Chapitre 3
Fonctionnement d'une application web

1. Introduction

Le *World Wide Web*, ou plus simplement le Web, est un système de consultation de l'information reposant sur plusieurs éléments : le principe d'hypertexte (ou hyperlien), les URL (*Uniform Resource Locator*), le protocole de communication HTTP (*HyperText Transfer Protocol*) et le langage informatique nommé HTML (*HyperText Markup Language*). La communication sur le Web s'effectue sur un mode de transmission client-serveur.

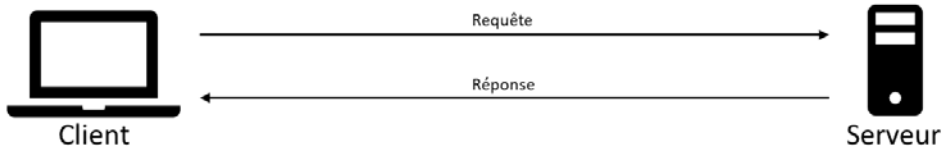
Remarque

Tim Berners-Lee est l'inventeur du Web tel que nous le connaissons aujourd'hui. Travaillant au CERN, l'Organisation européenne pour la recherche nucléaire, dans les années 1980, il proposa un projet permettant d'échanger de l'information de manière plus efficace. Les ordinateurs étaient déjà interconnectés entre eux à cette époque, mais il était difficile de trouver une information ou un document présent sur le réseau.

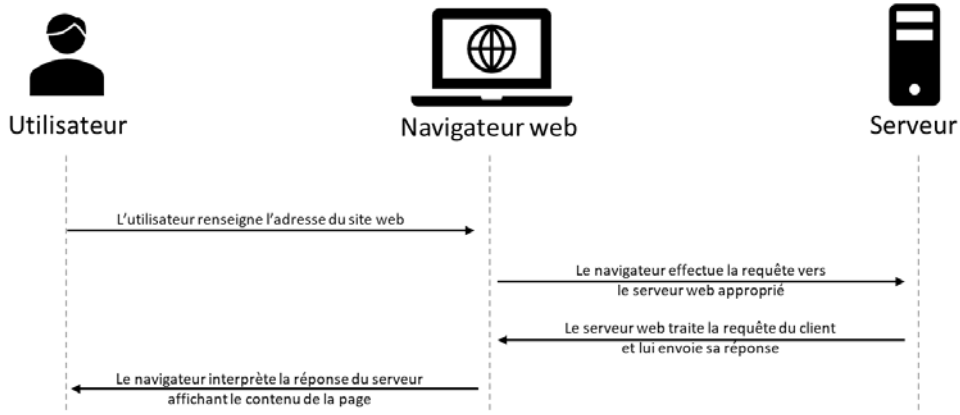
Avant d'aborder les aspects techniques de la sécurité des applications web, il est important de bien comprendre le fonctionnement du Web, car la présence de vulnérabilités est généralement le résultat d'une méconnaissance de ces mécanismes. De plus, l'exploitation de ces faiblesses, ainsi que leurs protections, s'appuie tout autant sur ces mêmes mécanismes. Comprendre le fonctionnement du Web facilite grandement l'appréhension des aspects de sécurité.

2. Le modèle client-serveur

Le modèle client-serveur est un mode de transmission de l'information entre deux entités. La première entité, le client, émet des requêtes vers le serveur. Le serveur, quant à lui, attend des requêtes provenant des clients, les traite et y répond.



Lorsqu'un internaute consulte une page web, le client, généralement un navigateur internet, va effectuer une requête au serveur hébergeant la page demandée. Le serveur traite la demande, puis envoie sa réponse au client. Ensuite, le client interprète cette réponse et affiche le résultat à l'utilisateur.



Remarque

En réalité, il n'y a pas qu'une seule requête envoyée pour afficher le contenu d'une page web, mais plutôt une requête pour chaque ressource présente sur la page. Une ressource peut être la page elle-même, une image, une vidéo, un fichier audio, etc.

2.1 Le navigateur web

Le navigateur web est un logiciel qui permet de consulter les informations présentes sur le Web. Il est souvent installé par défaut sur les ordinateurs des utilisateurs : Edge pour le système d'exploitation Microsoft Windows, Firefox pour le système Ubuntu (une distribution Linux) ou encore Safari pour MacOS par exemple. Bien sûr, rien n'interdit d'installer un autre navigateur afin de remplacer le premier, voire même d'en installer plusieurs en parallèle.

Un navigateur web est composé de plusieurs fonctionnalités. Il permet à l'utilisateur de renseigner les adresses des sites web à consulter, grâce à la barre d'adresse, de visualiser leurs contenus en s'aidant d'onglets pour mieux s'y retrouver, de gérer ses sites favoris ou encore d'étendre les capacités de navigation grâce à la notion d'extensions (ou *plugins*).

