

Bases de données II, Charleroi

Jef Wijsen

18 juin 2008

Répondez aux onze questions dans les espaces réservés. Durée : 2 heures et 30 minutes

Nom et prénom

Année

Question 1 (5 points) À la page 380.

For asymmetric binary data, measures that do not remain invariant under the inversion operator are preferred.

Qu'est-ce que *asymmetric binary data*? Donnez un exemple concret.

Question 2 (5 points) Exercice 2 (c) à la page 198.

Compute the Gini index for the Gender attribute.

Question 3 (5 points) Expliquez l'**utilité pratique** de la formule de Lance-Williams (il ne faut donc **pas expliquer la formule** elle-même).

Question 4 (5 points) Regardez la Figure 6.13 (b) *Number of frequent itemsets*, à la page 347. On pourrait penser que des *itemsets* fréquents de large taille sont plus rares. Néanmoins, la fonction pour le support = 0.1% atteint un maximum pour la taille = 10. Savez-vous expliquer ce comportement?

Question 5 (5 points) Considérez la base de transactions suivante:

<i>tid</i>	<i>items</i>
1	{ <i>a, b, c, d, e, f</i> }
2	{ <i>a, b, c, d, e</i> }
3	{ <i>a, d</i> }
4	{ <i>b, d, f</i> }
5	{ <i>a, b, c, e, f</i> }

Dessinez le *conditional FP-tree* for *f*, en utilisant un ordre dans lequel *f* est le dernier élément.

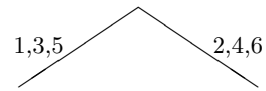
Question 6 (5 points) Exercice 12 à la page 562 présente l'algorithme du *leader*.

The leader algorithm represents each cluster using a point, known as a leader, and assigns each point to the cluster corresponding to the closest leader, unless this distance is above a user-specified threshold. In that case, the point becomes the leader of a new cluster.

Discutez **une faiblesse** de cet algorithme par rapport à K-means (limitez-vous à une seule faiblesse, même si vous en voyez plusieurs).

Question 7 (10 points) Faites l'exercice 10 à la page 408. Noter:

- Les candidats: $\{1, 2, 3\}, \{1, 2, 6\}, \{1, 3, 4\}, \{2, 3, 4\}, \{2, 4, 5\}, \{3, 4, 6\}, \{4, 5, 6\}$
- La fonction de hachage:



- Le nombre de candidats par nœud est ≤ 2 .

Le DTD suivant exprime que chaque pièce peut être composée de zéro ou plusieurs autres pièces. On peut supposer que chaque pièce a un nom qui est unique dans le document XML.

```
<!-- This file is called parts.dtd -->
<!ELEMENT composition (part)*>
<!ELEMENT part (part)*>
<!ATTLIST part name CDATA #REQUIRED>
<!ATTLIST part manufacturer CDATA #REQUIRED>
```

Les questions qui suivent concernent le document valide qui suit :

```
<?xml version="1.0"?>
<!DOCTYPE composition SYSTEM "parts.dtd">
<composition>
  <part name="voiture" manufacturer="renault">
    <part name="moteur" manufacturer="renault">
      <part name="bloc-moteur" manufacturer="renault"/>
      <part name="boite-de-vitesses" manufacturer="shimano"/>
      <part name="systeme-de-refroidissement" manufacturer="bosch"/>
    </part>
    <part name="roues" manufacturer="renault">
      <part name="jante" manufacturer="renault"/>
      <part name="pneu" manufacturer="michelin">
        <part name="soupape" manufacturer="bosch"/>
      </part>
    </part>
  </part>
  <part name="ordinateur" manufacturer="dell">
    <part name="processeur" manufacturer="intel"/>
    <part name="ecran" manufacturer="iiyama"/>
  </part>
</composition>
```

Question 8 (5 points) Écrivez une expression XPath qui rend tous les nœuds attributs “name” liés à une pièce fabriquée par renault. Il y en a cinq : name="voiture", name="moteur", name="bloc-moteur", name="roues", name="jante".

Question 9 (5 points) Écrivez une expression XPath qui rend tous les nœuds attributs “name” liés à une pièce qui ne contient aucune pièce fabriquée par michelin. Il y en a neuf : name="moteur", name="bloc-moteur", name="boite-de-vitesses", name="systeme-de-refroidissement", name="jante", name="soupape", name="ordinateur", name="processeur", name="ecran". Notez que name="voiture" n'est pas dans cette liste, parce qu'une voiture contient la “sous sous pièce” pneu fabriquée par michelin.

Question 10 (5 points) Écrivez une expression XPath qui rend tous les nœuds attributs “name” liés à une pièce qui elle-même n'est pas fabriquée par renault, mais qui est utilisée dans une pièce fabriquée par renault. Il y en a quatre : name="boite-de-vitesses", name="systeme-de-refroidissement", name="pneu", name="soupape".

Question 11 (10 points)

Donnez une feuille XSLT qui copie les nœuds élément, en remplaçant chaque balise `part` par le nom de la pièce tel qu'indiqué par l'attribut `name`, tout en supprimant les autres attributs. Voici le résultat désiré :

```
<composition-traduite>
  <voiture>
    <moteur>
      <bloc-moteur />
      <boite-de-vitesses />
      <systeme-de-refroidissement />
    </moteur>
    <roues>
      <jante />
      <pneu>
        <soupape />
      </pneu>
    </roues>
  </voiture>
  <ordinateur>
    <processeur />
    <ecran />
  </ordinateur>
</composition-traduite>
```

Aide : Voici deux façons pour créer un élément de type `mylabel` en XSLT :

<code><mylabel></code>		<code><xsl:element name="mylabel"></code>
<code><!-- Creation du contenu --></code>	a le même	<code><!-- Creation du contenu --></code>
<code></mylabel></code>	effet que	<code></xsl:element></code>

Néanmoins, l'usage de `xsl:element` permet la création d'un élément dont le nom est calculé à l'aide d'une expression, à condition de l'encadrer par des accolades (`{}`). Par exemple,

```
<xsl:element name="{@name}">
  <!-- Creation du contenu -->
</xsl:element>
```