

Université de Jijel
Faculté des Sciences Exactes et d'Informatique
Département d'Informatique
L3 – Systèmes Informatiques

Données Semi-Structurées

Chapitre 03

Noyau XML

Tarek Boutefara
t_boutefara@univ-jijel.dz
2020/2021

Plan

- Introduction à XML
- Structure XML de base
- Domaines nominaux
- Schémas XML

Plan

- **Introduction à XML**
- Structure XML de base
- Domaines nominaux
- Schémas XML

Introduction à XML

- Spécification
 - <http://www.w3.org/TR/REC-XML>

Introduction à XML

- Spécification
 - <http://www.w3.org/TR/REC-XML>
- The World Wide Web Consortium (W3C)
 - Une communauté internationale qui travaille pour développer le web
 - Dirigé par Tim Berners-Lee
 - 440 membres (Adobe, Amazon, Cisco, Facebook, Google, Microsoft, ...)

Introduction à XML

- Spécification
 - <http://www.w3.org/TR/REC-XML>
- The World Wide Web Consortium (W3C)
 - Processus
 - Discussion des sujets intéressants,
 - Elaboration d'une première proposition,
 - Participation des groupes de travail,
 - Elaboration des spécifications (et guidelines).

Introduction à XML

- Spécification
 - <http://www.w3.org/TR/REC-XML>
- The Extensible Markup Language (XML) is a subset of SGML that is completely described in this document. Its goal is to enable generic SGML to be served, received, and processed on the Web in the way that is now possible with HTML.
- XML has been designed for ease of implementation and for interoperability with both SGML and HTML.

Introduction à XML

- Spécification
 - <http://www.w3.org/TR/REC-XML>
- Le langage XML (Extensible Markup Language) est un sous-ensemble de SGML Son objectif est de permettre au SGML générique d'être envoyé, reçu et traité sur le Web de la manière qui est désormais possible avec HTML.
- XML a été conçu pour faciliter la mise en œuvre et pour l'interopérabilité avec SGML et HTML.

Introduction à XML

- Objectifs :
 - XML doit être directement utilisable sur Internet,
 - XML doit supporter une large variété d'applications,
 - XML doit être compatible avec SGML,
 - Les programmes qui lisent XML doivent être faciles à écrire,
 - Les éléments optionnels doivent être gardés au minimum,

Introduction à XML

- Objectifs :
 - Les documents XML doivent être lisibles par les humains,
 - La conception des documents XML doit être facile,
 - La conception des documents XML doit être formelle,
 - Les documents XML doivent être faciles à créer.

Introduction à XML

- Pourquoi XML ?
 - Définir des langages d'échange personnalisés et adaptés,
 - Modéliser des données
 - Document Type Definitions (DTD)
 - Schémas XML (XML Schema)
 - Publier des informations
 - Mise en forme CSS
 - Transformations XSL
 - Archiver des données
 - Bases de Données XML

Plan

- Introduction à XML
- **Structure XML de base**
- Domaines nominaux
- Schémas XML

Introduction à XML

- `<?xml version="1.0" encoding="ISO-8859-1" ?>`
- `<livre lang="fr">`
 - `<titre>Java</titre>`
 - `<anneepub>2015</anneepub>`
 - `<prix>3500</prix>`
- `</livre>`

Introduction à XML

- `<?xml version="1.0" encoding="ISO-8859-1" ?>`
 - Déclaration de la version de XML utilisée dans le document

