

Le format XML

16 janvier 2014

emeline.dorey@gmail.com

Sommaire

Le format XML

Historique

SGML
HTML
XML

Ce qu'XML va permettre

Exemple

Mise en oeuvre

Structure

Présentation
Le prologue
L'arbre d'éléments
Les commentaires
Bien former le document
Les sections littérales

Exemple complet

- 1 Historique
 - SGML
 - HTML
 - XML
- 2 Ce qu'XML va permettre
- 3 Exemple
- 4 Mise en oeuvre
- 5 Structure
 - Présentation
 - Le prologue
 - L'arbre d'éléments
 - Les commentaires
 - Bien former le document
 - Les sections littérales
- 6 Exemple complet

Historique : de SGML à XML

SGML

Le format XML

Historique

SGML
HTML
XML

Ce qu'XML va
permettre

Exemple

Mise en oeuvre

Structure

Présentation
Le prologue
L'arbre d'éléments
Les commentaires
Bien former le
document
Les sections littérales

Exemple complet

- SGML (Standard Generalized Markup Language)
 - Langage de balisage standard généralisé.
 - Adopté comme standard en 1986.
 - Trop général ⇒ difficile d'apprentissage et complexe d'emploi.
 - Inadapté à l'écriture de documents pour Internet
- Nécessité d'en dériver le langage HTML.

Historique : de SGML à XML

HTML

Le format XML

Historique

SGML
HTML
XML

Ce qu'XML va
permettre

Exemple

Mise en oeuvre

Structure

Présentation
Le prologue
L'arbre d'éléments
Les commentaires
Bien former le
document
Les sections littérales

Exemple complet

- HTML (HyperText Markup Language)
 - Langage de balisage permettant d'écrire de l'hypertexte.
 - Standard de 1995 à 1999.
 - Champ d'action limité : pas possible de définir autre chose qu'une page Web.
 - Langage conçu uniquement pour ça.
- HTML rencontre ses limites ⇒ demande de plus en plus forte pour définir un nouveau langage.

Historique : de SGML à XML

XML

Le format XML

Historique

SGML
HTML
XML

Ce qu'XML va
permettre

Exemple

Mise en oeuvre

Structure

Présentation
Le prologue
L'arbre d'éléments
Les commentaires
Bien former le
document
Les sections littérales

Exemple complet

- XML (eXtensible Markup Language)
 - Langage de balisage générique.
 - Tente d'allier la simplicité du HTML et la souplesse SGML.
 - Sert essentiellement à stocker/transférer des données de type texte, structurées en champs arborescents.
 - "eXensible" car il permet à l'utilisateur de définir son balisage.
- Langage adapté à beaucoup plus d'usages que ses concepteurs ne le pensaient au départ.

Ce qu'XML va permettre

Le format XML

Historique

SGML
HTML
XML

Ce qu'XML va permettre

Exemple

Mise en oeuvre

Structure

Présentation
Le prologue
L'arbre d'éléments
Les commentaires
Bien former le document
Les sections littérales

Exemple complet

- Dans un document XML, la mise en forme des données est totalement séparée des données elles-mêmes \Rightarrow séparation totale de l'information (le contenu) de son apparence (le contenant).
- XML va permettre aux logiciels de comprendre/exploiter au mieux le contenu de ces pages, rendu désormais explicite par un balisage spécifique, indépendant de toute application.

Ce qu'XML va permettre

Le format XML

Historique

SGML
HTML
XML

Ce qu'XML va permettre

Exemple

Mise en oeuvre

Structure

Présentation
Le prologue
L'arbre d'éléments
Les commentaires
Bien former le document
Les sections littérales

Exemple complet

XML va permettre aux utilisateurs :

- D'avoir un fichier plus lisible et modifiable :
 - informations étiquetées comme on veut,
 - ordonnancement laissé au choix de l'utilisateur.
- De saisir/mettre à jour une seule fois l'information pure (le contenu) :
 - sans s'occuper de la présentation ou des traitements futurs,
 - sans saisir de libellés, comme on alimenterait une base de données.

