

Bases de Données II, Charleroi, 6 janvier 2020

NOM + PRÉNOM :

Orientation + Année :

Cet examen contient 9 questions.

Le document XML de la Figure 1 est utilisé pour stocker les résultats du Tour de France. La DTD est incluse au début du document XML. Chaque coureur est identifié par un code unique; l'attribut `naissance` donne son année de naissance. Chaque édition du Tour de France est identifiée par l'attribut `annee`. Les étapes d'une édition sont stockées dans l'ordre qu'elles ont eu lieu. Par exemple, Binges-Epernay était la deuxième étape de l'édition 2019. À l'arrivée de chaque étape, les coureurs arrivants sont classés 1, 2, 3, ... sans ex-æquos; l'attribut `temps` est le temps parcouru en secondes.

Pour les questions 1 à 4, évitez, si possible, l'usage des axes *parent* et *ancestor*. **Il n'est pas permis d'utiliser des fonctions d'agrégation, telles que `count`, `max`, `min`...**

Question 1 Écrivez une expression XPath qui renvoie les villes d'arrivée (attribut `arrivee`) où Peter Sagan a gagné en 2019. Pour le document de la figure 1, la réponse contient deux villes :

```
arrivee="Bruxelles"  
arrivee="Epernay"
```

.../5

Question 2 Écrivez une expression XPath qui renvoie les années, sans doublons, dans lesquelles au moins deux coureurs sont nés. Pour le document de la figure 1, la réponse contient deux années :

```
naissance="1997"  
naissance="1945"
```

.../5

```

<?xml version="1.0" encoding="utf-8"?>
<!DOCTYPE TourDeFrance [
<!ELEMENT TourDeFrance (coureur*, edition*)>
<!ELEMENT edition (etape)*>
<!ELEMENT etape (place)*>
<!ELEMENT coureur (#PCDATA)>
<!ELEMENT place (#PCDATA)>
<!ATTLIST coureur cid CDATA #REQUIRED>
<!ATTLIST coureur naissance CDATA #REQUIRED>
<!ATTLIST edition annee CDATA #REQUIRED>
<!ATTLIST etape depart CDATA #REQUIRED>
<!ATTLIST etape arrivee CDATA #REQUIRED>
<!ATTLIST etape distance CDATA #REQUIRED>
<!ATTLIST place position CDATA #REQUIRED>
<!ATTLIST place cid CDATA #REQUIRED>
<!ATTLIST place temps CDATA #REQUIRED>
]>
<TourDeFrance>
<coureur cid="AG" naissance="1999">Andreas Goeman</coureur>
<coureur cid="BL" naissance="1997">Bjorg Lambrecht</coureur>
<coureur cid="DH" naissance="1945">Daan Holst</coureur>
<coureur cid="EB" naissance="1997">Egan Bernal</coureur>
<coureur cid="EM" naissance="1945">Eddy Merckx</coureur>
<coureur cid="FH" naissance="1945">Fernand Hermie</coureur>
<coureur cid="GS" naissance="1997">Gabriel Silva</coureur>
<coureur cid="JA" naissance="1992">Julian Alaphilippe</coureur>
<coureur cid="JJ" naissance="1945">Jan Jannsen</coureur>
<coureur cid="PG" naissance="1982">Philippe Gilbert</coureur>
<coureur cid="PS" naissance="1990">Peter Sagan</coureur>
<coureur cid="YL" naissance="1991">Yves Lampaert</coureur>
<edition annee="2018">
  <etape depart="Lyon" arrivee="Macon" distance="105">
    <place position="1" cid="JA" temps="9000"/> <place position="2" cid="PG" temps="9000"/>
    <place position="3" cid="PS" temps="9000"/> <place position="4" cid="YL" temps="9000"/>
  </etape>
  <etape depart="Macon" arrivee="Dijon" distance="151">
    <place position="2" cid="JA" temps="13600"/> <place position="1" cid="PG" temps="13590"/>
    <place position="4" cid="PS" temps="13600"/> <place position="3" cid="YL" temps="13600"/>
  </etape>
  <etape depart="Dijon" arrivee="Auxerre" distance="130">
    <place position="2" cid="JA" temps="9360"/> <place position="3" cid="PG" temps="9367"/>
    <place position="1" cid="PS" temps="9360"/>
  </etape>
</edition>
<edition annee="2019">
  <etape depart="Bruxelles" arrivee="Bruxelles" distance="27">
    <place position="4" cid="EB" temps="1805"/> <place position="2" cid="JA" temps="1800"/>
    <place position="5" cid="GS" temps="1805"/> <place position="1" cid="PS" temps="1755"/>
    <place position="3" cid="YL" temps="1800"/>
  </etape>
  <etape depart="Binges" arrivee="Epernay" distance="111">
    <place position="2" cid="EB" temps="9300"/> <place position="3" cid="JA" temps="9400"/>
    <place position="1" cid="PS" temps="9300"/> <place position="4" cid="YL" temps="9405"/>
  </etape>
  <etape depart="Reims" arrivee="Nancy" distance="185">
    <place position="4" cid="EB" temps="16007"/> <place position="1" cid="JA" temps="16000"/>
    <place position="2" cid="PS" temps="16000"/> <place position="3" cid="YL" temps="16000"/>
  </etape>
</edition>
</TourDeFrance>

