

# Fiche de TP : XSLT, SAX & DOM

---

## Contexte

Le corpus Sequoia est composé d'un ensemble de phrases annotées. Elle est disponible sur Madoc en plusieurs fichiers ou en un seul.

---

## Travail

### Question 0.

On cherche à écrire un programme permettant d'obtenir un document XML à partir d'une recherche ou d'un calcul sur le document. Ce programme sera proposé en plusieurs versions selon le paradigme utilisé (voir questions suivantes). Avec l'aide de votre enseignant, définir l'objectif et proposer un schéma permettant de mémoriser le résultat.

### 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 : SAX<sup>1</sup>, 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.

### Question 4. (Facultative)

Reprendre cette recherche en utilisant :

- les objets XMLReader<sup>2</sup> pour le traitement et XMLWriter<sup>3</sup> pour la génération ;
- l'API SimpleXML de PHP ;
- L'API JDOM en Java.

Comparez ces approches avec les solutions précédentes.

---

<sup>1</sup> Vous pouvez utiliser la bibliothèque Sax4PHP : <http://sourceforge.net/projects/sax4php>

<sup>2</sup> <http://fr2.php.net/manual/fr/book.xmlreader.php>

<sup>3</sup> <http://fr2.php.net/manual/fr/book.xmlwriter.php>

---

### 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 XLLINT: <http://xmlsoft.org/xmllint.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 :

Vous devrez rendre votre travail à l'enseignant qui encadre votre TP. Sous forme électronique (archive de nom "noms-binome.tar.gz"), vous donnerez :

- un fichier README permettant de donner les instructions nécessaires pour comprendre le code et l'exécuter correctement (éventuellement un fichier shell permettant l'exécution) ;
- votre code proprement présenté et commenté ;
- 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.