

Travaux dirigés et pratiques

Le formalisme XML

1 Première partie : données organisées sous forme d'arbre

1.1 Questions de cours concernant les bases du formalisme XML

Justifiez vos réponses si nécessaire.

1. XML pour... X... M... L...
2. Quel est le principal objectif de XML ?
3. En “voyant” un fichier XML seul, on sait tout de suite ce qu’il représente. V ou F ?
4. XML fournit une syntaxe et une sémantique. V ou F ?
5. Des données régulières (comme p.e. contenues dans une table relationnelle) peuvent être représentées avec XML. Par exemple, la table `person` suivante peut l’être.

name	phone	email
alan	3127786	alan@abc.com
sara	2136877	sara@xyz.edu
fred	7786312	fd@ac.uk

V ou F ? Si oui, donnez un exemple de représentation de la table `person` dans le formalisme XML.

6. Un document XML est bien formé ssi ...
7. Un document XML est valide ssi ...
8. Le document d est bien formé ... Le document d est valide (... À compléter par \rightarrow ou \leftarrow ou \leftrightarrow)

1.2 Les 7 erreurs du document XML bien formé

Voici un exemple de document XML qui devrait être bien formé :

```
<?xml version="3.0" encoding="utf-8"?>
<logiciel Comptabilite>
  <version>1</version/>
  <user>
    <uid>bgirot</uid>
  </user>
  <contact>
    <lastName>Giro</lastName>
    <firstName>Bernadette</FirstN</firstName>
    <email>gg</email>
  </contact>
  <author autId="ghuret" lastName="Huret" firstName="George" organism="">
    <company>
      <labId>3987</labId>
      <name>Analyse Informatique des Données & Information</name>
      <shortName></shortName>
    </company>
  </author>
  <author autId="mgirard" lastName="Girard" firstName="Matin" firstName="Matin" organism="">
    <company>
      <labId>6670</labId>
      <name>La compta en vogue</name>
      <shortName>Vogue la compta</shortName>
    </company>
    <company>
      <labId>4621</labId>
      <name>En avant les chiffres</name>
      <shortName>EAC</shortName>
    </company>
  </author>
</logiciel Comptabilite>
<date_mise_en_service>
  2007-09-17
</date_mise_en_service>
```

1. (Sur papier) Trouvez les 7 erreurs et corrigez-les.
2. (Sur machine) Vous pouvez récupérer le fichier "mal formé" à l'adresse http://raphael.fournier-sniehotta.fr/files/MedasAlgo/MEDAS-XML-TD_Formalism7erreurs.xml Corrigez-le et Vérifiez que votre document corrigé est bien formé. Pour cela, utilisez un éditeur de documents XML (Sera, Aquamacs, Xemacs, ... à vous de choisir) ou ouvrez tout simplement le document avec votre navigateur web qui intègre un parser XML.
3. (Sur papier) Représentez l'arbre XML du document corrigé.

1.3 La bibliothèque

On souhaite décrire le contenu d'une bibliothèque. Une bibliothèque est composée de livres. Chaque livre est composé d'un titre, d'un ensemble non vide d'auteurs, d'un éditeur, d'une année d'édition, d'un nombre de page, ainsi que d'un ensemble de parties (numéro de page, titre), contenant des chapitres (numéro de page, titre).

1. (Sur papier) Proposer une structuration XML de ce document, ne contenant pas d'attributs, en utilisant un exemple de votre choix.
2. (Sur papier) Représentez votre document sous forme d'arbre.
3. (Sur machine) Vérifiez que votre document est bien formé.

Pour tester si un document est bien formé, il suffit par exemple de l'ouvrir dans votre navigateur web, intégrant un parseur XML.

1.4 La bibliothèque... encore

Même contexte que pour l'exercice précédent mais:

1. (Sur papier) Proposer une structuration XML de ce document, contenant quelques attributs, en utilisant un exemple de votre choix.
2. (Sur papier) Représentez votre document sous forme d'arbre.
3. (Sur machine) Vérifiez que votre document est bien formé.

