



Dépendances et Normalisation

Composé avec L^AT_EX, révision numéro 3298.

1

Relation universelle...

— relation unique : un tuple opère toutes les liaisons

titre	annee	pays	capitale	date	acteur	a-pays	realisateur	r-pays
Troy	2004	USA	Wash. DC	Q1 04	Pitt	USA	Petersen	Allemagne
Troy	2004	USA	Wash. DC	Q1 04	Bloom	USA	Petersen	Allemagne
Troy	2004	France	Paris	Q2 04	Pitt	USA	Petersen	Allemagne
Troy	2004	France	Paris	Q2 04	Bloom	USA	Petersen	Allemagne
Pirates	2003	USA	Wash. DC	Q2 03	Bloom	USA	Verbinski	USA
Pirates	2003	France	Paris	Q3 03	Bloom	USA	Verbinski	USA

Le film *Troy* (2004) est sorti en France (capitale Paris) en Q2 04 avec l'acteur Pitt originaire des USA et réalisé par Petersen originaire d'Allemagne.

2

Problème : redondance des informations

acteur - a-pays, pays - capitale, titre - réalisateur

Conséquences : risques d'incohérences

mises à jour changement de nationalité d'un réalisateur...

insertions ajout d'un même acteur dans une autre nationalité

suppressions disparition d'un acteur si aucun film

3

Comment éviter la redondance ?

séparation des concepts, dans des relations différentes

modélisation approche ascendante avec E/A, UML...

à partir des entités, liées par des associations

normalisation approche descendante théorique

— relation universelle initiale

— analyse des dépendances fonctionnelles entre attributs

— décomposition en relations jointes (coût joint, éco mém)

— définition de formes normales avec bonnes propriétés

4

Dépendances Fonctionnelles (DF)

dépendance fonctionnelle d'une relation R :

soit deux sous-ensemble d'attributs S et D de R

ils dépendent *fonctionnellement* l'un de l'autre, noté $S \rightarrow D$:

$\forall t_1 \in R, \forall t_2 \in R, \Pi_S(t_1) = \Pi_S(t_2) \Rightarrow \Pi_D(t_1) = \Pi_D(t_2)$

exemples titre pays \rightarrow sortie

titre \rightarrow année réalisateur r-pays

dépendance sémantique réelle

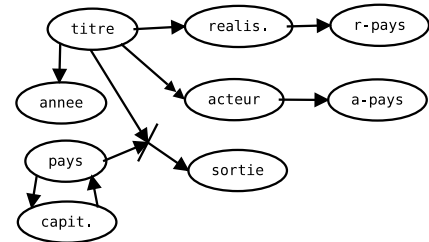
pas simplement constatées sur un jeu de données !

donc ne peuvent pas être dérivées des données !

exemple, on a pas : année \rightarrow titre

5

Exemple de dépendances (minimales) pour les films



— pas de DF entre un film et un acteur particulier

mais DMV (dépendance multivaluée) entre un film et ses acteurs

6

Propriétés des DF (axiomes d'Armstrong, 1974)

réflexivité auto-détermination des attributs

$X \subset S \Rightarrow S \rightarrow X$

e.g. titre année réalis. \rightarrow titre

augmentation ajout d'attributs

$\forall Z, X \rightarrow Y \Rightarrow XZ \rightarrow YZ$

e.g. titre \rightarrow année \Rightarrow titre nom \rightarrow année nom

transitivité des DF

$X \rightarrow Y \wedge Y \rightarrow Z \Rightarrow X \rightarrow Z$

e.g. titre \rightarrow pays \rightarrow capitale

7

Autres propriétés des DF

union de deux dépendances fonctionnelles

$(X \rightarrow Y \wedge X \rightarrow Z) \Rightarrow X \rightarrow YZ$

titre pays \rightarrow réalisateur \wedge titre pays \rightarrow sortie

\Rightarrow titre pays \rightarrow réalisateur sortie

preuve : augmentation et transitivité

pseudo-transitivité des DF

$(X \rightarrow Y \wedge WY \rightarrow Z) \Rightarrow WX \rightarrow Z$

capitale \rightarrow pays \wedge titre pays \rightarrow sortie

\Rightarrow titre capitale \rightarrow sortie

preuve : augmentation et transitivité

8

décomposition de DF

$$\forall Z \subset Y, X \rightarrow Y \Rightarrow X \rightarrow Z$$

titre pays \rightarrow réalisateur sortie \Rightarrow titre pays \rightarrow réalisateur

preuve : réflexivité et transitivité

Conséquence de l'union et de la décomposition

— on peut toujours proposer des DF vers un singleton

— $X \rightarrow A_1 \dots A_n \quad X \rightarrow A_1, \dots, X \rightarrow A_n$

