

Chapitre I: Rappel XML

I.1.Introduction à XML

Plan du cours

- Introduction
- XML
 - Principes XML
 - Langages apparentés
- STRUCTURE
 - L'entête :le prolog
 - Les commentaires
 - Type de document
 - Les nœuds éléments
 - Les attributs d'un élément
 - les nœuds texte
 - les nœuds texte :entité prédéfinies
 - les nœuds texte : Les entités du document
 - règles et syntaxe
- Conclusion

Introduction

- Le langage XML dérive de SGML (Standard Generalized Markup Language) et de HTML (HyperText Markup Language).
- SGML qui a été introduit en 1986 par C. Goldfarb. SGML a été conçu pour des documentations techniques de grande ampleur. Sa grande complexité a freiné son utilisation en dehors des projets de grande envergure
- Comme ces derniers, XML est un langage orienté texte et formé de *balises* qui permettent d'organiser les données de manière structurée

Introduction

- Le langage XML (eXtended Markup Language) est un format général de documents orienté texte. Il s'est imposé comme un standard incontournable de l'informatique. Il est aussi bien utilisé pour le stockage de document que pour la transmission de données entre applications.
- Sa simplicité, sa flexibilité et ses possibilités d'extension ont permis de l'adapter à de multiples domaines allant des données géographiques au dessin vectoriel en passant par les échanges commerciaux

- **SGML** riche, lourd, mal adapté au Web.
- **HTML** adapté à la présentation graphique, mais limité par un ensemble de balises figé, non extensible et sans typage.
- **Groupe de travail XML** à partir de août 1996 qui est composé essentiellement de membres du groupe de travail SGML.
 - **Recherche d'un langage assez simple** mais présentant une richesse proche de SGML.
 - **XML : un sous-ensemble de SGML**, qui élimine des points trop ciblés sur certains besoins.
 - Un document XML est **conforme** SGML.

XML

•Principes

- **séparer la structure d'un document de sa présentation** ☐
- distinguer le contenu d'un document et la présentation qui en est donnée. Un même contenu peut être rendu de façons très différentes. (html, pdf.....)

XML

•Principes

- Une structuration forte du document ☐

HTML	XML
Olivier Carton
 175, rue du Chevaleret
 75013 Paris
 <tt><u>Olivier.Carton@liafa.jussieu.fr</u></tt>	<pre><personne> <nomComplet> <nom> Carton </nom> <prenom> Olivier </prenom> </nomComplet> <dresse> 75013 Paris </dresse> <email> Olivier.Carton@liafa.jussieu.fr </email> </personne></pre>

XML

•Principes

- Une structuration forte du document ☐

XML

```
<personne>  
  <nomComplet>  
    <nom> Carton </nom>  
    <prenom> Olivier </prenom>  
  </nomComplet>  
  <dresse> 75013 Paris </dresse>  
  <email> Olivier.Carton@liafa.jussieu.fr </email>  
</personne>
```

Les balises (<personnes> ,<Nom> ...) sont des éléments **syntaxiques** destinés à structurer le contenu.

•Principes

- **Simplicité, universalité et extensibilité** ☐
- les noms des balises XML sont libres. Il appartient aux auteurs de documents de fixer les balises utilisées.
- il est seulement nécessaire que les auteurs s'entendent sur le vocabulaire, c'est-à-dire la liste des balises utilisées, lorsque des documents sont échangés.
- Cette liberté dans les noms de balises permet de définir des vocabulaires particuliers adaptés aux différentes applications.

XML

- Principes

- Interopérabilité

- Pouvoir échanger et traiter une donnée en utilisant de nombreux types de logiciels.

XML

•Principes

- Validation

- La liberté dans le choix des noms de balises implique une contrepartie. Il devient nécessaire de fixer des règles que doivent respecter les documents.
- Sans ces règles, il n'est pas possible d'échanger et de traiter de manière automatique ces documents.

Ces règles doivent d'abord fixer le vocabulaire mais aussi les relations entre les balises.

