

## Le service web

### 1. Introduction

Le Web (World wide web – la toile d’araignée modiale ou www ou w3) est un service d’internet permettant d’accéder à des documents au format HTML en utilisant pour la connexion et les échanges le protocole http (Hyper Text Transfer Protocol).

Le web a été inventé, en 1989, au CERN (Centre Européen de recherche Nationale) par Tim Berners-Lee, Robert Cailliau et leur équipe et constitue le service le plus populaire d’Internet. Ses principales caractéristiques :

- Une délocalisation totale du contenu, puisque les documents sont accessibles indifféremment depuis n’importe quel nœud du réseau.
- La possibilité de faire référence à n’importe quel objet ou document au sein d’un autre document (documents hypermedias).

### 2. Les URLs

Les documents sont accessibles par des URL (Uniforme Ressources Locator) comportant le nom du serveur http contenant le document, le chemin d’accès au document et le nom de celui-ci.

La syntaxe générale d’une URL est de la forme :

<http://machine:port/path#ancree>

Champs	Description
<i>machine</i>	Nom d'hôte (ou adresse IP) possédant la ressource à accéder. Ce nom d'hôte est divisé en domaine et sous domaine  Par exemple, www.isetmd.rnu.tn - est un serveur Web (www) situé en tunisie (tn) dans le sous-domaine universitaire (rnu) et il se nomme isetmd
<i>Port</i>	indique le numéro de port. Par défaut c’est le port TCP 80. Un port non standard peut être utilisé. Il doit être précisé dans l’URL. Exemple : <a href="http://www.site-exemple.com:8080">http://www.site-exemple.com:8080</a>
<i>path</i>	Permet de spécifier l’emplacement du fichier sur le serveur. Ce champ est en général constitué d’une suite de répertoires

	séparés par des '/' puis du nom du fichier à accéder.
<i>#ancree</i>	Détermine un endroit précis de la page Web.

### 3. Le protocole HTTP

#### 3.1. Généralités

HTTP est un protocole de transfert hypertexte (Hypertext Transfer Protocol), permettant essentiellement le transfert des fichiers HTML entre un client (navigateur) et un serveur web.

HTTPS (Hypertext Transfer Protocol secure) est la variante sécurisée du http permettant entre autre de :

- Vérifier l'identité du site visité grâce à un certificat d'authentification
- Assurer la confidentialité des données (communications chiffrées)

HTTPS est utilisé dans les applications de transactions financières en ligne (commerce électronique, banque en ligne, etc.), et de consultation de données privées (les webmail), etc.

Par défaut, les serveurs HTTPS sont connectés au port TCP 443.

#### 3.2. Méthodes HTTP

Les méthodes HTTP les plus fréquemment utilisées sont : GET, POST et HEAD

Méthode	Description
<b>Get</b>	Permet de demander un document
<b>Post</b>	Permet de transmettre des données (d'un formulaire par exemple) à l'URL spécifiée dans la requête. L'URL désigne en général un script Perl, PHP
<b>Head</b>	Permet de ne recevoir que les lignes d'en-tête de la réponse, sans le corps du Document

### 3.3. Requête et réponse http

#### 3.3.1. Requête http

La requête transmise par le client au serveur comprend :

- Une **ligne de requête** (*request-line*) contenant la méthode utilisée, l'URL du service demandé, la version utilisée de http
- Une ou plusieurs lignes d'en-têtes, chacune comportant un nom et une valeur.

```
GET /accueil.html HTTP/1.1
Host: www.isetmd.rnu.tn
Accept: */*
Accept-Language: fr
User-Agent: Mozilla/9.0 (MSIE 8.0; windows 7)
Connection: Keep-Alive
```

Dans cet exemple, le client demande le document à l'adresse <http://www.isetmd.rnu.tn/index.html>, il accepte tous les types de document en retour, préfère les documents en français, utilise un navigateur compatible Mozilla 9.0 sur un système *Windows 7*) et signale au serveur qu'il faut garder la connexion TCP ouverte à l'issue de la requête (car il a d'autres requêtes à transmettre).

#### 3.3.2. Réponse http

La réponse transmise par le serveur au client comprend :

- une ligne de statut (*status-line*) contenant la version de HTTP utilisée et un code d'état
- une ou plusieurs lignes d'en-têtes, chacune comportant un nom et une valeur
- Le corps du document retourné .

```
HTTP/1.1 200 OK
Date: Mon, 25 May 2015 20:11:33 GMT
Server: Microsoft-IIS/7.5
Cache-Control: max-age=600
Expires: Mon, 25 May 2015 20:21:33 GMT
Last-Modified: Thu, 30 Oct 2014 14:24:46 GMT
ETag: "26206-5b0-3af1f126"
Accept-Ranges: bytes
Content-Length: 1456
Content-Type: text/html
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0
Transitional//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-
transitional.dtd">
<html>
```

Dans cet exemple, le code 200 indique que le document demandé a été trouvé.

