

Le langage HTML

Pierre Letouzey
PPS Paris 7 / INRIA

15/2/2012

Lien avec la partie programmation

HTML est un langage informatique:

- ▶ il ne s'agit pas d'un langage de programmation
- ▶ mais d'un langage de description de document.
- ▶ On retrouve la notion de syntaxe précise
- ▶ Ainsi que celle d'interprétation
 - ▶ ici pour obtenir un document affiché

Autres exemples de langages de document

- ▶ PostScript (.ps) : on programme son document!
- ▶ Portable Document Format (.pdf)
- ▶ Bureautique : .doc / .rtf / .odt / ...
- ▶ Tex / Latex : .tex

Autres exemples de langages de document

A chacun ses spécificités :

- ▶ PDF : rendu uniforme entre machines, imprimantes, ...
- ▶ HTML : rendu adaptatif
 - ▶ smartphone
 - ▶ écran géant
 - ▶ tablette Braille

En pratique beaucoup de pages web "forcent" le rendu, et l'accessibilité est souvent très imparfaite

Lien avec la partie réseau

- ▶ notion d'URL :
`http://machine.dom.tld:p/chemin/fich.html?arg=v`
- ▶ Protocole HTTP:
 - ▶ Connexion à `machine.dom.tld` sur le port `p` (p.ex. 80)
 - ▶ Écriture de: `GET chemin/fich.html`
 - ▶ Lecture du fichier précédé d'une entête :
`Content-Type, Content-Length, ...`
- ▶ Démo...

HTML : un standard ouvert

- ▶ Un organisme de normalisation : le W3C
- ▶ Une norme accessible à tous : <http://www.w3.org/TR/>
- ▶ Ceci garantie une grande portabilité
- ▶ Malheureusement:
 - ▶ Manque de respect des normes (IE6...)
 - ▶ Extensions propriétaires : Flash

HTML : un standard en évolution

- ▶ 1989-1992 : Les origines du Web
- ▶ 1993 : NCSA Mosaic...
- ▶ 1994 : Netscape...
- ▶ 1995-1996 : HTML 2.0
- ▶ 1997 : HTML 3.2 et 4.0
- ▶ 2000-2006 : XHTML
- ▶ 2007-2012 : HTML 5

M pour Markup : structuration du document

- ▶ Un fichier HTML est un simple texte, structuré par des **balises**:
 - ▶ `<h1>Ceci est un titre</h1>`
- ▶ Ces balises peuvent être imbriquées, ce qui donne au document une structure d'arbre

$$\text{html} \left\{ \begin{array}{l} \text{head} \\ \text{body} \left\{ \begin{array}{l} \text{h1} \\ \text{p} \\ \text{table...} \end{array} \right. \end{array} \right.$$

Correction d'un document HTML

- ▶ Un fichier HTML peut être mal formé vis-à-vis de la norme.
- ▶ Un validateur est mis à disposition par le W3C
 - ▶ <http://validator.w3.org/>
- ▶ Comme pour un programme, attention à la syntaxe !
- ▶ Autres soucis fréquents : balises mal fermées, mal imbriquées...
- ▶ Malheureusement¹, les navigateurs web actuels acceptent tant bien que mal les documents mal formés.

¹ou pas...

H pour HyperText : les liens

- ▶ `Ceci est un lien`
- ▶ Cf. Wikipedia

Distinction fond / forme

- ▶ Les balises peuvent contenir des indications d'aspect
 - ▶ `<h1 style="color:red">Titre en rouge</h1>`
- ▶ Il est maintenant fortement recommandé de regrouper ces indications dans une **feuille de style** CSS
 - ▶ `h1 { color : red }`
- ▶ Intérêt:
 - ▶ Document plus "sémantique"
 - ▶ Indications de forme pour l'ensemble du document
 - ▶ Possibilité de changer de style plus facilement

HTML manuel / engendré

- ▶ La syntaxe HTML est peu commode à écrire à la main
- ▶ Les documents HTML sont souvent engendrés par des outils à partir d'autres langages sources
- ▶ Ex: OpenOffice, wiki, latex2html, ...
- ▶ La qualité de ces pages engendrés peut être très variable (ex: Frontpage)
- ▶ En particulier la séparation fond/forme est rarement respecté

HTML statique / dynamique

- ▶ **Page statique**
 - ▶ Le serveur web contient directement le fichier demandé
- ▶ **Page dynamique du coté serveur**
 - ▶ Le serveur web fabrique au vol la page (cf. php)
 - ▶ Exemple: interrogation d'une base de donnée sur le serveur
- ▶ **Page dynamique du coté client**
 - ▶ Le fichier fourni contient des extensions à html exécutés par le navigateur
 - ▶ Exemple: balise `<script>`, usage de Javascript
 - ▶ Page animée et/ou interactive (Web 2.0), etc
 - ▶ Lien vers les questions de droits, de vie privée, etc

Pour continuer

- ▶ Un cours complet avec exercices
 - ▶ www.gchagnon.fr/cours/html/