



HTML

HyperText Markup Language





Glossaire

- HTML: HyperText Markup Language
- SGML: Standard Generalized Markup Language
- URL: Uniform Ressource Locator
- URI: Uniform Ressource Identifier
- URN: Uniform Ressource Name
- ISBN: International Standard Book Number
- IETF: Internet Engineering Task Force
- DTD : Document Type Definition
- jpeg: Joint Photographic Experts Group
- gif: Graphic Interchange Format
- png: Portable Network Graphics





Plan

- Hypertexte
- Identification des ressources
- Le Langage HTML
 - Introduction
 - Structure et principes de base
 - Eléments





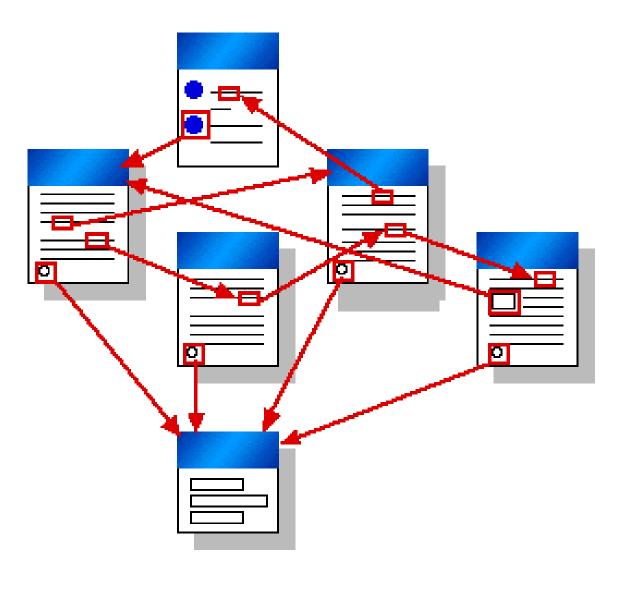
Hypertexte

- Système contenant des nœuds liés entre eux par des hyperliens
- Hyperlien : référence permettant de passer d'un document consulté à un document lié
 - Source: mot, phrase, image,...
 - Cible : partie du même document, document externe, partie d'un document externe, application,...
 - Basé sur les URL
- Langage HTML





Hypertexte







Identification des ressources

- Identification permanente d'une ressource (même en cas de déplacement)
- URI: Uniform Ressource Identifier
- URL: Uniform Ressource Locator
 - Localisation de la ressource de manière unique
 - Fournit les moyens d'agir sur la ressource et d'en obtenir une représentation
 - Utilisé dans le Web
- URN: Uniform Ressource Name
 - Identification par le nom (pas d'info sur l'emplacement ou la manière de l'obtenir)
 - Ex : numéro d'ISBN (International Standard Book Number)





Identification des ressources URL (1/3)

- Souvent appelé « adresse Web »
- Composé de :
 - Protocole de communication : http, ftp, Idap, mailto, telnet,...
 - Nom d'utilisateur (optionnel)
 - Mot de passe (optionnel)
 - Adresse IP ou nom de domaine
 - Numéro de port (optionnel)
 - Chemin d'accès à une requête (optionnel)
- Remarque : caractères spéciaux → % (code UTF-8, ASCI,...)





Identification des ressources URL (2/3)

URL absolue :

- EX: http://Dick:Rivers@www.dick-rivers.com:8888/chemin/d/acces.php?q=req&q2=req2#signet
- Schéma de représentation " http: "
- Localisation complète de la ressource " //"
 - > Données d'authentifications " Dick:Rivers @ "
 - > Nom de domaine " www.dick-rivers.com "
 - > Numéro de port " :8888 "
- Nom complet de la ressource
 - Chemin absolu " /chemin/d/"
 - Nom de la page (optionnel) " acces.php "
 - Données transmises (optionnelles) " ?q=req&q2=req2 "
- Données pour l'exploitation de la ressource (optionnelles)
 " #signet "





Identification des ressources URL (3/3)

- URL relative :
 - Ne contient pas
 - > Protocole
 - > Nom de domaine
 - Déduits à partir de la ressource contenant l'URL absolue
 - Souvent utilisés à l'intérieur d'un même site Web (évite la réécriture des URL en cas de déplacement...)
 - Inspiré du système de fichiers Linux :
 - > ./: dossier actuel
 - ../ dossier parent
 - /: dossier racine





Langage de balisage permettant de décrire un document basé sur SGML

- Permet :
 - De structurer et mettre en forme le contenu des pages Web
 - D'inclure des ressources multimédias, des formulaires de saisie et des éléments programmables
- Interopérabilité :
 - Fonctionne sur une large gamme de matériels
 - Rendu final dépend du logiciel client (navigateur,..)





