

## Structure d'une page Web

Une page Web est caractérisée par une adresse, des liens, des images, un code source et, éventuellement, des « frames ».

### L'adresse URL (*Uniforme Ressource Locator*)

C'est une sorte d'adresse universelle. Quand elle indique une page web, elle est composée d'une suite de mots et signes qui commence par *http://*

Il existe d'autres services sur Internet dont les URL commencent par :

- *gopher://* (outil Internet de recherche de sites)
- *ftp://* (protocole de transfert de fichiers)
- *file:///C:/* (fichiers html sur votre disque dur)
- *telnet://* (pilotage d'ordinateurs à distance)
- *News://* (Newsgroup)
- *mailto:* (courrier électronique).

La syntaxe, la casse (majuscule ou minuscule) et l'orthographe des adresses URL doivent absolument être respectées. La moindre erreur rendra l'accès impossible. Il faut donc se méfier des transcriptions que l'on trouve dans les publications : elles sont souvent erronées. Mais une bonne connaissance de leur structure vous permettra, la plupart du temps, de les corriger.

Après *http://*, la suite de l'URL dépend du serveur où se trouve l'information. On trouve souvent *www* : *http://www.multimania.com/Anasys* ou *http://www.cnamts.fr*

Mais ce n'est pas une généralité :

*http://members.aol.com/Anasys* ou *http://dicdoc.kb.inserm.fr:2010/basisMeSH/catn.html#N05*.

Ensuite on trouve des mots séparés par un point, une barre de fraction (/), couramment appelée un *slash*, un tiret (-), un tilde (~), un dièse (#).

Le *slash* penche toujours à droite dans une URL. Il fait passer d'un dossier (répertoire) à l'autre dans le chemin d'accès vers l'information. Le *slash* de fin d'URL est (en principe) facultatif.

Après *http://* et éventuellement les *www*, les mots qui suivent immédiatement représentent le nom de domaine. Le dernier mot avant le premier *slash* n'a pas la même signification s'il s'agit d'une URL sur un serveur américain ou bien si le serveur se situe dans le reste du monde.

Pour les serveurs américains, ce mot représente le type d'organisation qui le régit :

- *com* pour commercial
- *edu* pour éducation
- *gov* pour gouvernement

Pour le reste du monde, auparavant, ce mot représentait la nationalité de l'institution officielle auprès de laquelle le site était enregistré : *fr* (France), *uk* (United Kingdom), *de* (Deutschland), *ca* (Canada), etc.

Actuellement, le mot précédant le premier *slash* n'est plus systématiquement la nationalité.

Ce peut aussi être *com*, *org*, *net*, etc. En effet, les maîtres-toile (ou webmestres) se sont vite rendus compte que les résultats des requêtes des moteurs de recherche américains classent souvent, en premier, les sites dont l'extension du nom de domaine est « *com* ».

Pour ne pas apparaître en 1235<sup>e</sup> page (!), les sites choisissent donc souvent un nom de domaine se terminant par *com* (s'il est encore disponible au niveau de la planète...).

Le dernier mot d'une URL est parfois *.htm* ou bien *.html* (parfois un autre mot). Cette terminaison précise que le fichier de la page recherchée est en langage *html*.

Suivant le type de serveur où elle se trouve, il sera nécessaire d'écrire trois lettres pour l'extension (*htm*) ou bien quatre lettres (*html*). Une erreur à ce niveau vous empêchera d'ouvrir la page (on vous répondra que le site n'existe pas). Donc, méfiez-vous des adresses griffonnées que l'on a pu vous recommander : les erreurs de transcriptions sont fréquentes !

Connaissant le mode de constitution des adresses URL, il vous sera facile de trouver par déduction et avec peu de risques d'erreur l'URL de beaucoup de sites. Par exemple pour la CNAMTS :

*http://www.cnamts.fr*

Les sites de CHU ont une syntaxe particulière comportant toujours « *chu-* ». Par exemple :

*http://www.chu-rouen.fr/*

Il vous arrivera de cliquer sur un lien sans obtenir le transfert vers la page souhaitée. Cela arrive quand un lien est rompu, c'est-à-dire que l'URL a été modifiée (ou bien que le site a disparu).

Avant de vous lancer dans une recherche sur les annuaires que sont les moteurs de recherche, vous pourrez essayer de reconstituer vous-même la nouvelle adresse. Pour cela, vous remonterez dans le chemin d'accès en effaçant, progressivement, les contenus situés entre les *slashes*, en commençant par la fin.

### **Les liens hypertextes**

Dans une page Web, les mots soulignés sont, le plus souvent, des liens permettant d'accéder à d'autres pages du Web.

Sur un lien, le passage de la souris transforme sa flèche en une main avec un doigt pointant.

Quand la souris passe sur l'hypertexte, l'adresse URL, sur laquelle ce doigt est pointé, s'affiche en bas de l'écran, au-dessus de la barre des tâches de votre navigateur.

### **Les images**

Ce sont des éléments indépendants. Elles ont une adresse URL spécifique.

Elles peuvent provenir d'autres ordinateurs que celui qui héberge le texte qu'elles illustrent (voir chapitre « Navigateur »).

### **Le code source**

Le code source est écrit en langage *html*. Cet ensemble de commandes permet de reconstituer la page.

### **Les frames (en français : cadres ou cadres multiples)**

Certaines pages sont constituées de deux ou plusieurs pages indépendantes. Les commandes *html* arrivent à les rassembler sur un écran en réalisant une sorte d'incrustation.

Les *frames* sont très commodes pour présenter un sommaire dynamique. Il suffit de cliquer sur une rubrique pour que le document s'y rapportant s'affiche. Le sommaire reste à l'écran pour permettre un autre choix.

Il est facile d'ouvrir, indépendamment, une page que le concepteur a intégrée dans une *frame*. Pour cela, on effectue un clic droit sur son lien d'appel de la page puis on clique sur « *ouvrir dans une nouvelle fenêtre* ».

Si on n'a accès qu'à un lien vers une *frame*, il est possible d'ouvrir indépendamment l'une ou l'autre page qui compose la *frame*. Pour cela, il faut aller lire, dans le code source, l'adresse d'une des pages qui composent la *frame* et appeler cette page avec le navigateur. Cette dernière procédure nécessite une connaissance minimale des commandes *html*.