

Présentation

XML



Voici à quoi ressemble un document XML :

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<!DOCTYPE Cours SYSTEM "Cours.dtd">
<Cours>
  <Titre>Cours XML</Titre>
  <Auteur>
    <Nom>Poulard</Nom>
    <Prénom>Philippe</Prénom>
  </Auteur>
  <Description>Ce cours aborde les concepts de base
    mis en œuvre dans XML.
  </Description>
</Cours>
```

C'est presque comme du HTML !

Sauf que :

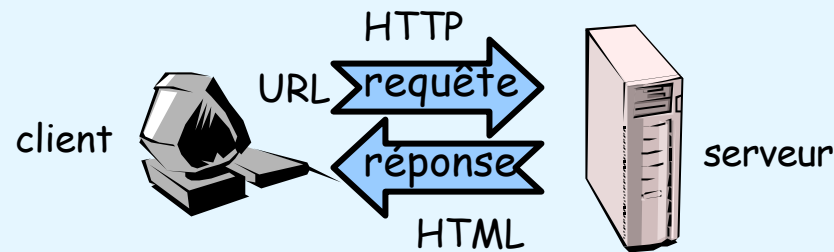
- j'ai créé **mes propres balises** pour mes propres besoins
- je suis **contraint** par une **syntaxe rigoureuse** qui n'accepte pas les erreurs

Et que :

- ce document **ne sait pas comment s'afficher** dans un navigateur

Mais que :

- je dispose d'une **boîte à outil** pour manipuler les **informations** de ce document



- Le Web peut transmettre n'importe quel type de document
- HTML s'est imposé comme le "format natif" du Web
- HTML a contribué au succès du Web
- HTML n'est qu'un **format de présentation**
- XML est un **format de structuration de l'information**

Rappels HTML

- Format ASCII-7 bits
- Une application SGML
- Simple
- Editable à la main
- Inclusion d'objets
- Liens hypertextes
- Structure documentaire
- Certains mauvais choix de conception : h1, h2, h3, h4, h5, h6, br
- Syntaxe laxiste → lourdeur du code des navigateurs
- Appropriation par les éditeurs (Netscape, Microsoft) → compatibilité difficile

```
<html>
  <head><title>Cours XML</title></head>
  <body>
    <p><b>Poulard</b> Philippe
    <p>Ce cours aborde les concepts
    de base mis en &oeilig;uvre dans XML.
  </body>
</html>
```

Une application sur le poste utilisateur,
le **navigateur** (Netscape, Internet Explorer)

Un client chez lui

Demande des documents à une machine distante,
le **serveur**

Commande un meuble

À une adresse (**URL**) `http://www.inria.fr`

Chez ACME

Via le protocole de transport **HTTP**


Par la poste

Elle reçoit un document,
en **HTML** (un "mode d'emploi")

En kit, avec le mode d'emploi

Pour s'afficher sur l'écran
grâce à des "balises" ou "tag"

Exprimé dans un langage connu (le français)
et qu'il peut interpréter
avec ses outils (tournevis, marteau)

Les balises décrivent l'effet "à peu près" souhaité  Rendus différents selon les navigateurs
 Rendus contrôlés par les navigateurs

Il y en a un nombre défini et fixe : on sait de quoi on parle

Marquage inclus dans le texte (embedded text markup)

foo bar

Données auto-descriptives

Autres solutions

• 2 fichiers séparés

bold: de 5 à 7

1234567

foo bar

• Fichier binaire

foo *■bar

ou

"foo " 27 3 "bar"

• Descriptif

(normal "foo ") (bold "bar")

ou

normal:foo ↵
bold:bar

• Etc...

Vocabulaire de balises fixes : les utilisateurs font du meccano pour exprimer ce qu'ils veulent

Navigateurs : course aux tags

- balises hétéroclites center, blink, frames...
- non-HTML : plug-in (flash, son...)

De plus en plus d'information disponible en HTML qu'on essaye de retraiter pour en extraire les informations

→ ça casse dès que les sites changent leur code HTML

Le Web en environnement professionnel, ce n'est plus du bricolage

1960	GML : Généralized Markup Language, IBM
1986	SGML : Standard Generalized Markup Language, ISO 8879, 1986 Meta-langage (permet de définir des langages de balisage) Utilisation industrielles lourdes : catalogues, documentations techniques
1992	HTML : HyperText Markup Language, CERN (Tim Berners-Lee) Une application de SGML, base du Web
Oct. 1994	Création du W3C ←
Juin 1996	Groupe de travail "SGML-light"
Fév. 1998	XML 1.0 , recommandation du W3C
Sept. 1998	Groupe de travail W3C XML divisé en 5 groupes Toutes les nouvelles normes seront en XML
1999-2000	Diversification des usages (web, database...), XML partout là où on ne l'attendait pas Augmentation de la complexité (toutes les normes prennent du retard)
2001	28 groupes de travail au W3C

SGML
Niche technologique
La communauté SGML a peur de ne pas profiter de l'essor de HTML

W3C
Le Web est un mélange de présentation et de structure

But commun :
Re-aiguiller le Web vers un successeur de SGML:
XML = SGML --

Groupe de travail constitué au sein du W3C par des gens de la communauté SGML
Conception rapide et homogène

XML, c'est SGML arrivé à maturité

XML n'est donc pas une techno JEUNE

- utilisable simplement sur internet
- internationalisation (ASCII + Unicode)
- applicable à un maximum de domaines
- compatible avec SGML
- facilité de traitement
- sans options (contrairement à SGML)
- document lisible par un humain
- spécification XML disponible rapidement
- spécification formelle et concise
- documents XML faciles à créer, même à la main
- ne cherche pas la concision du document

Pas de codage binaire
Se charge dans un bloc-note

(les documents sont d'ailleurs très verbeux)

