

Cours Web n°1

Introduction au World Wide Web et à XHTML

Pierre Senellart (pierre@senellart.com)



14 et 17 septembre 2007

Plan du cours

- 1 Internet et le Web
- 2 Le langage XHTML
- 3 XHTML et HTML
- 4 Références
- 5 Application

- Qu'est-ce que le Web (ou **World Wide Web**, Toile, WWW, W3) ?
 - ▶ Système **hypertexte** public : système contenant des documents liés entre eux par des hyperliens permettant de passer automatiquement d'un document à l'autre.
- Différence entre le Web et Internet ?
 - ▶ Internet : réseau mondial d'ordinateurs permettant aux utilisateurs de communiquer (courrier électronique), de publier des informations (Web), de transférer des données (FTP), de travailler à distance (telnet et ssh)...
 - ▶ Web : un aspect d'Internet.

Le client (navigateur : Internet Explorer, Mozilla, Firefox...)

- demande au serveur des informations
- affiche des pages pour l'utilisateur

Le serveur (Apache, Microsoft IIS...)

- reçoit en permanence les requêtes du client
- renvoie les documents correspondants

Le réseau Internet

- basé sur le protocole TCP/IP
- chaque machine dispose d'un numéro IP et d'un nom associé grâce aux serveurs de nom du domaine (DNS)

Exemple

Nom de machine	<code>www.ifips.u-psud.fr</code>
Adresse IP	<code>129.175.250.38</code>

Protocole Ensemble normalisé de règles décrivant la manière de transmettre des informations, par exemple sur un réseau comme Internet entre un client et un serveur.

HTTP HyperText Transfer Protocol, le plus utilisé des protocoles de communication sur le World Wide Web. Permet à un client Web d'indiquer quelle page il veut obtenir, et au serveur Web de lui répondre en lui donnant cette page.

- URL : *Uniform Resource Locator*
- Identifie l'**endroit** où se trouve une **ressource** sur le Web.
- Dans le cas du Web, ressource = **document** ou **fragment**
- `http://lea-linux.org/reseau/secu/firewall.html#intro`
protocole machine répertoire fichier fragment
- Principaux protocoles utilisés dans les URL : ftp, http, https, news, nntp, mailto, telnet, etc.

- Navigateurs graphiques
 - ▶ Internet Explorer (version 7 sortie fin 2006)
 - ▶ Moteur **Gecko** : Firefox, Mozilla, Camino, SeaMonkey (libres)
 - ▶ Opera (gratuit)
 - ▶ Moteur **KHTML** : Konqueror (libre), Safari
- Navigateurs textuels : w3m, lynx, links (libres)
- Autres navigateurs : navigateurs auditifs, etc.
- Mais aussi : robots des moteurs de recherche, logiciels de traduction automatique. . .

- Navigateurs graphiques
 - ▶ Internet Explorer (version 7 sortie fin 2006)
 - ▶ Moteur **Gecko** : Firefox, Mozilla, Camino, SeaMonkey (libres)
 - ▶ Opera (gratuit)
 - ▶ Moteur **KHTML** : Konqueror (libre), Safari
- Navigateurs textuels : w3m, lynx, links (libres)
- Autres navigateurs : navigateurs auditifs, etc.
- Mais aussi : robots des moteurs de recherche, logiciels de traduction automatique. . .

En théorie, tout site Web devrait être testé sur chacun de ces navigateurs, dans leurs différentes versions. Pour simplifier, on testera un site (au moins) :

- Avec des **validateurs**, pour vérifier que le code est correct.
- Avec 2 ou 3 des navigateurs **les plus courants** (p.ex. IE, Firefox, Opera).
- Si possible, avec un navigateur **en mode texte**.

Un serveur Web : hôte sur lequel fonctionne un serveur HTTP. Il héberge les ressources qu'il sert.

Un navigateur Web : logiciel client HTTP conçu pour accéder aux ressources du Web. Sa fonction de base est de permettre la consultation des documents HTML disponibles sur les serveurs HTTP.

Une page Web : document destiné à être consulté avec un navigateur Web. Elle est constituée d'une ressource centrale et d'éventuelles ressources liées automatiquement.

Un site Web : ensemble de pages Web et d'éventuelles autres ressources, liées dans une structure cohérente, publiées par un propriétaire (une entreprise, une administration, une association, un particulier, etc.) et hébergées sur un ou plusieurs serveurs Web.

Un hébergeur Web : entreprise de services informatiques hébergeant (mettant en ligne) sur ses serveurs Web les ressources constituant les sites Web de ses clients.

Hyperlien : lien logique, unidirectionnel (codé par une balise HTML) reliant une page Web à une autre.