- Les règles peuvent, par exemple, imposer qu'une balise `<nomComplet>` contiennent exactement une balise `<nom>` et une balise `<prenom>` sans pour autant fixer l'ordre de ces deux balises.

- Langages apparentés
- plusieurs technologies et de langages se sont développés autour de XML. Ceux-ci enrichissent les outils pour la manipulation des documents XML: Xpat, xpointer, xquery, Xpath, XSLT ...

Structure

- La syntaxe de XML est relativement simple. Elle est constituée de quelques règles pour l'écriture d'une entête et des balises pour structurer les données.
- Ces règles sont très similaires à celles du langage HTML utilisé pour les pages WEB mais elles sont, en même temps, plus générales (balises libres) et plus strictes (toute balise ouvrante correspond à une balise fermante.....).

Structure

Document XML

Un document XML est généralement contenu dans un fichier texte dont l'extension est .xml.

Les fichiers contenant des documents dans un dialecte XML peuvent avoir une autre extension qui précise le format. Les extensions pour les schémas XML, les feuilles de style XSLT et les dessins en SVG sont, par exemple, .xsd, .xsl et .svg.

Structure

Document XML

un document XML est un document texte lisible

EXEMPLE

```
<?xml version="1.0" encoding="ISO-8859-1"?>
```

```
<!-- Date de création : 30/09/07 -->
```

```
<cours titre="XML">
```

```
    <intervenant nom="alexandre brillant"> </intervenant>
```

```
    <plan> Introduction XML et la composition de  
documents
```

```
    </plan>
```

```
</cours>
```

Structure

structure générale

La composition globale d'un document XML est immuable. Elle comprend toujours les constituants suivants :

- **Prologue** : Il contient des déclarations facultatives.
- **Corps du document** : C'est le contenu même du document.
- **Commentaires et instructions de traitement** : Ceux-ci peuvent apparaître partout dans le document, dans le prologue et le corps

STRUCTURE

L'entete :le prolog

```
<?xml version="1.0" encoding="ISO-8859-1"?>
```

```
<!-- Date de création 30/09/07 -->
```

```
<code>
```

la première ligne d'un document XML servant à donner les caractéristiques globales du document :

- La version XML, soit 1.0 ou 1.1 (plus part 1.0)
- Le jeu de caractères employé (*encoding*) (standards ISO et Unicode, UTF-16 UTF-8)
- Le champ standalone désigne l'indépendance du document (YES:libre , NO :depend d'autres documents pour former la forme finale)
- le prologue n'est pas obligatoire par défaut (version 1.0 et encoding UTF-8).

Les commentaires

```
<?xml version="1.0" encoding="ISO-8859-1"?>
```

```
<!-- Date de création : 30/09/07 -->
```

```
<cours titre="XML"
```

```
  <intervenar      ="alexandre brillant">  </intervenar>
```

```
  <plan> Intro XML et la composition de  
  documents
```

```
  </p>
```

```
</cours>
```

- les mêmes qu'en HTML (parenté avec SGML).
- se positionnent n'importe où après le prologue
- peuvent figurer sur plusieurs lignes.

Type de document

```
<?xml version="1.0" encoding="ISO-8859-1"?>
```

```
<!-- Date de création : 30/09/07 -->
```

```
<!DOCTYPE cours SYSTEM "cours.dtd">
```

```
<cours titre="XML">
```

```
  <intervenant nom="alexandre brillant"> </intervenant>
```

```
  <plan> Introduction à XML et la composition de  
  document
```

```
</c>
```

- Optionnelle
- sert à attacher une grammaire de type DTD (DocumentType Definition) au document XML
- On va y revenir

Les nœuds elements

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<!-- Date de création : 30/09/07 -->
<!DOCTYPE cours SYSTEM "cours.dtd">
<cours titre="XML">
    <intervenant nom="alexandre brillant">
    </intervenant>
    <plan> Introduction XML et la composition de
        documents
    </plan>
</cours>
```

- Les éléments (tags, nœuds) gèrent la structuration des données d'un document XML, Exemple : Cours , Intervenant, Plan
- La syntaxe d'un element est : <element> contenu </element>

Les nœuds elements

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<!-- Date de création : 30/09/07 -->
<cours titre="XML">
  <intervenant nom="alexandre brillant">
  </intervenant>
  <plan> Introduction XML et la composition de
    documents
  </plan>
</cours>
```

