

Analyse conceptuelle – Première partie

Projet Base de Données – ENS Paris-Saclay

Guillaume Genestier
genestier@lsv.fr

25 janvier 2019

Motivation

- Utilité d'une base de données : modéliser un **univers** à propos duquel l'on stocke, extrait et gère des données.
- Univers modélisé par un schéma logique, en fonction de sa nature et de besoins.
- Le schéma logique est **indépendent** du SGBD utilisé.

Formalisme

Formalisme classique utilisé pour l'analyse conceptuelle : le modèle entité-association (EA).

Principe

- L'univers considéré est peuplé d'un nombre fini d'**entités**.
- Chacune de ces entités est une instance d'un concept abstrait, un **type d'entités** décrivant comment cette entité est caractérisée.
- Les entités peuvent être liées l'une à l'autre à travers des **associations**, qui décrivent une relation entre les entités impliquées.
- Un **type d'associations** permet de décrire, de manière abstraite, les associations entre entités d'un certain ensemble de types d'entité.

Définitions

- Chaque **type d'entités** est caractérisé par un **nom** (unique) et un ensemble d'**attributs**.
- Chaque attribut a un **nom** (unique pour ce type d'entités), un **type** (de données) et peut soit être **obligatoire**, soit être **optionnel**.
- Une **entité** est une instance d'un type d'entités associant, à chaque attribut, une valeur prise dans le domaine défini par son type de données ou la valeur non spécifiée NULL, si autorisée.

Représentation graphique

<nom>
<attribut_1> : <type_1>
...
<attribut_n> : <type_n>

PERSONNE
nss : uint
nom : varchar(30)
prenom : varchar(30)
date_naissance : date

nss \mapsto 3 34 06...

nom \mapsto Duck

prenom \mapsto Donald

date_naissance \mapsto 1934-06-09

INSTITUTION
nom : varchar(50)
ville : varchar(50)
budget[0-1] : uint

nom \mapsto U.Duckburg

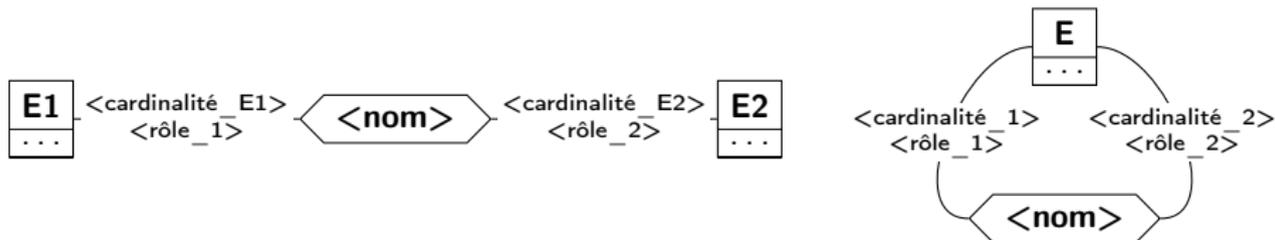
ville \mapsto Duckburg

budget \mapsto NULL

Définitions

- Chaque **type d'associations** est caractérisé par un **nom** (unique), une **paire de types d'entités** (non nécessairement distincts), chacun jouant un **rôle** (distinct) et ayant une **cardinalité** dans ce type d'associations.
- Pour un type d'associations non cyclique, le rôle joué par chaque type d'entités est par défaut implicitement son nom.
- Une **association** est une instance d'un type d'association, c.-à-d. simplement une paire d'entités des types impliqués.

Représentation graphique



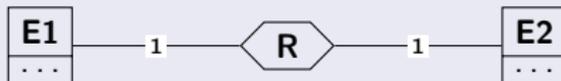
Prenons deux types d'entités E1 et E2, liés entre eux par le type d'associations R.

Classe fonctionnelle

- **Un-à-plusieurs** : chaque entité de E1 est associée à travers R à au plus une entité de E2, tandis que chaque entité de E2 est associée à travers R à zéro, une ou plusieurs entités de E1.

