

Fiche de TP : SAX & DOM

Contexte

La base de donnée MONDIAL est une base de donnée proposée par l'Université de Göttingen¹. Elle est présentée sous divers formats comme le modèle relationnel ou XML. Cette base a été construite à partir de données du Web telles que :

- CIA World Factbook² ;
- l'«International Atlas» de Kümmery Frey, Rand McNally et Westermann ;
- des données de la base de données géographique TERRA de Karlsruhe (Allemagne).

Cette base existe en XML. Elle est disponible (DTD, XML et d'autres documents complémentaires) dans Madoc.

Travail

On cherche à écrire un programme permettant d'obtenir un document XML présentant les pays (le nom du pays et le nom de la capitale) qui sont en Asie (continent dont l'identifiant est «asia») mais pas complètement (pas à 100%). Le document à générer doit respecter la DTD «liste-pays.dtd» suivante :

```
<?xml version="1.0" encoding="UTF-8"?>
<!ELEMENT liste-pays (pays)*>
<!ELEMENT pays EMPTY>
<ATTLIST pays
  nom CDATA #REQUIRED
  capitale CDATA #REQUIRED
  proportion-asie CDATA #REQUIRED
  proportion-autres CDATA #REQUIRED>
```

Sur le document «mondial.xml», cela donnera :

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<!DOCTYPE liste-pays SYSTEM "liste-pays.dtd">
<liste-pays>
  <pays nom="Russia" capitale="Moscow" proportion-asie="80" proportion-autres="20"/>
  <pays nom="Turkey" capitale="Ankara" proportion-asie="95" proportion-autres="5"/>
  <pays nom="Egypt" capitale="Cairo" proportion-asie="10" proportion-autres="90"/>
  <pays nom="Indonesia" capitale="Jakarta" proportion-asie="80" proportion-autres="20"/>
</liste-pays>
```

Question 1.

Donner le programme XSLT permettant d'obtenir le résultat précédent.

Question 2.

Donner les programmes PHP permettant d'obtenir le résultat précédent en utilisant les outils suivants : SAX³, DOM sans utiliser XPath puis en utilisant XPath.

Pour les versions DOM et la version SAX, le document produit devra être généré d'abord en DOM avant d'être retourné. Pour chaque version, faites en sorte que le traitement soit optimal en temps de traitement.

Question 3.

Discuter des différentes versions proposées (XSLT, DOM, DOM/XPath et SAX,) en termes de performance, d'adéquation au traitement, de clarté du code, de maintenance, etc.

¹ <http://dbis.informatik.uni-goettingen.de/Mondial/>

² <http://www.odci.gov/cia/publications/factbook/>

³ Vous pouvez utiliser la bibliothèque Sax4PHP : <http://sourceforge.net/projects/sax4php>

Question 4. (Facultative)

Reprendre cette recherche en utilisant :

- les objets XMLReader⁴ pour le traitement et XMLWriter⁵ pour la génération ;
- l'API SimpleXML de PHP ;
- L'API JDOM en Java.

Comparez ces approches avec les solutions précédentes.

Remarques :

- Une présentation de ces API est faite dans le cours e-miage ;
- Ce travail est à faire en binôme.
- En PHP, pour obtenir une sortie qui ne sera pas interprétée comme du HTML par votre navigateur, il faut mettre en début de programme «header('Content-type : text/plain') ;». Mettre «header('Content-type : text/xml') ;» pour qu'il soit traité correctement en tant que document XML.
- Il est aussi possible d'exécuter du PHP dans une fenêtre «terminal» en faisant : «php mon_exo.php».
- Éventuellement, pour évaluer le temps de traitement, il faut faire précéder la ligne par «time», par exemple : «time php mon_exo.php» ou «time java MaClasse». Le temps de traitement effectif de votre programme est la ligne «user».

Quelques références :

- Standard XPath : <http://xmlfr.org/w3c/TR/xpath>
- Standard DOM : <http://www.w3.org/DOM/>
- Documentation XMLELINT: <http://xmlsoft.org/xmlint.html> ;
- Documentation XSLTProc : <http://xmlsoft.org/XSLT/xsltproc.html>
- Documentation PHP DOM : <http://www.php.net/manual/fr/ref.dom.php>
- Documentation PHP «SAX» : <http://www.php.net/manual/fr/ref.xml.php>
- Documentation PHP SimpleXML : <http://php.net/manual/fr/book.simplexml.php>
- Site JDOM : <http://www.jdom.org/downloads/docs.html>

Travail à rendre :

Pour le vendredi 19 décembre 12h au plus tard, vous devrez rendre votre travail à l'enseignant qui encadre votre TP. Sous forme électronique (archive de nom "noms-binome.tar.gz"), vous donnerez :

- votre code ;
- un petit document de 4 à 5 pages présentant votre travail (analyse, conclusions), les subtilités de vos traitements et vos difficultés.

Une zone de dépôt sur Madoc sera disponible à cet effet. Une séance de tests pourra être organisée par l'enseignant de TP.



Cette création est mise à disposition selon le Contrat Paternité-Pas d'Utilisation Commerciale-Pas de Modification 2.0 France disponible en ligne <http://creativecommons.org/licenses/by-nc-nd/2.0/fr/> ou par courrier postal à Creative Commons, 171 Second Street, Suite 300, San Francisco, California 94105, USA.

⁴ <http://fr2.php.net/manual/fr/book.xmlreader.php>

⁵ <http://fr2.php.net/manual/fr/book.xmlwriter.php>