HAI725I : Modèle et algèbre relationnels

I.Mougenot

UM Faculté des Sciences Département Informatique

2021

Différents modèles

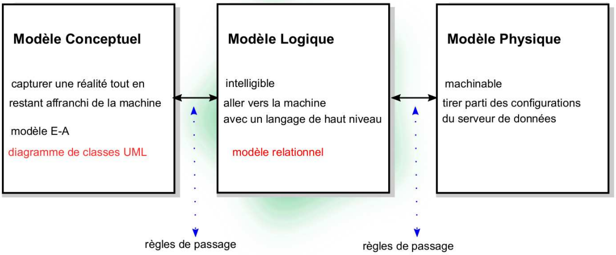


Figure: Une approche s'appuyant sur trois types de modélisation

Notion de relation

Edgar Codd (1970) : rupture avec la manière d'envisager l'organisation des données

- simplicité de la représentation : un seul concept de relation
- adossement à la théorie des ensembles
- mis en œuvre par les très nombreux SGBD relationnels (voir `https://db-engines.com/en/ranking`)
- efficacité de la représentation qui va couvrir 80% des besoins en matière de bases de données

Grands axes abordés autour du relationnel

Objectif : maîtrise de la définition et de la manipulation de tels modèles

- dériver un modèle relationnel à partir d'un modèle conceptuel
- formalisation / normalisation
- algèbre relationnelle
- langage standard SQL (Structured Query Language)

Notion de relation

Définition Relation

Sous-ensemble du produit cartésien d'une liste de domaines

Définition Domaine

Ensemble de valeurs

Domaine D1 = {'homme','femme'}

Domaine D2 = {'Montpellier','Lunel', 'Orange', 'Marseille'}

Domaine D3 = {10,11,12,13,14,15,16,20,22,24,26,28,30}

Notion de relation

Produit cartésien

D1 x D2 x D3

Une relation possible :

'femme'	'Lunel'	26
'homme'	'Orange'	22
'femme'	'Montpellier'	24

une ligne de la relation = un tuple ou n-uplet

Notion de tuple

Un tuple ou n-uplet

$\{v_1 \ , v_2, \ldots , v_n \ \}$

avec $v_1 \in$ D1, $v_2 \in$ D2, ..., $v_n \in$ Dn

Un exemple parmi les tuples présentés

{'femme','Montpellier',24}

Notion d'attribut

Attribut

couple (nom,domaine) - désignation de la propriété et ensemble de valeurs pouvant être prises par cette propriété

Attribut

genre : nom=genre et domaine = {'homme','femme'}

Attribut

ville : nom=ville et domaine = {'Montpellier','Lunel', 'Orange', 'Marseille'}

UNIVERSITÉ DE MONTPELLIER

Arité et cardinalité d'une relation

Arité ou encore degré

nombre d'attributs de la relation (amené à n'évoluer que rarement : schéma "figé")

Cardinalité

nombre de tuples de la relation (amené à évoluer fréquemment)

2 écritures de la relation

en intension (compréhension) et en extension (vision tabulaire comprenant les tuples)

Relation en intension

Schéma de la relation

nom de la relation + liste des attributs avec éventuellement leurs types de données et les contraintes qui peuvent s'y appliquer

Exemple de la relation Etudiant en intension

Etudiant(numINE, nom, prenom, genre, âge)

Exemple de la relation Etudiant avec types de données

Etudiant(numINE varchar(12), nom varchar(12), prenom varchar(12), genre varchar(8), âge integer)

Notion intuitive de clé de la relation

Une relation n'admet pas de doublons, les tuples se doivent d'être uniques. Un attribut ou une combinaison d'attributs vont jouer le rôle de clé primaire de la relation. Cet attribut ou cet ensemble d'attributs va garantir des valeurs uniques (différentes) et non nulles (toujours renseignées) pour chacun des tuples. Pour la relation, l'attribut numINE va permettre de garantir cette unicité de la valeur toujours renseignée

