

Web (Persistence)



Andrea G. B. Tettamanzi

Université de Nice Sophia Antipolis

Département Informatique

andrea.tettamanzi@unice.fr

CM - Séance 2

HTML

Plan

- Introduction à HTML
- Balises, entités, attributs
- DTD
- XHTML
- Quelques balises importantes
- Tables
- Formulaires
- Images cliquables

Langages de balisage

- Des langages spécialisés dans l'enrichissement (mark-up) d'information textuelle.
- Balise = unité syntaxique délimitant une séquence de caractères à l'intérieur d'un flux de caractères (par exemple un fichier texte).
- L'inclusion de balises permet de spécifier, à la fois
 - la structure (logique, graphique, ...) du document
 - son contenu.
- Les balises sont faciles à analyser par un programme, ce qui permet un traitement automatisé du contenu.
- Historiquement, les langages à base de balises servent surtout à structurer ou formater des documents.

SGML → HTML → XML

- Historiquement, HTML est une application dérivée de SGML
- SGML = Standard Generalized Markup Language
- Inventé par Charles Goldfarb pour rationaliser le système documentaire d'IBM ; devient un standard ISO en 1986
- Principe : séparation complète entre
 - structure logique d'un document (titres, chapitres, paragraphes, illustrations,...), qui est identifiée par des **balises** insérées dans le document lui-même
 - sa mise en page, qui dépend du support de présentation (livre, journal, écran, graphique même) et qui est définie en dehors du document dans une ou plusieurs **feuilles de style**
- XML est une extension de SGML pour structurer des données

Balises

- Une balise est définie par un identifiant/étiquette, par exemple **h1**
- Balise d'ouverture : **<h1>**
- Balise de fermeture : **</h1>**
- Un couple de balises (d'ouverture et de fermeture) délimite un bloc de texte
- Balises vides : **
**
- Les balises sont imbriquées hiérarchiquement

```
<body>
  <h1>Titre</h1>
  <p>Paragraphe</p>
</body>
```

Balises « de bloc » et « en-ligne »

- Distinction basée sur modèle de contenu et formatage
- Modèle de contenu :
 - les éléments de bloc peuvent contenir à la fois données, éléments de bloc et éléments en-ligne ;
 - les éléments en-ligne ne peuvent contenir que des éléments en-ligne et des données.
- Formatage :
 - En général, les éléments de bloc commencent sur une nouvelle ligne, et non les éléments en-ligne. »

Entités

- Les entités sont les éléments de base qui constituent le texte
- Les caractères sont des entités
- Les symboles sont des entités
- Entités nommées :
 - Délimitées par les caractères & et ;
 - Par exemple : & , é , etc.

Attributs

- Les balises peuvent avoir des attributs avec des valeurs
- Les valeurs sont affectés aux attributs dans la balise d'ouverture
- Par exemple :

```

```

Propriétés des balises HTML

- Propriétés générales (mais pas systématiques)
 - id, class (document-wide identifiers)
 - lang (language information), dir (text direction)
 - title (element title)
 - style (inline style information)
- onclick, ondblclick, onmousedown, onmouseup, onmouseover, onmousemove, onmouseout, onkeypress, onkeydown, onkeyup (intrinsic events)
- Définition dans la DTD de HTML.

Déclaration de DTD HTML 4.01

- HTML 4.01 strict DTD : tous les éléments et attributs déclarés et non dépréciés (deprecated) et qui ne sont pas lié au frameset.
- `<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01//EN" "http://www.w3.org/TR/html4/strict.dtd">`
- HTML 4.01 transitional DTD : strict DTD + deprecated (présentation visuelle).
- `<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN" "http://www.w3.org/TR/html4/loose.dtd">`
- HTML 4.01 Frameset DTD : transitional DTD + frame.
- `<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Frameset//EN" "http://www.w3.org/TR/html4/frameset.dtd">`

Fichier HTML minimum (HTML 4.01)

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01//EN"  
    "http://www.w3.org/TR/html4/strict.dtd">  
  
<HTML>  
    <HEAD>  
        <TITLE>  
        </TITLE>  
    </HEAD>  
  
    <BODY>  
    </BODY>  
  
</HTML>
```

XHTML

- Version « xml » de HTML
- Pour écrire un document xhtml (valide) à partir d'un document html 4 valide :
 - bien fermer les balises (ou tags)
 - bien fermer les balises dans le bon ordre
 - balises et attributs en minuscule (sensible à la casse).
 - fermer les éléments vides (e.g. `
`)
 - valeurs des attributs entre guillemets et chaque attribut a une valeur
 - (la minimisation n'est pas autorisée : `<input checked="checked" />`)
 - généralement l'attribut name est à remplacer par l'attribut id

Fichier XHTML minimum

