



Bases de données - Le modèle relationnel

Olivier Ridoux

► To cite this version:

| Olivier Ridoux. Bases de données - Le modèle relationnel. Licence. France. 2021. hal-03212428

HAL Id: hal-03212428

<https://univ-rennes.hal.science/hal-03212428>

Submitted on 29 Apr 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Prérequis : lire, comprendre, apprendre, puis procéder au pliage.

Pliage : autre face dessus, traits gris rentrants, traits rouges saillants. Découper selon le trait rouge entre les deux ●, puisachever le pliage.

① Théorie naïve des ensembles

Un **ensemble** est une collection d'objets où le rang et la multiplicité ne comptent pas. On appelle **éléments** les objets de la collection, et on dit qu'un élément **appartient** à un ensemble. Cela se note $e \in E$ pour « e appartient à E », et $e, e' \in E$ pour « e et e' appartiennent à E ».

On note une collection entre accolades (...) Ex. $\{a, b, c\}$ est une collection, et puisque le rang ne compte pas, la collection $\{b, c, a\}$ dénote le même ensemble, et puisque la multiplicité ne compte pas non plus, la collection $\{a, b, a, c\}$ dénote encore le même ensemble.

Dans la théorie naïve des ensembles on ne considère que des objets qu'il est toujours possible de distinguer les uns des autres, mais à part cela n'importe quoi peut être un objet, y compris un ensemble.

Une collection vide, {}, constitue un ensemble particulier qu'on appelle l'**ensemble vide**, et qu'on note \emptyset ou $\{\}$. Il n'y a qu'un ensemble vide. On appelle **singleton** un ensemble qui n'a qu'un élément. L'ensemble $\{\emptyset\}$ est donc un singleton. Ne pas confondre l'élément a et le singleton $\{a\}$. Les confondre revient à commettre une erreur de type en programmation. Ne pas confondre non plus \emptyset et $\{\emptyset\}$.

La collection des éléments d'un ensemble s'appelle l'**extension** de l'ensemble. On peut aussi spécifier les éléments d'un ensemble par une propriété qui les distingue de ceux qui n'appartiennent pas à l'ensemble. Cette propriété s'appelle l'**intension** de l'ensemble (l'intension n'a rien à voir avec l'intention, ou dessein). On exprime une intension soit par une phrase en prenant le risque d'être imprécis ou ambigu, soit par une formule logique en prenant le risque d'être fastidieux. N'importe quelle formule logique fausse (contradictoire ou absurde) est l'intension de \emptyset .

Ex. « **les lettres communes aux mots "bancal" et "tabac"** » est une intension possible pour l'ensemble constitué de la collection $\{a, b, c\}$. Les « **3 premières lettres de l'alphabet** » est une intension équivalente. Un ensemble peut avoir de nombreuses intensions, certaines même difficiles à comparer.

Si tous les éléments d'un ensemble E_1 appartiennent à un ensemble E_2 , on dit que E_1 est **inclus** dans E_2 , et on note $E_1 \subseteq E_2$. On dit aussi que E_1 est un **sous-ensemble** ou une **partie** de E_2 . On note $E_1 \subsetneq E_2$ pour signifier l'inclusion sans égalité possible (on parle alors de sous-ensemble strict), et $E_1 \subseteq E_2$ pour insister sur l'égalité possible.

② Algèbre des ensembles

Etant donné un univers U d'objets de références, on peut former une structure algébrique sur les sous-ensembles de U avec les opérations suivantes :

Produit cartésien (noté \times) : $A \times B$ est l'ensemble formé de tous les couples (a, b) où $a \in A$ et $b \in B$. Bien noter que l'ordre compte et que $(a, b) \neq (b, a)$ sauf si $a=b$. Noter aussi que $A \times \emptyset = \emptyset$.

Ensemble des parties (noté $\mathcal{P}(E)$ ou 2^E) : $\mathcal{P}(E)$ est l'ensemble formé de toutes les parties de E , donc un ensemble d'ensembles. Noter que $\mathcal{P}(\emptyset) = \{\emptyset\}$ et $\mathcal{P}(\{a\}) = \{\emptyset, \{a\}\}$.

Union (notée U) : $A \cup B$ est l'ensemble formé des éléments de A et de ceux de B . Comme la multiplicité ne compte pas, il est indifférent qu'un élément appartienne aux deux ensembles ou seulement à l'un d'entre eux. Par exemple, $\{a, b\} \cup \{c, b\} = \{a, b, c\} = \{a, b, c\}$. L'union est **commutative, associative, idempotente**, et a **∅ pour élément neutre**. Cela autorise la notation d'union étendue comme $\bigcup_{i \in [1, n]} E_i$. L'intension d'une union d'ensembles est la disjonction des intensions de ces ensembles. Noter que $A \subseteq B$ ssi $A \cup B = B$, et que $\bigcup_{i \in \emptyset} E_i = \emptyset$.

