

Introduction à XML et DOM

Code: xml-dom

Originaux

url: <http://tecfa.unige.ch/guides/tie/html/xml-dom/xml-dom.html>

url: <http://tecfa.unige.ch/guides/tie/pdf/files/xml-dom.pdf>

Auteurs et version

- [Daniel K. Schneider](#)
- Version: 1.11 (modifié le 2/10/07 par DKS)

Prérequis

- Internet et WWW de base, HTML de base
Module technique précédent: [internet](#)
Module technique précédent: [www-tech](#)
Module technique précédent: [xhtml-intro](#)

Modules

Module technique suivant: [xml-tech](#) (introduction technique à XML)

Module technique suivant: [xml-docu](#) (XML et documentation)

Objectifs

- Fournir une idée de la définition technique de XML
 - formalisme XML, la notion de méta-langage, la notion du "Domain Object Model" (DOM)
- Fournir une idée du XML framework du consortium WWW,
 - la place de ce nouveau framework par rapport aux autres outils (HTML, bases de données, etc.)
- Introduire les noms et fonctions de quelques langages XML, par exemple:
 - les langages les plus importants du W3C
 - XML pour la documentation
 - Le web sémantique
 - Webservices

1. Table des matières détaillée

1. Table des matières détaillée	3
2. Vers un nouveau paradigme Internet	4
2.1 Pointeurs supplémentaires:	4
2.2 XML sert à fabriquer des langages de balises	5
2.3 2 façons d'aborder XML	6
2.4 Le rapport avec HTML, PDF, etc.	7
2.5 Desiderata du monde Internet	8
3. Formats de données XML principaux	9
3.1 Eléments du "XML framework" du W3C pour les documents	9
3.2 Applications XML pour définir des "documents"	11
4. Quelques exemples de langages XML	13
4.1 La syndication des nouvelles et le social Web	13
4.2 Le Web sémantique	13
4.3 XML et documentation	14
4.4 Xlink - Vers un meilleur hypertexte ?	16
5. La notion de "service Web"	17
6. Sous le capot: Le Document Objet Model (DOM)	18
6.1 Le principe du DOM avec un exemple	19
7. XML dans le monde de l'éducation ?	21
7.1 Vocabulaires typiques	21
7.2 Organisations	22

2. Vers un nouveau paradigme Internet

- Ce chapitre contient quelques définitions,
- définit le rôle de XML par rapport à l'évolution des TIC,
- fait quelques réflexions concernant le "rapport" entre XML et HTML

2.1 Pointeurs supplémentaires:

Quelques indexes qui vous amèneront plus loin:

- XML@Tecfa: <http://tecfa.unige.ch/guides/xml/pointers.html>
 - standards, tutoriels, logiciels, autres pages ressources
- XML Applications@Tecfa: <http://tecfa.unige.ch/guides/xml/applications.html>
 - applications XML (grammaire variés)
- XSL(T)@Tecfa: <http://tecfa.unige.ch/guides/xml/xsl-pointers.html>
 - ressources pour les feuilles de style xslt et xsl-fo
- DOM@Tecfa: <http://tecfa.unige.ch/guides/dom/pointers.html> (faible!)
- RDF @Tecfa: <http://tecfa.unige.ch/guides/rdf/pointers.html> (faible!)
- SVG @Tecfa: <http://tecfa.unige.ch/guides/svg/pointers.html>
 - graphisme 2-D
- Voir le WDWL Acronym Expander pour l'explication de la plupart des sigles:
 - <http://wdvl.internet.com/Authoring/Languages/XML/Overview/acronym.html>

2.2 XML sert à fabriquer des langages de balises

XML = e**X**tended **M**arkup **L**anguage

- un langage "markup" ("langage de balises") sert à encoder/structurer des données (textes, formats vectoriels, messages entre machines, feuilles de style.....)
- XML fournit le formalisme + des mécanismes pour définir ces langages (grammaires, schémas)

Historique des langages "markup" Internet:

SGML (Standard Generalized Markup Language, ISO standard en 1986)

- meta-langage pour définir des langages de "markup"
- HTML (1990) est une application SGML ayant assez peu de balises (tags)

XML (1997, -)

- un meta-langage plus léger que SGML adapté au Web
- permet la définition de langages adaptés à des besoins très variés
- XML est un formalisme pour décrire des structures de données
 - sert à organiser l'échange d'informations
 - sert à remplacer HTML pour certaine tâches (mais fonctionne d'une autre façon !)
 - donc: XML **n'existe pas** au même sens que HTML, ce n'est qu'un formalisme !