```
<Titre>Cours XML</>  
<Auteur>  
  <Nom>Poulard</>  
  <Prénom>Philippe</>  
</>
```

et non !

On aurait pu être plus concis...

<http://w3.org>



World
Wide
Web
Consortium

Organisme à but non lucratif, 500 membres, 60 collaborateurs
3 centres :

MIT (Boston, USA)
INRIA (Sophia-Antipolis, France)
Keio University (Keio, Japon)



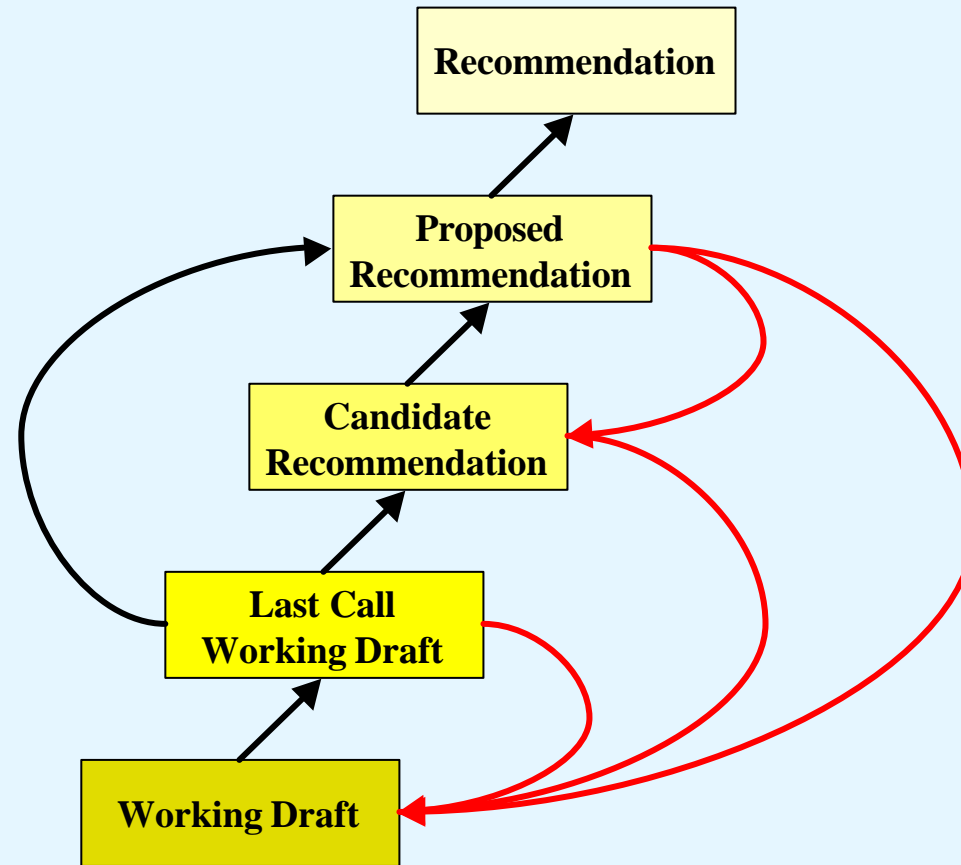
Ce qu'apporte le W3C :

- Arbitrage neutre (éditeurs de logiciel, utilisateurs)
- Expertise technique
- Processus de compromission
- Producteur de recommandations
- Producteur d'implémentations de référence (jigsaw, amaya...)

4 domaines :

- Architecture : XML, URL/URI, DOM, SOAP...
- User interface : HTML, SVG, PNG, device independance, MathML, CSS, Multimedia (SMIL), VoiceML...
- Technology & society : commerce électronique, confidentialité, web sémantique, XML encryption et signature
- WAI : Web Accessibility Initiative

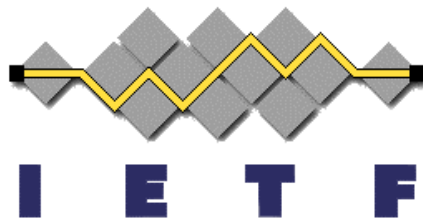
Cycle de production des standards au W3C



