

Le Web

Présentation du langage HTML et
du protocole HTTP.

Le web

Rapide historique :

1989 : première note relative au web « hypertexte et le Cern » qui jette les bases du World Wide Web.

1990 : disparition du réseau ARPAnet et apparition de l'Internet.

1991 : premier serveur web sur l'Internet, celui du CERN.

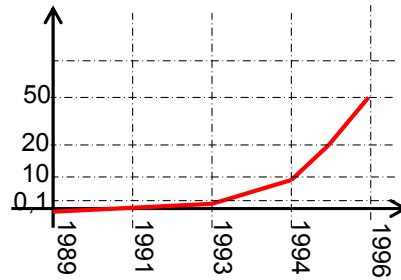
1993 : première version « libre » d'un navigateur web (celui utilisé par le CERN).

1994 : création du WWW Consortium par le CERN, le MIT et l'INRIA.

1995 : Java s'intègre au Web (et au navigateur Netscape).

1996 : début de l'exploitation commerciale du Web.

% de flux web sur l'internet

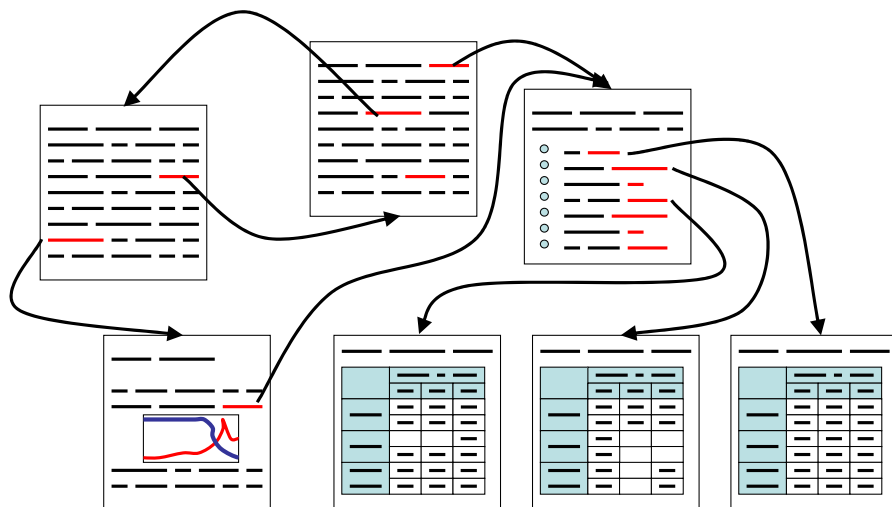


Objectif initial du web

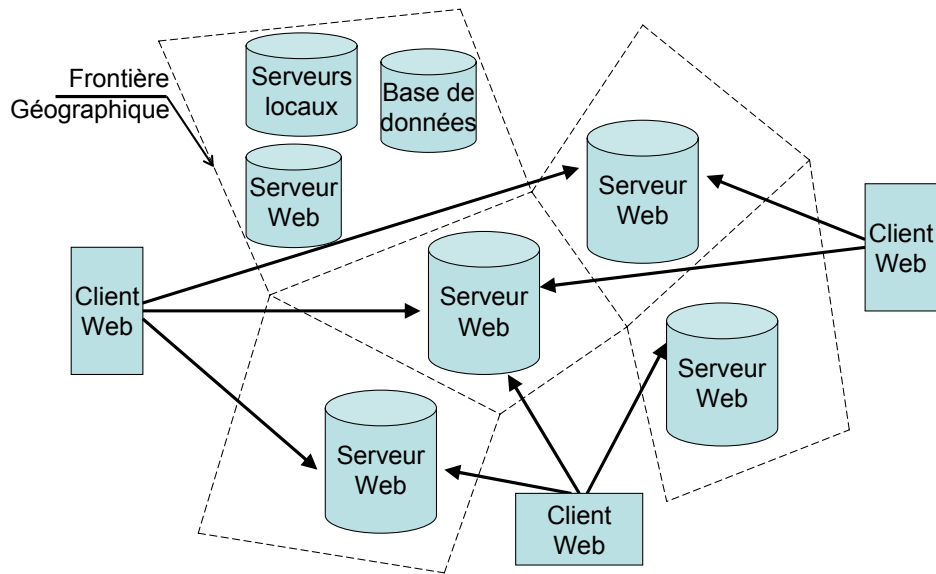
Pourquoi le Web est inventé au CERN ?
(Centre Européen de Recherche Nucléaire)

