

Création et Manipulation de documents

(Hélène Renard / Sylvain Schmitz)

Travaux Dirigés – Séance n°4

1 Objectifs du TD

Apprendre les bases du langage XHTML.

2 Introduction à XHTML

Le langage XHTML a remplacé en 1999 un HTML vieillissant en tant que recommandation officielle du W3C, qui est l'organisme chargé de développer les standards pour le web. La version initiale, XHTML 1.0, est quasiment identique à HTML 4.01, elle se veut simplement plus stricte et propre en adoptant la syntaxe des balises XML. Le langage XHTML a déjà été mis à jour avec une version 1.1 qui cherche à décourager les usages *présentationnels* des balises, qui sont du ressort d'une feuille de style, pour ne garder que les aspects *structurels* du document. La version 2.0 en cours de construction élimine définitivement les balises présentationnelles, et intègre les recommandations les plus récentes en matière de formulaires web XFORMS, les évènements XMLevents, les méta données RDF...

Un document XHTML décrit une page destinée à être affichée par un navigateur web. On décrit la page à l'aide d'éléments XHTML. Ces éléments expriment la façon de structurer la page.

2.1 Balisage XML

Normalement, un élément encadre son contenu à l'aide de balises. Par exemple, pour mettre en valeur un titre d'ouvrage *Critique de la raison pure*, on utilise l'élément `em` (pour *emphasize*), ce qui s'écrit en XHTML : `Critique de la raison pure`. Autrement dit, un élément `t` commence par la balise ouvrante `<t>` et se termine par la balise fermante `</t>`.

Suivant les règles d'un balisage XML, dans un document, tout élément ouvert doit être fermé. De plus, les éléments doivent être correctement imbriqués : `<h2>Résumé de la Critique de la raison pure</h2>` n'est pas correct, on doit écrire `<h2>Résumé de la Critique de la raison pure</h2>`. Toutefois, quelques éléments sont dits vides. Ils n'encadrent rien et n'ont qu'une balise (qui est donc ouvrante et fermante). C'est le cas de l'élément `br` (qui indique un passage à la ligne forcé). Un tel élément s'écrit `
`. Notons enfin que le nom des éléments dans les balises doit être en minuscule (autrement dit, on n'écrira pas `bonjour`).

La balise ouvrant un élément peut (et parfois doit) prendre un ou plusieurs attributs. Ceux-ci ont alors la forme `attribut="valeur"` (et on notera les lettres minuscules ainsi que les guillemets autour de la valeur). Par exemple, pour inclure une image `image.jpg` dans un document, on utilisera ``.

2.2 Document XHTML 1.1 minimal

Un document XHTML peut employer l'extension `.html`, mais on préfère `.xhtml`. De même, si un document XHTML est souvent servi avec le type mime `text/html` pour des raisons de compatibilité avec un certain navigateur pas très à jour, le type préféré est `application/xhtml+xml`.

Un document XHTML 1.1 minimal aura la forme :

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.1//EN"
    "http://www.w3.org/TR/xhtml1/DTD/xhtml11.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="fr">
  <head>
    <title>Document XHTML minimal</title>
  </head>
  <body>
    ...
  </body>
</html>
```

En-tête XML Un document XHTML est *a fortiori* un document XML, et il est recommandé de l'indiquer en commençant par la ligne :

```
<?xml version="1.0" encoding="UTF-8"?>
```

Cette indication devient obligatoire si jamais le codage de caractères employé devait différer de l'UTF-8 :

```
<?xml version="1.0" encoding="ISO-8859-1"?>
```

Déclaration de la DTD Un document XHTML doit commencer par la déclaration d'une DTD (*Document Type Definition*) suivi du tag `html` qui englobe tout le reste du document XHTML. La déclaration de la DTD sera toujours :

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.1//EN"
    "http://www.w3.org/TR/XHTML11/DTD/XHTML11.dtd">
```

Remarquons que cette déclaration n'est pas à proprement parler un élément XHTML et n'a pas de balise fermante, ni ne finit par `</>`. Cette DTD, qui s'appelle `-//W3C//DTD XHTML 1.1//EN` (produite par le W3C, et en anglais), est accessible à l'adresse <http://www.w3.org/TR/xhtml1/DTD/xhtml11.dtd>. Une DTD est un fichier qui décrit un ensemble de balises et comment celles-ci interagissent.

Élément racine `html` Enfin, l'élément `html` encadre la totalité du document. En tant qu'élément racine, on lui adjoint de plus la déclaration de l'espace de nom XHTML ; nous verrons par la suite que l'on peut exploiter ces espaces de nom pour mêler plusieurs dialectes XML dans un seul document.

```
<html xmlns="http://www.w3.org/1999/XHTML" xml:lang="fr">
```

L'élément `html` doit toujours contenir un élément `head` suivi par un élément `body`.

