

XHTML : introduction

1. Présentation et historique

Le HTML à l'origine est un langage simple destiné à la présentation des documents hypermédias. Toutefois au fil des ans, il s'est avéré que le HTML ne correspond pas à toutes les attentes des utilisateurs en terme de modularité, de maintenance et de présentation.

Le consortium w3 <http://www.w3.org> a décidé de refondre le langage HTML en basant sur les technologies XML. D'où la création du XHTML (eXtended HyperText Markup Language) en 2002

XHTML est un pont entre deux technologies, il est compatible avec les anciennes versions du HTML et accompagne le passage aux technologies modulaires et extensibles du XML

Les objectifs et avantages du XHTML sont :

- Un code simplifié, plus léger, moins bavard. Il est plus facile à manipuler et à mettre à jour.
- Une séparation entre la contenu et sa présentation qui permet de présenter un même document sous de nombreuses formes différentes sans le retoucher (version écran, print, téléphone portable, ...)
- Un standard qui peut être soumis à *validation* et permet de se décharger des spécificités des différents navigateurs
- Il facilite la mise en place d'un web *sémantique* qui a pour objectif de rendre accessible les informations d'une manière standard afin de faciliter l'indexation des informations aussi bien aux humains qu'aux robots.
- Une préoccupation accrue par rapports aux jeux de caractères et à l'internationalisation.

Les inconvénients du XHTML :

- Un langage plus rigoureux qui ne tolère pas les bidouillages.
- Le langage continue à évoluer et n'est pas figé.

Ce cours vise à apprendre les particularités du XHTML, il requiert une certaine connaissance préalable du HTML.

2. Introduction au langage

2.1. Encore du HTML...

Le XHTML va reprendre l'essentiel de la syntaxe du HTML en rajoutant les contraintes imposés par le XML.

Ces contraintes sont de plusieurs ordres :

- Déclaration obligatoire du document (DOM)
- Bien former ces balises

2.2. Déclaration d'un document

En XML, la déclaration du document est obligatoire. La déclaration vise à donner la référence syntaxique des balises utilisées dans le document. On parle de DTD (Définition du Type de Documents). Elles définissent sans ambiguïté toutes les balises et leurs syntaxes respectives.

Certaines DTD sont définis normalisées par le w3 . On choisira la DTD qui convient le mieux au niveau d'exigence et de comptabilité requis.

DTD XHTML stricte:

```
<!DOCTYPE html
PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
```

DTD XHTML de transition:

```
<!DOCTYPE html
PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
```

DTD XHTML avec support de jeu de cadres:

```
<!DOCTYPE html
PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-frameset.dtd">
```

Cela permettra de vérifier si le document respecte la syntaxe grammaticale donnée (le document est alors bien *formé*). Il est possible de vérifier si une page est bien formé grâce à l'adresse :

<http://validator.w3.org>

Il est aussi nécessaire d'associer la page à un espace de nom xmlns (xml namespace)

```
<html xmlns="http://www.w3.org/1999/xhtml"> ...  
</html>
```

2.3. Le document minimal XHTML

Il est obligatoire de placer un <head> et un <title> dans chaque page XHTML. Ce qui donne un document XHTML minimal du type :

```
<?xml version="1.0" encoding="UTF-8"?>  
<!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>Titre du document</title>  
  </head>  
  <body>  
    <!-- Contenu du document -->  
  </body>  
</html>
```

Le xml permet de spécifier le jeu de caractère de l'encode du document . Donc si vous êtes en unicode (UTF8) ou iso-latin, vous pouvez placer les caractères é, à, ç, ... directement sans recourir aux traditionnels é ; ,...

2.4. Règles de formatages des balises

XHTML est strict sur les points suivants et ne tolère plus les imprécisions du HTML

2.4.1. Tous les balises sont en minuscules

```
<a href="max.html">correct</a>  
<A HREF="max.html">incorrect</A>
```

2.4.2. Tous les balises sont fermés

```
<li>Élément liste correct</li>  
<li>Élément liste incorrect
```

Pour les éléments simples, on utilise des tags auto-fermants

```
  
un retour chariot: <br />
```

2.4.3. Emboîter les balises : *les poupées russes*

```
<a><b><c>correct</c></b></a>  
<a><b><c>incorrect</a></b></c>
```

2.4.4. Les attributs sont toujours avec des guillemets

```
  

```

2.4.5. La minimisation des attributs est interdite

```
<option sected="selected">correct</option>  
<option selected>incorrect</option>
```

