

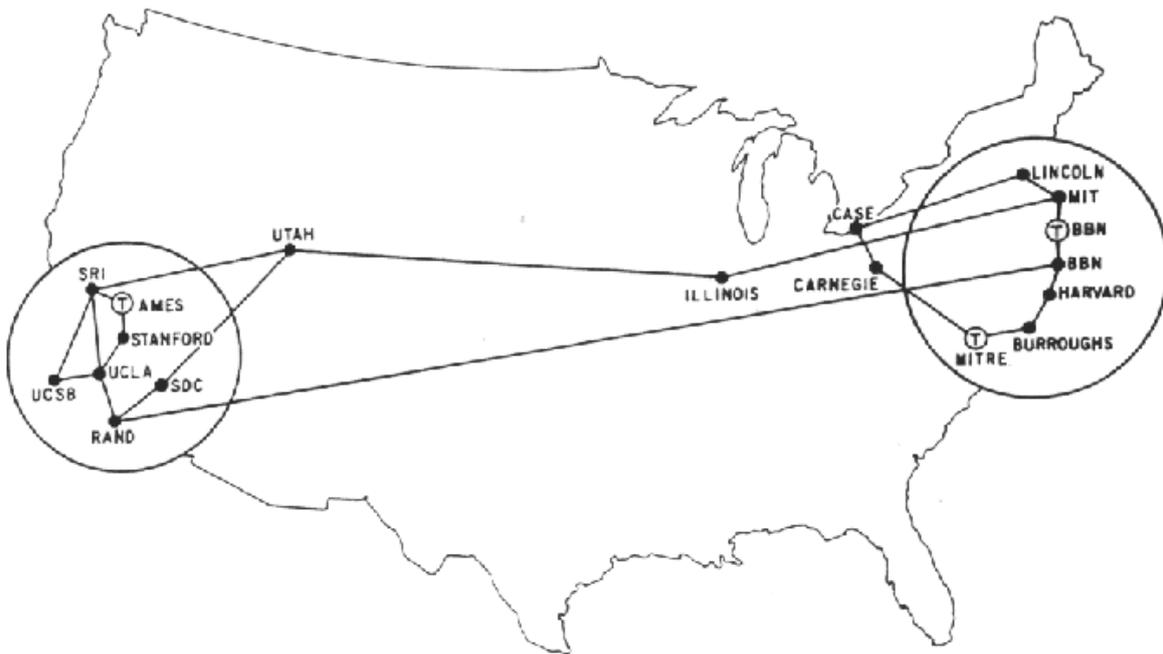
SEANCE 1 : Protocole Internet et principes de publications d'une page Web

1. Qu'est ce qu'internet ?

1.1. Arpanet : l'ancêtre d'Internet 1960

Le mythe voudrait que l'objectif d'Arpanet ait été de permettre aux réseaux de communication militaires (Arpanet a vu le jour en pleine guerre froide) de continuer à fonctionner malgré une attaque nucléaire massive de la part de l'Union soviétique, c'est-à-dire : « garder ouvertes des voies de communication quel que soit l'État de destruction du pays ». On crée ainsi un réseau qui relie quelques ordinateurs. On le nomme **ARPANET** (*ARPA était le nom du groupe de travail chargé de cette étude*).

Mais le Ministère des Armées américain ne retient pas le projet et ce sont les universités qui vont le récupérer : **en 1969, 4 universités** sont reliées entre elles. La première utilisation de ce réseau, c'est l'échange de courriers électroniques : les emails.