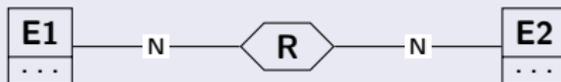


- **Un-à-un** : chaque entité de E1 est associée à travers R à au plus une entité de E2, et chaque entité de E2 est associée à travers R à au plus une entité de E1.



Hors classe fonctionnelle

- **Plusieurs-à-plusieurs** : chaque entité de E1 est associée à travers R à zéro, une ou plusieurs entités de E2, et chaque entité de E2 est associée à travers R à zéro, une ou plusieurs entités de E1.

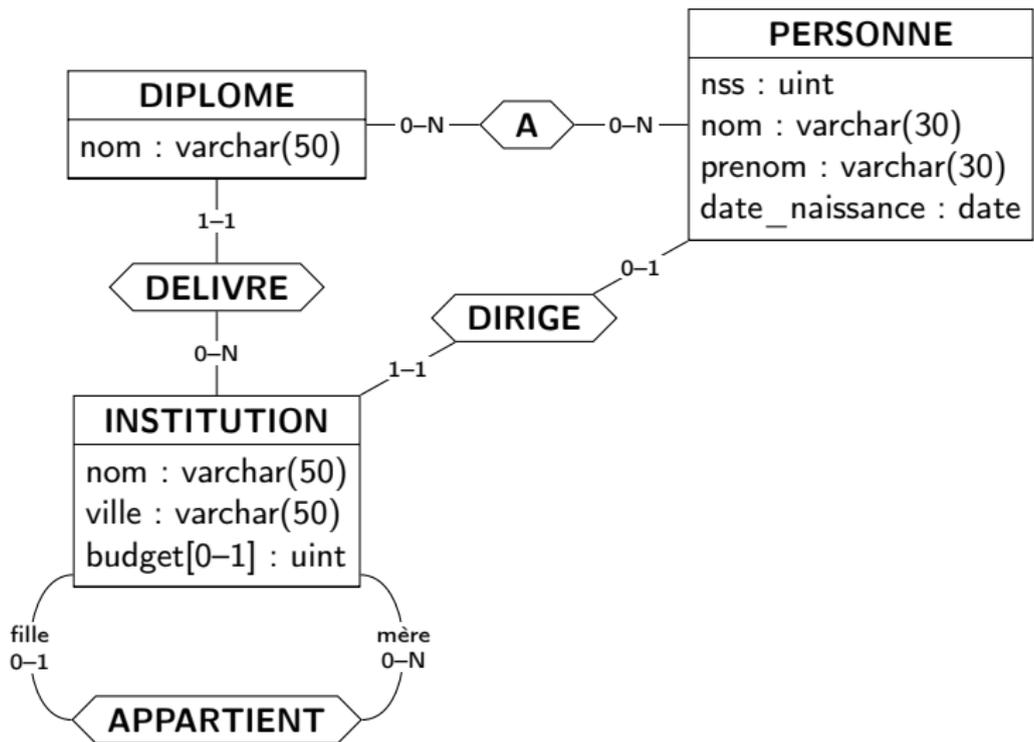


Cardinalité

- En plus de la classe fonctionnelle, on indique pour E1 et E2 si chaque entité concernée doit obligatoirement participer dans au moins une association de R ou non. Les cardinalités possibles sont donc : 0-1, 1-1, 0-N ou 1-N.
- Plus généralement, la cardinalité de E1 et E2 peut être tout intervalle $i-j$ où i est un entier naturel, j est $N (+\infty)$ ou un entier naturel plus grand que 0.

Types d'associations

Exemple



Définitions

- Chaque type d'entités a un ou plusieurs **identifiants**, chacun de ceux-ci identifiant de manière unique les entités de ce type.
- Étant donné un type d'entités E , un identifiant pour E est un **ensemble non vide minimal** pour l'inclusion :
 - d'attributs a_i de E ;
 - de rôles opposés $R_i.E_i$, où E_i est un type d'entités distinct de E associé à celui-ci via le type d'associations R_i , où $R_i.E$ a 0-1 ou 1-1 pour cardinalité ;

vérifiant que quelque soit la valuation donnée aux attributs a_i et les entités e_i de E_i , il existe au plus une entité de E ayant ces valeurs pour les attributs a_i et associée à chacune des entités e_i via R_i .