HTML: Introduction Historique

- 1990 : Création par Tim Berner-Lee au Centre Européen de Recherche Nucléaire (CERN)
- 1995 : HTML 2.0 normalisation par l'IETF
- Début 1997 : HTML 3.2 ajout des tables, des applets, etc...
- Fin 1997 : HTML 4.0 ajout des feuilles de styles, des frames, etc...
 Séparation de la présentation et du contenu
- Décembre 1999 : HTML 4.01 dernière version HTML
- 2000 : XHTML 1.0 reformulation de HTML 4 en XML 1.0
- 2006: XHTML 2.0
- Actuellement : Développement de HTML 5 (compatibilité, documents et applications en ligne, enrichir interface utilisateur)

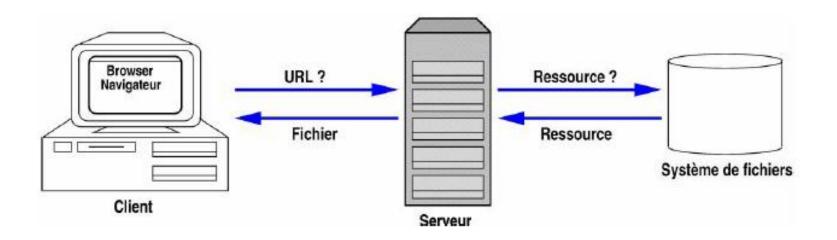
Normalisation par le W3C depuis 1996





Fonctionnement

- Le navigateur effectue une requête spécifiée à travers l'URL
- Le serveur retourne un flot typé de données
- Le navigateur interprète le flot de données et l'affiche







Outils de production de pages Web

- Editeurs :
 - Classiques
 - Orientés HTML : aide à la saisie des balises
 - WYSIWYG (What You See Is What You Get):
 - ➤ Composition visuelle
 - > Composition code source
 - Suppléments : modèles de document, projets, documentation sur les langages, remise en forme automatique, coloration,...
- Blogs: DotClear, WordPress, Blogger
- Système de publication de contenu (CMS) :
 - Ouverts (ex : Wikipedia)
 - Particuliers ou entreprises : SPIP, eZPublish, Joomla, Typo3





Structure du langage

Fichier HTML →

fichier texte constitué d'éléments eux-mêmes constitués de balises et d'attributs permettant d'apporter des informations sur le contenu





Séparation structure & présentation

- Structure ≠ apparence :
 - Paragraphe est une structure dont l'apparence est différente en français et en américain...
 - HTML décrit la structure du document :
 - Historique : initialement dérive du langage documentaire SGML
 - Portabilité : indépendant des navigateurs
 - ➤ Libre choix de l'utilisateur : document visuel ou sonore,... par exemple en cas de handicap
- Séparation :
 - Structure: HTML
 - Présentation : feuilles de style (CSS)
- Recommandé par le W3C
- Imposé dans le langage XHTML





HTML: Introduction Validation du code

- Outils de validation :
 - W3C

http://validator.w3.org/

 Extension Firefox : HTML Validator https://addons.mozilla.org/fr/firefox/addon/249/

• ...





Document Type Definition (DTD)

- Disponibles à l'adresse suivante : http://www.w3.org/QA/2002/04/valid-dtd-list.html
- Référence permettant le respect des règles
- Indispensable pour les validateurs (X)HTML
- Rendu différent en fonction de la DTD sur les navigateurs modernes (anciens et nouveaux codes...)
- Plusieurs types possibles :
 - Strict : suit rigoureusement la spécification de la version spécifiée
 - Transitional : autorise l'usage d'éléments caduques (venant des versions antérieures) → n'existe plus en XHTML 1.1
 - Frameset : transitional + éléments liés aux frames → n'existe plus en XHTML 1.1





HTML : Structure et bases Eléments

Partie structurée du document

- Contient des attributs et du contenu
- Représente diverses structures : en-têtes, paragraphes, liens hypertexte, listes, médias, ...

Types d'éléments :

- Bloc (block): grandes structures contenant d'autres blocs, éléments texte ou du texte (titres, paragraphes, tables,...)
 - agencement : l'un en dessous l'autre (succession verticale)
- En ligne (inline) : petites structures contenant d'autres éléments texte ou du texte (hyperliens, citations, images,...)
 - agencement : l'un après l'autre (succession horizontale)
- Auto-fermant : pas de contenu, uniquement pour la structure (saut de ligne, tabulation,...)





