

Modélisation de données

Glossaire

AAA_03

v100a

2022-01-04



Christina.Khnaisser@USherbrooke.ca

Luc.Lavoie@USherbrooke.ca

© 2018-2021, Μηττις (<http://info.usherbrooke.ca/llavoie>)

CC BY-NC-SA 4.0 (<https://creativecommons.org/licenses/by-nc-sa/4.0/>)

Plan

- Vocabulaire
- Unités de mesure
- Théorèmes, corollaires et constats
- Références
- Les colles du prof



Vocabulaire

Définition de quelques termes et concepts utiles en modélisation, conception et exploitation de bases de données

- informatique
- connaissance
- information
- signe
- donnée
- représentation
- signal
- valeur
- type
- type de base
- sous-type
- modèle
- schéma
- base de données
- ...

Vocabulaire – Définitions générales (1)

○ informatique

- 1a – Science du traitement de l’information.
- 1b – Science du traitement automatique et rationnel de l’information.
- 2 – Ensemble des techniques de la collecte, du tri, de la mise en mémoire, du stockage, de la transmission, et de l’utilisation des informations traitées automatiquement à l’aide de logiciels mis en oeuvre sur des ordinateurs.

Vocabulaire – Définitions générales (1')

○ informatique

- 1a – Science du traitement de l'*information*.
- 1b – Science du traitement automatique et rationnel de l'*information*.
- 2 – Ensemble des techniques de la collecte, du tri, de la mise en mémoire, du stockage, de la transmission, et de l'utilisation de *information* traitée automatiquement à l'aide de logiciels mis en oeuvre sur des ordinateurs.

Vocabulaire – Définitions générales (2)

○ information

- Élément de connaissance susceptible d'être transmis au moyen d'une suite de signes.
- Élément de connaissance représentable par une donnée.

○ signe

- *Le Robert* : Représentation matérielle simple qui se rapporte conventionnellement, dans une société donnée, à une réalité complexe (symbole).
- *Larousse* : Représentation matérielle d'une chose, dessin, figure ou son ayant un caractère conventionnel.
- *Wikipedia* : Marque, naturelle ou conventionnelle, désignant pour quelqu'un un objet ou un concept, et destinée à être interprétée par un tiers.
- *Antidote* : Ce qui représente quelque chose de complexe de façon simple et qui fait parfois partie d'un code.

Vocabulaire – Définitions générales (3)

○ connaissance

- *Faculté mentale produisant une assimilation par l'esprit d'un contenu objectif préalablement traduit en signes et en idées.*
- *Résultat de cette opération. La connaissance est une possession symbolique des choses. Elle comprend une infinité de degrés. La connaissance rationnelle, méthodique universelle a parfois été opposée au savoir empirique, chaotique, objectif.*

Références :

- Christian Godin, Dictionnaire de philosophie, Paris, Fayard, 2004, ISBN 978-2-213-62116-6.
- Cité par Wikipédia, <https://fr.wikipedia.org/wiki/Connaissance> [2020-08-16]

Vocabulaire – Définitions spécialisées (1)

○ donnée

- valeur associée à une représentation (apte à être traitée par ordinateur).

○ représentation

- suite de signaux

○ signal

- phénomène physique mesurable (donc suffisamment stable pour être mesuré et traité par un ordinateur).

Vocabulaire – Définitions spécialisées (2)

○ valeur

- élément d'un type de base.

○ type de base (type racine, domaine)

- ensemble fini de valeurs propres n'appartenant à aucun autre type de base.

○ sous-type (type dérivé)

- sous-ensemble d'un type déterminé par une contrainte explicite (qui restreint les valeurs acceptées).

○ type

- ensemble fini de valeurs ;
- dans le présent document, soit un type de base, soit un sous-type.

Vocabulaire – Remarques (1)

- Remarquons la circularité entre la définition de valeur et celle de type. Leur existence relève des postulats de base de la théorie de l'information.
- Remarquons également la circularité entre connaissance, information, donnée et signe.
- Quelqu'un a-t-il une meilleure suggestion ?
- Il y a peut-être un Ph. D. à la clé !

