

Systèmes de Gestion de Bases de Données - 2e

Chapitre 2 : Modèle relationnel

Daniel Schreurs

2 octobre 2021

Haute École de Province de Liège

Table des matières du chapitre i

1. Introduction
2. Relation, domaine et attribut
3. Clé primaire
4. Domaine primaire - clé étrangère
5. Intégrité de domaine
6. Intégrité d'entité ou de relation
7. Intégrité de référence

Table des matières du chapitre ii

8. Les opérateurs sémantiques

9. Les opérateurs ensemblistes

10. Les opérateurs ensemblistes

11. Les opérateurs relationnels

12. Les opérateurs additionnels

Introduction

Table des matières de la section : Introduction i

1. Introduction

1.1 Différents modèles

1.2 Le modèle relationnel

1.3 Base de données

2. Relation, domaine et attribut

3. Clé primaire

4. Domaine primaire - clé étrangère

5. Intégrité de domaine

Table des matières de la section : Introduction ii

6. Intégrité d'entité ou de relation

7. Intégrité de référence

8. Les opérateurs sémantiques

9. Les opérateurs ensemblistes

10. Les opérateurs ensemblistes

11. Les opérateurs relationnels

12. Les opérateurs additionnels

Introduction : Différents modèles

- Modèle réseau
- Modèle hiérarchique
- **Modèle relationnel**
- Modèle Objet - rel

Important

Nous n'étudierons, dans le cadre de ce cours, que le modèle relationnel.

En 1970, un mathématicien, [E.F. Codd](#), publie un article qui établit les bases du modèle relationnel. Ce modèle constitue un des apports les plus remarquables à la gestion de l'information que l'on peut résumer en 4 points :

- Rigueur des concepts de base
- Simplicité des concepts de base
- Puissance des opérateurs de manipulation
- Diminution des coûts de développement et de maintenance

Le modèle relationnel s'appuie sur une base formelle : la théorie des ensembles ou la théorie des prédicats.

- Un **ensemble** est une collection, un regroupement d'objets, de nombres, d'identités concrètes ou abstraites ;
- Les objets particuliers qui appartiennent à un ensemble sont appelés les **éléments** de cet ensemble ;
- On peut représenter graphiquement les ensembles et les opérations sur les ensembles par des **diagrammes de Venn**.

Le modèle relationnel : Rigueur des concepts de base

Important

Le modèle relationnel s'appuie sur une base formelle : la théorie des ensembles ou la théorie des prédicats¹.

- Des prédicats : P, Q
- Des connecteurs logiques : \wedge, \vee
- Des opérateurs : $+, -, \setminus$, etc..
- Des quantificateurs : \exists, \forall
- Des variables : x, y
- Des constantes : a, b
- Des fonctions : f, g

1. Un prédicat est une phrase qui peut comporter des paramètres et qui peut être vraie ou fausse

Un petit exemple :

$$\forall x, \exists y | amis(x, y) \wedge amis(x, mere(y))$$

Une relation est un ensemble, au sens mathématique, qui va être visualisé sous la forme d'une table. Il en résulte :

- Facilité d'apprentissage (par développeurs et utilisateurs)
- Plus grande communicabilité entre informaticiens et non-informaticiens
- Plus aucune référence à une méthode d'accès à un fichier ou organisation particulière des données sur les supports.

Le modèle relationnel : Puissance des opérateurs de manipulation

Les opérateurs relationnels sont des opérateurs ensemblistes.
L'application d'un opérateur à une ou plusieurs relations donne toujours une relation qui peut, à son tour, servir d'argument à un autre opérateur

Important

- notion de fermeture
- Non procédural : on dit ce que l'on veut obtenir, mais pas comment

Le modèle relationnel : Diminution des coûts de développement et de maintenance

Il a été estimé que le gain temps de développement d'une application de gestion varie entre 25 et 75% avec l'utilisation d'un SGBD relationnel.