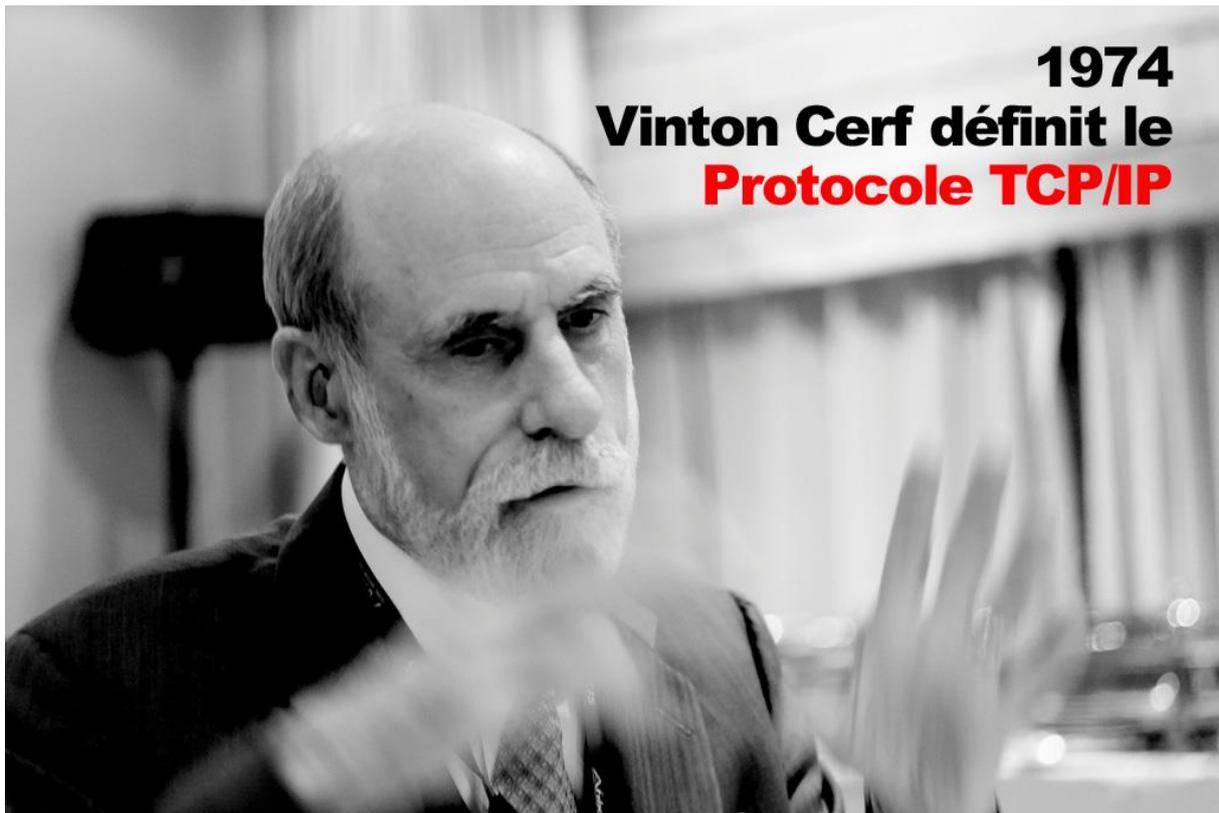
MAP 4 September 1971

1.2. Normalisation des protocoles de transfert - Extension du réseau 1970/1980

La difficulté principale, ce n'est pas de relier (par un fil !) des ordinateurs entre eux. Mais ces ordinateurs sont de conceptions ou de marques différentes : il faut qu'ils arrivent à se comprendre, à déchiffrer les mêmes données.

On décide donc dans les **années 70** de créer un « langage » commun, un standard ou une norme pour le transfert des données : c'est le **TCP/IP** (Protocole de Contrôle de Transfert/Internet Protocole), qui va s'imposer progressivement.

En 1977, une centaine d'ordinateurs sont reliés. Ça ne concerne évidemment pas le grand public, mais les institutions seulement. Au début des années 80, ce réseau se développe en Europe.