Vocabulaire – Remarques (2)

- Exigences applicables aux valeurs
 - Toute valeur doit avoir au moins une représentation.
 - Une même valeur peut donc avoir plusieurs représentations.
 - Lorsqu'une même représentation est associée à plus d'une valeur, le « contexte » doit permettre d'inférer la valeur que représente une donnée.
 - Par exemple, attendu un tuple, un attribut donne accès à une donnée dont la valeur appartient au type de l'attribut.
 - L'attribut fournit donc le contexte par le type qui lui est associé.

Vocabulaire – Définitions spécialisées (4)

- Un **modèle** est une représentation applicable à un objet (resp. personne, système) appartenant à ensemble d'objets (resp. personnes, systèmes).
- En science, un modèle doit en outre être formel, c'est-à-dire que la représentation doit être exprimée en respectant un langage formel bien défini.
- Le plus souvent, une théorie scientifique comprend (au moins) un modèle de référence.

Vocabulaire – Définitions spécialisées (5)

- En informatique, un *modèle conceptuel* comporte la description d'ensemble d'entités et de leurs relations, motivés et interprétables en regard d'un corpus de connaissances donné.
- La *portée* d'un *modèle conceptuel* est l'ensemble des informations dont il permet l'interprétation (en regard du corpus de connaissances pris en compte par le modèle conceptuel).
- Un *modèle logique* est un ensemble de prédicats logiques dont les propositions permettent de représenter toutes les informations de la *portée* du *modèle conceptuel* auquel il est associé.
- Un *modèle relationnel* est un *modèle logique* dont les prédicats sont représentés par des relations et les propositions par des tuples.

Vocabulaire – Remarques (4)

- En informatique, dans la pratique courante, il est commun :
 - de suffixer la locution «de données» aux termes modèle, modèle conceptuel et modèle logique :
 - modèle conceptuel de données (MCD),
 - modèle logique de données (MLD) ;
 - de remplacer le mot modèle par le mot schéma :
 - schéma conceptuel de données,
 - schéma logique de données ;
 - voire d’escamoter l’adjectif :
 - modèle de données ;
 - schéma de données.

Vocabulaire – Définitions spécialisées (6)

- Une **base de données** (BD) est une instance d'un *modèle logique de données*, c'est-à-dire :
 - un «contenant» regroupant les données structurées conformément à un modèle logique associé à un modèle conceptuel de référence ;
- Un **système de gestion de base de données** (SGBD) est un système qui permet
 - de gérer (définir, accorder les accès, etc.) des BD,
 - de définir et faire évoluer leur MLD ainsi que
 - d'en manipuler (insérer, modifier, retirer, extraire) les données tout en garantissant le respect de leur MLD par le biais de transactions concurrentes.

Vocabulaire – Usages (1)

○ schéma (d'une base de données) \equiv

- *modèle logique de données*

○ base de données \equiv

une instance d'un schéma \equiv

une instance d'un modèle logique de données \equiv

- *un contenant regroupant les données représentant un état de la réalité telle que modélisée grâce au modèle conceptuel à l'origine du schéma.*

Vocabulaire – Remarques (5)

- En général, un contrat très strict doit être établi entre le SGBD et les utilisateurs habilités à lui soumettre des transactions. Un type de contrat très répandu est le contrat ACID (atomicité, cohérence, indépendance, durabilité) qui sera présenté dans une leçon ultérieure.
- Dans un système d'information (SI), par l'entremise d'un SGBD, on tente en général de maintenir l'état d'une BD de telle sorte qu'elle reflète la réalité au mieux avec un minimum de décalage temporel.
- De plus en plus, on désire cependant retenir l'évolution de la réalité (son «histoire»), d'où la nécessité de définir des modèles logiques temporalisés, voire historicisés. Les entrepôts de données sont une tentative en ce sens qui sera présentée dans une leçon ultérieure.