Introduction à XML

- `<livre> ... </livre>`
 - La racine du document

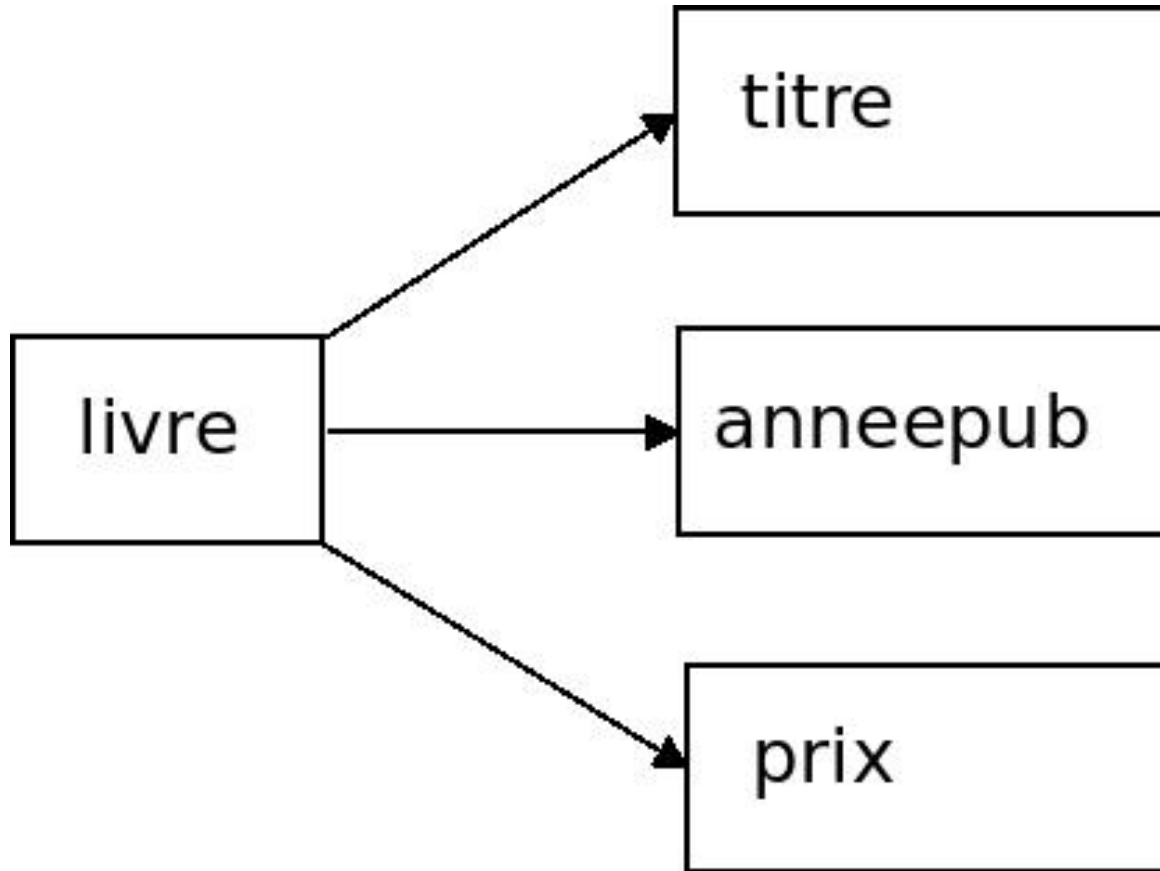
Introduction à XML

- `<livre> ... </livre>`
 - `lang= "fr"`
 - Attribut pour l'élément `<livre>`

Introduction à XML

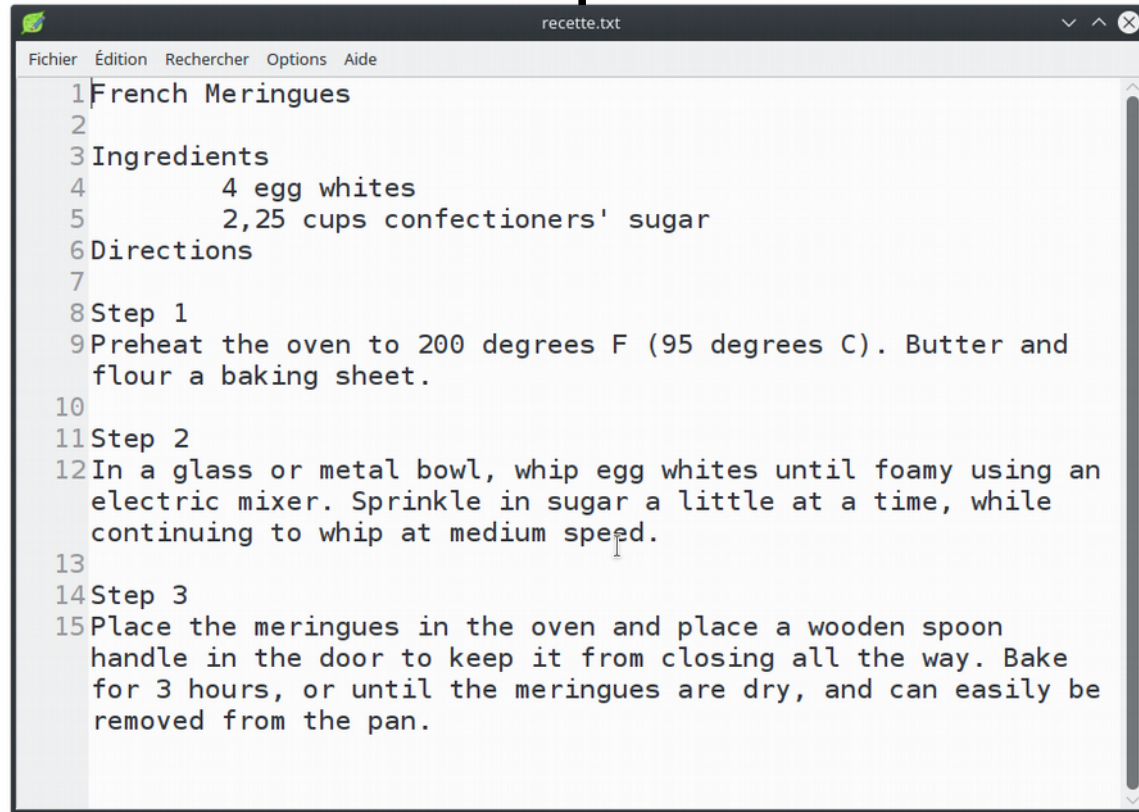
- Les éléments :
 - `<titre>Java</titre>`
 - `<anneepub>2015</anneepub>`
 - `<prix>3500</prix>`
- Trois éléments fils de l'éléments `<livre>`

Introduction à XML



Introduction à XML

- Un deuxième exemple



```
1 French Meringues
2
3 Ingredients
4     4 egg whites
5     2,25 cups confectioners' sugar
6 Directions
7
8 Step 1
9 Preheat the oven to 200 degrees F (95 degrees C). Butter and
  flour a baking sheet.
10
11 Step 2
12 In a glass or metal bowl, whip egg whites until foamy using an
  electric mixer. Sprinkle in sugar a little at a time, while
  continuing to whip at medium speed.
13
14 Step 3
15 Place the meringues in the oven and place a wooden spoon
  handle in the door to keep it from closing all the way. Bake
  for 3 hours, or until the meringues are dry, and can easily be
  removed from the pan.
```



/home/tarek/Enseignement/XML X



file:///home/tarek/Enseignement/XML ...



