $X \subseteq Y \Rightarrow \phi(X) \subseteq \phi(Y)$
- *idempotente* $\phi(\phi(X)) = \phi(X)$

1. Justifier que la fermeture X^+ est bien nommée. On utilisera la correction de l'algorithme.
2. En vous appuyant sur les propriétés précédentes calculer l'ensemble des fermés pour l'ensemble des dépendances Σ défini dans l'exercice 1, i.e., éviter d'énumérer des éléments « inutiles » parmi les $2^{|ABCDE|}$ solutions possibles.

Corrections

Solution de l'exercice 1

1. 1. $$\dfrac{\dfrac{\dfrac{D \to BE}{AD \to ABE}\text{aug.} \qquad \dfrac{AE \subseteq ABE}{ABE \to AE}\text{refl.}}{AD \to AE}\text{trans.} \qquad AE \to C}{AD \to C}\text{trans.}$$
   2. $AB \to C$
      $B \to DE$ donc $AB \to ADE$ par augmentation
      $AB \to AE$ par décomposition
      $AB \to C$ puisque $AE \to C$ par transitivité
   3. $AE \to BD$
      $AE \to C$ donc $AE \to AC$ par augmentation
      $AC \to B$ donc $AE \to B$ par transitivité
      $B \to DE$ donc $B \to D$ par décomposition
      On en déduit $AE \to D$ par transitivité
      Par composition, on a $AE \to BD$.
   4. $AC \to D$
      $B \to DE$ donc $B \to D$ par décomposition
      Or $AC \to B$ donc donc $AC \to D$ par transitivité
   5. $$\dfrac{\dfrac{\dfrac{D \to BE}{D \to B}\text{decomp.}}{CD \to BC}\text{aug.} \qquad BC \to A}{CD \to A}\text{trans.}$$
2. On se limite à la première et la dernière demandées :
   1. Pour $AD \to C$, les étapes successives de l'algorithme sont les suivantes :
      (a) *closure* $= AD$
      (b) *closure* $= ABDE$ en utilisant $D \to BE$
      (c) *closure* $= ABCDE$ en utilisant $AE \to C$
      (d) comme $C \subseteq AD^+$, on en déduit $AD \to C$ par correction de l'algorithme
   2. Pour $CD \to A$, les étapes successives de l'algorithme sont les suivantes :
      (a) *closure* $= CD$
      (b) *closure* $= BCDE$ en utilisant $D \to BE$
      (c) *closure* $= ABCDE$ en utilisant $BC \to A$
      (d) comme $A \subseteq CD^+$, on en déduit $CD \to A$ par correction de l'algorithme

Solution de l'exercice 2

On peut expliquer que pour prouver $A \subseteq B$ on considère $x \in A$ arbitraire et on montre que $x \in B$.

1. Pour le cas de base $i = 0$, pour tout $A \in X$ on obtient une preuve de $\Sigma \vdash X \to A$ par la règle de réflexivité et on a bien $X_0 \subseteq X^\star$. Pour l'induction, supposons que $X_i \subseteq X^\star$, il faut montrer que $X_{i+1} \subseteq X^\star$ également. Soit $W \to Z$ la dépendance utilisée pour passer de X_i à X_{i+1}. Considérons un attribut $A \in X_{i+1}$:
   – soit $A \in X_i$ et $A \subseteq X^\star$ par hypothèse d'induction ;
   – soit $A \notin X_i$ et donc $A \in Z$. On a $W \subseteq X_i$ et $X_i \subseteq X^\star$ par hypothèse d'induction. On a $W \subseteq X^\star$ par transitivité et ainsi $\Sigma \vdash X \to W$ d'après le lemme. Par transitivité avec $W \to Z$, on prolonge la preuve en $\Sigma \vdash X \to Z$. Comme $A \in Z$ on a $\Sigma \vdash Z \to A$ et on peut encore prolonger par réflexivité et transitivité pour avoir $\Sigma \vdash X \to A$ et donc $A \in X^\star$.
2. Par adéquation des axiomes d'Armstrong, on a $X^\star \subseteq X^+ = \{A \mid \Sigma \models X \to A\}$
3. Pour le cas de base on a $X = X_0 \subseteq Y_0 = Y$. Pour l'induction considérons un attribut $A \in X_{i+1}$:
   – soit $A \in X_i$ et $X_i \subseteq Y_i$ par induction. On a $A \in Y_i \subseteq Y_{i+1}$ et ainsi $X_{i+1} \subseteq Y_{i+1}$.

