

# HTML- 01- Présentation et structure d'un document

## Introduction

### Qu'est-ce que le World Wide Web ?

Le World Wide Web (Web) est un réseau de ressources et d'informations. Le Web repose sur trois mécanismes pour rendre disponibles ces données le plus largement possible :

1. Un système de nommage uniforme pour leur localisation sur le Web (par exemple les URL) ;
2. Des protocoles, pour accéder à des ressources nommées dont on possède l'adresse (par exemple HTTP) ;
3. L'hypertexte, pour faciliter la navigation entre ces ressources (par exemple HTML).

### Qu'est-ce que le HTML ?

Publier de l'information en vue d'une distribution mondiale nécessite l'usage d'un langage universellement compréhensible, d'une sorte de « langue maternelle » acceptée par tous les ordinateurs. Le langage de publication utilisé sur le World Wide Web est le langage de balise hypertexte HTML (Hyper Text Markup Language). HTML donne aux auteurs les moyens de :

- Publier des documents en ligne possédant des titres, du texte, des tables, des listes, des photos, etc.
- Ramener des informations en ligne via des liens hypertextes avec un clic ;
- Concevoir des formulaires pour conduire des transactions avec des services distants, dans le but de rechercher des informations, de faire des réservations, de prendre des commandes, etc.
- Inclure des feuilles de calculs, des séquences vidéos ou sonores et d'autres applications, directement dans les documents.

C'est un langage de balise, qui a comme objectif de représenter de l'information. Ces balises sont interprétables par les différents navigateurs (Internet Explorer/Edge, Firefox, Safari, Opera etc. et permettent l'affichage de l'information désirée.

Historique du HTML :

- 1993 : HTML (1.0)
- 1995 : HTML 2.0
- 1997 : HTML 3.2 et 4.0
- 2000 : XHTML

HTML 01

Réalisation : Guillaume DELACROIX Formateur AFPA

06 janvier 2023

**Afpa**

- 2009 : HTML 5

HTML 01

Réalisation : Guillaume DELACROIX Formateur AFPA

06 janvier 2023

**Afpa** 

# HTML 5

Le HTML 5, publié en 2009, est la version toujours en cours du HTML.

Le HTML 5 remplace le XHTML et en reprend les règles de base. Les cinq premières règles seront mises en œuvre de suite. N'omettre aucune des balises qui définissent la structure en arbre du document (<html>, <body>...) :

1. Fermer toutes les balises
2. Les balises et les attributs sont toujours en minuscules
3. Les valeurs des attributs sont entre guillemets
4. Tous les attributs ont des valeurs explicites
5. Les balises indiquent la structure du document et non pas sa présentation.  
Les balises et les attributs de présentation sont remplacés par l'application de styles.

Une 6<sup>ème</sup> règle est l'exigence de l'utilisation des feuilles de styles (C.S.S.).

Comme à chaque évolution du langage HTML, le HTML 5 apporte des nouveautés et supprime certaines balises. Cette page (en anglais) liste par ordre alphabétique l'ensemble des balises HTML 5 en précisant les nouvelles et celles devenues obsolètes.

# Architecture HTML

## HTML : Un langage de balises sémantique

Le langage HTML est formé de balises (*tag* en anglais) assemblées afin de déterminer le plan (découpage) d'une page (le *document*).

Une balise possède un nom qui permet de décrire son contenu : on appelle cela un langage sémantique. Par exemple si on utilise la balise `<article>`, cela indique (suppose) que le texte qu'elle contient est un article.

Le nom des balises est inséré entre les signes `<` et `>`, lesquels sont appelés chevrons :

Exemples : `<p>`, `<div>`, `<article>`, `<meta>`...

La plupart des balises doit être ouverte puis fermée. Les balises de fermeture sont les mêmes que les balises d'ouverture, mais sont précédées d'un slash (`/`)

: `<p>texte</p>`, autres balises `</head>`.

Dans d'autres cas, les balises sont *autofermantes*, c'est-à-dire fermées par un simple chevron fermant sans répéter le nom de la balise :

`<br>`, `<hr>`, `<img>`...

En HTML, la casse (lettres minuscules ou majuscules), n'a aucune importance. La bonne pratique est toutefois d'écrire les balises en minuscule.

Une balise possède des attributs, ce sont des paramètres supplémentaires à appliquer spécifiquement à la balise. Un attribut a un nom, auquel on affecte une valeur (une chaîne entre guillemets) via l'opérateur `=` :

- `<p id="toto">` : balise paragraphe avec un seul attribut, l'attribut `id` (identifiant unique)
- `<p id="toto" class="rouge">` : balise paragraphe plusieurs attributs, `id` et `class`.