Pour faciliter la gestion du cache du client, le serveur transmet la date actuelle, la date de dernière modification du document et la date d'expiration (après laquelle le document doit être demandé à nouveau).

L'en-tête Content-Type indique que le document retourné est de type HTML et l'en-tête Content-Length indique que le corps du document a une longueur de 1456 octets.

L'en-tête Server renseigne sur le logiciel serveur utilisé. L'envoi d'une telle information n'est pas recommandé pour des raisons de sécurité.

### 3.4. Statut des requêtes

Lorsque le serveur renvoie un document, il lui associe un code de statut renseignant ainsi le client sur le résultat de la requête (requête invalide, document non trouvé...).

Les principales valeurs des codes de statut HTTP sont détaillées dans le tableau ci-après

Code	Nom	Description
<b>Information 1xx</b>		
100	continue	le client peut envoyer la suite de la requête
<b>Succès 2xx</b>		
200	OK	Le document a été trouvé et son contenu suit
201	Created	Le document a été créé en réponse à un PUT
202	Accepted	Requête acceptée, mais traitement non terminé
<b>Redirection 3xx</b>		
301	Moved	Le document a changé d'adresse de façon permanente
302	Found	Le document a changé d'adresse temporairement
304	Not modified	Le document demandé n'a pas été modifié
<b>Erreurs du client 4xx</b>		
400	Bad request	La syntaxe de la requête est incorrecte
401	Unauthorized	Le client n'a pas les privilèges d'accès au document
403	Forbidden	L'accès au document est interdit
404	Not found	Le document demandé n'a pu être trouvé
<b>Erreurs du serveur 5xx</b>		
500	Internal error	Une erreur inattendue est survenue au niveau du serveur
501	Not implemented	La méthode utilisée n'est pas implémentée

502	Bad gateway	Erreur de serveur distant lors d'une requête proxy
-----	-------------	--

#### 4. http et le cache

Une page HTML est souvent constituée d'informations (textes ou images) ne subissant aucune modification pendant plusieurs jours.

Il devient alors intéressant de mettre en cache les objets pour lesquels nous avons déjà effectué une requête.

##### 4.1. Cache du navigateur

Les browsers actuels possèdent normalement tous la fonctionnalité de cache. Ce cache va permettre au navigateur de fournir immédiatement les objets présents dans le cache sans avoir à effectuer une requête auprès du serveur possédant l'objet..

##### 4.2. Cache du proxy

Le proxy cache, en plus d'un relais pouvant servir de filtre ou de firewall, archive les pages, lors d'une requête. S'il possède déjà la page, il la renvoie, sinon il va la chercher

- le client effectue une requête (GET) via un proxy cache
- le proxy vérifie s'il dispose de la page demandée
- si oui
  - le proxy vérifie la date d'expiration de l'URI (Expires)
  - interroge le serveur de l'URI afin de comparer les dates
    - de dernière modification de l'URI dans le cache
    - de dernière modification de l'URI sur le serveur
- si l'URI du cache est à jour alors le proxy retourne la ressource de son cache
- le proxy récupère la ressource du serveur, l'archive et la retourne au client.

#### 5. Serveurs et clients web

Un serveur HTTP n'est rien d'autre qu'un serveur de fichiers à l'écoute sur un port TCP (généralement le 80), et qui, lorsqu'il reçoit une demande de connexion suivie d'une requête (de la forme GET /chemin/nom.de.fichier) répond en expédiant le fichier demandé.

Les serveurs http les plus courants sont *Apache http Server* (logiciel libre et multiplateforme), *Apache Tomcat* (évolution de apache pour J2EE), *Internet Information Service* (serveur web de Microsoft).

Les clients http sont les navigateurs web. Il existe de nombreux navigateurs web, pour toutes sortes de matériels (ordinateur personnel, tablette tactile, téléphones mobiles, etc.) et pour différents systèmes d'exploitation (GNU/Linux, Windows, Mac OS, iOS et Android). Les plus utilisés à l'heure actuelle sont Google Chrome, Mozilla Firefox, Internet Explorer, Safari et Opera

#### **6. Installation et configuration du serveur web apache**

Les détails seront explicités dans la séance de TP correspondante à ce cours.