

Bases de Données II, Partie I, Charleroi, 5 janvier 2015

NOM + PRÉNOM :

Orientation + Année :

Cet examen contient 5 questions.

Un catalogue de livres est stocké dans un document XML. Les prix sont affichés en EUR ou USD, et peuvent varier d'un vendeur à l'autre.

La DTD est incluse au début du document XML de la figure 1. Le livre "XML Developer's Guide" coûte 44.95 USD auprès d'Amazon.com, et 34.00 EUR auprès de Proxis.com. **Le cours de change en vigueur est de 1 EUR = 1.34 USD.** Puisque $34.00 \text{ EUR} = 45.56 \text{ USD}$, Amazon.com offre un meilleur prix que Proxis.com. L'auteur de ce livre s'appelle Matthew Gambardella.

Pour les questions 1 à 4 évitez, si possible, l'usage des axes suivants : parent, ancestor, following-sibling, preceding-sibling, following et preceding.

Question 1 Écrivez une expression XPath (aussi simple que possible) qui rend chaque nœud de type **texte** dont la valeur est le titre d'un livre qui n'est pas disponible à un prix inférieur à 10.00 EUR (ou 13.40 USD ; le cours de change en vigueur est de 1 EUR = 1.34 USD). Pour le document de la figure 1, la réponse consiste en *XML Developer's Guide*, *Microsoft .NET: The Programming Bible* et *MSXML3: A Comprehensive Guide*.

.../5

Question 2 Écrivez une expression XPath (aussi simple que possible) qui rend chaque nœud de type `texte` dont la valeur est le nom de famille d'un auteur ayant écrit au moins deux livres. Essayez d'éviter d'afficher un même nom plusieurs fois. Pour le document de la figure 1, la réponse consiste en **Corets** et **O'Brien**.

.../5

Question 3 Écrivez une expression XPath (aussi simple que possible) qui rend chaque nœud de type `texte` dont la valeur est le titre d'un livre qui est écrit par l'auteur du livre intitulé "The Sundered Grail". Pour le document de la figure 1, la réponse consiste en **Maeve Ascendant**, **Oberon's Legacy** et **The Sundered Grail**. Notez que l'expression ne peut pas contenir des constantes différentes de **The Sundered Grail**.

.../5

Question 4 Écrivez une expression XPath (aussi simple que possible) qui rend chaque nœud de type `texte` dont la valeur est le nom de famille d'un auteur ayant écrit des livres dans deux ou plusieurs genres différents, parmi lesquels se trouve le genre "Fantasy". Notez que l'expression ne peut pas contenir des constantes différentes de **Fantasy**. Le même nom de famille peut apparaître plusieurs fois ; pas besoin d'éliminer les doublons. Pour le document de la figure 1, la réponse consiste en **Corets**.

.../5

```

<?xml version="1.0" encoding="utf-8"?>
<!DOCTYPE catalog [
<!ELEMENT catalog (book*)>
<!ELEMENT book (author, title, genre, sales)>
<!ELEMENT author (last, first)>
<!ELEMENT last (#PCDATA)>
<!ELEMENT first (#PCDATA)>
<!ELEMENT title (#PCDATA)>
<!ELEMENT genre (#PCDATA)>
<!ELEMENT sales (sale*)>
<!ELEMENT sale (#PCDATA)>
<!ATTLIST book id CDATA #REQUIRED>
<!ATTLIST sale bookstore CDATA #REQUIRED>
<!ATTLIST sale unit (EUR|USD) #REQUIRED>
]>
<catalog>
  <book id="bk101">
    <author><last>Gambardella</last><first>Matthew</first></author>
    <title>XML Developer's Guide</title><genre>Computer</genre>
    <sales><sale bookstore="Amazon.com" unit="USD">44.95</sale>
      <sale bookstore="Proxis.com" unit="EUR">34.00</sale>
      <sale bookstore="AbeBooks" unit="EUR">33.50</sale>
    </sales>
  </book>
  <book id="bk103">
    <author><last>Corets</last><first>Eva</first></author>
    <title>Maevae Ascendant</title><genre>Science Fiction</genre>
    <sales><sale bookstore="Biblio.com" unit="USD">11.95</sale>
      <sale bookstore="Proxis.com" unit="EUR">12.00</sale>
    </sales>
  </book>
  <book id="bk104">
    <author><last>Corets</last><first>Eva</first></author>
    <title>Oberon's Legacy</title><genre>Fantasy</genre>
    <sales><sale bookstore="Alibris" unit="EUR">11.95</sale>
      <sale bookstore="Proxis.com" unit="EUR">9.00</sale>
    </sales>
  </book>
  <book id="bk105">
    <author><last>Corets</last><first>Eva</first></author>
    <title>The Sundered Grail</title><genre>Fantasy</genre>
    <sales><sale bookstore="Amazon.com" unit="USD">5.95</sale>
      <sale bookstore="Books-A-Million" unit="USD">5.95</sale>
    </sales>
  </book>
  <book id="bk109">
    <author><last>Kress</last><first>Peter</first></author>
    <title>Paradox Lost</title><genre>Science Fiction</genre>
    <sales><sale bookstore="Amazon.com" unit="USD">14.95</sale>
      <sale bookstore="Proxis.com" unit="EUR">9.95</sale>
    </sales>
  </book>
  <book id="bk110">
    <author><last>O'Brien</last><first>Tim</first></author>
    <title>Microsoft .NET: The Programming Bible</title><genre>Computer</genre>
    <sales><sale bookstore="Biblio.com" unit="USD">36.95</sale>
      <sale bookstore="Proxis.com" unit="EUR">27.00</sale>
    </sales>
  </book>
  <book id="bk111">
    <author><last>O'Brien</last><first>Tim</first></author>
    <title>MSXML3: A Comprehensive Guide</title><genre>Computer</genre>
    <sales><sale bookstore="Amazon.com" unit="USD">36.95</sale>
      <sale bookstore="Alibris" unit="EUR">27.00</sale>
    </sales>
  </book>
</catalog>

