

Le langage XHTML

Emmanuel Bruno

février 2006

- 1 Présentation générale
- 2 Description du langage XHTML
 - La syntaxe et le modèle
 - L'entête du document (`head`)
 - Le corps du document (`body`)
 - Les éléments de niveau "bloc"
 - Les listes
 - Les éléments de niveau "texte"
 - Les liens
 - Les tableaux
 - Le arquage génériques

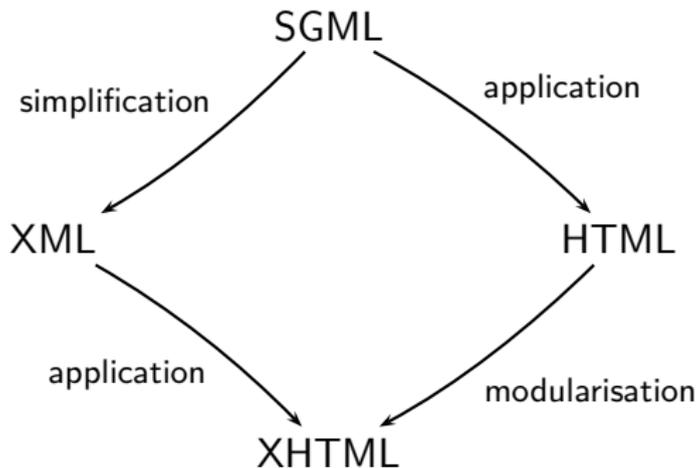
Éléments du langage XHTML

- (X)HTML (Extensible) Hypertext Markup Language
- **Hypertexte** : permet de créer des *liens hypertextuels*.
- C'est un **langage de marquage** de document, il permet de décrire la structure d'un document électronique sous forme lisible par un être humain et interprétable par une machine.

Definition

Dans un document, un *lien hypertexte* relie un mot, une expression ou une image à une autre partie de celui-ci, à un autre document ou une de ses parties. Habituellement, un clic sur un lien hypertexte permet d'accéder à la cible qu'elle soit locale ou sur un serveur distant.

Langages de balisage normalisés



Historique de HTML

- HTML est une application (parmi d'autres) de SGML.

Definition

SGML définit un langage de balisage général de documents, HTML est un ensemble de balises prédéfinies définies dans une grammaire appelée définition de type de document (DTD).

- HTML 1 : 1989 (texte, hyperliens) (Tim BERNERS-LEE, CERN)
- HTML 2 : 1994 (images, formulaires)
- HTML 3 : 1996 (son, applets)
- HTML 4 : 1998 (vidéo, CSS)
- Actuellement :
 - Recommandation du W3C (World Wide Web Consortium)
<http://www.w3.org/TR/html4/>
 - ISO-HTML : ISO/IEC 15445 :2000

XHTML

- XHTML 1.0 (2nd edition) :
<http://www.w3.org/TR/xhtml11/>
 - Extensible HyperText Markup Language Recommand. W3C janvier 2000 (rév. aout 2002) "A Reformulation of HTML 4 in XML 1.0"
 - Basé sur HTML 4.01 et XML (et non SGML)
- XHTML 1.1 : Module-based XHTML
<http://www.w3.org/TR/xhtml111/>
 - Recommandation W3C mai 2001
 - Introduit la possibilité d'extensions modulaires (par exemple, XForms pour les formulaires)

- (X)HTML – Avantages
 - Simple et portable
 - Formats textes et normalisés
 - Outils gratuits et faciles à utiliser
 - Navigateurs nombreux et présents sur toutes les plateformes
 - Parfait pour des documents simples
- (X)HTML – Limites
 - Difficile de contrôler parfaitement le rendu visuel des documents
 - PDF ou XSL-FO sont plus adaptés
 - Difficile de produire des documents imprimables
 - Ex : pas de notion de référence à un numéro de page

Fondamental : le respect de la norme

- Validité syntaxique
- Certains navigateurs sont trop laxistes :
 - permettent une auvaise imbrication des balises
 - autorise l'absence d'éléments obligatoires (head, body, etc.)
- Présence de balises propriétaires (non normalisées)
 - reconnues seulement par certains navigateurs spécifiques
- Conséquence : des pages non compatibles !
- Outils de validation : <http://validator.w3.org/>