Intersection (notée \cap) : $A \cap B$ est l'ensemble formé des éléments de A qui appartiennent aussi à B . Par exemple, $\{a, b\} \cap \{c, b\} = \{b\}$. L'intersection est **commutative, associative, idempotente**, et a **U pour élément neutre**. Cela autorise la notation d'intersection étendue comme $\bigcap_{i \in [1, n]} E_i$. L'intension d'une intersection d'ensembles est la conjonction des intensions de ces ensembles. Noter que $A \subseteq B$ ssi $A \cap B = A$, et que $\bigcap_{i \in \emptyset} E_i = U$.

Complémentation (notée \complement) : $\complement_A B$ est l'ensemble formé de la collection des éléments de A qui n'appartiennent pas à B ; le **complémentaire** de B dans A . On note parfois $A \setminus B$ ou même $A - B$. Il n'est pas nécessaire que B soit inclus dans A ! Quand l'ensemble de référence est U on peut se dispenser de le noter : $\complement B = \complement_U B$. Noter que $\complement (\complement B) = B$. L'intension du complémentaire d'un ensemble est la conjonction de l'intension de l'ensemble de référence et de la négation de l'intension de l'ensemble.

L'intersection et l'union sont **distributives** l'une par rapport à l'autre : pour tout triplet d'ensembles A , B et C , $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ et $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$. La complémentation est une sorte de négation, et suit les **lois de De Morgan** : $\complement(A \cup B) = \complement A \cap \complement B$ et $\complement(A \cap B) = \complement A \cup \complement B$.

Même si l'une fait penser à l'addition et l'autre à la soustraction, l'**union et la complémentation ne sont pas les opposées l'une de l'autre** ; pour certaines paires d'ensembles A et B , on a $(A \setminus B) \cup B \neq A$ ou $(A \setminus B) \setminus B \neq A$. Ex. $\{(a) \setminus (b)\} \cup \{b\} = \{a, b\}$ et $\{(a, b) \setminus \{b\}\} \setminus \{b\} = \{a\}$.

③ Cardinalité des ensembles finis

Le nombre d'éléments d'un ensemble fini est appelé la cardinalité de l'ensemble. La cardinalité d'un ensemble E est notée $\text{card}(E)$ ou $|E|$. La cardinalité des ensembles suit les lois suivantes :

Produit cartésien : $|A \times B| = |A| \times |B|$. Noter la surcharge du signe \times qui opère tantôt sur des ensembles (le produit cartésien) et tantôt sur des entiers (le produit arithmétique).

Ensemble des parties : $|\mathcal{P}(E)| = 2^{\text{card}(E)}$, ce qui explique un peu la notation 2^E pour l'ensemble des parties de E .

Union : $\max\{|A|, |B|\} \leq |A \cup B| \leq |A| + |B|$ car des éléments peuvent appartenir aux deux ensembles, mais ne doivent être comptés qu'une fois.

Intersection : $0 \leq |A \cap B| \leq \min\{|A|, |B|\}$. Plus précisément, $|A \cup B| = |A| + |B| - |A \cap B|$.

Complémentation : $|\complement_A B| = |A| - |A \cap B|$.

① Logique des prédictats

Dans sa définition la plus naïve la **logique des prédictats** est la formalisation de la logique employée tous les jours dans les activités un peu scientifiques. Elle permet d'exprimer des **jugements** sous forme de formules, et de décider formellement si elles sont **vraies** ou **fausses**. Les formules de la logique des prédictats sont définies de la façon suivante :

Prédicats atomiques (ou **atomes**) : Ce sont des formules élémentaires qui expriment des jugements sur des objets. Nous ne nous prononçons pas sur la notation à employer dans les atomes ; cela n'a pas vraiment d'importance. Par exemple,

- Rennes est une capitale, où Rennes est l'objet du jugement et être une capitale en est le prédictat.
- Rennes est la capitale de la France, où Rennes et la France sont des objets et être la capitale de est le prédictat.
- $x > y$, où x et y sont les objets, et où $>$ (être plus grand que) est le prédictat du jugement.
- $1273 + 556$ est un nombre premier, où $1273 + 556$ est l'objet et être un nombre premier est le prédictat.
- Le Grand théorème de Fermat est vrai, où le Grand théorème de Fermat est l'objet et être vrai est le prédictat.
- $P \neq NP$, où P et NP sont les objets et \neq (être différent) le prédictat.

