Chapitre I : Le calcul relationnel

Introduction : Le calcul relationnel est un langage non procédural basé sur le calcul de prédicats du premier ordre qui est un langage logique possédant une syntaxe et une sémantique formelles. Contrairement à l'algèbre relationnelle, le calcul relationnel permet de dire ce que l'on veut obtenir mais pas comment l'obtenir.

I. Calcul relationnel à variables n_uplets

I. 1. Définition

Une requête en calcul relationnel à variable n_uplet est telle que les variables qui y figurent sont des enregistrements.

I. 2. Notion de prédicat

Un prédicat **P** est une expression booléenne (évaluée à ***Vrai*** ou ***Faux***) qui peut comporter des paramètres. Les prédicats sont de la forme :

- r(t) <=> t ∈ r où r est l'instance d'une relation R ;
- t.A Θ Valeur ; // t est un n_uplet (enregistrement) et A est un attribut de la relation R ;
- t_1.Attribut_n Θ t_2.Attribut_m ; // Θ est un opérateur de comparaison ;
- Toute combinaison de ces prédicats est un prédicat.

Exemple

La relation **Etudiant** (<u>Matricule</u>, Nom, Prénom, Age, Adresse, VilleNaiss) d'instance **e**.

Les requêtes suivantes sont écrites en calcul relationnel comme suit :

1. Les étudiants dont le Prénom "Moustapha" ?

{t / e(t) ∧ (t.Prénom = 'Moustapha')}

2. Les étudiants qui habitent à Lyndiane ?

{t / e(t) ∩ (t.Adresse = 'Lyndiane')}

3. Nom et Prénom des étudiants habitant "Boucotte" ou "Tilene" et ayant moins de 22 ans.

{t.Nom, t.Prénom / e(t) ∩ ((t.Adresse = 'Boucotte') ∨ (t.Adresse='Tilene')) ∩ (t.Age < 22)}

I. 3. Notion de quantificateur

Il existe deux quantificateurs permettant de lier les variables :

I. 3. 1. Le quantificateur ∃ (il existe)

∃ t / P(t) est vrai s'il existe un n_uplet **t** dans la base qui vérifie le prédicat **P(t)**.

Exemple

Soit la relation **Salle** (<u>NumSalle</u>, <u>NumBat</u>, Capacité, AnnéeConst) ayant pour instance **s**.
Quels sont les bâtiments construits la même année ?

$\{t \ / \ s(t) \land \exists \ t_1 \in s \land (t_1.\text{AnnéeConst} = t.\text{AnnéeConst}) \land (t_1.\text{NumBat} \neq t.\text{NumBat})\}$

I. 3. 2. Le quantificateur ∀ (quelque soit)

∀ t P(t) signifie que pour tous les n_uplets le prédicat P(t) est vrai.

Exemple

On reprend la relation de l'exemple précédent :
Quels sont les bâtiments dont toutes les salles ont la même capacité ?

$\{t \ / \ s(t) \land \forall \ t_1 \in s, (t_1.\text{Capacité} = t.\text{Capacité})\}$

I. 4. Formalisme d'une requête en calcul relationnel à variable n-uplets

Une expression en calcul relationnel à variable n_uplet est de la forme :

{t / P(t)} => L'ensemble des n_uplets **t** tels que le prédicat **P(t)** soit vrai.

Dans cette expression, P est une formule et t est un n_uplet. Plusieurs variables n_uplets peuvent apparaître dans une même formule.

I. 4. 1. Formule

a) Formule atomique : Une formule est atomique si elle permet de comparer deux valeurs d'attribut ou une valeur d'attribut et une constante.

Les opérateurs de comparaison utilisés sont : =, <, >, <=, >=, <>.

Exemple

Avec toujours la même relation Salle d'instance s :

- x ∈ s ;
- x.Capacité = y.Capacité ;
- z.NumSalle = 5.

b) Formule complexe : Une formule complexe est une combinaison de formules atomiques reliées par des connecteurs (∧ ,∨,¬) et des quantificateurs (∃ , ∀).

Remarque : Soit P et Q des formules alors :

- ¬P, P ∧ Q, P ∨ Q sont des formules ;
- ∃ r, (P(r)) est une formule ;
- ∀ r, (P(r)) est une formule.

Remarque :

- Les quantificateurs ∃ r et ∀ r permettent de lier la variable r. Une variable non liée est dite variable libre.
- Une restriction importante s'impose à la définition d'une requête {t / P(t)} : Seules les variables t qui apparaissent à la gauche du signe "/" doivent être des variables libres dans la formule P.