(X)HTML strict et non strict

- Les objectifs de HTML 4 et de XHTML sont :
 - de recentrer le langage sur sa fonction principale
 - de séparer structure logique et formatage
 - de tendre vers le *web sémantique* <http://www.w3.org/2001/sw/>
 - *"The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation."*
T. Berners-Lee, J. Hendler, O. Lassila, *The Semantic Web*, Scientific American, May 2001
- Pour faciliter la transition deux versions :
 - Non strict :
 - Balisage orienté tantôt sur la structure logique de l'information, tantôt sur son formatage
 - Exemple : balises pour l'italique (`<i>...</i>`)
 - Strict :
 - Séparation claire entre contenu et présentation
 - Plus grande réutilisabilité de l'information

Exemple de document XHTML

```
<?xml version="1.0" encoding="iso-8859-1"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD_XHTML_1.0_Strict//EN"
  "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="fr-fr">
  <head>
    <title>Le langage XHTML</title>
  </head>
  <body>
    <h1>Le langage XHTML</h1>
    <hr />
    <p>Vous devriez aller sur le
      <a href="http://www.w3.org">site</a>
      du <strong>w3c</strong>
    </p>
  </body>
</html>
```

Listing 1 – ./Exemples/full.xhtml

Structuration avec des balises

- La structuration du document est indiquée dans le texte avec des *balises* (étiquettes, tags)
- Les balises commencent par < et se terminent par >
- Un *élément* XHTML possède un nom et est délimité par une balise ouvrante et une balise fermante.
- Par exemple :
 - `<title>Ma page personnelle</title>`
 - `<p>Un petit texte</p>`

Elements vides

- Certains éléments appelés *éléments vides* ne marquent pas du texte.
- Il n'ont pas de contenu, les balises de début et de fin peuvent être combinées :
 - `` devient ``

Ajout d'attributs

- Des *attributs* peuvent être ajoutés à un élément
- Ils précisent celui-ci sans faire partie de son contenu
- Le nom de l'attribut apparaît après le nom d'élément dans la balise ouvrante et est suivi du signe "=" puis de la valeur de l'attribut entre guillemets.
- Les attributs d'un élément forment un ensemble (nom unique, pas d'ordre).

```

```

Imbrication

- Les éléments peuvent être imbriqués les uns dans les autres :
`<p>Retenez ce mot <emph>abracadabra</emph></p>`
- Il ne peut pas y avoir de **chevauchement**. Tout élément doit être fermé avant ceux qui le contiennent.
- Un document ainsi structuré forme donc un arbre dont les nœuds sont les éléments et les feuilles des fragments du texte.

Structure d'un document XHTML

- Un document XHTML est un document XML composé de la façon suivante :
 - il commence par un prologue qui indique la version de XML et le codage des caractères :
 - `<?xml version="1.0"encoding="iso-8859-1"?>`
 - **OU** `<?xml version="1.0"encoding="utf-8"?>`
 - puis il indique la version du HTML ou XHTML
 - `<!DOCTYPE html PUBLIC "-//W3C//DTD_XHTML_1.0_Strict//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">`
 - le document lui-même est composé d'un élément `html` qui contient deux éléments :
 - `head` l'entête du document
 - `body` le contenu du document

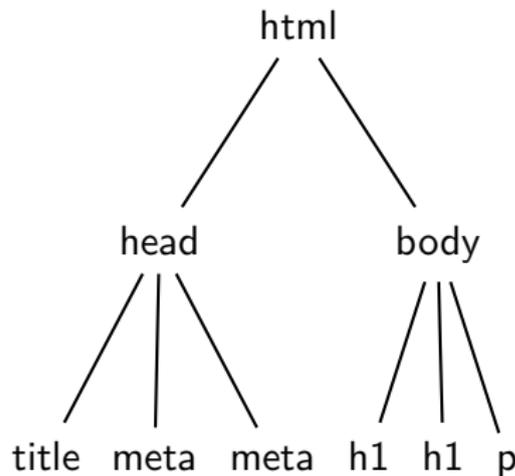
Un document XHTML

```
<?xml version="1.0" encoding="iso-8859-1"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD_XHTML_1.0_Strict//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
  <head>
    <title>Mon document</title>
  </head>
  <body>
    <h1>Mon document</h1>
    <p>Ce document est très simple.</p>
    <p>Nous verrons d'autres exemples ensuite.</p>
  </body>
</html>
```