2.3 Deux façons d'aborder XML

(1) XML comme formalisme

*Un formalisme pour
créer des grammaires
(aussi appelées schémas, modèles,
vocabulaires, applications, ...)*

Exemple d'une grammaire:

```
<!ELEMENT page (title, content,  
comment?)>  
<!ELEMENT title (#PCDATA)>  
<!ELEMENT content (#PCDATA)>  
<!ELEMENT comment (#PCDATA)>
```

Exemple d'un document XML:

```
<page>  
<title>Hello XML friend</title>  
<content>  
  Here is some content :)  
</content>  
<comment>  
  Written by DKS/Tecfa,  
</comment>  
</page>
```

(2) Le "XML framework" du W3C

*Un ensemble de langages/grammaires
pour créer un web plus puissant.
Ces formalismes sont définis avec XML !*

graphisme: SVG
formatage: XSL/FO
hypertexte: Xlink, Xpointer
recherche: XQL
transformations: XSLT
commerce: SOAP
.....

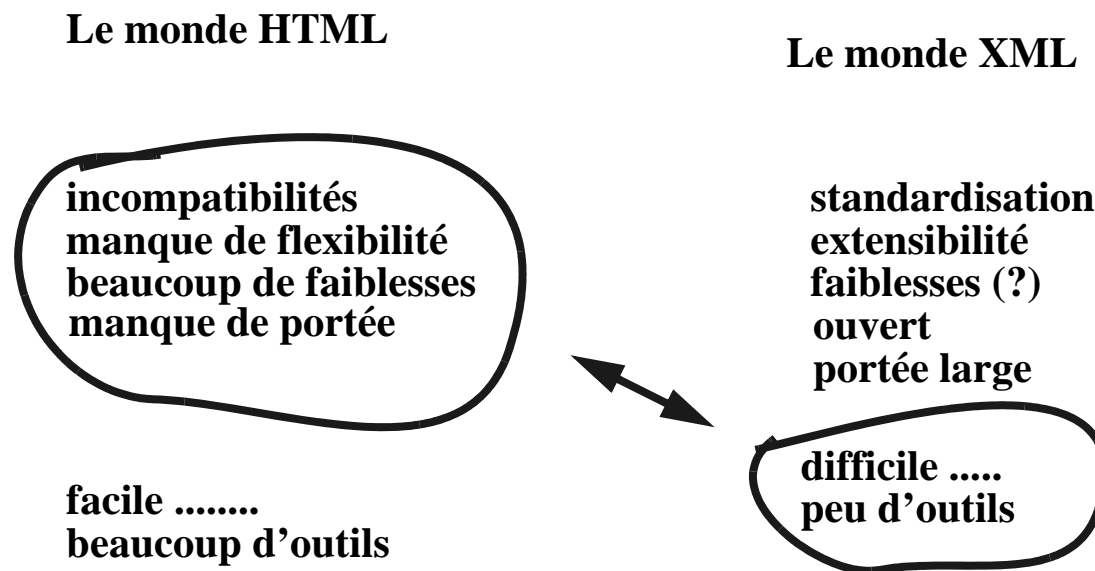
**Pour faire vivre ces langages
il faut des outils:**

- certains sont déjà en place
- certains sous développement
- certains pas encore adoptés

2.4 Le rapport avec HTML, PDF, etc.

- On ne peut pas vraiment comparer XML et HTML, on le fait puisque d'autres le font ...
- On l'aborde sous l'angle "le problème HTML et la solution XML"

HTML vs. XML



Attention:

- Aborder XML à travers les déficiences de HTML est une chose
- faire croire que XML se discute au même niveau est faux !
 - XML est un *formalisme* pour fabriquer des langages alors que HTML est un langage
 - la version moderne de HTML (c.a.d. XHTML) est défini avec XML ("est une application XML")!
 - HTML ancien est défini avec SGML ("est une application SGML")
 - XML est simplement un sous-ensemble de SGML.