Le modèle relationnel de Codd repose sur 3 piliers :

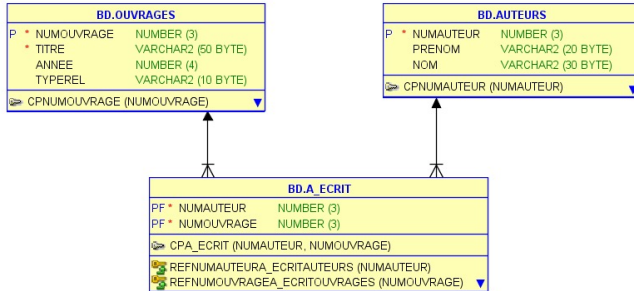
- Les **objets** : les éléments de base;
- Les **règles d'intégrité** : permettent de faire respecter le modèle des données;
- Les **opérateurs** : offrent la possibilité de manipuler la base de données;

- Relation Domaine/attribut
- Clé primaire
- Domaine primaire (clé étrangère)

- Contraintes d'intégrité de domaine
- Intégrité d'entité ou de relation
- Intégrité de référence

- Opérateurs sémantiques (liés aux domaines)
- Opérateurs ensemblistes : union, différence, produit cartésien
- Opérateurs relationnels : restriction (sélection), projection,
- Opérateurs additionnels : jointure, intersection, division

Base de données : Ouvrages



Base de données "Bandes Dessinées" utilisée dans les exemples

Base de données : Ouvrages

AUTEURS

Num Auteur	NOM	PRENOM
1	LELOUP	ROGER
2	CAUCIN	RAOUL
3	FRANKIN	ANDRE

AECRIT

NumAuteur	NumOuvrage
1	2
2	1
3	3
3	4

OUVRAGES

Num Ouvrage	TITRE	ANNEE	RELIURE
1	LE DAVID	1985	LUXE
2	LE TRIO DE L'ETRANGE	1980	NORMAL
3	LE CAS LAGAFFE	1980	NORMAL
4	LE GANG DES GAFFEURS	1981	LUXE

Base de données "Bandes Dessinées" utilisée dans les exemples

Relation, domaine et attribut

Table des matières de la section : Relation, domaine et attribut

1. Introduction

2. Relation, domaine et attribut

2.1 Relation

2.2 Domaine

3. Clé primaire

4. Domaine primaire - clé étrangère

5. Intégrité de domaine

Table des matières de la section : Relation, domaine et attribut

ii

6. Intégrité d'entité ou de relation

7. Intégrité de référence

8. Les opérateurs sémantiques

9. Les opérateurs ensemblistes

10. Les opérateurs ensemblistes

11. Les opérateurs relationnels

12. Les opérateurs additionnels

Relation, domaine et attribut : Relation

- La relation *AUTEURS* est représentée sous forme d'une table.
- Une ligne de la table constitue un élément de la relation.
- La relation ou table possède 3 colonnes ou attributs
- Chaque attribut a une valeur qui fait partie d'un ensemble de valeurs permises (domaine de l'attribut).
- Exemples :
 - NumAuteur : domaine = l'ensemble des entiers positifs ($D1$)
 - Nom : domaine = l'ensemble des chaînes de caractères de longueur 30 ($D2$)
- Une ligne quelconque de la table *AUTEURS* est constituée de $(V1, V2, V3) \mid V1 \in D1 \wedge V2 \in D2 \wedge V3 \in D3$
- La table *AUTEURS* est un sous-ensemble de toutes les combinaisons possibles, donc un sous-ensemble du produit cartésien $D1 * D2 * D3$

Important

Une RELATION R est un sous-ensemble du produit cartésien de n ensembles D_i appelés DOMAINES.

Une relation est donc un ensemble d'éléments de la forme

$$(v_1, v_2, v_3, \dots, v_n) | 1 \leq i \leq n, v_i \in D_i$$

que l'on appelle n -uplet ou tuple ou encore ligne. n est appelé le degré de la relation.

Une relation étant un ensemble, elle peut être définie de manière :

- Extensive : en donnant la liste de tous les tuples la composant :
 $AEcrit = \{(1, 2), (2, 1), (3, 3), (3, 4)\}$
- Intensive : en donnant le prédicat d'appartenance d'un tuple à
 $R : AEcrit = \{(x, y) | \dots\}$

Important

Un domaine représente l'ensemble des valeurs admissibles pour une composante d'une relation.