- Un élément racine par document : cours
- Un élément peut être:
 - Vide (pas de contenu)
 - Contient du texte (détaillé ultérieurement)
 - peut avoir plusieurs éléments fils : cours → intervenant, plan

Les attributs d'un element

```
<?xml version="1.0" encoding="ISO-8859-1"?>
```

```
<!-- Date de création : 30/09/07 -->
```

```
<!DOCTYPE cours SYSTEM "cours.dtd">
```

```
<cours titre="XML">
```

```
  <intervenant nom="alexandre brillant">
```

```
  </intervenant>
```

```
  <plan> Introduction XML et la composition de  
    document
```

```
  </plan>
```

```
</cours>
```

- Un attribut est un couple (clé, valeur) associé à la définition d'un élément.
- Apparaît dans la balise ouvrante de l'élément
- Un élément peut donc avoir de 0 à n attributs uniques
- complémentaire de l'élément (il ajoute une information à l'élément)

attributs ou element

```
<intervenant nom="alexandre brillant">  
</intervenant>
```

```
<intervenant>  
  <nom>"alexandre brillant"</nom>  
</intervenant>
```

- Un attribut peut se transformer en élément l'inverse n'est pas tj vrai
- un attribut ne peut pas être répété dans un élément

les nœuds texte

```
<?xml version="1.0" encoding="ISO-8859-1"?>
```

```
<!-- Date de création : 30/09/07 -->
```

```
<!DOCTYPE cours SYSTEM "cours.dtd">
```

```
<cours titre="XML">
```

```
  <intervenant nom="alexandre brillant">
```

```
  </intervenant>
```

```
  <plan> Introduction XML et la composition de  
    document
```

```
  </plan>
```

```
</cours>
```

- Une donnée est le texte (valeur) qui est associé à un attribut ou à un élément
- C'est le cœur du document (le formalisme ne sert qu'à classer les données)
- liée au prologue (encodage)

les nœuds texte : entité prédéfinies

<calcul>

if (a<b et b>c) ...

</calcul>



<calcul>

If (a<b et b>c)

</calcul>

Ambiguïté : <b b> sont il des balises ?

L'entité est un raccourci vers une autre valeur:

- < équivalent de < (*less than*) ;
- > équivalent de > (*greater than*) ;
- & équivalent de & (*ampersand*) ;
- " équivalent de " (*quote*) ;
- ' équivalent de ' (*apostrophe*).

les nœuds texte : *Les entités du document*

-- DTD --

```
<?xml version="1.0" encoding="ISO-8859-1"?>
```

```
<!DOCTYPE cours [
```

```
<!ENTITY auteur "Alexandre Brillant">
```

```
<!ELEMENT cours (#PCDATA)>
```

```
]>
```

.....

```
<?xml version="1.0" encoding="ISO-8859-1"?>
```

```
<cours>
```

```
Cours réalisé par &auteur;
```

```
</cours>
```

