



JEUX DE CARACTERES, SYMBOLES ET
METADONNEES EN HTML

Introduction aux langages HTML et CSS

El hadji Mamadou NGUER Enseignant chercheur
en Informatique UGB



Chapitre 3 : Jeux de caractères, symboles et métadonnées

Séquence 3 : (Jeux de caractères, symboles et métadonnées)

- **Objectifs spécifiques** : A la suite de cette séquence, l'étudiant doit être capable de :
 1. bien utiliser les jeux de caractères, les entités et les symboles dans une page web
 2. Décrire les différentes métadonnées d'un document HTML
 3. utiliser efficacement les méta-informations dans un document HTML
 4. décrire l'encodage d'url

Les jeux de caractères (Charset)

Un jeu de caractères (encodage de caractères) est une façon de représenter chaque caractère (et donc un texte) dans cette base binaire.

Pour afficher une page HTML correctement, un navigateur Web doit savoir le jeu de caractères à utiliser.

Les jeux de caractères usuelles :

- ASCII a été la première **norme d'encodage de caractères**. Il définit 128 différents caractères alphanumériques qui pourraient être utilisés sur Internet: les chiffres (0-9), les lettres anglaises (AZ) et certains caractères spéciaux comme ! \$ + - () @ < et >.
- ANSI (Windows 1252) était le jeu de caractères d'origine de Windows. Il supporte 256 codes de caractères différents.

Les jeux de caractères (Charset)

Les jeux de caractères usuelles :

- ANSI (Windows 1252) était le jeu de caractères d'origine de Windows. Il supporte 256 codes de caractères différents.
- ISO-8859-1 est le jeu de caractères par défaut pour HTML 4. Il prend en charge également 256 codes de caractères différents.
- ANSI et ISO-8859-1 étant si limitées, HTML 4 prend en charge également UTF-8.
- UTF-8 (Unicode) couvre presque tous les caractères et les symboles dans le monde.
- L'encodage de caractère par défaut pour HTML5 est UTF-8.

Les jeux de caractères (Charset)

Exemple d'encodage de caractères :

Caractère	ASCII	ISO-8859-1	UTF-8	UTF-16	UTF-32
e	01100101	01100101	01100101	00000000 01100101	00000000 00000000 00000000 01100101
é	—	11101001	11000011 10101001	00000000 11101001	00000000 00000000 00000000 11101001
?	—	—	11100001 10111010 10111001	00011110 10111001	00000000 00000000 00011110 10111001

Les entités et symboles

- Certains caractères sont réservés en HTML. Il n'est pas possible d'utiliser le signe inférieur (<) ou supérieur (>) dans votre texte, parce que le navigateur les prendra pour des balises.
- Pour afficher réellement les caractères réservés, nous devons utiliser les entités de caractères dans le code source HTML.
- Syntaxe d'une entité: `&entity_name;` ou `&#entity_number;`
- Exemple : Pour afficher le signe inférieur (<) il faut écrire: **<** ou **<**;
- **Avantage et inconvénient :**
L'avantage d'utiliser un nom d'entité, au lieu d'un nombre, c'est que le nom est facile à retenir. Mais l'inconvénient est que les navigateurs ne supportent pas tous les noms d'entité.

Les entités et symboles

Entités de quelques caractères usuels

- Les noms d'entités sont sensibles à la casse!

Résultat	Description	Nom	Numéro
	espace insécable	 	
<	Inférieur à	<	<
>	supérieure à	>	>
&	esperluette	&	&
¢	cent	¢	¢
£	livre	£	£
¥	yen	¥	¥
€	euro	€	€
©	copyright	©	©
®	registered trademark	®	®

Pour voir la liste complète de toutes les entités de caractères :
w3schools.com/tags/ref_entities.asp

Les métadonnées (l'élément head)

- L'élément **<head>** est un conteneur pour des métadonnées (données sur les données) et est placé entre la balise `<html>` et la balise `<body>`.
- Une métadonnée est une donnée sur le document HTML. Elles ne sont pas affichées.
- Elles définissent généralement le titre du document, le jeu de caractères, des styles, des liens, des scripts, et d'autres méta-informations.
- Les balises suivantes décrivent les métadonnées: `<title>`, `<style>`, `<meta>`, `<link>`, `<script>` et `<base>`.

```
<!DOCTYPE ht  
<html xmlns=  
  <head>  
    <meta ht  
    ht
```

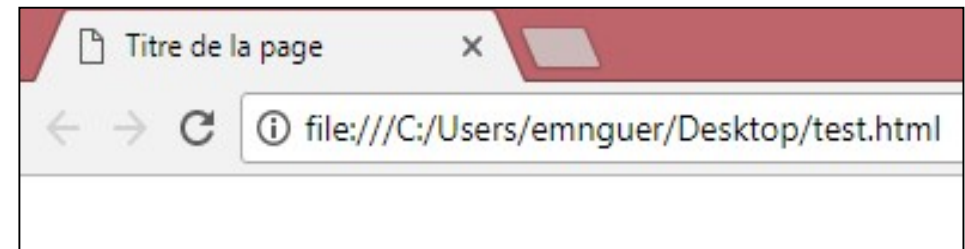

Les métadonnées (l'élément head)

L'élément <title>

- Il définit le titre du document et est nécessaire dans tous les documents HTML / XHTML.
- Il permet de :
 - définir un titre dans l'onglet du navigateur
 - fournir un titre de la page quand il est ajouté aux favoris
 - afficher un titre de la page dans les résultats des moteurs de recherche
- **Exemple:** `<head>`

