


-Introduction à XML-

Ichrak MEHREZ
(m_ichrak@hotmail.fr)

XML: Présentation

- Le langage XML (eXtensible Markup Language)
- « Langage de balisage extensible »
- « *Un langage de balisage* » : s'écrit grâce à des **balises**. Ces balises permettent de structurer de manière hiérarchisée et organisée les données d'un document.
- « *extensible* » : on a la possibilité de créer nos propres balises. Nous ne sommes pas obligés d'utiliser un ensemble de balises existantes comme c'est par exemple le cas en HTML.

```
<?xml version="1.0"?>
<questionnaire>
  <question>
    Qui était le premier
    empereur romain ?
  </question>
  <réponse>
    Auguste
  </réponse>
  <!-- Note : tu auras besoin
    de plus de questions.-->
</questionnaire>
```



XML: Présentation

- Le langage XML (eXtensible Markup Language) est un format général de documents orienté texte. Il s'est imposé comme un standard incontournable de l'informatique.
- Le XML a été créé pour faciliter les échanges de données entre les machines et les logiciels.
- Le XML est utilisé pour la transmission de données entre applications
- Le XML est un langage qui s'écrit à l'aide de **balises**.



XML: Présentation

```
<?xml version="1.0"?>
<questionnaire>
  <question>
    Qui était le premier
    empereur romain ?
  </question>
  <réponse>
    Auguste
  </réponse>
  <!-- Note : tu auras besoin
    de plus de questions.-->
</questionnaire>
```

XML

- Le XML est une recommandation du **W3C**, il s'agit donc d'une technologie avec des règles strictes à respecter.
- Le XML nous permet de créer notre propre vocabulaire grâce à un ensemble de règles et de balises personnalisables.
- Le XML est flexible, simple et extensible, ce qui a permis de l'adapter à de multiples domaines allant des données géographiques au dessin vectoriel en passant par les échanges commerciaux.
- De nombreuses technologies se sont développées autour de XML et enrichissent ainsi son environnement.

Syntaxe XML

- 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 et plus strictes.
- Elles sont plus générales car les *noms des balises sont libres*. Elles sont aussi plus strictes car elles imposent qu'*à toute balise ouvrante corresponde une balise fermante*.

Les balises

- les balises sont les éléments de base d'un document XML.
- Une balise porte un **nom** qui est entouré de **chevrons** (< , >).
- Balises par paires: composées de 2 balises: *ouvrante* et *fermante*
<NomBalise> contenu </NomBalise>
- Balises uniques: balise sans contenu **<NomBalise/>**

```
<Etudiant>  
  <Nom> Tounsi </Nom>  
  <prenom> Mohamed </prenom>  
  <cin>12345678</cin>  
  <passport/>  
</Etudiant>
```

Les balises : les noms

- Les noms peuvent contenir des lettres, des chiffres ou des caractères spéciaux.
- Les noms ne peuvent pas débiter par un nombre ou un caractère de ponctuation.
- Les noms ne peuvent pas commencer par les lettres XML (quelle que soit la casse).
- Les noms ne peuvent pas contenir d'espaces.
- On évitera les caractères -, ; , < et > qui peuvent être mal interprétés dans vos programmes.

Les attributs

- Il est possible d'ajouter à nos balises ce qu'on appelle des **attributs**.
- Tout comme pour les balises, c'est vous qui en choisissez le nom.
- Un **attribut** peut se décrire comme une option ou une donnée cachée.
- Ce n'est pas l'information principale que souhaite transmettre la balise, mais il donne des renseignements supplémentaires sur son contenu.

```
<prix moyen_paiement="chèque" devise="euro">25.3</prix>
```


Les commentaires

- Un **commentaire** est un texte qui permet de donner une indication sur ce que l'on fait.
- Il vous permet d'annoter votre fichier et d'expliquer une partie de celui-ci.
- En XML, les commentaires ont une syntaxe particulière. C'est une balise unique qui commence par **<!--** et qui se termine par **-->**.

```
<!-- ceci est un commentaire -->
```

Premier exemple

```
1  <?xml version="1.0" encoding="UTF-8"?>
2
3  <!-- Exemple d'un fichier XML -->
4  <Groupe>
5      <Etudiant>
6          <Nom> Tounsi </Nom>
7          <prenom> Mohamed </prenom>
8          <cin>12345678</cin>
9          <passseport/>
10     </Etudiant>
11
12     <Etudiant>
13         <Nom> Ben Salah </Nom>
14         <prenom> Amin </prenom>
15         <cin>98765432</cin>
16         <passseport type="P">Y123456</passseport>
17     </Etudiant>
18 </Groupe>
19
```

Premier exemple

```
1  <?xml version="1.0" encoding="UTF-8"?>
2
3  <!-- Exemple d'un fichier XML -->
4  <Groupe>
5    <Etudiant>
6      <Nom> Tounsi </Nom>
7      <prenom> Mohamed </prenom>
8      <cin>12345678</cin>
9      <passport/>
10   </Etudiant>
11
12   <Etudiant>
13     <Nom> Ben Salah </Nom>
14     <prenom> Amin </prenom>
15     <cin>98765432</cin>
16     <passport type="P">Y123456</passport>
17   </Etudiant>
18 </Groupe>
19
```

Entête XML avec la version 1.0 et l'encodage UTF-8 des caractères.