Search



Most Visited



Linux Mint



Blog



Forums



Community



News



Bookmark this page



Other Bookmarks

This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
-<recette id="1336">
  <titre>French Meringues</titre>
  -<ingredients>
    <ingredient qte="4">egg whites</ingredient>
    <ingredient qte="2.25" unite="cup">confectioners' sugar</ingredient>
  </ingredients>
  -<directions>
    -<step n="1">
      Preheat the oven to 200 degrees F (95 degrees C). Butter and flour a baking sheet.
    </step>
    -<step n="2">
      In a glass or metal bowl, whip egg whites until foamy using an electric mixer. Sprinkle in sugar a little
      at a time, while continuing to whip at medium speed.
    </step>
    -<step n="3">
      Place the meringues in the oven and place a wooden spoon handle in the door to keep it from closing
      all the way. Bake for 3 hours, or until the meringues are dry, and can easily be removed from the pan.
    </step>
  </directions>
</recette>
```

Introduction à XML

- Un deuxième exemple
 - Visualisation :
 - Plusieurs outils pour la visualisation sous forme d'arbre,
 - Beaucoup moins pour la visualisation sous forme de graphe,
 - Dans cet exemple :
 - <https://codebeautify.org/xmlviewer>

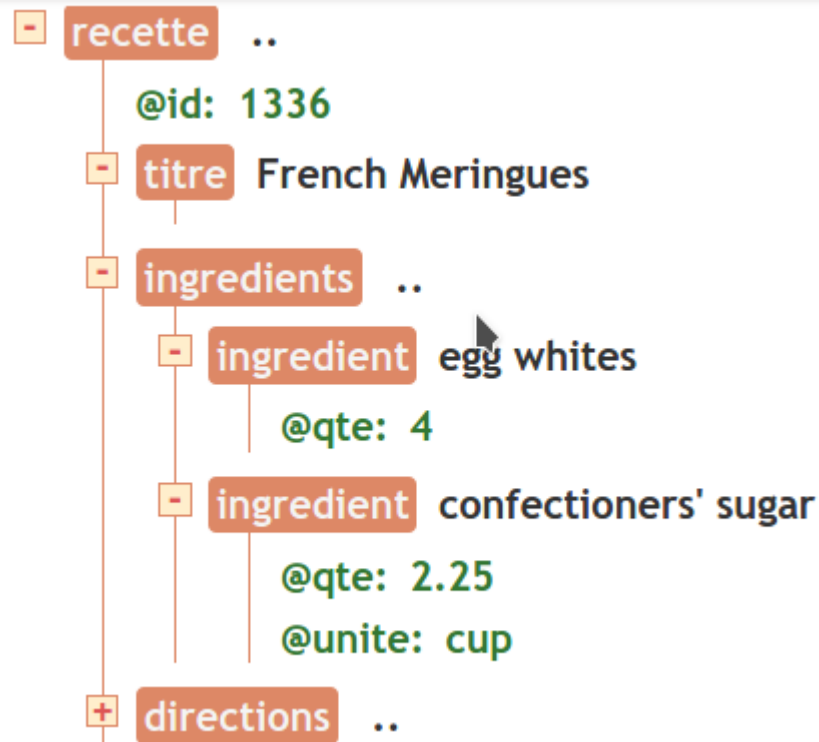
Introduction à XML

- Un deuxième exemple



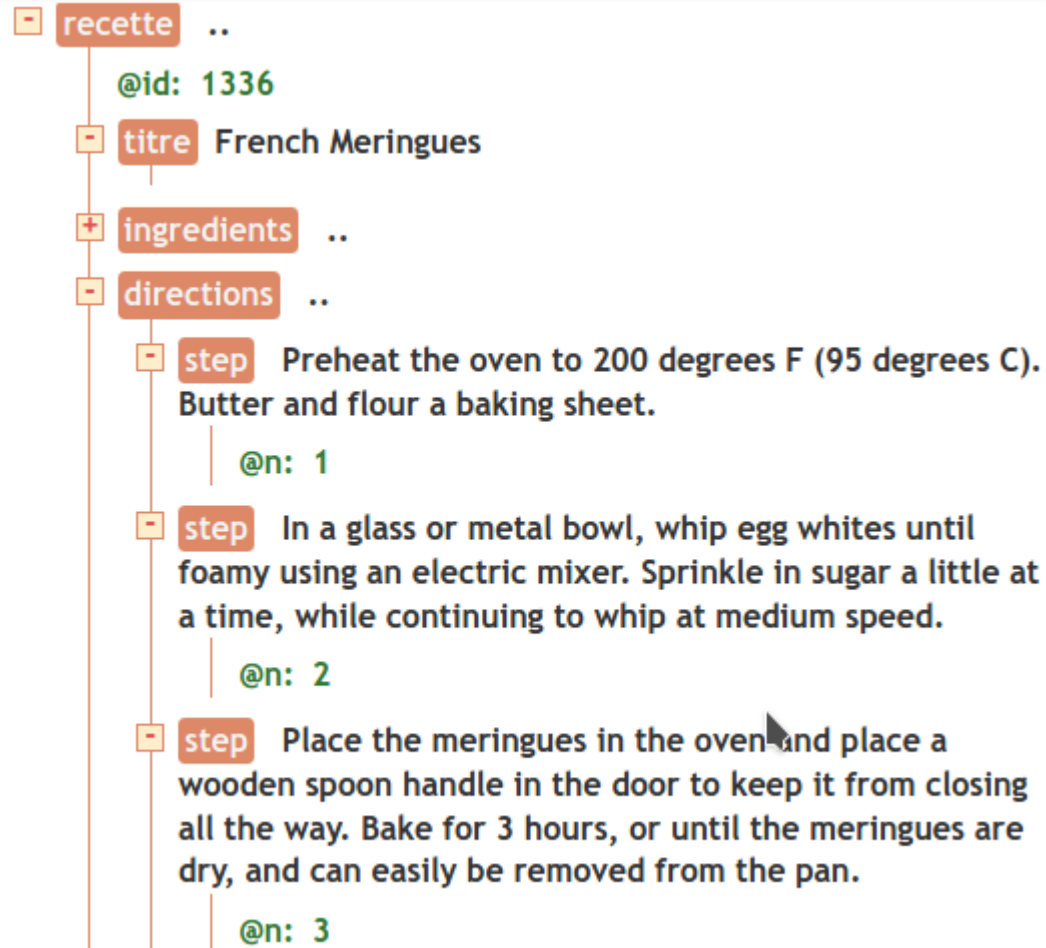
Introduction à XML

- Un deuxième exemple



Introduction à XML

- Un d



Plan

- Introduction à XML
- Structure XML de base
- **Domaines nominaux**
- Schémas XML

Domaines nominaux

- Liberté en définition des balises :
 - Risques des conflits au moment de fusion,
- Exemple :
 - Livre
 - Client
 - Commande (?)

Domaines nominaux