Vocabulaire – Langages et notations

- Un **script** est la description textuelle d'un modèle (de données) conforme à une théorie (entité-association, relationnelle, etc.).
- Un **diagramme** est une représentation graphique (le plus souvent incomplète) d'un modèle.
- Nous utiliserons souvent un langage de description textuel inspiré de l'algèbre relationnelle (sous le nom « **langage relationnel** »).
- **SQL** comprend un langage de description textuelle de modèles relationnels... c'est l'un des langages de programmation les plus utilisés au monde.
- Nous utilisons en cours deux notations graphiques
 - la notation graphique classique pour les **diagrammes relationnels** (des rectangles pour les relations, des flèches pour les clés référentielles) ;
 - la notation graphique de Chen pour les **diagrammes entité-association** (des rectangles, des ovales, des losanges et divers types de flèches) ;
 - d'autres notations, en particulier celle de **Merise**, sont couramment utilisées.
- **UML** est une autre notation graphique fréquemment utilisée pour les diagrammes de classes or, plusieurs l'utilisent aussi pour représenter des diagrammes entité-association ce qui conduit fréquemment à des confusions.

Vocabulaire – SQL et le modèle relationnel

- En SQL, les modèles relationnels sont décrits à l'aide du langage de définition des données (LDD) ou *Data Definition Language (DDL)*.
- Le LDD permet de déclarer (CREATE), de modifier (ALTER) ou de retirer (DROP) :
 - un type de base (TYPE),
 - un sous-type (DOMAIN),
 - une (variable de) relation de base (TABLE),
 - une (variable de) relation virtuelle (VIEW),
 - une contrainte (CONSTRAINT, ASSERTION),
 - une fonction (FUNCTION),
 - une procédure (PROCEDURE),
 - un automatisme (TRIGGER),
 - etc.

Vocabulaire – Mise en garde

- Polysémie des mots, dont
 - domaine
 - type
 - variable
 - modèle
 - schéma

Unités de mesure

Définition de quelques unités de mesure utiles en modélisation, conception et exploitation de bases de données

- volume

A venir :

- temps

- durée

- latence

- débit

Vocabulaire – Mesure d’encombrement et symboles ISO (1)

Unités internationales (symboles indépendants de la langue)

- bit (b)
- octet (o)

Préfixes (symboles indépendants de la langue)

Préfixe	Base2	Sym2	Base10	Sym10
kilo	2^{10}	ki	10^3	k
mega	2^{20}	Mi	10^6	M
giga	2^{30}	Gi	10^9	G
tera	2^{40}	Ti	10^{12}	T
exa	2^{50}	Ei	10^{15}	E
peta	2^{60}	Pi	10^{18}	P
zetta	2^{70}	Zi	10^{21}	Z
yotta	2^{80}	Yi	10^{24}	Y

Vocabulaire – Mesure d’encombrement et symboles ISO (2)

En 2022

- La taille d’une base de données se mesure souvent en Mo, Go ou To.
- La taille d’un entrepôt de données se mesure souvent en To ou en Eo.
- La taille du stockage disponible dans un centre d’hébergement se mesure souvent en Eo ou en Po.

Remarques

- Plusieurs usagers confondent les échelles binaires et décimales et utilisent de façon inappropriée, par exemple, Mo pour Mio et inversement.
- Dans certains pays anglo-saxons, le B (b majuscule) est utilisé en lieu et place du o (o minuscule) pour octet ; ce qui introduit fréquemment une confusion entre bit et octet (octet se disant *byte*, en américain), particulièrement dans le domaine des télécommunications. Ce sont essentiellement les mêmes pays qui continuent à utiliser le système impérial (IS) de poids et mesure plutôt que le système international (SI).

Théorèmes, corollaires et constats

Rappel de quelques théorèmes, corollaires et constats utiles en modélisation, conception et exploitation de bases de données

- Théorèmes d'incomplétude de Gödel
- Théorie, modèle et langage

Théorèmes d'incomplétude de Gödel

○ Synthèse

- *Toute théorie cohérente ayant un nombre fini d'axiomes exprimés dans un langage qui permet de décrire l'arithmétique comprend des propositions indécidables.*