On peut attribuer une valeur de vérité, **Vrai** ou **Faux**, à un atome, en se référant à une réalité de terrain, comme pour Rennes, à des théories et des calculs, comme pour la primalité de $1273 + 556$, ou à des preuves externes, comme celle du grand théorème de Fermat. Parfois on ne peut pas, par manque d'information, comme pour $x > y$, ou parce qu'on ne sait vraiment pas, comme pour $P \neq NP$.

Formules propositionnelles : Ce sont des formules, élémentaires ou non, qui sont reliées par des connecteurs logiques, généralement \wedge , \vee , \neg , ou \Rightarrow .

- **conjonction**, $\phi_1 \wedge \phi_2$: relie deux formules pour en former une troisième qui est vraie si les deux premières le sont. Le connecteur \wedge est commutatif, associatif, et a pour élément neutre la valeur de vérité **Vrai**. Cela permet la notation de conjonction étendue comme $\bigwedge_{i \in [1, n]} \phi_i$. Si on convenait que Faux < Vrai, $\phi_1 \wedge \phi_2$ serait le minimum de ϕ_1 et ϕ_2 . Si on convenait que Faux = 0 et Vrai = 1, $\phi_1 \wedge \phi_2$ serait $\phi_1 \times \phi_2$.
- **disjonction**, $\phi_1 \vee \phi_2$: relie deux formules pour en former une troisième qui est vraie si au moins une des deux premières l'est. Le connecteur \vee est commutatif, associatif, et a pour élément neutre la valeur de vérité **Faux**. Cela permet la notation de disjonction étendue comme $\bigvee_{i \in [1, n]} \phi_i$. Si on convenait que Faux < Vrai, $\phi_1 \vee \phi_2$ serait le maximum de ϕ_1 et ϕ_2 . Si on convenait que Faux = 0 et Vrai = 1, $\phi_1 \vee \phi_2$ serait $1 - (\neg \phi_1) \times (1 - \phi_2)$.
- **implication**, $\phi_1 \Rightarrow \phi_2$: relie deux formules pour en former une troisième qui est fausse si ϕ_1 est vraie alors que ϕ_2 est fausse. Le connecteur \Rightarrow n'est ni commutatif ni associatif. Si on convenait que Faux < Vrai, $\phi_1 \Rightarrow \phi_2$ serait Vrai si $\phi_1 \leq \phi_2$.
- **négation**, $\neg \phi$: constitue une formule qui est fausse si ϕ est vraie. Si on convenait que Faux = 0 et Vrai=1, $\neg \phi$ serait $1 - \phi$.

Formules quantifiées : Ce sont des formules, élémentaires ou non, dont la valeur de vérité s'évalue par rapport à un ensemble d'objets, plutôt que par rapport à des objets pris individuellement.

- **quantification universelle**, $\forall x . \phi(x)$: constitue une formule qui est vraie si ϕ est vraie de tous les éléments du domaine. Si le domaine est fini, la quantification universelle est équivalente à une conjonction des applications de ϕ à tous les éléments du domaine : $\forall x \in \{e_1, e_2, \dots, e_n\} . \phi(x)$ si $\bigwedge_{i \in [1, n]} \phi(e_i)$.
- **quantification existentielle**, $\exists x . \phi(x)$: constitue une formule qui est vraie si ϕ est vraie d'au moins un élément du domaine. Si le domaine est fini, la quantification existentielle est équivalente à une disjonction des applications de ϕ à tous les éléments du domaine : $\exists x \in \{e_1, e_2, \dots, e_n\} . \phi(x)$ si $\bigvee_{i \in [1, n]} \phi(e_i)$.

Parenthèses et priorité des opérateurs : Il est courant d'assigner des priorités d'opérateurs aux différents connecteurs et quantificateurs afin de spécifier comment se lisent les formules complexes en l'absence de parenthèses. Nous pensons que c'est utile mais absolument pas fondamental car ces conventions dépendent trop des outils, et même quand ce n'est pas le cas, il n'est jamais très prudent de se reposer trop lourdement sur les priorités des opérateurs quand le coût de quelques parenthèses est si faible devant le coût d'une grosse erreur. Ce genre d'expertise est à réservé aux experts. Il vaut mieux ne pas être avare de parenthèses, et nous proposons d'utiliser les parenthèses rondes (...) pour structurer les connecteurs, et les parenthèses carrées (ou crochets, [...]) pour les quantificateurs.

N'oubliez jamais !

Un **modèle** est toujours **imparfait**, parfois **utile**, et c'est tout ce qu'on peut lui demander (d'après George Box 1978, voir aussi a contrario les cartes à l'échelle 1/1 de Jorge Luis Borges 1946 et Lewis Carroll 1893, si parfaites qu'elles sont inutiles).