Listing 2 – ./Exemples/premier.xhtml

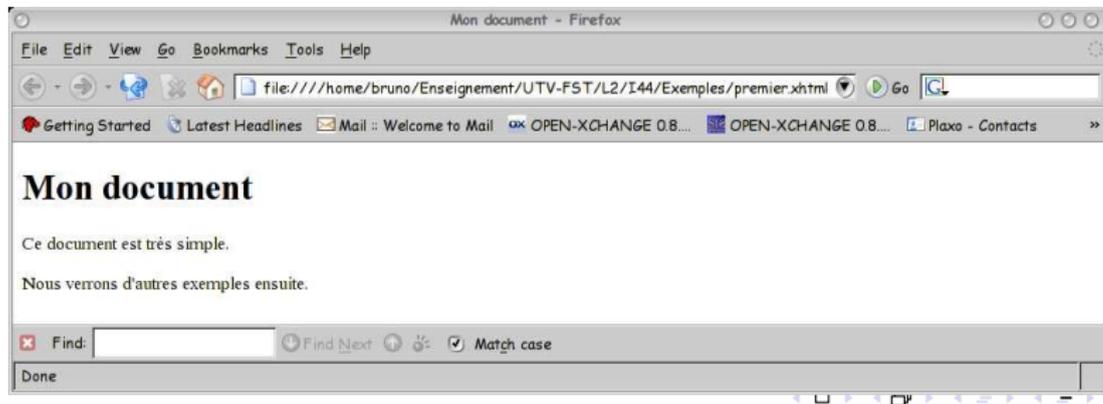
Structure arborescente d'un document

- Au delà de l'aspect syntaxique, un document peut être vu formellement comme un arbre
- L'ordre des nœuds et leur imbrication précisent la structure du document
- Cette structure est exploitable :
 - Quels sont les titre et auteur du document ?
 - Combien y-a-t-il de paragraphes ?
 - Quel est le titre de niveau 2 qui précède le 2^{ème} paragraphe



Un document visualisé dans un navigateur

- Le même document peut être visualisé dans un navigateur, la structure logique est interprétée pour fournir une présentation
- Un document peut donc être considéré à trois niveaux : physique (syntaxe), logique (structure d'arbre) et présentation (le rendu visuel, audio, ...)



Structure de l'entête (head) (1/2)

- L'entête doit comprendre :
 - Un titre (élément `title`) qui apparaît dans le haut de la fenêtre du navigateur.
- Il peut aussi comprendre
 - Une ou plusieurs références à d'autres documents (élément `link`)

- en indiquant une relation

`http://www.w3.org/TR/REC-html40/types.html#type-links` (Alternate, Stylesheet, Start, Next, Prev, Contents, Index, Glossary, Copyright, Chapter, Section, Subsection, Appendix, Help, Bookmark)

- Stylesheet : feuilles de styles pour la présentation

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01//EN"
    "http://www.w3.org/TR/html4/strict.dtd">
<HTML>
<HEAD>
  <TITLE>Chapter 2</TITLE>
  <LINK rel="Index" href="../index.html">
  <LINK rel="Next" href="Chapter3.html">
  <LINK rel="Prev" href="Chapter1.html">
</HEAD>
...the rest of the document...
```

Structure de l'entête (head) (2/2)

- L'entête peut aussi contenir des métadonnées (élément meta)

```
<HEAD profile="http://www.acme.com/profiles/core">  
  <TITLE>How to complete Memorandum cover sheets</TITLE>  
  <META name="author" content="John Doe">  
  <META name="copyright" content="© 1997 Acme Corp.">  
  <META name="keywords" content="corporate, guidelines, cataloging">  
  <META name="date" content="1994-11-06 T08:49:37+00:00">  
</HEAD>
```

- Ces informations ne sont généralement pas affichées, mais sont très utiles (avec des réserves) en particulier pour les moteurs de recherche qui indexent les documents sur le web.
- <http://www.w3.org/TR/REC-html40/appendix/notes.html#recs>

Exemple d'entête

```
<?xml version="1.0" encoding="iso-8859-1"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD_XHTML_1.0_Strict//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
  <head>
    <title>Page personnelle de E. Bruno</title>
    <meta name="author" content="Emmanuel_Bruno"/>
    <meta name="description"
          content="un_premier_exemple_de_document"/>
    <link rel="stylesheet" type="text/css"
          href="styles.css" />
  </head>
  <body> <h1>Mon document</h1> </body>
</html>
```

Listing 3 – ./Exemples/entete.xhtml

éléments de niveau “bloc”

- Les éléments de niveau bloc définissent la structure générale du document
- Ils se succèdent mais ne peuvent pas s'imbriquer
- Ils causent pour la plupart un saut de ligne avant et après
- Exemples : p, address

Entêtes de section

- Les éléments `h1`, `h2`, `h3`, ..., `h6` sont des entêtes de section (titre et sous-titres)
- Les chiffres réfèrent au niveau hiérarchique : `h1` est le titre de plus haut niveau et ne doit apparaître qu'une seule fois dans un document ; `h2` à `h6` sont de plus bas niveau et chacun peut apparaître plus d'une fois
- `<h1>Le langage XHTML</h1>`

Divers blocs

- `P` : un paragraphe
- `address` : une adresse
- `pre` : du texte preformaté (sauts de ligne conservés, police courrier)
- `blockquote` : une citation (Le texte est affiché en retrait par rapport aux paragraphes).

Des exemples de bloc

