

TP1.1 Bases de XML

L'objectif de ce TP est de se familiariser avec les notions suivantes :

- Annotation descriptive et balisage
- Structuration de document
- Syntaxe de XML
- XML bien formé
- DTD

Exercice 1 - Structuration de documents

XML permet de structurer un document selon les informations qu'il contient. Il est donc nécessaire, avant d'envisager d'utiliser ce format, de se familiariser avec le problème de la structuration en lui-même.

1.1 Structuration d'information

Le texte suivant contient de l'information 'en vrac' :

Une bouteille d'eau Cristaline de 150 cl contient par litre 71 mg d'ions positifs calcium, et 5,5 mg d'ions positifs magnésium. On y trouve également des ions négatifs comme des chlorures à 20 mg par litre et des nitrates avec 1 mg par litre. Elle est recueillie à St-Cyr la Source, dans le département du Loiret. Son code barre est 3274080005003 et son pH est de 7,45. Comme la bouteille est sale, quelques autres matériaux comme du fer s'y trouvent en suspension.

Une seconde bouteille d'eau Cristaline a été, elle, recueillie à la source d'Aurèle dans les Alpes Maritimes. La concentration en ions calcium est de 98 mg/l, et en ions magnésium de 4 mg/l. Il y a 3,6 mg/l d'ions chlorure et 2 mg/l de nitrates, pour un pH de 7,4. Le code barre de cette bouteille de 50 cl est 3268840001008.

Une bouteille de même contenance est de marque Volvic, et a été puisée à... Volvic, bien connu pour ses sources donnant un pH neutre de 7. Elle comprend 11,5 mg/l d'ions calcium, 8,0 mg/l d'ions magnésium, 13,5 mg/l d'ions chlorures et 6,3 mg/l d'ions nitrates. Elle contient également des particules de silice. Son code barre est 3057640117008.

PS : Volvic est dans le Puy-de-Dôme...

1. Réorganiser l'information sans format particulier de manière à mettre en évidence sa structure logique. En s'attachant à ne perdre aucune information par rapport à la source d'information initiale (le texte).

```
Bouteille 1:
  Marque: Cristaline
  Volume: 150 cl
  Source: St-Cyr la Source
  Département: Loiret
  Composition:
    Ions positifs par litre :
      Calcium : 71 mg
      Magnésium : 5,5 mg
    Ions négatifs par litre :
      Chlorures : 20 mg
      Nitrates : 1 mg
  Contaminant: fer
  pH : 7,45
  Code barre : 3274080005003
```

```
Bouteille 2:
  Marque: Cristaline
  Volume: 50 cl
  Source : Aurèle
  Département: Alpes Maritimes
  Composition:
    Ions positifs par litre :
      Calcium : 98 mg
      Magnésium : 4 mg
    Ions négatifs par litre :
      Chlorures : 3,6 mg
```

Nitrates : 2 mg
pH : 7,4
Code barre : 3268840001008

Bouteille 3:
Marque: Volvic
Volume: 50 cl
Source: Volvic
Département: Puy-de-Dôme
Composition:
Ions positifs par litre :
Calcium : 11,5 mg
Magnésium : 8,0 mg
Ions négatifs par litre :
Chlorures : 13,5 mg
Nitrates : 6,3 mg
Contaminant: silice
pH neutre : 7
Code barre : 3057640117008