Exemple de la relation Etudiant avec la clé soulignée

Etudiant(<u>numINE</u>, nom, prenom, genre, âge)

Relation en extension

La relation est présentée dans sa vision tabulaire avec les tuples qui la composent. On parlera aussi d'instance de la relation

Exemple de la relation Etudiant en extension

numINE	nom	prénom	genre	âge
'2016564'	'Dusol'	'Marie'	'femme'	22
'2014564'	'Dusol'	'Paul'	'homme'	27
'2020564'	'Bony'	'Paul'	'homme'	20
'2020568'	'Balard'	'Zoé'	'femme'	20

Relation en extension

Toutes les permutations colonne/ligne sont possibles. Il s'agit toujours de la même instance de relation

Exemple de la relation Etudiant en extension

numINE	nom	prénom	âge	genre
'2016564'	'Dusol'	'Marie'	22	'femme'
'2020568'	'Balard'	'Zoé'	20	'femme'
'2014564'	'Dusol'	'Paul'	27	'homme'
'2020564'	'Bony'	'Paul'	20	'homme'

UNIVERSITÉ DE MONTPELLIER

Schéma de la base de données relationnelle

Ensemble des schémas relationnels

Un exemple réduit

Etudiant(<u>numINE</u>, nom, prenom, genre, âge)
Formation(<u>codeF</u>, libellé, départementEnseignement)
Inscrit_dans(<u>numINE,codeF,année</u>)

Algèbre relationnelle

La relation est abordée en terme de structure à laquelle on applique des opérations pour retourner d'autres relations

Opérateurs ensemblistes

Produit cartésien (X), Union ($\cup$), Intersection ($\cap$), Différence (-), Division ($\div$)

Opérateurs spécifiques

Sélection (Σ), Projection (Π), Jointure ($\bowtie$)

Le résultat d'une opération algébrique est toujours une relation

Notion de langage complet

Cinq opérateurs qui permettent d'obtenir les autres opérateurs

Produit cartésien (X), Union ($\cup$), Différence (-), Sélection (Σ), Projection (Π)

Rôles dévolus à l'algèbre relationnelle

- Spécifier les opérations à agencer pour arriver à produire le résultat d'une requête
- Socle de requêtage du schéma qui sous-tend le langage SQL
- utile pour l'optimisation de requêtes

L'union

notée $\cup$; opérateur binaire

R3 $=$ R1 $\cup$ R2 : opération qui s'applique à deux relations opérandes **de mêmes schémas**, notées ici R1 et R2, pour restituer une relation de même schéma R3 qui contient à la fois les tuples de R1 et les tuples de R2 (les doublons ne sont notés qu'une fois)

Opération commutative : R1 $\cup$ R2 $\equiv$ R2 $\cup$ R1
Opération associative : R1 $\cup$ (R2 $\cup$ R3) $\equiv$ R2 $\cup$ (R1 $\cup$ R3)

Visions ensembliste et arborescente

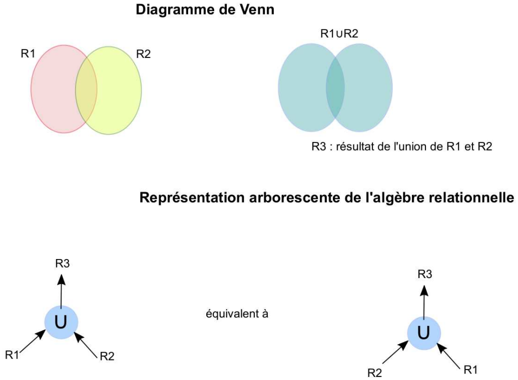


Figure: Union R3 s'appliquant à deux relations R1 et R2

Un exemple concret : Personne = Etudiant $\cup$ Enseignant

Etudiant(nom,prénom) et Enseignant(nom,prenom)

Etudiant en extension

nom	prénom
'Dusol'	'Marie'
'Dusol'	'Paul'
'Bony'	'Paul'
'Balard'	'Zoé'