Les formules sont constituées à partir d'atomes de la manière suivante :

a) r(t) : **t** est une variable n_uplet et **r** une instance relationnelle ;

b) s.A ϴ r.B : **s** et **r** sont des variables n_uplets, **A** est un attribut de la relation dont **s** appartient à l'instance et **B** est un attribut de la relation dont **r** appartient à l'instance. ϴ est un opérateur de comparaison.

c) r.A ϴ C : **C** est une constante, ϴ un opérateur de comparaison, **r** un n_uplet et **A** un attribut.

I. 5. Opérations algébriques en calcul relationnel

Toutes les opérations de l'algèbre relationnelle peuvent être exprimées à l'aide des formules du calcul relationnel.

I. 5. 1. La sélection

σ_C (R) ⇔ {t / r(t) ∧ C}

- r est l'instance de la relation R ;
- **C** une condition de sélection ;
- t un n_uplet appartenant à r.

I. 5. 2. La projection

$\prod_{A1,A2,\ldots,Aj}(R) \Leftrightarrow \{t.A_1, t.A_2,\ldots,t.A_j \,/\, t \in r\}$

- r est l'instance de la relation R ;
- t un n_uplet tel que $j < n$;
- $A_1, A_2,\ldots, A_j$ des attributs de la relation R.

I. 5. 3. Le complément

Soit R($A_1, A_2, \ldots, A_n$**)** une relation d'instance r, le complément de la relation R noté –R s'exprime comme suit :

$-R \Leftrightarrow \{t.A_1, t.A_2, \ldots, t.A_n \,/\, \neg(r(t)) \wedge \exists\, t_1 \in r \wedge \exists\, t_2 \in r \wedge \exists\, t_3 \in r \wedge \exists\, t_4 \in r \wedge (t.A_1 = t_1.A_1) \wedge (t.A_2 = t_2.A_2) \wedge \ldots \wedge (t.A_n = t_n.A_n)\}$

I. 5. 4. L'union

$R_1 \vee R_2 \Leftrightarrow \{t \,/\, t \in r_1 \vee t \in r_2\}$

- r_1 et r_2 sont respectivement les instances des relations R_1 et R_2 ;
- t un n_uplet.

I. 5. 5. L'intersection

$R_1 \wedge R_2 \Leftrightarrow \{t \,/\, t \in r_1 \wedge t \in r_2\}$

- r_1 et r_2 sont respectivement les instances des relations R_1 et R_2 ;
- t est un enregistrement.

I. 5. 6. La différence

$R_1 - R_2 \Leftrightarrow \{t \,/\, r_1(t) \wedge \neg(r_2(t))\}$

- r_1 et r_2 sont les instances respectives des relations R_1 et R_2 ;
- t un n_uplet.

I.5.7. La division

Soient R_1 et R_2 deux relations de schéma respectif $(A_1, A_2, \ldots, A_n)$, $(A_3, A_4,\ldots, A_m)$ (avec $m < n$) et d'instance respective r_1 et r_2 : $R_1 (A_1, A_2, \ldots, A_n)$ et $R_2 (A_3, A_4,\ldots , A_m)$.

La division de R_1 par R_2 notée $R_1 \div R_2$ s'exprime comme suit en calcul relationnel :

$R_1 \div R_2 = \{t.A_1, t.A_2, t.A_{m+1} , \ldots, t.A_n \,/\, \forall\, t_1 \in r_2, \exists\, t_2 \in r_1 \wedge (t.A_1 = t_2.A_1) \wedge (t.A_2 = t_2.A_2) \wedge \ldots \wedge (t.A_{m+1} = t_2.A_{m+1}) \wedge (t_1.A_3 = t_2.A_3) \wedge (t_1.A_4 = t_2.A_4) \wedge \ldots \wedge (t_1.A_m = t_2.A_m)\}$

$R_1 \div R_2$ a pour schéma ($A_1, A_2, A_{m+1}, \ldots, A_n$)

I. 5. 8. Le produit cartésien

Soient R_1 et R_2 deux relations dont les attributs sont respectivement $A_1, A_2, \ldots, A_n$ et $B_1, B_2, \ldots, B_m$: $R_1 (A_1, A_2, \ldots, A_n)$ et $R_2 (B_1, B_2, \ldots, B_m)$

Le produit cartésien $R_1 * R_2$ avec r_1 et r_2 les instances respectives de R_1 et R_2 s'exprime comme suit :

$$\mathbf{R_1 * R_2 \Leftrightarrow \{t / \exists u \in r_1 \wedge \exists v \in r_2 \wedge (t.A_1 = u.A_1) \wedge \ldots \wedge (t.A_n = u.A_n) \wedge (t.B_1 = v.B_1) \wedge \ldots \wedge (t.B_m = v.B_m)\}.}$$

