

La formation au numérique des élèves :

Le langage HTML

Internet : de l'opacité à la clarté

Jean-Alain Roddier

IA-IPR de mathématiques

Le langage HTML est un langage incontournable de notre quotidien. À la différence d'un langage qui pourrait être chanté comme une partition de musique ou qui pourrait être récité comme une belle poésie, le langage HTML est d'une discrétion remarquable. Cette discrétion ne saurait perdurer pour nos élèves et notre intention avec ce document est de faire « parler » ce langage plutôt opaque.

Gutenberg créateur d'une technologie

Johannes Gutenberg en inventant la typographie ne savait pas forcément qu'il avait inventé un système visant à faire intervenir un processus technologique dans le cadre de la transmission de l'information.



Portrait de Johannes Gensfleisch zur Laden zum Gutenberg

Tableau de 14^e siècle réalisé par un peintre inconnu. Source : site Wikipédia



Depuis plus de cinq siècles, l'imprimerie offre à tout écrivain la possibilité de diffuser son texte à grande échelle. Tout ce qui est réalisé à l'intérieur d'une imprimerie importe peu ; seul le résultat obtenu à la sortie importe. En d'autres termes, aussi bien l'écrivain que le lecteur du texte n'ont que faire de ce qu'il se passe concrètement entre les murs de l'imprimerie.

Ceci étant dit, rien ne vous empêche de pousser un jour les portes d'une imprimerie pour découvrir ce qu'il s'y passe.



Planche datée de 1568. Source : site Wikipédia

À l'identique de ce côté parfaitement utilitaire de l'imprimerie, toutes les tâches réalisées par « Internet » lorsque nous l'utilisons n'ont aucune importance à nos yeux. Nous nous plaçons systématiquement dans une position d'utilisateur : seules la vitesse à laquelle on nous livre l'information et la qualité de cette information sont jugées. En d'autres termes, nous attendons qu'Internet nous fournisse un produit fini sans jamais nous pencher sur sa façon d'agir.

Changeons aujourd'hui de paradigme, observons simplement ce qu'il se passe lorsque nous ouvrons – par exemple – la page d'accueil de notre site académique.

La capture d'écran ci-contre montre une page de notre site académique qui porte – comme fait exprès – sur l'éducation à la citoyenneté mondiale. Cette page Internet a été « co-construite » à partir d'une réflexion poussée sur les différentes composantes qui doivent en assurer une parfaite lisibilité afin de transmettre notre message sur la mobilisation forte de notre académie autour des valeurs de la République.



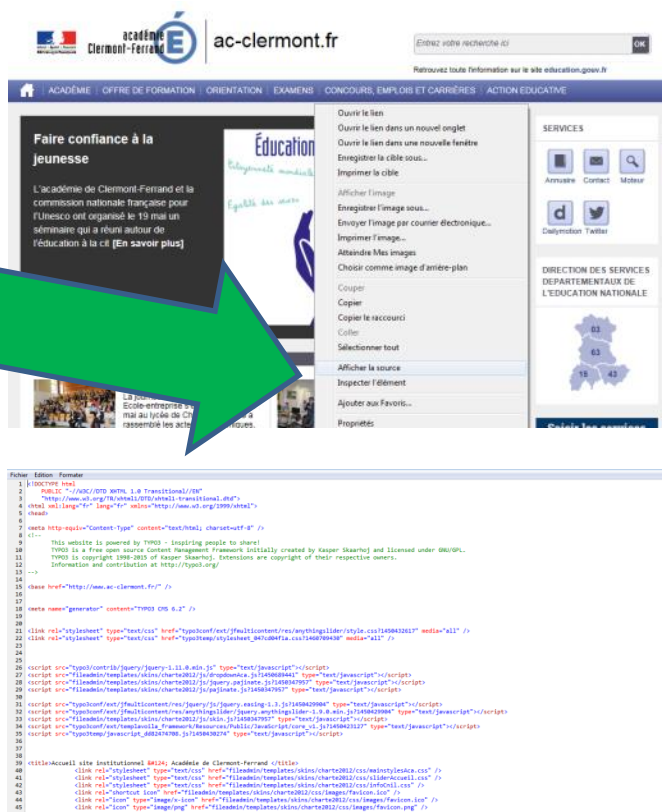
Construire une page à diffuser sur Internet, c'est avant tout produire **une réflexion sur le contenu de cette page**. Vient ensuite la construction technique de cette même page avec l'utilisation incontournable du langage HTML que nous allons découvrir à présent.