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

(liste non exhaustive !)

•W3C	Standards XML	World Wide Web Consortium	http://w3.org
•ISO	SGML	International Organization for Standardization	http://www.iso.ch
•IETF	RFCs	The Internet Engineering Task Force	http://www.ietf.org/
•Unicode	Caractères		http://www.unicode.org
•IANA	Registres	Internet Assigned Numbers Authority	http://www.iana.org/
•OASIS	Catalogues SGML		http://www.oasis-open.org/
•ECMA	ECMAScript	European Computer Manufacturers Association	http://www.ecma.ch



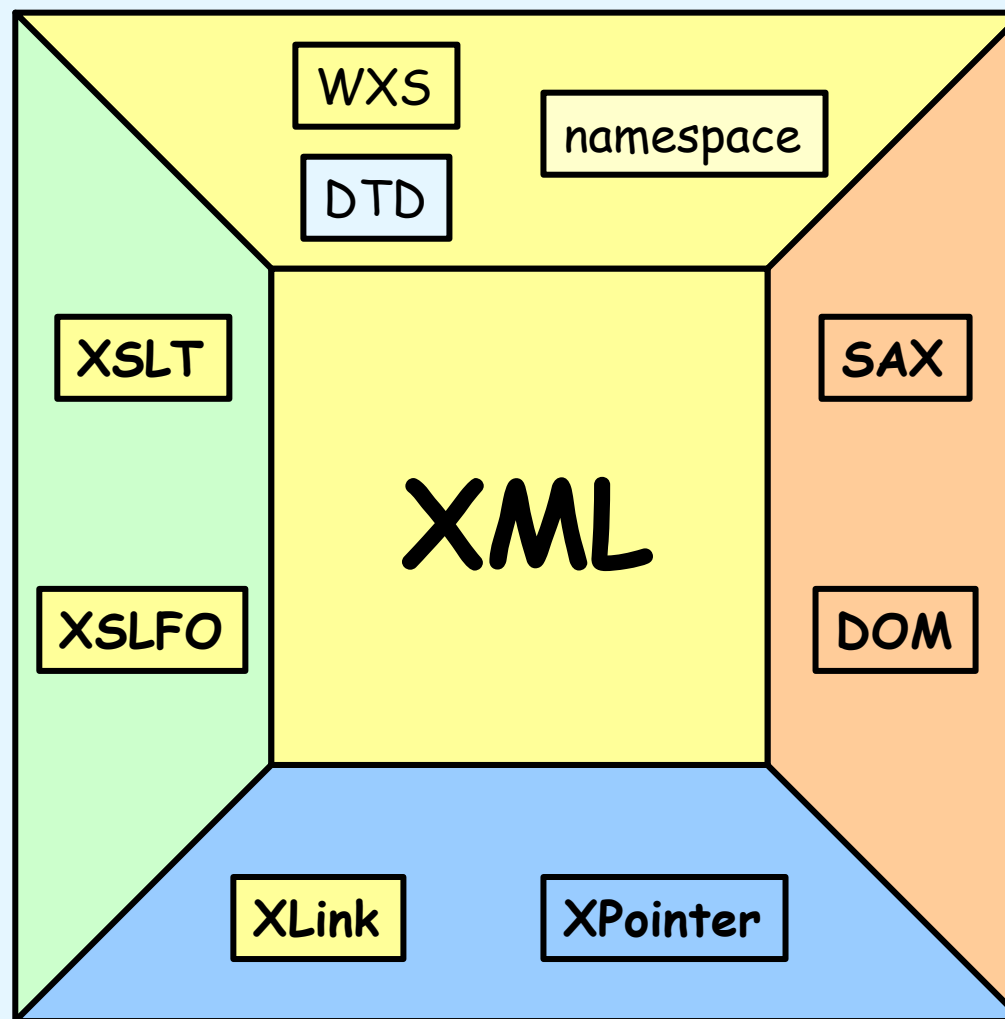
XML ne sait pas faire grand chose tout seul, à part servir pour les échanges

Standards complémentaires

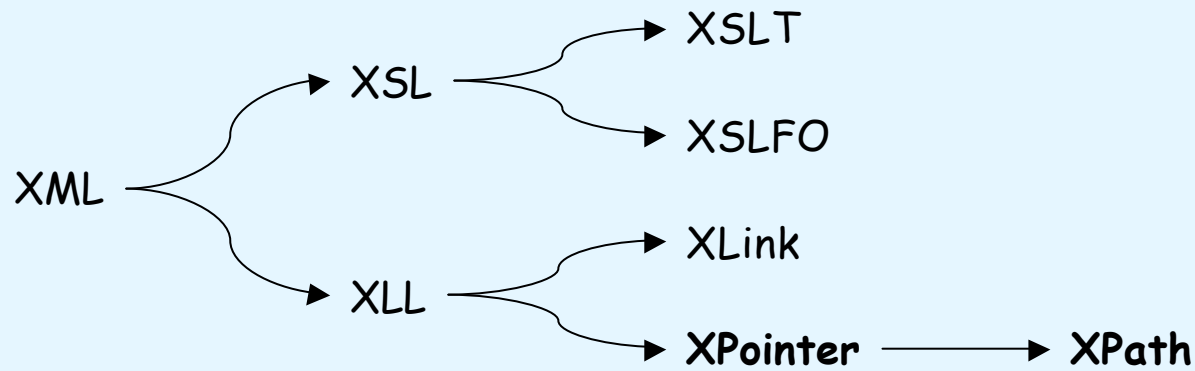
transformation et formatage

definition

manipulation



liaison



DOM
SAX

Namespace
XML Infoset

DTD
WXS
Relax-NG

XForms
XMLQuery
...

Applications de présentation :

- XHTML
- WML
- XSLFO
- SVG
- SMIL

Applications sectorielles :

- MathML
- CML
- DocBook

Applications d'échange :

- SOAP
- UDDI

Applications de description :

- RDF
- RSS
- WSDL

Affichage dans Internet Explorer → représentation arborescente du document
(pas vraiment utile pour un utilisateur)



Association d'une feuille de style à un document XML

- CSS (basique)
- XSLT (beaucoup plus puissant)

Les CSS (Cascading Style Sheets)

Permet d'associer une présentation à chaque élément d'un document

Problème CSS :

- s'applique seulement aux éléments
- dans l'ordre du document

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
- <Fable>
  <Titre>Le corbeau et le renard</Titre>
  - <Auteur>
    <Nom>De la Fontaine</Nom>
    <Prénom>Jean</Prénom>
  </Auteur>
  - <Contenu>
    <p>Maître corbeau sur un arbre
      perché, tenait dans son bec un
      fromage.</p>
  </Contenu>
</Fable>
```

```
<html>
  <head>
    <title>Fable</title>
  </head>
  <body>
    <h1>Le corbeau et le renard</h1>
    <p><i>Jean De la Fontaine</i></p>
    <p>Maître corbeau sur un arbre perché,
tenait dans son bec un fromage.</p>
.../...
  </body>
</html>
```