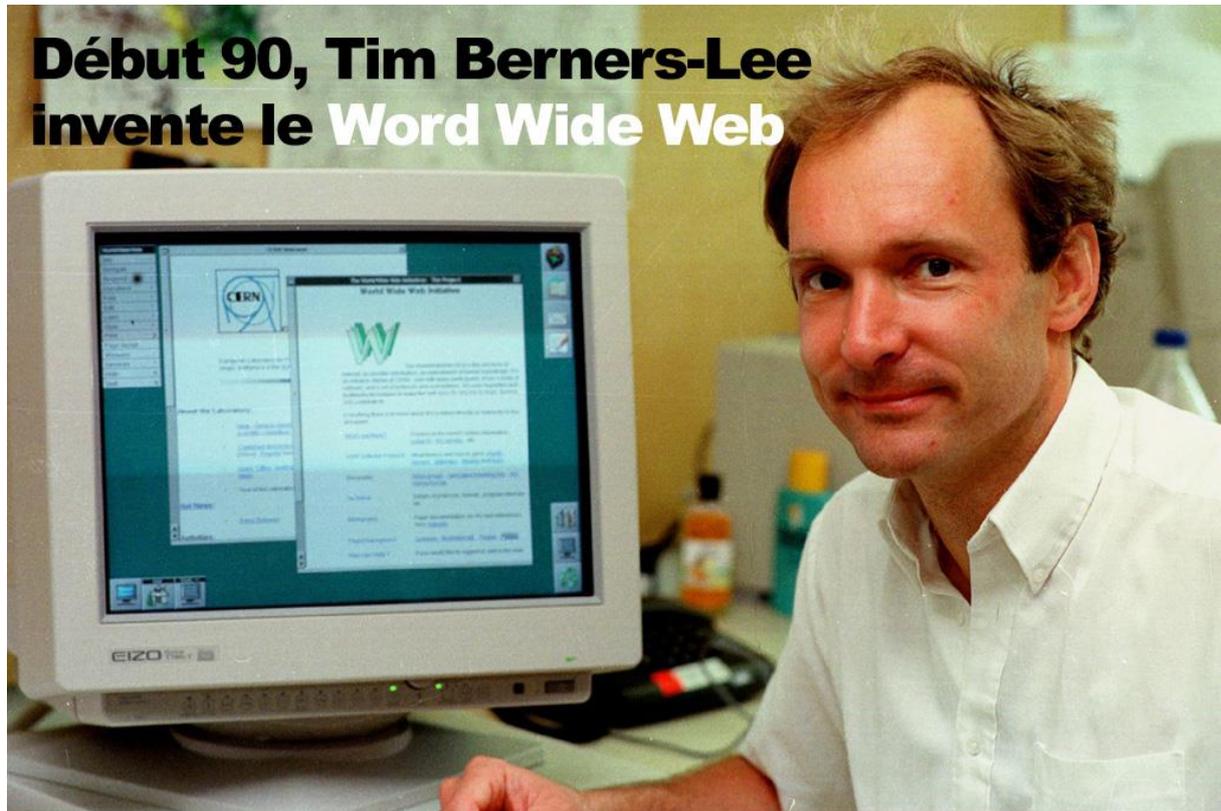
- **Naissance du « web » - Création du W3C 1990**

L'extension significative se produit au début des années 90. Là, un informaticien du CERN (Centre Européen de Recherche Nucléaire), Tim Berners-Lee, met en place le premier véritable site internet (encore destiné aux chercheurs).

On crée à cette époque le langage html qui permet l'affichage et la mise en forme de pages. On conçoit alors les « navigateurs » qui lisent ce langage et restitue sur un ordinateur privé l'affichage de la page d'origine. C'est ce qu'on nomme alors le World Wide Web ou Web (web signifie « toile », cela évoque les multiples ramifications d'une toile d'araignée).

La grande innovation c'est aussi ce qu'on nomme les « hyperliens » (et l'« hypertexte ») le jeu des liens pour passer d'une page à l'autre, ou d'un site à un autre.

En 1994 Tim Berners-Lee s'installe aux USA et participe à la création du **World Wide Web Consortium (W3C)**. Cet organisme de standardisation va promouvoir la compatibilité des technologies et des différents langages utilisés. Ses spécifications sont aujourd'hui la référence pour tous les développeurs.



