Lifbdw2 – Bases de données avancées
TD4 – énumération, Relation d'Armstrong

Licence informatique – Automne 2019-2020

Les questions marquées du symbole (†) sont à préparer pour la séance

Algorithme 1: calcul d'une relation d'Armstrong

```
Data : R un schéma de relation, F un ensemble de DF sur R.
Result : Une relation d'Armstrong r pour F.
for A ∈ R do
    t[A] := 0
r := {t}
i := 1
for X ∈ Cl(F) \ R do
    for A ∈ R do
        if A ∈ X then
            t[A] := 0
        else
            t[A] := i
    r := r ∪ {t}
    i := i + 1
return r
```

Exercice 1 : dépendances fonctionnelles satisfaites par une instance

Soit la relation de la table 1 dans laquelle ont indique pour chaque employé sont numéro, le numéro de son département, l'année d'entrée de l'employé dans le département, le nom du département en question et le numéro du responsable.

1. Écrire la liste de toutes les DF que vous pouvez trouver. Proposer une méthode pour ne pas en oublier.

	A	B	C	D	E
t_1	1	1	85	Biochimie	5
t_2	1	5	94	Admission	12
t_3	2	2	92	Informatique	2
t_4	3	2	98	Informatique	2
t_5	4	3	98	Géophysique	2
t_6	5	1	75	Biochimie	5
t_7	6	5	88	Admission	12

Table 1 – Affectation des employés à un département

2. Indiquer à votre avis quelles sont les DFs précédentes qui n'ont pas d'interprétation naturelle vis-à-vis des données.

Exercice 2 : relation d'Armstrong

Soit un schéma de relation R et un ensemble F de DF sur ce schéma. Une *relation d'Armstrong* pour F est une relation r sur R vérifiant *exactement* F^+. Soit le schéma $R = ABCDE$ et l'ensemble de DF $F = \{A \rightarrow BC, D \rightarrow E, C \rightarrow D\}$ pour la suite.

1. Calculer l'ensemble des fermés de F, défini par $Cl(F) = \{X^+ \mid X \subseteq R\}$. Pour cela, utilisez une exploration systématique des sous-ensembles de R en commençant par ceux de taille 1, puis 2 ... Penser à ne pas visiter les sur-ensembles des clés pour limiter le nombre de calculs.
2. Utiliser l'algorithme 1 pour calculer une relation d'Armstrong r.
3. Supposons que l'on supprime le tuple correspondant au fermé BE de la relation r obtenue à la question précédente. Il y a-t-il des DFs nouvellement vérifiées qui ne l'étaient pas ?
4. Même question que précédemment avec le fermé DE.
5. Quand $F = \emptyset$, combien il y a-t-il de tuples dans la relation d'Armstrong ? Même question quand F est l'ensemble de toutes les DFs possibles sur R ?

Corrections

Solution de l'exercice 1

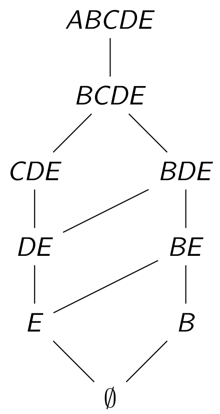
1. Utiliser une approche par niveaux croissants sur les parties gauches. Pour trouver quels sont les attributs qui dépendent de l'ensemble d'attributs X, on va regarder toutes les couples (s, t) de tuples tels que $s[X] = t[X]$ et on élimine tous les attributs qui n'ont pas de valeurs identiques pour au moins un couple. Par exemple, pour $X = A$, on a $t_1[A] = t_2[A]$ mais tous les autres attributs sont différents, on a donc seulement $A \to A$.
   On commence par les attributs seuls, ensuite, on va regarder toutes les combinaisons possibles (sans ordre entre les attributs). Au niveau 3, on ne considère que les parties gauches dont aucun sous-ensemble n'est une clé, le seul à considérer est donc BDE. Il n'y a pas à aller plus loin car en ajoutant soit A soit C à BDE on obtient une clé.

partie gauche	fermeture
A	A
B	BDE
C	CE
D	DBE
E	E
AB	ABCDE
AC	ABCDE
AD	ABCDE
AE	ABCDE
BC	ABCDE
BD	BDE
BE	BDE
CD	ABCDE
CE	CE
DE	BDE
BDE	BDE

2. On a $B \to DE$ et $D \to BE$, autrement dit on peut identifier un département par son nom ou son numéro et connaître l'un ou l'autre permet de déterminer qui est le reponsable du département. Que AB et AD soient clés est donc logique. En revanche, les clés faisant intervenir C ou E en partie gauche sont douteuses. Typiquement $C \to E$ n'est pas sensée, c'est juste un effet de bord dû au fait qu'informatique et géophysique aient le même directeur.

Solution de l'exercice 2

1. On trouve les fermés suivants : $ABCDE$, $BCDE$, B, CDE, DE, E, BDE, BE. Muni de l'inclusion, on peut représenter graphiquement cet ensemble ordonné par son diagramme de Hasse :



2. On trouvera, en les considérant dans l'ordre où les fermés sont donnés :

fermé	A	B	C	D	E
$ABCDE$	0	0	0	0	0
$BCDE$	1	0	0	0	0
B	2	0	2	2	2
CDE	3	3	0	0	0
DE	4	4	4	0	0
E	5	5	5	5	0
BDE	6	0	6	0	0
BE	7	0	7	7	0

3. La relation à deux tuples composée de $\langle 7, 0, 7, 7, 0\rangle$ associé à BE et de $\langle 0, 0, 0, 0, 0\rangle$ associé à $ABCDE$ falsifie les DFs de la forme $BE \rightarrow X$ avec $X \in ACD$. On peut voir que parmi celles-ci c'est la DF $BE \rightarrow D$ qui n'a plus de contre exemple si on supprime $\langle 7, 0, 7, 7, 0\rangle$. Les DFs $BE \rightarrow A$, $BE \rightarrow C$ restent quant à elles encore violées grâce à $\langle 6, 0, 6, 0, 0\rangle$ associé à BDE.
4. Dans ce cas là, il n'y a aucun contre-exemple qui est perdu. En effet le tuple de DE permet d'invalider $DE \rightarrow A$, $DE \rightarrow B$ et $DE \rightarrow C$ mais chacune de ces DFs trouve encore un contre-exemple soit via le tuple associé à CDE soit via celui de BDE. En fait, la méthode utilisée pour générer une relation d'Armstrong produit des relations avec un nombre maximal de tuples. Pour avoir un nombre minimal il ne faut pas considérer la famille de Moore $\{ABCDE, BCDE, B, CDE, DE, E, BDE, BE\}$ mais seulement son sous-ensemble $\{ABCDE, BCDE, B, CDE, BDE, BE\}$ qui l'engendre par intersection (théorème 4.1 de *On the Structure of Armstrong Relations for Functional Dependencies*).
5. Si $F = \emptyset$ alors chaque $X \subseteq R$ est fermé et il y a ainsi $2^{|R|}$ tuples dans la relation d'Armstrong. A l'inverse, si F contient toutes les DFs possibles, alors il n'y a qu'un seul fermé $ABCDE$ et la relation obtenue ne contient qu'un seul tuple.