

Bases de Données II, Charleroi

Jef Wijsen

20 janvier 2011

Partie XML. Durée : 1 heure

Nom et prénom
Année

La figure 2 montre une base de données XML concernant des propriétaires de voitures. Chaque voiture (balise `car`) est identifiée de façon unique par son modèle (balise `model`, avec valeurs *323i*, *a4*, *q7*,...). La figure 1 montre le DTD.

```
<!-- This file is called cars.dtd -->
<!ELEMENT carsdb (car*, customer*)>
<!ELEMENT car (make, model, category, price)>
<!ELEMENT customer (owns*)>
<!ELEMENT owns (model, color, marketValue)>
<!ELEMENT make (#PCDATA)>
<!ELEMENT model (#PCDATA)>
<!ELEMENT category (#PCDATA)>
<!ELEMENT price (#PCDATA)>
<!ELEMENT color (#PCDATA)>
<!ELEMENT marketValue (#PCDATA)>
<!ATTLIST car country CDATA #REQUIRED>
<!ATTLIST customer cid CDATA #REQUIRED>
<!ATTLIST customer cname CDATA #REQUIRED>
<!ATTLIST customer age CDATA #REQUIRED>
```

Figure 1: DTD.

```

<?xml version="1.0"?>
<!DOCTYPE carsdb SYSTEM "cars.dtd">
<carsdb>
  <car country="germany"><make>bmw</make>
    <model>323i</model>
    <category>sport</category><price>25000</price></car>
  <car country="germany"><make>audi</make>
    <model>a4</model>
    <category>sport</category><price>30000</price></car>
  <car country="germany"><make>audi</make>
    <model>q7</model>
    <category>suv</category><price>10000</price></car>
  <car country="germany"><make>mercedes</make>
    <model>ml</model>
    <category>suv</category><price>80000</price></car>
  <car country="england"><make>aston martin</make>
    <model>bb9</model>
    <category>sport</category><price>100000</price></car>
  <car country="japan"><make>subaru</make>
    <model>outback</model>
    <category>wagon</category><price>15000</price></car>
  <car country="usa"><make>ford</make>
    <model>f100</model>
    <category>suv</category><price>10000</price></car>
  <customer cid="1" cname="john" age="20">
    <owns><model>a4</model>
      <color>white</color><marketValue>30000</marketValue></owns>
    <owns><model>q7</model>
      <color>black</color><marketValue>40000</marketValue></owns>
    <owns><model>323i</model>
      <color>red</color><marketValue>20000</marketValue></owns></customer>
  <customer cid="2" cname="mary" age="18">
    <owns><model>q7</model>
      <color>pink</color><marketValue>40000</marketValue></owns></customer>
  <customer cid="3" cname="jane" age="28">
    <owns><model>a4</model>
      <color>silver</color><marketValue>30000</marketValue></owns></customer>
  <customer cid="4" cname="ann" age="40" />
  <customer cid="5" cname="joyce" age="33">
    <owns><model>f100</model>
      <color>green</color><marketValue>25000</marketValue></owns>
    <owns><model>outback</model>
      <color>green</color><marketValue>10000</marketValue></owns></customer>
  <customer cid="6" cname="terry" age="25">
    <owns><model>ml</model>
      <color>white</color><marketValue>50000</marketValue></owns></customer>
  <customer cid="7" cname="claire" age="80">
    <owns><model>ml</model>
      <color>brown</color><marketValue>40000</marketValue></owns></customer>
  <customer cid="8" cname="bob" age="60">
    <owns><model>bb9</model>
      <color>yellow</color><marketValue>10000</marketValue></owns>
    <owns><model>ml</model>
      <color>brown</color><marketValue>20000</marketValue></owns>
    <owns><model>outback</model>
      <color>brown</color><marketValue>10000</marketValue></owns></customer>
</carsdb>

```

Figure 2: Fichier XML avec des informations sur des propriétaires de voitures.