Ce qu'XML va permettre

Le format XML

Historique

SGML
HTML
XML

Ce qu'XML va permettre

Exemple

Mise en oeuvre

Structure

Présentation
Le prologue
L'arbre d'éléments
Les commentaires
Bien former le document
Les sections littérales

Exemple complet

- D'en générer automatiquement :
 - de multiples présentations (tableau, graphique, image, HTML, PDF...)
 - éventuellement, couplé aux feuilles de style XSLT : des tris, des sélections, des réorganisations, des générations automatiques de libellés, de tables des matières, index...
 - le tout sur de multiples médias (écran de téléphone portable, ordinateur de bureau, base de données, papier...)

Conclusion

Tout cela grâce à l'indépendance du balisage par rapport à la présentation.

Exemple : bibliographie

Code source HTML

Le format XML

Historique

SGML
HTML
XML

Ce qu'XML va
permettre

Exemple

Mise en oeuvre

Structure

Présentation
Le prologue
L'arbre d'éléments
Les commentaires
Bien former le
document
Les sections littérales

Exemple complet

```
<h2>Bibliographie XML</h2>
<ul>
  <li>Thierry Boulanger et Sebastien Lecomte,
  <br /><i>XML par la pratique - Bases
  indispensables, concepts et cas pratiques</i>,
  <br />Paris, Editions ENI, 2008</li>
  <li>Alexandre Brillant,
  <br /><i>XML : Cours et exercices</i>,
  <br />Paris, Eyrolles, 2007 </li>
  <li>Kevin Howard Goldberg,
  <br /><i>Manuel de Prise en Main de XML</i>,
  <br />Paris, Pearson Education, 2009</li>
</ul>
```

Exemple : bibliographie

Résultat sous Firefox

Le format XML

Historique

SGML
HTML
XML

Ce qu'XML va
permettre

Exemple

Mise en oeuvre

Structure

Présentation
Le prologue
L'arbre d'éléments
Les commentaires
Bien former le
document
Les sections littérales

Exemple complet

Bibliographie XML

- Thierry Boulanger et Sebastien Lecomte,
XML par la pratique - Bases indispensables, concepts et cas pratiques,
Paris, Editions ENI, 2008
- Alexandre Brillant,
XML : Cours et exercices,
Paris, Eyrolles, 2007
- Kevin Howard Goldberg,
Manuel de Prise en Main de XML,
Paris, Pearson Education, 2009

Exemple : bibliographie

Une partie en XML

Le format XML

Historique

SGML
HTML
XML

Ce qu'XML va
permettre

Exemple

Mise en oeuvre

Structure

Présentation
Le prologue
L'arbre d'éléments
Les commentaires
Bien former le
document
Les sections littérales