```

FIGURE 1 – Les résultats des étapes du Tour de France.

```

<?xml version="1.0" encoding="utf-16"?>
<PALMARES>
<COUREUR nom="Egan Bernal">
  <DEUXIEME annee="2019" etape="Binges-Epernay" />
</COUREUR>
<COUREUR nom="Julian Alaphilippe">
  <GAGNEUR annee="2018" etape="Lyon-Macon" />
  <DEUXIEME annee="2018" etape="Macon-Dijon" />
  <DEUXIEME annee="2018" etape="Dijon-Auxerre" />
  <DEUXIEME annee="2019" etape="Bruxelles-Bruxelles" />
  <TROISIEME annee="2019" etape="Binges-Epernay" />
<GAGNEUR annee="2019" etape="Reims-Nancy" />
</COUREUR>
<COUREUR nom="Philippe Gilbert">
  <DEUXIEME annee="2018" etape="Lyon-Macon" />
  <GAGNEUR annee="2018" etape="Macon-Dijon" />
  <TROISIEME annee="2018" etape="Dijon-Auxerre" />
</COUREUR>
<COUREUR nom="Peter Sagan">
  <TROISIEME annee="2018" etape="Lyon-Macon" />
  <GAGNEUR annee="2018" etape="Dijon-Auxerre" />
  <GAGNEUR annee="2019" etape="Bruxelles-Bruxelles" />
  <GAGNEUR annee="2019" etape="Binges-Epernay" />
  <DEUXIEME annee="2019" etape="Reims-Nancy" />
</COUREUR>
<COUREUR nom="Yves Lampaert">
  <TROISIEME annee="2018" etape="Macon-Dijon" />
  <TROISIEME annee="2019" etape="Bruxelles-Bruxelles" />
  <TROISIEME annee="2019" etape="Reims-Nancy" />
</COUREUR>
</PALMARES>

```

FIGURE 2 – Output du programme XSLT de la question 5.

Question 3 Écrivez une expression XPath qui renvoie les noms des coureurs nés après 1980 et n'ayant pas participé au Tour de France en 2019. Pour le document de la figure 1, la réponse contient trois noms : **Andreas Goeman, Bjorg Lambrecht et Philippe Gilbert.**

.../5

Question 4 Écrivez une expression XPath qui renvoie les noms des coureurs ayant participé à au moins deux éditions du Tour de France. Pour le document de la figure 1, la réponse contient trois noms : **Julian Alaphilippe, Peter Sagan et Yves Lampaert.**

.../5

Question 5 Écrivez un programme XSLT qui liste le palmarès de chaque coureur. La liste contient tous, et seulement, les coureurs qui se sont classés parmi les trois premiers d'une étape . Le format est celui de la figure 2. Pour chaque coureur, les étapes sont listées dans l'ordre qu'elles apparaissent dans le document d'entrée. L'ordre dans lequel les coureurs sont affichés n'a pas d'importance.

La position des blancs et retours à la ligne n'a pas d'importance. Le programme ne peut pas contenir les balises `xsl:for-each`, `xsl:if` ou `xsl:with-param`.

.../15

Question 6 Expliquez en détail l'image suivante.

FP-tree : une structure de données pour calculer les intersections en RAM (Random Access Memory)

TID	Items
1	{a,b}
2	{b,c,d}
3	{a,c,d,e}
4	{a,d,e}
5	{a,b,c}
6	{a,b,c,d}
7	{a}
8	{a,b,c}
9	{a,b,d}
10	{b,c,e}

Parmi les 8 transactions qui contiennent *a*, il y en a 5 qui contiennent aussi *b*. Parmi ces 5 transactions qui incluent *ab*, il y en a 3 qui contiennent aussi *c*. Etc.

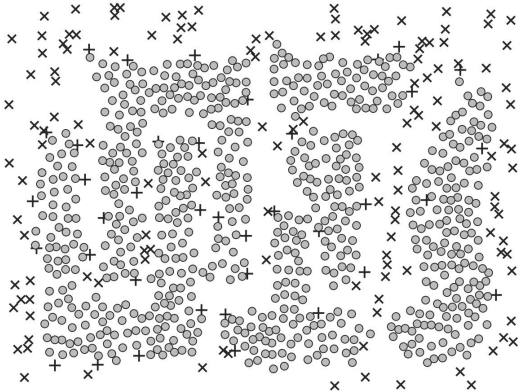
J. Wijsen Association Rules 20 / 21

Vous devez détailler

1. le problème qui est à la base de cette image ;
2. comment le *FP-tree* peut contribuer à la résolution de ce problème ;
3. les étapes du calcul qui suivent la construction du *FP-tree*. Illustrez ces étapes à l'aide de cet exemple concret ;
4. d'autres éléments que vous jugez pertinents.

.../10

Question 7 Expliquez la figure suivante de façon détaillée et précise.



- Quel algorithme aboutit à ce clustering ?
- Quel est le nombre de clusters ?
- Comment sont définis \times , $+$ et \circ ?

.../10

Question 8 Décrivez le fonctionnement de la méthode de *cross-validation* dans la classification et expliquez l'objectif de cette méthode.

.../10

Question 9 Situez chaque terme dans le cursus et expliquez de façon succincte mais précise.

Closed itemset.

.../5

Oblique decision tree.

.../5
