

Ma page ouèbe en une leçon

Exposé au Séminaire Étudiant du LSP

Djalil CHAFAÏ

<http://www.lsp.ups-tlse.fr/Chafai/>

Mai 1999

Ce texte est destiné aux doctorant(e)s du Laboratoire de Statistique et Probabilités de Toulouse. Seuls les rudiments sont abordés, les choses sont beaucoup plus complexes bien évidemment...

Des histoires, toujours des histoires...

Internet existe depuis les années soixante-dix. Né des lubies des militaires puis développé pour les besoins des scientifiques, il véhiculait alors essentiellement le courrier électronique, des fichiers et les messages des forums de discussion. Le web¹ tel que nous le connaissons aujourd'hui n'est apparu que beaucoup plus tard. Il est un peu inspiré d'un système analogue mais moins puissant appelé Gopher qui a aujourd'hui quasiment disparu.

Le web a été imaginé en 1989 par Tim BERNERS-LEE, un physicien du CERN² qui a pensé à mettre au point un système de pages mêlant texte mis en forme, images et liens pointant vers des fichiers de toute sorte ou menant à d'autres pages du même type situées sur d'autres machines (d'où le mot Web). Son but était alors de faciliter l'échange d'information entre physiciens nucléaires et plus généralement entre scientifiques. La première démonstration publique eu lieu en 1991. Le web a largement dépassé ces objectifs et a entraîné par sa facilité d'utilisation et son côté attrayant la démocratisation de l'Internet que nous connaissons aujourd'hui.

HTTP, HTML, URL...

Les pages web sont écrites en un langage de mise en forme appelé HTML (Hyper Text Markup Language). La récupération des pages sur les serveurs web³ se fait par un protocole de transmission appelé HTTP (Hyper Text Transfert Protocol). L'ensemble des pages écrites en HTML disponibles via HTTP sur les serveurs web reliés à l'Internet s'appelle le World Wide Web. En réalité, HTTP permet d'échanger des documents de types très variés (images, fichiers de toutes sortes...). Le web est très convivial car il permet de présenter l'information de façon agréable et de sauter d'une page à une autre à travers le réseau mondial avec une facilité déconcertante.

Un document sur le réseau mondial est référencé par son URL (Universal Resource Locator). Un URL a la forme suivante [protocole://nom.du.serveur/chemin/document](http://nom.du.serveur/chemin/document). Voici un exemple d'URL :

<http://www.lsp.ups-tlse.fr/Chafai/SE/index.html>

La partie <http://> correspond au protocole⁴ utilisé pour obtenir le fichier en question. La partie

¹World Wide Web en anglais, ce qui signifie « toile d'araignée mondiale »

²Centre Européen de Recherches Nucléaires à Genève. En 1994, le CERN et le MIT signèrent un protocole d'accord pour la promotion et la standardisation du Web et créèrent le World Wide Web Consortium (W3C), voir <http://www.w3c.net/>. Le W3C est actuellement administré par le MIT et l'INRIA.

³Un serveur web est un logiciel répondant aux requêtes HTTP. Par abus de langage, on désigne également par ce nom la machine faisant tourner ce type de logiciel. Le logiciel libre Apache <http://www.apache.org/> est sans doute le serveur web le plus répandu à l'heure actuelle.

⁴D'autres protocoles existent bien sûr. Par exemple le FTP (File Transfert Protocol), qui permet d'échanger des fichiers de façon plus performante que HTTP. Cela dit, HTTP a d'autres avantages, mais c'est une autre histoire.

`www.lsp.ups-tlse.fr` n'est rien d'autre que le nom du serveur à contacter en utilisant le protocole HTTP. Quant à la partie `Chafai/SE/index.html`, elle constitue le chemin d'accès au document dans l'arborescence du serveur. Les parties protocole et nom de serveur ne sont pas sensibles aux majuscules et minuscules. Lorsque vous lancez Netscape⁵ et que vous lui demandez l'URL `http://www.lsp.ups-tlse.fr/Chafai/SE/index.html`, Netscape interroge le serveur `www.lsp.ups-tlse.fr` en HTTP et lui demande de lui envoyer le document `/Chafai/SE/index.html`. Ce document contient en l'occurrence un texte rédigé en HTML comme son nom l'indique. C'est ce que l'on appelle une page web. Une fois la page reçue, Netscape interprète le code HTML, récupère en HTTP les éventuelles images qui y sont référencées puis affiche la page correctement formatée.

En fait, le `http://` peut être omis avec Netscape car c'est ce qu'il fait par défaut, en bon navigateur web ! Quant au `www` dans le nom du serveur `www.machine.fr` signifie juste que ce serveur répond aux requêtes HTTP. De la même manière que le serveur nommé `mail.cict.fr` est responsable du courrier électronique. Ce `www` est indicatif, tous les noms de serveurs web ne commencent pas comme ça.