Exemple complet

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<biblio>
  <livre langue="fr" sujet="xml">
    <titre>XML : Cours et exercices</titre>
    <auteur>
      <nom>Brillant</nom>
      <prenom>Alexandre</prenom>
    </auteur>
    <editeur>
      <nom>Eyrolles</nom>
      <ville>Paris</ville>
    </editeur>
    <annee>2007</annee>
  </livre>
</biblio>
```

Mise en oeuvre

Le format XML

Exemple : production d'un document HTML à partir de données formatées en XML.

- Il faut écrire au moins deux fichiers :
 - Dans l'un, les données à mettre en forme ⇒ document XML.
 - Dans l'autre, les informations pour la mise en forme ⇒ document XSLT.
- Un troisième fichier peut être nécessaire pour définir les balises reconnues pour ce document XML :
 - Une DTD (Document Type Definition)
 - Un schéma.
 - Ce fichier sera par exemple écrit pour contraindre la saisie/mise à jour du document XML.
- Pour le document HTML, mieux vaut ajouter une feuille de style CSS.

Historique

SGML
HTML
XML

Ce qu'XML va
permettre

Exemple

Mise en oeuvre

Structure

Présentation
Le prologue
L'arbre d'éléments
Les commentaires
Bien former le
document
Les sections littérales

Exemple complet

Mise en oeuvre

Le format XML

Historique

SGML
HTML
XML

Ce qu'XML va
permettre

Exemple

Mise en oeuvre

Structure

Présentation
Le prologue
L'arbre d'éléments
Les commentaires
Bien former le
document
Les sections littérales

Exemple complet

Résumé

Pour produire un fichier HTML à partir de données formatées en XML :

- 1/ Créer un fichier définissant les balises utilisables : DTD (Optionnel)
- 2/ Créer un fichier de données XML.
- 3/ Créer une feuille de style XSL (permet de produire le fichier HTML).
- 4/ Créer une feuille de style CSS (Optionnel).

Structure d'un document XML

Présentation

Le format XML

Historique

SGML
HTML
XML

Ce qu'XML va
permettre

Exemple

Mise en oeuvre

Structure

Présentation

Le prologue
L'arbre d'éléments
Les commentaires
Bien former le
document
Les sections littérales

Exemple complet

- Un fichier XML doit être composé :
 - d'un prologue,
 - d'un arbre d'éléments,
 - de commentaires et instructions de traitement (facultatif).

Structure d'un document XML

Le prologue

Le format XML

Historique

SGML
HTML
XML

Ce qu'XML va
permettre

Exemple

Mise en oeuvre

Structure

Présentation
Le prologue
L'arbre d'éléments
Les commentaires
Bien former le
document
Les sections littérales

Exemple complet

- La déclaration XML :

```
<?xml version='1.0' encoding='ISO-8859-1' standalone='yes'?'>
```

- xml version="1.0" : décrit la version utilisée
 - encoding="ISO-8859-1" : code les caractères utilisés dans le document
 - standalone : indique si le document fait référence à d'autre documents (optionnel). Par défaut il est à "no"
- Cette déclaration fait partie des instructions de traitements.

Structure d'un document XML

Le prologue

Le format XML

Historique

SGML
HTML
XML

Ce qu'XML va
permettre

Exemple

Mise en oeuvre

Structure

Présentation
Le prologue
L'arbre d'éléments
Les commentaires
Bien former le
document
Les sections littérales

Exemple complet

Cette déclaration XML :

- les attributs "version", "encoding" et "standalone" doivent être placés dans cet ordre ;
- elle doit être positionnée en toute première ligne du document XML. Par exemple, il ne faut ni commentaire, ni même une simple ligne code avant elle.

Structure d'un document XML

L'arbre d'éléments

Le format XML

Historique

SGML
HTML
XML

Ce qu'XML va
permettre

Exemple

Mise en oeuvre

Structure

Présentation
Le prologue
L'arbre d'éléments
Les commentaires
Bien former le
document
Les sections littérales

Exemple complet

- Un document a un unique élément racine :
 - c'est la base du document XML,
 - il est unique et englobe tous les autres éléments,
 - il s'ouvre juste après le prologue et se ferme à la toute fin du document.
- Dans notre exemple, l'élément racine est **biblio**.

Structure d'un document XML

Les éléments

Le format XML

Historique

- SGML
- HTML
- XML

Ce qu'XML va permettre

Exemple

Mise en oeuvre

Structure

- Présentation
- Le prologue

L'arbre d'éléments

- Les commentaires

- Bien former le document

- Les sections littérales

Exemple complet

- Les éléments forment la structure même du document : ce sont les branches et les feuilles de l'arborescence.
- Un élément :
 - possède un nom (son type) ;
 - il ne doit pas contenir d'espace, ni commencer par un chiffre
 - il peut comprendre des lettres de l'alphabet, des chiffres et les caractères - et _
 - contient du texte (une chaîne de caractères), des attributs (autres éléments), des instructions de traitement ;
 - il ne peut pas avoir deux attributs de même nom ;
 - il peut aussi être vide.
- L'élément contenant est nommé "élément parent".

Structure d'un document XML

L'arbre d'éléments

Le format XML

Historique

SGML
HTML
XML

Ce qu'XML va permettre

Exemple

Mise en oeuvre

Structure

Présentation
Le prologue

L'arbre d'éléments

Les commentaires

Bien former le document

Les sections littérales

Exemple complet

Exemple d'élément contenant du texte :

