

XML, Langage de balisage extensible

Rédaction : Pierre Attar

Recommandation(s) liée(s) : [Canonical](#) - [XInclude](#) - [Infoset](#) - [Fragment Interchange](#) - [XML Base](#) - [DTD](#) - [XML Language](#)

XML est un **langage de codage de données** dont l'objectif est, dans un échange entre systèmes informatiques, de transférer, en même temps, des données et leurs structures.

Permettant de coder n'importe quel type de donnée, depuis l'échange EDI jusqu'aux documents les plus complexes en passant par les échanges de données inter-applications, son potentiel est de devenir le standard universel et multilingue d'échange d'informations. Appliqué aux documents textuels, il permet d'identifier, de façon logique, la structure et l'organisation de l'information textuelle.

Objectifs

XML est un format textuel très flexible dérivé de SGML. Initialement conçu pour relever les défis de l'édition électronique de grande puissance, XML joue également un rôle de plus en plus important dans l'échange d'une grande variété de données, que ce soit sur le Web ou pour n'importe quel échange inter-applicatif.

XML permettra, comme le souligne le W3C :

- l'édition électronique internationalisée, de façon indépendante des logiciels et des systèmes ;
- aux industries de définir des protocoles, indépendants des logiciels et des systèmes, pour l'échange des données (particulièrement les données du commerce électronique) ;
- de fournir de l'information aux agents utilisateurs sous une forme qui permette un traitement automatique après réception, par exemple pour toutes les applications de téléphonie mobile ;
- de faciliter le développement logiciel dès lors qu'il s'agit de manipuler l'information spécialisée et répartie ;
- de faciliter les traitements de données avec des logiciels peu coûteux ; à ce titre, l'avenir d'XML et d'un certain nombre de ses recommandations associées (XSLT, XPath, Infoset, etc.) est de devenir partie intégrante des couches hautes des systèmes d'exploitation ;
- aux utilisateurs du Web d'afficher l'information reçue avec la feuille de styles qu'ils souhaitent ;
- de faciliter la fourniture de Métadonnées (données descriptives de documents) qui aide à trouver de l'information.

Principes

Un document XML se compose, d'une part, de texte, et, d'autre part, d'informations de structure. Les informations de structure servent le plus souvent à délimiter du texte, pour en identifier la sémantique. Ainsi, `<métier>Consultant</métier>` permet de dire que la chaîne de caractères "Consultant" doit être comprise comme étant une définition de métier. Il est possible de délimiter des chaînes de caractères ; il est aussi possible de délimiter tout ensemble d'informations mélangeant texte et structure. Par exemple, dans le document suivant :

```
<métier>
  <nom>
    Consultant
  </nom>
  <descr>
    Le consultant travaille pour des consultés qui ...
  </descr>
</métier>
```

La notion de métier introduit, d'une part, un nom de métier et, d'autre part, une description de métier. C'est l'appartenance hiérarchique qui définit que tout cela parle bien du même métier : elle permet de spécialiser des description : un nom de métier et une description de métier.

Pour compléter cette description, il est parfois nécessaire de *valuer* la signification d'un objet. Par exemple, `<auteur affiliation="rennes2">Jean Dupont</auteur>`, permet d'identifier un chaîne de caractères comme étant un auteur et, en plus, de décrire cet auteur, par *valuation*, comme appartenant à l'Université de Rennes II.

Les balises (`<maBalise> ... </maBalise>`) délimitent des objets typés ; les attributs (`<maBalise type="standard">`) définissent des *valuations* d'objets. Un document XML est alors un arbre d'objets typés et *valués*.

Pour finir, un document XML doit définir le jeu de caractères qu'il utilise, ainsi que la version de la recommandation XML. Du coup, l'exemple précédent s'écrira :

```
<?xml version="1.0" encoding="utf-8"?>
<métier>
  ...
</métier>
```

Ce document est parfaitement décrit, il représente, au sens de la recommandation, un document "bien formé" (*wellformed*). Si nécessaire, un modèle documentaire peut lui être adjoint (voir DTD). Celui-ci définira les contraintes associées à ce document : le document devra alors être "valide", au regard de ce modèle. Définis de façon électronique, les applicatifs de type *parsers* seront capables de valider, de façon automatisée, la conformité d'un document à sa classe, à son modèle. L'avantage ? Il sera possible d'appliquer des processus automatiques sur une classe de documents (les algorithmes étant écrits au regard de la classe et non pas au regard de chaque instance de document).

Les différents aspects de la recommandation XML

La recommandation XML s'intéresse à des notions fort différentes :

- un langage de codage, basé sur Unicode XML, de documents, avec des éléments, des attributs, un jeu de caractères, etc. ;
- un langage de définition de modèles documentaires, avec les DTD ;
- un langage d'expression d'inclusions de fichiers, avec les `internal` et `external subset` des DTD ;

l'expression d'informations applicatives, comme les deux attributs `xml:lang` et `xml:space`, dont le rôle est, respectivement, d'indiquer la langue de rédaction d'un fragment de contenu et la façon de traiter les caractères d'espacement d'un contenu ;
un langage de prise en compte extrêmement limité des espaces de noms.

Le fait de mettre tous ces aspects dans une même recommandation a un aspect politique important, car cela oblige à prendre *tout* en compte sans différenciation. D'un point de vue technique, cela peut parfois poser des problèmes de compréhension : par exemple, pourquoi prendre en compte seulement les notions de langue et pas celles d'URI ? Cela peut aussi poser des problèmes d'architecture d'application, dès lors que, par exemple, un [Schema](#) permet d'exprimer la même chose, et davantage qu'une [DTD](#)... sauf les notions d'inclusions de données spécifiées de façon indépendante des documents eux-mêmes, au travers des entités générales.

Recommandations(s)

- ■ *Le langage de balisage extensible (XML) 1.1*
Recommandation, version 1.1, du 04-02-2004
Document sur <http://www.yoyodesign.org/doc/w3c/xml11/>
- 🇺🇸 *Extensible Markup Language (XML) 1.1 (Second Edition)*
Recommandation, version 1.1, du 16-08-2006
Document sur <http://www.w3.org/TR/xml11>