Premier exemple

```
1  <?xml version="1.0" encoding="UTF-8"?>
2
3  <!-- Exemple d'un fichier XML -->
4  <Groupe>
5      <Etudiant>
6          <Nom> Tounsi </Nom>
7          <prenom> Mohamed </prenom>
8          <cin>12345678</cin>
9          <passport/>
10     </Etudiant>
11
12     <Etudiant>
13         <Nom> Ben Salah </Nom>
14         <prenom> Amin </prenom>
15         <cin>98765432</cin>
16         <passport type="P">Y123456</passport>
17     </Etudiant>
18 </Groupe>
19
```

Commentaire délimité
par les chaînes de
caractères <!-- et -->.

Premier exemple

```
1  <?xml version="1.0" encoding="UTF-8"?>
2
3  <!-- Exemple d'un fichier XML -->
4  <Groupe>
5      <Etudiant>
6          <Nom> Tounsi </Nom>
7          <prenom> Mohamed </prenom>
8          <cin>12345678</cin>
9          <passport/>
10     </Etudiant>
11
12     <Etudiant>
13         <Nom> Ben Salah </Nom>
14         <prenom> Amin </prenom>
15         <cin>98765432</cin>
16         <passport type="P">Y123456</passport>
17     </Etudiant>
18 </Groupe>
19
```

Balise ouvrante de
l'élément *racine* Groupe

Premier exemple

```
1  <?xml version="1.0" encoding="UTF-8"?>
2
3  <!-- Exemple d'un fichier XML -->
4  <Groupe>
5    <Etudiant>
6      <Nom> Tounsi </Nom>
7      <prenom> Mohamed </prenom>
8      <cin>12345678</cin>
9      <passport/>
10   </Etudiant>
11
12   <Etudiant>
13     <Nom> Ben Salah </Nom>
14     <prenom> Amin </prenom>
15     <cin>98765432</cin>
16     <passport type="P">Y123456</passport>
17   </Etudiant>
18 </Groupe>
19
```

Balise ouvrante de l'élément *passport* avec un attribut de nom *type* et de valeur *P*

Premier exemple

```
1  <?xml version="1.0" encoding="UTF-8"?>
2
3  <!-- Exemple d'un fichier XML -->
4  <Groupe>
5      <Etudiant>
6          <Nom> Tounsi </Nom>
7          <prenom> Mohamed </prenom>
8          <cin>12345678</cin>
9          <passseport/>
10     </Etudiant>
11
12     <Etudiant>
13         <Nom> Ben Salah </Nom>
14         <prenom> Amin </prenom>
15         <cin>98765432</cin>
16         <passseport type="P">Y123456</pas
17     </Etudiant>
18 </Groupe>
19
```

Balise fermante de
l'élément racine Groupe

Structure d'un document XML

- Un document XML peut être découpé en 2 parties :
 - le **prologue**
 - le **corps**

Structure d'un document XML

Le Prologue

- **Le prologue:** la première ligne de votre document XML. Il donne des informations de traitement.
- Le prologue est une balise unique qui commence par **<?xml** et qui se termine par **?>**

```
<?xml version = "1.0" encoding="UTF-8" standalone="yes" ?>
```

Structure d'un document XML

Le Prologue

```
<?xml version = "1.0" encoding="UTF-8" standalone="yes" ?>
```

- ❑ **version:** permet d'indiquer la version de XML que l'on utilise pour décrire nos données
- ❑ **encoding:** Il s'agit du **jeu de caractères** utilisé dans mon document XML. Par défaut, l'encodage de XML est l'UTF-8
- ❑ **standalone:** permet de savoir si votre document XML est autonome ou si un autre document lui est rattaché.

Structure d'un document XML

Le Corps

- Le **corps** d'un document XML est constitué de l'ensemble des balises qui décrivent les données.
- *NB: une balise en paires unique doit contenir toutes les autres.* Cette balise est appelée **élément racine** du corps.
- Le plus souvent, la **balise racine** aura pour mission de décrire ce qu'elle contient.

```
<racine>  
  <balise1>texte</balise1>  
    <balise11>texte</balise11>  
    <balise12>texte</balise12>  
  <balise2>texte</balise2>  
</racine>
```

Un document bien formé (well-formed)

- Décrire un document XML qui est *conforme aux règles syntaxiques*:
 - ❑ Le prologue est bien renseigné.
 - ❑ Le document XML ne possède qu'une seule balise racine.
 - ❑ Le nom des balises et des attributs est conforme aux règles de nommage.
 - ❑ Toutes les balises en paires sont correctement fermées.
 - ❑ Toutes les valeurs des attributs sont entre guillemets simples ou doubles.
 - ❑ Les balises ne se chevauchent pas, il existe une arborescence dans votre document

A voir

- les **fichiers de définition**: définir une structure stricte aux documents XML : les **DTD** et les **schémas XML**.
- Extraire et exploiter les informations contenues dans un document XML: l'**API DOM** et **XPath**.
- Transformer les informations d'un document XML vers un autre type de document comme un autre document XML ou encore une page web: **XSLT**