```
<?xml version="1.0" ?>
<livre>
  <titre>Java</titre>
  <auteurs>
    <auteur>
      <nom>Jamie Chan</nom>
      <nationalite>américaine</nationalite>
    </auteur>
  </auteurs>
  <prix>11.99</prix>
</livre>
```

Domaines nominaux

```
<?xml version="1.0" ?>  
<client>  
  <code>1226</code>  
  <nom>Benmohammed Mohammed</nom>  
  <adresse>Jijel</adresse>  
</client>
```

Domaines nominaux

```
<?xml version="1.0" ?>
```

```
<commande>
```

```
  <client>
```

```
    <code>1226</code>
```

```
    <nom>Benmohammed Mohammed</nom>
```

```
    <adresse>Jijel</adresse>
```

```
  </client>
```

```
  <livres>
```

```
    <livre>
```

```
      <titre>Java</titre>
```

```
      <auteurs>
```

```
        <auteur>
```

```
          <nom>Jamie Chan</nom>
```

```
          <nationalite>américaine</nationalite>
```

```
        </auteur>
```

```
      </auteurs>
```

```
      <prix>11.99</prix>
```

```
    </livre>
```

```
  </livres>
```

```
</commande>
```

Domaines nominaux

- Namespaces
 - Définition de collections de nom,
 - Applicables sur les balises et les attributs
 - Chaque collection est définie par un URI,
 - L'URI est associé à un préfixe
 - xmlns:préfixe="URI"
 - le préfixe doit être déclaré par un ascendant

Domaines nominaux

