Matrise Informatique 2001
Bases de Données 2 Feuille de Partiel n° 1 (Correction)

Partiel de BD II, 24 avril 2001

La durée du partiel est de deux heures. Tout document manuscrit (notes de cours ou de TD de BD2 de cette année) est autorisé. Le barème est donné à titre indicatif, et il est susceptible d'être modifié. Tout résultat déjà établi en cours et/ou TD peut être cité sans être redémontré. Les questions sont indépendantes les unes des autres.

S'il vous plaît, rédigez les trois premiers exercices (1-3) et les trois autres (4-6) sur deux copies séparées : **un point de plus sera donné à toute personne respectant cette indication !**

Exercice 1 (4 points)

Soit le schéma de base de données:
REPRESENTATION(Opéra, Ville, Anne)
OEUVRES(Opra, Auteur).
Formuler chacune des requêtes suivantes en calcul relationnel à variable domaine et en calcul relationnel à variable n-uplet.

1. Quels sont les auteurs dont certaines œuvres ont été jouées à Paris en 1996 ?
2. Quelles sont les villes où il n'y a eu aucune représentation d'un opéra de Verdi en 1999 ?
3. Quelles sont les villes où ont été représentés tous les opéras de Wagner ?
4. En quelle année a-t-on représenté à Paris au moins deux opéras différents de Mozart ?

Correction :

1. *En* calcul relationnel à variable domaine*:*

$$\{a \ \mid \exists o\,(OEUVRES(o,a) \wedge REPRESENTATION(o,Paris,96))\}$$

2. *En* calcul relationnel à variable domaine*:*

$\{v \ \mid \exists o\, \exists a\, REPRESENTATION(o,v,a) \wedge$
$\neg\, \exists o'(REPRESENTATION(o',v,96) \wedge OEUVRES(o',Verdi))\}$

NB : la sous-formule $\exists o \ \exists a \ REPRESENTATION(o,v,a)$, qui borne v à être une ville, est nécessaire, car sinon tout v qui n'est pas une ville fera partie de la réponse.

3. *En* calcul relationnel à variable domaine*:*

$\{v \ \mid \ \exists o\, \exists a\, REPRESENTATION(o,v,a) \wedge$
$\forall o' \ (OEUVRES(o',Wagner) \ \rightarrow \ \exists a' \ (o',v,a'))\}$

NB : la sous-formule $\exists o \ \exists a \ REPRESENTATION(o,v,a)$, qui borne v à être une ville, est la pour gérer aussi le cas où Wagner n'aurait écrit aucun opéra (que ce cas ne soit par réaliste, n'a aucune importance). Dans ce cas la, en absence de cette formule, tout *v ferait partie de la réponse, même si ce n'est pas une ville.*

4. *A ècrire.*

Exercice 2 (3 points)

Considérez la propriété[1] : $\sigma_\psi(E_1 \bowtie_\phi E_2) \ \equiv \sigma_\psi(E_1) \bowtie_\phi E_2$

[1] Par définition, la notation $E_1 \bowtie_\phi E_2$ est une formulation abrégée de $\sigma_\phi(E_1 \times E_2)$.

1. Donner un exemple montrant que cette propriété n'est pas vraie, dans le cas général.
2. Indiquer des conditions sur ϕ et ψ suffisantes pour assurer la validité de la propriété. Vos conditions ne doivent pas être *triviales* : par exemple, n'indiquez pas la condition $\psi = \phi = True$! Justifiez très rapidement votre réponse.
3. En général, laquelle des deux expressions respectivement à gauche et à droite de $\equiv$ est la moins coûteuse à évaluer, en termes de la somme des tailles des relations intermédiaires ? Justifiez très rapidement votre réponse.

Exercice 3 (6 points)

Rappel et Indication: Etant donnée une requête R du calcul relationnel à variable domaine, R est indépendante du domaine si et seulement si la réponse à R ne dépend que des valeurs des relations et des constantes qui apparaissent dans R. En d'autres termes, même si on fait varier le domaine, la réponse reste la même une fois qu'on a fixé les valeurs des relations et des constantes qui apparaissent dans R.

