

Développer des projets en XML

PXM - Contenu du cours

1. Introduction à XML et aux technologies associées
2. XML et la composition de documents
 - a. Documents bien-formé et valide
 - b. Les namespaces
 - c. Les schémas, DTD et XML Schema.
 - d. Les parseurs XML
 - e. La gestion des liens en XML, XLink/XPointer.

PXM - Contenu du cours

3. XML et la publication de documents

- a. XML et les langages de style
- b. XSL-T, langage de transformation
- c. XSL-FO, langage de formatage de sortie

PXM - Contenu du cours

5. XML et les bases de données

- a. Stockage de documents XML
- b. Solutions avec SGBDR
- c. Base de données XML native

6. XML et les échanges entre applications

7. XML et le commerce électronique

Développer des projets en XML

Introduction à XML et aux technologies associées

Introduction à XML

- ◆ XML signifie eXtensible Markup Language
- ◆ XML n'est pas un langage mais un **méta-langage**.
- ◆ Ce méta-langage permet de concevoir de nouveaux langages (eXtensible) à balises (markup) (ex.: WML, xhtml, SVG,...).
- ◆ XML trouve son origine dans la technologie SGML, méta-langage standardisé par l'ISO et à l'origine de HTML.

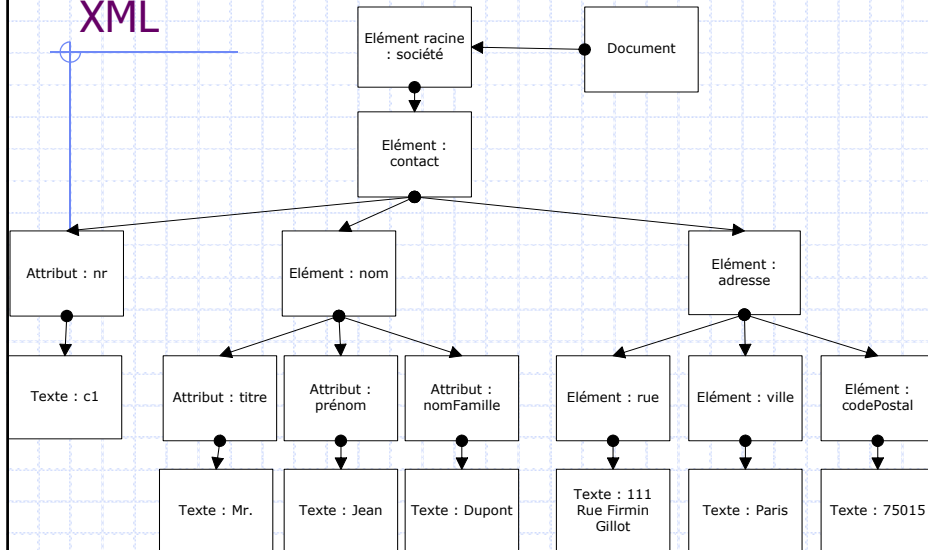
Introduction à XML

- ◆ XML est une norme développée par le World Wide Web Consortium (W3C)
 - consortium qui contrôle l'évolution des technologies du Web (HTML, HTTP, CSS, ...)
 - N'émet pas des standards mais des recommandations
- ◆ Technologies normalisées complémentaires indispensables pour l'exploitation (DTD, XML Schema, XSL-T, XSL-FO, XLink, ...)
- ◆ Accepté par le marché.
- ◆ Supportés par de nombreux outils.

Exemple de document XML bien formé

```
<?xml version="1.0" encoding="UTF-8" ?>
<société>
  <contact nr="c1">
    <nom titre="Mr."
      prénom="Jean" nomFamille="Dupont" />
    <adresse>
      <rue>111 Rue Firmin Gillot</rue>
      <ville>Paris</ville>
      <codePostal>75015</codePostal >
    </adresse>
  </contact>
</société>
```

Structure hiérarchique d'un document XML



20/11/2008

Page 9

Caractéristiques du méta-langage XML

- ◆ Codification du contenu (recommandation XML) :
 - Minimum UNICODE (UTF-8, UTF-16)
 - autres jeux de caractères supporté
- ◆ Les balises sont des marqueurs qui délimitent les données caractères.
- ◆ Elles sont des méta-données accompagnant les données et renseignant sur la sémantique du contenu.