2.5 Desiderata du monde Internet

Les besoins en matière des TIC ont beaucoup évolué depuis une décennie

On veut par exemple:

- mieux structurer de l'information
 - et la retrouver facilement,
 - la traiter électroniquement,
 - et l'imbriquer facilement dans des applications.
- faire des systèmes de documentation et d'hypertextes puissants
- afficher et imprimer de façon flexible et jolie
 - écrans, papier, portables, etc.
- un format universel pour toute sortes de données et d'usages
 - diffuser/échanger/stocker/chercher/.....
 - donc on peut réutiliser les mêmes outils informatiques pour certaines tâches
- un format lisible par les humains et les machines
 - Note: "humains" = spécialistes

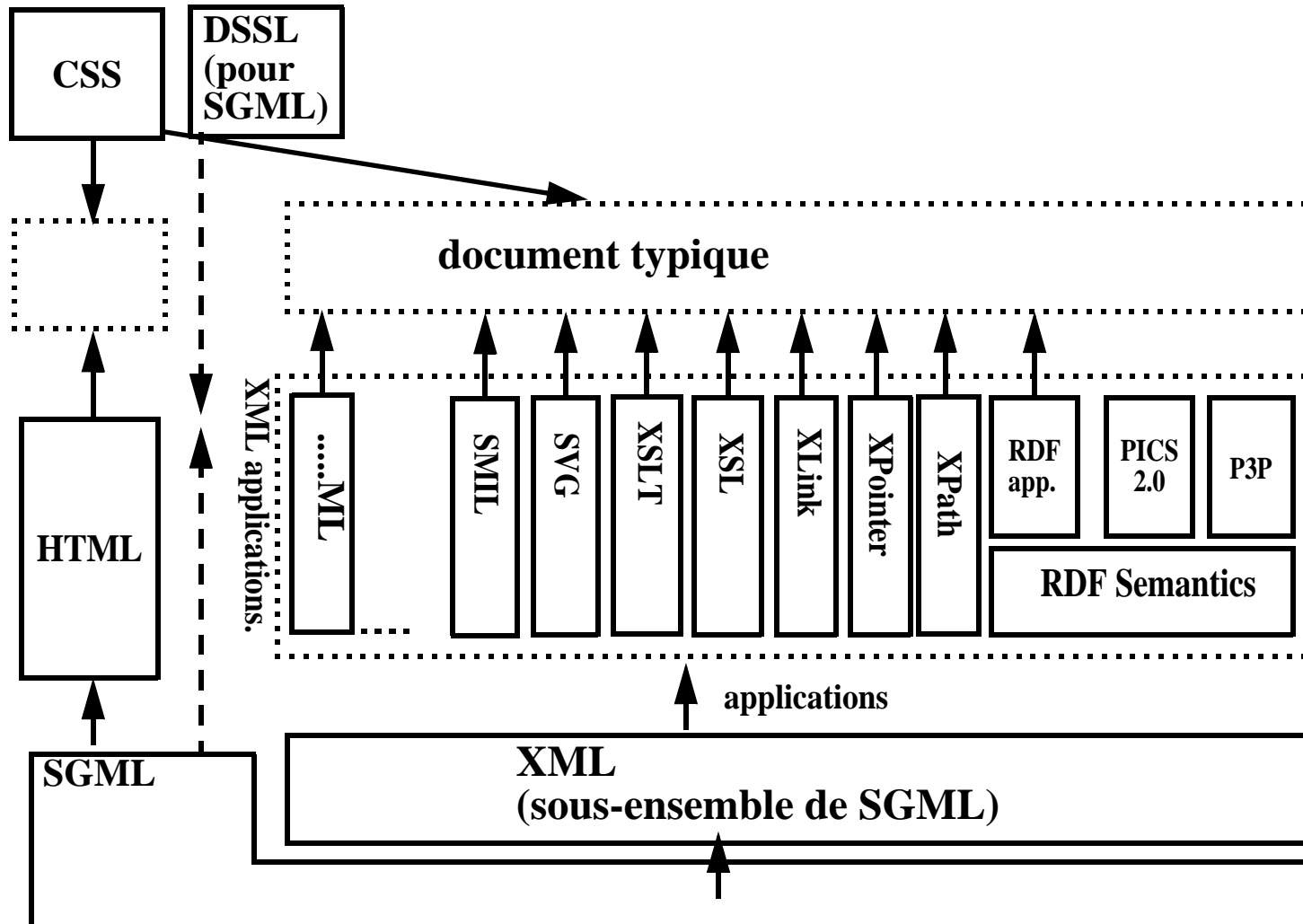
Il fallait un meta-langage propre qui permette de créer des "langages" variés

Rappel (!)