Selon la balise, il y a des attributs obligatoires ou facultatifs. Certains attributs peuvent recevoir une valeur librement définie par le développeur (comme les attributs `id` et `class` ci-dessus), d'autres doivent avoir une valeur proposée par les spécifications du HTML (W3C).

**Le nombre d'attributs peut donc être variable.**

Certaines balises peuvent être **imbriquées**, cela forme un **ensemble hiérarchique** appelé *balisage*, on a donc des balises *enfant* d'autres balises. La hiérarchie peut s'établir sur plusieurs niveaux.

Les balises imbriquées peuvent être multiples et différentes.

Exemple 1 : une liste d'options : les balises <li> (= list item) héritent de la balise <ul> (= unordered list) :

```
<ul>
```

```
<li>1er élément</li>
```

```
<li>2ème élément</li>
```

```
<li>3ème élément</li>
```

```
</ul>
```

Exemple 2 : des paragraphes et un <div> inclus dans un <div> :

```
<div>
```

```
<p>Texte du 1er paragraphe</p>
```

```
<p>Texte du 2ème paragraphe</p>
```

```
<div>Un autre div dans le div précédent qui contient <a href="page.html">un  
lien</a></div>
```

```
</div>
```

# Structure d'un document HTML

Un document HTML 5 se compose de plusieurs parties dont **l'ordre est à respecter**.

Une ligne de déclaration de type contenant les informations de la version HTML de la page : en HTML 5, cette déclaration doit être `<!DOCTYPE html>`

Attention : la casse de cette balise est à respecter.

- Un élément `<html>` qui comprend :
- Une section en-tête déclarative (enclose par la balise `<head>`)
- Un corps, qui emporte le contenu effectif du document. Le corps est généralement enclos par l'élément `<body>`

Chaque balise `<head>` ou `<body>` peut elle-même contenir des balises incluses.

Voici un exemple de document HTML 5 simple :

```
<!DOCTYPE html>

<html lang="fr">

<head>

<meta charset="utf-8">

<title>Ma première page en HTML 5</title>

</head>

<body>

<h1>Le HTML 5</h1>

<p>Le langage HTML, c'est fantastique.</p>

</body>

</html>
```

Attention : en HTML 5, toutes les balises ainsi que leurs attributs doivent être en minuscule (sauf DOCTYPE).

Dans la balise <html>, ne pas omettre l'attribut lang qui permet d'indiquer la langue dans laquelle la page web est rédigée aux robots de référencement ou aux logiciels de synthèse vocale.

# Entête du document : `<head></head>`

Le bloc `<head></head>` contient des informations sur le document courant, tels que son titre, des mots-clés que les moteurs de recherche peuvent exploiter et qui permettront d'indexer (référencer) votre site, et d'autres données qui ne sont pas considérées comme faisant partie du contenu du document : ce sont des métadonnées qui, pour leur grande majorité, seront contenues dans des balises `<meta>`.

Les balises de la partie `<head>` ne sont pas affichées par le navigateur à l'internaute, à l'exception de la balise `<title>` qui apparaît dans les onglets du navigateur et dans les moteurs de recherche et de la meta « description » reprise elle aussi par les moteurs de recherche (ci-dessous, copie d'écran de la page de résultats de Google) : Cependant, le contenu de la partie `<head>` est visible dans le code source (Ctrl+U dans un navigateur). Il ne doit donc pas contenir d'informations sensibles pouvant compromettre la sécurité de votre site.



# Les balises <meta>

## Principe

Les balises <meta> permettent de fournir ce que l'on appelle des metadonnées d'une page web, c'est-à-dire des indications supplémentaires différentes du contenu de la page.

## Evolution des balises <meta>

Certaines balises <meta> sont obligatoires, d'autres sont facultatives et certaines (par exemple keywords, refresh) ont désormais une utilité controversée ou obsolète.

## Obligatoire : la balise meta charset

La balise <meta charset> permet de spécifier aux navigateurs l'encodage de la page web. La valeur standard est l'**utf-8** :

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

Spécifier en UTF-8 évite de se retrouver avec des problèmes de caractères spéciaux ou d'accents mal affichés du genre aprÃ"s avoir Ã© lancÃ©.

Cette balise doit donc se trouver avant tout texte affiché à l'internaute, soit en pratique avant la balise <title> (et donc immédiatement après la balise d'ouverture <head>).

