

Initiation aux serveurs Web

Patrick Cégielski
cegielski@u-pec.fr

Janvier 2012

Pour Irène et Marie

Legal Notice

Copyright © 2012 Patrick Cégielski

Université Paris Est Créteil - IUT

Route forestière Hurtaut

F-77300 Fontainebleau

cegielski@u-pec.fr

Table des matières

1	Mise en place d'un serveur Web	1
1.1	Le Web	1
1.2	Mise en place d'un serveur Apache	1
1.3	Historique	2
1.3.1	Hypertexte	2
1.3.2	Création du WWW	2
1.3.3	Diffusion du WWW	4
1.3.4	De nouveaux navigateurs	4
1.3.5	Développement des serveurs Web	5
2	Le langage HTML	7
2.1	Page de texte	8
2.1.1	Premier exemple	8
2.1.2	Structure d'un fichier HTML	8
2.1.3	Structuration du texte	9
2.1.4	Des textes en couleur	12
2.2	Les liens	15
2.2.1	Les liens hypertexte	15
2.2.2	Les liens hypermédia	17
2.3	Formulaires	20
2.3.1	Premier exemple	20
2.3.2	Syntaxe	21
2.3.3	Deuxième exemple	22
2.3.4	Autres éléments de syntaxe	23
2.4	Les tableaux	24
2.5	Les cadres	26
2.6	Historique	28
2.6.1	Les origines 1989-1992	28
2.6.2	L'apport de Mosaic (1993)	28
2.6.3	L'apport de Netscape Navigator (1994)	29
2.6.4	HTML 2.0 (1995-1996)	29
2.6.5	HTML 3.2 et 4.0 (1997)	29
3	Introduction à HTTP	31
3.1	HTTP 0.9	31
3.2	HTTP 1.0	32
3.2.1	Requête HTTP	32
3.2.2	Réponse HTTP	33

3.3	HTTP 1.1	33
3.3.1	Connexions persistantes	33
3.3.2	Négociation de contenu	34
3.3.3	Transfert par morceaux	34
3.3.4	Nouvelles méthodes	35
3.4	Exemples	35
3.5	Codes de statut	38
3.6	Historique	39
4	Exécution de code par le serveur HTTP	41
4.1	Mise en place	41
4.1.1	Configuration du serveur	41
4.1.2	Premier exemple	42
4.2	Les directives SSI	43
4.2.1	Syntaxe	43
4.2.2	Directive d'affichage	44
4.2.3	Directive d'inclusion de fichiers	44
4.3	Inconvénient des SSI	45
5	Le langage de programmation Perl	47
5.1	Installation et premier programme	48
5.1.1	Installation	48
5.1.2	Premier programme	48
5.2	Premiers éléments de langage Perl	49
5.2.1	Commentaires	49
5.2.2	Scalaires	49
5.2.3	Entrées et sorties standard	50
5.2.4	Structures de contrôle	50
5.2.5	Modularité	51
5.3	Structures de données	51
5.3.1	Tableaux et listes	51
5.3.2	Chaînes de caractères	51
5.3.3	Tableaux indicés ou associatifs	52
5.4	Historique	52
6	Programmation CGI	53
6.1	Principe : architecture à trois niveaux	53
6.2	Paramétrisation du serveur HTTP	54
6.2.1	Principe	54
6.2.2	Cas d'un serveur Apache	54
6.2.3	Installation de Perl	54
6.2.4	Premier exemple	55
6.3	Les variables d'environnement	55
6.3.1	Variables d'environnement standard	55
6.3.2	Paramètres passés par l'utilisateur	57
6.4	Acquisition des données d'un formulaire	59
6.4.1	Paramètres passés par la méthode GET	59
6.4.2	Paramètres passés par la méthode POST	60
6.4.3	Décodage des données	61
6.4.4	Cas de plusieurs champs	63

Chapitre 1

Mise en place d'un serveur Web

1.1 Le Web

Le *World Wide Web* (**WWW**), littéralement la « toile (d'araignée) mondiale », communément appelé le **Web**, parfois la *Toile*, est un système hypertexte fonctionnant sur Internet qui permet de consulter, avec un logiciel appelé **navigateur** (*browser* en anglais), des informations appelées **pages Web** accessibles sur des sites (Web). L'image de la toile d'araignée vient des **hyperliens** qui lient les pages web entre elles.