Syntaxe \neq Sémantique

Les relations sont définies à partir de domaines. Pour définir une base de données relationnelle, on commence par définir les domaines.

- Domaine NumeroAuteur = entier compris entre 1 et 100
- Domaine NomAuteur = chaîne de caractères
- Domaine PrenomAuteur = chaîne de caractères
- Domaine NumeroOuvrage = entier compris entre 1 et 500
- Domaine TitreOuvrage = chaîne de caractères
- Domaine AnneeEdition = entier ≥ 1900
- Domaine TypeReliure =
{*'NORMAL', 'LUXE', 'CARTONNE', 'BROCHE'*}

Définition des relations sur base des domaines : Relation AUTEURS composée des attributs :

- NumAuteur : défini sur NumeroAuteur,
- Nom : défini sur NomAuteur,
- Prenom : défini sur PrenomAuteur;

Relation OUVRAGES composée des attributs :

- NumOuvrage : défini sur NumeroOuvrage,
- Titre : défini sur TitreOuvrage,
- Année : défini sur AnneeEdition,
- Reliure : défini sur TypeReliure

...

Deux domaines sont déclarés compatibles s'ils sont sémantiquement comparables, c'est-à-dire si les ensembles qui les définissent ne sont pas disjoints.

- En particulier, deux domaines identiques ou liés par inclusion sont compatibles. Exemple : les domaines VilleEurope et VilleBelge, tous deux de type chaîne de caractères sont liés par inclusion et donc compatibles².
- Exemple de domaines non compatibles : NumeroAuteur et NumeroOuvrage sont incompatibles même s'ils sont définis au moyen de types de données comparables (des nb entiers)

2. Puisque toutes les villes de la Belgique sont aussi de villes de l'Europe.

- Les noms des domaines et des attributs correspondants ne sont pas les mêmes.
- Il est possible d'avoir dans une même table deux attributs différents issus du même domaine
- Par contre, les attributs d'une relation doivent tous être différents.

Clé primaire

1. Introduction

2. Relation, domaine et attribut

3. Clé primaire

3.1 Objectifs

3.2 Définition

3.3 Exercice

4. Domaine primaire - clé étrangère

5. Intégrité de domaine

Table des matières de la section : Clé primaire ii

6. Intégrité d'entité ou de relation

7. Intégrité de référence

8. Les opérateurs sémantiques

9. Les opérateurs ensemblistes

10. Les opérateurs ensemblistes

11. Les opérateurs relationnels

12. Les opérateurs additionnels

- Une relation est un ensemble. Il doit être possible de distinguer tous les éléments (tuples) de la relation.
- Un attribut ou un groupe d'attributs va jouer le rôle d'identifiant de la relation : c'est la **clé primaire**.
- Une valeur de clé primaire permet d'identifier de manière unique un tuple d'une relation.

Clé primaire : Définition

Important

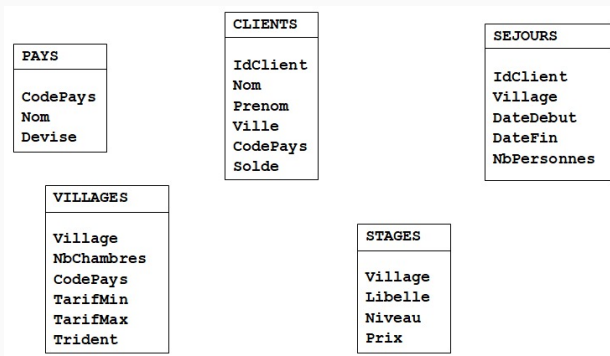
Une clé primaire est un ensemble d'attributs, K, vérifiant la double propriété :

- Unicité : les valeurs de clés primaires sont uniques et non nulles³;
- Minimalité : aucun attribut composant K ne peut être enlevé sans perdre la propriété d'unicité.

