Support de cours HTML/XML

Pierre POMPIDOR

Table des matières

Première partie
HTML

1 HTML : présentation générale

HTML est un langage à balises pouvant comporter des attributs :
<balise attribut1=... attribut2=... ...> ... </balise>

- Généralement une balise de fin </balise> correspond à une balise de début <balise>
- Les balises peuvent être indifféremment écrites en miniscules ou en majuscules.
- Les valeurs des attributs peuvent être circonscrites par des guillemets.

2 Liste des balises courantes

2.1 Début et fin du document

```
Balise de début du document : <html>
Balise de fin du document   : </html>
```

2.2 Entête d'un document HTML

```
Balise de début d'en-tête                        : <head>
Titre (banière, bookmarks ..)                    : <title> ... </title>
Pour insérer des directives dans la réponse HTTP : <meta>
   attributs http-equiv et content               : <meta http-equiv=... content=...>
   (meta est une balise unique)                    <meta http-equiv=Refresh content="5; http://...">
Balise de fin                                    : </head>
```

2.3 Corps d'un document HTML

```
Balise de début                                 : <body>
   attribut fixant la couleur d'arrière plan    : bgcolor=...
   attribut fixant l'image d'arrière plan       : background=...
   attribut fixant la couleur du texte          : text=...
   attribut fixant la couleur des liens         : link=...
   attribut fixant la couleur des liens visités : vlink=...
Balise de fin                                   : </body>
```

2.4 Présentation des chaînes de caractères

2.4.1 Tailles prédéfinies

```
De la plus grosse : <h1> ... </h1>
...
à la plus petite  : <h6> ... </h6>

Par rapport à une échelle de 1 à 7 et d'une taille moyenne de 3 : <font size=+n> ... </font>
```

2.4.2 Styles

```
En gras           : <b> ... </b>
En italique       : <i> ... </i>
En souligné       : <u> ... </u>
Machine à écrire  : <tt> ... </tt>
Pré-formaté       : <pre> ... </pre>
```

2.4.3 Couleurs

```
Prédéfinie : <font color=...> ... </font>
(Aqua, Black, Blue, Gray, Green, Lime, Maroon, Navy, Olive, Purple, Red, Silver, Teal, White, Yellow)
RGB        : <font color=#rrggbb> ... </font>
```

2.4.4 Présentation de code informatique

```
Pour des lignes de code     : <code> ... </code>
Pour des entrées au clavier : <kbd> ... </kbd>
Pour des variables          : <var> ... </var>
```

2.5 Figures de présentations

2.5.1 Saut de ligne

```
Balise unique : <br>
```

2.5.2 Paragraphes

```
<p> ... </p>
```

2.5.3 Traits horizontaux

```
Balise unique                       : <hr>
   attribut fixant l'épaisseur      : size=...
   attribut fixant la longueur      : width=...
    (par rapport à un pourcentage)  : width=..%
   attribut pour un trait plein     : noshade
```

2.5.4 Clignotement

```
<blink> ... </blink>
```

2.5.5 Centrage

```
<center> ... </center>
```

2.6 Structures de présentation

2.6.1 Listes

Non ordonnées :

```
<ul>
<li>
...
</ul>
```

Ordonnées :

```
<ol>
<li>
...
</ol>
```

De description :

```
Balise de début  : <dl>
Balise de titre  : <dt>
Balise de donnée : <dd>
                   ...
Balise de fin    : </dl>
```

2.6.2 Tableaux

```
Balise de début                                        : <table>
   attribut fixant l'épaisseur des traits              : border=...
   attribut/valeur affichant le cadre                  : frame=all
   attribut/valeur affichant les lignes de séparations : rules=all
Légende                                                : <caption> ... </caption>
Passage à la ligne                                     : <tr>
Définition d'une cellule d'entête                      : <th>
Définition d'une cellule du corps du tableau           : <td>
Balise de fin                                          : </table>
```