Surfer, Naviguer : parcourir le Web en suivant les hyperliens, de page Web en page Web.

Plan du cours

- 1 Internet et le Web
- 2 Le langage XHTML**
- 3 XHTML et HTML
- 4 Références
- 5 Application

Caractéristiques :

- successeur de HTML (langage simple utilisé pour générer des documents portables),
- conforme aux normes XML (eXtended Markup Language),
- développé par le W3C (World Wide Web Consortium) regroupant industriels (Microsoft, IBM, Apple. . .) et académiques (INRIA, MIT. . .)

Principes :

- format **ouvert** : lecture possible dans des conditions correctes sans contrainte matérielle ou logicielle
- description de la **structure** d'un document, accent sur l'**accessibilité**

Conséquence :

votre code doit pouvoir être validé par un validateur tel que <http://validator.w3.org/>

- XHTML permet de décrire la **structure** d'un document. Ne pas chercher à décrire la **présentation** (ce sera fait avec CSS)!
- XHTML est un langage qui alterne texte et **balises** (`<blabla>` ou `</blabla>`)
 - ▶ Les balises définissent les propriétés de chaque partie du document et sont analysées par le navigateur pour réaliser la mise en page du document.
- Les fichiers XHTML
 - ▶ sont des fichiers texte
 - ▶ ont (conventionnellement) pour extension *.html*
 - ▶ sont structurés en deux parties principales : l'en-tête `<head> ... </head>`) et le corps `<body> ... </body>`)
- En XHTML, les blancs (espace, tabulations, retours à la ligne) sont en général équivalents et servent juste à délimiter mots, balises... Leur nombre n'a pas d'importance.

- Leur syntaxe est (balises ouvrante et fermante)

```
<balise attributs>contenu</balise>
```

ou (balise sans contenu)

```
<balise />
```

où :

balise mot clé associé à une propriété particulière

contenu peut représenter du texte ou d'autres balises

attributs représente les différents paramètres associés à la balise, sous la forme d'une liste de `nom="valeur"` ou `nom='valeur'`, séparés par des espaces

- Les noms des balises et les types des attributs sont toujours écrits en minuscule.
- Les balises sont ouvertes et refermées dans l'ordre (`<i></i>` et non `<i></i>`).
- Des règles strictes déterminent quelles balises peuvent être mises à l'intérieur de quelles balises.

Exemples

- `
` pour obtenir un saut de ligne (en poésie, par exemple)
- `<title>coucou</title>` pour attribuer le titre *coucou* au document
- `cuicui` pour mettre en *emphase* le texte *cuicui* (cela sera rendu, le plus souvent, par une mise en italique).
- `cuicui` pour indiquer que le texte **cuicui** est important (cela sera rendu, le plus souvent, par une mise en gras).
- `<!--zut.-->` pour ajouter un commentaire, ici *zut.*, dans le document, qui ne sera ni affiché ni interprété par le navigateur.

Contre-exemple

```
<strong><em>bouh</strong></em>
```

```
<!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"
    lang="fr" xml:lang="fr">
  <head>
    <!-- En-tête du document -->
  </head>
  <body>
    <!-- Corps du document -->
  </body>
</html>
```

- La déclaration `<!DOCTYPE ...>` précise la version d'(X)HTML utilisée.
- La langue du document est précisée avec les deux attributs (redondants!) `lang` et `xml:lang` de la balise principale `<html>`.

- L'**en-tête** du document est délimitée par les balise `<head> ... </head>` qui n'admet pas d'attributs.
- L'en-tête contient des **méta-informations** concernant le document telles que son titre, son encodage, les fichiers annexes, etc. Pour commencer, on indiquera :
 - ▶ Le jeu de caractères de la page, à mettre **tout au début** de l'en-tête

```
<meta http-equiv="Content-Type"  
      content="text/html; charset=utf-8" />
```

- ▶ Le titre de la page, l'information la plus importante (et la seule requise); celui-ci sera par exemple affiché dans la barre de titre du navigateur, il n'apparaît pas dans la page elle-même.

```
<title>Le site des adorateurs de la vache</title>
```

Unicode : **répertoire de caractères**, assignant à chaque caractère, de quelque langue que ce soit, un nombre entier.

Exemples

A	→	65		ε	→	949
é	→	233		ℵ	→	1488

Jeu de caractères : moyen de représenter concrètement, par une suite de 0 ou de 1, un caractère Unicode.

Exemples (é)

iso-8859-1	11101001		Seulement pour certains caractères
utf-8	11000011	10101001	
utf-16	11101001	00000000	

On utilisera uniquement **utf-8**.