- XML n'est pas un langage comme HTML,
.... mais un formalisme qui permet de fabriquer des langages !

3. Formats principaux de données XML

3.1 Éléments du "XML framework" du W3C pour les documents



Petite explication de quelques notions et sigles

XML et SGML sont des meta-Langages:

- Un meta-langage est un langage qui permet de définir d'autres langages (appelés souvent "applications" dans ce contexte)
- SGML: Standardized Generalized Markup Language (ISO 8879)
- XML (eXtensible Markup Language): version simplifiée de SGML)

Langages SGML (en voie de disparition):

- il y en a plein (notamment dans le domaine du "document processing")
- HTML: Jeu de balises pour structurer un texte selon des catégories plutôt "typographiques" (titres, paragraphes, listes, etc.), inclure un peu de multimédia et un peu d'hypertexte.

Langages de Style:

- CSS (1/2/3): Langage de style pour HTML et XML
- DSSL: Langage de Style le plus populaire pour SGML (Scheme like)
- XSL/FO (application XML): Langage de style pour XML
- XSLT (application XML): Langage de transformation pour XML

La notion de "application":

- Structure: le document est associé à un "modèle d'information" / "grammaire" / "application"
- Définition de grammaires XML: Il existe 3 standards populaires les DTD (Document Type Definitions), XSchema et Relax-NG
- Validation (on peut contraindre les auteurs à obéir une grammaire)

3.2 Applications XML pour définir des "documents"

Quelques applications XML du W3C (consortium WWW)

- XSL/FO (application XML): Langage de style pour XML
- XSLT (application XML): Langage de transformation pour XML
- XLink: Hypertext links
 - <http://www.w3.org/TR/xlink/>
- XPointer (pointeurs vers une ressource) et XPath (chemins dans la structure)
 - (utilisés par XSLT, XInclude, XLink, XQuery, etc.)
- Applications RDF: (langage pour définir des relations entre objets)
 - voir: <http://tecfa.unige.ch/guides/rdf/pointers.html>
- PICS 2.0: Platform for Internet Content Selection
 - <http://www.w3.org/PICS/>
- SMIL: Synchronized Multimedia Integration Language
 - <http://www.w3.org/AudioVideo/>
- P3P: Platform for Privacy Preferences
 - <http://www.w3.org/P3P/>
- SVG: Scalable Vector Graphics
 - <http://www.w3.org/TR/SVG/>
- MathML: Mathematical Markup Language
 - <http://www.w3.org/Math/>
- XHMTL
 - <http://www.w3.org/TR/html/>

Autres applications

- “XML est un formalisme universel, utilisé dans beaucoup de contextes pour encoder un contenu spécifique (formules chimiques, échange de nouvelles, contenus e-learning, etc.):
 - CML (Chemical Markup Language)
 - X3D (VRML-Xmlisé)
 - NML (News Markup Language) et NITF (News Interchange Text Format)
 - IMS Content packaging (structure et emballage de contenus e-learning)
 - vous pouvez créer votre propre langage.....
- Certains langages XML ajoutent des fonctionnalités à ce type de vocabulaire:
 - feuilles de style
 - hyper-liens
 - graphismes et animations: SVG, X3D, Smile
 - catégorisations / méta-donnés: RDF/Dublin Core
 - échange de données
 - échange de requêtes entre programmes: XML-RPC, SOAP
 - etc.
- XML permet de définir des langages de programmation:
 - par exemple XSLT, XQuery
- XML définit des langages de communication entre machines (webservices)

Résumé

- Il existe déjà des centaines d'applications XML....
- Il faut distinguer entre applications plutôt "text-centric" et "data-centric"
 - "text-centric": contenu destiné à la lecture
 - "data-centric": contenu destiné à un logiciel (qui utilise, traite, met en forme, communique, etc.)