- On peut par exemple poser que :
 - $\{nss\}$ est un identifiant pour PERSONNE ;
 - $\{\text{nom, DELIVRE.INSTITUTION}\}$ est un identifiant pour DIPLOME.

- Les composants d'un identifiant peuvent soit être tous obligatoires, soit être tous facultatifs.
- L'un de ces identifiants est déclaré **primaire**, et doit nécessairement être obligatoire ; les autres sont déclarés **secondaires**.

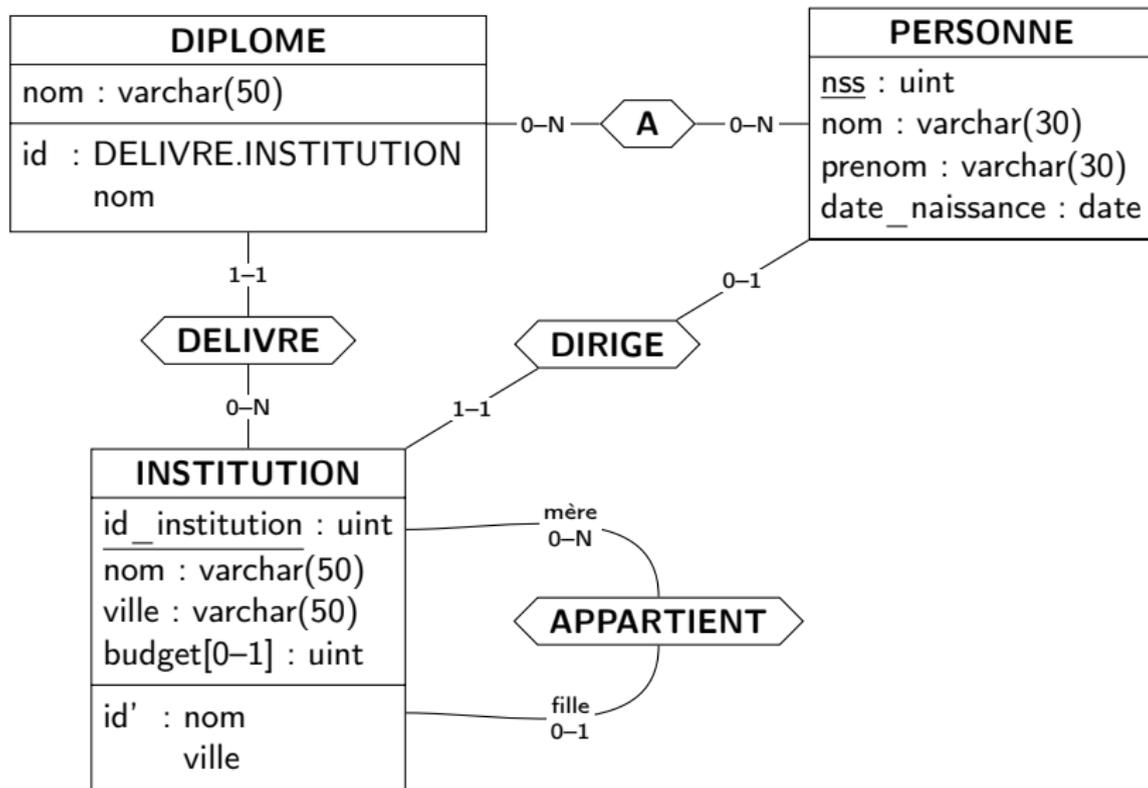
Représentation graphique

E
...
id : <liste_identifiant_primaire>
id' : <liste_identifiant_secondaire_1>
...
id' : <liste_identifiant_secondaire_k>

Lorsque l'identifiant primaire est composé uniquement d'attributs, ces derniers sont soulignés.

