

Leçon 1.1 – HTML Structure et balises

s1 -----

Dans cette partie, nous allons donc découvrir HTML, acronyme de Hyper Text Markup Language. Nous allons voir comment est structurée une page web et introduire quelques balises, parmi les plus courantes.

s2 -----

HTML est le langage qui permet de décrire une page web. Qu'est-ce qu'une page web ? Une page web est tout simplement un fichier texte transmis par un serveur après une requête.

Qu'est-ce qu'un serveur ? C'est une machine connectée au réseau et dont la tâche principale est d'attendre des requêtes et renvoyer des fichiers.

Comment faire une requête ? C'est par exemple taper l' URL - c'est-à-dire ajouter l'adresse web - de la page d'accueil d'un site web dans la barre d'adresses d'un navigateur ou le fait de cliquer sur un lien dans une page d'un site web.

Un navigateur c'est donc - entre autres - une application qui effectue des requêtes, les transmet à un serveur et en interprète le résultat pour l'afficher sur un écran.

Mais les navigateurs ne sont pas les seules applications à interpréter les fichiers HTML. Citons les robots qui parcourent les pages web et indexent leur contenu pour enrichir les moteurs de recherche ou les lecteurs braille qui traduisent des éléments textuels en langage Braille.

s3 -----

Les extensions les plus communes pour un fichier HTML sont .html ou .htm.

On distingue deux types de pages : les pages dites statiques, les pages dynamiques.

Les premières ont un contenu figé : on va voir que du code informatique permet de modifier l'aspect de cette page une fois qu'elle est affichée par le navigateur, cependant elle est figée dans le sens où le serveur se contente de transmettre un fichier qui existe déjà tel quel ; les secondes ont un contenu qui est élaboré par le serveur, par exemple lorsqu'on affiche un compte en banque. Cette exécution met en oeuvre d'autres langages tels que PHP ou Java : les extensions peuvent alors être différentes de HTML ou HTML. Par exemple .php ou .jsp...

s4 -----

Comment crée-t-on une page web ?

On peut créer une page web tout simplement avec un éditeur de texte tel que NotePad, TextEdit, Emacs. C'est ce que nous ferons dans ce module.

On peut aussi utiliser des nombreux éditeurs dédiés à ce langage qui permettent de travailler en mode WYSIWYG (What You See Is What You Get), tel que Dreamweaver d'Adobe, par exemple : en effet, il est bien plus rapide de dessiner un tableau par exemple que d'écrire à la main le texte HTML correspondant. Pour autant, même avec ces outils, il est important de savoir comment

fonctionne HTML afin de garder le contrôle sur ce qui est produit et de comprendre l'origine de possibles erreurs d'affichage.

Comment s'exécute sur notre ordinateur un fichier HTML statique ? Tout simplement avec un navigateur. La lecture d'une page dynamique que l'on viendrait de créer nécessite, comme on l'a dit plus haut, la présence d'un serveur sur notre machine.

Ainsi la création et le test d'une page statique peuvent se faire sans connexion internet. Lorsque le fichier sera prêt, il faudra bien entendu le déposer sur un serveur web pour le rendre accessible aux internautes.

Comment est-on sûr que ce fichier est correctement écrit et sera donc correctement lu par l'ensemble des navigateurs ? Le mieux est de le tester sur un panel de navigateurs, voire de versions différentes. Notons que la plupart des navigateurs (Firefox, Chrome, Opéra...) possèdent des fonctionnalités utiles pour déboguer, que l'on découvrira à la fin de ce module.

On peut aussi utiliser un validateur en ligne, comme celui proposé par le consortium www.w3.org.

s5 -----

Voyons maintenant dans le détail ce que contient un fichier HTML. Un fichier HTML contient un ensemble de balises. Une balise est un ensemble de caractères entre crochets, plus exactement entre le signe < et le signe >. Une balise caractérise un contenu, par exemple un paragraphe, un tableau, une partie de texte en italique

dans cet exemple le mot bateau est situé entre la balise ouvrante <i> et la balise fermante </i>, l'ensemble signifiant au navigateur que le mot doit être affiché en italique

Comment est structurée une page web en fonction de ces balises ?

D'abord, il est conseillé d'indiquer dans la page HTML le prologue du type de document, c'est-à-dire une référence à la norme HTML utilisée afin de préciser dans quel type de HTML la page a été écrite. Cette déclaration se fait a minima par une ligne du type <!DOCTYPE HTML>. L'intérêt est de pouvoir faire passer les pages au validateur du consortium W3C, afin de vérifier qu'on n'a pas fait d'erreur sur la syntaxe du HTML. Nous n'entrerons pas dans les détails de cet élément.