Un **serveur HTTP** ou **démon HTTP** ou **HTTPd** (pour *HTTP daemon*) ou (moins précisément) **serveur Web**, est un logiciel servant des requêtes respectant le protocole de communication client-serveur *Hypertext Transfer Protocol* (**HTTP**), qui a été développé pour le World Wide Web.

Un ordinateur sur lequel fonctionne un serveur HTTP est également appelé **serveur Web**.

1.2 Mise en place d'un serveur Apache

Le serveur Apache est un logiciel libre disponible à l'adresse :

<http://httpd.apache.org>

pour plusieurs systèmes d'exploitation Unix. Il existe une version pour Windows. C'est celle-ci que nous proposons d'installer dans une première étape.

Première étape : Récupération des binaires.- Sur le site indiqué ci-dessus, cliquer sur la colonne

de gauche en haut sur « *Download!* » (en fait sur le sous-menu *from a mirror*). Cliquer ensuite sur la dernière version stable (*Stable release*), la version 2.2.21 du 13 septembre 2011 au moment où j'écris ceci. Cliquer enfin sur le lien qui suit *Win32 Binary including OpenSSL 0.9.8r (MSI Installer)*. On obtient le fichier :

```
httpd-2.2.21-win32-x86-openssl-0.9.8r.msi
```

à placer dans le répertoire de votre choix.

Deuxième étape : installation des binaires.- Il suffit de double-cliquer sur ce fichier pour installer les binaires correspondants (si vous possédez les droits d'administrateur de la machine) dans le répertoire que vous voulez.

Troisième étape : vérification.- Si tout s'est bien déroulé alors en mettant :

```
http://localhost
```

dans votre navigateur préféré, on voit apparaître une fenêtre avec le simple affichage : *It works!*

Si le chemin du répertoire dans lequel Apache est installé est `repertoire` alors le fichier `index.html` qui est utilisé se trouve dans le répertoire `repertoire\htdocs`.

Vous pouvez placer votre site web préféré à cet emplacement.

1.3 Historique

1.3.1 Hypertexte

En 1945, Vannevar BUSH écrit un article [Bus-45] dans *Atlantic Monthly* sur un outil mécanico-photo-électrique appelé *Memex* (pour *MEMory EXtension*), qui pourrait créer et suivre des liens entre des documents sur microfiche.

Dans les années 1960, Doug ENGELBART crée un prototype de ce qu'il appelle un *oNLine System* (NLS) qui permet de naviguer à travers des liens hypertexte. Il invente la souris pour en faciliter la manipulation.

Ted NELSON introduit le terme **hypertexte** (*hypertext* en anglais) dans l'article [Nel-65].

Beaucoup de recherches sont effectuées sur les hypertextes. Il y a eu deux conférences, *HyperText'87* [HT-87] et *HyperText'89* [HT-89]. Le NST (*National Institute of Standards and Technology*) a tenu un atelier à Washington sur la standardisation des hypertextes en 1990. La revue *Communications of the ACM* a fait paraître en juillet 1988 un numéro spécial sur les hypertextes contenant un grand nombre de références [ACM-88]. Une bibliographie est donnée dans [NIST-90] et un newsgroup *alt.hypertext* fut créé.

1.3.2 Création du WWW

Tim BERNERS-LEE occupe un emploi temporaire à l'organisation européenne pour la recherche nucléaire (CERN pour *Conseil Européen pour la Recherche Nucléaire*) en 1980. Il y conçoit un logiciel, *Enquire*, pour conserver des données sur l'organisation du CERN. Il n'y a pas d'interface graphique mais est multi-utilisateur [BL-80].

Il revient au CERN sur un emploi permanent en 1984. Il propose, le 13 mars 1989, de créer un système hypertexte distribué sur le réseau informatique pour que les collaborateurs puissent

partager les informations au sein du CERN [BL-89]. Tim BERNERS-LEE utilise le terme **hypermédia** (*hypermedia* en anglais) pour désigner les documents incluant du texte, certes, mais aussi des graphiques, du son et des vidéos.

Cette même année, les responsables du réseau du CERN décident d'utiliser la suite de protocoles TCP/IP et le CERN ouvre sa première connexion extérieure avec Internet.