```
<?xml version="1.0" encoding="iso-8859-1"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD_XHTML_1.0_Strict//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
  <head><title>Mon document</title></head>
  <body>
    <p>Ceci est le texte de mon premier paragraphe .
    Il comprend une ou plusieurs phrases.</p>
    <pre>
      ce texte
      est formatté
    </pre>
    <blockquote>Eurêka , j'ai trouvé</blockquote>
    <address>Université du Sud Toulon-Var</address>
  </body>
</html>
```

Listing 4 – ./Exemples/bloc.xhtml

Les séparateurs

- Deux éléments vides :
 - `br` : force un saut de ligne.
 - `hr` : ligne horizontale

Les listes

- Liste simple ou ordonnée
 - Liste non numérotée (à puces) : `ul`
 - Liste numérotée : `ol`
 - Contiennent l'élément `li` utilisé pour chacun des items de la liste
- Liste de définitions : `dl`
 - Contient des `dt` (le terme) et des `dd` (la définition) en alternance

Un exemple de listes

```
<!DOCTYPE html PUBLIC "-//W3C//DTD_XHTML_1.0_Strict//EN"  
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">  
<html xmlns="http://www.w3.org/1999/xhtml">  
  <head><title>Les listes</title></head> <body>  
    <h2>Une liste non ordonnees</h2> <ul>  
      <li>Un item</li>  
      <li>Un autre item</li>  
    </ul>  
    <h2>Une enumeration</h2> <ol>  
      <li>le premier item</li>  
      <li>le second item</li>  
    </ol>  
    <hr/>  
    <h2>Une liste de definitions</h2> <dl>  
      <dt>XML</dt><dd>Extensible Markup Language</dd>  
      <dt>XHTML</dt><dd>Extensible HyperText Markup Language</dd>  
    </dl>  
  </body> </html>
```

Listing 5 – ./Exemples/listes_xhtml

Les éléments de niveau texte

- Les éléments de niveau texte permettent d'attribuer un rôle spécifique à une partie d'un passage textuel
- Ils n'interrompent pas le flot du texte (pas de retour à la ligne)
- En particulier :
 - `em` : emphase, mise en évidence de mots dans un texte (italique?)
 - `strong` : emphase forte (gras?)
 - `cite` : titre d'un document cité en référence (italique?).

Exemple de structuration du texte

```
<?xml version="1.0" encoding="utf-8"?>
<?xml version="1.0" encoding="iso-8859-1"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD_XHTML_1.0_Strict//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
  <head>
    <title>Le texte en HTML</title>
  </head>
  <body>
    <emph>M. Tim Berners-Lee</emph> est l'inventeur
    du <strong>Web</strong>, lisez son livre
    <cite>Weaving the Web</cite>
  </body>
</html>
```

Listing 6 – ./Exemples/texte.xhtml

Les ancres

- Pour pouvoir référencer des parties d'un document, on définit des ancres
- En XHTML, on ajoute l'attribut `id` à un élément qui sera la destination de la référence. La valeur de cet attribut doit être unique dans tout le document.
 - `<p id="intro">Voilà l'introduction</p>`
- Par compatibilité avec HTML, on peut aussi ajouter un élément `A` portant un attribut `name`.
 - `<p>Voilà l'introduction</p>`

Les liens internes

- Il est possible de définir des liens internes, c'est-à-dire vers une section du même document. On utilise l'élément `A`, l'attribut `href` indique l'ancre (préfixée par `#`), le texte (ou l'image) contenu dans l'élément est utilisé comme étiquette (présenté comme un lien cliquable).
 - `<p>Ce point est abordé dans`
l'`introduction</p>`

Les liens internes exemple