Clair-opaque-clair

Internet est construit sur un ensemble de pages dites HTML du fait du langage qu'elles utilisent (le langage HTML). Avec Internet, on assiste ainsi à la transmission d'une information d'une personne à d'autres personnes à l'identique de l'imprimerie. Cette information est claire au départ et claire à l'arrivée ; elle est en revanche transmise avec une phase de transmission complètement opaque. (à l'identique là-aussi de l'imprimerie). C'est en résumé un trio du style : **clair-opaque-clair** dont deux simples clics vont nous permettre de lever le voile sur ce côté opaque.

En effectuant un « clic droit » sur notre page Web, on obtient un menu déroulant dans lequel nous prenons l'option **Afficher la Source**.

Après cette facilité d'accès à cette page, nous sommes contents d'obtenir ce code informatique en langage HTML, nous sommes en revanche vraiment déçus de constater que ce langage est parfaitement inaccessible. Nous reconnaissons bien en effet dans cette page de code quelques mots en anglais (que nous arrivons à traduire) mais comment les choses s'agencent-elles pour fournir notre belle page d'accueil de notre site académique, cela relève - pour l'instant - du mystère.



Afin de lever ce mystère, nous avons sollicité des auteurs afin d'une part de nous apporter un éclairage général sur le langage HTML et d'autre part de nous donner des exemples simples et précis que les élèves pourront mettre en œuvre et modifier à souhait.

Monsieur Massih-Reza Amini professeur des universités nous livre ainsi un texte remarquable car parfaitement explicatif des bases du langage HTML. Ce texte nous permet ainsi de lever cette opacité sur le langage HTML.

Monsieur Sébastien Hamon professeur agrégé de mathématiques au lycée Théodore de Banville fournit une construction balisée d'une page HTML en sept étapes. Ce cheminement porteur pour nos élèves leur permettra de se placer dans un registre différent de celui d'être uniquement utilisateurs d'internet ; en ce sens, ce premier pas vers la création d'une page web leur sera utile.

Nous remercions vivement ces experts qui œuvrent ici de façon magistrale à la formation de nos élèves sur cet outil majeur du numérique qui ne reste cependant qu'un outil ; il ne faudrait pas laisser penser – en effet – à nos élèves que ce langage HTML est inscrit dans le marbre.

Introduction à la Toile et au langage HTML

Massih-Reza Amini

Professeur des Universités

Université Grenoble Alpes

Directeur du groupe de recherche

dAta analysis, Modeling and mAchine learning (AMA)

du Laboratoire Informatique de Grenoble (LIG)

Les problèmes principaux liés au développement de la Toile (ou *Web* en anglais), sont (a) la création, la modification et la suppression spontanées d'informations sans aucune coordination au préalable et (b) le grand nombre de données existantes.

L'évolution du *Web* est principalement due à trois inventions centrales :

1. le format standardisé pour représenter les données avec le langage de balisage HTML (*Hypertext Markup Language* en anglais) ;
2. le protocole de communication client-serveur pour le transfert hypertexte, HTTP (*HyperText Transfer Protocol* en anglais) ;
3. . les adresses *Web* qui sont à la base des liens hypertextes et qui permettent d'identifier les données sur la Toile.

Ces inventions ont permis le développement de nouvelles technologies à deux niveaux distincts : collecte et indexation des données, et conception de nouvelles stratégies qui permettent d'améliorer l'efficacité de la recherche. Dans ce document, nous allons donner les grandes lignes de la recherche sur le Web ainsi qu'une description succincte du langage HTML.