- Les balises `<body> ... </body>` délimitent le **corps** du document.
- Le corps est **structuré** en sections, paragraphes, listes, etc.
- Il existe 6 balises permettant de représenter les titres de **sections**, par importance décroissante :
 - ▶ `<h1>Titre de la page</h1>`
 - ▶ `<h2>Titre de section principale</h2>`
 - ▶ `<h3>Titre de sous-section</h3>`
 - ▶ `<h3>Titre de sous-sous-section</h3>`
 - ▶ ...
- Les balises `<p> ... </p>` permet de délimiter un **paragraphe**. Tous les paragraphes de texte doivent être balisés ainsi.
- La balise `<hr />` (**horizontal rule**) indique une séparation majeure dans le document (rendue par exemple graphiquement par une ligne horizontale).

- XHTML possède plusieurs balises permettant de présenter le texte sous forme de listes.
- On en distingue trois types :
 - ▶ les listes non numérotées,
 - ① les listes numérotées,
- les listes de définitions (ou lexiques)
- Ces listes peuvent être emboîtées les unes à l'intérieur des autres.

- Les listes classiques :
 - ▶ Les listes non numérotées délimitées par les balises ` ... ` (**unordered list**).
 - ▶ Les listes numérotées délimitées par les balises ` ... ` (**ordered list**).
 - ▶ Tous les éléments d'une liste numérotée ou non sont délimités par les balises ` ... ` (**list item**)
- Les lexiques sont délimités par les balises `<dl> ... </dl>` (**definition list**) et leurs entrées par les balises `<dt> ... </dt>` (**term**) et `<dd> ... </dd>` (**definition**).

Exemples

```
<ol> <li>un</li> <li>deux</li> </ol>
```

```
<dl> <dt>lapin</dt> <dd>rongeur à oreilles</dd> </dl>
```

- Les tableaux sont délimités par les balises `<table>... </table>` .
- Les balises `<tr>... </tr>` (**table row**) délimitent les lignes.
- Les balises `<td>... </td>` (**table data**) délimitent les cellules.
- **Attention !** On déclare les lignes à l'intérieur du tableau, les cellules à l'intérieur des lignes.

Exemple

```
<table>
  <tr> <td> 11, c1 </td> <td> 11, c2 </td> </tr>
  <tr> <td> 12, c1 </td> <td> 12, c2 </td> </tr>
</table>
```

Contre-exemple

```
<td>
  <tr> c1, 11 </tr>
  <tr> c1, 12 </tr>
</td>
```

Ajouter de la structure à un tableau en :

- donnant une **légende** au tableau avec les balises `<caption>... </caption>` juste après la balise ouvrante `<table>`.
- remplaçant les `<td> ... </td>` qui contiennent des en-têtes (de ligne, de colonne) par des `<th> ... </th>` (**table header**).

- Pour insérer une **image** dans un document XHTML, on utilise la balise ``.
 - ▶ L'attribut `src` permet de préciser où se trouve l'image.
 - ▶ L'attribut `alt` permet de remplacer l'image par un texte quand elle n'est pas disponible. Il est obligatoire de l'utiliser, pour que tout agent (malvoyants, navigateur texte, incidents techniques, robots) ne pouvant voir votre image puisse avoir un **texte alternatif**.

```
  

```

- Les formats autorisés sont :
 - ▶ Le JPEG (.jpg), un format adapté aux photos
 - ▶ Le GIF (.gif) et le PNG (.png), des formats adaptés aux autres types d'image ; le format GIF permet d'avoir des images animées (à utiliser avec parcimonie!), le PNG est à utiliser sinon (transparence, profondeur de couleurs...).

- Ce qui différencie une page Web (page HyperTexte) d'un banal document : ce sont les **liens** !
- Ils sont introduits par la balise `<a> ... `.
- En cliquant sur un lien, on peut se déplacer vers :
 - ▶ un autre serveur ou un fichier du même serveur
 - ▶ une autre partie du même document
- Repérage d'un hyperlien (texte ou graphique) dans un navigateur :
 - ▶ le curseur change de forme en passant sur le lien
 - ▶ le texte est en couleurs et souligné (la couleur change si le lien a déjà été visité)
 - ▶ l'image est encadrée

(mais tout cela est paramétrable avec CSS).

