

Bases de données II, Charleroi, 19 janvier 2012

Partie cahier ouvert.

Nom et prénom
Année

La figure 2 montre une base de données XML pour stocker les CD possédés par mes amis. Chaque CD (balise CD) est identifié de façon unique par un identifiant (balise ID, avec valeurs *bd85*, *bt88*, *dp82*,...). La base de données stocke, par exemple, que mon ami *Jean* possède un exemplaire de *dh73* et un exemplaire de *rs90*. La figure 1 montre le DTD.

```
<!-- This file is called cd.dtd -->
<!ELEMENT COLLECTIONS (CATALOG, FRIENDS)>
<!ELEMENT CATALOG (CD*)>
<!ELEMENT FRIENDS (FRIEND*)>
<!ELEMENT CD (TITLE, ARTIST, COUNTRY, COMPANY, YEAR)>
<!ELEMENT FRIEND (DISK*)>
<!ELEMENT TITLE (#PCDATA)>
<!ELEMENT ARTIST (#PCDATA)>
<!ELEMENT COUNTRY (#PCDATA)>
<!ELEMENT COMPANY (#PCDATA)>
<!ELEMENT YEAR (#PCDATA)>
<!ELEMENT DISK (#PCDATA)>
<!ATTLIST CD ID CDATA #REQUIRED>
<!ATTLIST FRIEND NAME CDATA #REQUIRED>
<!ATTLIST DISK ID CDATA #REQUIRED>
```

FIGURE 1 – DTD.

Question 1 Écrivez une expression XPath (aussi simple que possible) qui rend chaque nœud de type `texte` dont la valeur est le nom d'un artiste qui a sorti un album avant 1975. Pour le document de la figure 2, la réponse contient `Dr.Hook`.

.../2

Question 2 Écrivez une expression XPath (aussi simple que possible) qui rend chaque nœud de type `attribute` dont la valeur est le nom d'un ami qui ne possède pas de CD. Pour le document de la figure 2, *An* ne possède pas de CD.

.../2

Question 3 Écrivez une expression XPath (aussi simple que possible) qui rend chaque nœud de type `attribute` dont la valeur est le nom d'un ami qui possède un CD de Joe Cocker. Pour le document de la figure 2, *Jean* et *Ed* possèdent un CD de Joe Cocker.

.../4

Question 4 Écrivez une expression XPath (aussi simple que possible) qui rend chaque nœud de type `attribute` dont la valeur est le nom d'un ami qui possède un CD qui se trouve aussi dans la collection de Jean. Pour le document de la figure 2, *Ed* possède un CD que Jean possède aussi (notamment *dh73*).