2. Puis transformer cette structure logique sous forme d'un document XML.

```
<eaux>
  <bouteille>
    <marque>Cristaline</marque>
    <volume>150 cl</volume>
    <source>St-Cyr la Source</source>
    <departement>Loiret</departement>
    <composition>
      <ions_positifs>
        <calcium>71 mg</calcium>
        <magnesium>5,5 mg</magnesium>
      </ions_positifs>
      <ions_negatifs>
        <chlorures>20 mg</chlorures>
        <nitrates>1 mg</nitrates>
      </ions_negatifs>
    </composition>
    <pH>7,45</pH>
    <contaminant>
      fer
    </contaminant>
    <code_barre>3274080005003</code_barre>
  </bouteille>
  <bouteille>
    <marque>Cristaline</marque>
    <volume>50 cl</volume>
    <source>Aurèle</source>
    <departement>Alpes Maritimes</departement>
    <composition>
      <ions_positifs>
        <calcium>98 mg</calcium>
        <magnesium>4 mg</magnesium>
      </ions_positifs>
      <ions_negatifs>
        <chlorures>3,6 mg</chlorures>
        <nitrates>2 mg</nitrates>
      </ions_negatifs>
    </composition>
    <pH>7,4</pH>
    <code_barre>3268840001008</code_barre>
  </bouteille>
  <bouteille>
    <marque>Volvic</marque>
    <volume>50 cl</volume>
    <source>Volvic</source>
    <departement>Puy-de-Dôme</departement>
    <composition>
      <ions_positifs>
        <calcium>11,5 mg</calcium>
        <magnesium>8,0 mg</magnesium>
      </ions_positifs>
      <ions_negatifs>
        <chlorures>13,5 mg</chlorures>
        <nitrates>6,3 mg</nitrates>
      </ions_negatifs>
    </composition>
    <pH>7</pH>
    <contaminant>
```

```

silice
</contaminant>
<code_barre>3057640117008</code_barre>
</bouteille>
</eaux>

```

3. Commenter la possibilité de structurer directement le texte (faisabilité, difficultés...).

Pour la question 1 on voit que les données sont structurées plutôt dans une façon de type array avec des keys (marque, source ...). C'est une façon de structurer de la donnée malgré le fait qu'elle masque une partie du texte, elle permet de voir instantanément les différentes informations données par le texte. La question 2 nous demande de structurer cette fois l'information en xml.

L'information est moins claire visuellement qu'à la question 1 mais l'information est toujours claire.

Les deux façons contiennent les mêmes informations.

1.2 Description de texte

Nous cherchons à décrire le texte de ce sujet de TP lui-même (ce que vous lisez en ce moment). Proposer un document XML qui rende compte de son organisation hiérarchique (inutile de recopier l'intégralité du texte).

```

<tp>
  <intro>
    <titre></titre>
    <description></description>
  </intro>
  <exercice>
    <titre></titre>
    <description></description>
    <partie>
      <sous-titre></sous-titre>
      <consigne></consigne>
      <question></question>
      <question></question>
      <question></question>
    </partie>
    <partie>
      <sous-titre></sous-titre>
      <consigne></consigne>
    </partie>
  </exercice>
  <exercice>
    <titre></titre>
    <description></description>
    <question></question>
    <question></question>
  </exercice>
  <exercice>
    <titre></titre>
    <description></description>
    <partie>
      <sous-titre></sous-titre>
      <document></document>
    </partie>
    <partie>
      <sous-titre></sous-titre>
      <document></document>
    </partie>
    <partie>
      <sous-titre></sous-titre>
      <document></document>
    </partie>
  </exercice>
  <exercice>
    <titre></titre>
    <description></description>
    <partie>
      <sous-titre></sous-titre>
      <schema></schema>
    </partie>
    <partie>
      <sous-titre></sous-titre>
      <document></document>
    </partie>
  </exercice>

```

```

<partie>
  <sous-titre></sous-titre>
  <document></document>
</partie>
</exercice>
<exercice>
  <titre></titre>
  <description></description>
</exercice>
<exercice>
  <titre></titre>
  <description></description>
  <question></question>
  <question></question>
  <question></question>
</exercice>
</tp>

```

Exercice 2 - Quelques recherches sur la syntaxe xml

Trouver les réponses aux questions suivantes dans la spécification du w3c sur <http://www.w3.org/TR/2008/REC-xml-20081126/> :

1.Comment écrit-on des commentaires en XML ?

Grâce aux balise `<!-- -->`, voici un exemple :

```

<!-- Se positionner sur la racine du document -->

```

2.A quoi servent les CDATA ? Donner un exemple.