Question 1 Disons que la *richesse* d'une personne est la somme des valeurs marchandes (`marketValue`) de ses voitures. Par exemple, la richesse de *joyce* est égale à $25000 + 10000 = 35000$. Disons qu'une personne est *riche* si sa richesse est supérieure à 45000. Écrivez une expression XPath qui rend chaque nœud de type `attribute` dont la valeur est le nom d'un client riche. Pour le document de la figure 2, *john* et *terry* sont riches.

.../5

Question 2 Écrivez une expression XPath **qui ne fait pas appel à la fonction count** et qui rend les catégories de voiture qui contiennent au moins deux modèles. Pour le document de la figure 2, les catégories *sport* et *suv* contiennent plus d'un modèle.

.../5

Question 3 Écrivez une expression XPath qui rend chaque nœud de type `attribute` dont la valeur est le nom d'un client qui possède une voiture de sport. Pour le document de la figure 2, *john*, *jane* et *bob* possèdent une voiture de sport.

.../5

Question 4 Écrivez une expression XPath qui rend chaque nœud de type `attribute` dont la valeur est le nom d'un client qui ne possède aucune voiture allemande. Pour le document de la figure 2, *ann* et *joyce* ne possèdent pas de voiture allemande.

.../5

Question 5 Écrivez un programme XSLT qui affiche pour chaque catégorie les noms des personnes qui possèdent une voiture de cette catégorie. Le résultat doit être formaté comme suit :

```
<?xml version="1.0"?>
<answer><sport>
  <name>john</name><name>jane</name><name>bob</name>
</sport>
<wagon>
  <name>joyce</name><name>bob</name>
</wagon>
<suv>
  <name>john</name><name>mary</name><name>joyce</name>
  <name>terry</name><name>claire</name><name>bob</name>
</suv>
</answer>
```

.../10

Partie Data mining. Durée : 2 heures

Nom et prénom

Année

Situez chaque terme dans le cursus et expliquez de façon succincte mais précise.

Question 6 Leave-one-out.

.../5

Question 7 Fuzzy clustering.

.../5

Question 8 Ward's method.

.../5

Question 9 Gini index.

.../5

Question 10 À la page 356, les auteurs disent :

“We can use the closed frequent itemsets to determine the support counts of the non-closed frequent itemsets.”

Expliquez comment cela est possible. Utilisez un exemple concret et bien choisi pour illustrer votre explication.

.../10

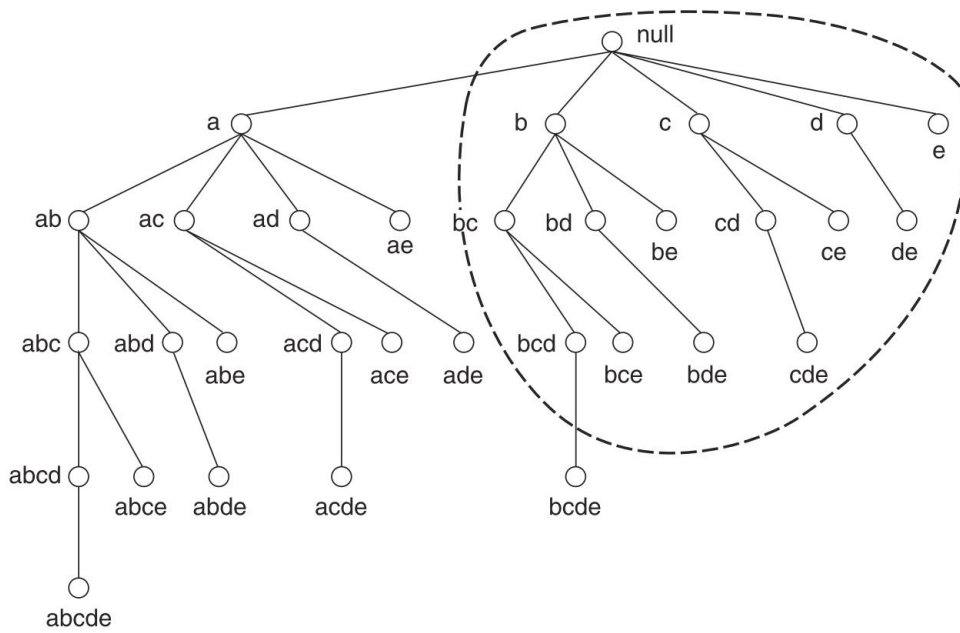


Figure 3: Generating candidate itemsets using the \boxed{X} approach.