4. Quelques exemples de langages XML

4.1 La syndication des nouvelles et le web social

- Internet subit depuis 2-3 ans une profonde mutation en ce qui concerne l'organisation de ses "espaces d'information"
- RSS (sous ses formes variées) permet l'échange automatique de "titres" et de "résumés" entre différents portails et weblogs
- FOAF (et autres formats) permettent de définir des profils de personnes qui sont ensuite utilisés pour organiser des réseaux sociaux en ligne

4.2 Le Web sémantique

- Le Web devient plus "smart"
- Les Topic Maps (standard ISO) permettent d'organiser des collections de ressources sous forme de réseau sémantique (donc on ne retrouve pas juste les arbres, mais on a un "carte" de la forêt.
- RDF est un langage qui permet de décrire des relations entre objets
 - exemple: standards de "méta-données" qui décrivent ce qu'il y a dans une ressource
- OWL, le "Web Ontology Language" (construit avec RDF) est un formalisme qui permet de décrire les relations qui existent entre des choses. Il existe un lien conceptuel avec les "topic maps".

4.3 XML et documentation

- Aujourd'hui on peut directement présenter des informations codés en XML (n'importe quelle grammaire) dans un navigateur, mais à condition d'utiliser une feuille de style.

Les style-sheets permettent de:

- préparer/arranger un contenu pour une "présentation"
- définir le "layout" (mise en forme, formattage) d'un "texte" écrit en XML

L'utilité des style-sheets est donc de:

- séparer contenu et présentation
- rationaliser le travail (un style-sheet pour beaucoup de documents)

A. XSL: Extensible Stylesheet Language

- XSL est un langage recommandé par le W3C

url: <http://www.w3.org/TR/xsl>

XSL possède 2 fonctions principales

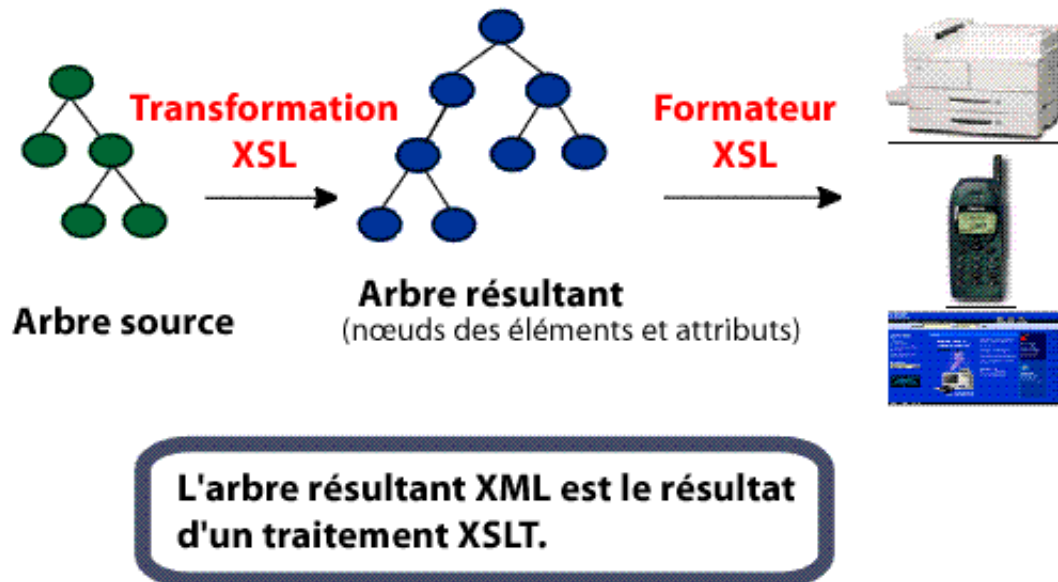
1. langage de transformation (XSLT) d'éléments XML
- Par exemple: création de tables de matières, traduction de XML vers HTML

url: <http://www.w3.org/TR/xslt>

2. langage de mise en page (formattage) (XSL/FO)

url: <http://www.w3.org/TR/xsl/>

Avec un GIF (tiré de la recommandation XSL-1)



Le formatage avec XSL-FO:

- formatage sophistiqué, aussi selon héritage, descendance, position etc.
- génération de textes et graphiques
- possibilité de définir des macros
- tout ce que l'on trouve dans CSS et plus

B. CSS (Cascading Style Sheets)

- CSS1 (pour HTML seulement, mais prédominant)
- CSS2 (support pour XML, marche avec IE 6 / Firefox, etc.)