Balises

- Permettent de délimiter des éléments :
- Syntaxe :
 - Balise ouvrante (en début) : <balise>
 - Balise fermante (en fin) : </balise>

Ceci est un paragraphe.

Balise seule : <balise />





Attributs

- Caractéristiques de l'élément : référence vers une adresse HTML (lien hypertexte), emplacement d'un fichier (image), ...
- Syntaxe :

```
<balise attribut1="valeur1" attribut2="valeur2" attribut3="valeur3">texte</balise>
```






HTML 4/XHTML 1.0

- XHTML 1.0 est une reformulation de HTML 4 en tant qu'application XML 1.0
- Principales différences dans XHTML 1.0 :
 - Noms d'éléments et d'attributs en minuscules
 - Valeurs des attributs entre guillemets ou apostrophes
 - Un attribut a toujours une valeur
 - Toute balise ouverte doit être fermée
 - Cohérence dans les imbrications de balises





Commentaires

- Importance :
 - Compréhension du code
 - Maintenance du site
 - Evolution du site
- Temps de chargement de la page augmenté
- Syntaxe :

<!-- Commentaires -->





HTML: Structure et bases DOCTYPE

- Carte d'identité de la page Web
- Référence au langage utilisé à travers le DTD
- Permet de valider le code
- Référence au HTML 4.01 strict :

Référence au XHTML 1.0 :

Référence au XHTML 1.1 :

<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.1//EN" "http://www.w3.org/TR/xhtml11/DTD/xhtml11.dtd">





HTML: Structure et bases Squelette HTML 4

- <html> : délimite le document HTML
- <head> : en-tête contenant les informations sur le document
- <body> : corps du document





HTML: Structure et bases Squelette XHTML 1.0

Balises en minuscule...





HTML: Structure et bases Squelette HTML 5





HTML: Structure et bases En-tête (1/2)

- title : titre affiché dans la barre de titre du navigateur
- meta :
 - méta-informations sur le document

```
<meta name="author" content="Odile Deray" />
<meta name="keywords" content="Cannes, festival" />
<meta name="description" content="Présentation du film Red is dead" />
<meta name="date" content="1994-03-09T16:55:37+01:00" />
<meta http-equiv="Content-Type" content="text/html; charset=iso-8859-1" />
```

Encodage du fichier sur le serveur

- Informations à destination du navigateur via HTTP
 - > Ex : Mise à jour du cache :

<meta http-equiv="Expires" content="Sat, 08 Jan 2011 07:30:27 GMT" />





HTML: Structure et bases En-tête (2/2)

• **script**: introduction de programmes permettant l'utilisation d'effets dynamiques dans la page

```
<script type="text/javascript" src="fonctions.js"/>
```

link: lien vers des fichiers utilisés par le document

style : insérer du code CSS

```
<style type="text/css">
/* Votre code CSS ici */
</style>
```





HTML: Structure et bases Caractères spéciaux (1/2)

- Par défaut : codage des caractères en ASCII (128 caractères)
- Possibilité d'ajouter des caractères :
 - Impossible à insérer sous forme de texte (car interprétés comme caractère de contrôle)

Caractère	Code iso	Code HTML
"	& #34;	"
&	& #38;	&
<	& #139;	<
>	& #155;	>





Caractères spéciaux (2/2)

 Non compris dans l'encodage choisi (ex : caractères accentués en français)

Caractère	Code iso	Code HTML
€	€	€
à	à	à
Ç	ç	ç
é	é	é





Documentation

- Spécification HTML 4.01
 (contient la liste des éléments HTML) : http://www.w3.org/TR/html4/
- Particularités du XHTML 1.0 : http://www.w3.org/TR/xhtml1/
- Particularités du XHTML Basic 1.1 : http://www.w3.org/TR/2008/REC-xhtml-basic-20080729/





Groupements d'éléments

- Organisation du document, mise en forme
- 2 balises :
 - <div>: type bloc
 - : type en ligne
- Attributs principaux :
 - class : classe CSS
 - id : nom permettant de repérer la balise
 - style : intégration directe de CSS (à éviter...)





Titres (ou en-têtes) et paragraphes

- 6 niveaux de titres :
 - Balises <h1>, <h2>, ..., <h6>
 - Présentation hiérarchisée (taille de police,...)
 - Génération automatique de la table des matières
- Paragraphes :
 - Balise : saut de ligne à l'intérieur du paragraphe avec la balise
br/>
 - Balise : préformaté → le navigateur affiche le texte tel qu'il est formaté dans le code source (espaces, tabulations, sauts de ligne,...)