20/11/2008

Page 10

Caractéristiques du méta-langage XML

- ◆ Les langages ou vocabulaires à base de XML (également appelés instances de XML) sont conçus selon les besoins des applicatifs.
- ◆ Un nouveau langage conçu à l'aide de la technologie XML est destiné à un type d'application ou un secteur d'activité :
 - formule mathématique (MathML), feuille de style (XSL-T), médecine, commerce, finance, ...

Caractéristiques du méta-langage XML

- ◆ Méta-langage :
 - pas d'éléments prédéfinis
 - l'auteur crée les éléments dont il a besoin
 - les éléments sont donc adaptés au besoin de structuration du contenu du document
- ◆ Règles de conception très stricts :
 - Règles de syntaxe et de structuration
 - Simplification des logiciels consommateurs (traitement d'erreurs réduit)
 - Correspond à l'évolution : le code est généré par des éditeurs spécialisés

XML par rapport à HTML

- ◆ XML et HTML ne sont pas 2 technologies vraiment comparables:
 - XML est un méta-langage
 - HTML, xhtml est un langage
- ◆ XML apporte une technologie de structuration de contenu pouvant s'appliquer, notamment, dans le contexte Web.
- ◆ HTML n'offre pas de possibilités pour renseigner sur la sémantique du contenu du document.
- ◆ HTML est destiné principalement à la présentation des informations de pages Web sur navigateur d'ordinateur.

XML et HTML

- ◆ XML ne devrait pas, au sens propre, remplacer HTML.
- ◆ HTML, xhtml et CSS continueront d'être utilisés comme technologie de présentation à l'écran pour les navigateurs d'ordinateurs.
- ◆ La nouvelle version du langage HTML, xhtml est principalement une "xmlisation" du langage, xhtml est une application du méta-langage XML.

XML et HTML

◆ HTML

```
■ <HTML>
  <HEAD>
    <TITLE>
      Le seigneur des
anneaux</TITLE>
  </HEAD>
  <BODY>
    <H1> Le seigneur des
anneaux</H1>
    <P>J.R.R. Tolkien
    <P>When Mr. Bilbo
      Baggings of Bad
      End...
    </BODY>
  </HTML>
```

◆ XML

```
■ <?xml version="1.0"?>
  <livre>
    <info>
      <titre>
        Le seigneur des
anneaux
      </ titre >
      <auteur>
        J.R.R. Tolkien
      </auteur>
    </info>
    <chapitre nr="ch1">
      <p>When Mr. Bilbo
        Baggings of Bad
        End...</p>
    </chapitre>
  </livre>
```

Exemple HTML et XML

◆ En HTML :

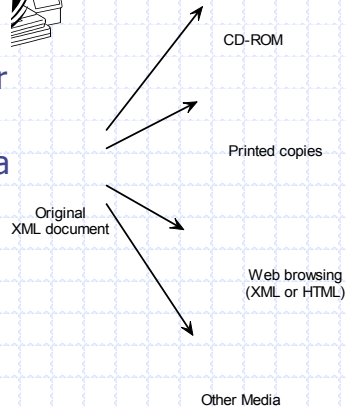
```
<P>Portable P300</P>
<BR>Centre des portables
<BR>20.000,-FF
```

◆ En XML :

```
<produit>
  <modèle>Portable P300</modèle>
  <vendeur>Centre des portables</vendeur>
  <prix>20.000,-FF</prix>
</produit>
```


Avantage

- ◆ XML se concentre sur la structure
 - HTML se concentre sur la présentation
- ◆ présentation déduite de la structure et selon les besoins du medium
 - un document (une structure), plusieurs utilisations



Utilisation du langage XML

- ◆ XML offre un moyen de structuration et de représentation de l'information
- ◆ Indépendant de toutes plates-formes et applications
- ◆ Aussi bien côté serveur que client
- ◆ Indépendant des technologies de persistance (système de fichiers, base de données relationnelles ou objet)
- ◆ Dans toute application pour laquelle l'information doit être transmise, XML trouve sa place.