En septembre 1990, le patron de BERNERS-LEE lui permet d'acheter un Next Cube et de l'utiliser pour implémenter son système hypertexte. En octobre, l'ingénieur système Robert CAILLIAU se joint au projet d'hypertexte au CERN, immédiatement convaincu de son intérêt, et collabore à une révision de la proposition [BC-90] pour la rendre plus attractive. Tim BERNERS-LEE et Robert CAILLIAU sont reconnus comme les deux personnes à l'origine du *World Wide Web*.



FIG. 1.1 – Le premier serveur Web, installé sur un Next Cube

Le premier serveur Web est prêt en novembre, il se trouve à l'adresse `nxoc01.cern.ch`; la première page Web est :

`http://nxoc01.cern.ch/hypertext/WWW/TheProject.html`

(non conservée); la plus ancienne page conservée date du 3 novembre et est encore accessible :

`http://www.w3.org/History/19921103-hypertext/hypertext/WWW/Link.html`

Le premier navigateur Web, appelé *WorldWideWeb* (plus tard rebaptisé *Nexus*), développé en Objective C sur NeXT, sort le 25 décembre 1990 et est présenté publiquement le 25 février 1991 :

`http://www.w3.org/People/Berners-Lee/WorldWideWeb.html`

En plus d'être un navigateur, *WorldWideWeb* est un éditeur Web. Le navigateur mode texte (*line-mode* en anglais) est développé en langage C pour être portable sur les nombreux modèles d'ordinateurs et terminaux de l'époque. *WorldWideWeb* est le premier logiciel à utiliser non seulement le protocole FTP mais également le protocole HTTP (*HyperText Transfer Protocol*), inventé par Tim BERNERS-LEE en 1989, les adresses Web et HTML.

1.3.3 Diffusion du WWW

Plusieurs versions circulent parmi les collègues de BERNERS-LEE avant d'être mises à disposition du public (par l'intermédiaire des newsgroups). Le 6 août 1991, Tim BERNERS-LEE rend le projet *WorldWideWeb* public dans un message sur *Usenet* :

`http://groups.google.com/group/alt.hypertext/msg/06dad279804cb3ba`

On pourra remarquer qu'il y est fait alors référence au projet *World Wide Web* (avec des espaces entre les mots) et à l'acronyme WWW.



FIG. 1.2 – Logo historique du Web par Robert CAILLIAU

En 1992, le *WorldWideWeb* est promu sur le Web :

`http://www.w3.org/History/19921103-hypertext/hypertext/WWW/TheProject.html`

On pourra remarquer qu'il y est fait alors référence aussi bien à *WorldWideWeb*, à *World Wide Web* qu'à W3. La notion de **navigateur** (*browser* en anglais) est introduite, comme « un programme qui permet à une personne de lire de l'hypertexte ; le navigateur fournit un moyen de visualiser le contenu des nœuds et de naviguer d'un nœud à l'autre », ainsi que le **web** comme un « ensemble de nœuds interconnectés par des liens ».

Jusqu'en 1993, le Web est essentiellement développé sous l'impulsion de Tim BERNERS-LEE et Robert CAILLIAU. BERNERS-LEE et Jean-François GROFF adaptent certaines composantes de *WorldWideWeb* dans une version C, en créant l'API `libwww`, qui existe toujours (dans une version qui a évolué) :

`http://www.w3.org/Library/`

Le 30 avril 1993, le CERN décide de placer le code source de *WorldWideWeb* dans le domaine public, faisant de *WorldWideWeb* un logiciel libre. Plusieurs versions du logiciel peuvent encore être téléchargées, dont les sources :

`http://www.w3.org/History/1991-WWW-NeXT/Implementation/`

et les binaires (pour Next) :

`http://browsers.evolt.org/?worldwideweb/`

Fin 1993, les médias grand public remarquent Internet et le WWW.