– soit $A \notin X_i$ et donc $A \in Z$ avec $W \to Z$ la dépendance utilisée pour passer de X_i à X_{i+1}. On a $X \subseteq X_i$ et $X_i \subseteq Y_i$ par induction. On peut donc passer de Y_i à Y_{i+1} en utilisant $W \to Z$ et ainsi $A \in Y_{i+1}$.

4. Pour le cas de base, on considère une preuve $\langle f_0 = X \to Y \rangle$ de longueur un. Ainsi,
   – soit $X \to Y$ est déduit par réflexivité et $Y \subseteq X = X_0$;
   – soit $X \to Y \in \Sigma$, on a ainsi $X_1 = X_0 \cup Y$ car $X \subseteq X$ et donc $Y \subseteq X_1$
   Pour l'induction on suppose que $X^\star \subseteq X_n$ pour une preuve de longueur p et on va montrer qu'il est de même pour preuve $\Sigma \vdash X \to Y$ de longueur $p+1$. Une des trois règles d'Armstrong peut être utilisée à la dernière étape d'une preuve de longueur $p+1$:

   pour le cas de la *réflexivité* ou le cas où $f_{p+1} \in \Sigma$: l'argument est le même que pour le cas de base (l'argument inductif est inutile) ;

   pour le cas de la *transitivité* : on a deux preuves $\Sigma \vdash X \to Z$ et $\Sigma \vdash Z \to Y$ de longueurs inférieurs ou égale à p. Pour la première on a $Z \subseteq X^\star \subseteq X_n$ par le lemme et par induction et comme la fermeture est croissante, on obtient $Z_n \subseteq X_n$. Pour la seconde on a $Y \subseteq Z^\star \subseteq Z_n$ avec le même argument (on lance l'algorithme pour calculer Z_n). On obtient ainsi $Y \subseteq X_n$.

   pour le cas de l'*augmentation* : on a une preuve $\Sigma \vdash V \to W$ de longueur inférieure ou égale à p avec $X = ZV$ et $Y = ZW$ et il faut montrer que $ZW \subseteq (ZV)_n$. Par induction on a $W \subseteq V_n$. On a $V_n \subseteq (ZV)_n$ et $Z_n \subseteq (ZV)_n$ car la fermeture est croissante (avec $V \subseteq ZV$ et $Z \subseteq ZV$). Par transitivité de l'inclusion on a $W \subseteq (ZV)_n$. Comme $Y = Z \cup W$ et qu'on a à la fois $Z \subseteq Z_n \subseteq (ZV)_n$ et $W \subseteq (ZV)_n$ on conclu que $Y \subseteq (ZV)_n$

5. On a prouvé que $X^\star \subseteq X_n$, par complétude on a $X^+ \subseteq X^\star$ et ainsi $X^+ \subseteq X_n$. Comme d'autre part on a également prouvé que $X_n \subseteq X^+$ on obtient $X^\star = X^+$ et l'algorithme est correct.

Solution de l'exercice 3

1. – extensive : à chaque étape X_i croit et $X_0 = X$ on a donc $X \subseteq X_n$.
   – croissante : c'est la question 3 de l'exercice précédent.
   – idempotente : intuitivement on essaie de saturer à nouveau un ensemble qui l'est déjà, plus formellement, la séquence $\langle X_0 \dots X_n \rangle$ de la première exécution de l'algorithme étant de longueur maximale, la prochaine exécution sur X^+ ne fera rien de plus.
2. Faire une approche par niveaux, calculer les fermetures de tous les singletons, puis des couples d'attributs ...

niveau 1 :
$A^+ = A$
$B^+ = BDE$
$C^+ = CE$
$D^+ = BDE$
$E^+ = E$
niveau 2 :
$AB+ = R$
$AC^+ = R$
$AD^+ = R$
$AE^+ = R$
$BC+ = R$
$BD+ = BDE$
$BE+ = BDE$
$CD+ = BCDE$
$CE+ = CE$

Se convaincre qu'il est inutile de continuer (niveau 3) car on ne pourra pas trouver de nouveaux fermés.

$Cl(\Sigma) = \{A, E, CE, BDE, BCDE, R\}$

Eventuellement finir avec la construction de la relation d'Armstrong.