Conclusion

- Aucun problème pour afficher un document XML de type "texte"

4.4 Xlink - Vers un meilleur hypertexte ?

- reste une proposition, pas d'implémentation complète pour le moment.
- utilisation d'une version "simple" dans SVG, X3D etc.
- Xlink = comment insérer un lien dans un document XML (le lien exprime une relation entre deux ou plusieurs objets)

url: <http://www.w3.org/TR/xlink>

Xlink repose sur d'autres standards (partagés avec XSLT par exemple)

- XPointer = comment identifier un fragment XML (utilisable par des liens)
- XPointer repose sur XPath = comment identifier un chemin vers une ressource

url: <http://www.w3.org/TR/xptr>

url: <http://www.w3.org/TR/xpath>

Caractéristiques principales:

- Liens multi-directionnels, liens à multiple destinations
- Occlusions, inclusions, remplacements de contenus dans un document,

D'où vient ce standard ?

- HTML
 - Ancres: href (attribut de A), src (attribut de IMG et NOTE) ...
 - Targets: name attributs (A), id (attribut dans HTML 4.x)
- surtout: langages SGML HyTime et TEI Extended Pointers (extension à HyTime)

5. La notion de "service Web"

- Un service Web est un ensemble de protocoles et de normes informatiques utilisés pour échanger des données entre les applications
- Les données échangées sont en format XML

Exemples:

(il existe toute une panoplie d'instruments)

- Le protocole SOAP (Simple Object Access Protocol) définit un langage d'échange. Il est composé de deux parties :
 - une enveloppe, informations permettant son acheminement et son traitement,
 - un modèle de données pour les informations à transmettre.
- WSDL (Web Services Description Language)
 - décrit une interface publique d'accès à un Service Web
 - indique comment communiquer pour utiliser le service

6. Sous le capot: Le Document Objet Model (DOM)

Le DOM est la représentation informatique (dans la machine) d'un document XML

- Le DOM est un standard pour lire et manipuler des documents par le biais d'un programme
- Voir W3C Data Formats (<http://www.w3.org/TR/NOTE-rdfarch>)

L'API (application programming interface) pour documents HTML et XML sert à

- construire des documents (navigateurs)
- naviguer dans leur structure avec un programme
- ajouter, modifier ou détruire des éléments avec un langage de programmation comme PHP (du côté serveur) ou JavaScript (dans votre navigateur)
- Cela a permis de standardiser ce qu'on appelle "DHTML"
 - Le terme DHTML ne se retrouve dans aucun standard, mais indique qu'on manipule les propriétés CSS et le contenu des éléments avec JavaScript

Définition officielle du terme "DOM":

De l'abstract de la spécification (<http://www.w3.org/TR/REC-DOM-Level-1/>): a platform- and language-neutral interface that allows programs and scripts to dynamically access and update the content, structure and style of documents. The Document Object Model provides a standard set of objects for representing HTML and XML documents, a standard model of how these objects can be combined, and a standard interface for accessing and manipulating them. Vendors can support the DOM as an interface to their proprietary data structures and APIs, and content authors can write to the standard DOM interfaces rather than product-specific APIs, thus increasing interoperability on the Web.

- Note: Le DOM est compliqué: Il existe des alternatives plus simples !

6.1 Le principe du DOM avec un exemple

Exemple 6-1: Un simple table HTML comme "DOM tree"

- voir: REC-DOM-Level-1.19981001

Les données XML:

```
<TABLE>  <TBODY>
<TR> <TD>Pierre Muller</TD>
      <TD>http://pm.com/</TD> </TR>
<TR> <TD>Elisabeth Dupont</TD>
      <TD></TD> </TR>
</TBODY> </TABLE>
```