3. *NULL* ne veut pas dire "" ou 0 mais *absence de valeur*.

Clé primaire : Exercice

Spécifier les clés primaires dans le schéma suivant :



Structure d'une DB pour des séjours

Domaine primaire - clé étrangère

Domaine primaire - clé étrangère : Définition

Important

Un domaine primaire est un domaine sur lequel une clé primaire est définie.

Exemple : NumeroAuteur et NumeroOuvrage sont des domaines primaires.

Domaine primaire - clé étrangère : Définition

Important

Un attribut qui n'est pas clé primaire, mais qui est défini sur un domaine primaire est appelé une clé étrangère⁴.

Exemple : NumAuteur et NumOuvrage, dans la relation AEcrit, sont des clés étrangères.

4. La notion de clé étrangère permet d'exprimer les associations entre entités.

Intégrité de domaine

Il existe deux grandes classes de contraintes d'intégrité :

- Les **contraintes structurelles** dépendant du modèle de données (intégrité de domaine, d'entité ou de relation et de référence);
- Les **contraintes applicatives** liées à l'univers réel modélisé.

Important

L'intégrité de domaine porte sur le contrôle syntaxique et sémantique des valeurs présentes dans un attribut : seules les valeurs appartenant au domaine de l'attribut sont autorisées⁵.

5. Ce type de vérification se fait lors du chargement initial de la base de données comme pendant toute manipulation de celle-ci.

Intégrité d'entité ou de relation

Intégrité d'entité ou de relation : Définition

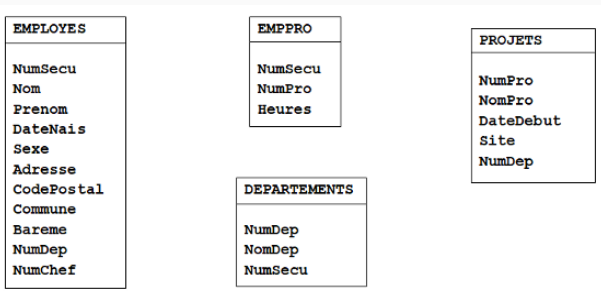
Important

L'intégrité d'entité ou de relation concerne les valeurs de la clé primaire d'une relation qui doivent être uniques et toujours définies (non nulles).

Intégrité de référence

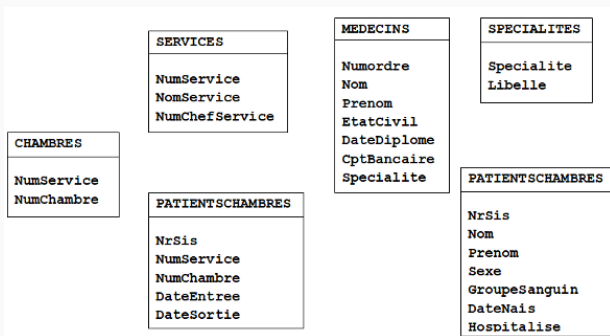
Intégrité de référence : Exercice

Spécifier les clés primaires et les clés étrangères dans le schéma suivant :



Intégrité de référence : Exercice

Spécifier les clés primaires et les clés étrangères dans le schéma suivant :



Les opérateurs sémantiques

Important

Les opérateurs sémantiques permettent la création et la manipulation des domaines⁶.

Remarque importante : la notion de domaine n'existe pas en Oracle : on pourra préciser le type des valeurs permises pour les attributs et spécifier des contraintes de domaine pour restreindre les valeurs permises et ainsi correspondre à la réalité.⁷

6. Nous ne les étudierons pas plus avant ici.

7. Donc en Oracle, pas d'objet domaine, mais on pourra définir des contraintes de domaine pour limiter les valeurs permises dans les différents champs des tables.

Les opérateurs ensemblistes

Les opérateurs ensemblistes : Définition

Important

L'algèbre relationnelle ou langage algébrique inventé par Codd est considéré comme une collection d'opérateurs (ensemblistes et relationnels) portant sur des relations.

