BDD3 : Algèbre relationnelle

Afin de manipuler un ensemble de relations, nous allons définir des opérateurs opérant sur les relations dans un cadre mathématique. Dans un second temps nous verrons comment, à l'aide de ces opérateurs, il est possible d'effectuer des requêtes très précises sur une base et étudierons un langage informatique de requêtes s'appuyant sur ces opérations.

1 Exemple

Nous travaillerons dans ce cours sur un exemple. Il s'agit d'une base de données intitulée "cinema" qui contient les tables : films, Cine1 et Cine2.

Films						
id_film	Nom	Realisateur	Pays	Annee	NoteCritique	NotePublic
1	"Star Wars"	Lucas	USA	1977	7.0	8.3
2	"Asterix : Mission Cleopatre"	Chabat	France	2012	4.1	8.0
3	"La mort aux trousses"	Hitchcock	USA	1959	9.0	8.1
4	"Les oiseaux"	Hitchcock	USA	1963	9.4	8.1
5	"La dolce vita"	Fellini	Italie	1960	8.8	8.8
6	"Bienvenu chez les chtis"	Boon	France	2008	4.2	9.1
7	"La cité de Dieu"	Meirelles	Bresil	2002	8.8	8.2
8	"Le loup de Wall Street"	Scorsese	USA	2013	8.2	7.6

Cine1	
Titre	**Seance**
"Asterix : Mission Cleopatre"	20 :00
"Star Wars"	20 :00
"La cité de Dieu"	20 :30
"La dolce vita"	20 :30
"Le loup de Wall Street"	21 :00
"Les oiseaux"	22 :00
"Bienvenu chez les chtis"	22 :00

Cine2	
Titre	**Seance**
"La mort aux trousses"	20 :00
"Star Wars"	20 :00
"La cité de Dieu"	20 :15
"La dolce vita"	20 :30
"Le loup de Wall Street"	20 :40
"Asterix : Mission Cleopatre"	22 :00
"Bienvenu chez les chtis"	22 :00
"Le loup de Wall Street"	22 :15

2 Opérateurs usuels sur les ensembles

2.1 Union

Il s'agit de connaître pour un même schéma relationnel donné, l'ensemble des valeurs dans un schéma ou l'autre.

Dans l'exemple de la base de données "cinema", il s'agira de connaître l'ensemble des séances existantes.

définition 1 (union) :
Soit $R_1(S)$ et $R_2(S)$ deux relations de même schéma S.
On appelle union de R_1 et de R_2 la relation de schéma S dont l'ensemble des valeurs est constitué des valeurs comprises dans R_1 ou dans R_2
Cette relation est notée $R_1 \cup R_2$

Exercice 1 : Que vaut #(cine1 $\cup$ cine2) ?

2.2 Intersection

Il s'agit de connaître pour un même schéma relationnel donné, l'ensemble des valeurs dans un schéma et l'autre.

Dans l'exemple de la base de données "cinema", il s'agira de connaître l'ensemble des séances ayant lieu à la même heure dans les deux cinémas.

définition 2 (intersection) :
Soit $R_1(S)$ et $R_2(S)$ deux relations de même schéma S.
On appelle intersection de R_1 et de R_2 la relation de schéma S dont l'ensemble des valeurs est constitué des valeurs comprises dans R_1 et dans R_2
Cette relation est notée $R_1 \cap R_2$

Exercice 2 : Détailler la relation cine1 $\cap$ cine2.

2.3 Différence

Il s'agit désormais de connaître les séances ayant lieu au cine1 mais non au cine2.

définition 3 (différence) :
Soit $R_1(S)$ et $R_2(S)$ deux relations de même schéma S.
On appelle différence de R_1 et de R_2 la relation de schéma S dont l'ensemble des valeurs est constitué des valeurs comprises dans R_1 mais pas dans R_2
Cette relation est notée R_1 - R_2

Exercice 1 : Que vaut #(cine1 - cine2) ?

3 Projection

Toutes les informations concernant les films ne sont pas forcément pertinentes pour tous les utilisateurs. Ainsi l'attribut id_film est interne à la table et ne sera pas affiché lors d'une consultation de cette table. De même, certains utilisateurs peuvent ne pas être intéressés par certaines données (comme le réalisateur, les notes des critiques etc...)

La projection est un opérateur qui permet à partir d'une relation R de créer une relation identique à R mais qui ne possède que certaines colonnes de R.

définition 4 (projection) :
Soit R(S) une relation de schéma S et X $\subset$ S.
On appelle projection de R selon X la relation :
$\Pi_X(R) = \{e(X)/e \in R\}$

Le schéma de $\Pi_X(R)$ est donc X.

Exercice 4 : On considère la relation Films et P={Nom, Annee}. Détailler Π_P(Films)

Une projection ne contient pas forcément autant de valeurs que la relation de départ. En effet, plusieurs valeurs peuvent être fusionnées.

Exercice 5 : Soit E={Pays}. Que vaut $\#\Pi_E$(Films) ?