1 Architecture de la Toile

L'architecture de la Toile repose sur deux éléments fondamentaux, appelés, d'une part, *interaction et recherche*, et d'autre part, *collecte et indexation* dans la figure 1. En amont de toute recherche, un robot d'indexation (*web crawler* ou *spider* en anglais) récupère les pages de la Toile suivant différentes stratégies, et les transfère à un sous-système d'indexation qui les analyse et les stocke dans les fichiers d'index [1]. Dans la phase de recherche, un utilisateur exprime son besoin d'information via l'interface d'un navigateur, qui est généralement constituée d'une page *Web* avec une zone de saisie. Le navigateur analyse ensuite la demande dans une forme que son moteur de recherche associé peut comprendre et lui envoie une requête. Le moteur de recherche exécute l'opération de recherche sur les fichiers d'index construits dans la phase précédente et trie les données jugées pertinentes par rapport à la demande initiale qui sont au final montrées à l'utilisateur via une autre interface du navigateur (on peut citer le programme Lucene <http://lucene.apache.org/> issu d'un projet *open source* qui permet d'indexer et de rechercher du texte suivant l'architecture de la figure 1).

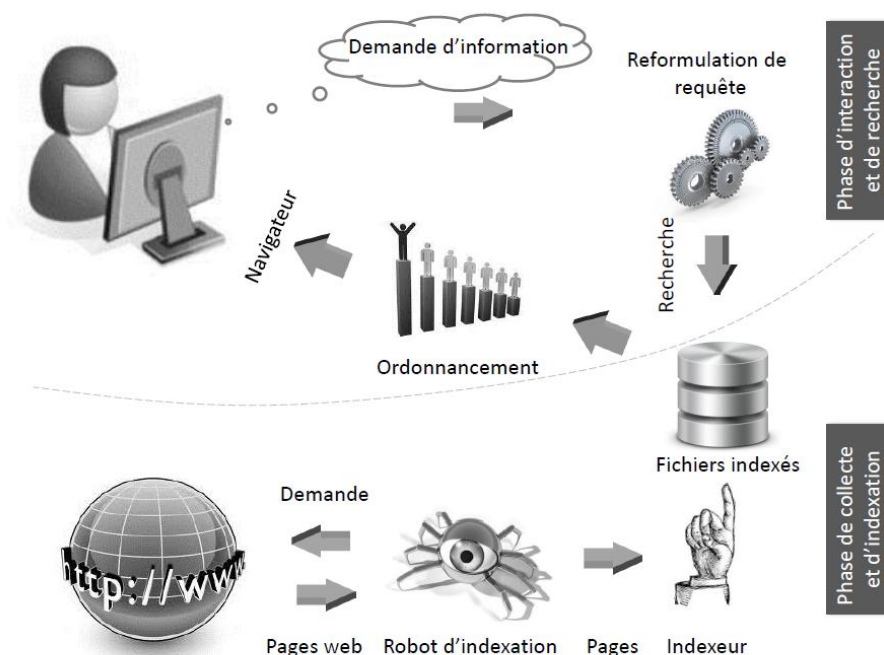


Figure 1: Différentes étapes de la recherche sur la Toile avec les deux éléments (a) de collecte et d'indexation, et (b) d'interaction et de recherche, séparés par des pointillés.

2 Langage HTML

Le développement du *Web* est en grande partie dû à la mise en place d'un environnement client-serveur simple et accessible à tous. Dans cet environnement, un client, généralement un navigateur, interagit avec un serveur par l'intermédiaire d'une interface graphique simple. Un serveur communique ensuite avec un client de manière asynchrone via le protocole HTTP qui permet de transporter différents types d'informations (textes, images, des fichiers audio et vidéo) codées dans le langage à balises, HTML, ou *HyperText Markup Language*. Ce langage est le principal langage à balises utilisé pour afficher des pages *Web* et d'autres informations fournies par un serveur dans un navigateur.

- le terme *Hypertext* dans HTML, fait référence à la manière dont les pages *Web* sont reliées entre elles sur la Toile;
- le terme *Markup Language* définit comment les documents sont formés.