- Pour faire un lien, on utilise l'attribut `href` de la balise `<a>` dont le contenu formera le lien :

```
<a href="http://www.cnrs.fr/">  
    
</a>
```

```
<a href="bio/indexbioinfo.html">Bioinformatique</a>
```

- Les **ancres** servent à atteindre un endroit précis dans le document.
 - ▶ On commence par définir les ancres, soit sur une balise existant déjà grâce à l'attribut `id`, soit avec un `` :

```
<h3 id="tutorials">Tutorials</h3>  
<a id="tutorials" />
```

- ▶ Ensuite, on fait le lien avec cette ancre.

```
<a href="#tutorials">tutorials</a>  
<a href="http://www.w3.org/#tutorials">tutorials</a>
```

Plan du cours

- 1 Internet et le Web
- 2 Le langage XHTML
- 3 XHTML et HTML**
- 4 Références
- 5 Application

- HyperText Markup Language
- Dernière version : HTML 4.01
- Décrit par un formalisme très compliqué (SGML)
- En pratique : pages Web ne respectent pas ce formalisme \implies
navigateurs ne respectent pas ce formalisme \implies soupe de balises !

- Les balises sans contenus `
`, s'écrivent `
` en HTML.
- Certaines balises peuvent ne pas être refermées (` un deux `).
- Les valeurs des attributs peuvent ne pas être entre guillemets (``).
- Les noms des balises et des attributs sont insensibles à la casse (`<HTML laNg=fr>`).
- ```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01//EN"
"http://www.w3.org/TR/html4/strict.dtd">
```
- Pas de `xmlns`, pas de `xml:lang`.
- Et quelques autres petites subtilités...

## Avantages de HTML 4.01

- Meilleur support par les navigateurs (Internet Explorer 6/7 comprennent mal XHTML).
- Moins de contraintes. . .

## Avantages de XHTML 1.0

- Plus de contraintes. . . (donc plus simple!).
- Syntaxe claire, sans ambiguïté.
- Familiarité avec XML, utile dans d'autres contextes.
- Facilité d'utilisation dans des contextes XML (p.ex. XSLT).

- Théoriquement, `<p />` et `<p></p>` sont synonymes en XML, donc en XHTML. En pratique, on utilisera la notation `<balise />` uniquement pour les balises n'ayant jamais de contenu (p.ex. `<br />`, `<hr />` ...).
- Théoriquement, `<br />` et `<br/>` sont synonymes. En pratique, on utilisera toujours `<br />`.
- Théoriquement, un document XHTML peut commencer par une ligne `<?xml version="1.0"encoding="utf-8"?>`. En pratique, on l'omettra.
- cf. <http://www.w3.org/TR/xhtml1/#guidelines>
- On pourra utiliser <http://qa-dev.w3.org/~bjoern/appendix-c/validator/> pour vérifier le respect de ces quelques règles.

# Plan du cours

- 1 Internet et le Web
- 2 Le langage XHTML
- 3 XHTML et HTML
- 4 Références**
- 5 Application

- Spécification de XHTML 1.0  
<http://www.w3.org/TR/xhtml1/>
- Spécification de XML 1.0  
<http://www.w3.org/TR/REC-xml/>
- Spécification de HTML 4.01  
<http://www.w3.org/TR/REC-html40/>
- *HTML et XHTML : La Référence*, O'Reilly
- <http://www.tuteurs.ens.fr/internet/>

# Plan du cours

- 1 Internet et le Web
- 2 Le langage XHTML
- 3 XHTML et HTML
- 4 Références
- 5 Application**

- Allumez l'ordinateur et connectez-vous.
- Lancez le navigateur de votre choix.
- Allez sur le site du LRI : <http://www.lri.fr/>
  - ▶ regardez le source de la page principale
- Allez sur votre site préféré (ou <http://www.w3.org/MarkUp/>)
  - ▶ regardez le source
  - ▶ comparez
- Remarques ?

- Ouvrez un éditeur de texte
- Créez la page donnée en modèle
- Enregistrez la sous le nom *essai.html*
- Validez à l'aide de <http://validator.w3.org/>

Rajoutez dans votre fichier *essai.html*

- dans la rubrique *Exemples* :
  - ▶ une liste non-numérotée de 4 pays
  - ▶ pour chaque pays, une liste numérotée de 2 villes
  - ▶ une liste numérotée de 2 fruits
  - ▶ pour chaque fruit, une liste non-numérotée de 2 prix possibles au kilo
- une rubrique *Annexe*, contenant un lexique de 5 balises définies depuis le début du cours.
- Validez !

Rajoutez dans la rubrique *Exemples* de votre fichier *essai.html* une sous-rubrique contenant la table de multiplication de 1 à 4 (avec légende). Validez !

- Insérez une image à la fin de *essai.html* pointant vers <http://pierre.senellart.com/ifips.jpg>. Ne pas oublier un texte alternatif approprié.
- Enregistrez ce fichier image dans le même répertoire que *essai.html* et insérez à nouveau une image pointant vers ce fichier local.
- Transformez cette image en lien pointant vers <http://www.ifips.u-psud.fr/>.
- Ajoutez une ancre à la rubrique Annexe, et un lien vers cette ancre au tout début du document.