Exercice 3 - Respect de la syntaxe du XML

Voici une série de documents XML, vous devrez indiquer si ces documents sont bien formés. S'il ne le sont pas, expliquez les erreurs et corrigez le document afin d'obtenir des documents XML bien formé.

3.1 Document 1

```

<annuaire>

  <personne id='1'>

    <nom>HEUTE</nom>

    <prenom>Thomas</prenom>

    <1email>webmaster@xmlfacile.com</1email>

  </personne>

  <personne id='2'>

    <nom>CANTAT</nom>

    <prenom>Bertrand</prenom>

    <$2email$>noir@desir.fr</2email>

  </personne>

```

Pour ce xml, il est mal formé car les balise email ont des caractère spéciaux, et la balise fermante `</annuaire>` est manquante :

```

<annuaire>
  <personne id='1'>
    <nom>HEUTE</nom>
    <prenom>Thomas</prenom>
    <email>webmaster@xmlfacile.com</email>
  </personne>
  <personne id='2'>
    <nom>CANTAT</nom>
    <prenom>Bertrand</prenom>

```

```

    <email>noir@desir.fr</email>
  </personne>
</annuaire>

```

3.2 Document 2

```

<?xml version="1.0" encoding="ISO-8859-1"?>
<!-- encoding="ISO-8859-1" n'est en fait pas nécessaire pour un fichier en anglais -->

<xsl:stylesheet xmlns:xsl="http://www.w3.org/TR/WD-xsl">

  <!-- Se positionner sur la racine du document -->

  <xsl:template match="/">

    <xsl:apply-templates >

  </xsl:template>

  <!-- Reproduire le document à l'identique -->

  <xsl:template match="@* | * | text()">

    <xsl:copy>

      <xsl:apply-templates select="@* | * | text()"/>

    <!-- Attention : ne pas couper la ligne suivante pour ne pas introduire d'espaces parasites -->

  </xsl:template>
</xsl:stylesheet>

```

Pour corriger ce document il faut :

- Mettre un url valide dans l'attribut xmlns:xsl dans la balise xsl:stylesheet.
- Ajouter des balises de fermeture pour les éléments xsl:template afin que le XML soit bien formé.

```

<?xml version="1.0" encoding="ISO-8859-1"?>
<!-- encoding="ISO-8859-1" n'est en fait pas nécessaire pour un fichier en anglais -->

<xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform">

  <!-- Se positionner sur la racine du document -->

  <xsl:template match="/">

    <xsl:apply-templates/>

  </xsl:template>

  <!-- Reproduire le document à l'identique -->

  <xsl:template match="@* | * | text()">

    <xsl:copy>

      <xsl:apply-templates select="@* | * | text()"/>

    </xsl:copy>

  </xsl:template>
</xsl:stylesheet>

```

3.3 Document 3

```

<?xml version ="1.0" ?>

<article>

  <editeur nom ="EYROLLES">

```

```

<adresse>

<rue>1, rue Thénard</rue>

<cp>75005</cp>

<ville>Paris</ville>

</Adresse>

</editeur>

<editeur nom=First Interactive>

<adresse>

<rue>13, rue Buffon</rue>

<cp>75005</cp>

<ville>Paris</ville>

</Adresse>

</editeur>

<topic>

<title>XML : le guide de l'utilisateur</title>

<$author>Elliott Rusty Harold</author>

<editeur nom="EYROLLES"> </editeur>

</topic>

<topic>

<title>Moteurs de recherche pour le web</title>

<author>Alfred, Emily Glossbrenner</author>

<editeur nom= First Interactive> </editeur>

</topic>

</article>

```

Pour corriger ce document il faut :