- ⇒ un grand nombre d'universités et de laboratoires européens.
- ⇒ des documents produits sur une grande variété de sites, dans une grande variété de formats.
- ⇒ volonté de constituer une base unique, et globale composée de l'ensemble de ces documents qui sont néanmoins éparpillés sur le réseau.

Notion de document Hypertexte



L'architecture du Web



http://www.unsite.com:80/unRep/unDoc.cgi?arguments

protocole utilisateur Adresse du serveur port Chemin du document arguments

6

Documents hypertexte et Langage HTML

Objectif :

Présentation d'un texte « formaté » qui
contient différents supports de média dont :

- Du texte, des tables, des listes, ...
- Des images, des sons, des vidéos, ...
- Des fichiers binaires ;
- Des références vers des hyperliens.

Documents hypertexte et Langage HTML

Principe du langage :

Section de textes à afficher structurés à l'aide de *tag HTML* notés `<tagName> </tagName>`.

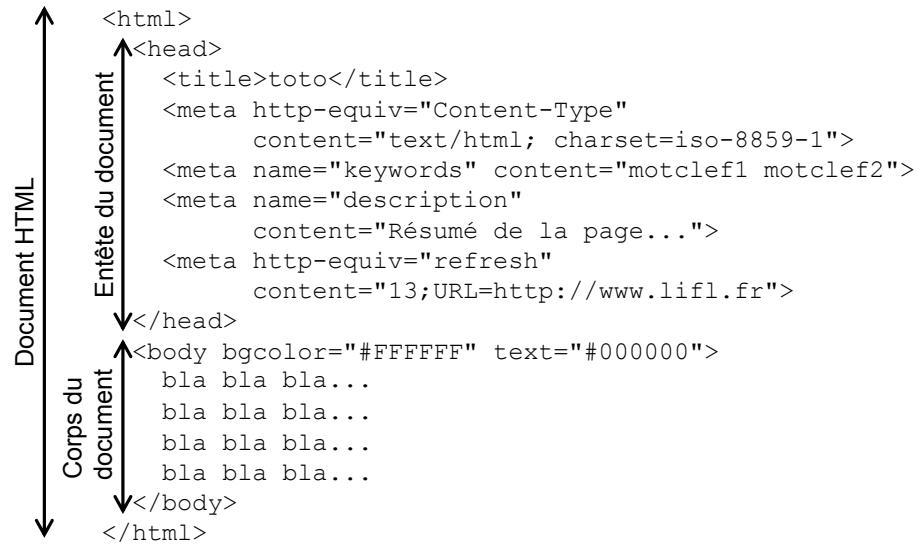
Le premier tag (`<tagName>`) définit le début de la zone de texte sur laquelle il porte, le second (`</tagName>`) définit la fin de la zone de texte concernée.

Les tags HTML peuvent être encapsulés les uns dans les autres.

Exemple :

```
<p> texte affiché à l'intérieur d'un  
paragraphe.</p>
```


Structure générale d'une page HTML



Tags clefs du langage HTML

Format du texte :

`<h1>` titre du document `</h1>`

`<h2>` sous-titre `</h2>`

`<h3>` titre de section `</h3>` (jusqu'à `<h6>` `</h6>`)

`<p>` Corps d'un paragraphe, paragraphe qui peut contenir plusieurs lignes, comme tout texte entre deux tags. `</p>`

`
` pour passer **à** la ligne à l'intérieur d'un paragraphe.