1.3.4 De nouveaux navigateurs

En 1993 apparaissent des navigateurs concurrents : *NCSA Mosaic* et *Lynx*. Disponible d'abord sur *X Window*, puis sur Windows et MacOS, *Mosaic* cause un phénoménal accroissement de la popularité du Web. Il s'agit d'un navigateur Web développé par Eric BINA et Marc ANDREESSEN, alors étudiant, au *National Center for Supercomputing Applications* (NCSA), à l'université de

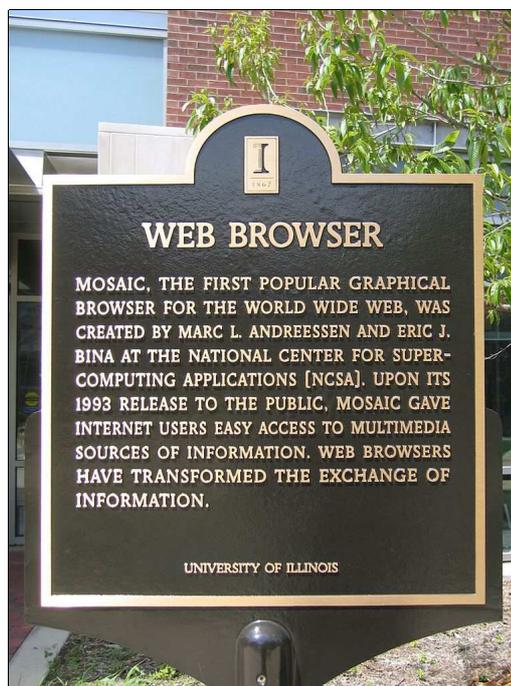


FIG. 1.3 – Plaque rappelant la création du navigateur *Mosaic*

Champaign dans l'Illinois. *NCSA Mosaic* jette les bases de l'interface graphique des navigateurs modernes et est à l'origine de la popularité du Web. Les images apparaissent dans les pages Web (*Mosaic 0.10*) alors qu'elles étaient affichées dans une fenêtre à part sur *WorldWideWeb*. Apparaissent également les formulaires interactifs (*Mosaic 2.0pre5*).

Le NCSA produit également le *NCSA HTTPd*, un serveur HTTP qui évoluera en Apache HTTP Server (le serveur HTTP le plus utilisé à partir de 1996).

En 1994, *Netscape Communications Corporation* est fondée avec une bonne partie de l'équipe de développement de *NCSA Mosaic*. Sorti fin 1994, *Netscape Navigator* supplante *NCSA Mosaic* en quelques mois.

Cette même année Tim BERNERS-LEE fonde le *World Wide Web Consortium (W3C)*, dont la première recommandation est la RFC 1738 (sur *Uniform Resource Locators*).

En 1995, Microsoft essaie de concurrencer Internet avec *The Microsoft Network (MSN)* mais échoue. Fin 1995, après la sortie de Windows 95 sans navigateur Web préinstallé, Microsoft lance le navigateur *Internet Explorer* et ouvre la guerre des navigateurs contre *Netscape Navigator*.

1.3.5 Développement des serveurs Web

Le premier serveur HTTP fut *CERN httpd*, développé par Tim BERNERS-LEE en 1990, avec une version 0.1 en juin 1991. Son développement fut repris par le consortium W3C, la dernière version fut la 3.0A sortie le 15 juillet 1996 sous une licence MIT :

<http://www.w3.org/Daemon/>

Il fut suivi de *NCSA HTTPd*, développé au NCSA en même temps que le navigateur *NCSA Mosaic*, abandonné mi-1994. Ce dernier a donné lieu à *Apache HTTP Server* de la *Apache Software Foundation*, qui est maintenant le serveur dominant sur le marché.

Ce dernier serveur HTTP est un logiciel libre, souvent abrégé en **Apache**, apparu en avril 1995, avec un type spécifique de licence, nommée *licence Apache*. Au début, il s'agissait d'une collection de correctifs et d'additions au serveur *NCSA HTTPd* 1.3, qui était dans le domaine public. De cette origine, de nombreuses personnes affirment que le nom « Apache » vient de *a patchy server*, soit « un serveur rafistolé ». Par la suite, *Apache* a été complètement réécrit, de sorte que, dans la version 2, il ne reste pas de trace de *NCSA HTTPd*.

La *fondation Apache* (*Apache Software Foundation* ou ASF) a été créée en 1999 à partir du groupe Apache (*Apache Group*) à l'origine du serveur en 1995. Depuis, de nombreux autres logiciels utiles au *World Wide Web* sont développés à côté du serveur HTTP.

Le livre de référence sur l'installation d'Apache est [LL-97] mais la documentation en ligne sur le site Web d'Apache, en plusieurs langues dont le français, est bien faite et tenue à jour.