- Ajout de guillemets autour des valeurs des attributs nom dans les éléments <editeur> et <topic>.
- Supprimer le caractère '\$' dans l'élément <author>.

```

<?xml version="1.0" ?>

<article>
  <editeur nom="EYROLLES">
    <adresse>
      <rue>1, rue Thénard</rue>
      <cp>75005</cp>
      <ville>Paris</ville>
    </adresse>
  </editeur>

  <editeur nom="First Interactive">
    <adresse>
      <rue>13, rue Buffon</rue>
      <cp>75005</cp>
      <ville>Paris</ville>
    </adresse>
  </editeur>

  <topic>
    <title>XML : le guide de l'utilisateur</title>
    <author>Elliott Rusty Harold</author>
    <editeur nom="EYROLLES"></editeur>
  </topic>

  <topic>
    <title>Moteurs de recherche pour le web</title>

```

```
<author>Alfred, Emily Glossbrenner</author>
<editeur nom="First Interactive"></editeur>
</topic>
</article>
```

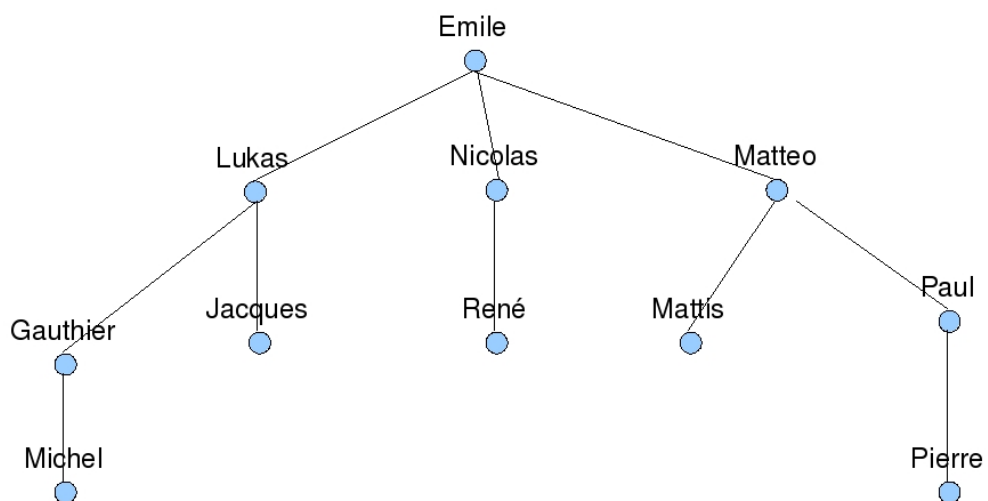
Exercice 4 - Représentation arborescente

Pour chacune des représentations suivantes, il faudra indiquer :

1. Le grand père de Gauthier
2. Le père de Michel
3. Le fils de Jacques
4. L'oncle de Pierre
5. Qui est Lukas par rapport à Matteo ?

Dérouler le parcours à réaliser pour trouver la réponse dans différentes conditions pour appréhender leur mérite respectifs :

4.1 dans un arbre



1. Le grand père de Gauthier

Il faut prendre le deuxième noeud au dessus : Emile

2. Le père de Michel

Il faut prendre le noeud au dessus : Gauthier

3. Le fils de Jacques

Il n'en a pas car Jacques n'a pas de noeud descendant.

4. L'oncle de Pierre

Il faut prendre le deuxième noeud au dessus et prendre le descendant qui n'a pas Pierre comme descendant : Mathis

5. Qui est Lukas par rapport à Matteo ?

Ils sont frère car ils ont le même noeud parent

4.2 dans un document XML non-indenté

```
<?xml version ="1.0" ?>

<personne nom='Mattis'>

<personne nom='Lukas'><personne nom='Matteo'>

<personne nom='Emile'>

<personne nom='Nicolas'>

<personne nom='Michel' /> </personne>

<personne nom='René' /> </personne>

</personne>

<personne nom='Gauthier' />

<personne nom='Jacques'>

<personne nom='Paul' />

<personne nom='Pierre' />

</personne>

</personne> </personne>
```