Ce langage permet ainsi de concevoir des documents structurés en désignant la structure des textes avec des balises, ou *tags* entre crochets, qui sont interprétées par un navigateur, mais non affichées dans l'interface de ce dernier.

De façon plus spécifique, un document HTML commence et se termine avec des balises `<html>` et `</html>`. Ces balises indiquent à un navigateur que l'ensemble d'un document donné est écrit suivant le standard HTML.

À l'intérieur de ces deux balises, le document est divisé en deux parties:

- La première partie, ou l'en-tête, est délimitée par des balises `<head>` ... `</head>` et contient des informations générales comme le titre du document, l'auteur de l'information, etc. qui ne sont pas affichées par un navigateur.

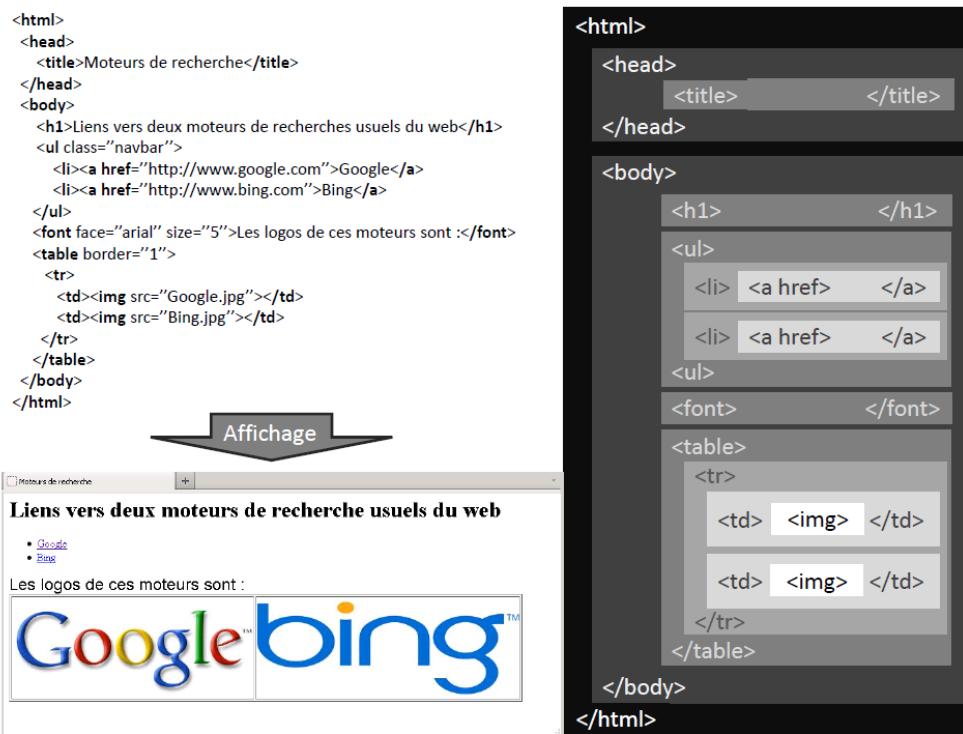


Figure 2 : Exemple d'un document HTML (en haut à gauche), de sa structure interne (à droite) et de la page interprétée et affichée par un navigateur (en bas à gauche). Dans la structure du document, la portée de chaque élément est montrée par la couleur du rectangle le contenant.

- La seconde partie, délimitée par les balises **<body> ... </ body>**, contient le corps réel de ce document qui sera visible aux utilisateurs. Les balises HTML sont toutes désignées par des noms. Les plus courantes viennent en paires, bien que certains tags connus comme des éléments vides ne sont pas appariés. La première balise dans une paire est la *balise ouvrante* et la seconde est appelée *balise fermante*. Entre ces balises, on met généralement du texte, des tags, des commentaires ou d'autres types de contenu textuel. Selon les standards HTML, les noms des balises ne sont pas sensibles à la casse et une propriété commune à tous les navigateurs est que les éléments mal orthographiés, ou dont le nom n'existe pas dans l'ensemble des noms des éléments existants, sont ignorés. Il n'y aura donc pas d'erreurs d'affichage.