Début 90, Tim Berners-Lee invente le Word Wide Web

1.3. Extension du web et développements technologiques 1990

En 1995 on dénombrait plus de 2 millions d'ordinateurs connectés, dans 146 pays. En 2000, on approche les 200 millions !

Les chiffres explosent ensuite, en 2007, en Chine, pays émergeant dans ce domaine, on compte 137 millions de « connectés » ! On planifie plus de 2 milliards de machines connectées d'ici à 2016.

Les aspects technologiques évoluent rapidement, avec notamment la généralisation du haut débit et la portabilité (liaisons sans fil).

Question souvent posée: quelle est la différence entre Internet et le Web ?

-> Internet c'est **le réseau physique**, les câbles (ou ondes radio aujourd'hui) qui relient les divers ordinateurs, c'est du courant électrique qui circule. Ce sont les ordinateurs reliés entre eux.

-> Le Web, c'est **le contenu**, ce sont les pages sur lesquelles on « surfe » à l'aide des liens.

2. Principes de production et de publication d'un site web

Rappel : Qu'est ce qu'un site web ?

1.1 Production d'une page web

Une page web est un document composé de code, d'écritures. L'extension de ce document définit le langage de programmation employé.

Prenons l'exemple d'un document intitulé « index.html ». Index, désigne la première page d'un site web. Le langage de programmation utilisé est « html », du nom que porte son extension.

La page web est produite par un développeur. Son mode de composition pourrait être comparé à un document rédigé avec un logiciel de traitement de texte comme « Microsoft Office Word® ». Seulement, pour produire ces pages, le développeur utilise un **éditeur de code**. Il existe 2 grandes familles d'éditeurs.

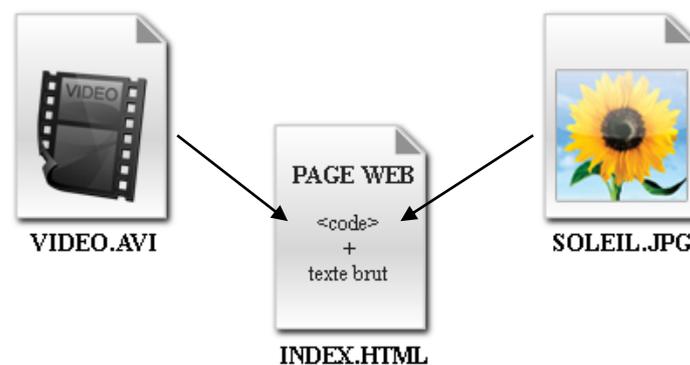
- **Les éditeurs payants** sont généralement des logiciels qui permettent une production automatique du code en cliquant sur des boutons. Ils intègrent dans la plupart des cas une représentation graphique du document web produit en temps réel, sans passer par le navigateur. Malheureusement le rendu est assez souvent faussé et limité aux seuls langages HTML et CSS... Ces éditeurs s'intitulent « **WYSIWYG** », qui est l'acronyme de « what you see is what you get ».

- **Les éditeurs gratuits** ont une présentation du code plus claire, accès sur la lisibilité mais n'ont pas forcément ces fonctions d'automatisations.

Le choix d'un éditeur dépend du niveau en programmation, de l'utilisation que l'on va en avoir, du système d'exploitation (linux / windows / mac) et de son prix (libre ou gratuit). La majorité des développeurs chevronnés auront tendance à privilégier l'éditeur gratuit « Notepad++ » qui est performant et utilise un système de coloration syntaxique. Un amateur passionné plutôt Dreamweaver.

(La liste des éditeurs est disponible en *annexe 1*).

Le code produit dans la page fait appel à des contenus externes (image / vidéo / son) afin de composer une mise en page numérique intégrant **des ressources multimédias** (des fichiers) et **du texte** (écrit directement dans le code de la page). La page web est ensuite interprétée par un « navigateur » ou « browser » en anglais, qui est la seule application capable d'interpréter et de représenter graphiquement le code, la mise en page produite.