2 Deuxième partie : références, espaces de noms et DTD

2.1 Questions de cours concernant les espaces de noms et les DTD

1. Je peux n'associer ni DTD ni namespace à un document XML. V ou F ?
2. Je peux n'associer qu'une DTD à un document XML. V ou F ?
3. Je peux associer un ou plusieurs namespace(s) sans DTD à un document XML. V ou F ?
4. Je peux associer à la fois une DTD et un namespace à un document XML. V ou F ?
5. Je veux à la fois garantir une certaine syntaxe et désambiguer la sémantique de mon document :
 - (a) Si j'associe une DTD à mon document, je n'ai pas besoin de namespace. V ou F ?
 - (b) Si j'associe un namespace à mon document, je n'ai pas besoin de DTD. V ou F ?
6. Indiquer qu'un élément appartient à un namespace suffit à donner une sémantique à l'élément. V ou F ?

2.2 Les 7 erreurs de la DTD

1. Cette DTD contient six erreurs et une maladresse. Trouvez-les.

```
<!DOCTYPE mon_annuaire [  
<!ELEMENT annuaire (personne*)>  
<!ELEMENT personne (nom, prenom, email+, equipe)>  
<!ATTLIST personne type (etudiant | enseignant-chercheur | chercheur) "student">  
<!ATTLIST annee_naissance NUMBER #REQUIRED>  
<!ELEMENT id (integer)>  
<!ELEMENT nom (#PCDATA)>  
<!ELEMENT prenom (#PCDATA)>  
<!ELEMENT equipe (nom!, responsable)>  
<!ELEMENT responsable (personne)>  
<!ELEMENT email (#PCDATA)>  
<!ELEMENT note (#PCDATA)>  
>
```

2. Cette DTD ne pourra jamais être satisfaite. Pourquoi ?

2.3 Le mail

L'objectif de cet exercice est d'écrire un fichier XML représentant un mail et sa DTD. Un mail est constitué de deux parties.

1. Un entête constitué de :
 - un émetteur,
 - un ou plusieurs destinataires principaux,
 - éventuellement des destinataires secondaires,
 - éventuellement des destinataires cachés,
 - un sujet,
2. Un corps constitué d'un message.

Ecrivez une DTD ainsi qu'une instance XML codant un mail intitulé "MEDAS - proposition d'un sujet de réflexion" à votre chargé de cours avec les autres membres du groupe en copie. Le mail doit annoncer le titre du sujet choisi, les noms et prénoms des membres du groupe ainsi que leurs adresses mail. Validez votre document en ligne à l'adresse suivante : <http://validator.w3.org> (il vous suffit de copier-coller le contenu de votre fichier avec DTD déclarée en interne par *direct input* (troisième onglet)).

(Optionnel) Tentez de valider votre document avec le schéma de votre voisin. Que constatez-vous ? Qu'en déduisez-vous ? Ici, il y aura sûrement non validation car pb de syntaxe.

2.4 Un dialecte XML : MathML

MathML est un dialecte XML permettant de décrire des expressions mathématiques. Une description complète de MathML (syntaxe et sémantique) se trouve sur <http://www.w3.org/TR/MathML2/>. Afin de vous éviter de lire la documentation entière, voici quelques informations suffisantes mener à bien l'exercices : `mo` désigne un opérateur, `mrow` désigne une expression, `mi` désigne une variable, `mn` désigne un nombre, `⁢` désigne l'opérateur de multiplication et indique que celui-ci ne doit pas être affiché, `msup` désigne l'opérateur exposant et contient deux éléments : le nombre mis à la puissance puis l'exposant.

Voici un exemple de document mathML :

```
<math>
  <mrow>
    <mrow>
      <msup>
        <mi>x</mi>
        <mn>2</mn>
      </msup>
      <mo>+</mo>
      <mn>4</mn>
    </mrow>
    <mo>=</mo>
    <mn>0</mn>
  </mrow>
</math>
```

1. Donner une interprétation (expression mathématique) de ce document.
2. Le document suivant a-t-il la même interprétation ?
`<mrow><mrow><msup><mi>x</mi><mn>2</mn></msup><mo>+</mo><mn>4</mn></mrow><mo>=</mo><mn>0</mn></mrow>`

3. Le document suivant a-t-il la même interprétation ?

```

<math>
  <mrow>
    <mrow>
      <msup>
        <mi>x</mi>
        <mn>2</mn>
      </msup>
      <mo>+</mo>
      <mn>4</mn>
    </mrow>
    <mo>=</mo>
    <mn>0</mn>
  </mrow>
</math>

```

4. Vérifier que ce document est bien formé.

5. Namespace et des DTD.

- (a) Associez la DTD de MathML au fichier et ouvrez-le dans votre navigateur (Firefox). Que constatez-vous ?
- (b) Associez le namespace de MathML aux balises du document et ouvrez le document dans votre navigateur (Firefox). Que constatez-vous ?
- (c) Qu'en déduisez-vous ?