HTTP est « la langue » (protocole, on dit) qui permet aux navigateurs web comme Netscape de communiquer avec les serveurs web pour échanger des fichiers au format HTML, des images et plus si affinités. Vous n'avez pas à apprendre HTTP, ça ne vous concerne pas. C'est un jargon entre programmes. Par contre, **le langage HTML dans lequel les pages sont rédigées vous concerne.**

Récapitulons. Pour faire votre page web, il vous faut donc savoir **comment mettre en forme un texte en HTML**, ça vous rappellera \LaTeX , puis **trouver un serveur sur lequel mettre votre page web** : au CICT, avec le reste du labo, ça vous va ?

L'arborescence du serveur ouèbe

Sur le serveur, votre page web principale est un document HTML stocké dans un fichier nommé `index.html`. Il faudra le mettre dans un répertoire particulier (votre répertoire web) pour que le serveur comprenne qu'il s'agit de votre page web. Vous pourrez ensuite créer d'autres fichiers HTML, ajouter des images, des fichiers PostScript etc dans votre répertoire web et faire des liens entre eux.

Sur un serveur web, toutes les pages, les images et plus généralement les fichiers qui seront disponibles aux navigateurs en HTTP sont stockés dans un répertoire spécial (racine du serveur web). Ce répertoire possède des sous répertoires et ainsi de suite. Votre répertoire web fait partie du lot. Le chemin dans un URL est donné par rapport à ce répertoire racine de l'arborescence de serveur web. Il ne faut pas confondre cette racine avec celle de la machine elle même. La racine du serveur ouèbe `www.serveur.fr` peut très bien être `/var/www/` dans l'arborescence de la machine, ce qui fait que le document d'URL `http://www.serveur.fr/images/trombine.gif` correspondra en réalité au fichier `/var/www/images/trombine.gif` sur la machine faisant tourner le serveur web. **Ne confondez donc pas le répertoire racine du serveur web et le répertoire racine de la machine.**

Je veux apprendre l'HTML !

Il vous fera penser à \LaTeX . Il permet de mettre en forme un texte. C'est à dire de spécifier la forme et la taille des caractères, l'agencement du texte en paragraphes, l'inclusion d'images et de liens vers des URL. Il permet aussi tout plein d'autres choses (sons, animations...) car il en est à sa quatrième version.

Tout comme en \LaTeX , le HTML consiste en des séquences de caractères que l'on ajoute au texte et qui permettent de spécifier la façon de l'afficher. Voici quelques exemples :

```
<b>ceci est un petit texte qui sera affiché en gras</b>
<i>celui-ci sera affiché en italique</i>
```

⁵Netscape est loin d'être le seul navigateur ouèbe. Citons par exemple W3m, Lynx, Konqueror, Mozilla, Galeon, Amaya, Internet Explorer... On dit navigateur ou fureteur en français.

Comme vous le voyez, on a mis le texte entre deux balises (l'une ouvrante, l'autre fermante) `` pour le gras et `<i></i>` pour l'italique. La balise `<p></p>` permet de délimiter un paragraphe. Des tas d'autres balises existent. Certaines ne sont pas constituées d'un couple. Par exemple pour insérer une image au format GIF⁶ on fera

```

```

Le chemin en question pourra être soit relatif par rapport au répertoire dans le quel le fichier HTML contenant la balise se trouve soit absolu par rapport à la racine du serveur web (donc commençant par un /). Supposons par exemple que la racine du serveur web soit `/var/www/` et que votre répertoire web est `/var/www/me/`. Si le fichier `/var/www/me/index.html` contient la balise ``, le serveur pensera que l'image se trouve en `/var/www/me/images/nom.gif` alors que si la balise est ``, il pensera que l'image se trouve en `/var/www/images/nom.gif`

On peut aussi faire insérer une image située sur un autre serveur en utilisant son URL

```

```

De même, pour faire un lien vers un autre document, on fera

```
<a href="chemin/nom.du.doc">texte à cliquer</a>
```

Pour faire un lien vers un document situé sur un autre serveur, on utilisera son URL

```
<a href="http://nom.du.serveur/chemin/nom.du.doc">texte à cliquer</a>
```

Des commentaires peuvent être insérés comme suit

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

Bon, vous commencez à comprendre n'est-ce pas! Voici un exemple complet de fichier HTML. La première ligne est un commentaire spécial qui permet d'annoncer le standard de conformité du document, ici « HTML 4.0 transitionnal ».