2.4.6. Les éléments *script* et *style* sont déclarés comme possédant un contenu de données textuelles analysées (PCDATA : Parsed Character DATA)

```
<script language="javascript" type="text/javascript">  ![CDATA[  
    prix = compte <= 1000 ? 500 : 450;  
  ]]> </script>
```

2.4.7. L'attribut *name* doit être progressivement abandonné au bénéfice de l'identificateur *id*

```
<form id="formulaire_correct">...</form>  
<form name="formulaire_incorrect">...</form>
```

2.4.8. Automatiquement, les agents utilisateurs supprimeront les espaces de début et de fin des valeurs d'attributs et si plusieurs espaces se suivent, ils seront remplacés par un unique espace

```
attribut=" valeur de l'attribut "  
<!-- après analyse -->  
attribut="valeur de l'attribut"
```

2.4.9. XHTML interdit l'inclusion de certains éléments dans d'autres

- *a* ne peut contenir d'autres éléments *a*
- *pre* ne peut contenir les éléments *big*, *img*, *object*, *small*, *sub* ou *sup*
- *button* ne peut contenir les éléments *button*, *form*, *fieldset*, *iframe*, *input*, *isindex*, *label*, *select* ou *textarea*
- *label* ne peut contenir d'autres éléments *label*
- *form* ne peut contenir d'autre éléments *form*

2.5. Encore quelques règles pour être compatible....

Voici maintenant quelques recommandations qui seront bientôt obligatoires. Cette liste n'est pas exhaustive, se reporter à la documentation du w3 pour plus de détails.

2.5.1. L'utilisation de l'esperluette dans une valeur d'attribut doit se faire par l'intermédiaire de sa référence d'entité caractère, &

```
<a href="http://www.site.com/script.html?d=1&v=val">Lien correct</a>  
<a href="http://www.site.com/script.html?d=1&v=val">Lien incorrect</a>
```

2.5.2. Les retours à la ligne dans les valeurs d'attributs sont à proscrire

```
alt="Une image de mon album"  
<!-- construction correcte -->  
alt="Une image  
de mon album"  
<!-- construction incorrecte -->
```

2.6. Changement au niveau des balises

Pour éviter la concurrence d'informations avec les feuilles de style, certaines balises ou attributs de balises ont disparu. Voir la document officielle pour les détails précis. Pour mémoire, toutes les informations concernant la présentation graphique ne sont plus disponibles.

Résumé des principaux changements :

- `body` : suppression de *background*, *bgcolor*, *text*, *link*, *alink*, *vlink*
- `img`: *alt* obligatoire et suppression de *align*, *border*, *hspace*, *vspace*
- `table` : suppression de *align*, *bgcolor*
- `td`, `th` : suppression de *bgcolor*, *height*, *nowrap*, *width*

3. XHTML, langage modulaire

Nous avons utilisé pour l'instant des DTD prédéfinis. XHTML a aussi pour objectif d'être supporté sur nombreux supports (mobile, console, ...). Il est possible d'écrire ses propres DTD grâce aux 28 modules de bases pour définir précisément les balises que l'on envisage d'employer.

4. Passer du HTML au XHTML

Passer du HTML au XHTML est relativement facile pour les sites en dynamique, cela peut être plus fastidieux pour des sites en statiques . Voici une petite liste des opérations à effectuer :

1. Ajouter les déclarations de document (doctype....)
2. Convertir tous les tags en minuscules
3. Fermer toutes les tags
4. Supprimer tous les attributs superflus
5. Ajouter les attributs obligatoires comme le alt des images

Il est aussi possible d'utiliser des outils comme convertir automatiquement vos pages. Le logiciel TIDY du w3 peut notamment vous y aider : <http://www.w3.org/People/Raggett/tidy/>

5. XHTML 2.0

La version 2 est encore à l'état de brouillon (working draft) et non de recommandation.

Voici quelques unes des nouveautés à venir <http://www.w3.org/TR/xhtml2/>

- L'attribut href sur d'autres balises que a
- La balise object remplace img, applet, embed
- Apparition de nl (navigation list)
- Apparition de l'attribut edit pour indiquer le statut d'une info

6. Ressources

Quelques liens pour en savoir plus :

http://openweb.apinc.org	XHTML et CSS en francais
http://fr.selfhtml.org	Cours assez complet sur XHTML
http://www.latruiste.com	Autre cours sur XHTML
http://www.w3schools.com	Introduction en anglais
http://www.w3.org	Site de référence et normatif