```
<?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>...</title>
</head>

<body>

</body>

</html>
```

Balises HTML / XHTML

- Balises structurelles
- Balises contenant du texte (phrase)
- Balise hypertexte
- Balises de titres
- Balises de paragraphes et citations
- Balises de listes
- Balises de table (tableaux)
- Balises de formulaire
- les images cliquables, objets, applet
- Balises de frame
- Balises meta (entêtes)
- Balises de mise en forme (déconseillées !)

Balises Structurelles

- html : définition du document
 - un head, un body
- head : partie comprenant les entêtes et meta-données du document
 - Un `<title> ... </title>` obligatoire
 - Les balises « meta », « script », « style », etc. optionnelles
- title : titre du document
 - Titre de la page (en haut du navigateur)
 - Contient du texte
- body : corps du document
 - Partie centrale du document

Balises de bloc

- Regroupement générique : div | fieldset
- Paragraphes : p | pre
- Titres : h1 | h2 | h3 | h4 | h5 | h6
- Listes : ol | ul | dl
- Tableau : table
- Formulaire : form
- Divers : noscript | blockquote | hr | address

Balises en-ligne

- texte (PCDATA)
- phrase : em | strong | dfn | code | samp | kbd | var | cite | abbr | acronym
- special : a | img | object | br | script | map | q | sub | sup | span | bdo
- formulaire : input | select | textarea | label | button
- fontstyle : tt | i | b | big | small (déconseillées)

Balise <a> (ancre)

- Référencer une page par son URI
 - `....`
- référencer à l'intérieur d'une page
 - référence : `href= "URI#nom de l'ancre"` (lien local si URI = "")
 - "nom de l'ancre" = la valeur d'un attribut `id` d'une balise (n'importe laquelle)
 - `http://www.i3s.unice.fr/~tettaman/Classes/IntroWeb/html/exo/index.html#exo2`
- Contient des éléments html « inline » excepté `<a>`

Tables

- `<table>...</table>`
- Table row : `<tr>...</tr>`
- Table data : `<td>...</td>`
- (caption?, (col*|colgroup*), thead?, tfoot?, tbody+)
 - Le tableau peut avoir une légende `<caption>`
 - On peut regrouper les colonnes (`<col>` et `<colgroup>`)
 - Il peut y avoir une entête (répété pour impression) : `<thead>`
 - Il peut y avoir un pied de tableau (répété pour impression) : `<tfoot>`
 - Il doit y avoir au moins un corps de tableau `<tbody>`

Colspan et Rowspan

- L'attribut `colspan= n` permet de définir une cellule qui mesure n colonnes de large
- L'attribut `rowspan= n` permet de définir une cellule qui mesure n ligne de haut

Formulaires

```
<FORM action="http://somesite.com/prog/adduser" method="post">
  <P>
    <LABEL for="firstname">First name: </LABEL>
      <INPUT type="text" id="firstname"><BR>
    <LABEL for="lastname">Last name: </LABEL>
      <INPUT type="text" id="lastname"><BR>
    <LABEL for="email">email: </LABEL>
      <INPUT type="text" id="email"><BR>
    <INPUT type="radio" name="sex" value="Male"> Male<BR>
    <INPUT type="radio" name="sex" value="Female"> Female<BR>
    <INPUT type="submit" value="Send"> <INPUT type="reset">
  </P>
</FORM>
```

Images cliquables

- des images contenant des zones servant de liens
- Pour la mise en place d'une image cliquable, on doit composer 2 balises :
 - une balise Image `` pour charger l'image
 - une balise `<map>` pour définir les zones et les liens concernés
 - la référence à la balise `<map>` est faite par l'attribut « usemap »
- Deux types d'images cliquables
- Côté client (le plus fréquent et préférable) : traitement par le navigateur
- Côté serveur : traitement sur le serveur

Pour approfondir...

- Specification officielle de HTML 4.01
 - Disponible sur le site du W3C
 - Lien sur la page du cours
- Service de validation syntaxique
 - <http://validator.w3.org/>

Merci de votre attention