Il est caractérisé par les propriétés suivantes :

- Fermeture
- Ensembliste
- Non procédural
- Universel
- Indépendance

L'application d'un opérateur relationnel à une ou des relations génère TOUJOURS une relation qui peut à son tour être utilisée comme argument de nouveaux opérateurs.

Le résultat d'une requête est toujours un sous-ensemble d'une ou plusieurs relations.

L'utilisateur spécifie quoi (le résultat qui l'intéresse), le système détermine comment.

L'étude de l'algèbre relationnelle constitue un tremplin pour l'étude des langages supportés par n'importe quel SGBD relationnel.

Les opérateurs sont basés sur des valeurs d'attributs : seul moyen d'accès. Les accès multi-relations sont effectués par des comparaisons entre valeurs d'attributs ce qui permet de très grandes potentialités d'accès totalement indépendantes de l'implémentation.

Les opérateurs ensemblistes

Les opérateurs ensemblistes de base sont binaires : à partir de deux relations, ils en génèrent une troisième. Les opérateurs que nous allons étudier :

- Union
- Différence
- Intersection (qui peut aussi être définie à partir de la différence)
- Produit cartésien

L'union, la différence et l'intersection ne s'appliquent qu'à des relations "union-compatibles". Deux relations sont union-compatibles si :

- Elles ont le même nombre d'attributs (le même degré);
- Les attributs associés deux à deux sont définis sur des domaines compatibles.

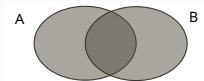
Les opérateurs ensemblistes : L'union

$$X = R_1 \cup R_2$$

L'union de deux relations R_1 et R_2 union-compatibles est une relation X contenant l'ensemble des tuples appartenant à R_1 ou à R_2 ou aux deux relations.

- Elles ont le même nombre d'attributs (le même degré);
- Les attributs associés deux à deux sont définis sur des domaines compatibles.

$$A \cup B = \{x | x \in A \vee x \in B\}$$



L'union de l'ensemble A et B

Exemple

- $A = \{2, 3, 4, 5\}$
- $B = \{1, 3, 6, 8\}$
- $C = \{2, 5, 10\}$

8

$$A \cup B = \{1, 2, 3, 4, 5, 6, 8\}$$

$$B \cup C = \{2, 3, 4, 5, 10\}$$

(1)

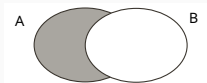
8. Attention, pas de doublons dans les ensembles, contrairement à Oracle.

Les opérateurs ensemblistes : Différence

$$X = R_1 - R_2$$

La différence de deux relations R_1 et R_2 union-compatibles (dans l'ordre $R_1 - R_2$) est une relation X contenant les tuples appartenant à R_1 et n'appartenant pas à R_2 .

$$A \setminus B = \{x | x \in A \wedge x \notin B\}$$



La différence de l'ensemble A et B

Exemple

- $A = \{2, 3, 4, 5\}$
- $B = \{1, 3, 6, 8\}$
- $C = \{2, 5, 10\}$

9

$$\begin{aligned} A \setminus B &= \{2, 4, 5\} \\ A \setminus C &= \{1, 3, 6, 8\} \end{aligned} \quad (2)$$

9. $B \setminus C = B$ puisqu'aucune donnée commune entre B et C.

$$X = R_1 \cap R_2$$

L'intersection pouvant s'exprimer en fonction de la différence, cet opérateur n'est pas indispensable¹⁰.

10. Nous reviendrons sur l'intersection de deux relations dans le paragraphe des opérateurs additionnels.

$$X = R_1 \times R_2$$

Le produit cartésien de deux relations R_1 et R_2 (de schéma quelconque) est une relation ayant pour attributs tous les attributs de R_1 et de R_2 et dont les tuples sont constitués de toutes les concaténations possibles d'un tuple de R_1 à un tuple de R_2 .

$$A \times B = \{(x, y) | x \in A \wedge y \in B\}$$

11

-
11. Attention, si une relation compte 0 tuple et l'autre x tuples, le produit cartésien comportera 0 tuple!