Eléments sémantiques

- : Mise en valeur
- : Mise en valeur forte
- <cite> : extrait ou référence à une autre source
- <code> : code d'un programme
- <abbr> : forme abrégée
- ...





Listes

- 3 types de listes :
 - Listes à puces () :
 - ➤ Chaque item est introduit par la balise
 - Listes ordonnées ()
 - ➤ Chaque item est introduit par la balise
 - Listes descriptives (<dl>):
 - ➤ Chaque item est composé :
 - d'un terme : balise <dt>
 - d'une description : balise <dd>





Tableaux

- Eviter de les utiliser pour la mise en page du document
- Elément composé des éléments suivants :
 - Légende : <caption>
 - En-tête : <thead>
 - > Cellules d'en-tête :
 - Corps:
 - ➤ Lignes : >
 - ➤ Cellules : (attribut *headers* fait le lien entre les en-têtes et la cellule)
 - Fin : <tfoot>





Images

- Temps de chargement (les dernières versions de certains formats permettent de charger les images par parties...)
- Formats principaux :
 - Jpeg : photographies, images scannées
 - Gif: images simples avec des traits et de larges zones de couleurs, possibilité d'animer l'image
 - Png : adapté au Web (transparence plus aisée qu'ne GIF)
- Balise dont les attributs principaux sont :
 - src : emplacement de l'image
 - alt : texte alternatif à afiicher à la place de l'image (affichage des images désactivé)





Liens hypertexte

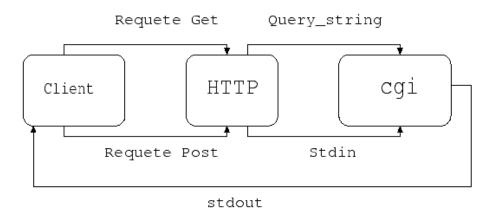
- Balise <a> avec l'attribut href
 - Liens externes
 - Liens internes
- Types de liens :
 - Document (fichier HTML, texte, image, son, film,...)
 - Autre service Internet (FTP, messagerie électronique,...)
- Lien sur différents éléments : texte, image,...





Formulaires (1/4)

- Introduction de l'interactivité
- Elément <form> dont les attributs sont :
 - action : URL chargée du traitement des données
 - method : méthode d'acheminement des données
 - ➤ GET (par défaut) : transmis dans la partie query de l'URL → visible
 - ➤ POST : transmis au serveur dans le flot de données → invisible







Formulaires (2/4)



Injection SQL

SELECT * from admins WHERE login='\$login' AND password='\$password'

SELECT * from admins WHERE login=" OR 1=1





Formulaires (3/4)

- **Enctype**: méthode d'encodage pour la forme que prendra le contenu du formulaire (uniquement pour POST)
 - application/x-www-form-urlencoded (par défaut) : contenu du formulaire sous forme URL (exemple : nom=valeur_nom&prenom=valeur_prenom)
 - > Text/plain : contenu du formulaire sous forme de texte lisible (utilisé pour l'action mailto)
 - Multipart/form-data : permet d'expédier un fichier attaché au message





Formulaires (4/4): Contenu

Balises:

- <fieldset> : regroupement dans un formulaire
- <legend> : légende du formulaire
- <label> : étiquette
- <input> : champ de saisie et différents types de boutons
- <buton> : boutons
- <select> : menu déroulant
- <textarea> : zone de saisie de texte libre (multiligne)





Multimédia

- Insertion de fichiers audio, vidéo, flash, java,...
- Utilisation de programmes externes : plug-in
- Liste des plug-in installés sous Firefox :
 - Outils/Modules complémentaires





Multimédia : contenu

- Balise <embed> : obsolète
- Balise<object> (inclusion de tout type d'objet : images, vidéos, applets Java, interfaces VRML,...) dont les attributs principaux sont :
 - type : type MIME
 - data: URI du document
 - standby: en attendant le chargement...
- Balise<param> : définition des paramètres des objets
 - src : nécessaire pour localiser la source sous IE
 - autoplay, autostart : démarrage automatique
 - ...





Bibliographie

- Cours d'Olivier Rossigny (ISEN Lille)
- Cours d'Alexandre Pauchet (INSA Rouen)
- Cours de Gilles Changon
- Wikipedia
- http://www.iana.org
- http://www.w3.org
- http://xhtml.le-developpeur-web.com
- http://www.w3.org/QA/2002/04/valid-dtd-list.html
- http://www.alsacreations.com/article/lire/560-dtdhtml-xhtml-comment-choisir.html