Élément `head` L'élément `head` rassemble les méta informations du document (balises `title`, `meta` et `link`), ainsi que les déclarations embarquées de présentation (balise `style`) et de scripts (balise `script`). L'élément `title` en particulier est obligatoire.

```
<head>
  <title>Document XHTML minimal</title>
</head>
```

Élément body L'élément `body` est la racine du contenu à proprement parler du document.

```
<body>
... corps du document ...
</body>
```

Exercice n°1 :

1. Écrire une page XHTML de titre « Première page » contenant simplement le texte « Bienvenue dans le *Meilleur des mondes!* » (où « Meilleur des mondes » est un titre à mettre en valeur).
2. Il faut maintenant s'assurer que votre première page est bien un document XHTML valide. Le W3C fournit un validateur que l'on trouve à l'adresse <http://validator.w3.org/>. Vérifiez si votre page est valide. Si ce n'est pas le cas, rendez-la valide.

3 Des balises par millions

Vous trouverez dans cette section un certain nombre d'éléments XHTML répartis en plusieurs modules. Dans chaque module, vous aurez accès aux principaux éléments dont vous aurez besoin pour construire des pages XHTML valides.

Commençons par la présentation du module de texte :

Module de texte		
Éléments	Nom	Description
<code>h1 . . . h6</code>	<i>heading</i>	Définit un entête (titre pour une section, paragraphe, . . .). L'élément <code>h1</code> représente des titres de plus grande importance que <code>h2</code> , <code>h2</code> que <code>h3</code> , etc.
<code>blockquote</code>	<i>blockquote</i>	La balise <code><blockquote></code> est un bloc de citation au sein d'une page web XHTML. On utilise cette balise dans le cas de citation longue et/ou sur plusieurs lignes.
<code>br</code>	<i>forced line break</i>	La balise <code>
</code> est un élément vide qui force un retour à la ligne.
<code>div</code>	<i>div</i>	La balise <code><div></code> est souvent employée pour regrouper d'autres balises dans des zones par thème. Par exemple, nous pouvons regrouper une balise <code><h3></h3></code> et <code><p></p></code> dans un <code><div></code> de manière un former un bloc.
<code>kbd</code>	<i>keyboard</i>	Délimite un texte prévu pour être saisi au clavier.
<code>p</code>	<i>paragraph</i>	Délimite un paragraphe de texte.
<code>pre</code>	<i>preformatted</i>	Affiche un texte en préservant les espaces et les sauts de ligne.
<code>q</code>	<i>quote</i>	La balise <code><q></code> représente une citation courte dans une page web XHTML. L'attribut <code>cite</code> contient la référence métrique de la citation. L'attribut <code>lang</code> définit la langue dans laquelle la citation est écrite.
<code>strong</code>	<i>strong</i>	Insiste fortement sur le texte contenu, qui peut être par exemple une consigne de sécurité, un rappel de règlement, etc. Son affichage peut-être fait en gras, en majuscules, en rouge. . .
<code>em</code>	<i>emphasize</i>	Met en valeur le texte contenu, qui peut être un titre d'ouvrage ou de journal, un titre honorifique, une portion de texte surlignée, etc. Son affichage est généralement en italique ou souligné.
<code>abbr</code>	<i>abbreviation</i>	La balise <code><abbr></code> est une abréviation. L'attribut <code>title</code> permet de donner la signification de l'abréviation et l'attribut <code>lang</code> la langue de l'abréviation.
<code>code</code>	<i>computer code</i>	La balise <code><code></code> définit du code informatique.
<code>cite</code>	<i>cite</i>	La balise <code><cite></code> représente la source d'une citation.

Module de texte, suite		
Éléments	Nom	Description
dfn	<i>definition</i>	La balise <dfn> représente une définition de terme. En fait, elle entoure le terme à définir et le reste du texte du bloc représente la définition.
samp	<i>sample</i>	La balise <samp> définit une sortie de code informatique.
span	<i>span</i>	Rassemble quelques mots. Par exemple, Jean Sébastien Bach.
var	<i>variable</i>	La balise <var> définit une variable de code informatique.

Le module de texte perdra quelques éléments dans la version XHTML 2.0. Ils se verront réattribués pour certains au module de structure.

Le module des liens hypertextes est le plus petit module du XHTML 1.1. Il contient un seul élément (l'élément **a** qui peut avoir différents attributs.

Module des liens hypertextes		
Éléments	Nom	Description
a	<i>anchor</i>	Déclare un lien hypertexte. Ce tag contient en général l'attribut href pour déclarer la page vers laquelle pointe le lien. La balise <a> peut avoir différents attributs : title (info bulle), href (page de destination), name (pour nommer une ancre), id ...

Comme son nom le laisse sous-entendre, le module des listes fournit les éléments utilisables dans les listes (à puces, numérotées, etc.).

Module des listes		
Éléments	Nom	Description
dl	<i>definition list</i>	La balise <dl> est une liste de définition. Elle s'utilise avec les balises <dd> et <dt>.
dt	<i>definition term</i>	La balise <dt> est un terme à définir.
dd	<i>definition description</i>	La balise <dd> est une description de définition
ol	<i>ordered list</i>	Déclare une énumération. Chaque élément de l'énumération est déclaré avec l'élément li .
ul	<i>unordered list</i>	Déclare une liste. Chaque élément de la liste est déclaré avec l'élément li .
li	<i>list item</i>	La balise est un élément d'une liste. Elle peut faire partie d'une liste non-ordonnée () ou ordonnée ().