Le parseur, composant logiciel de base pour XML

- ◆ Toute application (navigateur, éditeur) exploitant XML est développée autour d'un parseur.
- ◆ Un parseur est un composant logiciel effectuant l'analyse lexicale et syntaxique (avec validation ou non) d'une structure XML.
- ◆ Disponibles sous différentes formes et pour différents langages (package Java, composant ActiveX, librairies C, ...)
- ◆ Disponibles gratuitement (Apache, IBM, Lotus, Microsoft, Oracle, Sun, ...).

SGML

- ◆ HTML et XML ont la même origine, SGML
- ◆ SGML est un standard international, publié en 1986, contenant toutes les règles pour définir de nouveau langage de markup.
- ◆ Utilisé depuis par de grandes organisations pour gérer leurs documentations techniques ou autres.
- ◆ DSSSL est le langage de style utilisé avec SGML.
- ◆ Utilisés directement, SGML et DSSSL sont complexes et peu utilisés.

SGML

- ◆ Un document décrit en SGML est composé:
 - Document Type Definition (DTD) : ensemble d'instructions indiquant à l'agent de l'utilisateur la signification des balises ou tags. (Les tags sont définis dans le DTD et peuvent varier d'un document à l'autre)
 - Feuille de style (rédigée en DSSSL) : indiquant à l'agent de l'utilisateur que faire avec ces balises.
 - Le document avec le contenu utile.

XML versus SGML

- ◆ XML est une version simplifiée de SGML.
- ◆ XML simplifie la grammaire du SGML, permettant de créer aisément de nouveau langage à balises.
- ◆ Situation de SGML:
 - outil difficile à mettre en œuvre
 - difficulté d'une conformité totale
- ◆ XML ramène à 26 pages la définition standard du langage, alors qu'en SGML, il s'agissait de 400 pages (norme ISO 8879)

Principes de conception du XML

- ◆ XML sera directement utilisable sur Internet
- ◆ XML supportera une large variété d'applications
- ◆ XML sera compatible avec SGML
- ◆ Facilité de création de programmes de traitement de documents XML
- ◆ Un document XML sera raisonnablement lisible aisément par un humain
- ◆ Facilité de création de document XML
- ◆ Design XML rapidement disponible, formel et concis.

Exemple complet de document XML valide

- ◆ Schéma du langage dans un fichier externe:
ListeAdresses.dtd

```
<!ELEMENT ListeAdresses ( Adresse )+>
<!ELEMENT Adresse ( (Travail| Maison)+)>
<!ATTLIST Adresse nom CDATA #REQUIRED>
<!ELEMENT Travail ( Rue, Ville, CodePostal, Pays)>
<!ELEMENT Maison ( Rue, Ville, CodePostal, Pays)>
<!ELEMENT Rue ( #PCDATA )>
<!ELEMENT Ville ( #PCDATA )>
<!ELEMENT CodePostal ( #PCDATA ) >
<!ELEMENT Pays ( #PCDATA )>
```

- ◆ Ce schéma est défini par un DTD, il existe un autre langage de définition de schéma, XML Schema.

Exemple complet XML (2)

◆ Fichier adresses.xml (suite)

```
<?xml version="1.0"?>
<!DOCTYPE ListeAdresses SYSTEM "ListeAdresses.dtd">
< ListeAdresses >
  <Adresse nom="Dupont">
    <Travail>
      <Rue>22, Rue de la station</Rue>
      <Ville>Dunkerque</Ville>
      <CodePostal>12345</CodePostal>
      <Pays>France</Pays>
    </ Travail>
    < Travail>
      <Rue>44, Rue Neuve</Rue>
      <Ville>Bruxelles</Ville>
      <CodePostal>1000</CodePostal>
```

Exemple complet XML (3)

◆ Fichier adresses.xml (suite)

```
<Pays>Belgique</Pays>
  </Travail>
</Adresse>
<Adresse nom="Durand">
  <Travail>
    <Rue>34 Av Brigand</Rue>
    <Ville>Lausanne</Ville>
    <CodePostal>3456</CodePostal>
    <Pays>Suisse</Pays>
  </Travail>
```