`<title>`Titre de la page`</title>`

`</head>`



Les métadonnées (l'élément head)

L'élément <style>

- L'élément **<style>** est utilisé pour définir les informations de style pour une seule page HTML.
- Exemple :

```
<style>  
  body {background-color: powderblue;}  
  h1 {color: red;}  
  p {color: blue;}  
</style>
```

L'élément <link>

- L'élément **<link>** est utilisé pour créer un lien vers des feuilles de style externes:
- Exemple : `<link rel="stylesheet" href="mystyle.css">`

Les métadonnées (l'élément head)

L'élément <meta>

- L'élément **<meta>** est utilisé pour spécifier le jeu de caractères utilisé, la description de la page, les mots clés, l'auteur et d'autres métadonnées.
- Les métadonnées sont utilisées par les navigateurs (pour savoir comment afficher le contenu), par les moteurs de recherche (pour les mots clés) et d'autres services Web.
- **Exemple**

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

```
<meta name="description" content="Tutoriel HTML">
```

```
<meta name="keywords" content="HTML,CSS,XML,JavaScript">
```

```
<meta name="author" content="Jean Marie Dembelé">
```

```
<meta http-equiv="refresh" content="30">
```

Les métadonnées (l'élément head)

Réglage de la fenêtre

- HTML5 a introduit une méthode permettant aux concepteurs de sites web de prendre le contrôle de la fenêtre à travers la balise <meta>.
- Le **viewport** est la zone visible à l'utilisateur d'une page Web. Elle varie selon l'appareil. Il est plus petit sur un téléphone que sur un ordinateur.
- L'élément <meta> viewport est à inclure dans toutes vos pages web:
`<meta name="viewport" content="width=device-width, initial-scale=1.0">`
- Il donne les instructions au navigateur sur la façon de contrôler les dimensions et mise à l'échelle de la page.

Les métadonnées (l'élément head)

Réglage de la fenêtre

- La partie `width=device-width` définit la largeur de la page selon la largeur de l'écran du dispositif (qui varie en fonction de l'appareil).
- La partie `initial-scale=1.0` définit le niveau de zoom initial lorsque la page est chargé par le navigateur.
- **Exemple :**
 - 1) Sans la balise meta viewport
 - 2) Avec la balise meta viewport



Encodage d'URL

URL - Uniform Resource Locator

- Une URL est utilisée pour traiter un document (ou d'autres données) sur le web.
 - C'est ce qu'utilise un navigateur Web demande des pages de serveurs Web.
 - Aussi ce qui est utilisé quand on clique sur un lien dans une page HTML, une balise <a> sous-jacente pointe vers une adresse sur le world wide web.
- Une URL peut être composée de mots, tels que "ugb.sn", ou l'adresse IP d'un protocole Internet (IP) : 192.68.20.50. La plupart des gens entre le nom du site lors de la navigation, car les noms sont plus faciles à mémoriser que des chiffres.

Encodage d'URL

Composition d'une URL:

- Une URL comme `http://www.uvs.sn/pats/index.html` suit ces règles de syntaxe:
uri://host.domain:port/path/filename Où :
- **uri** - définit le **type de service** Internet. Le type le plus commun est **http**
- **host** - définit l'**hôte du domaine** (l'hôte par défaut pour HTTP est **www**)
- **domaine** - définit le **nom de domaine** comme `ugb.sn`
- **Port** - définit le **numéro de port** sur l'hôte (le numéro de port par défaut pour HTTP est **80**)
- **path** - définit un **chemin d'accès** au serveur (S'il est omis, le document doit être stocké dans le répertoire racine du site web)
- **filename** - définit le nom d'un document / ressource

Encodage d'URL

Les uri usuels

- Le tableau ci-dessous répertorie certains uri usuels:

uri	Signification	Utilisé pour...
http	HyperText Transfer Protocol	Pages web ordinaires, commence par http://. Non chiffré
https	Secure Hypertext Transfer Protocol	Sécuriser les pages Web. Toutes les informations échangées sont cryptées
ftp	File Transfer Protocol	Pour uploader ou télécharger des fichiers à un site Web. Utile pour la maintenance de domaine
file		Un fichier sur votre ordinateur

Encodage d'URL

L'encodage

- Une URL ne peut être envoyée sur Internet qu'en utilisant le jeu de caractères ASCII .
- Comme les URL contiennent souvent des caractères hors du jeu ASCII, l'URL doit être converti dans un format ASCII valide.
- L'encodage d'URL convertit les caractères dans un format qui peut être transmis par Internet.
- Autrement dit, il remplace les caractères non ASCII avec un "%" suivi de deux chiffres hexadécimaux.
- Une URL ne peut pas contenir d'espace. L'encodage d'URL remplace normalement un espace avec un signe %20.
- **Exemple:** Sénégal est encodé par S%E9n%E9gal

Encodage d'URL

L'encodage de quelques caractères

- Le tableau ci-après montre l'encodage de quelques caractères usuels.
- Vous trouverez une liste complète des encodages des caractères à l'adresse suivante:

http://www.w3schools.com/tags/ref_urlencode.asp

Caractère	Encodage
€	80%
£	A3%
©	% A9
®	AE%
À	% C0
Á	% C1
Â	% C2
Ã	C3%
Ä	C4%