Enseignant en extension

nom	prénom
'Dubois'	'Alice'
'Drapier'	'Paul'
'Balard'	'Zoé'

Personne (nom,prenom) en extension

nom	prénom
'Dubois'	'Alice'
'Drapier'	'Paul'
'Balard'	'Zoé'
'Dusol'	'Marie'
'Dusol'	'Paul'
'Bony'	'Paul'

La différence

notée - ; opérateur binaire

R3 = R1 - R2 : opération qui s'applique à deux relations opérandes **de mêmes schémas**, notées ici R1 et R2, pour restituer une relation de même schéma R3 qui contient les tuples de R1 privés des tuples qui appartiennent aussi à R2 (les tuples que R1 partage avec R2 sont enlevés du résultat)

Attention : R1 - R2 $\neq$ R2 - R1
La différence est une opération non commutative et non associative

Visions ensembliste et arborescente

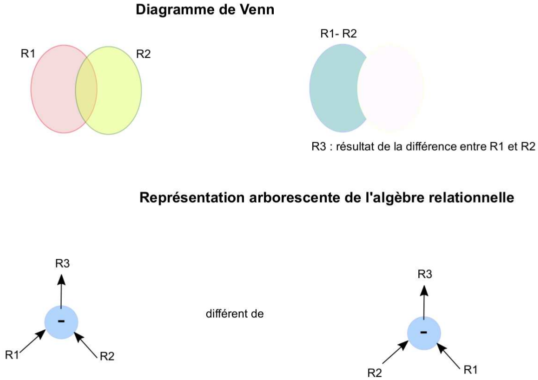


Figure: Différence R3 s'appliquant à deux relations R1 et R2

Un exemple concret : Etudiant - Enseignant

Sémantique : étudiants qui ne sont pas aussi enseignants

Etudiant en extension

nom	prénom
'Dusol'	'Marie'
'Dusol'	'Paul'
'Bony'	'Paul'
'Balard'	'Zoé'

Enseignant en extension

nom	prénom
'Dubois'	'Alice'
'Drapier'	'Paul'
'Balard'	'Zoé'

Etudiant - Enseignant en extension

nom	prénom
'Dusol'	'Marie'
'Dusol'	'Paul'
'Bony'	'Paul'

L'intersection

notée $\cap$; opérateur binaire

$R3 = R1 \cap R2$: opération qui s'applique à deux relations opérandes **de mêmes schémas**, notées ici R1 et R2, pour restituer une relation de même schéma R3 qui contient les tuples que R1 partage avec R2

Opération commutative : $R1 \cap R2 \equiv R2 \cap R1$
Opération associative : $R1 \cap (R2 \cap R3) \equiv R2 \cap (R1 \cap R3)$

Visions ensembliste et arborescente

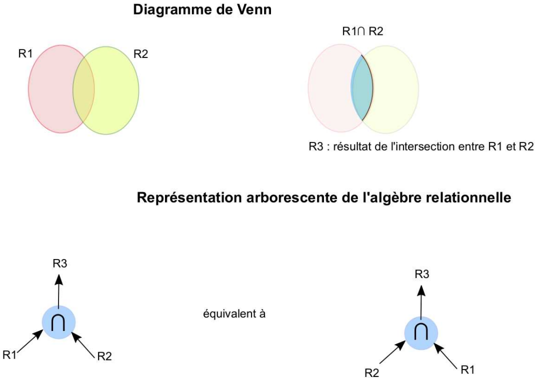


Figure: Intersection R3 s'appliquant à deux relations R1 et R2

Un exemple concret : Etudiant ∩ Enseignant

Sémantique : étudiants qui sont aussi enseignants

Etudiant en extension

nom	prénom
'Dusol'	'Marie'
'Dusol'	'Paul'
'Bony'	'Paul'
'Balard'	'Zoé'