Hormis les noms, les *attributs* constituent l'autre partie importante des balises HTML. Un attribut est utilisé pour définir les caractéristiques d'un élément et il est placé à l'intérieur de la balise ouvrante de cet élément. Tous les attributs sont constitués de deux parties :

- le nom de la propriété associée à un élément qu'on souhaite utiliser ;
- la valeur de cette propriété.

Par exemple si on souhaite que les textes d'une page soient de couleur bleu et écrits avec la police Arial, on place ces textes entre les balises ** ... ** en spécifiant les propriétés recherchées dans la balise ouvrante, par exemple ****.

Dans la version 4 de HTML, 91 éléments déclarés sont classés soit au niveau bloc, soit au niveau texte :

- Les éléments de niveau bloc sont des superstructures qui contiennent d'autres blocs et éléments de niveau texte. Ils sont généralement séparés des autres blocs par des marges verticales.
- Les éléments de niveau texte représentent des textes relativement courts et ne peuvent contenir que d'autres éléments de niveau texte.

Parmi tous les éléments HTML, l'élément hyperlien joue un rôle central.

En effet, cet élément permet de pointer la page *Web* contenant la balise `<a href="adresse web"> ... </a>`, vers la page dont l'adresse est spécifiée entre les guillemets de la balise ouvrante de cet élément. Entre les deux balises ouvrante et fermante, il y a aussi une zone soulignée, appelée zone *d'ancrage*, qui permet d'inclure du texte ou de l'image et qui définit le lien ainsi créé dans le document HTML initial². Dans la version 5 de HTML, il a été inclus une couche application avec de nombreuses interfaces de programmation (ou *Application Programming Interface* - API) et un algorithme de transfert conçu pour traiter les documents de la Toile à la syntaxe non conforme. La figure 2 montre l'exemple d'un document HTML ainsi que sa structure interne et la page interprétée et affichée par un navigateur.

Références

[1] Massih-Reza Amini and Éric Gaussier *Recherche d'Information - applications, modèles et algorithmes*. Eyrolles, 2013.

[2] L'index de toutes les balises et leur compatibilité sont répertoriés à l'adresse <http://meiert.com/en/indices/html-elements/> et il existe une multitude de cours et de tutoriaux sur le langage HTML en libre accès sur la Toile (voir par exemple <http://j-willette.developpez.com/tutoriels/html/les-bases-du-html/>).

Ma page web « en deux temps trois mouvements » et 7 étapes.

Sébastien HAMON

Professeur agrégé de mathématiques au lycée Théodore de Banville

Nous vous offrons ici la possibilité de créer votre propre page web en suivant ce chemin balisé qui - nous l'espérons- vous amènera à bon port.

I) Première étape : Première page

- 1) Ouvrir le bloc note, écrire : « C'est la première page ! ».
- 2) Sélectionner enregistrer sous, entrer comme nom de fichier « page1.html », valider.
- 3) À présent double-cliquer sur ce fichier : il s'ouvre dans votre navigateur !

Un fichier HTML (***Hypertext Markup Language***) est donc un fichier texte, mais à l'aide de balises on va aller beaucoup plus loin. Avant de passer à la suite, deux suggestions : utiliser [Firefox](#) comme navigateur et [Notepad](#) comme éditeur.

II) Deuxième étape : Liens hypertexte

- 1) Ajouter à votre fichier `<a href="http://ac-clermont.fr"> Site de l'académie </a>`
- 2) Enregistrez-le puis ouvrez-le avec le navigateur.

Entrer les balises `<a>` et `</a>` se trouvent les mots sur lesquels on pourra cliquer, on a ainsi un texte dit « cliquable ».

À l'intérieur, après le paramètre href, se trouve l'adresse du lien.

D'autres paramètres existent, par exemple le paramètre target désignera la cible du lien.