9

DF élémentaire**définition** $S \rightarrow D$ avec S le plus petit pour D S tel que $D \cap S = \emptyset$ et $\forall X [X \subset S \wedge X \rightarrow D \Rightarrow X = S$ *plus petite clef d'un tuple, séparation des attributs* S attributs sources, D attributs déduits**remarque** pas de solution unique à partir de $D \cup S$

e.g. titre année pays sortie réalisateur capitale

titre pays \rightarrow année capitale réalisateurtitre capitale \rightarrow année pays réalisateur

10

Fermeture transitive des DF d'une relation**définition** F^+ déduit de F par réflexivité, augmentation et transitivité...*construction itérative de toutes les DF possibles***La fermeture transitive définit toutes les DF vraies**

— validité des DF dérivées par examen des trois règles

— complétude des DF dérivées...

11

6 Formes Normales hiérarchisées**1NF** attributs atomiques**2NF** 1NF + attributs non clef en DF de la clef entière**3NF** 2NF + attributs non clef en dépendances directe des clefs**Boyce-Codd NF/3.5NF** 3NF + tous les déterminants sont des clefs**4NF** BCNF + sans dépendances multivaluées non-triviales**5NF** 4NF + ...

— les NF mesurent divers degrés de redondance...

— en pratique, on vise BCNF ou 4NF

— moins si motivé par performances : éviter des jointures

12

1NF : première forme normale— **attributs atomiques** (pas de tableaux ou d'ensembles)

— lié à une modélisation insuffisante (attribut vs entité)

— tableaux (Access, PostgreSQL) à éviter...

Cine(*titre,acteurs...*)

titre	acteurs
Troy	{Pitt,Bloom}
Pirates	{Bloom}

— requêtes non relationnelles avec tableaux... ou texte...

relation dont l'attribut est le nom d'un acteur ?— solution : Joue(*titre,nom*)

13

2NF : seconde forme normale— 1NF + **attributs non clef en DF de la clef entière***non clef : attribut n'appartenant à aucune des clefs*Cine(*titre,nom, a-pays*) est 1NF

titre	nom	a-pays
Troy	Pitt	USA
Troy	Bloom	USA
Pirates	Bloom	USA

— problème : redondance du *a-pays* qui ne dépend que du *nom*— solution : Film(*titre,nom*), Acteur(*nom,a-pays*)

14

3NF : troisième forme normale (Codd, 1970)— 2NF + **attributs non clefs en dépendances directe des clefs***pas de transitivité des DF pour les attributs non clefs !*Cine(*nom,a-pays,capitale*) est 2NF

nom	a-pays	capitale
Pitt	USA	Wash. DC
Bloom	USA	Wash. DC
Petersen	Allemagne	Berlin
Verbinski	USA	Wash. DC

— problème : lien *pays-capitale* répété pour chaque *nom*— solution : Acteur(*nom,a-pays*), Pays(*pays,capitale*)

15

3.5/BCNF : forme normale de Boyce-Codd (Heath 1971...)— 3NF + **tous les déterminants sont des clefs potentielles**— motif $AB \rightarrow C, C \rightarrow B$ — *cinéma d'un type le plus proche*Proche(*personne,type,lieu*)Proche(*personne,type,lieu*)

est 3NF

prénom	type	plus proche
Calvin	Cinéma	Le Club
Calvin	Multiplex	Alticiné
Hobbes	Cinéma	Le Club
Hobbes	Multiplex	Alticiné
Susie	Cinéma	Studio des Ursulines
Susie	Multiplex	MK2 Odéon

— **pas** solution : Proche(*personne,lieu*) + Lieu(*lieu, type*)— solution partielle : Proche(*personne,type,lieu*) + Lieu + contrainte ?

16

Décomposition d'une relation

- regroupement des attributs en sous relations
- schéma d'une relation $R = \{A_1, \dots, A_n\}$
- vers un ensemble de schémas : $R = R_1 \cup \dots \cup R_k$
- attributs communs entre les relations !