## Référencement

Le référencement fera l'objet d'une séance complète.

Aux débuts du web, l'une des principales utilisations des balises <meta> était l'indexation des sites web par les moteurs de recherche, à l'aide des attributs keywords et description. Mais beaucoup de webmasters renseignaient ces balises avec des mots-clés sans aucune pertinence par rapport au contenu réel de la page mais qui attireraient énormément de trafic. Désormais, l'hégémonie de Google (90% des recherches web en France) a imposé que seules les balises <title> (cf. ci-après) et <meta description> restaient utiles pour le référencement et toute velléité d'abus est sanctionnée par un bannissement des résultats de recherche (la page disparaît des moteurs).

Voici un exemple d'indexation commenté. A noter que ces balises doivent être placées entre les balises <head> et </head>, juste après le titre de votre page. Si votre site est en plusieurs langues, vous devez créer autant de balises méta que de langues.

Pour spécifier l'auteur du site :

```
<meta name="author" content="Dave Loper">
```

Pour décrire le site en une phrase de référence :

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

Une liste de mots-clés correspondant au mieux à votre site :

HTML 01

Réalisation : Guillaume DELACROIX Formateur AFPA

06 janvier 2023

**Afpa**

```
<meta name="keywords" content="Tapez ici les mots-clés séparés par une virgule">
```

Pour diriger les robots de référencement :

```
<meta name="robots" content="index, nofollow">
```

Ici, le robot indexe la page, mais n'indexe pas les pages qui ont un lien sur cette page... Les différentes valeurs possibles pour cet attribut sont :

## La balise <title>

Dans l'ordre, après les balises doctype, html, head et meta charset, on trouve la balise <title> dans laquelle vous devez indiquer le titre de votre document. Comme nous l'avons vu plus haut, la balise <title> est absolument essentielle pour le référencement, c'est celle qui est prise en compte en priorité par Google. Exemple

```
<title>Ma première page HTML</title>
```

## La balise <link>

L'élément HTML <link> définit la relation entre le document courant et une ressource ex-terne (CSS, icône). Cet élément peut être utilisé pour définir un lien vers une feuille de style ou un cadre de navigation (accéder à la même page dans une langue différente par exemple).

Cette balise est principalement utilisée pour lier les feuilles de styles CSS (que nous aborderons à la séquence suivante) à la page web :

```
`<link href="style.css" rel="stylesheet">`
```

Autre usage courant de la balise <link> : indiquer une icône de favori (« favicon ») pour un site :

```
`<link rel="icon" type="image/png" href="favicon.png">`
```

Contrairement aux balises vues jusqu'ici, il est possible d'avoir plusieurs <link> car il en faut une par ressource appelée.

Si vous souhaitez connaître les autres usages - spécifiés via l'attribut rel - de cette balise, vous pouvez consulter cette page.

# Le corps du document: `<body></body>`

Le contenu réel de la page (textes, images, tableaux etc.) visible par l'internaute se trouve dans le bloc `<body></body>` : c'est le corps du document.

Exemple

```
<body>
<h1>Mon contenu</h1>
<p>Le HTML 5 c'est fantastique.</p>
</body>
```

La partie `<body>` accueille à la fois :

- les éléments (balises) qui structurent la page (colonnes, blocs, paragraphes, menus, entête, pied de page etc.)
- les éléments relevant davantage du formatage (mise en exergue, gras, italique, citations, médias etc.).

Ces différentes balises seront vues au prochain chapitre.

## Commentaires

Vous pouvez commenter votre code source, l'internaute ne verra pas ce commentaire.

Exemple :

```
<!-- Ceci est un commentaire -->
```

Les commentaires seront visibles dans le code source HTML de votre page (afficher la page dans un navigateur puis touches `Ctrl + U`). Par conséquent, éviter trop d'indications pouvant compromettre la sécurité.

Les commentaires seront affichés dans le code source de la page. Par conséquent, évitez trop d'indications pouvant compromettre la sécurité.

Pour voir le code source d'une page, afficher la page dans un navigateur web puis touches `Ctrl + U` ou, si vous connaissez déjà, inspectez-le via les outils de développement des navigateurs (vus plus loin dans la formation).

# Exercice d'application

---

- A l'aide d'un éditeur de texte, saisissez le code HTML indiqué de la section *Architecture HTML > Structure d'un document HTML*.
- Enregistrez le fichier en lui donnant l'extension `.html`, puis visualisez-le à l'aide d'un navigateur web.
- Inclure les différentes balises vues dans ce document.