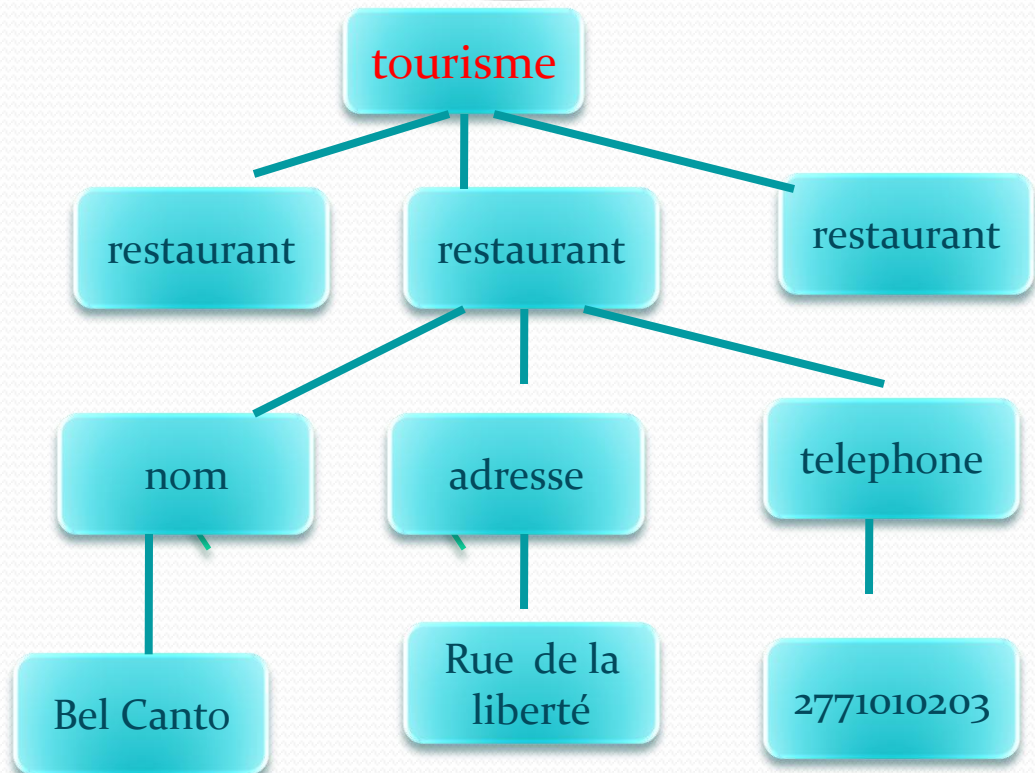
Arbre XML

Un document XML peut se représenter sous deux formes:

- La forme sérialisée est la forme courante (contenu marqué par des balises). Elle est utilisée pour:
 - Stocker un document dans un fichier
 - Echanger des documents
- La forme arborescente met en évidence la structure du document (facilite la conception des traitements).
 - Elle permet de situer la place de l'élément dans la hiérarchie
 - Elle permet de spécifier des manipulations de données XML

Arbre XML

```
<tourisme>
<restaurant>
  <nom>Bel Canto </nom>
  <adresse>
    Rue de la Tombe-Issoire
  </adresse>
  <telephone>
    01 45 46 47 48
  </telephone>
</restaurant>
<restaurant>
  .....
</restaurant>
<restaurant>
  .....
</tourisme>
```



REGLES et SYNTAXE

- Le nom d'un élément ne peut commencer par un chiffre.
- Si le nom d'un élément est composé d'un seul caractère il doit être dans la plage [a-z A-Z]
- Avec au moins 2 caractères, le nom d'un élément peut contenir `_`, `-`, `.` et `:` plus les caractères alphanumériques (attention, le caractère `:` est réservé à un usage avec les espaces de nom que nous aborderons par la suite).
- Tous les éléments ouverts doivent être fermés.
- Un élément parent est toujours fermé après la fermeture des éléments fils. `<a><b></a></b>`.

STRUCTURE

Exemple complet

```
<?xml version="1.0" encoding="iso-8859-1"?>

<!-- Time-stamp: "bibliography.xml 3 Mar 2008
16:24:04" -->

<!DOCTYPE bibliography SYSTEM
"bibliography.dtd" >

<bibliography>

  <livre key="Michardo1" lang="fr">
    <titre>XML langage et applications</titre>
    <auteur>Alain Michard</auteur>
    <annee>2001</annee>
    <editeur>Eyrolles</editeur>
    <isbn>2-212-09206-7</isbn>
    <url>http://www.editions-
      eyrolles/livres/michard/</url>
  </livre>

  <livre key="Zeldmano3" lang="en">
    <titre>Designing with web standards</titre>
    <auteur>Jeffrey Zeldman</auteur>
    <annee>2003</annee>
    <editeur>New Riders</editeur>
    <isbn>0-7357-1201-8</isbn>
  </livre>

  <livre>
    .....
  </livre>
  ...
  ...
  ...
</bibliography>
```