Identifiants

Exemple



Une **contrainte d'intégrité** est une propriété que les instances décrites par un schéma (les entités, les associations et les valeurs d'attributs) doivent respecter. Deux types :

- les **contraintes statiques** : propriétés devant être vérifiées à tout instant par les instances ;
- les **contraintes dynamiques** : propriétés devant être vérifiées lors de l'ajout, la suppression ou la modification d'une ou plusieurs instances.

Contraintes de base

Elles sont **déjà données** directement dans le schéma entité-association, à savoir :

- les identifiants ;
- les attributs obligatoires ;
- les contraintes de cardinalité ;
- le type de données des attributs.

Contraintes additionnelles

Elles sont données **en supplément** du schéma entité-association et sont exprimées de manière semi-formelle ou en langage courant. On peut trouver :

- des **contraintes de domaine** sur les valeurs données aux attributs (ex. : $\forall i \in \text{INSTITUTION}, 10000 \leq i.\text{budget} \leq 10000000000$);
- des **contraintes entre valeurs** d'attributs (ex. : $\forall (m, f) \in \text{APPARTIENT}, m.\text{id_institution} < f.\text{id_institution}$);
- des **contraintes d'existence** (ex. : $\forall m \in \text{INSTITUTION}, (\exists f \in \text{INSTITUTION t.q. } (m, f) \in \text{APPARTIENT}) \Rightarrow m.\text{budget} \neq \text{NULL}$);
- des **dépendances fonctionnelles**;
- etc.

Dépendances fonctionnelles

Soit un type d'entités E et deux ensemble A, B non vides

- d'attributs de E ;
- de rôles opposés $R_i.E_i$, où E_i est un type d'entités distinct de E associé à celui-ci via le type d'associations R_i , où $R_i.E$ a 0-1 ou 1-1 pour cardinalité.

On a la **dépendance fonctionnelle** $A \rightarrow B$ ssi pour toute entité de E , la valeur donnée aux attributs de A et les entités associées des rôles opposés de A déterminent de manière unique la valeur donnée aux attributs de B et les entités associées des rôles opposés de B .

Par exemple :

- dans DIPLOME, on a $\{\text{nom, DELIVRE.INSTITUTION}\} \rightarrow \{\text{nom}\}$ (identifiant) ;
- on pourrait imposer que dans INSTITUTION, $\{\text{nom}\} \rightarrow \{\text{budget}\}$.

Objectif

À partir d'une **spécification informelle** d'un certain univers, produire un **schéma EA** modélisant les **concepts** ainsi que les **faits** au sujet de ces concepts donnés par la spécification.

Procédure suggérée

La construction d'un schéma EA peut se faire de manière itérative au cours de la lecture de la spécification informelle, où chaque pas consiste à :

- **repérer** la prochaine **proposition élémentaire irréductible** de la forme $C_1 L_1 C_2 \cdots L_{k-1} C_k$ où les C_i sont des concepts et les L_i expriment des liens entre ces concepts ;
- **intégrer**, si ce n'est déjà le cas, le **fait exprimé** par l'ajout d'attributs, de types d'entités, de types d'associations, en restructurant éventuellement le schéma.

Contraintes d'intégrité

À ajouter ensuite. Deux sources possibles :

- celles données explicitement par la spécification informelle ;
- celles issues d'une analyse critique ou d'un certain choix de conception (nécessaire pour les contraintes de base).

Normalisation du schéma

- **Simplification du schéma** : éliminer les constructions inutilement compliquées.
- **Élimination des redondances internes** : pour chaque type d'entités E , les seules dépendances fonctionnelles $A \rightarrow B$ doivent être telles que A est un sur-ensemble d'un identifiant de E , autrement il faut décomposer.

Documentation du schéma

À chaque type d'entités, attribut et type d'associations est associée une **description** indiquant quel concept ou fait de la spécification informelle il représente, ou, à défaut, son utilité en regard de la spécification.

Complétude

Simplement **vérifier** que tous les concepts et faits de l'énoncé formel sont bien pris en compte et que que tout ce qui est nécessaire est présent dans le schéma (identifiants, types de données, cardinalité, etc.).