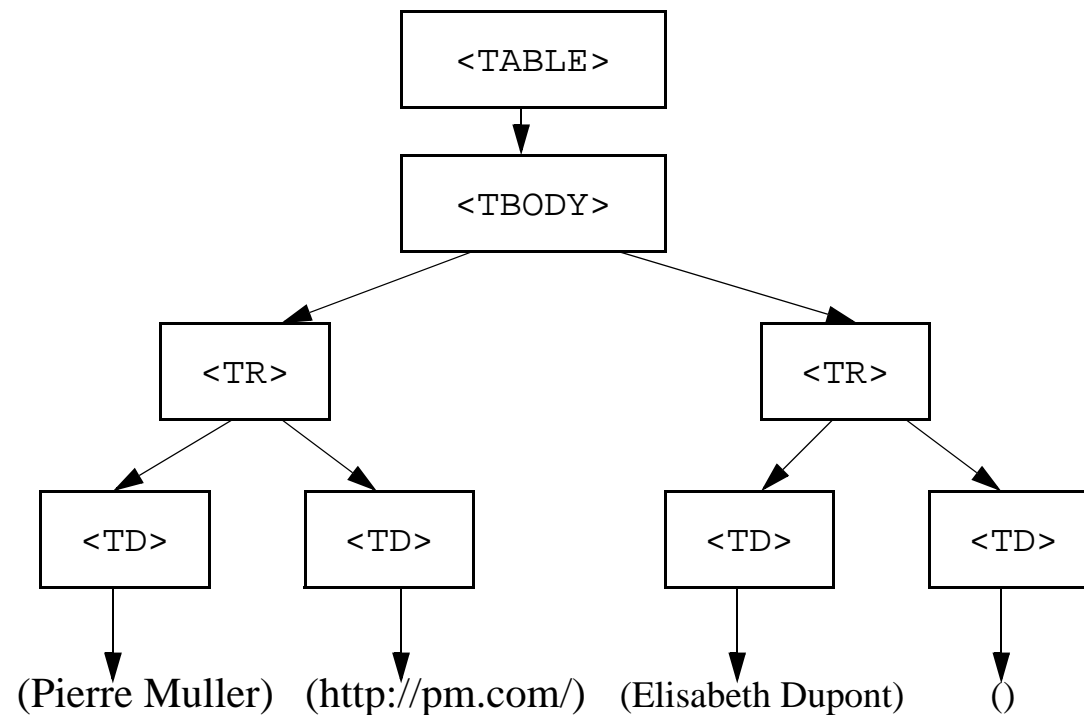
Affichage typique dans un browser:

Pierre Muller	http://pm.com/
Elisabeth Dupont	

Représentation alternative dans un browser ou un applet:

- line 1:
 - Pierre Muller
 - http://pm.com/
- line 2:
 - Elisabeth Dupont

Représentation interne très APPROXIMATIVE de l'arbre dans le DOM



- Un script peut manipuler cette structure au niveau du contenu, de l'affichage et de l'interface utilisateur
 - Javascript dans les navigateurs Web
 - Java, Flash (et autres) dans des extensions (applets, Flash player).
 - PHP, Java, ASP etc. au niveau du serveur
- Grâce au DOM: changement de paradigme pour les pages Internet:
Display quasi-statique => filtrages, multiples feuilles de style, applications, etc. (web 2.0)

7. XML dans le monde de l'éducation ?

7.1 Vocabulaires typiques

- Contenus e-learning
 - par exemple norme "simple sequencing" de l'IMS
- "Quiz Markup"
- Langages pour structurés des activités d'écriture (faits maison)
- Informations sur les étudiants
 - format d'échange de données concernant les étudiants
- Textes structurés dans des bibliothèques en ligne, par ex:
 - The Oxford Text Archive
 - The Humanities Text Initiative (Univ of Michigan)
- Catalogues:
 - "Learning objects" indexés avec les normes Dublin Core ou IMS/IEEE
 - Catalogues bibliothécaires comme MARC
- "Content packaging"
 - Emballage de "learning objects"
- "Instructional design" (comme la norme "learning design" de l'IMS)

Certains standards sont déjà bien déployés (voir les "normes" IMS qui dominent le monde du e-learning).

7.2 Organisations

- Il existe plusieurs organisations qui développent des standards pédagogiques:
 - IMS (Meta-data, persons, some content, student work)
url: <http://www.imsproject.org/>
 - DoD ADL / SCORM (meta-data, contents)
url: <http://www.adlnet.org/>
 - IEEE LTSA (LOM meta-data)
url: <http://edutool.com/ltsa/>
 - Ariadne (meta-data)
url: <http://ariadne.unil.ch/>
- se coordonnent quelque peu,
- se sont concentrés d'abord principalement sur les "meta-data" et le "content packaging" (catégoriser et emballer des ressources),
- travaillent sur le concept du "reusable learning object" (des éléments que l'on peut réutiliser dans différents contextes et dans différents systèmes).
- utilisent des designs pédagogiques très simples ("simple-sequencing")
- une norme plus sophistiqué publié est "learning design", mais il manque des implémentations