
TD-4 : Entités, DocBook et XSL-FO

Benoît Valiron

1. Encore des entités

(reprise du TD 3)

Considérez le fichier XML suivant.

```
<?xml version="1.0" encoding="utf-8"?>
<!DOCTYPE ville SYSTEM "ville.dtd">
<ville xmlns:geo="http://www.w3.org/2003/01/geo/wgs84_pos#">
  <nom>&grenoble;</nom>
  <geo:lat>45.196349</geo:lat>
  <geo:long>5.73226</geo:long>
</ville>
```

Si le fichier valide, que contient la DTD ?

Même question, avec une entité paramètre pour le préfixe geo. Comment changeriez-vous l'appel à la DTD dans le fichier XML pour pouvoir changer de préfixe ?

2. XSL-FO

Voilà une DTD simplifiée pour XSL-FO (la racine est root).

```
<!ELEMENT root (layout-master-set, page-sequence*)>
<!ELEMENT layout-master-set (page-master-sequence+)>
<!ELEMENT page-master-sequence (region-body)>
<!ATTLIST page-master-sequence master-name ID #REQUIRED
                                page-height CDATA #REQUIRED
                                page-width CDATA #REQUIRED
                                margin-top CDATA #REQUIRED
                                margin-bottom CDATA #REQUIRED
                                margin-left CDATA #REQUIRED
                                margin-right CDATA #REQUIRED>
<!ELEMENT region-body EMPTY>
<!ATTLIST region-body region-name ID #REQUIRED>

<!ELEMENT page-sequence (flow)>
<!ATTLIST page-sequence master-reference IDREF #REQUIRED>

<!ELEMENT flow (block+)>
<!ATTLIST flow flow-name IDREF #REQUIRED>

<!ELEMENT block (#PCDATA | inline)*>
<!ATTLIST block font-family CDATA #REQUIRED
                font-size CDATA #REQUIRED
                font-weight (normal|bold) "normal"
                font-style (normal|italic) "normal"
                text-align (left|right|center|justify) "left"
                color CDATA "black">
<!ELEMENT inline (#PCDATA)>
<!ATTLIST inline font-family CDATA #REQUIRED
                font-size CDATA #REQUIRED
                font-weight (normal|bold) "normal"
```

```
font-style    (normal|italic) "normal"
text-align   (left|right|center|justify) "left"
color        CDATA "black">
```

On vous demande de concevoir une page de texte A4 avec des marges de 1cm. Vous placerez un titre de police Helvetica, centré, de taille 18pt "Titre", un sous-titre de même police mais de taille 14pt disant "Auteur" et un paragraphe de texte justifié, de taille 12pt, en police "Times" disant "ceci est un mot en rouge". Le mot "mot" sera mis en gras et en rouge.

3. DocBook

Voilà un document DocBook simple.

```
<?xml version="1.0" encoding="utf-8"?>
<!DOCTYPE article PRIVATE
  "/usr/share/xml/docbook/schema/dtd/4.4/docbookx.dtd">
<article>
  <articleinfo>
    <title>Titre</title>
    <author>
      <firstname>Prénom auteur</firstname>
      <surname>Nom-auteur</surname>
    </author>
  </articleinfo>

  <sect1>
    <title>Titre de section</title>
    <para>
      Ceci est un <emphasis>mot</emphasis> emphasé.
    </para>
  </sect1>
</article>
```

Écrivez une DTD simple pour ce document.

Comparez avec le même texte en XSL-FO. En TP, vous allez utiliser un outils de conversion automatique de DocBook en XSL-FO puis de XSL-FO vers PDF.

On veut rajouter une section de titre "Un symbole", avec un paragraphe contenant la chaîne de caractère "±" (caractère unicode numéro 177). Où le texte sera-t-il placé ? Écrivez la partie manquante.

Maintenant, on veut définir une entité &plusmoins; pour faire appel au caractère spécial. Comment faire ?

4. Encore des DTDs

Le moteur de recherche google propose un service aux webmasters pour l'indexation : les sitemaps. Il s'agit d'un format XML pour donner des informations succinctes au robot d'indexation sur les pages qu'il peut rencontrer. Le format est le suivant (pris sur wikipedia) :

- Élément <urlset>. Obligatoire. Racine du document.
- Élément <url>. Obligatoire. Élément parent pour chaque entrée. Les éléments restant sont tous fils de cet élément.
- Élément <loc>. Obligatoire. Contient l'adresse internet d'une page, incluant le protocole (http:// ou https://). Doit faire au maximum 2048 caractères de long.

- Élément `<lastmod>`. Facultatif. La date de dernière modification du fichier, en format ISO ou plus simplement YYYY-MM-DD.
- Élément `<changefreq>`. Facultatif. Fréquence à laquelle la page est modifiée en générale : `always`, `hourly`, `daily`, `weekly`, `monthly`, `yearly`, `never`.
- Élément `<priority>`. Facultatif. L'importance relative de cette page par rapport aux autres. Valeur entre 0.0 et 1.0, valeur par défaut de 0.5.

Écrivez une DTD pour ce format, puis un document simple pour un site avec 3 pages : une page d'accueil à `http://www.mon-site.com`, changeant tous les jours ; une page `http://www.mon-site.com/infos.html`, ne changeant jamais ; une page `http://www.mon-site.com/doc.html`, changeant tous les mois.

On souhaite modifier le format pour que les informations contenues dans les éléments `loc`, `lastmod`, `changefreq` et `priority` soient maintenant attributs de `url`. Comment changeriez-vous la DTD ?

Les DTDs empêchent-elles l'écriture de documents sitemap non traitables par le robot d'indexation ?