```

FIGURE 1 – Catalogue de livres.

```

<bookstores>
  <bookstore name="AbeBooks">
    <title Euros="33.50">XML Developer's Guide</title>
  </bookstore>
  <bookstore name="Books-A-Million">
    <title Euros="4.76">The Sundered Grail</title>
  </bookstore>
  <bookstore name="Biblio.com">
    <title Euros="9.56">Maeve Ascendant</title>
    <title Euros="29.56">Microsoft .NET: The Programming Bible</title>
  </bookstore>
  <bookstore name="Proxis.com">
    <title Euros="34.00">XML Developer's Guide</title>
    <title Euros="12.00">Maeve Ascendant</title>
    <title Euros="9.00">Oberon's Legacy</title>
    <title Euros="9.95">Paradox Lost</title>
    <title Euros="27.00">Microsoft .NET: The Programming Bible</title>
  </bookstore>
  <bookstore name="Amazon.com">
    <title Euros="35.96">XML Developer's Guide</title>
    <title Euros="4.76">The Sundered Grail</title>
    <title Euros="11.96">Paradox Lost</title>
    <title Euros="29.56">MSXML3: A Comprehensive Guide</title>
  </bookstore>
  <bookstore name="Alibris">
    <title Euros="11.95">Oberon's Legacy</title>
    <title Euros="27.00">MSXML3: A Comprehensive Guide</title>
  </bookstore>
</bookstores>

```

FIGURE 2 – Output du programme XSLT.

Question 5 Écrivez un programme XSLT qui génère un document XML affichant les titres de livre par vendeur, dans le format illustré par la figure 2. Tous les prix sont affichés en EUR, en utilisant le cours de change de 1 EUR = 1.34 USD.

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

Bases de Données II Partie II, Charleroi, 5 janvier 2015

Cahier fermé. Cet examen contient 6 pages.

NOM + PRÉNOM :

Orientation + Année :

Question 1 Situez chaque terme dans le cursus et expliquez de façon succincte mais précise.

Leave-one-out.

.../5

Oblique decision tree.

.../5

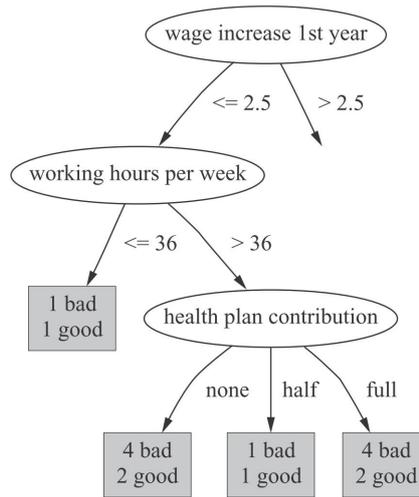


FIGURE 1 – Arbre de décision

Question 2 Pour l'arbre de décision de la figure 1, la question est de savoir si, oui ou non, il y aurait lieu d'élaguer le nœud "health plan contribution".

1. Dessinez l'arbre de décision que résulterait de cet élagage.
2. Décrivez une méthode scientifique pour déterminer si cet élagage est bénéfique ou non.

Question 3 Une agence de presse lance l'application payante *MonHéro* pour smartphone. Cette application demande à son utilisateur de sélectionner sa célébrité favorite dans une liste (e.g., *Vincent Company*, *François Hollande*, *Madonna*...), et envoie alors à l'utilisateur des articles de presse qui traitent de sa célébrité sélectionnée. L'application résout un problème de classification supervisée : elle s'appuie sur un arbre de décision pour décider, par exemple, si un article traite, oui ou non, de *François Hollande*. Détaillez des mesures statistiques que l'agence de presse puisse utiliser pour évaluer la qualité de l'application *MonHéro*. Intuitivement, la qualité est bonne si un utilisateur qui s'abonne sur *François Hollande* reçoit tous et seulement les articles qui traitent du président de la République française. La qualité est mauvaise si l'application envoie des articles non pertinents (par exemple, sur la Province de Hollande aux Pays-bas) ou n'envoie pas certains articles pertinents.

.../7

Question 4 Le principe de Minimum Description Length (MDL) est un critère permettant de déterminer entre deux modèles élaborés à partir de données empiriques, lequel est le plus simple. Détaillez comment ce critère peut être appliqué dans la comparaison entre deux arbres de décision M_1 et M_2 .

.../7

Question 5 Dans la construction d'un arbre de décision pour la table ci-dessous, on souhaite déterminer la valeur de v telle que le gain d'information du test binaire $A \leq v$ ou $A > v$ soit maximal. Détaillez une méthode efficace pour déterminer cette valeur de v . Illustrez cette méthode à l'aide de la table ci-dessous.

...	A	...	C
	16		yes
	23		yes
	44		yes
	49		yes
	52		yes
	27		no
	32		no
	36		no
	41		no
	56		no
	60		no
	62		no

.../7
