

XML (*eXtensible Markup Language*) une synthèse

Christian Soutou
soutou@iut-blagnac.fr



Plan

- Généralités
- Langage
- Grammaires DTD
- Espaces de noms
- Flux d'informations (syndication de contenus)

• Généralités

« *La connaissance de XML, un prérequis indispensable à toute embauche* »

01 Informatique – 27 Août 2004

« *There is essentially no computer in the world, desktop, handheld, or backroom, that doesn't process XML sometimes...* »

Tim Bray – Sun – 2008

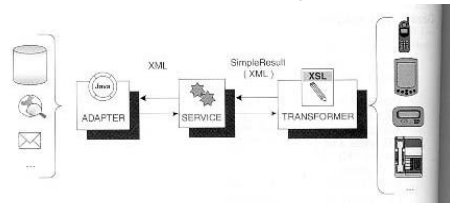
Généralités

- Langage de balises permettant de structurer des données en format texte
- Impulsé par Sun en 1986, standardisé depuis 1998 par le W3C (*World Wide Web Consortium*)
<http://www.w3.org/XML/>
- Successeur de SGML (norme en 1986) et de HTML (1990) à terme avec XHTML
- Vocabulaire de certains langages spécialisés (MATHML, SVG graphisme vectoriel, etc.)
- Pas de limitation dans la description
- Les balises définissent la sémantique des données

Objectifs



- Bus de données
 - Échanger, réutiliser, indexer et transformer des données entre applications ou plates-formes hétérogènes
 - Pas adapté pour visualiser un document
- Stockage des données
 - Qui ne nécessitent pas un SGBD
 - Bases de données XML (Tamino, XIndice)
- Séparation du contenu et de la présentation (couples {XML,XSLT})



HTML / XML



- HTML
 - ``
 - `Jean Benard`
 - `Yves Dupré`
 - `...```
 - ✓ *Présentation Web*
 - ✓ *Syntaxe déchiffrée par les navigateurs (affichage à tout prix...)*
- XML : une forme plus adaptée aux données

```
<ENTREPRISE>
<EMPLOYE SECU_SOC="1.80.12.75.120.058/51">
  <NOM>Benard</NOM>
  <PRENOM>Jean</PRENOM>
</EMPLOYE>
<EMPLOYE SECU_SOC="1.51.02.38.032.181/18">
  <NOM>Dupré</NOM>
  <PRENOM>Yves</PRENOM>
</EMPLOYE>
...
</ENTREPRISE>
```

- ✓ *S'intéresse au fond et pas à la forme*
- ✓ *Syntaxe rigoureuse (pas de validation à tout prix)*

Structure des documents



- Informations mises en forme par un humain : fichier **XML orienté documents** (exemple : livre, article, etc.)
- Informations mises en forme par un programme: fichier **XML orienté données** (exemple : extraction d'une base de données, résultat d'un programme, etc.)
- Vérification de la syntaxe d'un document grâce au DTD (*Document Type Definition*) : fichier décrivant la structure des documents y faisant référence grâce à un langage adapté
- Un document suivant les règles de XML est appelé document **bien formé**
- Un document XML possédant une DTD et étant conforme à celle-ci est appelé document **valide**
- Valide => Bien formé

Manipulation des documents



- Le receveur d'un document XML doit utiliser un **parseur**
- Deux manières d'analyse :
 - **SAX** (*Simple Application Programming interface for XML*) : parcours séquentiel de l'arbre XML (*forward only*). Ne lit le document que par petites portions ce qui lui évite de le charger en mémoire intégralement (intéressant pour la lecture des gros documents)
 - **DOM** (*Document Object Model*) : le document XML est totalement chargé en mémoire. Chaque noeud peut être accessible par des méthodes permettant de parcourir l'arbre, de façon hiérarchique ou transversale. Il existe des implémentations de DOM dans pratiquement tous les langages
- Outils payants : XMLSpy (<http://www.altova.com/>), Stylus Studio (<http://www.stylusstudio.com/>), Editix (<http://www.editix.com/>)
- Outils gratuits : Cooktop (<http://www.xmlcooktop.com/>) mais plus supporté, XMLNotePad 2007 (<http://www.microsoft.com/downloads>)

➤ Dialogue

```
<?xml version="1.0" encoding="UTF-8"?>
<vxml version="2.0" xmlns="http://www.w3.org/2001/vxml"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.w3.org/2001/vxml
    http://www.w3.org/TR/voicexml20/vxml.xsd">
<menu>
  <prompt> Bienvenue à la maison. Dis ton choix parmi :
  <enumerate/>
</prompt>
  <choice next="http://www.sports.example.com/vxml/debut.vxml">
    Sports </choice>
  <choice next="http://www.meteo.example.com/intro.vxml">
    Météo </choice>
  <choice next="http://www.astronomie.com/voice/astrojournal.vxml">
    Le journal d'astrophysique</choice>
</menu>
</vxml>
```

- Libre de droits
- Langage autodescriptif et extensible
- Aucune connaissance théorique n'est *a priori* prérequis pour comprendre le contenu d'un document XML
- Universalité et portabilité (les différents jeux de caractères sont pris en compte)
- Déployable : peut être facilement distribué par n'importe quels protocoles à même de transporter du texte, comme HTTP

- Langage XML, généralités

Structure d'un document

- Fichier d'extension « .xml » en 3 parties
 1. Le prologue indique les caractéristiques globales du document (norme, jeu de caractères et informations éventuelles sur des traitements particuliers : grammaire, transformation, etc.)
 2. Déclaration de l'éventuelle grammaire DTD
 3. Arbre des éléments
- Les données sont soit :
 - encadrées par des balises (**éléments**)
 - placées à l'intérieur de balises (**attributs**)
 - définies à l'aide d'entités