$R_1 * R_2$ a pour schéma ($A_1, A_2, \ldots, A_n, B_1, B_2, \ldots, B_m$)

I. 5. 9. La jointure naturelle

Soient R_1 d'attributs $A_1, A_2, \ldots, A_n, C_1, \ldots, C_p$ et R_2 d'attributs $B_1, B_2, \ldots, B_m, D_1, \ldots, D_p$. r_1 et r_2 étant respectivement les instances de R_1 et R_2.

$$\mathbf{R_1 \mid\!\!><\!\!\mid_C R_2 \Leftrightarrow \{t / \exists\ u, \exists\ v \wedge r_1(u) \wedge r_2(v) \wedge (t.A_1 = u.A_1) \wedge \ldots \wedge (t.A_n = u.A_n) \wedge (t.B_1 = v.B_1) \wedge \ldots \wedge (t.B_m = v.B_m) \wedge (u.C_1 = v.D_1) \wedge (t.C_1 = u.C_1) \wedge (t.D_1 = v.D_1) \wedge \ldots \wedge (u.C_p = v.D_p) \wedge (t.C_p = u.C_p) \wedge (t.D_p = v.D_p)\}}$$

$R_1 \mid\!\!><\!\!\mid_C R_2 = R (A_1, A_2, \ldots, A_n, C_1, \ldots, C_p, B_1, B_2, \ldots, B_m, D_1, \ldots, D_p)$.

C est la condition de jointure telle que : $C = ((C_1 = D_1) \wedge (C_2 = D_2) \wedge \ldots \wedge (C_p = D_p))$

Exemple : Soient les relations suivantes :

Etudiant (<u>NumCarte</u>, Nom, Prénom, Adresse, Filière, Age) d'instance **e**

Cours (<u>Code</u>, Libellé, VolHoraire) d'instance **c**

Suivre (<u>#Etudiant</u>, <u>#Cours</u>) d'instance **s**

Questions

1. Donnez les noms des étudiants âgés de 24 ans et qui habitent Diourbel ou Ziguinchor.
2. Donnez l'adresse et le nom de chaque étudiant qui suit le cours de base de données.
3. Quels sont les cours dont le volume horaire dépasse 24h ?
4. Quels sont les libellés des cours suivis par l'étudiant Moustapha SAMB ?
5. Quels sont les cours qui ont le plus grand volume horaire ?

Réponses

1. {t.Nom / e(t) ∧ (t.Age = 24) ∧ ((t.Adresse = 'Diourbel') ∨ (t.Adresse = 'Ziguinchor'))}

2. {t.Nom, t.Adresse / e(t) $\wedge$ $\exists$ $t_1 \in$ s, $\exists$ $t_2 \in$ c $\wedge$ (t.NumCarte = t_1.NumCarte) $\wedge$ (t_1.Code = t_2.Code) $\wedge$ (t_2.Libellé = 'Base de données')}
3. {t / c(t) $\wedge$ (t.Volume > 24)}
4. {t.Libellé / c(t) $\wedge$ $\exists$ $t_1 \in$ e $\wedge$ $\exists$ $t_2 \in$ s $\wedge$ (t.Code = t_2.Code) $\wedge$ (t.NumCarte = t_1.NumCarte) $\wedge$ (t_1.Nom = 'SAMB') $\wedge$ (t_1.Prénom = 'Moustapha')}
5. {t / c(t) $\wedge$ $\forall$ t_1 c(t_1) $\wedge$ (t.Volume >= t_1.Volume}

Remarque

De telles expressions peuvent générer un nombre infini de n_uplets. Ainsi, on restreint la définition des formules en introduisant la notion de domaine d'une formule. Le domaine d'une formule P est l'ensemble des valeurs qui y apparaissent explicitement. Le domaine de P est noté dom(P)

II. Calcul relationnel à variables domaines

A la différence du calcul relationnel à variables n_uplets, les variables portent sur les valeurs d'es attributs. Il utilise donc des variables domaines provenant d'un domaine d'attribut plus que d'une variable n_uplet.

II. 1. Formalisme d'une requête en calcul relationnel à variable domaine

Une requête en calcul relationnel à variable domaine est de la forme :

$$\{<x_1, x_2, \ldots, x_n> / P(<x_1, x_2, \ldots, x_n>)\}$$

1. $x_1, x_2, \ldots, x_n$ représentent des variables domaines ;
2. P est une formule composée d'atomes de la forme :
   - R($x_1, x_2, \ldots, x_n$) où R est une relation de n attributs et $x_1, x_2, \ldots, x_n$ sont des variables domaines ;
   - x Θ y où x et y sont des variables domaines et Θ un opérateur de comparaison. x et y doivent être des domaines compatibles ;
   - x Θ C où C est une constante du domaine de l'attribut dont x est une variable domaine.