4 Sélection

Un cinéphile n'est intéressé que par les films réalisés par Hitchcock. Il souhaite voir les informations de l'ensemble des films de ce réalisateur, et de lui seul. L'opérateur de sélection va répondre à cette problématique.

définition 5 (sélection simple) :
Soit R(S) une relation de schéma S ; A $\in$ S et a $\in$ dom(A)
On appelle $\sigma_{A=a}(R)$ la sélection de R selon A=a la relation e $\in$ R / e $\cdot$ A = a
Elle est donc obtenue en sélectionnant les éléments de R dont l'attribut A vaut a

Exercice 6 : Détailler $\sigma_{\text{Realisateur="Hitchcock"}}(Films)$

Si le domaine de l'attribut permet d'autres comparaisons que l'égalité on peut étendre la notion de sélection à ces comparaisons. Par exemple, $\sigma_{A \geqslant 1}(R)$ sélectionnera les valeurs e de R telle que e· A $\geqslant$ 1.

On sera souvent amené à effectuer des sélections plus complexes ; on peut par exemple vouloir les informations sur les films américains antérieurs à 2000, ou bien les informations sur les films dont la note du public est supérieure à 8.5 ou la note des critiques supérieure à 8.0

On est alors amené à faire une sélection composée : l'opérateur de sélection appliqué à une relation R produit une nouvelle relation définie à partir d'un sous-ensemble de valeurs de R. Les valeurs sélectionnées sont celles qui vérifient une propriété P impliquant les attributs de R.

P est une expression logique, dont la valeur est soit vraie, soit fausse pour chaque valeur de R.

On peut définir une expression logique formellement de la façon suivante :

définition 6 (expression logique) :
Soit a et b deux attributs ou constantes. Alors :
- a<b, a⩽b, a=b, a>b, a⩾b et a≠b sont des termes
- Si t_1 et t_2 sont des termes t_1 ET t_2 et t_1 OU t_2 sont des expressions logiques
- Si e_1 et e_2 sont des expressions logiques e_1 ET e_2 et e_1 OU e_2 sont des expressions logiques

On est alors en mesure de définir la sélection composée selon une expression logique :

définition 7 (sélection composée) :
Soit R(S) une relation de schéma S ; et E une expression logique définie à l'aide d'attributs de S
On appelle $\sigma_E(R)$ la sélection de R selon E la relation {e ∈ R / E est vraie pour e}

$\sigma_E(R)$ a la même schéma que R.

Exercice 7 : On considère P=(Annee < 2000 ET Pays = USA) OU (Pays = Bresil). Détailler σ_P(Films)

On considère P'=(Note Critique >= 8.0) OU (Note Public >= 9.0). Que vaut $\#\sigma_{P'}$(Films) ?

Exercice 8 : Écrire l'expression logique qui permet de sélectionner les films dont la note du public est plus grande ou égale que la note de la critique et tel que l'écart des deux notes soit inférieur ou égal à 1 point ?

Écrire l'expression logique qui permet de sélectionner les films pour lesquels l'écart en valeur absolue entre la note de la critique et celle du public est strictement inférieur à 1 ?

5 Renommage

Il est possible, souvent pour des raisons pratiques afin de lever une ambiguïté, de renommer un attribut d'une relation à l'aide d'un opérateur, dit de renommage.

La relation obtenue après avoir effectué cette opération est alors identique à la relation de départ, mis à part le schéma qui a été changé pour présenter le nouveau nom.

définition 8 (renommage) :
Soit S=(A_1,...,A_n) un schéma, i $\in [|1, n|]$ et B un attribut tel que dom(B)=dom(A_i)
On note $S_{A_i \leftarrow B}$ = (A_1,...,A_{i-1},B,A_{i+1},...,A_n) le schéma déduit de S en renommant A_i en B

6 Produit et division cartésiennes

définition 9 (produit cartésien) :
Soit R(S) et R'(S') deux relations de schémas disjoints, leur produit cartésien noté R × R' est :
{ (v_1,...,v_n,v'_1,...,v'_m) / (v_1,..., v_n) $\in$ R et (v'_1,...,v'_m) $\in$ R' }
Son schéma est S $\uplus$ S' = (A_1,...,A_n,B_1,...,B_m) où S =(A_1,...,A_n) et S' = (B_1,...,B_m)

R × R' contient donc l'ensemble des possibilités d'association entre une valeur de R et une valeur de R'. La notation S $\uplus$ S' rappelle qu'il ne s'agit pas seulement de prendre l'union des schémas mais aussi de s'assurer qu'ils soient disjoints. Notons qu'il est toujours possible de s'y ramener par l'intermédiaire d'un renommage des attributs présents dans les deux schémas.

Prenons par exemple les deux tables suivantes :

Viande
Type Viande
Steak
Poulet
Magret

Accompagnement
Type Accompagnement
Frites
Salade
Légumes

Exercice 9 : Détailler le produit cartésien des deux relations précédentes.

*Exercice 10 :*Une relation R comporte n enregistrements et une relation R' en comporte p. Combien d'enregistrements comporte R × R' ?

