

Validation des documents HTML et XHTML

Andrei Popescu-Belis
TIM / ETI, Université de Genève

Cours n°4

<http://moodle.unige.ch>

Qu'est-ce que HTML ?

- *Hyper-Text Markup Language*
 - langage de balisage utilisé pour les documents consultables sur le web
 - NB: consultables via le protocole HTTP (Hyper-Text Transfer Protocol) utilisé par un navigateur web
- HTML
 - est une **application** de SGML
 - un ensemble de balises et de règles de syntaxe (DTD)
 - objectif : affichage graphique des documents
 - hypertexte = texte + mise en page + images + ...

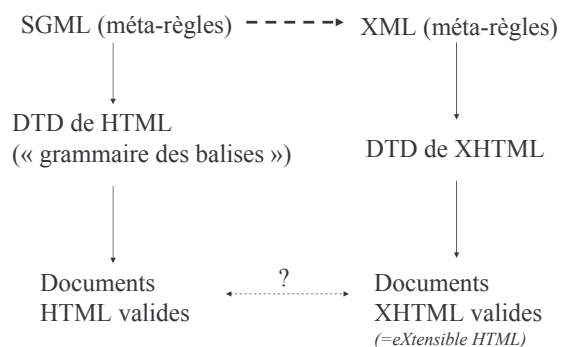
2

Principe de HTML et XHTML

- Les balises (X)HTML indiquent comment un document doit être affiché, et rien d'autre
- Le navigateur web interprète les balises qu'il trouve dans le fichier (dit « source ») et affiche la meilleure représentation graphique possible en fonction des balises
- S'il y a des erreurs, l'affichage peut être très perturbé, ou au contraire pas du tout...

3

Peut-on considérer HTML comme application d'XML ? Pas tout à fait...



4

Théorie et pratique

- En théorie
 - la différence entre HTML et XHTML est très faible si le document HTML est « strictement » valide
- En pratique
 - XHTML respecte toujours la syntaxe XML et sa DTD
 - mais tous les navigateurs ne les comprennent pas encore
 - HTML (du fait de sa diffusion) connaît de nombreuses variantes et formes acceptables
 - les navigateurs web peuvent afficher des documents HTML contenant diverses « erreurs » au sens SGML/XML
 - les normes HTML ont peu à peu intégré les « erreurs »
 - donc HTML est beaucoup plus « flou » que XHTML

5

Intérêt des pages (X)HTML valides

- La page aura le même aspect avec tous les navigateurs qui respectent précisément le standard
- La page pourra être éditée avec n'importe quel outil qui respecte le standard
- Les informations structurelles apparaîtront plus clairement
 - par exemple : la langue, l'encodage, le titre
 - cela facilite l'indexation automatique par les moteurs de recherche
- Il existe un affichage par défaut, mais l'utilisation des feuilles de styles CSS offre plus d'homogénéité à un site

6

Les DTD de HTML et XHTML

- HTML : versions 2.0, 3.2, 4.01
 - variantes : Strict, Transitional, Frame
 - 4.01 : dernière spécification de HTML (1999) en principe définitive
- XHTML : version 1.0 (parue en l'an 2000)
 - aussi Strict, Transitional, Frame
 - également suivie de 1.1 et 2
- Les spécifications et les DTD elles-mêmes sont consultables à <http://www.w3.org/MarkUp/>
- Note : il n'y aura pas de HTML 5 car les développements technologiques concernent désormais XHTML

7

Versions de HTML 4.01

- **HTML 4.01 Strict**
 - permet tous les éléments et attributs qui ne sont pas *deprecated* (dépassés)
 - ne permet pas d'utiliser les *frames*

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01//EN"
"http://www.w3.org/TR/html4/strict.dtd">
```
- **HTML 4.01 Transitional**
 - permet tout ce que permet « Strict » plus les éléments anciens (dépassés) pour des raisons de rétro-compatibilité, sauf les *frames*
 - la plupart des éléments dépassés concernent le formatage graphique
 - devrait en principe être géré uniquement par des feuilles de style

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN"
"http://www.w3.org/TR/html4/loose.dtd">
```
- **HTML 4.01 Frameset**
 - permet tout ce que permet « Transitional » plus les *frames* (cadres)
 - inconvénient des cadres
 - ils perturbent les moteurs de recherche
 - ils ont souvent des problèmes d'emboîtement (affichage des cadres dans les cadres)

