

Master TRIED - XML

Travaux Pratiques

Nicolas Travers
CNAM Paris

TABLE DES MATIÈRES

1 XML	3
1.1 Syntaxe XML	3
1.2 Arbre XML	4
1.3 Schéma DTD	4
1.4 XML et DTD	4
2 XPath	5
2.1 Évaluation XPath	5
2.2 Rédaction de requêtes XPath	5
2.3 Films et Artistes	6
2.4 DocX	6

Télécharger le logiciel BaseX¹ qui nous servira pour regarder les documents XML et faire des requêtes XPath.

1.1 Syntaxe XML

Corrigez le document XML suivant que vous pouvez télécharger à l'adresse suivante (faire 'enregistrer sous...') :

<http://cedric.cnam.fr/~traversn/teaching/TRIED/7erreursXML.xml>

Vérifiez que votre document corrigé est bien formé. Pour cela, utilisez un éditeur de documents XML² comme BaseX : Création d'une base, ouvrir le fichier XML où vous l'avez sauvegarder, la base est chargée s'il n'y a pas d'erreurs dans le document.

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

1. <http://basex.org/products/download/all-downloads/>

2. Editeurs : BaseX, Sera, Aquamacs, Xemacs, ... à vous de choisir) ou ouvrez tout simplement le document avec votre navigateur web qui intègre un parser XML

1.2 Arbre XML

Une fois le document chargé avec BaseX, vous pouvez visualiser l'arbre (Bouton 'arborescence', ou Ctrl+2). Naviguer dans celui-ci, ajouter des éléments via l'éditeur (sur un élément, bouton droit -> 'New').

1.3 Schéma DTD

1. Créer une DTD permettant de valider ce document XML. En commençant par le fragment ci-dessous :

```
<!DOCTYPE logicielComptabilite [
<!ELEMENT logicielComptabilite (version...)>
```

Pour vérifier la validation, veuillez proposer le document XML+DTD sur le site du W3C :
<http://www.xmlvalidation.com/index.php>

2. Trouvez dans la DTD suivante les erreurs (télécharger à l'adresse ci-dessous) :
<http://cedric.cnam.fr/~traversn/teaching/TRIED/7erreursDTD.xml>

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

1.4 XML et DTD

Pour le texte suivant, proposez un document XML et la DTD associée correspondant, et ce, sans redondance.

Le LRI, dont le directeur est Michel Beaudouin-Lafon, dispose de plusieurs équipes de recherche. Parmi elles, on trouve l'équipe Programmation (directeur Marie-Claude Gaudel), l'équipe Demonstration (directeur Christine Paulin), l'équipe Intelligence Artificielle et Systemes d'Inference (directeur Chantal Reynaud), l'équipe Inference et Apprentissage (directeur Michele 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 Shoenuer) implique l'équipe Inference et Apprentissage, le projet InSitu (directeur Wendy Mackay) implique l'équipe Programmation et les projets Proval et Logical (directeur Gilles Dowek) impliquent l'équipe Demonstration.

2.1 Évaluation XPath

Pour chaque expression XPath suivante, donner une phrase exprimant le résultat obtenu sur le document XML de référence¹. Vous testerez ces requêtes sur le logiciel *BaseX*².

1. /child : :biblio/child : :book/child : :author
2. /child : :biblio/child : :book[position()=1]/child : :author[position()=1]
3. /child : :biblio/child : :book/child : :title/child : :text()
4. /child : :biblio/child : :book/attribute : :ISBN
5. /biblio/book/author
6. count(/biblio/book/author)
7. count(/biblio/book[1]/author)
8. biblio/book[@lang="fr"]/title
9. /biblio/book[not(position()=1)]/author

2.2 Rédaction de requêtes XPath

À partir de chaque phrase suivante, trouvez la requête XPath correspondante.

1. Livres de la "biblio" n'ayant qu'un seul auteur
2. Toute balise "book" ayant un fils avec un seul auteur
3. Tous les livres avant le deuxième livre
4. Tous les livres après le deuxième livre
5. Tous les auteurs de livres français
6. Les titres de livres écrits en français écrit par "Knab"
7. Nombre de livres dont le sujet est "databases"
8. count(//book)
9. Premier et deuxième auteurs du premier livres
10. Titre des livres écrit par Philippe Rigaux avec au moins un autre auteur
11. Titre des livres dont le titre contient XML et publié chez Eyrolles
12. Auteurs ayant écrit au moins deux livres
13. Premier auteur de chaque livre
14. Nom des auteurs étant dans au moins deux livres
15. Noms des "publisher" de François Knab
16. Nom et Prénom des auteurs ayant écrit plusieurs livres sur XML

1. Document de référence : <http://cedric.cnam.fr/~traversn/teaching/nfe204/bibliotheque.xml>

2. <http://basex.org/products/download/all-downloads/>

2.3 Films et Artistes

Trouver les expressions XPath sur le document XML suivant :
<http://cedric.cnam.fr/~traversn/teaching/TRIED/FilmsArtistes.xml>

1. Titre des films Français
2. Films dans lesquels Gérard Darmond à joué
3. Nombre de films "coeur des hommes"
4. Metteur en scène du film "Gladiator"
5. Titre des films de Steven Spielberg
6. Titre des films de Steven Spielberg qu'il avait réalisé avant ses 40 ans
7. Titre des films dont le réalisateur est également acteur
8. Genre des films dans lesquels jouent ensemble De Funès et Galabru
9. Nombre de rôles joués par Johnny Deep
10. Nombre de rôles différents joués par Johnny Deep
11. Rôles (intitule) joués dans plusieurs films

2.4 DocX

Télécharger le document Microsoft Word au format XML :
<http://cedric.cnam.fr/~traversn/teaching/TRIED/docxTest.xml>

Pour le comparer avec BaseX, ouvrir le document avec Word (Microsoft ou Open office).

1. Trouver tous les éléments dont le namespace appartient à Word (w :)
2. Trouver le corps du document (body)
3. Afficher le texte du document
4. Afficher le titre du document
5. Trouver les éléments en italiques / gras
6. Donner les couleurs utilisées dans le document