Enseignant en extension

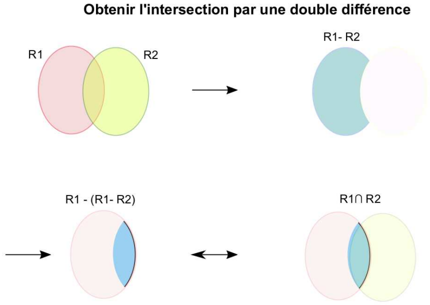
nom	prénom
'Dubois'	'Alice'
'Drapier'	'Paul'
'Balard'	'Zoé'

Etudiant ∩ Enseignant en extension

nom	prénom
'Balard'	'Zoé'

Intersection & différence

L'intersection se traduit par une double différence : $R1 \cap R2 \equiv R1 - (R1 - R2)$ ou bien $\equiv R2 - (R2 - R1)$



Le produit cartésien

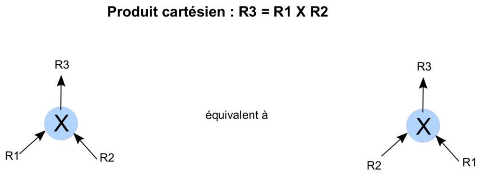
notée X ; opérateur binaire

R3 = R1 X R2 : opération qui s'applique à deux relations opérandes, notées ici R1 et R2, pour restituer une relation de schéma R3 juxtaposant les schémas de R1 et de R2, et qui combine les tuples des deux relations

Opération commutative : R1 X R2 ≡ R2 X R1
Opération associative : R1 X (R2 X R3) ≡ R2 X (R1 X R3)

Représentation visuelle

Multiplication ensembliste



Un exemple concret : Etudiant X Inscrit_Dans

Sémantique : pas forcément de sémantique attachée

Etudiant en extension

numINE	nom	prénom
'20164545'	'Dusol'	'Marie'
'20165546'	'Bony'	'Paul'
'20174533'	'Balard'	'Zoé'

Inscrit_Dans en extension

numEtudiant	codeModule
'20164545'	'HMIN112M'
'20164545'	'HMIN110M'
'202095995'	'HMIN110M'

Produit cartésien de schéma (numINE, nom, prénom, numEtudiant, codeModule) en extension : 3 X 3 = 9 tuples

numINE	nom	prénom	numEtudiant	codeModule
'20164545'	'Dusol'	'Marie'	'20164545'	'HMIN112M'
'20165546'	'Bony'	'Paul'	'20164545'	'HMIN112M'
'20174533'	'Balard'	'Zoé'	'20164545'	'HMIN112M'
'20164545'	'Dusol'	'Marie'	'20164545'	'HMIN110M'
'20165546'	'Bony'	'Paul'	'20164545'	'HMIN110M'
'20174533'	'Balard'	'Zoé'	'20164545'	'HMIN110M'
'20164545'	'Dusol'	'Marie'	'202095995'	'HMIN110M'
'20165546'	'Bony'	'Paul'	'202095995'	'HMIN110M'
'20174533'	'Balard'	'Zoé'	'202095995'	'HMIN110M'

La sélection

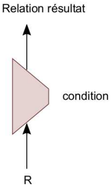
notée σ ; opérateur unaire

$R1 = \sigma_{condition}(R)$: opération qui s'applique à une relation opérande, notée ici R, pour restituer une relation de même schéma, qui retourne les tuples qui satisfont la condition posée en indice.