2.7 Liens

2.7.1 Liens internes

```
<a name="mot_clef"> ... </a>
```

```
<a href="#mot_clef"> ... </a>
```

2.7.2 Liens externes

```
<a href="http://..."> ... </a>
```

2.7.3 Rapatriement d'un document par ftp

```
<a href="ftp://..."> ... </a>
```

2.7.4 Envoi de mail

```
<a href="mailto://..."> ... </a>
```

2.7.5 Intégration d'images

```
<img src="..."> ... </a>
   attribut fixant la largeur de l'image                    : width
   attribut fixant la hauteur de l'image                    : height
   attribut définissant le texte de remplacement de l'image : alt
   attribut fixant l'épaisseur du cadre entourant l'image   : border
   attribut fixant la marge en haut et en bas               : vspace
   attribut fixant la marge à gauche et à droite            : hspace
   attribut fixant l'alignement par rapport au texte suivant : align (top, bottom, middle, left, right)
```

2.8 Formulaires

2.8.1 Structure d'un formulaire

```
Balise de début                         : <form method=GET action="http://...">
Zones de saisie, listes déroulantes, etc : ...                                    ...
Bouton de lancement du formulaire       : <input type=submit value=Envoyez>
Bouton de réinitialisation              : <input type=reset value=Réinitialiser>
Balise de fin                           : </form>
```

Quand l'utilisateur sélectionnera le bouton Envoyez, le script dont l'URL est associée à l'attribut action sera appelé.
Les données seront concaténées à la suite de l'URL :

```
http://.../script.cgi?nom_1=valeur_1&nom_2=valeur_2&...
```

2.8.2 Saisie de texte

Champs de saisie

```
Balise unique                                : <input type=text name=...>
   attribut fixant le nb de caractères affichés : size
```

Zones de saisie

```
Balise de début                            : <textarea ...>
   attribut nommant la zone                : name
   attribut fixant le nombre de lignes     : rows
   attribut fixant le nombre de colonnes : cols
Balise de fin                              : </textarea>
```

2.8.3 Listes déroulantes

```
<select name=...>
        <option> item1 </option>
        ...
</select>
```

2.8.4 Cases à cocher

```
<input type=checkbox name=nom_case_a_cocher value=...>
<input type=checkbox name=nom_case_a_cocher value=...>
...
```

2.8.5 Radio boutons

```
<input type=radio name=nom_bouton_radio value=...>
<input type=radio name=nom_bouton_radio value=...>
...
```

2.9 Insertion d'une applet Java

```
Appel de l'applet                  : <applet code="....class" width=... height=...>
Passage d'un paramètre à l'applet : <param name="..." value="...">
Balise de fin                      : </applet>
```

Récupération des paramètres de la pge HTML dans l'applet :

```
import java.applet.*;

public class Affichage extends Applet {
  public void init ()
  {
     String [] argv = {getParameter("..."), getParameter("...")};
     ...
  }
...
}
```

Deuxième partie
XML

2.10 XML : présentation générale

XML est une représentation textuelle de données. Le composant de base de XML est **l'élément**, c'est à dire un morceau de texte (PCDATA : Parsed Character Data) encadré par deux balises (tags), telles que `<personne>` et `</personne>` :

```
<personne>
   <prénom> pierre </prénom>
   <nom> pompidor </nom>
   <email> pompidor@lirmm.fr </email>
</personne>
```

Comme avec des données semi-structurées, nous pouvons répéter des éléments encadrés par les mêmes balises pour représenter des collections :

```
<lirmiens>
   <description> Les gens du second étage </description>
   <personne>
      <prénom> pierre </prénom>
      <nom> pompidor </nom>
      <email> pompidor@lirmm.fr </email>
   </personne>
   <personne>
      <prénom> jean-françois </prénom>
      <nom> pons </nom>
      <email> pons@lirmm.fr </email>
   </personne>
   ...
</lirmiens>
```

