LifBDW2 – Bases de données avancées

TD5 – Couvertures, dépendances et logique

Licence informatique – Automne 2020–2021

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

Algorithme 1: calcul d'une couverture minimale

Data : F un ensemble de DF
Result : G une couverture minimale 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 $G \setminus \{X \to X^+\} \vdash X \to X^+$ **then**
    $G := G \setminus \{X \to X^+\}$;
return G;

Algorithme 2: réduction des attributs d'un ensemble minimal de DFs

Data : Un ensemble *minimal* de DF F sur R.
Result : F avec un nombre réduit 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 1 : comparaisons des couvertures minimales

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.

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

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

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

FIGURE 1 – Axiomatisation de Casanova pour les DIs

2. En utilisant l'algorithme 1 déterminer une couverture minimale de F_1. Comparer la couverture ainsi obtenue à F_2.
3. En utilisant l'algorithme 1 déterminer si $F = \{A \to BC, D \to E, C \to D\}$ est minimal.

Exercice 2 : couverture minimale et réduite

Soit Σ l'ensemble de DFs $\Sigma = \{A \to B; A \to C; D \to E; C \to D; B \to C; BC \to A\}$

1. Donner une couverture minimum de Σ en utilisant l'algorithme 1.
2. Réduire la couverture minimum en appliquant l'algorithme 2 de réduction des parties droites et gauches.

Exercice 3 : couverture minimale et réduite

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 minimale et réduite de F.

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

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

Exercice 5 : dépendances fonctionnelles et logique propositionnelle ([1])

Dans cet exercice on va s'intéresser au lien entre les dépendances fonctionnelles $X \to Y$ et les formules de la logique propositionnelle classique de la forme $X \Rightarrow Y$ pour montrer que la correspondance entre les notations n'est pas simplement superficielle.

Soit Σ un ensemble de DFs sur un schéma $\mathbf{R}$. A chaque attribut $A \in \mathbf{R}$ on associe une *variable propositionnelle* $\underline{A} \in \mathcal{P}$. On étend cette transformation aux DFs en faisant correspondre à une DF $A_1 \dots A_p \to B_1 \dots B_q \in \Sigma$ la formule de logique propositionnelle $\underline{A}_1 \wedge \dots \wedge \underline{A}_p \Rightarrow \underline{B}_1 \wedge \dots \wedge \underline{B}_q$ sur l'ensemble de propositions $\mathcal{P}$. Enfin, on note $\underline{\Sigma}$ l'ensemble de formules logiques correspondant aux DFs de Σ en étendant la transformation à un ensemble de dépendances. Soit $\alpha = A_1 \dots A_p \to B_1 \dots B_q$ une DF. On va montrer que les trois propositions suivantes sont équivalentes :

$$\Sigma \models \alpha \tag{1}$$
$$\Sigma \models_2 \alpha \tag{2}$$
$$\underline{\Sigma} \models_P \underline{\alpha} \tag{3}$$

Dans ces formules, $\Sigma \models \alpha$ désigne l'implication logique entre un ensemble de DFs et une DF, $\Sigma \models_2 \alpha$ désigne l'implication logique de DFs limitée *aux relations ne contenant que deux tuples*, autrement dit $\Sigma \models_2 \alpha \equiv \forall r.(|r| = 2 \wedge r \models \Sigma) \Rightarrow r \models \alpha$. Enfin, $\underline{\Sigma} \models_P \underline{\alpha}$ désigne l'implication logique classique, c'est-à-dire, que pour toute valuation des variables propositionnelles $\nu : \mathcal{P} \to \{0, 1\}$ telle que $\nu \models \underline{\Sigma}$ on a aussi $\nu \models \underline{\alpha}$. On donne le lemme suivant :

Lemme 1. *Soient $\nu : \mathcal{P} \to \{0, 1\}$ une valuation des variables propositionnelles, $r = \{t_1, t_2\}$ la relation à deux tuples tels que $t_1[A] = 1$ et $t_2[A] = \nu(\underline{A})$ pour tout $A \in \mathbf{R}$ et $\alpha = A_1 \dots A_p \to B_1 \dots B_q$ une dépendance arbitraire. Alors on a $\nu \models_P \underline{\alpha}$ si et seulement si $r \models \alpha$.*

1. Montrer que la proposition (1) est équivalente à la proposition (2). Pour le sens (2) $\Rightarrow$ (1) on procédera par l'absurde en montrant que supposer $\Sigma \models_2 \alpha$ et $\Sigma \not\models \alpha$ conduit à une contradiction.
2. Montrer le lemme 1 en construisant une relation à deux tuples « à la Armstrong » à partir d'une interprétation ν donnée et vice-versa.
3. Montrer que la proposition (3) est équivalente à la proposition (2) en utilisant le lemme 1. Procéder par l'absurde pour les deux implications.
4. Conclure.

Références

[1] R. Fagin. Functional dependencies in a relational database and propositional logic. *IBM J. Res. Dev.*, 21(6) :534–544, Nov. 1977.