```
<titre>XML : Cours et exercices</titre>
```

Exemple d'élément contenant d'autres éléments :

```
<auteur>  
  <nom>Brillant</nom>  
  <prenom>Alexandre</prenom>  
</auteur>
```

Exemple d'élément vide (il ne contient pas d'élément enfant) :

```
<couverture couleur='rouge' />
```

Structure d'un document XML

Les attributs

Le format XML

Historique

SGML
HTML
XML

Ce qu'XML va
permettre

Exemple

Mise en oeuvre

Structure

Présentation
Le prologue
L'arbre d'éléments
Les commentaires
Bien former le
document
Les sections littérales

Exemple complet

- Un attribut permet de stocker des informations supplémentaires sur un élément, sans ajouter de texte au contenu de l'élément lui-même.
- Tous les éléments peuvent contenir de 0 à n attributs uniques.
- Un attribut a un nom et une valeur.
- Un attribut ne peut être présent que dans la balise ouvrante de l'élément (par exemple, il est interdit d'écrire `</livre language="en">`).

Structure d'un document XML

Les attributs

Le format XML

Historique

- SGML
- HTML
- XML

Ce qu'XML va permettre

Exemple

Mise en oeuvre

Structure

- Présentation
- Le prologue

L'arbre d'éléments

- Les commentaires
- Bien former le document
- Les sections littérales

Exemple complet

Exemple d'un élément avec attribut :

```
<fruit coeur='noyau'>abricot</fruit>
```

Exemple d'élément vide avec attributs :

```
<img src='lapin.png' alt='Petit lapin'  
width='50' height='50' />
```

Structure d'un document XML

Elément/Attribut

Le format XML

Historique

SGML
HTML
XML

Ce qu'XML va
permettre

Exemple

Mise en oeuvre

Structure

Présentation
Le prologue
L'arbre d'éléments
Les commentaires
Bien former le
document
Les sections littérales

Exemple complet

Faut-il plutôt déclarer un élément ou un attribut ?

- Le choix de la structure d'un document XML n'est pas complètement arbitraire. Il faut réfléchir à l'utilisation des informations :
 - information sur le contenu du document XML ⇒ préférer l'attribut,
 - information à utiliser régulièrement ⇒ préférer l'élément.

Structure d'un document XML

Les commentaires

Le format XML

Historique

SGML
HTML
XML

Ce qu'XML va
permettre

Exemple

Mise en oeuvre

Structure

Présentation
Le prologue
L'arbre d'éléments
Les commentaires
Bien former le
document
Les sections littérales

Exemple complet

- Les commentaires XML se déclarent de la même manière qu'en HTML :
 - ils commencent par `<!--`
 - ils se terminent par `-->`
- Ils peuvent être placés n'importe où, du moment qu'ils sont à l'extérieur d'une autre balise.

Exemple de commentaires valides :

```
<!--commentaire correct -->  
<texte><!-- commentaire correct -->  
petit texte</texte>
```

Structure d'un document XML

Bien former le document

Le format XML

Historique

SGML
HTML
XML

Ce qu'XML va permettre

Exemple

Mise en oeuvre

Structure

Présentation
Le prologue
L'arbre d'éléments
Les commentaires

Bien former le document

Les sections littérales

Exemple complet

Le document doit également respecter des critères syntaxiques : la casse, des cotes simples ou doubles autour des attributs...

Exemples :

- `<LIVRE>...<livre>` : mal formé
- `<livre>...</livre>` : bien formé
- `<livre langue=fr>...</livre>` : mal formé
- `<livre langue="fr">...</livre>` : bien formé

Structure d'un document XML

Bien former le document

Le format XML

Historique

SGML
HTML
XML

Ce qu'XML va
permettre

Exemple

Mise en oeuvre

Structure

Présentation
Le prologue
L'arbre d'éléments
Les commentaires

Bien former le
document

Les sections littérales

Exemple complet

Un document est bien formé si c'est un arbre d'éléments respectant les règles précédentes.

Exemple d'un document mal formé (on a deux arbres) :