Exemple

- $A = \{2, 3\}$
- $B = \{1, 3, 6\}$

$$A \times B = \{(2, 1), (2, 3), (2, 6), (3, 1), (3, 3), (3, 6)\}^{12} \quad (3)$$

12. Attention, toujours pas de doublons dans les ensembles, contrairement à Oracle.

Les opérateurs relationnels

Les deux opérateurs unaires sélection et projection combinés avec les opérations ensemblistes union, différence et produit cartésien étudiés au paragraphe précédent permettent de définir toutes les expressions correctes de l'algèbre relationnelle.

$$X = Projection(R/C_1, C_2, \dots, C_p)$$

La projection d'une relation R de schéma $R(C_1, C_2, \dots, C_n)$ sur les attributs $C_{i_1}, C_{i_2}, \dots, C_{i_p}$ (avec $i, j \neq i, k$ et $p \leq n$) est une relation R de schéma $R(C_{i_1}, C_{i_2}, \dots, C_{i_p})$ dont les tuples sont obtenus par élimination des valeurs des attributs de R n'appartenant pas à R et par suppression des tuples en double.

L'opérateur de projection permet donc d'extraire certains attributs d'une relation. On parle de sélection verticale.

Les opérateurs relationnels : La projection

C1	C2	C3	C4	C5
1	X	1	A	B
1	Y	1	A	D
2	Z	4	B	E
3	Z	5	C	A

C1	C3	C4
1	1	A
1	1	A
2	4	B
3	5	C

C1	C3	C4
1	1	A
1	1	A
2	4	B
3	5	C

Pas de doublon !!

Résultat final :

C1	C3	C4
1	1	A
2	4	B
3	5	C

Exemple : PROJECTION (R/C1,C3,C4)

Les opérateurs relationnels : La sélection

$$X = \text{SELECTION}(R/\text{prédicat})$$

L'opération de sélection, selon un critère C , appliquée à une relation R donne une relation R de même schéma dont les tuples sont ceux de R satisfaisant le critère C .

Critère de sélection : prédicat ou expression logique de prédicats. Chaque prédicat exprime une comparaison entre une colonne d'une table et une constante au moyen d'un opérateur de comparaison.

$$\begin{aligned} &\text{SELECTION}(\text{livre}/\text{Annee} = 2010 \wedge \\ &\quad \text{NumOuvrage} > 3 \text{ ET } \text{Annee} \geq 2005 \wedge \\ &\quad (\text{NumOuvrage} < 2 \text{ ET } \text{Annee} = 2006) \vee \text{Annee} = 2002) \end{aligned} \quad (4)$$

Les opérateurs relationnels : La sélection

Résultat final :				
C1	C2	C3	C4	C5
1	X	1	A	B
1	Y	1	A	D
2	Z	4	B	C
3	Z	5	C	A

Exemple : SELECTION (R/C4=A)

Les opérateurs additionnels

Les cinq opérateurs vus jusqu'à présent sont suffisants pour exprimer une requête quelconque de l'algèbre relationnelle. Cependant, certaines requêtes, même banales, sont très longues à exprimer.

- Intersection
- Jointure
- Jointure externe
- Division

Les opérateurs additionnels : Intersection

$$X = R_1 \cap R_2$$

L'intersection de deux relations R_1 et R_2 union-compatibles est une relation X contenant les tuples appartenant à R_1 et à R_2 .

$$A \cap B = \{x | x \in A \wedge x \in B\}$$

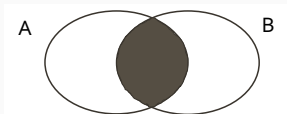


FIGURE 1 – L'intersection

Les opérateurs additionnels : Intersection

$$X = R_1 \cap R_2$$

On peut aussi définir l'intersection au moyen de la différence :

$$x = R_1 \setminus (R_1 \setminus R_2) = R_2 \setminus (R_2 \setminus R_1)$$

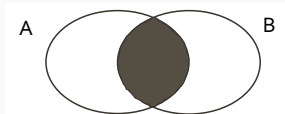


FIGURE 2 – L'intersection

Exemple

- $A = \{2, 3, 4, 5\}$
- $B = \{1, 3, 6, 8\}$
- $C = \{2, 5, 10\}$

$$A \cap B = \{3\}$$

$$A \cap C = \emptyset$$

(5)

