BD TD3.

Calcul de la couverture minimale et formes normales.

Algorithme de calcul d’une fermeture d'un ensemble d'attributs :

1 Comment trouver la fermeture d'un ensemble d'attributs :

R un schéma de relation, R=(A1,A2,…An)

F un ensemble de dépendances fonctionnelles (dfs) définies sur R.

Nous donnons en langage algorithmique, un algorithme permettant de calculer la fermeture d'un ensemble d'attributs X notée **X+**.

**Algorithme de calcul de X+ :**

X+:= X

Répéter

AUX := X+

Pour chaque df Y → Z de F faire

si Y est inclus dans X+ alors X+ := X+U Z

Jusqu'à AUX = X+ ou X+= R

Exemple

Soit R = (A, B, C, D, E) et F= {AB → C, B → D, CD → E}. Calculons (AB)+.

Appliquons d'abord l'algorithme en prenant les df dans l'ordre suivant:

CD→ E, B → D, AB → C

X+:= AB

1ere boucle du répéter

AUX := AB

CD→ E n'augmente pas X+= AB

B→ D donne X+:= ABD

AB→C donne X+:= ABCD

Comme on a AUX≠X+ on continue.

2eme boucle du répéter

AUX:= ABCD

CD→ E augmente X+ := ABCDE

B→ D et AB→ C ne peuvent rien changer puisque X+= R.

Comme X+= R, l'algorithme s'arrête.

Remarquons que nous venons de montrer que AB est une clé de R.

L'ordre choisi pour les dfs influe sur la rapidité de l'algorithme.

Appliquons maintenant l'algorithme en prenant les dfs dans l'ordre suivant :

AB-> C, B --> D, CD --> E

X+:= AB

1ere boucle du répéter:

AUX := AB

AB→ C donne X+:= ABC

B→ D donne X+:= ABCD

CD→ E donne X+:= ABDCDE

Comme X+ = R, l'algorithme s'arrête.

2 Algorithme pour trouver une couverture minimale F° d'un ensemble F de df

Notations: X, Y désignent des ensembles d'attributs,

A, B, C,... désignent des attributs.

Données: R un schéma de relation, R= (A1, A2, . . . An).

F un ensemble de df définies sur R.

Résultat: G une couverture minimale de F, vérifiant :

(1) G+ = F+.

(2) Tout membre droit d'une df de G est réduit à un seul attribut.

(3) Pour aucune df X→A de G on n'a G - (X→ A) |=X→ A.

(4) Pour aucune df X→A de G on n'a G |= Y→ A avec Y partie stricte de X.

Algorithme

On ordonne les df de F, supposons F= {X1→Y1, X2→Y2, . . ., Xm→Ym}

Etape 1: on décompose les membres droits Yi des dfs de F

Pour i de 1 à m faire

si Yi = A1 A2... As, avec s>1

alors F := F –{Xi → Yi} U {Xi→A1, Xi→A2, . . ., Xi→As}

Nous supposerons par la suite que l'on a

F= {Xi →Y1, X2 →Y2,..., Xp →Yp}

Etape 2: on regarde si on peut enlever des dfs de F sans modifier sa fermeture.

Pour i de 1 à p faire,

Si F- {Xi →Ai}|= {Xi →Ai} alors F := F - {Xi →Ai}

Quitte à renuméroter les dfs, nous supposerons qu'à la fin de cette étape nous avons:

F= {Xi →Y1, X2 →Y2,..., Xp →Yq} avec q≤p.

Etape 3: on cherche si on peut remplacer les membres gauches des dfs de F qui sont formés de plus d'un attribut par des membres gauches contenant moins d'attributs sans changer la fermeture de F.

Pour i de 1 à q faire

Si Xi = B1 B2... Br, avec r >1 alors

Pour j de 1 à r faire

Si F |= {Xi-Bj} → Ai alors Xi := Xi -Bj

Exemple 1

Considérons un schéma de relation R concernant des cours C, des étudiants E, des professeurs P, des notes N, des jours et horaires de cours Jet H, et des salles de cours S. Soit R = (C, E, P, N, J, H, S) muni de l'ensemble F des dfs

CEP → N…...(1)

JHS → PC.….(2)

EP → C……..(3)

EPS → C……(4)

Appliquons l'algorithme.