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.0 Transitional//EN">
```

```
<!-- La balise qui suit dit que le code HTML commence -->
```

```
<html lang="fr">
```

```
<!-- La balise head permet de donner le titre, etc. -->
```

```
<head>
```

```
<!-- Le titre de la page -->
```

```
<title>Un exemple de page</title>
```

```
<!-- Le nom de l'auteur de la page -->
```

```
<meta name="author" content="Jojo Lapin, jojo@lapin.com ">
```

```
<!-- La langue du document -->
```

```
<meta HTTP-EQUIV="content-Language" content="fr">
```

```
<!-- Description du document -->
```

```
<meta name="Description" lang="fr" content="Exemple de page">
```

```
<!-- Mots clés pour les moteurs de recherche -->
```

```
<meta name="Keywords" lang="fr" content="exemple, www, html">
```

```
<!-- Le copyright -->
```

```
<meta name="Copyright" lang="en" content="Jojo Lapin, 1999">
```

```
<!-- Directives pour les robots de recherche -->
```

```
<meta name="Robot" lang="en" content="index,follow">
```

```
<!-- On dit que l'on utilise les accents directement -->
```

⁶GIF (Graphic Image Format) est un format d'image compressé conservatif très répandu basé sur une palette de 256 couleurs mis au point par CompuServe. Il permet la transparence, l'entrelacement et de petites animations. D'autres formats d'images sont utilisés sur le web, comme par exemple le format compressé non conservatif JPEG (Joint Photographic Experts Group) mieux adapté aux images en milliers de couleurs ou encore le plus récent PNG (Portable Network Graphic).

```
<meta HTTP-EQUIV="content-Type" content="text/html; charset=iso-8859-1">
</head>
```

```
<!-- La balise body marque le debut du texte de la page proprement dit -->
<!-- Le fond sera blanc, pour mettre une image de fond, -->
<!-- utilisez plutôt background="chemin/image" -->
<body bgcolor="white">
```

```
<h1>Ceci est un gros titre</h1>
<h4>Ceci est un sous-titre</h4>
<h5>Ceci est un sous-sous-titre</h5>
```

```
<p>
Voici un paragraphe contenant une petite image

ainsi qu'un lien vers une
<a href="http://nom.du.serveur/chemin/nomdufichier.html">autre page</a>.
<b>Ceci est un petit texte qui sera affiché en gras</b> quant à
<i>celui-ci, il sera affiché en italique</i>.
<!-- La balise br permet de forcer un retour à la ligne -->
<br>
Voici un lien qui, lorsque l'on clique dessus dans Netscape, provoque le
lancement d'une fenêtre qui permet d'envoyer un courrier électronique
<a href="mailto:nom@machine.fr">clicker ici pour m'écrire...</a>.
</p>
```

```
<p>
Voici un lien vers un fichier PostScript stocké dans le même répertoire que ce
fichier HTML :
<a href="nomdufichier.ps">ma première publi</a>
</p>
```

```
<p>
Voici un lien vers un fichier PostScript stocké dans le répertoire /publis
d'un serveur web distant :
<a href="http://nom.du.serveur/publis/nomdufichier.ps">la publi d'un pote</a>
</p>
```

```
<p>
<font color="green" size="4">Ce texte apparaîtra en vert et en gros</font>.
<center>Ceci est un texte centré</center>
Voici un petit tableau 2x2
<table>
<tr> <!-- Ligne 1 -->
  <td>case 1,1</td> <td>case 1,2</td>
</tr>
<tr> <!-- Ligne 2 -->
  <td>case 2,1</td> <td>case 2,2</td>
</tr>
</table>
</p>
```

```

<p>Voici une liste ordonnée</p>
<ol>
  <li>Premier élément</li>
  <li>Deuxième élément</li>
</ol>
<p>et voici une liste non ordonnée</p>
<ul>
  <li>un élément</li>
  <li>un autre élément</li>
</ul>