Question 11 Voir la figure 3. Quel mot a été remplacé par \boxed{X} dans l'intitulé de cette figure ?

.../1

Question 12 Voir la figure 3. Dans quel ordre les itemsets sont-ils traités dans l'approche \boxed{X} ? Cochez la case qui précède la phrase correcte.

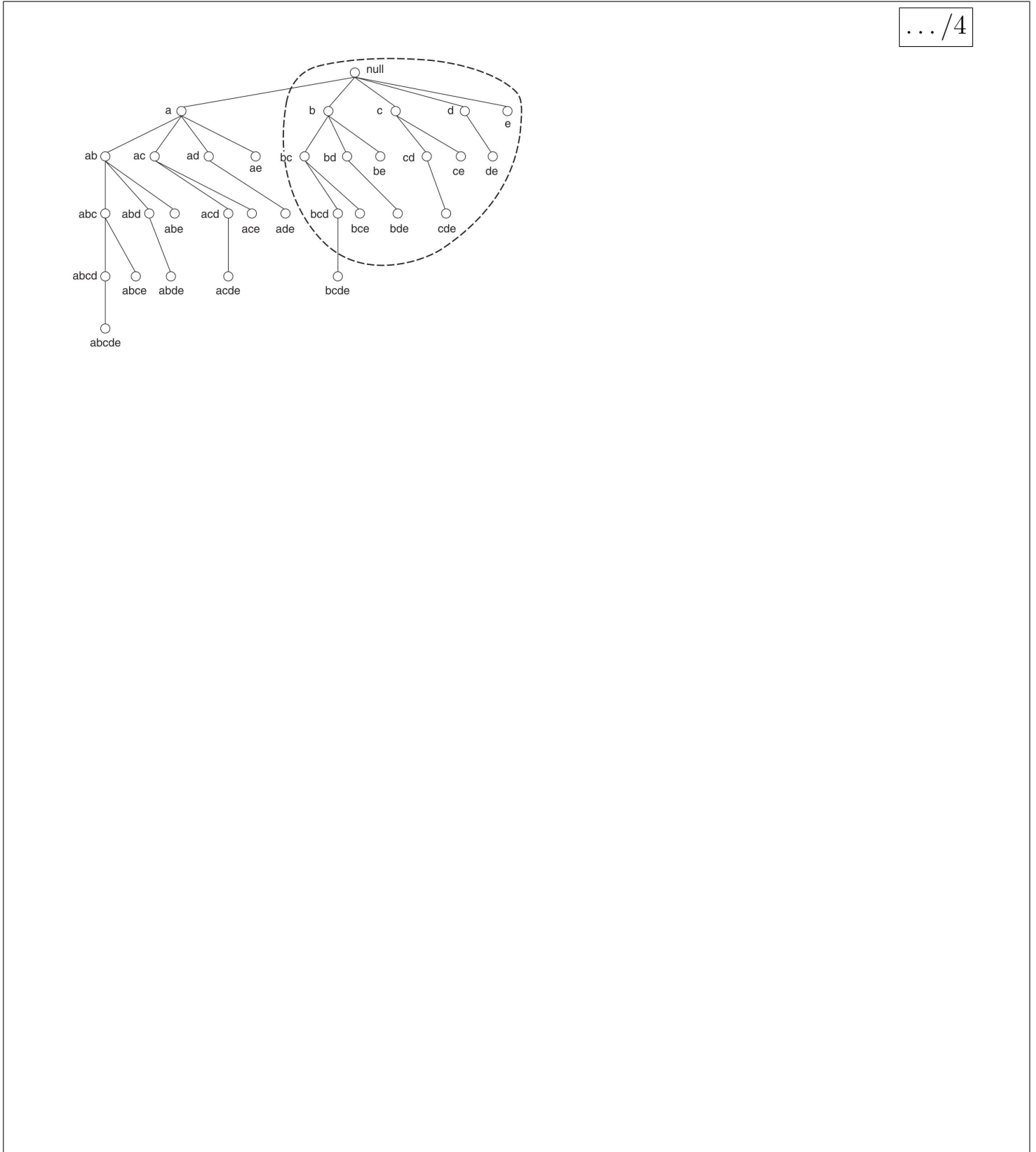
.../1

- Tous les 1-itemsets seront traités avant ab.
- bde sera traité avant c.

Question 13 Voir la figure 3. À la page 361, les auteurs disent :

“The $\square X$ approach is often used by algorithms designed to find maximal frequent itemsets. This approach allows the frequent itemset border to be detected more quickly than using a breadth-first approach. Once a maximal frequent itemset is found, substantial pruning can be performed on its subsets.”

Illustrez cette phrase de façon précise à l’aide de la figure 3.



Question 14 Voir la figure 3. À la page 362, les auteurs disent :