`<i>` texte en italique `</i>`

`` texte en gras ``

``
texte avec une couleur et une fonte
particulière **à** ``

`` texte associé à un
hyperlien vers www.lifl.fr ``

Tags clefs du langage HTML

Liste de puces et listes de points :

Exemple de code :

```
<p>Liste de puces et de points :</p>
<ul>
  <li> puce numero 1 ;</li>
  <li> puce numero 2 ;</li>
  <li> puce numero 3 dont deux
    points principaux :
    <ol>
      <li> point 3.1 ;</li>
      <li> point 3.2 .</li>
    </ol>
  </li>
  <li> puce numero 4.</li>
</ul>
```

Rendu par le navigateur :

Liste de puces et de points :

- puce numero 1 ;
- puce numero 2 ;
- puce numero 3 dont deux points principaux :
 1. point 3.1 ;
 2. point 3.2 .
- puce numero 4.

Tags clefs du langage HTML

Tableaux est tables HTML

Exemple de code :

```
<p>Exemple de table :</p>
<table width="400" height="92" border="1">
  <tr>
    <td>T(1,1)</td> <td>T(2,1)</td>
    <td>T(3,1)</td>
  </tr>
  <tr>
    <td rowspan="2">T(1,2) et T(1,3)</td>
    <td>T(2,2)</td> <td>T(3,2)</td>
  </tr>
  <tr>
    <td colspan="2" height="50%">
      T(2,3) et T(3,3)</td>
  </tr>
</table>
```

Rendu par le navigateur :

Exemple de table :

T(1,1)	T(2,1)	T(3,1)
T(1,2) et T(1,3)	T(2,2)	T(3,2)
T(2,3) et T(3,3)		

HTML est les zones d'images

Placer une image dans un texte :

```

```

Placer un programme java (« applet ») dans le document :

```
<applet code="testFTP2.class" codebase =  
"http://www.lifl.fr/applets" width="320" height="200">  
  <param name="arg1" value="13">  
  <param name="arg2" value="texte dans une String">  
</applet>
```

Un protocole pour transporter des documents Hypertextes

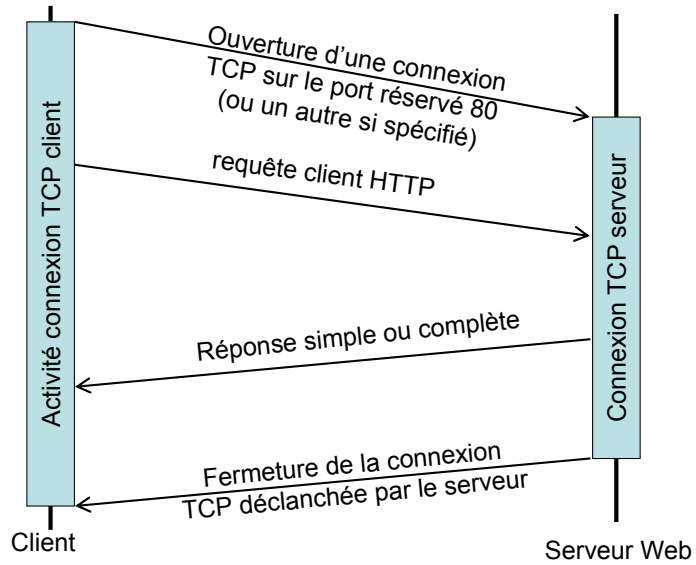
Pourquoi définir un nouveau protocole :

Initialement, permettre que n'importe quel serveur connecté au réseau IP puisse délivrer des documents textes de la manière la plus simple possible aux clients qui les réclament.
⇒ HTTP 0.9