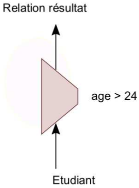
La condition prend la forme générale :
attribut opérateur (arithmétique) valeur
avec pour opérateurs : $<$, $>$, $<>$ ou $!=$, $=$, $>=$, $<=$

Représentation visuelle

La sélection est un filtre horizontal sur la relation



Etudiant (nom, prenom, age)



Un exemple concret autour de la sélection

Sémantique : Etudiants de plus de 24 ans :
Résultat $= \sigma_{age>24}(\text{Etudiant})$

Etudiant en extension

nom	prénom	age
'Dusol'	'Marie'	22
'Bony'	'Paul'	25
'Balard'	'Zoé'	29

Résultat

nom	prénom	age
'Bony'	'Paul'	25
'Balard'	'Zoé'	29

UNIVERSITÉ DE MONTPELLIER

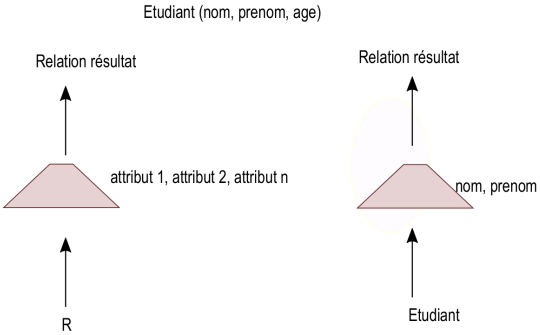
La projection

notée Π ; opérateur unaire

$R1 = \Pi_{(listeattributs)}(R)$: opération qui s'applique à une relation opérande, notée ici R, pour restituer une relation de sous-schéma de R, qui retourne les tuples de taille réduite qui ne consistent qu'aux valeurs des attributs mentionnés.

Représentation visuelle

La projection est un filtre vertical sur la relation



Un exemple concret autour de la projection

Sémantique : nom et prénom des étudiants :

Résultat $= \Pi_{nom,prenom}(Etudiant)$

Etudiant en extension

nom	prénom	age
'Dusol'	'Marie'	22
'Bony'	'Paul'	25
'Balard'	'Zoé'	29

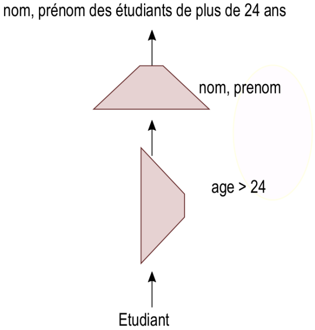
Résultat

nom	prénom
'Dusol'	'Marie'
'Bony'	'Paul'
'Balard'	'Zoé'

Concerter l'action de plusieurs opérateurs

Retourner le nom et le prénom des étudiants de plus de 24 ans

Etudiant (nom, prenom, age)



Action combinée σ et Π

Sémantique : nom et prénom des étudiants de plus de 24 ans :

$$\text{Résultat} = \Pi_{nom,prenom}(\sigma_{age>24}(\text{Etudiant}))$$

Etudiant en extension

nom	prénom	age
'Dusol'	'Marie'	22
'Bony'	'Paul'	25
'Balard'	'Zoé'	29

Résultat

nom	prénom
'Bony'	'Paul'
'Balard'	'Zoé'

La jointure

notée $\bowtie$; opérateur binaire

R3 $=$ R1 $\bowtie$ R2 : opération qui s'applique à deux relation opérandes qui possèdent au moins un attribut commun sur lequel opérer une comparaison. Le schéma résultant sera la juxtaposition des schémas de R1 et de R2, et les tuples résultants correspondront à la combinaison des tuples de R1 et de R2, lorsqu'ils satisfont la condition de jointure posée.

La jointure peut aussi s'écrire à partir de X et σ :
R1 $\bowtie$ R2 $\equiv \sigma_{conditiondejointure}$(R1 X R2)

Un exemple concret : Etudiant $\bowtie$ Inscrit_Dans

Sémantique : Etudiants et modules dans lesquels ils sont inscrits (lorsqu'ils sont inscrits à au moins un module) numINE = numEtudiant

Etudiant en extension

numINE	nom	prénom
'20164545'	'Dusol'	'Marie'
'20165546'	'Bony'	'Paul'
'20174533'	'Balard'	'Zoé'