Il est possible d'afficher le code d'une page web en allant dans la rubrique « code source » de votre navigateur. Ainsi il est possible de consulter le code de la page web en cours de consultation.

Il est important de comprendre que les médias et les textes sont deux éléments bien distincts. Images, vidéos, et autres médias, sont appelés dans la mise en page lorsque celle-ci est interprétée. Mais ils ne sont jamais intégrés directement dans le code. A l'inverse les textes d'un document sont directement inscrits dans la page.

Dans la production d'un site web, le développeur va dans un premier temps tester les fichiers sur sa machine avant de les mettre en ligne. On appelle cette phase, le développement « **locale** ». Celle-ci permet d'effectuer des tests avant la mise en ligne de la page. Cette phase finale s'intitule **migration**.

1.2 La structure standard d'une page web

Quelle soit dynamique ou statique, la **structure des écritures de base du document web sera toujours la même**, ce qui change c'est le langage de programmation que nous allons utiliser pour produire ces écritures. Toutes les pages obéissent au même principe de composition :

1.	< !doctype html>	<i>Doctype</i>
2.	<html>	<i>Début du document html</i>
3.	<head>	<i>Tête</i>
4.	</head>	<i>Fin de tête</i>
5.	<body>	<i>Corps de page</i>
6.	Contenu de la page web (mise en page)	<i>Contenu</i>
7.	</body>	<i>Fin corps de page</i>
8.	</html>	<i>Fin du document html</i>

Exemple : Créer un document texte, renommer l'extension de votre fichier en ".html". Insérer le code ci-dessus à l'intérieur de votre document. Remplacer la ligne 6 par « Hello world ! ». Sauvegarder et ouvrir votre document sur un navigateur web. Vous venez de créer votre premier document web.

Ce qui différencie une page statique d'une page dynamique est la production du contenu (6).

Dans une page web statique tous les contenus, les textes, les images, sont spécifiés par le développeur. **Ces contenus sont figés et ne peuvent pas évoluer.**

Une page dynamique va afficher les contenus en fonction de la demande (appelée **requête**) effectuée par l'internaute, construisant ainsi une page sur-mesure aux possibilités infinies. Dans une autre mesure il est possible de définir une page dynamique comme offrant des possibilités d'interactions avec le consultant que l'on appelle **le client**. Client = Internaute.

Nous y reviendrons plus tard.

1.3. La page web est lue et interprétée par un navigateur

Une fois les écritures produites, la page web est interprétée par un navigateur. Conventionnellement les navigateurs sont utilisés pour consulter du contenu en ligne « online », mais il peuvent également permettre d'interpréter du contenu « offline », c'est-à-dire n'importe qu'elle page html, certains fichiers multimédias, disponibles sur notre machine.

On compte 5 grands navigateurs à l'heure actuelle :

- Mozilla / Firefox
- Google / Chrome
- Microsoft / Internet Explorer
- Opera Software / Opera
- Apple / Safari

Il convient donc de tester le rendu de notre page web sur tous ces navigateurs mais également sur les différentes versions de chacun de ces navigateurs, car **ces derniers interprètent le code différemment**.

En effet, le langage html est en perpétuelle évolution, il est donc logique que les fonctionnalités actuelles ne soient pas interprétées par les versions antérieures du navigateur, car celles-ci n'existaient pas. Dans une moindre mesure des correctifs et des mises à jours permettent de faire évoluer les anciennes versions sans les rendre pour autant entièrement compatible avec les derniers standards. Cela explique en partie la nécessité d'avoir un navigateur mis à jour pour profiter pleinement de l'ensemble des applications web.

La version du navigateur, notamment « Internet Explorer 8, 9, 10 » qui représente 33% du marché, dépend du système d'exploitation installé (Windows XP / Vista...) sur la machine de l'internaute. Ce qui signifie qu'un internaute ayant fait l'acquisition d'une machine avec Windows XP il y a 7 ans, ne pourra par conséquent pas avoir accès à la dernière version de son navigateur et ne pourra pas interpréter les dernières applications HTML5.