format de présentation
HTML

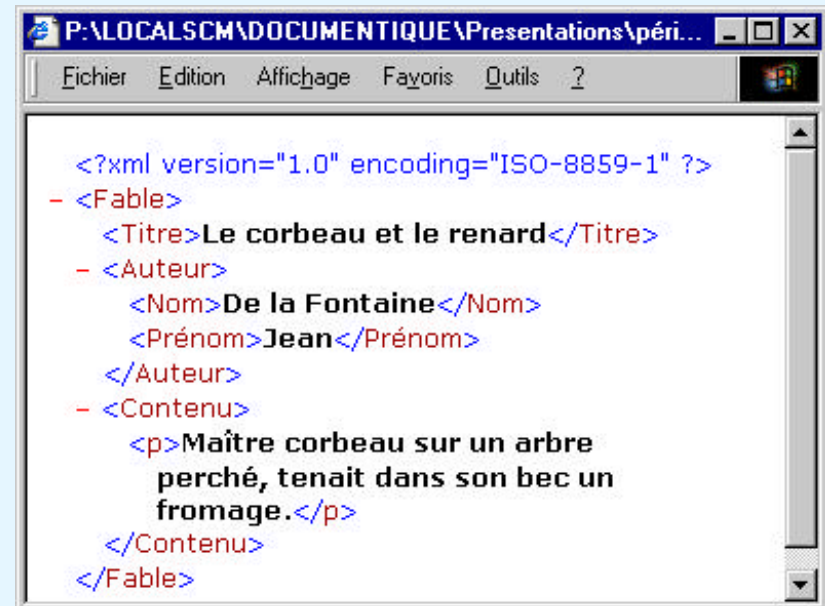
XML

format de structuration sémantique

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<!DOCTYPE Fable SYSTEM "Fable.dtd">
<Fable>
  <Titre>Le corbeau et le renard</Titre>
  <Auteur>
    <Nom>De la Fontaine</Nom>
    <Prénom>Jean</Prénom>
  </Auteur>
  <Contenu>
    <p>Maître corbeau sur un arbre perché,
tenait dans son bec un fromage.</p>
.../...
  </Contenu>
</Fable>
```



XSLT



Document HTML

jeu de balises fini et connu

structure orienté présentation exclusivement

wysiwyg qui dépend du navigateur (!!)

réutilisation limitée (par la quantité de documents HTML)

cible : web exclusivement

écriture anarchique, pas de respect de charte graphique possible, temps perdu alloué à la présentation

stockage : SGF

Document XML

jeu de balises extensible, à priori non connu

toutes structures (sémantique) données auto-descriptives

pas de présentation intrinsèque possible : obligation de feuille de style

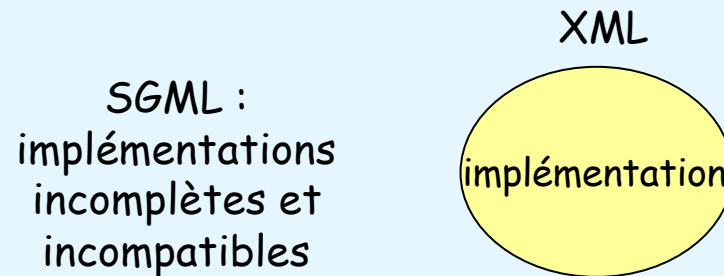
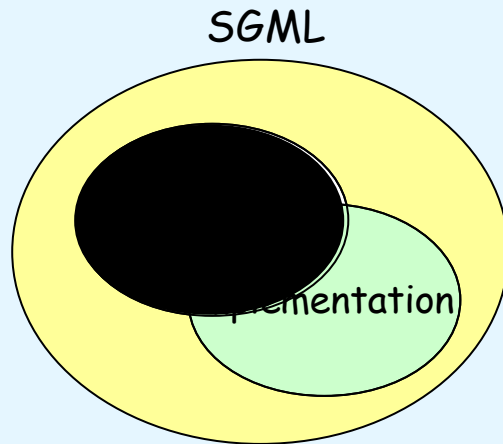
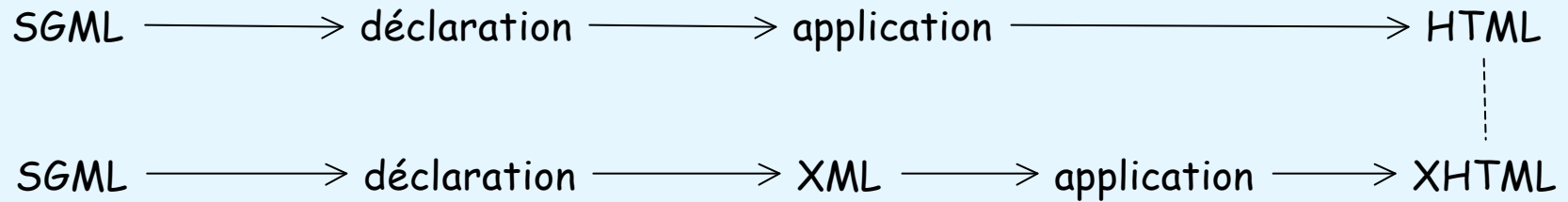
exploitation sémantique possible, transformation de structure

cible : Web, Wap, PDF, texte, autres échanges entre applications

rédaction de contenu exclusivement

stockage : SGF, BdDR, natif (XND)

XML est un sous-ensemble de SGML



Profils d'applications

Exemple pour les feuilles de style : DSSSL, FOSI os CALS-B

XML :

- Déclaration fixe (syntaxe fixe)
- Infiniment plus simple à analyser (parser) que SGML