Etape 1

F devient l'ensemble suivant

CEP → N (1)

JHS → P (2')

JHS → C (2")

EP → C (3)

EPS → C (4)

Etape 2

On ne peut retirer la df (1) de F car les df (2'), (2"), (3), (4) ne contiennent pas N au second membre, et par suite ne peuvent impliquer (1). Par un argument similaire, on montre que l'on ne peut retirer (2') à F. Pour voir que l'on ne peut retirer (2") de F, on calcule (JHS)+ relativement F- (2"). On trouve (JHS)+= JHSP, donc C n'est pas dans (JHS)+.De même, (EP)+ relativement à F -(3) vaut EP et ne contient pas C, donc on ne peut supprimer (3). Par contre, la df (3) implique la df (4), donc F- (4) implique (4) et on peut enlever (4) à F.

A la fin de l'étape 2, on a donc F° ={ (1), (2), (2"), (3) ).

Etape 3

On considère la df (1). On teste si F implique la df EP → N (1').

Sans calculer (EP)+, on voit que les df (3) et (1) impliquent EP → N par pseudo-transitivité.

Donc on peut remplacer F par {F- (1)} U {(1')}.

Par contre E+= E et P+= P, donc on ne peut pas faire "maigrir" davantage le membre gauche de la df (1').

On recommence la même recherche avec la df (2'). Mais comme aucune df de F n'a de membre gauche inclus dans une partie stricte de JHS, on a (HS)+= HS, (JS)+= JS et (JH)+= JH, et on ne peut remplacer la df (2') par

HS → P ou JS → P ou JH → P.

On obtient les mêmes résultats avec les df (2") et (3).

A la fin de l'algorithme, on a donc la couverture minimale F° = { (1'), (2'), (2"), (3) }.

Exemple 2

Soit R = (A, B, C, D, E) muni de l'ensemble F de df suivant

C→A (1)

CE → B (2)

ACD→ B (3)

CE → D (4)

D→ E (5)

BC→ D (6)

Si on applique l'algorithme sans changer l'ordre des dfs ni des étapes, on obtient à la fin de l'étape 2

F = { (1), (3), (4), (5), (6) }

Puis à la fin de l'étape 3, la couverture minimale

F1°= { (1), CD → B, (4), (5), (6) }.

Modifions l'ordre des dfs, échangeons les df (2) et (3) avant d'appliquer l'algorithme. A la fin de l'étape 2 nous avons

F2° = { (1), (2), (5), (6)}

L'étape 3 ne modifie pas F2 qui est donc une couverture minimale différente de F1.

Cet exemple montre qu'il n'y a pas en général unicité de la couverture minimale et que le nombre de dfs dans une couverture minimale n'est pas fixe.

Ici nous trouvons la même couverture que précédemment si nous inversons les étapes 2 et 3 de l'algorithme.

Reprenons F avec l'ordre (1), (3), (2), (4), (5), (6). Si nous appliquons l'étape 3 avant l'étape 2 nous trouvons à la fin de l'étape 3 F= { (1), CD → B, (2), (4), (5), (6) }

A la fin de l'étape 2 appliquée ensuite, la couverture minimale obtenue est

F3° = { (1), (2), (5), (6) }.

Exercices :.

1. Soit R (A, B, C, D, E, F)

Monter en appliquant l’algorithme de calcul d’une couverture minimale que :

1. Si

T1 = {A -> B; B -> CDE; C -> E; DE -> F; F -> G}

Alors

T1° = {A -> B; B -> C; C -> E; B -> D; DE -> F; F -> G}

Quelle est la clé de R ?

1. Si

T2 = {A -> BCD; B -> AE; C -> D; D -> A}

Alors

T2°= {A -> B; A -> C; C -> D; B -> A; B -> E; D -> A}

1. Quelle est la clé de R ?
2. Quelle est la FN de R ? Justifier.
3. Proposer une décomposition de R si elle n’est pas en 3FN
4. Soit le schéma relationnel R(A,B,C,D,E,H) et l’ensemble des dépendances fonctionnelles associé **F**={AB**→C ;B→D;D→E ; A→C**}
5. Déterminer les attributs clés et les attributs non clés
6. Quelles sont les clés candidates de R
7. Quelle est la FN de R ? Justifier.
8. Proposer une décomposition de R si elle n’est pas en 3FN