C'est dans cette problématique de compatibilité et d'homogénéité qu'évolue le travail du développeur.

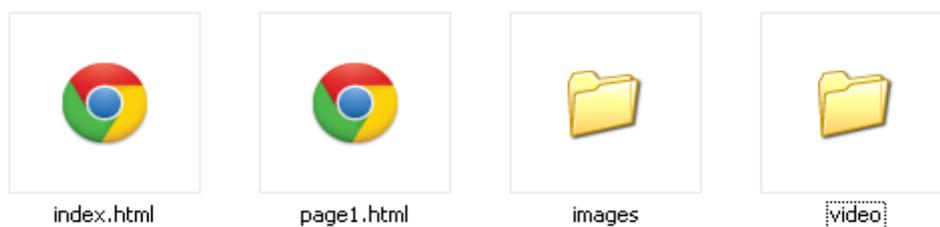
1.4. Un site web est un dossier composé de pages web

L'action de naviguer sur internet, appelée « surf » symbolise la possibilité de passer d'une page web à une autre, en réalité il s'agit de consulter un fichier puis un autre...

Un site internet est un ensemble de pages web développées dans plusieurs langages de programmation en interactions. L'ensemble de ces pages sont placées dans un dossier qui peut éventuellement inclure d'autres dossiers, c'est ce que l'on appelle **l'arborescence des fichiers**.

Le niveau 0, du dossier s'appelle **la racine**, « **root** » en anglais.

Par ailleurs l'ensemble des fichiers multimédias (photo, vidéos, son...) sont également inclus à l'intérieur de ces dossiers, généralement séparés des pages web, dans des dossiers dédiés. Lorsqu'une page web est interprétée par le navigateur, ce dernier va intégrer, en fonction des informations contenues dans le code, les différents contenus externes au sein de la mise en page. Mais code et ressources sont toujours séparés.



Ce site est composé de 2 pages. La page d'accueil, ainsi qu'une autre page « page1.html ». L'ensemble des contenus images et vidéos sont stockés dans leurs dossiers respectifs. Ces contenus peuvent être appelés, intégrés, dans la mise en page de « index.html » ou « page1.html »

C'est en partie ce qui explique qu'il soit potentiellement possible de récupérer tout type de contenu visible dans votre navigateur. Et c'est d'ailleurs cette possibilité qui pose des problèmes juridiques autour du droit d'auteur, droit à l'image etc...

«Faites clique droit + enregistrer sous » sur votre navigateur pour copier le contenu en cours de

lecture.

1.5. Publier ce dossier (votre site web) sur internet

Pour rendre la consultation de ce dossier accessible aux internautes du monde entier, il va falloir stocker notre dossier de site web sur un ordinateur qui partagera le contenu de ce dossier. En fait, lorsque nous consultons un site internet, **nous consultons les fichiers disponibles dans un dossier d'ordinateur distant, ouvert a tous**. Plus précisément, nous demandons à un ordinateur distant de nous envoyer une page html et les contenus auxquels elle fait référence, afin de les interpréter avec notre navigateur. **C'est le principe du partage de fichier et du réseau.**