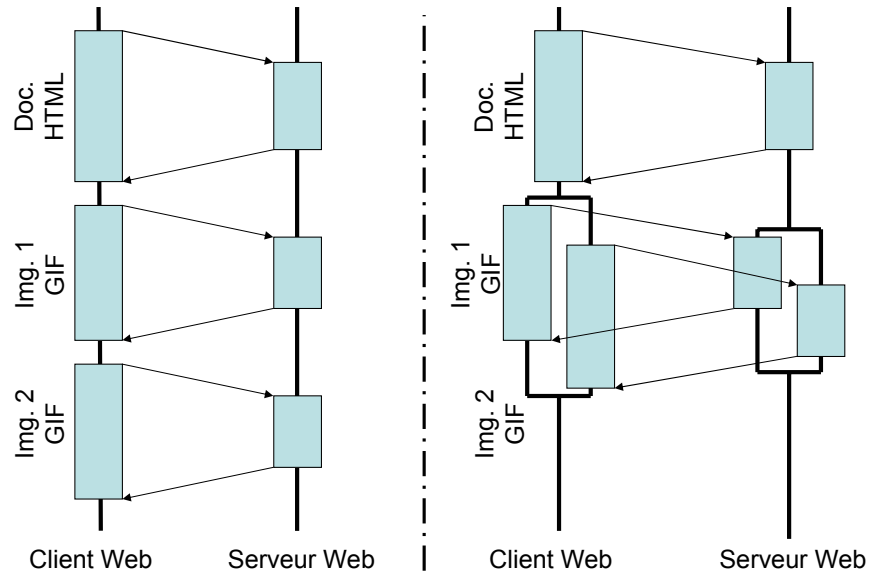
Aujourd'hui HTTP est pensé comme un support pour la transmission de documents distribués et multimédia à travers un système d'information multi-utilisateurs.

HTTP : *HyperText Transfert Protocol* : Protocole de transfert d'hypertextes

HTTP sur TCP/IP : architecture des serveurs Web



Cas de fonctionnement d'un service Web



les requêtes du client HTTP

Format des requêtes :

Trois requêtes de base :

```
GET <SP> <URL> <SP> [HTTP/1.0 <CRLF>
{"Authorization: Basic QWxhZGRpbjpvcGVuIHNlc2FtZQ==" | // utilisateur / mot de passe
"From:" adress_mail | // adresse de l'utilisateur
"If-modified-since:" date | // notification 2 changement ?
"Refered:" URL | // source du lien.
"User-Agent:" id de client Web) ] // nom du client (navigateur)

HEAD <SP> <URL> <SP> [HTTP/1.0 <CRLF>
{"Authorization: Basic QWxhZGRpbjpvcGVuIHNlc2FtZQ==" | // utilisateur / mot de passe
"From:" adress_mail | // adresse de l'utilisateur
"Refered:" URL | // source du lien.
"User-Agent:" id de client Web) ] // nom du client (navigateur)

POST <SP> <URL> <SP> [HTTP/1.0 <CRLF>
{"Authorization: Basic QWxhZGRpbjpvcGVuIHNlc2FtZQ==" | // utilisateur / mot de passe
"From:" adress_mail | // adresse de l'utilisateur
"If-modified-since:" date | // notification 2 changement ?
"Refered:" URL | // source du lien.
"User-Agent:" id de client Web | // nom du client (navigateur)
"Content-Type:" TypeDeMedia | // type des données transmises
"Content-Length:" TailleDuContenu | // taille des données
"Content-Encoding:" typeDeCodage } <CRLF> // exemple x-gzip
{OCTET} // corps des données transmises
```

les requêtes du client HTTP

Format des requêtes :

Autres commandes parfois implantées :

PUT <SP> URL <CRLF> document au format MIME

DELETE <SP> URL <CRLF> HTTP/1.0 <CRLF> en-têtes

Requêtes génériques :

Méthode <SP> URL <SP> HTTP/1.0 <CRLF> en-têtes et
document MIME

Format des réponses du serveur Web

Pour les serveurs « HTTP 0.9 » :
Document HTML

Depuis « HTTP 1.0 » :
"HTTP/" une DIGIT "." <SP> trois chiffres <SP> raison <CRLF>
en-tête générale <CRLF>
en-tête réponse <CRLF>
en-tête entité <CRLF>
corps entité

Les trois chiffres désignent un code d'erreur.