.../4

```

<?xml version="1.0"?>
<!DOCTYPE COLLECTIONS SYSTEM "cd.dtd">

<COLLECTIONS>
<CATALOG>
  <CD ID="bd85"><TITLE>Empire Burlesque</TITLE><ARTIST>Bob Dylan</ARTIST>
    <COUNTRY>USA</COUNTRY><COMPANY>Columbia</COMPANY><YEAR>1985</YEAR></CD>
  <CD ID="bt88"><TITLE>Hide your heart</TITLE><ARTIST>Bonnie Tyler</ARTIST>
    <COUNTRY>UK</COUNTRY><COMPANY>CBS Records</COMPANY><YEAR>1988</YEAR></CD>
  <CD ID="dp82"><TITLE>Greatest Hits</TITLE><ARTIST>Dolly Parton</ARTIST>
    <COUNTRY>USA</COUNTRY><COMPANY>RCA</COMPANY><YEAR>1982</YEAR></CD>
  <CD ID="gm90"><TITLE>Still got the blues</TITLE><ARTIST>Gary Moore</ARTIST>
    <COUNTRY>UK</COUNTRY><COMPANY>Virgin records</COMPANY><YEAR>1990</YEAR></CD>
  <CD ID="er97"><TITLE>Eros</TITLE><ARTIST>Eros Ramazzotti</ARTIST>
    <COUNTRY>EU</COUNTRY><COMPANY>BMG</COMPANY><YEAR>1997</YEAR></CD>
  <CD ID="bg88"><TITLE>One night only</TITLE><ARTIST>Bee Gees</ARTIST>
    <COUNTRY>UK</COUNTRY><COMPANY>Polydor</COMPANY><YEAR>1998</YEAR></CD>
  <CD ID="dh73"><TITLE>Sylvias Mother</TITLE><ARTIST>Dr.Hook</ARTIST>
    <COUNTRY>UK</COUNTRY><COMPANY>CBS</COMPANY><YEAR>1973</YEAR></CD>
  <CD ID="rs90"><TITLE>Maggie May</TITLE><ARTIST>Rod Stewart</ARTIST>
    <COUNTRY>UK</COUNTRY><COMPANY>Pickwick</COMPANY><YEAR>1990</YEAR></CD>
  <CD ID="jc87"><TITLE>Unchain my heart</TITLE><ARTIST>Joe Cocker</ARTIST>
    <COUNTRY>USA</COUNTRY><COMPANY>EMI</COMPANY><YEAR>1987</YEAR></CD>
  <CD ID="jc96"><TITLE>Organic</TITLE><ARTIST>Joe Cocker</ARTIST>
    <COUNTRY>USA</COUNTRY><COMPANY>EMI</COMPANY><YEAR>1996</YEAR></CD>
</CATALOG>
<FRIENDS>
  <FRIEND NAME="Jean">
    <DISK ID="dh73"/><DISK ID="rs90"/><DISK ID="jc87"/><DISK ID="jc96"/>
  </FRIEND>

  <FRIEND NAME="Ed">
    <DISK ID="gm90"/><DISK ID="dh73"/><DISK ID="jc87"/>
  </FRIEND>

  <FRIEND NAME="Pierre">
    <DISK ID="gm90"/>
  </FRIEND>

  <FRIEND NAME="An"/>
</FRIENDS>
</COLLECTIONS>

```

FIGURE 2 – Fichier XML avec des informations sur des propriétaires de CD.

Question 5 Écrivez un programme XSLT qui affiche, pour chacun de mes ami(e)s, les CD dans sa collection. Le résultat doit être formaté comme suit (la position des blancs et retours à la ligne n'a pas d'importance) :

```
<?xml version="1.0" encoding="UTF-16"?>
<answer>
  <Jean>  <OWNS>Sylvias Mother (Dr.Hook)</OWNS>
          <OWNS>Maggie May (Rod Stewart)</OWNS>
          <OWNS>Unchain my heart (Joe Cocker)</OWNS>
          <OWNS>Organic (Joe Cocker)</OWNS>      </Jean>
  <Ed>    <OWNS>Still got the blues (Gary Moore)</OWNS>
          <OWNS>Sylvias Mother (Dr.Hook)</OWNS>
          <OWNS>Unchain my heart (Joe Cocker)</OWNS>  </Ed>
  <Pierre> <OWNS>Still got the blues (Gary Moore)</OWNS> </Pierre>
  <An></An>
</answer>
```

.../10

Partie cahier fermé.

Nom et prénom

Année

Question 6 À la page 356, les auteurs mentionnent la propriété suivante :

If a rule $X \rightarrow Y \setminus X$ with $X \subseteq Y$ does not satisfy the confidence threshold, then any rule $X' \rightarrow Y \setminus X'$ with $X' \subseteq X$ must not satisfy the confidence threshold as well.

Donnez une preuve de cette propriété.

.../5

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

Question 7 Cross-validation.

.../3

Question 8 Closed itemset.

.../3

Question 9 Ward's method.

.../3

Question 10 Bisecting K-Means.