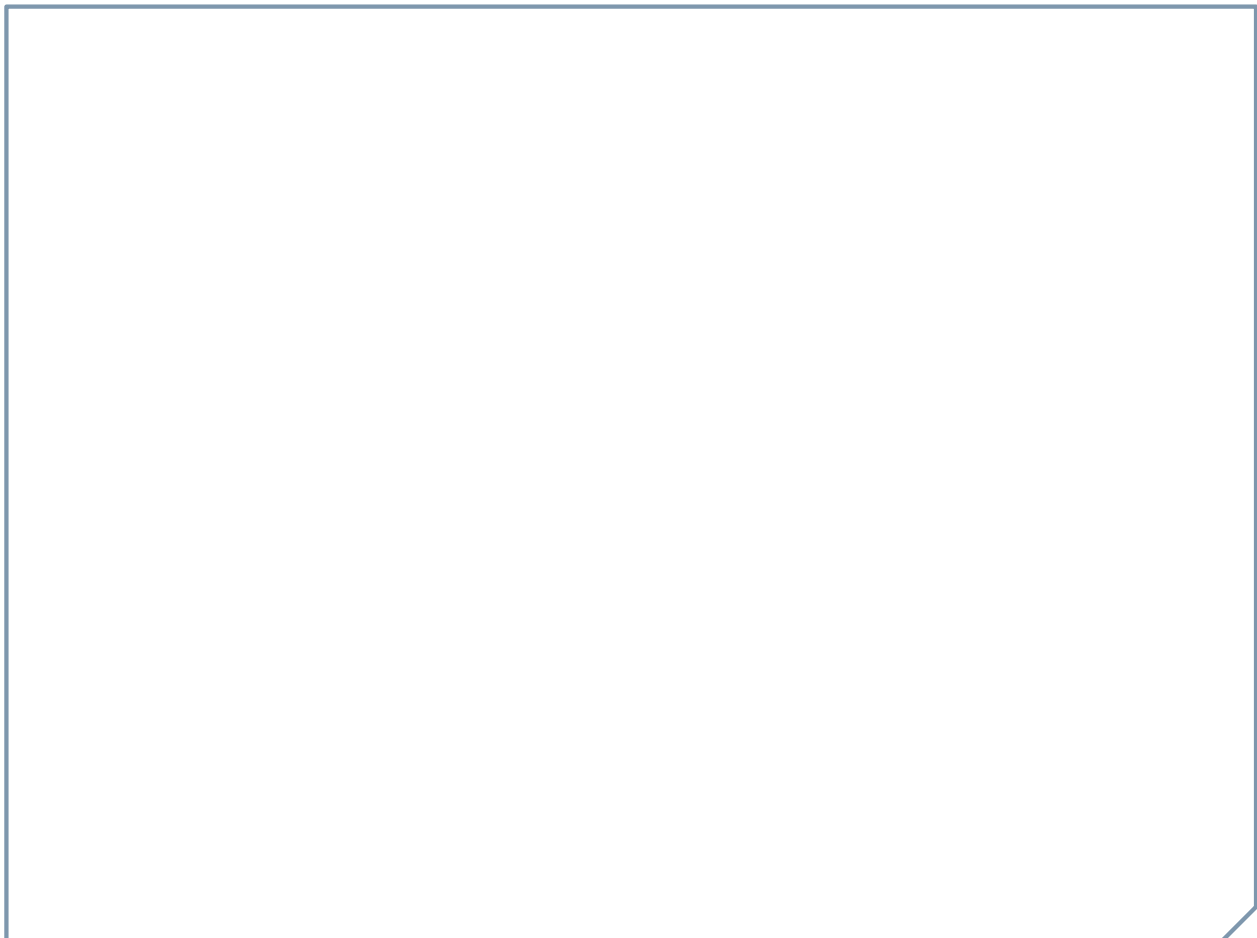
○ Exemples

- l'arithmétique de Peano,
 - la théorie des ensembles.
- Par ailleurs, un modèle n'est pas forcément complet relativement à une théorie.
- Par ailleurs, un langage n'est pas forcément complet relativement à un modèle.