2.11 Eléments vides

XML permet d'abréger la notation d'éléments vides :
`<marié> </marié>` → `<marié/>`

2.12 Emploi des attributs

XML permet d'associer des **attributs** à des éléments (équivalents à des propriétés dans les modèles de données) :
`<prix monnaie="Euro"> 100 </prix>`

Une ambiguïté existe entre éléments et attributs :

```
<personne>
   <prénom> pierre </prénom>
   <nom> pompidor </nom>
   <email> pompidor@lirmm.fr </email>
</personne>
```

ou

```
<personne prénom="pierre" nom="pompidor" email="pompidor@lirmm.fr"/>
```

2.13 Référencement des éléments

Les éléments peuvent être référencés :

```
<adresse id="0">
    <établissement> LIRMM </établissement>
    <rue> 161 rue ADA </rue>
</adresse>
```

```
<personne>
   <prénom> pierre </prénom>
   <nom> pompidor </nom>
   <email> pompidor@lirmm.fr </email>
   <résidence_administrative idref="0"/>
</personne>
```

2.14 Document Type Definition (DTD)

La grammaire sous-jacente au document XML peut être explicitée dans une **DTD (Document Type Definitions)** : Soit un document XML présentant un nombre arbitraire de personnes :

```
<lirmiens>
   <personne>
      <prénom> pierre </prénom>
      <nom> pompidor </nom>
      <email> pompidor@lirmm.fr </email>
   </personne>
   <personne>
      <prénom> jean-françois </prénom>
      <nom> pons </nom>
      <email> pons@lirmm.fr </email>
   </personne>
   ...
<lirmiens/>
```

la DTD associée peut être la suivante :

```
<!DOCTYPE lirmiens [
   <!ELEMENT lirmiens (personne*)>
   <!ELEMENT personne (prénom, nom, email)>
   <!ELEMENT prénom (#PCDATA)>
   <!ELEMENT nom (#PCDATA)>
   <!ELEMENT email (#PCDATA)>
]>
```

2.14.1 Opérateurs intervenant dans les expressions régulières

Voici les opérateurs intervenant dans les expressions régulières :
- `*` : 0 à n
- `+` : 1 à n
- `?` : 0 ou 1
- `|` : alternative
- `,` : concaténation

Une DTD est précisèment une grammaire context-free pour le document considéré. En particulier, cet exemple impose que les éléments *prénom*, *nom* et *email* apparaissent dans cet ordre dans un élément *personne*. Ces grammaires peuvent être récursives (et décrire des arbres ...) :

```
<!ELEMENT noeud (feuille | (noeud,noeud))>
<!ELEMENT feuille (#PCDATA)>
```

2.14.2 Types des attributs

Les DTD permettent aussi de déclarer le type des attributs :

```
<!ATTLIST prix monnaie CDATA #IMPLIED>
```

Voici les différents types d'attributs possibles :
- `CDATA` : texte
- `ID` : identifiant de l'élément
- `IDREF` : référence
- `IDREFS` : liste de références

Voici un exemple employant quelques uns des types précédents :

```
<!DOCTYPE famille [
   <!ELEMENT famille (personne*)>
   <!ELEMENT personne (nom)>
   <!ELEMENT nom (#PCDATA)>
   <!ATTLIST personne id      ID     #REQUIRED>
                      mère    IDREF  #IMPLIED>
                      père    IDREF  #IMPLIED>
                      enfants IDREFS #IMPLIED>
]>
```

2.14.3 Limitations des DTD

Limitations des DTD :
- imposition d'un ordre
- pas de types atomiques (simplement `#PCDATA`)
  et donc pas d'intervalles de valeurs
- ...

2.15 Navigation à partir d'un document XML

Des **liens XML (XLink)** et des **pointeurs XML (Xpointer)** permettent de naviguer sur le Web en référant des positions arbitraires dans les documents.

Index