```
<?xml version=' '1.0' '>  
<livre><auteur>Zola</auteur></livre>  
<livre><auteur>Dumas</auteur></livre>
```

Exemple acceptable en HTML et mal formé en XML
(imbrication incorrecte des balises) :

```
<p>petit texte en <i>italique</p></i>
```

Structure d'un document XML

Les sections littérales

Le format XML

- Il n'est pas permis de placer directement dans le contenu d'un document XML des caractères tels que "<", ">" ou "&".
- Il faut utiliser un section CDATA. Elle permet d'inclure du texte qui n'est pas analysé par le parseur.
- Cette instruction est à utiliser avec parcimonie.

Exemple incorrect :

```
<test>if(i<4 && j>10){println(i);}</test>
```

Exemple correct :

```
<test><![CDATA[if(i<4 && j>10){println(i);}]]></test>
```

Historique

SGML
HTML
XML

Ce qu'XML va permettre

Exemple

Mise en oeuvre

Structure

Présentation
Le prologue
L'arbre d'éléments
Les commentaires
Bien former le document
Les sections littérales

Exemple complet

Exemple d'un document XML

Le format XML

Historique

SGML
HTML
XML

Ce qu'XML va
permettre

Exemple

Mise en oeuvre

Structure

Présentation
Le prologue
L'arbre d'éléments
Les commentaires
Bien former le
document
Les sections littérales

Exemple complet

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

<!-- Element racine -->
<biblio>
  <!-- Premier enfant -->
  <livre langue="fr" sujet="xml">
    <!-- Element enfant titre -->
    <titre>XML par la pratique - Bases
      indispensables, concepts et cas pratiques</
      titre>
    <auteur>
      <nom>Boulangier</nom>
      <prenom>Thierry</prenom>
    </auteur>
```

Exemple d'un document XML

Le format XML

```
<editeur>
  <nom>Editions ENI</nom>
  <ville>Paris</ville>
</editeur>
<annee>2008</annee>
</livre>
<!-- Second enfant -->
<livre langue="fr" sujet="exemples">
  <titre>XML : Cours et exercices</titre>
  <auteur>
    <nom>Brillant</nom>
    <prenom>Alexandre</prenom>
  </auteur>
```

Historique

SGML
HTML
XML

Ce qu'XML va
permettre

Exemple

Mise en oeuvre

Structure

Présentation
Le prologue
L'arbre d'éléments
Les commentaires
Bien former le
document
Les sections littérales

Exemple complet

Exemple d'un document XML

Le format XML

```
<editeur>
  <nom>Eyrolles</nom>
  <ville>Paris</ville>
</editeur>
<annee>2007</annee>
</livre>
<!-- Troisième enfant -->
<livre langue="fr" sujet="manuel">
  <titre>Manuel de Prise en Main de XML</titre>
  <auteur>
    <nom>Goldberg</nom>
    <prenom>Kevin Howard</prenom>
  </auteur>
```

Historique

SGML
HTML
XML

Ce qu'XML va
permettre

Exemple

Mise en oeuvre

Structure

Présentation
Le prologue
L'arbre d'éléments
Les commentaires
Bien former le
document
Les sections littérales

Exemple complet

Exemple d'un document XML

Le format XML

Historique

SGML
HTML
XML

Ce qu'XML va
permettre

Exemple

Mise en oeuvre

Structure

Présentation
Le prologue
L'arbre d'éléments
Les commentaires
Bien former le
document
Les sections littérales

Exemple complet

```
<editeur>
  <nom>Pearson Education</nom>
  <ville>Paris</ville>
</editeur>
<annee>2009</annee>
</livre>
</biblio>
```