1. Le grand père de Gauthier

Mathis car ses balises englobent Lukas qui englobe Gauthier

2. Le père de Michel

Nicolas car ses balises englobent Michel

3. Le fils de Jacques

Les fils de Jacques sont **Pierre** et **Paul** car ses balises les englobent

4. L'oncle de Pierre

Gauthier car il est sur le même niveau que Jacques qui est son père

5. Qui est Lukas par rapport à Matteo ?

Lukas est le père de Matteo

4.3 dans un document XML indenté

```
<?xml version ="1.0" ?>
  <personne nom='René'>
    <personne nom='Jacques'>
      <personne nom='Matteo'>
        <personne nom='Lukas'>
          <personne nom='Nicolas' />
        </personne>
      </personne>
    </personne>
    <personne nom='Emile' />
  </personne>
  <personne nom='Michel'>
    <personne nom='Paul'>
      <personne nom='Mattis' />
      <personne nom='Pierre'>
        <personne nom='Gauthier' />
      </personne>
    </personne>
  </personne>
</personne>
```

1. Le grand père de Gauthier

Le père de Gauthier est **Paul** car il englobe Pierre qui englobe Gauthier

2. Le père de Michel

Le père de Michel est René

3. Le fils de Jacques

Matteo et Emile

4. L'oncle de Pierre

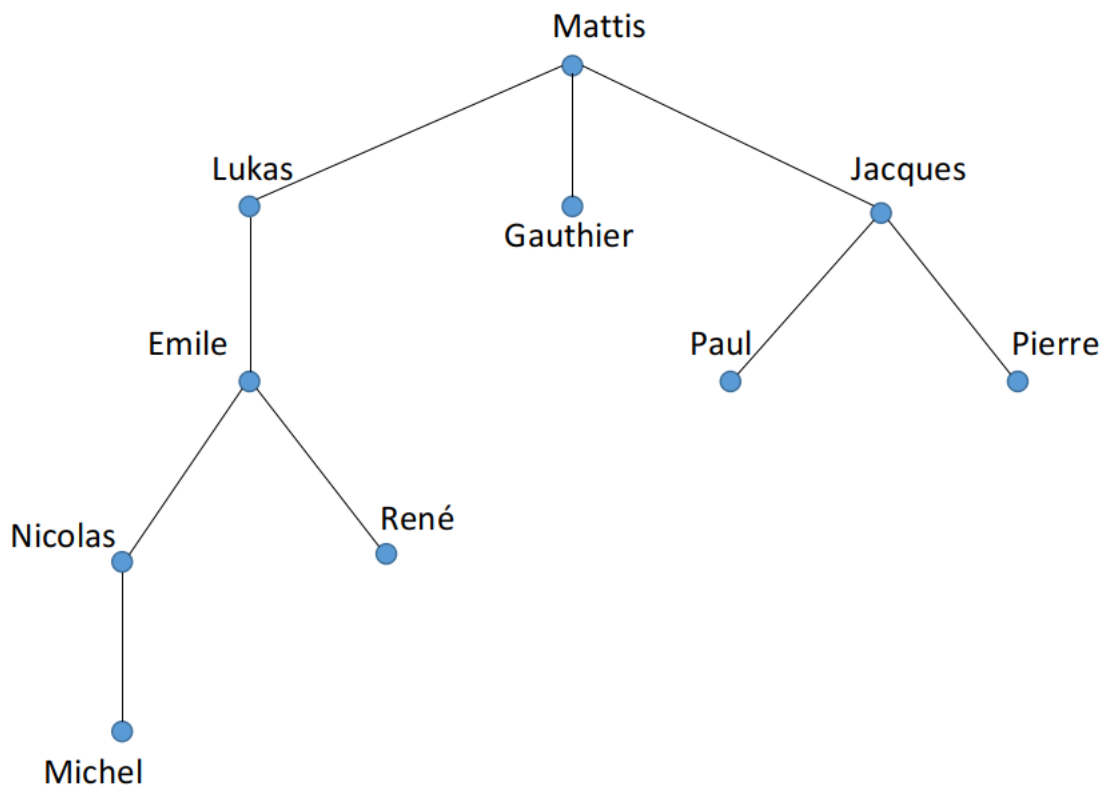
Il n'a pas d'oncle

5. Qui est Lukas par rapport à Matteo ?

Lukas est le fils de Matteo

Exercice 5 - Représentation en arbre

Reprendre l'exercice 4 et représenter le document 2 sous forme d'un arbre.