Exemple complet XML (4)

◆ Fichier adresses.xml (suite)

```
<Maison>  
  <Rue>23 Rue Couture</Rue>  
  <Ville>Milan</Ville>  
  <CodePostal>98765</CodePostal>  
  <Pays>Italie</Pays>  
</Maison>  
</Adresse>  
</ListeAdresses>
```

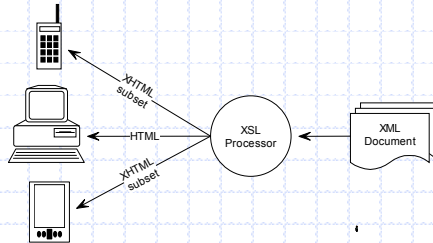
Exemples d'utilisation de XML dans un contexte Web

- ◆ Document avec recherche basée sur la sémantique du contenu
- ◆ Longs documents à éclater en plusieurs sections, fusion de documents
- ◆ Présentation des données dans des formats différents
- ◆ Filtrage d'accès au contenu des documents par la connaissance du contenu.
- ◆ Lorsque la puissance de XSL-T et/ou XSL-FO est nécessaire par rapport à CSS.
- ◆ Traitements automatisables nécessaires.

Exemple : site web

◆ gestion de site web

- à l'épreuve du changement
- parce que basé sur la structure (stable)
- site optimisé pour l'utilisateur et son matériel



Langages à base de XML

- ◆ XML est un outil de base pour définir un langage commun d'expression et de sémantique des données:
 - MathML, CML, AML
 - SMIL, SVG
 - WebDAV, ICE, CDF
 - VML, PGML...
- ◆ XML permet de définir de nouveau langage, applicables dans des domaines variés, à l'aide d'une syntaxe simple.
- ◆ La définition de ces langages est reprise dans un schéma (DTD, Document Type Definition ou XML Schema) d'accès public et à distance.

Exemple MathML

```
<math>
  <mrow>
    <mi>f</mi>
    <mo>:</mo>
    <msubsup>
      <mi>R</mi>
    </msubsup>
    <mn>2</mn>
  </mrow>
  <mo>&to;</mo>
  <msubsup>
    <mi>R</mi>
  </msubsup>
  <mn>2</mn>
</math>
```

20/11/2008

Page 31

Exemple DocBook

```
<!DOCTYPE article PUBLIC "-//OASIS//DTD DocBook V3.1//EN">
<article>
  <artheader>
    <title>My Article</title>
    <author><honorific>Dr</honorific><firstname>Emilio</firstname>
      <surname>Lizardo</surname></author>
  </artheader>
  <para> ... </para>
  <sect1><title>On the Possibility of Going Home</title>
  <para> ... </para>
</sect1>
  <bibliography> ... </bibliography>
</article>
```

20/11/2008

Page 32

Exemple fichier SMIL

```
<par>
  <img href="background"/>
  <seq> streams played sequentially
  <par> streams played in parallel
  <img href="graph" loc="left-video" dur="45s"/>
  dur defines duration of presentation
  <text href="graph-text" loc="left-text"/>
  </par> loc defines location in <head> where it plays
  <par>
    <a href="http://www.w3.org/People/Berners-Lee">
      <video href="tim-video" loc="left-video"/> </a>
      <text href="tim-text" loc="left-text"/>
    </par>
  </seq> image, text, audio, video channels, href for links
  <seq>
    <audio href="joe-audio"/>
    <audio href="tim-audio"/>
  </seq>
  <video id="jv" href="joe-video" loc="right-video"/>
  <text href="joe-text" loc="right-text"/>
</par>
```

XML pour les textes et les données

- ◆ XML peut aussi bien être utilisé pour la conception de:
 - langage à balises orienté texte (document-centric)
 - langage à balises orienté donnée (data-centric)
- ◆ La notion de document en XML doit être prise au sens large, un document est une structure de données et non pas une pièce physique.