II. 2. Propriétés sur les formules

- Si P_1 et P_2 sont des formules alors $\neg P_1$, $P_1 \vee P_2$ et $P_1 \wedge P_2$ sont des formules ;
- Si P(x) est une formule alors $\exists$ x $\wedge$ (P(x)) et $\forall$ x $\wedge$ (P(x)) sont des formules.

Exemple : Soient les relations suivantes :

Etudiant (<u>NumCarte</u>, Nom, Prénom, Age, Adresse) représentée par E_1

Enseignant (<u>Matricule</u>, Nom, Prénom, Salaire, Grade, Age) représentée par E_2

Questions

1. Quels sont les étudiants qui habitent à Dakar ?
2. Quels sont les enseignants plus âgés que tous les étudiants ?
3. Quels sont les étudiants plus âgés qu'un de leurs enseignants ?
4. Matricule, Nom et Prénom des enseignants qui gagnent plus de 200000 par mois.
5. Numéro de carte, Age et Adresse des étudiants ayant le même prénom qu'un enseignant.
6. Numéro de carte, Nom et Prénom des étudiants les plus âgés.

Réponses

1. $\{<x_1, x_2, x_3, x_4, x_5> / E_1(<x_1, x_2, x_3, x_4, \text{'Dakar'}>)\}$
2. $\{<x_1, x_2, x_3, x_4, x_5, x_6> / E_2(<x_1, x_2, x_3, x_4, x_5, x_6>) \wedge \forall E_1(<y_1, y_2, y_3, y_4, y_5>) \wedge (x_6 > y_4)\}$
3. $\{<x_1, x_2, x_3, x_4, x_5> / E_1(<x_1, x_2, x_3, x_4, x_5>) \wedge \exists E_2(<z_1, z_2, z_3, z_4, z_5, z_6>) \wedge (x_4 > z_6)\}$
4. $\{<x_1, x_2, x_3> / E_2(<x_1, x_2, x_3, x_4, x_5, x_6>) \wedge (x_4 > 200000)\}$
5. $\{<x_1, x_4, x_5> / E_1(<x_1, x_2, x_3, x_4, x_5>) \wedge \exists E_2(<y_1, y_2, y_3, y_4, y_5, y_6>) \wedge (x_3 = y_3)\}$
6. $\{< x_1, x_2, x_3> / E_1(<x_1, x_2, x_3, x_4, x_5>) \wedge \forall\ E_1(<y_1, y_2, y_3, y_4, y_5>) \wedge (x_4 >= y_4)\}$

III. Calcul relationnel Vs Algèbre relationnelle

L'algèbre relationnelle et le calcul relationnel ont la même puissance d'expression ; donc toutes les requêtes qui peuvent être formulées en utilisant l'un peuvent aussi l'être en utilisant l'autre. Ce fut vérifié en premier par E.F. Codd en 1972. La preuve est fondée sur un algorithme (appelé « algorithme de réduction de Codd ») par lequel une expression arbitraire du calcul relationnel peut être réduite à une expression au sens équivalent de l'algèbre relationnelle. Certains déclarèrent que les langages basés sur le calcul relationnel sont de « haut niveau » ou « plus déclaratifs » que les langages basés sur l'algèbre relationnelle parce que ce dernier spécifie (partiellement) l'ordre des opérations tandis que le calcul laisse le compilateur ou l'interpréteur déterminer l'ordre d'évaluation le plus efficace.

IV. Exercice d'application

Soit le schéma de la base de données agence de location immobilière suivant :

Immeuble (Adresse, Nb_niveau, Annee) représenté par I

Appartement (Numero, #Immeuble, Nb_piece, Prix, Niveau) représenté par A

Locataire (Numero, Nom, Prenom, Age, Sexe, Profession) représenté par L

Louer (#Appartement, #Immeuble, #Locataire, Date, Duree) représenté par R

1. Écrivez les requêtes suivantes en calcul relationnel à variable domaine et à variables n-uplets ;
2. Donnez l'équivalent de chaque requête en algèbre relationnelle

a. Donner la liste des immeubles.
b. Donner la liste des appartements de l'immeuble situé au 250 Tilène.
c. Donner le nombre de pièces et le prix des appartements de l'immeuble situé au 10 Castor.
d. Quels sont les appartements les plus chers ?
e. Qui a occupé le plus longtemps l'appartement 5 d'un immeuble construit en 2018 ?