Exercice 6 - Structuration de données

Nous voulons rédiger une DTD pour une filmothèque. Cette filmothèque contient des films, et les fiches des participants (réalisateurs, acteurs, producteurs...).

Les informations nécessaires pour un film sont :

- le titre général ;
- le nom du réalisateur ;
- la liste des acteurs principaux ;
- les producteurs ;
- l'année de sortie ;

- son résumé ;

Les informations nécessaires pour un participant sont :

- son nom ;
- sa date de naissance ;
- le lien vers sa fiche wikipédia ;
- ses fonctions (parmi réalisateur, acteur, producteur) ;

Certains de ces éléments peuvent bien sur être vides dans certains cas. On réservera aussi un champ optionnel pour un avis personnel.

1.Proposer une DTD rendant compte de la description précédente. A l'aide de l'utilitaire ExchangeXML, tester cette DTD avec un fichier XML d'exemples mettant en évidence toutes les possibilités de la DTD.

DTD :

```
<!ELEMENT filmothèque (film+, participant*)>

<!ELEMENT film (titre, realisateur, acteurs?, producteurs?, annee, resume, avis?)>
<!ELEMENT titre (#PCDATA)>
<!ELEMENT realisateur (#PCDATA)>
<!ELEMENT acteurs (acteur+)>
<!ELEMENT acteur (#PCDATA)>
<!ELEMENT producteurs (producteur+)>
<!ELEMENT producteur (#PCDATA)>
<!ELEMENT annee (#PCDATA)>
<!ELEMENT resume (#PCDATA)>
<!ELEMENT avis (#PCDATA)>

<!ELEMENT participant (nom, date_naissance?, wikipedia?, fonctions)>
<!ELEMENT nom (#PCDATA)>
<!ELEMENT date_naissance (#PCDATA)>
<!ELEMENT wikipedia (#PCDATA)>
<!ELEMENT fonctions (#PCDATA)>

<!ATTLIST film
  id ID #REQUIRED>

<!ATTLIST participant
  id ID #REQUIRED>
```

XML d'exemple:

```
<!DOCTYPE filmothèque SYSTEM "films.dtd">
<filmothèque>
  <film id="film1">
    <titre>The Big Lebowski</titre>
    <realisateur>Joel Coen, Ethan Coen</realisateur>
    <acteurs>
      <acteur>Jeff Bridges</acteur>
      <acteur>John Goodman</acteur>
      <acteur>Julianne Moore</acteur>
    </acteurs>
    <producteurs>
      <producteur>Joel Coen</producteur>
      <producteur>Ethan Coen</producteur>
    </producteurs>
    <annee>1998</annee>
    <resume>Jeffrey Lebowski, dit le "Dude," se retrouve entraîné dans une série d'événements étranges après qu'on lui confonde avec un autre.
    <avis>Un film culte rempli d'humour.</avis>
  </film>
  <film id="film2">
    <titre>Las Vegas Parano</titre>
    <realisateur>Terry Gilliam</realisateur>
    <acteurs>
      <acteur>Johnny Depp</acteur>
      <acteur>Benicio Del Toro</acteur>
      <acteur>Tobey Maguire</acteur>
    </acteurs>
    <annee>1998</annee>
    <resume>Deux amis partent à Las Vegas pour un voyage hallucinant à la recherche du rêve américain.</resume>
    <avis>Un film fou.</avis>
  </film>
  <participant id="participant1">
    <nom>Joel Coen</nom>
```

```

        <date_naissance>29 novembre 1954</date_naissance>
        <fonctions>realisateur, Producteur</fonctions>
    </participant>
    <participant id="participant2">
        <nom>Ethan Coen</nom>
        <date_naissance>21 septembre 1957</date_naissance>
        <fonctions>realisateur, Producteur</fonctions>
    </participant>
    <participant id="participant3">
        <nom>Terry Gilliam</nom>
        <date_naissance>22 novembre 1940</date_naissance>
        <fonctions>realisateur</fonctions>
    </participant>
</filmotheque>