L'informatique ne travaille que sur la **représentation** des choses. Il faut toujours se demander ce qui a du sens par rapport à ce qui est représenté.

② Idiomes logiques

En pratique, on n'utilise pas n'importe quelle formule de la logique des prédictats. On a tendance à utiliser des expressions idiomatiques qu'il convient de reconnaître et interpréter correctement au premier coup d'œil.

Quantifications dans un domaine : Très souvent, on écrit des formules comme $\forall x \in E . \phi(x)$ ou $\forall x \text{ tq } \psi(x) . \phi(x)$. C'est une façon de dire dans quel domaine doit être évaluée la quantification. Ces formules **doivent** être lues $\forall x . [x \in E \Rightarrow \phi(x)]$ et $\forall x . [\psi(x) \Rightarrow \phi(x)]$. Une conséquence directe est qu'une quantification universelle sur un domaine vide est trivialement vraie. Cela paraît une situation étrange, mais c'est banal en informatique, spécialement quand on considère les cas d'initialisation : ex. **Tous les utilisateurs sont ...** lorsqu'il n'y a pas d'utilisateurs. De la même façon, on écrit des formules comme $\exists x \in E . \phi(x)$ ou $\exists x \text{ tq } \psi(x) . \phi(x)$. Ces formules **doivent** être lues $\exists x . [x \in E \wedge \phi(x)]$ et $\exists x . [\psi(x) \wedge \phi(x)]$. On voit alors qu'une quantification existentielle sur un domaine vide est trivialement fausse.

Cascades d'implications : On écrit parfois des formules comme $\phi_1 \Rightarrow \phi_2 \Rightarrow \phi_3$, qui pourrait être lue soit $(\phi_1 \Rightarrow \phi_2) \Rightarrow \phi_3$ soit $\phi_1 \Rightarrow (\phi_2 \Rightarrow \phi_3)$. Dans le premier cas, on exprime que $\phi_1 \Rightarrow \phi_2$ est un théorème qui pourrait servir à démontrer ϕ_3 . Dans le second cas, la formule est équivalente à $(\phi_1 \wedge \phi_2) \Rightarrow \phi_3$, donc $(\phi_1 \wedge \phi_2) \Rightarrow \phi_3$, et donc $\phi_2 \Rightarrow (\phi_1 \Rightarrow \phi_3)$. Ce qui n'a rien à voir avec la première lecture. Conclusion, faire des économies de bouts de parenthèses avec discernement !

Formules de De Morgan (Augustus De Morgan, 1806-1871) : La négation a à voir avec le complémentaire d'une situation, mais le complémentaire d'une situation compliquée est souvent encore plus compliqué que la situation. Les formules de De Morgan (qui ne sont pas toutes dues à De Morgan) peuvent être mises à profit pour pousser les négations vers l'intérieur des formules où elles s'appliqueront à des formules plus petites.

- $\neg(\phi_1 \wedge \phi_2)$ est équivalent à $\neg \phi_1 \vee \neg \phi_2$ et $\neg(\phi_1 \vee \phi_2)$ est équivalent à $\neg \phi_1 \wedge \neg \phi_2$.
- $\neg(\phi_1 \Rightarrow \phi_2)$ est équivalent à $\phi_1 \wedge \neg \phi_2$.
- $\neg \neg \phi$ est équivalent à ϕ . Pose beaucoup plus de difficultés que la taille de l'identité ne le laisse penser, mais est vrai pour la logique usuelle. Se rappeler combien nous humains sommes peu doués pour les doubles négations.
- $\neg \forall x . \phi(x)$ est équivalent à $\exists x . \neg \phi(x)$ et $\neg \exists x . \phi(x)$ est équivalent à $\forall x . \neg \phi(x)$.

Cascades de quantifications : On imbrique souvent les quantifications. Certaines imbrications doivent faire réfléchir, et d'autres moins.

- $\exists x . \exists y . \phi(x, y)$ est équivalent à $\exists y . \exists x . \phi(x, y)$ et on note souvent $\exists x, y . \phi(x, y)$.
- $\forall x . \forall y . \phi(x, y)$ est équivalent à $\forall y . \forall x . \phi(x, y)$ et on note souvent $\forall x, y . \phi(x, y)$.
- $\forall x . \exists y . \phi(x, y)$ exprime que pour chaque x il y a un y , qui peut dépendre de x , qui a la propriété désirée. Ex. dans $\forall x . \exists y . x \times y = 0$, le même y convient pour tous les x , mais dans $\forall x . \exists y . x \times y = 1$, à tout x correspond un y différent. Un y existentiel est donc implicitement une fonction de tous les x universels qui le précédent.
- $\exists x . \forall y . \phi(x, y)$ exprime qu'un x unique a la propriété désirée pour tous les y . Ex. $\$