En-tête entité pour un document HTML :

`Content-Type:text/html`

Le corps est alors un document HTML.

Codes d'erreur HTTP

Listes des erreurs prédéfinis :

Code ; Raison	description
"200" ; OK	OK
"201" ; Created	Créé
"202" ; Accepted	Accepté
"204" ; No Content	pas de contenu
"301" ; Moved Permanently	Changement définitif
"302" ; Moved temporarily	Changement temporaire
"304" ; Not modified	non modifié
"400" ; Bad request	requête incorrecte
"401" ; Unauthorized	non autorisé
"403" ; Forbidden	Interdit
"404" ; Not Found	Non trouvé
"500" ; Internal server error	Erreur interne du serveur
"501" ; Not implemented	Non implanté
"502" ; Bad Gateway	Erreur de routeur
"503" ; Service Unavailable	Indisponible

Ou autres code 3 digits + texte retour d'erreur

Entête de réponse d'un serveur HTTP

Entête HTTP

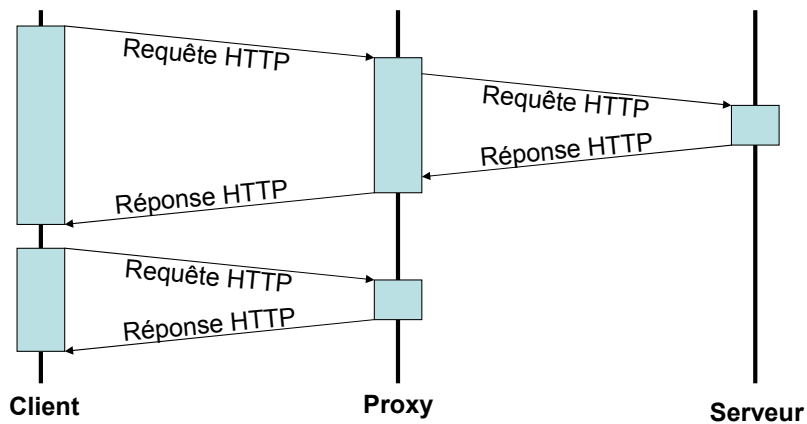
```
HTTP/1.0 <SP> 200 <SP> OK <CRLF>
Server: <SP> Apache/1.3.22 (Unix) PHP/4.0.6 <CRLF>
Date: <SP> Sun, 28 Apr 2002 20:57:46 GMT <CRLF>
Expires: <SP> Tue, 08 May 2000 13:41:16 GMT <CRLF>
Last-Modified: <SP> Sun, 02 Feb 2002 20:57:46 GMT <CRLF>
Content-Type: <SP> text/html <CRLF>
Content-Length: <SP> 2717 <CRLF>
```

corps

```
<html>
  <head>
    <title>toto</title>
    <meta http-equiv="Content-Type"
      content="text/html; charset=iso-8859-1">
    <meta name="keywords" content="motclef1 motclef2">
    ...
```

Le corps est une suite d'octets. Si on a dans l'en-tête la ligne content-Type: image/gif
Le corps du document sera la suite des octets qui constitue l'image au format GIF.

Tunnel (ou Proxy) HTTP



Les requêtes HTTP ne sont plus envoyées au serveur, mais à une machine proxy qui relaie les messages HTTP.

Intérêt :

- Tunnel d'accès à l'Internet (pour contrôle) : Protection ;
- Cache des documents les plus consultés : Performance.

Documentation relative au web

- Protocole HTTP 1.0 et HTTP 1.1
RFC 1945 et RFC 2616
- grammaire des URL, URI et URN :
RFC 1738
- Le Web et le Langage HTML :
<http://www.w3c.org>
- Valideur de code en ligne :
<http://validator.w3.org/>