.../3

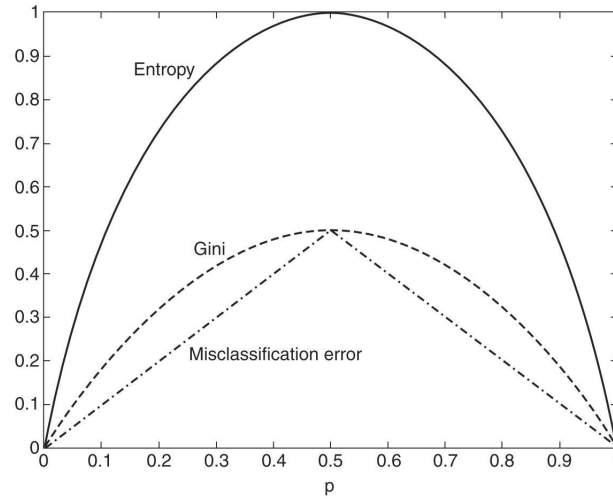


FIGURE 3 –

Question 11 Voir la figure 3.

1. Quelle est la valeur exacte de **Gini** pour $p = 0.3$?
2. Quelle est la valeur exacte de **Misclassification Error** pour $p = 0.373$?

Détaillez les calculs.

.../3

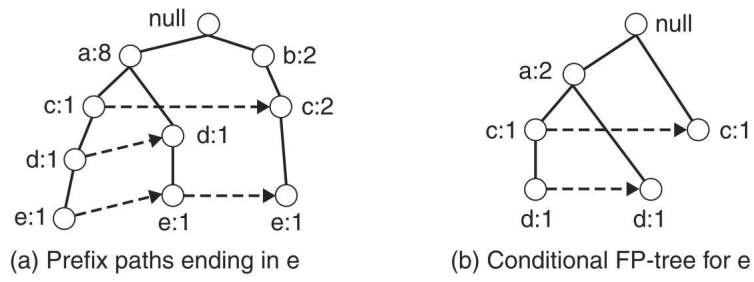


FIGURE 4 –

Question 12 Quel est le jeu de transactions représenté par l'arbre de gauche de la figure 4 ?

.../3

Question 13 Pour la figure 4, expliquez en détail comment on passe de l'arbre de gauche à l'arbre de droite. Expliquez aussi pourquoi le produit b disparaît.

.../6

Question 14 Le chapitre 8 contient une section « Unsupervised Evaluation of Hierarchical Clustering ». Expliquez d'abord le problème : que veut-on évaluer et pourquoi ? Expliquez ensuite une approche pour réaliser cette évaluation.

.../10

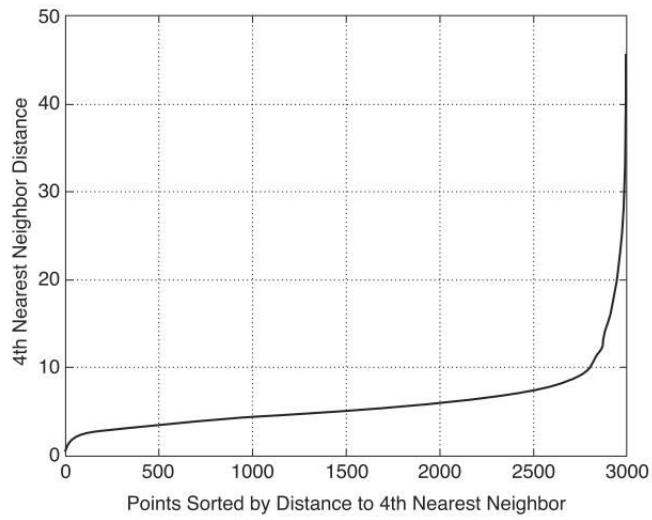


Figure 8.23. K-dist plot for sample data.

FIGURE 5 –

Question 15 Voir la courbe de la figure 5. Expliquez en détail (par exemple, en pseudocode de haut niveau) comment cette courbe est construite à partir d'un jeu de données.

.../3

Question 16 Expliquez la forme de la courbe de la figure 5. Pourquoi cette courbe est-elle croissante ? Pourquoi cette courbe a-t-elle la forme d'un genou ?

.../3

Question 17 Quelle est l'utilité de la courbe de la figure 5 pour le data mining ?

.../3