La seconde balise est la balise <HTML> elle indique le début du document proprement dit ; on retrouvera tout à la fin du document la balise fermante </HTML> ; le contenu HTML est donc délimité par ces deux balises.

On trouve ensuite l'en-tête. Cette partie, qui n'est pas visible contient des propriétés globales, comme le titre du fichier, les mots-clés que recueillent les navigateurs, etc.

Vient en suite le corps. Le contenu proprement dit de la page, celui qui fait l'objet de l'affichage, est situé entre les balises ouvrantes et fermantes body.

s6 -----

Ces jeux de balises ouvrantes et fermantes définissent à la fois une sémantique (les balises ont un sens, par exemple p pour paragraphe), un comportement

(certaines balises impliquent une interaction, comme le lien hypertexte a), et éventuellement une apparence prédéfinie (pour la plupart des navigateurs, le lien a une apparence bleue et soulignée).

Prenons l'exemple de la balise p. La sémantique indique qu'on est en présence d'un paragraphe. Ici pas de comportement particulier. L'apparence du texte est ici régie par des propriétés prédéfinies par le navigateur, par exemple : Times 12 points, noir.

On trouve des balises autofermantes telles que (pour image) <hr> (pour ligne horizontale)
 (saut de ligne)...

Certaines balises possèdent des attributs (certains obligatoires, d'autres facultatifs) qui précisent ou modifient ses propriétés. Par exemple, au sein de la balise img, on définit le chemin vers l'image à afficher.

```

```

s7 -----

On distingue trois types de balises.

Des balises de contenu de flux : ces balises sont des éléments de structuration de la page à afficher.

On trouve par exemple les balises

- a pour lien hypertexte
- p pour paragraphe
- div pour une partie de contenu
- header pour l'en-tête
- footer pour le pied de page
- table pour tableau
- form pour formulaire
- etc..

On trouve aussi des balises de contenu de phrasé. Ces balises en général imbriquées dans les précédentes confèrent au contenu qu'elles enserrent une valeur sémantique

- a pour lien hypertexte (a est un cas particulier : à la fois contenu de flux et de phrasé)
- em pour une mise en relief du texte en italique
- strong pour un contenu en gras
- span pour un contenu à isoler
- input pour une entrée de texte
- select pour une liste déroulante

On trouve enfin des balises de Métadonnées qui introduisent des données complémentaires

- link, qui permet de définir un lien avec une page extérieure (en pratique, un fichier de style pour la mise en page du document)
- style, qui définit directement la mise en page des éléments, position, couleur, etc.
- meta, qui définit les métadonnées du fichier, les mots-clés, par exemple
- script, qui définit du code additionnel, par exemple JavaScript que nous découvrirons plus tard

s8 -----

Un mot sur les commentaires. Ils ne sont pas affichés dans le navigateur. Ils donnent des indications sur la nature du contenu à un futur développeur, mais aussi souvent à soi-même lorsqu'on reprend le

code quelques mois plus tard. Les commentaires débutent par <!-- et finissent par -->

s9 -----

Un petit mot sur l'encodage. L'encodage est la manière de stocker en binaire (c'est-à-dire grâce à des 0 et des 1, des bits) les caractères. C'est bien entendu de cette façon que les données vont circuler sur le réseau.

Depuis les débuts du web, l'encodage a évolué en fonction des besoins. Par exemple, s'est fait sentir rapidement le besoin d'afficher des caractères accentués, des caractères asiatiques ou arabes...

Le consortium W3C prescrit actuellement l'encodage UTF-8 (qui est un codage des caractères sur 1 à 4 octets, 1 octet valant 8 bits).

Une des façons d'indiquer au navigateur de quelle façon le document a été encodé est de le préciser dans le <head> par une balise

```
<meta charset= "UTF-8">
```

s10 -----

Pour finir sur HTML, un mot sur le type MIME.

Le type MIME (acronyme de Multipurpose Internet Mail Extensions) est utilisé pour préciser le type des documents attachés à un courrier, mais aussi pour typer les documents transférés par le protocole HTTP pour que le navigateur puisse savoir de quelle manière afficher un document, que ce soit un document HTML, PDF, Flash, etc.

Le navigateur cherche le type MIME dans l'entête du message HTTP envoyé par le serveur (mais il n'est pas toujours présent)

s11 -----

Comme je le disais en introduction de ce module, nous n'allons pas ici examiner une à une toutes les balises. Mon intention est de vous présenter les principes et de vous rendre autonome afin de découvrir au fil de vos réalisations futures la richesse de ce langage.

Je vous propose deux exercices qui vous permettront de vous familiariser avec HTML.

Dans la prochaine partie, nous introduirons le langage CSS, qui sert notamment à décrire le rendu des documents HTML.

Bon travail.