La division cartésienne se définit vis à vis du produit cartésien, de manière similaire à la façon dont la division euclidienne se définit vis à vis de la multiplication d'entiers, il s'agit à partir des ensembles A et B d'obtenir un ensemble C, où la combinaison de chaque valeur de B avec chaque valeur de C (produit cartésien de B et C) existe dans A. Plus formellement

définition 10 (division cartésienne) :
La relation R ÷ R' est la plus grande relation, vis-à-vis de l'inclusion, telle qu'il existe une relation R" vérifiant :
$[(R \div R') \times R'] \cup R'' = R$ (ou $[R' \times (R \div R')] \cup R'' = R$) et
$[(R \div R') \times R'] \cap R'' = \emptyset$ (ou $[R' \times (R \div R')] \cap R'' = \emptyset$)

Exercice 11 : On considère les relations suivantes :

Plat

Type Viande	**Type Accompagnement**
Steak	Frites
Steak	Salade
Poulet	Frites
Poulet	Légumes
Magret	Frites
Magret	Légumes

Viande

Type Viande
Steak
Poulet
Magret

Détailler Plat ÷ Viande

L'opérateur de division cartésienne existe pour des raisons théoriques de complétude de l'algèbre relationnelle, c'est-à-dire pour s'assurer qu'il soit possible d'exprimer toutes les requêtes possibles, mais n'est pas présent en pratique dans la plupart des langages de requête.

7 Jointure

On souhaite effectuer des requêtes portant sur plusieurs tables ayant un attribut commun, par exemple trouver l'ensemble des films américains dont la séance a lieu entre 20 heures et 21 heures. Le produit cartésien ne suffit pas à exprimer seul cette requête, en effet nous l'avons explicitement défini sur deux relations de schémas disjoints (même si le renommage suffit à s'y ramener). De plus la taille des tables obtenues par produit cartésien fait que cet opérateur est difficile à mettre en oeuvre en pratique.

Les opérateurs de jointures sont des opérateurs qui permettent de construire une relation à partir de deux relations qui ont (en général) une propriété commune. Il existe différents types de jointures, nous allons étudier la jointure symétrique ou jointure naturelle.

définition 11 (jointure naturelle) :
Soit R(S) et R'(S') deux relations de schémas respectifs S et S'.
La jointure (naturelle) de R et R' notée $R \bowtie R'$ est la relation de schéma $S \cup S'$ définie par :
$R \bowtie R' = \{ x \text{ sur } S \cup S' \;/\; \exists u \in R \text{ et } \exists v \in R',\ x_{/S}=u \text{ et } x_{/S'}=v \}$

Quand S=S' on a $R \bowtie R' = R \cap R'$
Quand S $\cap$ S' = $\emptyset$ on a $R \bowtie R' = R \times R'$

Exercice 12 : On renomme l'attribut Titre de la table cine1 en Nom.
Détailler alors la relation Films $\bowtie$ cine1

On conçoit une base de données constituée de plusieurs relations lorsqu'on veut isoler des informations concernant différentes entités (les films et les séances dans notre exemple) ; chaque relation est propre à une de ces entités et ne contient que les informations qui lui sont propres.

Si les valeurs d'une relation R_1 doivent faire référence aux valeurs d'une relation R_2, on prévoit dans R_2 un attribut A_2, si possible une clé, auquel la relation R_1 pourra faire référence par le biais d'un de ses propres attributs A_1 qui sera une clé étrangère.

La jointure symétrique permet de faire correspondre à chaque valeur de R_1 toutes les informations de la valeur qui lui correspond dans R_2, et donc par exemple d'effectuer une sélection selon ces informations.

Dans une même base de données, on peut avoir plus de deux relations, éventuellement reliées deux à deux par le biais de divers attributs. On peut alors écrire des requêtes élaborées au moyen de plusieurs jointures symétriques.

Exercice 13 : Expliciter les opérateurs de sélection qui :
- permettent de trouver l'ensemble des films américains dont la séance a lieu entre 20 heures et 21 heures :

- permettent de trouver l'ensemble des films dont la note des critiques est supérieure ou égale à 8.0 ; dont la note du public est supérieure ou égale à 8.2 et dont la séance commence à 21 heures ou plus tard :

8 Fonctions d'agrégation

Les opérateurs d'agrégation permettent d'effectuer des opérations sur certaines colonnes d'une relation. Les opérateurs d'agrégation que nous allons considérer sont :
- les fonction min et max qui retourne respectivement le minimum et le maximum des valeurs d'une colonne (contenant des valeurs numériques)

- la fonction somme qui permet de sommer les valeurs d'une colonne (contenant des valeurs numériques)
- la fonction moyenne qui permet de calculer la moyenne des valeurs d'une colonne (contenant des valeurs numériques)
- la fonction de comptage qui retourne le nombre de valeurs d'une colonne

Exercice 14 : Quels sont les résultats de ces cinq fonctions d'agrégation pour les valeurs de la colonne NoteCritique dans la table Films ?