Inscrit_Dans en extension

numEtudiant	codeModule
'20164545'	'HMIN112M'
'20164545'	'HMIN110M'
'202095995'	'HMIN110M'

jointure sur la condition numINE = numEtudiant : un seul attribut sur les deux est conservé

numINE	nom	prénom	codeModule
'20164545'	'Dusol'	'Marie'	'HMIN112M'
'20164545'	'Dusol'	'Marie'	'HMIN110M'

Un exemple concret de jointure

Explication par rapport au produit cartésien

Etudiant en extension

numINE	nom	prénom
'20164545'	'Dusol'	'Marie'
'20165546'	'Bony'	'Paul'
'20174533'	'Balard'	'Zoé'

Inscrit_Dans en extension

numEtudiant	codeModule
'20164545'	'HMIN112M'
'20164545'	'HMIN110M'
'202095995'	'HMIN110M'

Produit cartésien juxtaposant les schémas avec condition de jointure numINE = numEtudiant

numINE	nom	prénom	numEtudiant	codeModule
'20164545'	'Dusol'	'Marie'	'20164545'	'HMIN112M'
~~'20165546'~~	~~'Bony'~~	~~'Paul'~~	~~'20164545'~~	~~'HMIN112M'~~
~~'20174533'~~	~~'Balard'~~	~~'Zoé'~~	~~'20164545'~~	~~'HMIN112M'~~
'20164545'	'Dusol'	'Marie'	'20164545'	'HMIN110M'
~~'20165546'~~	~~'Bony'~~	~~'Paul'~~	~~'20164545'~~	~~'HMIN110M'~~
~~'20174533'~~	~~'Balard'~~	~~'Zoé'~~	~~'20164545'~~	~~'HMIN110M'~~
~~'20164545'~~	~~'Dusol'~~	~~'Marie'~~	~~'202095995'~~	~~'HMIN110M'~~
~~'20165546'~~	~~'Bony'~~	~~'Paul'~~	~~'202095995'~~	~~'HMIN110M'~~
~~'20174533'~~	~~'Balard'~~	~~'Zoé'~~	~~'202095995'~~	~~'HMIN110M'~~

La jointure

Différentes catégories de jointure

- équijointure : opérateur arithmétique est l'égalité
- jointure naturelle : opérateur arithmétique est l'égalité et les deux attributs mis en correspondance sont les attributs communs présents dans les deux tables
- theta-jointure ou Θ-jointure : opérateur arithmétique est autre chose que l'égalité

Opération commutative : $R1 \bowtie R2 \equiv R2 \bowtie R1$
Opération associative : $R1 \bowtie (R2 \bowtie R3) \equiv R2 \bowtie (R1 \bowtie R3)$

En visuel

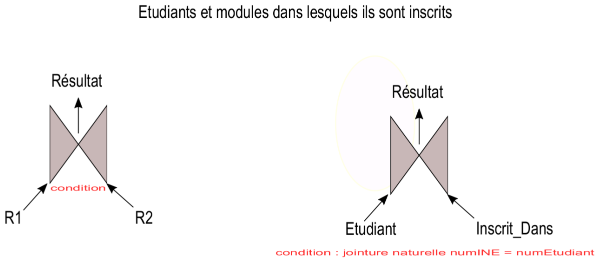


Figure: Algèbre relationnelle sous forme d'arbre

La division ou quotient

notée $\div$

Play Sound

R3 $=$ R1 $\div$ R2 : l'opérateur s'applique à deux relation opérandes notées R1 et R2, avec R2, un sous-schéma de R1. Le schéma résultant sera le schéma de R1 privé de celui de R2, et les tuples résultants seront les "sous-tuples" de R1 qui trouvent une correspondance avec tous les tuples de R2

Soit R1 de schéma $(a_1, a_2, \ldots, a_p, a_{p+1}, \ldots, a_n)$ et R2 de schéma $(a_{p+1}, \ldots, a_n)$, R3 aura pour schéma $(a_1, a_2, \ldots, a_p)$ et aura pour tuples, tous les tuples de R1 notés t1, tel que pour tout tuple t2 dans R2, il existe t1t2 dans R1.