2.5 Documents satisfaisant une DTD

1. Pour chaque ensemble de documents, donner une DTD que respectent **tous** des documents de l'ensemble. Il peut exister plusieurs solutions (plus ou moins contraignantes) pour chacun des ensembles, donnez-en quelques-unes.

Ensemble 1 :

<A> 	<A> <C/>
---------------------	---------------------

Ensemble 2 : (une correction)

<A> <C/> <C/> <C/> <C/> 	<A> <C/> 	<A> <C/> <C/> <C/> 	<A> <C/>
---	---------------------	---	---

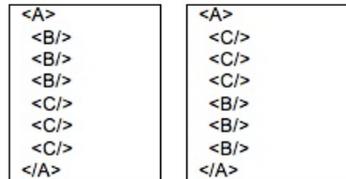
Ensemble 3 : (une correction)

<A> <C/> <C/> <C/> 	<A> 	<A> <C/> <C/>
---	-------------	---

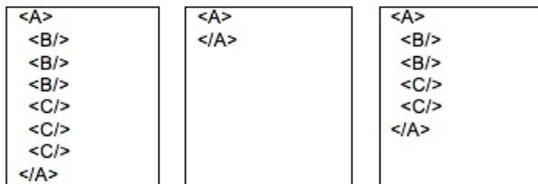
Ensemble 4 : (une correction)



Ensemble 5 : (une correction)



Ensemble 6 : (une correction)



2. On désire maintenant modifier la DTD de l'ensemble 4 en ajoutant les informations suivantes :

- B n'est plus vide mais contient des entiers.
- A a un attribut D dont la valeur est 3, 4 ou 5

Ecrivez la nouvelle DTD.

2.6 Des projets et des gens

Le LRI, dont le directeur est Michel Beaudouin-Lafon, dispose de plusieurs équipes de recherche. Parmi elles, on trouve l'équipe Programmation (directrice Marie-Claude Gaudel), l'équipe Démonstration (directrice Christine Paulin), l'équipe Intelligence Artificielle et Systemes d'Inference (directrice Chantal Reynaud), l'équipe Inference et Apprentissage (directrice Michèle Sebag) et l'équipe Bases de données (directeur Nicolas Spyrtos). Ces équipes participent à différents projets avec l'INRIA. Le projet GEMO (directeur Serge Abiteboul) implique l'équipe Intelligence Artificielle et Systemes d'Inference et l'équipe Bases de données, le projet TAO (directeur Marc Shoenauer) implique l'équipe Inference et Apprentissage, le projet InSitu (directrice Wendy Mackay) implique l'équipe Programmation et les projets Proval et Logical (directeur Gilles Dowek) impliquent l'équipe Démonstration.

Représentez les informations données dans l'énoncé ci-dessus, et ce sans redondance, dans le formalisme XML. Vérifiez que votre document est bien formé.

2.7 Des projets et des gens... suite

Ajouter une DTD au fichier issu de la traduction de l'énoncé "Des projets et des gens" (exercice 2.6). Validez votre document en ligne.

2.8 CineDVDbanque (optionnel)

L'objectif de cet exercice est de développer une base de données répondant aux besoins en stockage de distributeurs automatiques de DVDs en location appelé CineDVDBanque.

Les DVD mis à la location sont achetés via une centrale d'achat. Les informations disponibles sur ces DVD sont : le code barre numérique de l'oeuvre, le titre, le réalisateur, le genre et éventuellement une URI (Uniform Resource Identifier) indiquant le fichier XHTML correspondant à l'image du boîtier commercial du DVD. Les genres possibles sont Action, Comédie, Drame, Fantastique, Suspens et Terreur. Ces DVD sont en location dans des distributeurs localisés sur tout le territoire. Pour ces distributeurs, il est nécessaire de stocker leur adresse, les DVD en location. Les loueurs de DVD se sont nécessairement inscrits préalablement à toute location. Les informations requises pour cette inscription sont le nom, l'adresse, le téléphone. De plus, chaque loueur possède une clef alphanumérique correspondant au cryptage d'un mot de passe.

1. A l'aide d'une DTD, représentez les informations données dans l'énoncé ci-dessus. Donnez une instance valide de ce schéma. Validez votre document en ligne.
2. Vous allez mettre les données à disposition sur le site web de CineDVDBanque (<http://www.cinedvdbanque/>). Dans cette optique, associez un namespace à vos données. Validez votre document en ligne.