```
<?xml version="1.0" ?>
```

```
<commande
```

```
  xmlns="http://www.univ-jijel.dz/commandes/"
```

```
  xmlns:l="http://www.univ-jijel.dz/livres/"
```

```
  xmlns:c="http://www.univ-jijel.dz/clients/">
```

```
  <c:client>
```

```
    <c:code>1226</c:code>
```

```
    <c:nom>Benmohammed Mohammed</c:nom>
```

```
    <c:adresse>Jijel</c:adresse>
```

```
  </c:client>
```

```
  <livres>
```

```
    <l:livre>
```

```
      <l:titre>Java</l:titre>
```

```
      <l:auteurs>
```

```
        <l:auteur>
```

```
          <l:nom>Jamie Chan</l:nom>
```

```
          <l:nationalite>américaine</l:nationalite>
```

```
        </l:auteur>
```

```
      </l:auteurs>
```

```
      <l:prix>11.99</l:prix>
```

```
    </l:livre>
```

```
  </livres>
```

```
15/03/23</commande>
```

Données Semi-Structurées

Plan

- Introduction à XML
- Structure XML de base
- Domaines nominaux
- **Schémas XML**

Schémas XML

- Pourquoi ?
 - Éléments autorisés
 - Définition de la hiérarchie
 - Attributs autorisés
 - Modèle de la valeur d'un attribut

DTD

- 5 « composantes » à définir :
 - Élement,
 - Attribut,
 - Entités
 - L'ensemble des caractères spéciaux
 - PCDATA
 - parsed character data : va être parsé
 - CDATA
 - character data : ne va pas être parsé

DTD

- Élément
 - Déclaration
 - `<!ELEMENT nom_element catégorie>`
 - `<!ELEMENT nom_element (sous_elements)>`

DTD

- Élément
 - Opérateurs :
 - + : minimum un,
 - * : 0 ou plusieurs,
 - ? : 0 ou 1,
 - , : liste d'éléments,
 - | : OU,
 - #PCDATA
 - EMPY (sans parenthèses)
 - ANY (sans parenthèses)

DTD

- Attributs
 - Déclaration
 - `<!ATTLIST nom_element nom_attribut type_attribut valeur_attribut>`

DTD

- Attributs
 - Types Attributs
 - CDATA : chaîne de caractères,
 - (val1|val2|val3) : une valeur de la liste,
 - ID : valeur unique,
 - ...

DTD

- Attributs
 - Valeurs Attributs
 - « valeur » : valeur par défaut,
 - #REQUIRED : obligatoire,
 - #IMPLIED : optionnel,
 - #FIXED « valeur » : valeur fixe.

DTD

- Exemple :

- Recette :

```
<!DOCTYPE recette [  
  <!ELEMENT recette (titre,ingredients,directions)>  
  <!ELEMENT titre (#PCDATA)>  
  <!ELEMENT ingredients (ingredient+)>  
  <!ELEMENT ingredient (#PCDATA)>  
  <!ELEMENT directions (step+)>  
  <!ELEMENT step (#PCDATA)>  
  
  <!ATTLIST recette id CDATA #REQUIRED>  
  <!ATTLIST ingredient qte CDATA #REQUIRED>  
  <!ATTLIST ingredient unite CDATA #IMPLIED>  
  <!ATTLIST step n CDATA #REQUIRED>  
  

```


DTD

- Déclaration
 - Début du document
 - Dans un fichier séparé
 - `<!DOCTYPE recette SYSTEM "recette.dtd">`

XSD : XML Schema

- Le schéma est lui même un fichier XML,
- Permet de définir des types plus complexes,
- Permet de préciser des types de données :
 - Entier,
 - Date,
 - ...

XSD : XML Schema

- Exemple : Recette

```
<xs:schema attributeFormDefault="unqualified" elementFormDefault="qualified"
xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <xs:element name="recette">
    <xs:complexType>
      <xs:sequence>
        <xs:element type="xs:string" name="titre"/>
        <xs:element name="ingredients">
          <xs:complexType>
            <xs:sequence>
              <xs:element name="ingredient" maxOccurs="unbounded" minOccurs="0">
                <xs:complexType>
                  <xs:simpleContent>
                    <xs:extension base="xs:string">
                      <xs:attribute type="xs:float" name="qte" use="optional"/>
                      <xs:attribute type="xs:string" name="unite" use="optional"/>
                    </xs:extension>
                  </xs:simpleContent>
                </xs:complexType>
              </xs:element>
            </xs:sequence>
          </xs:complexType>
        </xs:element>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
```

XSD : XML Schema

- Exemple : Recette (suite)

```
<xs:element name="directions">
  <xs:complexType>
    <xs:sequence>
      <xs:element name="step" maxOccurs="unbounded" minOccurs="0">
        <xs:complexType>
          <xs:simpleContent>
            <xs:extension base="xs:string">
              <xs:attribute type="xs:byte" name="n" use="optional"/>
            </xs:extension>
          </xs:simpleContent>
        </xs:complexType>
      </xs:element>
    </xs:sequence>
  </xs:complexType>
</xs:element>
<xs:attribute type="xs:short" name="id"/>
</xs:complexType>
</xs:element>
</xs:schema>
```

Université de Jijel
Faculté des Sciences Exactes et d'Informatique
Département d'Informatique
L3 – Systèmes Informatiques

Données Semi-Structurées

Chapitre 03

Noyau XML

Tarek Boutefara
t_boutefara@univ-jijel.dz
2020/2021