Théorie, modèle et langage

- Une théorie (mathématique) est un ensemble d'affirmations dont certaines sont des axiomes et les autres des théorèmes démontrables à partir de ces axiomes et au moyen de règles d'inférence (exprimée à l'aide de la) logique.
 - <http://fr.wikipedia.org/wiki/Théorie>
- Un modèle est une représentation conforme à une théorie.
 - <http://fr.wikipedia.org/wiki/Modèle>
- Un langage (formel) est un formalisme permettant de décrire des propositions sémantiquement interprétables en termes d'un modèle.
 - http://fr.wikipedia.org/wiki/Langage_formel
- Une théorie peut être à l'origine de plusieurs modèles, un modèle de plusieurs langages, un langage de plusieurs dialectes.
- Pour en savoir plus sur le calcul des prédicats :
 - http://fr.wikipedia.org/wiki/Calcul_des_prédicats

Références



Références (1)

○ Théorie relationnelle

- E.F. Codd (1990) :
The Relational Model for Database Management: Version 2.
Boston, MA, USA : Addison-Wesley Longman Publishing Co., Inc.
- C.J. Date, H. Darwen (2007) :
Databases, types and the relational model: the third manifesto.
Reading, Mass.: Addison-Wesley.
- F. de Sainte Marie (2013) :
Bases de données relationnelles et normalisation : de la première à la sixième forme normale.
<https://fsmrel.developpez.com/basesrelationnelles/normalisation/>
- H. Darwen (2006) :
How To Handle Missing Information Without Using NULL.
<http://www.dcs.warwick.ac.uk/~hugh/TTM/Missing-info-without-nulls.pdf>

Références (2)

○ Théorie des types

- Luca Cardelli, Peter Wegner (1985) :
On understanding types, data abstraction, and polymorphism.
ACM Computing Surveys 17, 4; pp. 471–523.
DOI:<http://dx.doi.org/10.1145/6041.6042>
- Chris J. Date (2016) :
Type inheritance and relational theory: subtypes, supertypes, and substitutability.
O'Reilly Media, Sebastopol, CA.
ISBN:978-1-4919-5999-2

Les colles du prof

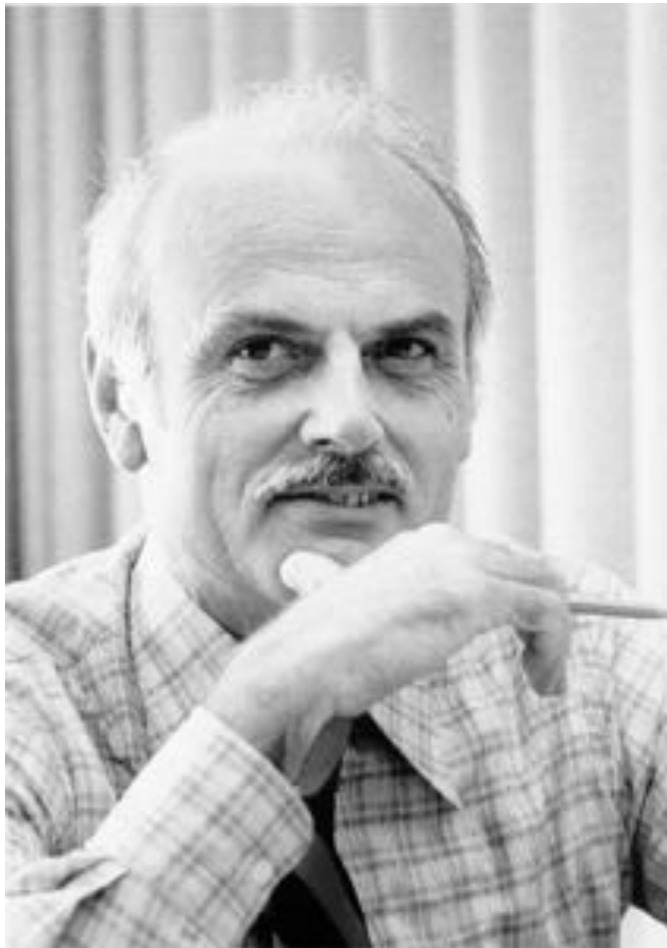




Les colles du prof

- Quelles différences existe-t-il entre
 - un type de base et sous-type ?
 - un tuple et une relation ?
 - un modèle logique et une base de données ?
 - une théorie et un modèle ?
 - un modèle et un langage ?

Edgar Frank Codd et Christopher J. Date



https://en.wikipedia.org/wiki/Edgar_F._Codd



Photo of Chris Date by Douglas Robertson, Edinburgh

https://en.wikipedia.org/wiki/Christopher_J._Date