```
<?xml version="1.0" encoding="iso-8859-1"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD_XHTML_1.0_Strict//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
  <head> <title>Mon document</title> </head>
  <body> <h1>Mon document</h1>
    <h2>Table des matières</h2> <ul>
      <li><a href="#intro">Introduction</a></li>
      <li><a href="#concl">Conclusion</a></li>
    </ul>
    <h2>Contenu</h2>
    <h3 id="intro">Introduction</h3>
    <p>bla bla</p>
    <h3 id="concl">Conclusion</h3>
    <p>allez lire la <a href="http://www.w3.org/TR/xhtml1/">
norme</a> et plus précisément la parties sur les
<a href="http://www.w3.org/TR/xhtml1/#h-4.10">ancres</a>.
</p> </body> </html>
```

Listing 7 - ./Exemples/liens.xhtml

Les URL

- Une URL est une expression qui permet d'adresser une ressource sur Internet
- `<protocole>://[user[:password]@<machine>[:port]][/<path>[\#label|?liste paramêtres >]]`
- `<protocole>` : méthode d'accès au document
- `<machine>` : adresse de la machine
- `<port>` : numéro de port sur la machine
- `<path>` : chemin d'accès au document sur la machine

Les hyperliens

- Il est possible de définir des liens vers d'autres documents XHTML (ou des parties) d'autres protocoles.
- On utilise toujours l'élément `A`, l'attribut `href` est utilisé pour spécifier l'adresse (URL) de la ressource vers laquelle on veut pointer (absolue ou relative)
 - ` Site de l'USTV`
 - ` Page de E. Bruno`
 - ` Publications de E. Bruno`
 - `contact`
 - `Publications de E. Bruno`
 - `Téléchargement`

Les images

- On insère une image avec l'élément `img`.
- L'attribut `src` indique l'URL de l'image
- L'attribut `alt` donne une description textuelle alternative de l'image.
- ``

Les tableaux

- L'élément `table` permet de décrire un tableau ("bloc")
- Il contient :
 - Un attribut `summary` qui décrit le contenu.
 - Eventuellement un élément `caption` qui donne sa légende
 - Eventuellement un élément `thead` qui décrit la l'entête
 - Eventuellement un élément `tfoot` qui décrit la le pied
 - des éléments `tbody` qui décrivent le contenu, composés de
 - Un suite de `TR` (lignes) qui contiennent des `TD` (cellules) ou des `TH` (cellules d'entête).
 - Les attributs `colspan` et `rowspan` indiquent combien de colonnes ou de lignes les `TD` et `TH` occupent.

Un exemple de tableau

```
<?xml version="1.0" encoding="iso-8859-1"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
  <head><title>Un tableau</title></head> <body>
    <table summary="Ce tableau décrit l'âge des enfants">
      <caption>Age des garçons et filles</caption>
      <thead>
        <tr> <td colspan="2">Identité</td><td>Autre</td></tr>
        <tr> <th>Nom</th> <th>Prénom</th> <th>Age</th></tr>
      </thead>
      <tfoot>
        <tr> <th>Nom</th> <th>Prénom</th> <th>Age</th></tr>
      </tfoot>
      <tbody>
        <tr><td>Pierre</td><td>Martin</td><td>12 ans</td></tr>
        <tr><td>Jean</td> <td>Durand</td><td>13 ans</td></tr>
      </tbody>
      <tbody>
        <tr><td>Nicole</td><td>Martin</td><td>15 ans</td></tr>
        <tr><td>Sophie</td><td>Durand</td><td>16 ans</td></tr>
      </tbody>
    </table>
  </body></html>
```

Listing 8 – ./Exemples/tablex.xhtml

DIV et SPAN

- Dans le cas où aucun élément standard ne correspond à une zone qui doit être marquée, on utilise `div` (niveau "bloc") ou `span` (niveau "texte")
- On leur associe l'attribut `class` ou l'attribut `id` qui permet de préciser quoi il s'agit
- Ces deux éléments seront très utilisés plus tard lors de la présentation avec des feuilles de style CSS.

Un exemple d'utilisation de DIV et SPAN

```
<?xml version="1.0" encoding="iso-8859-1"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD_XHTML_1.0_Strict//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
  <head><title>DIV et SPAN</title></head>
  <body>
    <div class="entete">Ma page oueb</div>

    <div class="corps">Bienvenue sur la page d'
    <span class="personne">Emmanuel Bruno</span>
    vous pouvez me contacter au
    <span class="tel">555-456-589</span>
    </div>
  </body>
</html>
```

Listing 9 – ./Exemples/divspan.xhtml