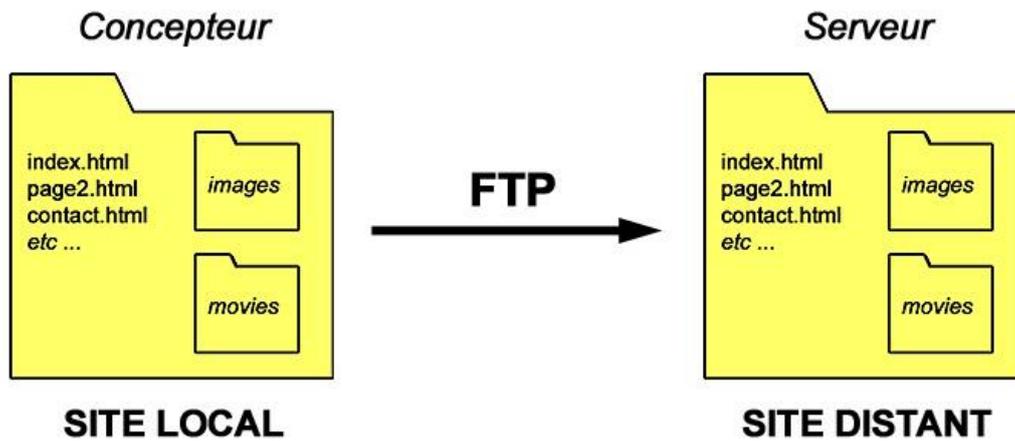
Les ordinateurs chargés de stocker ces fichiers se nomment « **serveur** ». Et cette prestation de stockage de fichiers s'appelle l'hébergement. Des millions d'ordinateurs hébergent actuellement les fichiers de millions de sites web. Certaines compagnies de renom développent des « data center », des fermes de stockages d'ordinateurs qui contiennent des milliards de données.

Google data center en video : <http://youtu.be/avP5d16wEp0>



Il va donc falloir envoyer les fichiers de notre machine sur cet ordinateur distant (le serveur). Pour procéder nous avons recours à un **logiciel FTP**. (File Transfer Protocol) qui sert de médiateur entre les machines pour envoyer des fichiers.

On développe donc le site sur son propre ordinateur (on parlera de « site local »), puis on le transfère sur le serveur (« site distant »). Il faudra, bien sûr respecter strictement l'imbrication des différents dossiers et fichiers, faute de quoi les liens deviendraient caducs et la navigation serait impossible.



Pour communiquer entre elles et partager des fichiers, les machines utilisent un protocole intitulé « **TCP/IP** ». «Transmission Control Protocol/Internet Protocol». Ce protocole fut construit à l'origine dans un but militaire « Arpanet ».

Quand vous êtes sur un site web vous consultez donc les fichiers qui sont stockés sur une machine distante ouverte à tous.

Il n'est pas nécessaire de comprendre le fonctionnement du protocole, comme il n'est pas nécessaire de connaître le fonctionnement d'un grille pain pour pouvoir faire des toasts... Toutefois ces connaissances sont indispensables pour les personnes en charge de maintenir et d'administrer un réseau, de partage de fichier en entreprise, ou de site web sur internet. Ce corps de métier s'intitule « administrateur réseau ».

Chaque ordinateur possède une adresse IP. Cette adresse IP est un numéro d'identification qui est attribué de façon permanente ou provisoire à chaque appareil connecté. Ainsi lorsque nous accédons à une page web, notre machine demande en réalité l'accès à un fichier de la machine qui a pour adresse IP 98.155.41.44, par exemple...

L'adresse IP d'une machine est un numéro fixe, comme une adresse de boîte aux lettres. Elle est constituée de 4 numéros de 0 à 255. Exemple : 10.44.53.17

1.6. L'internaute / client

On appelle le navigateur de l'utilisateur le **Client**, puisque c'est lui qui effectue une requête.

Lorsque le client tape dans une adresse dans la barre de son navigateur il va en fait demander l'affichage d'une page hébergée sur un serveur. Le serveur va alors renvoyer le code de la page demandée au navigateur de l'utilisateur qui va l'interpréter à son tour.

Le serveur renvoi la page html et les contenus qui sont associés (images / vidéos etc...). Certains de ces contenus sont mis en cache, c'est-à-dire qu'ils sont stockés sur la machine du client pour avoir un affichage plus rapide des médias ayant déjà été chargés

Comment traduire l'url dans la barre d'adresse ?

Lorsque nous tapons un nom de domaine dans la barre d'adresse de notre navigateur le serveur de nom fait le lien entre le nom de domaine « www.monsite.com » et l'adresse IP de l'ordinateur qui héberge les fichiers.

On parle de DNS (Domain Name System) pour le nom de domaine.