“The $\square X$ approach also allows a different kind of pruning based on the support of itemsets. For example, suppose the support for $\{a, b, c\}$ is identical to the support for $\{a, b\}$. The subtrees rooted at abd and abe can be skipped because they are guaranteed not to have any maximal frequent itemsets. The proof of this is left as an exercise to the readers.”

Donnez la preuve qui est laissée aux lecteurs.

.../4

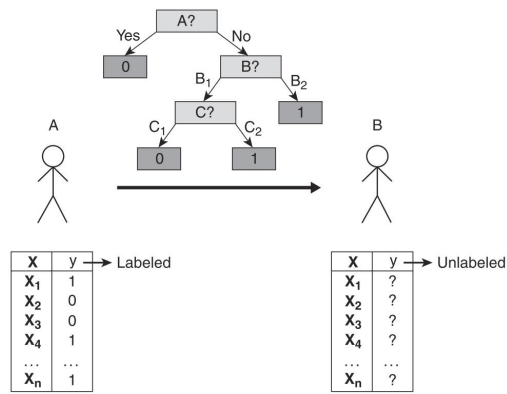


Figure 4: The \boxed{Y} principle.

Question 15 Voir la figure 4. Quel mot a été remplacé par \boxed{Y} dans l'intitulé de cette figure ?

.../1

Question 16 Expliquez la figure 4 de façon détaillée. Évitez des explications trop générales qui ne sont pas spécifiques pour la figure en question.

.../9

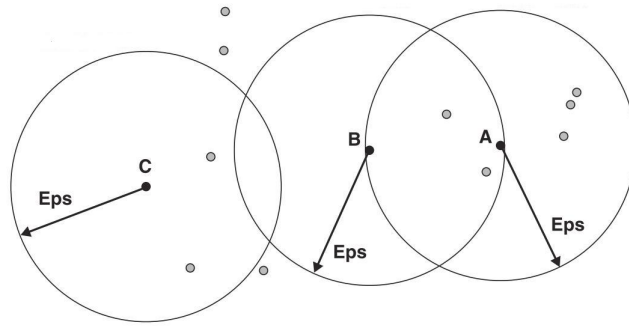


Figure 5: Z .

Question 17 Voir la figure 5. Pour quelle(s) valeur(s) de $MinPts$ cette-image est-elle correcte ?

.../2

Question 18 Expliquez la figure 5 de façon détaillée. Évitez des explications trop générales qui ne sont pas spécifiques pour la figure en question.

.../8