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Frameset//EN"
"http://www.w3.org/TR/html4/frameset.dtd">
```

8

Exemple : document HTML 4.01 valide

```
<!DOCTYPE html PUBLIC
"-//W3C//DTD HTML 4.01 Transitional//EN">
<html>

<head>
  <meta http-equiv="Content-Type"
        content="text/html; charset=ISO-8859-1">
  <title>Page de Jean Dupont</title>
</head>

<body>
  <h1>Jean Dupont</h1>
  <p align="center">Jean Dupont est étudiant à
    l'ETI, Université de Genève.</p>
</body>

</html>
```

9

Versions de XHTML 1.0

- Semblables aux 3 versions de HTML

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

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

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

10

Pour écrire du (X)HTML valide

- Choisir une version de DTD publique adaptée
- Modes de rédaction
 - utilisation d'un éditeur HTML/XHTML : le paramétrer si possible sur la bonne DTD
 - rédaction manuelle (éditeur texte) : respecter les spécifications de la documentation et la DTD
- Sources d'invalidité des documents HTML/XHTML
 - absence de déclaration DOCTYPE
 - absence de déclaration du jeu de caractères (si différent de UTF-8)
 - balises non conformes (trop anciennes, trop spécifiques, ...)
 - balises incorrectement fermées, p.ex. :
, <hr>, <p>
 - balises (éléments) qui se croisent
 - balises apparaissant dans un ordre non permis
 - absence de guillemets autour des valeurs des attributs
 - ...

11

Statut de la déclaration XML

- La déclaration XML

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

devrait être obligatoire pour les documents XHTML (application XML)
- Difficulté
 - La déclaration n'est pas forcément comprise par les navigateurs
 - Mais elle est nécessaire pour la validité XML et pour indiquer l'encodage
- Solution
 - les navigateurs récents ne l'affichent pas
 - on ne l'utilise pas pour les documents HTML (logique !)
 - on peut en théorie l'omettre pour XHTML, mais elle est conseillée
 - si omise, il faudra indiquer l'encodage ailleurs

12

Déclarations d'encodage en HTML