Soit R une relation unaire de schéma A, S une relation de schéma AD, et T une relation de schéma EFG. Pour chacune des requêtes du calcul relationnel à variable domaine suivantes, calculer leurs formes normales saines R_1, R_2 et R_3. Pour chacune des requêtes R_1, R_2 et R_3, indiquer si elle saine ou non, si elle est indépendante du domaine ou non. Si elle est saine, la réécrire en algèbre. Si elle n'est pas indépendante du domaine, exhibez un exemple permettant de le montrer. Si, par contre, elle est indépendante du domaine, justifiez brièvement votre réponse.

1. $\{< x, y, u > \ \mid \ \exists z\, T(x, y, z) \wedge \exists v[(R(u) \vee S(u, v)) \wedge R(v)]\}$
2. $\{< y > \ \mid \ \forall x\ (R(x) \to (R(x) \wedge P(y)))\}$
3. $\{< x, y > \ \mid \ \exists z\, T(x, y, z) \wedge \exists u, v([S(u, v) \wedge R(v)] \to \exists w\ T(w, w, v))\}$

Correction :

1. *R_1 est la requête elle même. Elle est saine, et, par conséquent, indépendante du domaine (résultat vu en cours). Algèbre :*$\pi_{EF}(T) \times \pi_A((R \times R) \cup (S \bowtie_{A=D} R))$,
2. *R_2 est*
$$\{< y > \ \mid \ \neg\exists x\ (R(x) \wedge (\neg R(x) \vee \neg P(y)))\}$$
Elle n'est pas saine. Voici un exemple montrant qu'elle n'est pas indépendante du domaine :
Soit une instance de base telle que R est vide, P est $\{a\}$. Si le domaine est $\{a\}$, alors la réponse est $\{a\}$, si le domaine est $\{a, b\}$, alors la réponse est $\{a, b\}$.
3. *en FNS, cela donne $\exists\ z\ T(x, y, z) \wedge \exists\ u, v\ [\neg S(u, v) \vee \neg R(v) \vee \exists\ w\ \ (T(x, w, v) \wedge T(u, v, y))]$.La formule n'est pas saine (on dérive $\perp$ dès qu'on traverse $\exists\ u, v$). Elle est dépendante du domaine. Par exemple, appelons R_2 la formule $\exists\ u, v\ [\neg S(u, v) \vee \neg R(v) \vee \exists\ w\ \ (T(x, w, v) \wedge T(u, v, y))]$. Cette formule est vraie si uniquement $\exists\ u, v\ [\neg S(u, v)$ est vraie, et dans ce cas, u et v peuvent prendre n'importe quel élément du domaine pour valeur.*

Exercice 4 (3 points)

Soit U l'ensemble d'attributs $\{A, B, C, D\}$ et soit d la décomposition $\{AB, ACD, BCD\}$ de U. Soit F l'ensemble de dépendances fonctionnelles : $\{A \to C, D \to C, BD \to A\}$.

- Quel est le résultat de l'algorithme de poursuite pour d et F ?
- Si ceci est possible, donner un exemple de relation r qui montre que d n'est pas SPI par rapport à F et proposer une nouvelle décomposition d' de U qui soit SPI par rapport à F. Sinon, expliquer rapidement pourquoi une telle r ne peut pas exister.

Exercice 5 (3 points)

Soit ARM_1 la formulation suivante du système de preuve d'Armstrong pour les dépendances fonctionnelles :

- Toute dépendance $X \to Y$ telle $Y \subseteq X$ est un *axiome*.
- On a deux règles d'inférence :
    1. Si on a $X \to Y$, alors on peut dériver $XZ \to YZ$ (*Augmentation*)
    2. Si on a $X \to Y$ et $Y \to Z$, alors on peut dériver $X \to Z$ (*Transitivité*)

Soit ARM_2 la formulation alternative du système d'Armstrong :

- Toute dépendance $X \to Y$ telle $Y \subseteq X$ est un *axiome*
- On a une seule règle d'inférence : si on a $X \to Y$ et si on a $YW \to Z$ alors on peut dériver $XW \to Z$ (*Pseudo Transitivité*)

Démontrer que les deux formulations sont équivalentes. Vous pouvez raisonner syntaxiquement ou sémantiquement, comme vous voulez, à condition d'être concis.

Exercice 6 (3 points)

Démontrer la règle d'union pour les dépendances multivaluées :
Si X, Y et Z sont des ensembles d'attributs et si X $\to\to$ Y et X $\to\to$ Z, alors X $\to\to$ Y $\cup$ Z.