

TP 2 : Validation des documents XML -DTD

Objectif : valider un document XML

Lab. 0 : Généralité sur la validation :

- La validation est un moyen pour vérifier que votre document est conforme à une grammaire.
- La validation va renforcer la qualité des échanges en contraignant l'émetteur de données et le consommateur de données à vérifier la cohérence des données structurées en XML.
- la validation revient à établir un visa sur le document XML.
- Il existe deux manières de validation : DTD et les schémas

Lab. 1 : la validation par DTD :

1. DTD :

- « DTD » (Document Type Definition)
- La DTD définit :
 - Quelle est la racine du document ?
 - Qui doit/peut avoir quels enfants ?
 - Combien d'enfants ?
 - Qui peut contenir du texte ?
 - Qui a des attributs particuliers ?
 - Quelles sont les valeurs de ces attributs ?
- On parle de :
 - Document bien formé : Respect de la syntaxe XML
 - Document valide : Respect de la DTD spécifiée
- Une DTD peut être interne ou externe au document XML. L'usage voudra que l'on privilégie la forme externe pour des raisons de maintenance et de facilité d'accès.
Par exemple, un document XML ayant une DTD externe cours.dtd, située dans le même Répertoire que notre document XML (accès relatif), se présente sous la forme :

```
<?xml version="1.0"?>
<!DOCTYPE cours SYSTEM "cours.dtd">
<cours>
...
</cours>
```

2. Création d'une DTD :

a. Déclaration des éléments :

- `<!ELEMENT nom_element EMPTY>` Élément vide `<Bidon/>`
- `<!ELEMENT nom_element ANY>` Peut contenir n'importe quoi
- `<!ELEMENT nom_element (#PCDATA)>` Element de données

b. Déclaration des sous éléments :

- `<!ELEMENT nom_element (sous_element)>`
`< nom_element >` contient un seul élément `< sous_element >`
- Un ensemble d'éléments séparés par des opérateurs, le tout placé entre parenthèses.
 - `|` : Opérateur de choix
 - `,` : opérateur de suite`<!ELEMENT enfant (age,sexe)>` `<enfant>` contient deux éléments `<age>` et `<sexe>` DANS CET ORDRE
`<!ELEMENT enfant (garcon|fille)>` `<enfant>` contient obligatoirement l'un ou l'autre des éléments `<garcon>`,`<fille>`

c. Déclarations des occurrences :

Les contenus (élément ou groupe d'éléments) peuvent être quantifiés par les opérateurs *, + et ?. Ces opérateurs sont liés au concept de cardinalité. Lorsqu'il n'y a pas d'opérateur, la quantification est de 1 (donc toujours présent).

Voici le détail de ces opérateurs :

- `:` 0 à n fois ;
- `+` : 1 à n fois ;
- `?` : 0 ou 1 fois.

Exemple :

`<!ELEMENT plan (introduction?,chapitre+,conclusion?)>`

L'élément plan contient un élément introduction optionnel, suivi d'au moins un élément chapitre et se termine par un élément conclusion optionnel également.

d. Déclaration des attributs :

Les attributs sont précisés dans l'instruction ATTLIST. Cette dernière, étant indépendante de l'instruction ELEMENT, on précise à nouveau le nom de l'élément sur lequel s'applique le ou les attributs. On peut considérer qu'il existe cette forme syntaxique :

nom TYPE OBLIGATION VALEUR_PAR_DEFAULT

Le TYPE peut être principalement :

- CDATA : du texte (Character Data) ;
- ID : un identifiant unique (combinaison de chiffres et de lettres) ;
- Une énumération de valeurs : chaque valeur est séparée par le caractère |.

L'obligation peut être :

- #REQUIRED : attribut obligatoire.
- #IMPLIED : attribut optionnel.
- #FIXED : attribut toujours présent avec une valeur.

Lab. 3 : Mise en pratique

Exercice 1

Utilisation d'une DTD

Créez la DTD carnet.dtd suivante :

```
<!ELEMENT carnet (personne+)>
<!ELEMENT personne EMPTY>
<!ATTLIST personne
  nom CDATA #REQUIRED
  prenom CDATA #IMPLIED
  telephone CDATA #REQUIRED>
```

Créez un document XML qui soit valide par rapport à cette DTD.

Exercice 2

Création d'une DTD

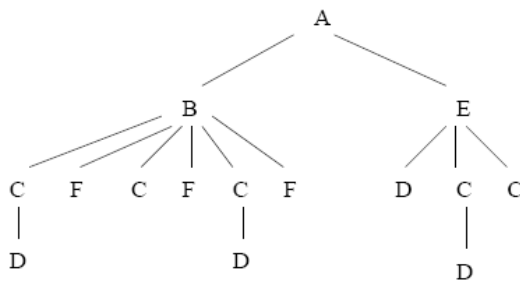
Créez une DTD livre.dtd à partir du document livre2.xml créé dans le TP précédent.

Exercice 3

Création d'une DTD

Créez une DTD annuaire.dtd à partir du document annuaire.xml créé dans le TP précédent.

Exercice 4 :



Définissez une DTD pour l'arbre : arbre.dtd

Ecrire le Document XML correspondant

Exercice 5 :

Écrire une DTD pour des documents XML - examen.

- Un examen contient un code de cours, un titre et une date composée du mois et de l'année.
- Ces éléments sont suivis par une liste de questions.
- Un examen a de 5 à 6 questions
- chaque question a une ou plusieurs parties.
- Une partie est un mélange de texte et d'autres parties.
- La valeur de mois doit être une chaîne de caractère valide.

Donnez une instance de document valide par rapport à la DTD