```
<HTML>
<HEAD>
<META HTTP-EQUIV="Content-Type" CONTENT="text/html; charset=XXX">
</HEAD>...
```

- Le 'charset' (XXX) peut être
 - UTF-8 = Unicode, le jeu de caractères « universel »
 - ISO-8859-1 ou ISO-LATIN-1 = langues d'Europe Occidentale, i.e. « nos » alphabets courants
 - ISO-8859-2 = langues d'Europe Centrale et Orientale
 - beaucoup d'autres encodages possibles (→ cours 5 et 6)

13

Extensions des fichiers HTML et XHTML

- Rappel (*sous Windows mais aussi sous Unix/Linux*)
 - forme des noms de fichiers : **XXXXXX.YYY**
 - **YYY** est l'extension indiquant le type du document
 - elle peut être cachée par Windows mais elle détermine l'icône du fichier
- Comment distinguer les documents XHTML des documents HTML et des documents XML ?
 - document HTML 4.01 → **mondoc.html** ou **.htm**
 - document XHTML 1.0 → **mondoc.xhtml**
 - document XML autre → **mondoc.xml**
- Note : d'un point de vue strictement logique :
 - on pourrait écrire **.xml** pour un document **.xhtml**
 - on pourrait écrire **.html** pour un document **.xhtml**
 - mais utiliser toujours l'extension la plus précise

14

Rédaction de documents HTML/XHTML

- Options :
 1. Utilisez votre éditeur HTML/XHTML WYSIWYG* favori
 - indiquez-lui la version de HTML ou de XHTML que vous voulez utiliser (DTD)
 2. Écrivez directement la source (les balises) pour un document simple
 - respecter les spécifications de la documentation et la DTDDans les deux cas, n'oubliez pas de vérifier que le document est valide
- HTML Tidy
 - <http://tidy.sourceforge.net/>
 - peut « nettoyer » votre fichier HTML
 - par exemple s'il a été généré par Word
 - peut l'enregistrer au format XHTML
- Amaya
 - <http://www.w3.org/Amaya/>
 - navigateur + éditeur
 - enregistre des documents HTML en XHTML

* WYSIWYG :
*What you see is
what you get*

15

Validation de HTML et XHTML avec <http://validator.w3.org>

- Soumettre le fichier dont on veut tester la validité – deux options :
 - mettre le fichier accessible sur Internet (page web)
 - télécharger le fichier vers le site (= *upload*)
- L'interface validator.w3.org est assez riche
 - lit l'entête du document
 - choisit la DTD et l'encodage observés (si absents → erreur)
 - on peut aussi les ajuster dans l'interface
- Affichage des erreurs trouvées (dans la mesure du possible)
 - suivre les indications sur la page pour corriger les erreurs
- Lorsqu'il n'y a plus d'erreur → le fichier est valide
 - suggestions (bas de page) pour ajouter une vignette de validité (logo « *Valid!* ») dans le fichier validé
 - le logo permet de re-valider le fichier par un simple clic

16

HTML et XHTML sous XML Spy

- Type du document (HTML ou XHTML)
 - si nouveau document, on peut le choisir lors de la création
 - si document existant, indiqué par l'extension (.xhtml ou .html)
 - d'où son importance !
- Document HTML
 - pas de validation possible ☹
 - affichage en mode texte ou « browser » ☺

17

Applications

Création de pages web en XHTML

(ou en d'autres mots : création de documents XML valides selon la DTD de XHTML)

Devoir pour le lundi 20 novembre (voir fin)

Création d'un document XHTML valide... très simplifié

- Dans un premier temps vous allez travailler avec une version très simplifiée de la DTD de XHTML qui vous facilitera la création d'un document XHTML
- Téléchargez depuis <http://moodle.unige.ch> :
 - la DTD : xhtml1-strict-extraits.dtd
 - le fichier à compléter : index-test.xhtml
- Développez le fichier index-test.xhtml avec XML Spy pour en faire une page d'accueil personnelle (ou une autre page web de votre choix)
 - écrivez des balises pour faire varier l'affichage selon les explications fournies dans la DTD elle-même
 - vérifiez régulièrement la **validité** (F8)
 - vérifiez l'**aspect graphique** dans l'onglet Browser ou en consultant le document index-test.xhtml avec Firefox ou IE
 - utilisez au maximum les possibilités offertes par la DTD simplifiée
 - si vous souhaitez les étendre (en fonction de vos connaissances de HTML), vous pouvez modifier la DTD fournie (mais avec précaution...)

19

Affichage du document créé comme une page web

- Copiez ou enregistrez votre fichier index-test.xhtml dans votre dossier H:\PRIVE\WWW
 - le document est visible sur le web à l'adresse <http://home.etu.unige.ch/~VOTRELOGIN/index-test.xhtml>
 - note : l'adresse ne peut pas être trouvée par des inconnus...
- Testez le résultat avec Firefox ou IE
- Testez aussi le résultat de votre voisine ou voisin !
- Note : si vous changez le nom du fichier en index.xhtml et que vous renommez l'ancien index.html avec un autre nom (p.ex. index.old.html), votre page web deviendra simplement <http://home.etu.unige.ch/~VOTRELOGIN>

20

Pour aller plus loin (facultatif)

- On souhaite maintenant passer à la vraie DTD de XHTML (version selon votre choix)
- Faites une copie du fichier précédent (gardez l'extension .xhtml) et changez le DOCTYPE pour indiquer une DTD publique de XHTML (voir cours, T10)
 - vous pouvez continuer à utiliser XML Spy
 - la DTD intégrale offre beaucoup plus de balises
 - vous pouvez aussi valider le fichier à <http://validator.w3.org> en indiquant son adresse URL
- S'il est valide, faites comme indiqué (sur le site ou sur T16) pour afficher le logo « Valid! »
- **Devoir pour le lundi 20 novembre**: indiquez sur moodle l'URL du meilleur résultat obtenu, par exemple <http://home.etu.unige.ch/~VOTRELOGIN/XXXXXX.YYY>

21