

TP1 – XML, DTD, XML Schema

Indications

Pour lire ou modifier le contenu d'un fichier XML, il suffit de l'ouvrir avec un éditeur de texte (e.g., **gedit**). On peut également visualiser un fichier XML dans un navigateur ; si le fichier est bien formé on voit l'arbre du document, sinon on a une erreur. Dans le navigateur il est possible de voir le contenu du fichier initial en demandant à voir le code source de la page.

On va utiliser **xmllint** (disponible sous Linux) pour tester la validité d'un document XML.

Par exemple :

- Pour valider un document contre un DTD (qui lui est déjà associé), il suffit de faire :

```
xmllint --valid fichier.xml --noout
```

- Pour valider un document contre un XML Schema, il faut explicitement donner le schéma :

```
xmllint --schema fichier_schema.xsd fichier.xml --noout
```

Dans les deux cas, l'option **--noout** empêche d'afficher le document XML. Plus d'infos sur l'outil **xmllint** sont disponibles via la commande "**man xmllint**".

Pour plus de détails de syntaxe DTD et XML Schema, n'hésitez-pas de consulter les tutoriels W3Schools :

- http://www.w3schools.com/xml/xml_dtd_intro.asp
- http://www.w3schools.com/xml/schema_intro.asp

Exercice 1

La France, l'Italie et l'Espagne ont pour capitales respectives Paris, Rome et Madrid. Leurs populations respectives sont de 64 102 000 habitants (en 2007), de 58 133 509 et 44 708 964 habitants (en 2006). Structurez ces informations sous la forme d'un document XML bien formé.

Exercice 2

1. Validez le fichier **cd-1.xml** contre le DTD **cd.dtd**.
2. Essayez de valider le fichier **cd-2.xml** contre le même DTD **cd.dtd** et vous allez constater des erreurs. Faites un nombre minimal de changements sur le fichier **cd-2.xml** pour le rendre valide.
3. Validez le fichier **note-1.xml** contre le XML Schema **note.xsd**.
4. Essayez de valider le fichier **note-2.xml** contre le même XML Schema **note.xsd** et vous allez constater des erreurs. Faites un nombre minimal de changements sur le fichier **note-2.xml** pour le rendre valide.

Exercice 3

On désire écrire un DTD pour des documents décrivant des familles. Une famille porte un nom et est constituée d'une ou plusieurs personnes. Pour chaque personne de la famille, on a le prénom, l'âge, le poids en kilos (kg) ou le poids en livres (lb), et éventuellement la taille. Les liens de parenté (père et mère) sont gérés grâce à des attributs de type ID et IDREF.

Écrire le DTD correspondant `famille.dtd`. Celle-ci devra accepter par exemple le document suivant qui est disponible dans l'archive sous le nom `famille.xml`. L'attribut `pnumber` est obligatoire.

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE famille SYSTEM "famille.dtd">
<famille>
  <nom>Martin</nom>
  <personne pnumber="a1">
    <prenom>Juliette</prenom>
    <age>30</age>
    <poids-kg>58</poids-kg>
  </personne>
  <personne pnumber="a2">
    <prenom>Romeo</prenom>
    <age>31</age>
    <poids-lb>97</poids-lb>
  </personne>
  <personne pnumber="a3" mere="a1" pere="a2">
    <prenom>Max</prenom>
    <age>4</age>
    <poids-kg>12</poids-kg>
    <taille>1.25</taille>
  </personne>
  <personne pnumber="a4" mere="a1" pere="a2">
    <prenom>Marie</prenom>
    <age>3</age>
    <poids-lb>18</poids-lb>
    <taille>1.10</taille>
  </personne>
</famille>
```

Exercice 4

En plein championnat de football, on veut mémoriser les scores et les dates du championnat dans un fichier XML.

- Dans le championnat de France, il y a 20 clubs. On mémorisera la liste des clubs, indépendamment du calendrier du championnat. Chaque club possède un nom long (comme "Olympique de Marseille", un nom court (comme "OM").
- Le championnat est divisé en journées, et chaque journée comporte 10 rencontres. On mémorisera la date de chaque journée, mais pas la date de chaque match (parfois un match se joue en avance ou en retard par rapport à la date officielle). Il y a au total 38 journées.
- Les clubs se rencontrent tous 2 fois, une fois à domicile et une fois à l'extérieur. Les 19 premières journées, chaque club rencontre tous les autres (on parle de matches aller). Les 19 journées suivantes, chaque club rencontre aussi tous les autres mais pas sur le même terrain (on parle de matches retour). Une rencontre est donc caractérisée par un couple de (noms courts de) clubs. On mémorisera tous les scores des matches du championnat.

Définissez un XML Schema qui permet de représenter le championnat de France de football. La qualité d'un schéma est aussi liée à sa lisibilité. Évitez les imbrications trop profondes (schéma de type *poupées russes*). Un fichier `championnat.xml` est donné à titre d'exemple. Vous pouvez modifier ce fichier pour l'adapter à votre schéma et pour tester celui-ci.