Bases de données avancées

Dépendances d'inclusion

Équipe pédagogique BD

https://perso.liris.cnrs.fr/marc.plantevit/doku/doku.php?id=lifbdw2_2018a

Version du 7 octobre 2018

Dépendances d'inclusion

Inférence de dépendances d'inclusion

Dépendances d'inclusion

Inférence de dépendances d'inclusion

Dépendances d'inclusion

Les Dépendances d'Inclusion (DI) se différencient des fonctionnelles sur plusieurs points :

1. les DI peuvent être définies entre attributs de relations différentes et possèdent un caractère global à la base, pas local à une relation.
2. Les DI sont définies non pas entre deux ensembles d'attributs, mais entre deux séquences d'attributs de même taille.

L'ordre des attributs est donc important pour les dépendances d'inclusion !

Définition des dépendance d'inclusions

Syntaxe des dépendances d'inclusion

Soit **R** un schéma de base de données. Une Dépendance d'Inclusion (DI) sur **R** est une expression de la forme

$$R[X] \subseteq S[Y]$$

avec $R, S \in \mathbf{R}$, X et Y des séquences d'attributs distincts respectivement de R et de S, avec $|X| = |Y|$.

Définition des dépendance d'inclusions

Intuitivement, une DI est satisfaite dans une base de données si toutes les valeurs prises par la partie gauche apparaissent dans la partie droite.

Sémantique des dépendances d'inclusion

Soit $d = \{r_1, r_2, \ldots, r_n\}$ une base de données sur un schéma $\mathbf{R} = \{R_1, \ldots, R_n\}$. Une dépendance d'inclusion $R_i[X] \subseteq R_j[Y]$ sur $\mathbf{R}$ est *satisfaite* dans d, noté $d \models R_i[X] \subseteq R_j[Y]$, *ssi*

$$\forall t_i \in r_i, \exists t_j \in r_j . t_i[X] = t_j[Y]$$

(ou de manière équivalente *ssi* $\pi_X(r_i) \subseteq \pi_Y(r_j)$)

Exemple

Tout titre projeté actuellement de la relation Programme *est le titre d'un film apparaissant dans la relation* Films

$$Programme[Titre] \subseteq Films[Titre]$$

Exemple (1)

- Supposons des schémas de relation pour décrire les modules :

$$MODULE = \{\underline{NUMM}; INTITULE; DESC\}$$

- et un schéma de relation pour décrire les séances de cours :

$$SEANCE = \{\underline{DATE; NUMM}; NUMSALLE\}$$

- Pour imposer que les numéros de modules dans les séances *soient bien des modules qui existent*, on définit la dépendance d'inclusion :

$$SEANCE[NUMM] \subseteq MODULE[NUMM]$$

Exemple (2)

- Supposons maintenant un schéma de relation *EMPRUNT* modélisant les réservations de vidéos pour les séances, sous la forme :

$$EMPRUNT = \{DATE, NUMM; NUMPROF; NUMVIDEO\}$$

- Pour assurer la cohérence de la base, on doit préciser que les vidéos doivent être réservés pour des séances existantes dans la base on définira alors la DI :

$$EMPRUNT[DATE, NUMM] \subseteq SEANCE[DATE, NUMM]$$

Si on inverse les attributs à gauche par exemple,
la DI change complètement de signification !

Clé étrangère

Une *contrainte d'intégrité référentielle* est une DI dont la partie droite est une clé

- La partie gauche d'une contrainte d'intégrité référentielle est appelée *clé étrangère*

Les DI ne sont pas toutes des clés étrangères

- On souhaite imposer que tous les cours possèdent au moins une séance dans l'année, on définira alors une DI :

$$COURS[NUMCOURS] \subseteq SEANCE[NUMCOURS]$$

- Tous les cours apparaîtront donc au moins une fois dans la relation des séances
- Mais $NUMCOURS$ n'est pas une clé de $SEANCE$
- Cette DI n'est donc pas une contrainte d'intégrité référentielle.

Dépendances d'inclusion

Inférence de dépendances d'inclusion

Système d'inférence de Casanova *et al.*

Soit I un ensemble de DI sur un schéma de base de données **R**. Les règles d'inférence suivantes sont appelées système d'inférence de Casanova *et al.* pour les DI, dans lequel σ est une permutation d'un sous-ensemble de $\{1..n\}$:

- Réflexivité

$$R[X] \subseteq R[X]$$

- Permutation & projection

$$\frac{R[A_1 ... A_n] \subseteq S[B_1 ... B_n]}{R[A_{\sigma(1)} ... A_{\sigma(k)}] \subseteq S[B_{\sigma(1)} ... B_{\sigma(k)}]}$$

- Transitivité

$$\frac{R[X] \subseteq S[Y] \qquad S[Y] \subseteq T[Z]}{R[X] \subseteq T[Z]}$$

Propriétés du système

- Ce système d'inférence est lui aussi
  - correct
  - complet
- Contrairement aux DF, le problème d'inférence des DI est très difficile à traiter dans le cas général (PSPACE-complet) ;
- La notion I^+ s'applique aussi pour noter la fermeture d'un ensemble de DI.
- En revanche, la notion de fermeture d'un *ensemble* d'attributs par rapport à un ensemble de DI ne s'applique pas, car les DI manipulent des séquences et non des ensembles.

Combinaison DF et DI

- soit on inférait des DF à partir d'un ensemble de DF,
- soit on inférait des DI à partir d'un ensemble de DI.

Mais il existe des interactions entre DF et DI

Proposition

Si $|X| = |T|$, les propriétés suivantes sont vérifiées

1. $\{R[XY] \subseteq S[TU], S : T \to U\} \models R : X \to Y$
2. $\{R[XY] \subseteq S[TU], R[XZ] \subseteq S[TV], S : T \to U\} \models R[XYZ] \subseteq S[TUV]$
3. $\Sigma = \{R[XY] \subseteq S[TU], R[XZ] \subseteq S[TV], S : T \to U\}$, si $r \models \Sigma$, $s \models \Sigma$ et $u \in r$ alors $u[Y] = u[Z]$

Fin.