XML pour les textes et les données

◆ XML pour le texte :

```
<doc>
<titre>Titre du document</titre>
<auteur>Auteur du document</auteur>
<para>Ceci est du texte en <gras>format varié</gras>
</para>
<para>Encore du texte
</para>
</doc>
```

XML pour les textes et les données

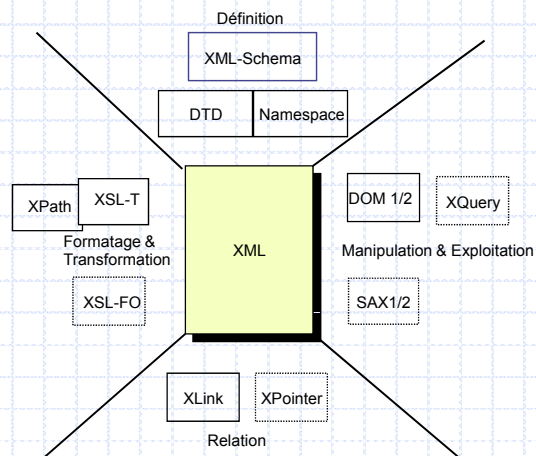
◆ XML pour les données:

```
<bonCommande commDate="2001/04/04">
  <client nom="Dupont" codePostal="1040" />
  <item quantite="12">
    <descrProd>Gadgets</descrProd>
  </item>
  <item quantite="8">
    <descrProd>Gadgets</descrProd>
  </item>
</bonCommande>
```

Les schémas

- ◆ Le schéma d'un langage est défini à l'aide de DTD ou XML Schema
- ◆ La validité et la qualité de l'échange des données peuvent être contrôlées sans obligation de définir, pour chaque application, les règles lexicales et syntaxique.
- ◆ Les méta-informations se retrouvent dans le schéma.
- ◆ Le schéma permet de comprendre et d'exploiter l'information.

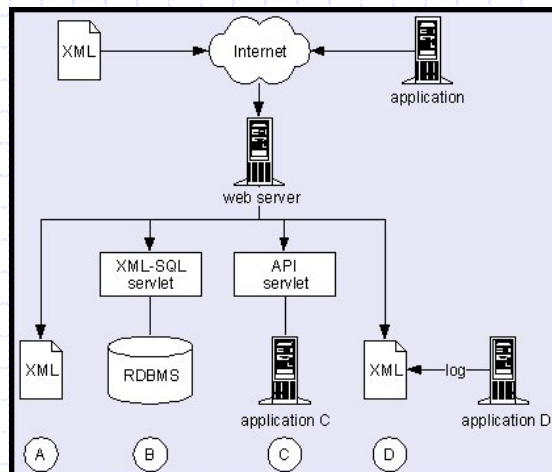
XML et les technologies associées



XML et les technologies associées

- ◆ XML est accompagné de technologies complémentaires permettant d'exploiter les données selon les besoins spécifiques.
- ◆ Définition : DTD, XML Schema, Namespace
- ◆ Interface et langage de manipulation : DOM, SAX, XQuery
- ◆ Langage de transformation : XSL-T avec XPath, XSL-FO
- ◆ Relation entre documents : XLink et XPointer

XML, élément de solution MiddleWare



XML, élément de solution MiddleWare

- ◆ XML dépasse le simple contexte d'utilisation du Web
- ◆ XML est utilisée dans les solutions de middleware, (cfr SOAP) technologie de messaging et de communication entre composants.
- ◆ XML est un format uniforme permettant de fédérer les accès à des sources d'information multiples et hétérogènes.