<!-- Fin du texte de la page -->
</body>
<!-- Fin du code html -->
</html>
<!-- FIN DE L'EXEMPLE -->

```

Les noms des balises HTML sont inspirés de la langue anglaise : **b**=« bold », **h1**=« header 1 », **br**=« break », **ul**=« unordered list », **ol**=« ordered list », **li**=« list item », **a**=« anchor », **tr**=« table row », **td**=« table data »... Les balises HTML peuvent être écrites en mêlant majuscules et minuscules, mais dans les derniers standards (XHTML), seules les minuscules sont acceptées. De plus, les espaces supplémentaires et les retours à la ligne ne sont pas pris en compte par défaut. Pour insérer un espace de force, il faudra utiliser ` ` (pour « non breakable space »). Pour insérer un retour à la ligne de force, il faudra utiliser `
`. Comme vous l'avez sans doute remarqué, certaines balises acceptent des paramètres de réglages. Pour obtenir par exemple un paragraphe centré, on pourra faire `<p align="center"></p>`. En fait, on peut séparer le rendu du document de son contenu en utilisant des feuilles de style (CSS), dont je ne vais pas parler ici. Enfin, la directive `Content-Type` dans l'entête du fichier HTML permet de taper les accents directement, tel quels. C'est très pratique et beaucoup plus lisible que les ignobles `´` que l'on devrait utiliser sinon.

Je vous renvoie à la grande quantité de guides et autres références sur le langage HTML. Vous y trouverez beaucoup d'autres balises et des tas d'astuces amusantes. Une bonne référence pour commencer est par exemple « Un Nouveau Guide Internet » à l'URL <http://www.imagnet.fr/ime/fr\protect\T1\textunderscoreungi2.htm>.

Beurk, c'est trop compliqué !

Faire du HTML sans le (sa)voir, voilà ce qu'il faudrait. Pour les JOURDAIN du web qui trouvent que taper du HTML sous Emacs ne fait pas partie des tâches humainement envisageables, je tiens à signaler l'existence d'un éditeur HTML intégré à Netscape 4 qui permet de fabriquer des pages sans se soucier du HTML (cf. menu File). Vous travaillez directement sur la page telle qu'elle apparaîtra⁷ (ou presque). Le code HTML étant généré automatiquement par derrière. Pour faire un lien, il suffit de cliquer sur une icône qui insère un lien, idem pour les images et le reste.

Cette méthode est agréable mais limitée. De plus, le code HTML généré automatiquement n'est pas forcément très lisible par un humain, mais à quoi bon après tout... Cela dit, le jour où l'éditeur montrera ses limites, vous aurez beaucoup de mal à modifier à la main le code HTML ainsi généré.

Faites vous une idée vous même et choisissez votre méthode. L'éditeur en question est accessible via le menu File des dernières versions de Netscape (version 4). D'autres éditeurs du même type mais plus puissants existent mais ne sont pas disponibles au CICT à ma connaissance.

Philippe CARMONA a fait des canevas de pages, avec différentes possibilités. Vous pouvez les copier chez vous et les modifier et adapter à loisir.

⁷On dit alors qu'il s'agit d'un éditeur WYSIWYG (What You See Is What You Get).

D'accord, mais je le met où mon fichier HTML ?

Il vous faut demander à Philippe CARMONA de vous créer un répertoire qui accueillera vos fichiers HTML. Ce sera votre répertoire web. Le fichier principal, celui qui contiendra votre page de garde HTML, devra s'appeler `index.html`.

N'oubliez pas que les documents situés dans votre répertoire web sont destinés à être lus par le serveur web du CICT qui le transmettra en HTTP aux navigateurs des internautes. Il faut donc que les droits d'accès soient correctement positionnés. Pour ce faire, placez vous **dans votre répertoire web** et exécutez la commande UNIX suivante⁸

```
chmod ugo+r *
```

Comment piquer des images et des idées dans d'autres pages

Lorsque vous naviguez (surfez, furetez. . .) sur la toile, il vous arrive de tomber sur des images que vous aimeriez bien récupérer pour vos propres pages. Rien n'est plus simple avec Netscape, cliquez sur le bouton droit de la souris sur l'image désirée et choisissez « Save this image as. . . ». Netscape vous demandera alors sous quel nom et à quel endroit vous voulez la sauvegarder. Il ne vous reste plus ensuite qu'à l'inclure dans vos propres pages ! Gâre cependant aux droits d'auteurs. Mais qui viendra embêter un petit thésard, hein ? Je vous le demande. Une parade consiste à faire référence à l'image en question dans votre page sans la copier chez vous :

```

```

Dans le même ordre d'idées, si vous trouvez une page bien agencée et que vous voulez comprendre comment faire pareil, vous pouvez visualiser le code HTML de la page. Dans Netscape 4, il suffit d'utiliser l'option de menu « Document source. . . » du menu View. Vous pouvez aussi sauvegarder la page chez vous via l'option de menu « Save As. . . » du menu File.

Épilogue

À moins que vous ayez la fibre artistique très tendue, je vous déconseille de passer des heures à fignoler vos pages web. À mon avis, une page sobre et bien agencée vaut beaucoup mieux qu'une page remplie de GIF animés et autres babioles flashantes. Donnez la priorité à l'information plutôt qu'à la forme. Si vous n'avez rien à dire ni rien à montrer, pourquoi faire une page web !? Enfin, c'est vous qui voyez de toute manière. Tous les (dé)goûts sont dans la (triste) nature. . .

Compilé avec L^AT_EX 2_ε le 6 novembre 2001. Merci à Seb qui a bien voulu relire la version préliminaire.

⁸Si vous avez créé des sous répertoires dans votre répertoire web, faites plutôt `find . -type d -print | xargs chmod ugo+rx` puis `find . -type f -print | xargs chmod ugo+r`