Module des tables		
Éléments	Nom	Description
caption	<i>caption</i>	Contient la légende d'un tableau.
table	<i>table</i>	Déclare un tableau. Ce tableau peut contenir optionnellement une légende caption . On définit ensuite les lignes avec l'élément tr , et dans chaque ligne, les cases avec l'élément td .
td	<i>table data cell</i>	La balise <td> est une cellule d'un tableau de données. Elle contient toutes sortes d'autres balises et éléments orphelins. Les attributs colspan et rowspan permettent de définir des cellules qui débordent sur les colonnes ou lignes suivantes.
th	<i>table header cell</i>	La balise <th> est une cellule d'en-tête d'un tableau de données.
tr	<i>table row</i>	La balise <tr> est une ligne d'un tableau de données. Elle permet de regrouper les cellules d'une même ligne.

Module des images		
Éléments	Nom	Description
<code>img</code>	<i>image</i>	La balise <code></code> doit obligatoirement contenir un attribut <code>src</code> (source) et <code>alt</code> (texte alternatif). Les attributs optionnels sont <code>height</code> , <code>longdesc</code> et <code>width</code> . De plus, <code></code> ne peut pas être imbriquée dans une balise <code><pre></code> .

Le détail du module de présentation ne sera pas fait ici car c'est un module qui est amené à disparaître dans les versions futures d'XHTML. Pour les exercices, il ne vous manque que l'élément vide `<hr />` (*horizontal rule*) qui affiche une ligne horizontale.

D'autre part, certains caractères ne sont pas accessibles directement dans un document XHTML car ils ont une signification particulière (par exemple le caractère `<`). Pour afficher ces caractères, on utilise des entités :

Caractère	Entités
espace	<code>&nbsp;</code> ;
<code><</code> (less than)	<code>&lt;</code> ;
<code>></code> (greater than)	<code>&gt;</code> ;
<code>&</code> (ampersand)	<code>&amp;</code> ;
"	<code>&quot;</code> ;

Précisons également que les commentaires commencent par `<!--` et finissent par `-->`.

Exercice n°2 : La page correspondant au fichier `~hrenard/tp_errone.xhtml` n'est pas valide. Copiez ce fichier dans votre répertoire `CMDocs/xhtml` et corrigez les erreurs afin de rendre la page valide.

Quelques remarques sur l'exercice précédent : vous avez peut-être remarqué que la page erronée s'affiche « presque » de la même façon que la page valide (essayez de voir le pourquoi du « presque »). Toutefois, seule la validité de vos pages vous assure d'une part, que l'affichage est « réellement » comme vous le voulez et surtout qu'il sera conforme à vos attentes sur tous les navigateurs.

Exercice n°3 : Essayez de reproduire la page web représentée par l'image suivante le plus fidèlement possible : `~hrenard/pageXHTML.png`. Faites bien attention à tous les détails, tels que le titre de la page, les espaces insécables etc. Vous devez ensuite valider le résultat.

Exercice n°4 : Vous allez reprendre votre « Première page » et lui ajouter quelques citations :

- Tout est bien, tout va bien, tout va le mieux qu'il soit possible.
- Les malheurs particuliers font le bien général ; de sorte que plus il y a de malheurs particuliers et plus tout est bien.

en sachant que la source de ces citations est : « Candide ou l'Optimiste, Voltaire ». Une fois ces citations insérées correctement, illustrez-les avec une image appropriée (par exemple, une image du livre)

Exercice n°5 : Rajoutez un lien depuis votre « Première page » vers la page de l'exercice n°3. Vérifiez que votre page est valide, puis ajoutez le logo de validité.

Page à reproduire (exo 3 du TD n°4, CMDocs) - Mozilla Firefox <2>

Echier Edition Affichage Historique Marque-pages Outils Aide

file:///home/frenard/Polytech/2006_2007/CIP1/CMDocs/TD4/page_a_re

Page de démonstration des balises XHTML

Liste

Il est possible d'utiliser des listes d'éléments dits non-ordonnés :

- en voici un 1er élément
- en voici un autre
 - qui nous montre qu'on peut imbriquer les listes
 - et on remarque qu'alors les puces changent
- et on peut même en mettre un troisième

Mais il est aussi possible d'avoir des listes numérotées :

1. dont voici le 1ère élément
2. et dont voici le 2ème
3. en notant que la numérotation est automatique

Tableau

Un titre d'importance moindre

1ère ligne, 1ère colonne	1ère ligne, 2nde colonne
2nde ligne, 1ère colonne	2nde ligne, 2nde colonne

D'autres choses

Notons que certains caractères ne peuvent être directement inclus dans une page web. C'est le cas de < ou >. Toutefois, inclus directement, ils ne sont pas rejetés par le validateur qui donne simplement un avertissement.



Terminé