Jointure sans perte d'information vis à vis de dépendances

- r relation de R , R décomposé,
- $r = \pi_{R_1}(r) \bowtie \dots \bowtie \pi_{R_k}(r)$

17

Algorithme de Tsou et Fisher (1982) : DF vers BCNF

- suppose d'avoir le graphe des DF...
- nécessite donc une modélisation précise et juste !

lemme toute relation a 2 attributs est BCNF, et

si R non BCNF, $\exists AB \in R : (R - AB) \rightarrow A$

algorithme décomposition récursive

18

Dépendances multivaluées (DMV)

- notées $X \twoheadrightarrow Y$
- plusieurs valeurs possible d'un attribut dans une clef (1 pour n)
- ce n'est pas une *dépendance fonctionnelle* (1 pour 1)

Générique(*titre,acteur*)

titre	acteur
Troy	Pitt
Troy	Bloom
Pirates	Bloom

- exemple : film \twoheadrightarrow acteur
- l'ensemble des acteurs dépend du film

19

4NF : quatrième forme normale (Fagin, 1977)

- BCNF + **sans dépendances multivaluées non-triviales**
- triviale : $X \twoheadrightarrow Y \Rightarrow X$ est une clef

Générique(*titre,acteur,réalisateur*)

est BCNF

titre	acteur	réalisateur
Troy	Pitt	Petersen
Troy	Bloom	Petersen
Pirates	Bloom	Verbinski

- problèmes : relations indépendantes liés dans une même relation
- pas de film sans acteur, réalisateur répété...*
- pas de lien direct entre réalisateur et acteur
- solution : Joue(*titre,acteur*), Réalise(*titre,réalisateur*)

20

5NF : cinquième forme normale ?

- pas de décomposition possible d'une relation ?
- similaire à 4NF, mais décomposition en 3 relations ?
- exemples un peu tordus...

Dépendances de jointure (Rissanen 1979)

Dépendances algébriques (Yannakakis et Papadimitriou 1980)

Dépendances généralisées (Sadri et Ullman 1981)

Dépendances implicationnelles (Fagin 1982)

...

21

Conseil de lecture

- Jeffrey D. Ullman
- Principles of Database and Knowledge-Base Systems
- Volume 1
- Chapitre 7
- Computer Science Press, 1988

22

List of Slides

- 1 Dépendances et Normalisation
- 2 Relation universelle...
- 3 Problème : redondance des informations
- 3 Conséquences : risques d'incohérences
- 4 Comment éviter la redondance ?
- 5 Dépendances Fonctionnelles (DF)
- 6 Exemple de dépendances (minimales) pour les films
- 7 Propriétés des DF (axiomes d'Armstrong, 1974)
- 8 Autres propriétés des DF
- 9 Conséquence de l'union et de la décomposition
- 10 DF élémentaire

Dénormalisation ?

- normalisation réduit les redondances
- facilite insertion, mise à jour et effacement
- mais impact sur les performances de consultation
- rétablissement des liens par jointures
- optimisation : utilisation d'index...
- problème de l'évolution du schéma
- ajout a posteriori de concepts
- en général, pas de reprise de la modélisation
- dénormalisation parfois utile pour performances
- systèmes distribués, SQL ou NoSQL...

23

- 11 Fermeture transitive des DF d'une relation
- 11 La fermeture transitive définit toutes les DF vraies
- 12 6 Formes Normales hiérarchisées
- 13 1NF : première forme normale
- 14 2NF : seconde forme normale
- 15 3NF : troisième forme normale (Codd, 1970)
- 16 3.5/BCNF : forme normale de Boyce-Codd (Heath 1971...)
- 17 Décomposition d'une relation
- 17 Jointure sans perte d'information vis à vis de dépendances
- 18 Algorithme de Tsou et Fisher (1982) : DF vers BCNF
- 19 Dépendances multivaluées (DMV)
- 20 4NF : quatrième forme normale (Fagin, 1977)
- 21 5NF : cinquième forme normale ?
- 21 Dépendances de jointure (Rissanen 1979)
- 21 Dépendances algébriques (Yannakakis et Papadimitriou 1980)
- 21 Dépendances généralisées (Sadri et Ullman 1981)
- 21 Dépendances implicationnelles (Fagin 1982)
- 21 ...
- 22 Conseil de lecture
- 23 Dénormalisation ?