Représentation visuelle

La division est l'opération la plus compliquée de l'algèbre relationnelle

Play Sound

Division : R3 = R1 ÷ R2

Un exemple concret de division

Les étudiants inscrits dans tous les modules : $\Pi_{numEtudiant,codeModule}(\text{Inscrit\_Dans}) \div \Pi_{codeModule}(\text{Module})$

Module en extension

codeModule	nom
HMIN112M	'SI BD'
HMIN111M	'Programmation'

Inscrit_Dans en extension

numEtudiant	codeModule
'20164545'	'HMIN112M'
'20164545'	'HMIN111M'
'202095995'	'HMIN111M'

Play Sound

Relation résultat de schéma (numEtudiant)

numEtudiant
'20164545'

La division

La division répond intuitivement à une question comprenant le mot "tous", par exemple donner les fournisseurs qui fournissent TOUS les produits, ou encore donner les étudiants qui sont inscrits dans TOUS les modules

Play Sound

La division peut s'obtenir à partir d'une double différence

$R(X,Y) \div S(Y) = \Pi_X(R) - \Pi_X(\Pi_X(R) \times \Pi_Y(S) - \Pi_{X,Y}(R))$

UNIVERSITÉ DE MONTPELLIER

Retour sur l'exemple

$\Pi_{numEtudiant,codeModule}$(Inscrit_Dans)$\div\Pi_{codeModule}$(Module)

Play Sound

équivaut à :

$\Pi_{numEtudiant}$(Inscrit_Dans) - $\Pi_{numEtudiant}$($\Pi_{numEtudiant}$(Inscrit_Dans) X $\Pi_{codeModule}$(Module) - $\Pi_{numEtudiant,codeModule}$(Inscrit_Dans))

UNIVERSITÉ DE MONTPELLIER

Un exemple concret de division

Les étudiants inscrits dans tous les modules :
$\Pi_{numEtudiant,codeModule}(\text{Inscrit\_Dans}) \div \Pi_{codeModule}(\text{Module})$

$\Pi_{numEtudiant}$(Inscrit_Dans) X $\Pi_{codeModule}$(Module) : tous les possibles

numEtudiant	Module.codeModule
'20164545'	'HMIN112M'
'20164545'	'HMIN111M'
'202095995'	'HMIN112M'
'202095995'	'HMIN111M'

Moins ceux qui existent vraiment dans Inscrit_Dans

numEtudiant	codeModule
'20164545'	'HMIN112M'
'20164545'	'HMIN111M'
'202095995'	'HMIN111M'

Play Sound

Résultat de $\Pi_{numEtudiant}$(Inscrit_Dans) X $\Pi_{codeModule}$(Module) - $\Pi_{numEtudiant,codeModule}$(Inscrit_Dans))

numEtudiant	codeModule
'202095995'	'HMIN112M'

Un exemple concret de division

Les étudiants inscrits dans tous les modules : $\Pi_{numEtudiant,codeModule}(\text{Inscrit\_Dans}) \div \Pi_{codeModule}(\text{Module})$

Play Sound

$\Pi_{numEtudiant}$(Inscrit_Dans)

numEtudiant
'20164545'
'202095995'

Moins ceux qui existent dans

numEtudiant
'202095995'

Résultat final

numEtudiant
'20164545'

Autres opérateurs

Il est à noter l'existence d'opérateurs complémentaires, à l'exemple de la semi-jointure ($\ltimes$), l'anti-projection et le complément ($\neg$) qui peuvent venir appuyer l'implémentation et le code d'opérateurs physiques associés

Play Sound

Semi-jointure : opération portant sur deux relations notée R1 $\ltimes$ R2 qui consiste à projeter sur les attributs de R1 le résultat de la jointure naturelle notée R1 $\bowtie$ R2

$R1 \ltimes R2 = \Pi_{attdeR1}(R1 \bowtie R2)$