Lorsque l'on tape un nom de domaine dans la barre d'adresse on accède en fait à la racine du serveur, le niveau 0 du FTP. Dans une adresse internet, l'imbrication des dossiers successifs est marquée par un slash, exemple : « www.monsite.com/images/page.html »

1.7. Interprétation du serveur et envois des informations

Lorsqu'une page s'intitule « index », elle est par défaut la première page du site ou d'un dossier.

Lorsque vous tapez www.baidu.com la page que vous consultez est en réalité www.baidu.com/index.html

De même si vous consultez www.monsite.com/dossier, vous consultez en réalité www.monsite.com/dossier/index.html

Encore faut-il que la page index existe...

La « racine » ou « root » d'un site internet correspond généralement au point de départ, au niveau 0 du FTP et correspond donc à l'écriture du nom de domaine strict.

Pour afficher sa page il suffit de taper l'adresse du site, puis l'arborescence des dossiers à suivre pour atteindre le fichier et enfin le nom du fichier (si ce dernier n'est pas l'index) depuis la racine du FTP.

ANNEXE 1 : Editeurs de code html

1. Les éditeurs à éviter

Frontpage / Word / Publisher

Le code est généralement invalide et pose des problèmes de compatibilités avec certains navigateurs. Génère du code invalide : balises en majuscules, propriétés propriétaires (c'est-à-dire non standards, invalides et ne fonctionnant pas sous tous les navigateurs).

2. Les éditeurs WYSIWYG

Dreamweaver

Le plus performant des WISIWIG, un logiciel professionnel qui fait partie de la suite Adobe®. C'est celui qui génère le code le plus propre en mode création, et l'auto-complétion des balises et attributs en mode code est appréciable.

Web Expert

Il est relativement complet et génère un code propre il ne prend pas en compte les anciens navigateurs. Toutefois son prix reste un frein à son accession.

Nvu / Kompozer

Deux projets basés sur le Composer de Mozilla.

Nvu est le premier à avoir vu le jour, mais le développement de ce logiciel est désormais suspendu et un certain nombre de défauts lui sont reprochés par des utilisateurs avertis.

Tout n'est pas perdu pour autant et le projet "non officiel" Kompozer a vu le jour. Il est fortement recommandé de préférer cette nouvelle version de l'éditeur, qui pour sa part, reste en développement, corrige des bugs de Nvu et apporte de nouvelles fonctionnalités à l'éditeur.

Amaya

Un autre éditeur web respectueux des standards, créé par le W3C, gratuit et open-source. Sans doute l'un des plus simples à prendre en main, il convient très bien aux débutants.

Vulcain

Récent, ce logiciel (version 2.0 uniquement) fait ses preuves et offre la possibilité de créer des sites web dynamiques en HTML, CSS, PHP, JavaScript. L'utilisation de ce WYSIWYG requiert quelques notions de base en programmation. Le code généré est propre, il est possible de visionner votre site dans le navigateur intégré au logiciel mais l'affichage reste très médiocre! Le logiciel vous permet aussi d'envoyer votre site sur le net avec la fonction FTP.

Webacappella

Ce WYSIWYG possède une version d'essais, il ne nécessite aucune connaissance en programmation et offre un design facile à gérer pour l'utilisateur.

3. Editeurs en mode code

Notepad++

Notepad++ est un éditeur de code source qui supporte plusieurs langages. Il utilise la même composante que SciTE (Scintilla). Il offre la possibilité d'auto-complétion de texte avec certains langages en proposant une liste de fonctions de base utiles. Tout est paramétrable. La coloration syntaxique par défaut est très pertinente.

Distribué gratuitement sous licence GPL.

PSPad

Un excellent éditeur en mode code sous Windows. Possède un tas de fonctionnalités et détails pratiques. En plus, il est très léger. Il ressemble beaucoup à Notepad++

On peut lui ajouter des extensions, disponibles sur le site officiel.

SciTE

Excellent, léger, simple et efficace (libre surtout :)).

Sources : « *Comment ça marche* » :

<http://www.commentcamarche.net/faq/4671-webmaster-choisir-un-bon-editeur-les-editeurs-a-eviter>