```

2.Modifier la DTD précédente :

- en ajoutant un attribut optionnel soustitre à l'élément titre
- en ajoutant un élément identifiant aux participants qui doit être utilisé ensuite pour les associer aux films dans les éléments réalisateur, acteurs et producteurs
- en faisant de l'élément année un attribut de type énuméré, prenant comme valeurs possibles 2023, 2022, 2021, "avant 2021" et "inconnue" et proposant comme valeur par défaut inconnue

```

<!ELEMENT filmotheque (film+, participant*)>

<!ELEMENT film (titre, realisateur, acteurs?, producteurs?, annee, resume, avis?)>
<!ELEMENT titre (#PCDATA)>
<!ELEMENT soustitre (#PCDATA)>
<!ELEMENT realisateur (#PCDATA)>
<!ELEMENT acteurs (acteur+)>
<!ELEMENT acteur (#PCDATA)>
<!ELEMENT producteurs (producteur+)>
<!ELEMENT producteur (#PCDATA)>
<!ELEMENT annee (#PCDATA)>
<!ELEMENT resume (#PCDATA)>
<!ELEMENT avis (#PCDATA)>

<!ELEMENT participant (nom, date_naissance?, wikipedia?)>
<!ELEMENT nom (#PCDATA)>
<!ELEMENT date_naissance (#PCDATA)>
<!ELEMENT wikipedia (#PCDATA)>
<!ELEMENT association (film*, participant*)>

<!ATTLIST film
    id ID #REQUIRED>

<!ATTLIST participant
    id ID #REQUIRED>

<!ATTLIST film annee (2023|2022|2021|avant_2021|inconnue) "inconnu">
<!ATTLIST titre soustitre CDATA #IMPLIED>

```

3.A partir de cette nouvelle DTD, créer un fichier XML exemple valide

```

<!DOCTYPE filmotheque SYSTEM "films.dtd">
<filmotheque>
    <film id="film1">
        <titre soustitre="Fury Road">The Big Lebowski</titre>
        <realisateur>Joel Coen, Ethan Coen</realisateur>
        <acteurs>
            <acteur>Jeff Bridges</acteur>
            <acteur>John Goodman</acteur>
            <acteur>Julianne Moore</acteur>
        </acteurs>
        <producteurs>
            <producteur>Joel Coen</producteur>
            <producteur>Ethan Coen</producteur>
        </producteurs>
        <annee>2021</annee>
        <resume>Jeffrey Lebowski, dit le "Dude," se retrouve entraîné dans une série d'événements étranges après qu'on lui confonde av
        <avis>Un film culte rempli d'humour.</avis>
    </film>
    <film id="film2">
        <titre soustitre="Fury Road">Las Vegas Parano</titre>
        <realisateur>Terry Gilliam</realisateur>
        <acteurs>

```

```

    <acteur>John Goodman</acteur>
    <acteur>Julianne Moore</acteur>
  </acteurs>
  <annee>2022</annee>
  <resume>Deux amis partent à Las Vegas pour un voyage hallucinant à la recherche du rêve américain.</resume>
  <avis>Un film fou.</avis>
</film>
<participant id="participant1">
  <nom>Joel Coen</nom>
  <date_naissance>29 novembre 1954</date_naissance>
  <wikipedia></wikipedia>
</participant>
<participant id="participant2">
  <nom>Ethan Coen</nom>
  <date_naissance>21 septembre 1957</date_naissance>
  <wikipedia></wikipedia>
</participant>
<participant id="participant3">
  <nom>Terry Gilliam</nom>
  <date_naissance>22 novembre 1940</date_naissance>
  <wikipedia></wikipedia>
</participant>
</filmotheque>

```