Exemple

- $A = \{2, 3, 4, 5\}$
- $B = \{1, 3, 6, 8\}$
- $C = \{2, 5, 10\}$

$$A \cap B = A - (A - B) = B - (B - A)$$

$$A - B = \{2, 4, 5\} \quad A - (A - B) = \{3\} \quad (6)$$

$$B - A = \{1, 6, 8\} \quad B - (B - A) = \{3\}$$

$$X = \text{jointure}(R_1, R_2 / C)$$

La jointure de deux relations R_1 et R_2 selon un critère généralisé C est l'ensemble des tuples du produit cartésien $R_1 \times R_2$ satisfaisant le critère C .

Le critère C est une expression logique de prédicats dans laquelle chaque prédicat exprime une comparaison entre un attribut et une constante ou un attribut et un autre attribut. Les attributs comparés doivent impérativement être définis sur des domaines compatibles!

$$\begin{aligned} X &= \text{jointure}(R_1, R_2/C) \\ x &= \text{jointure}(\text{AUTEURS}, \text{A_ECRIT} / \\ &\quad \text{AUTEURS.NumAuteur} = \text{A_ECRIT.NumAuteur}) \end{aligned} \quad (7)$$

13

-
13. Quelles sont les colonnes de l'ensemble X ? Toutes les colonnes de *Auteurs* et *A_Ecrit*, sans répétition de la colonne *NumAuteurs*

Plusieurs cas particuliers de jointures :

- L'équi-jointure de R_1 et R_2 sur les attributs C_{R1} et C_{R2} est la jointure selon le critère $C_{R1} = C_{R2}([INNER]JOIN)$ ¹⁴
- L'auto-jointure de R selon C_i est la jointure de R avec elle-même selon le critère $C = C$
- La jointure naturelle de R_1 et R_2 est l'équijointure de R_1 et R_2 sur tous les attributs ayant le même nom dans R_1 et R_2 , suivie d'une projection qui permet de conserver un seul de ces attributs égaux de même nom. (NATURAL JOIN)

14. Le mot-clé INNER JOIN sélectionne les enregistrements qui ont des valeurs correspondantes dans les deux tables.

$$X = JOINEXT(R_1, R_2/C)$$

La jointure externe de deux relations R1 et R2 est obtenue en deux étapes :

1. On effectue une jointure de R1 et R2
2. On ajoute à la relation obtenue en (1) les tuples de R1 et R2 qui ne participent pas à la jointure complétés avec des valeurs nulles pour les attributs de l'autre relation

$$(\{LEFT|RIGHT|FULL\}OUTERJOIN)$$

$$\begin{aligned} X &= JOINEXT(R_1, R_2/C) \\ x &= JOINEXTLEFT(AUTEURS, A_ECRIT / \\ &\quad AUTEURS.NumAuteur = A_ECRIT.NumAuteur) \end{aligned} \quad (8)$$

15

-
15. Liste de tous les auteurs et pour ceux qui ont écrit un ou plusieurs ouvrages, un enregistrement pour chaque ouvrage avec l'identifiant de celui-ci en plus des informations concernant l'auteur.

$$X = R_1 \div R_2$$

Le quotient de la relation R_1 de schéma $R_1(C_1, C_2, \dots, C_p, C_{p+1}, \dots, C_n)$ par la sous-relation R_2 de schéma $R_2(C_{p+1}, \dots, C_n)$ est la relation D de schéma $D(C_1, C_2, C_p)$ formée de tous les tuples qui concaténés à chacun des tuples de R_2 donne toujours un tuple de R_1 .

$PROJECTION_2(\$
 $JOIN_1(Dept, Emp$
 $/Dept.NrDept = Emp.NrDept)$
 $/Emp.nom, Dept.nom)$ (9)

- Donner le résultat de cette expression.
- Quelle est la question à poser pour obtenir ce résultat ?