Créer puis tester un fichier page2.html avec comme code :

```
<a href='page1.html' target='_blank'> Cliquer ici </a>
```

III) Troisième étape : **Balise de texte**

Essayer de copier/coller ce morceau de code, tester, modifier, tester

```
<h1>Titre 1</h1>
<h2>Titre aligné à gauche</h2>
<h3 align="right">Titre aligné à droite</h3>
<p align="center">Paragraphe centré</p>
<p>Mon texte <font size="5" face="georgia" color="red">en plus gros, en rouge et en georgia
ici</font> et puis c'est tout. </p>
```

Les balises h1, h2, h3, ... correspondent à des titres de plus en plus petits.

La balise <p>...</p> est utilisée pour les paragraphes.

Remarque : le langage HTML ne tient pas compte de l'indentation et des passages à la ligne dans votre code. Ce problème est en partie résolu avec la balise
 qui permet de passer à la ligne.

IV) Quatrième étape : **Image**

Prendre une image extraite de votre dossier image (une belle photo par exemple).

La sauvegarder sous le nom `essai.gif` si elle est au format GIF (pour d'autres formats prendre le nom correspondant : `essai.jpeg`, `essai.jpg` ou `essai.png`).

La déposer au même endroit que votre fichier, puis taper :

```

```

Après vient le temps des combinaisons, cette nouvelle balise :

- lorsqu'elle est placée entre <a> et devient « cliquable » ;
- lorsqu'elle est placée entre <p align="center"> et </p>, l'image apparaîtra de façon centrée ;
-

V) Cinquième étape : **Structure de page**

À cette étape, notre code fonctionne mais n'est pas conforme aux standards W3C. En clair, un professionnel du Web serait scandalisé.

En effet nous n'avons créé que la partie qui sera affichée, elle correspond au corps de la page (body).

Mais il existe une autre partie, située entre les balises <head>, où se trouvent des informations non affichées mais utiles au navigateur (des métadonnées : des indications pour les moteurs de recherche, le nom de l'auteur, le codage, ...), des liens vers des fichiers css, javascript, le titre, ...

```
<html>
  <head>
    <title>Titre de la page</title>
    <link rel="stylesheet" href="style.css" />
    <meta name="keywords" content="HTML,simple">
    <meta name="author" content="Hamon S">

  </head>
  <body>
    Texte de la page ici.
  </body>
</html>
```

Remarque : la ligne <link rel="stylesheet" href="style.css" /> indique au navigateur de chercher (au même endroit que la page html) un fichier appelé `style.css` et d'intégrer au code

de la page html les informations qui s'y trouve. Nous allons à présent apprendre à créer ce type de fichier.

VI) Sixième étape : **feuille de style CSS**

Créer un nouveau fichier appelé style.css et y taper :

```
body{ background-color : yellow;
      color : blue; }
```

Relancer votre fichier HTML : La couleur du fond et de la police a changé.

L'avantage de mettre le style dans un fichier séparé est de pouvoir utiliser ce fichier sur plusieurs pages. Ainsi si vous changez un élément du style, cela se fait automatiquement pour toutes les pages.

VII) Septième étape : **Création d'un site avec un menu**

L'idée est de créer en HTML un conteneur (balise <div>) dans lequel on insère une liste (balise et) pour le titre. Ensuite dans le fichier CSS, on donne un style au conteneur : il occupera 20 % de la largeur à droite.

<pre><html> <head> <link rel="stylesheet" href="style.css" type="text/css"> </head> <body> <h1> Ici c'est le Titre ! </h1>
 <div id='menu'> page 1 page 2 </div> texte page 1
 </body> </html></pre>	<pre>body { text-align: center; background-color : cyan; color : purple; } #menu{ float: left; width: 20%; background-color : pink; } div#menu a:hover{ color: white;} div#menu a{text-decoration: none; }</pre>
--	--

Pour plus d'informations :

Un site tous simple reprenant en détail tous ceci : <http://formation.upyupy.fr/>

Un site pour apprendre à coder en ligne : <https://www.codecademy.com/fr/learn/web>

Un site pour approfondir : <https://openclassrooms.com/courses/apprenez-a-creeer-votre-site-web-avec-html5-et-css3>