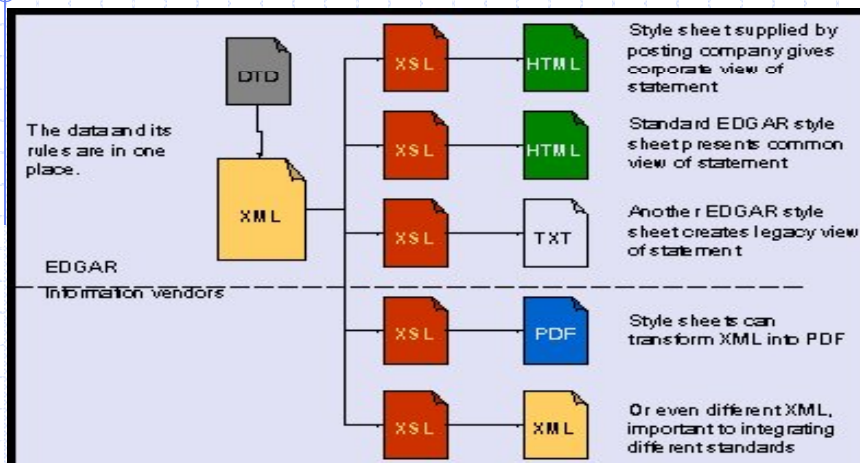
XML, l'ASCII du 21ème siècle

- ◆ XML est un langage accepté universellement
- ◆ W3C a adopté et standardisé XML
- ◆ Sun, IBM et Microsoft, entre autres, favorisent et participent à l'évolution de cette technologie
- ◆ XML permet de récupérer les investissements des systèmes hérités, mainframe...
- ◆ XML permet d'exposer sous une forme facilement exploitable l'information existante
- ◆ XML donne une deuxième jeunesse à l'EDI

Avantages de XML

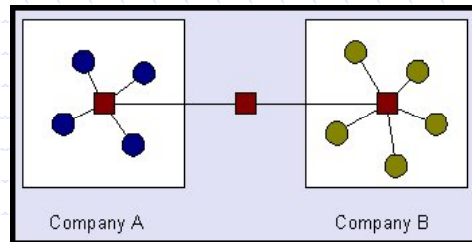
- ◆ Afficher les mêmes données dans des formats différents (personnalisation)
- ◆ Afficher les mêmes données sur des média différents (xhtml, wml, voiceXML, smil, ...)
- ◆ Permet d'exploiter les infrastructures existantes sans nécessité de remettre en question les connaissances.

Avantages de XML

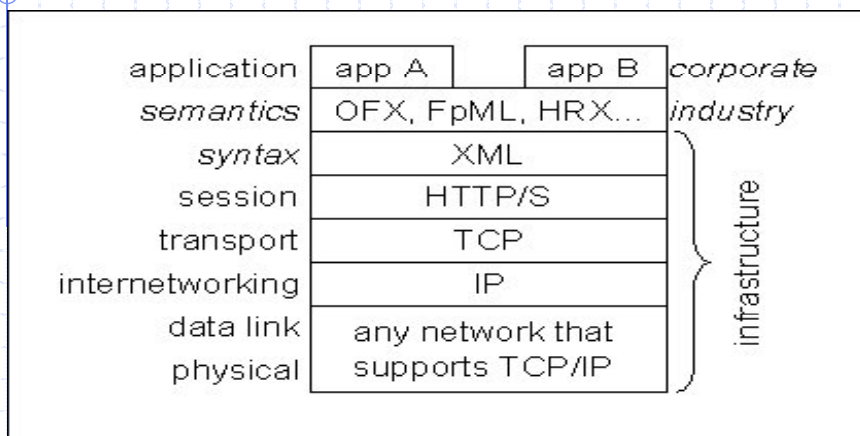


Avantages de XML

- XML permet de normaliser la communication au sein d'un système et entre des systèmes hétérogènes



XML, indépendant des plates-formes



Outils pour XML

- ◆ Pour rédiger un document en XML, un simple éditeur de texte suffit (supportant Unicode, idéal).
- ◆ Pour analyser un document, il faut un parseur (ou parser, processor), module logiciel qui analyse le document
- ◆ De nouveaux éditeurs, outils de programmation et de vérification arrivent sur le marché (XMetal, XML Spy, Turbo XML, Epic Editor, ...)
- ◆ Pour consulter un document XML
 - conversion de XML en HTML, côté client ou serveur
 - navigateur compatible (IE5 et 6, NN6, ...)

Outils pour XML

- ◆ Evolution des base de données traditionnelles: SGBDR et SGBDOO
- ◆ Base de données XML native
- ◆ Système de gestion documentaire, système de gestion de contenu