Techniquement, le navigateur va effectuer les requêtes auprès de différents serveurs web, puis interpréter les réponses de ces derniers afin d'afficher le contenu à l'utilisateur.

Remarque

Les navigateurs web font partie d'un ensemble plus vaste d'applications nommées clients web. Un client web est un logiciel capable d'envoyer des requêtes HTTP à un serveur web et d'en traiter le résultat.

2.2 Le serveur web

Mis à part son rôle, un serveur web n'est pas une machine très différente d'un ordinateur de bureau. Il est destiné à stocker les fichiers nécessaires au bon fonctionnement du site web et à répondre aux demandes des visiteurs. C'est tout de même une machine souvent plus puissante qu'un ordinateur de bureau et il travaille rarement seul. En effet, une seule machine serveur ne peut répondre à une volumétrie très importante de demandes, il est donc possible de multiplier les serveurs web afin de répartir la charge de travail.

Au début du Web, le contenu des pages renvoyées par le serveur était statique. Le serveur ne faisait que transmettre le contenu des fichiers (représentant le contenu de la page web), stocké dans sa mémoire, sans aucun traitement spécifique. Ce contenu était donc le même pour tous les utilisateurs désirant accéder à ces pages, ce qui n'est plus le cas de nos jours.

Remarque

Cela ne signifie pas qu'il n'est plus possible de proposer des pages statiques de nos jours, mais que les besoins et les possibilités ont évolué.

3. Le principe d'hypertexte

Derrière ce nom un peu barbare se cache la possibilité de naviguer de document en document, qu'ils soient présents sur le même site web ou non, grâce à des liens cliquables.

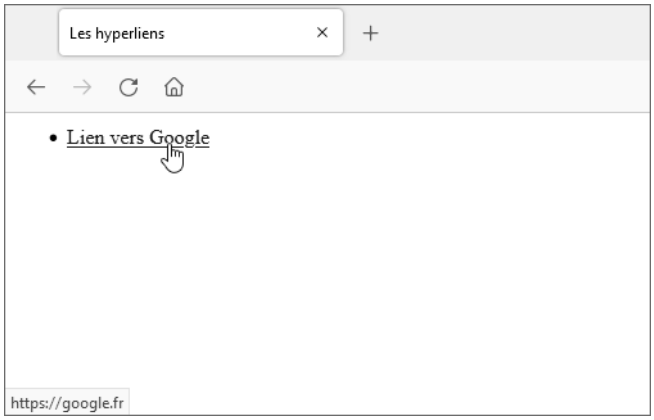
Un lien hypertexte (ou hyperlien, ou plus simplement, lien) est donc une manière d'accéder à un document qui peut être hébergé sur n'importe quel serveur. Mais comment localiser l'emplacement du document que l'utilisateur souhaite consulter ? Cette problématique est résolue par la notion d'adresse web ou plus exactement d'URL.

Remarque

Le principe d'hypertexte a été inventé par Ted Nelson dans les années 1960 dans le cadre du projet Xanadu. Ce projet avait pour but de partager des données informatiques de façon instantanée tout en respectant le principe de droit d'auteur et de suivi de modifications. Tim Berners-Lee a donc repris ce travail en l'adaptant à son projet pour en faire des hyperliens, mais l'amputant au passage des deux derniers principes.

Les Uniform Resource Locator (URL)

Les URL (*Uniform Resource Locator*) sont des chaînes de caractères qui permettent d'identifier et de localiser un document (ou plus généralement, une ressource) présent sur la toile. Un hyperlien référence une URL permettant ainsi de connaître la destination à atteindre lorsque l'utilisateur clique sur celui-ci. Les navigateurs indiquent, généralement en bas à gauche de la fenêtre, l'URL de destination lors du survol de l'hyperlien avec la souris.



Cela facilite grandement la consultation de documents puisque cela permet de naviguer de document en document. Il est également possible de créer un document listant une multitude de documents comme le ferait une table des matières par exemple.

La syntaxe d'une URL

Une URL possède une syntaxe spécifique qui peut être décomposée de la façon suivante :

- **Le protocole** : spécifie le protocole qui sera utilisé afin d'accéder à la ressource. Il existe plusieurs protocoles, mais le principal protocole du Web est le protocole HTTP (et sa version sécurisée HTTPS).
- **Le nom de domaine** : il s'agit d'un identifiant indiquant le serveur (ou groupe de serveurs) à contacter. Un nom de domaine est en fait composé de plusieurs éléments hiérarchisés, le plus haut dans la hiérarchie étant l'élément le plus à droite :
  - L'élément le plus à droite est nommé domaine de premier niveau (TLD pour *Top-Level Domain*) ou parfois extension (par analogie à l'extension des noms de fichiers). Dans l'exemple ci-dessus, le TLD est **.com**, mais il en existe bien d'autres : .fr, .org, .net...

  - Le domaine principal (ou dit de second niveau) est ici **exemple**, il se situe donc directement à gauche du TLD.
  - Peut venir ensuite une liste de sous-domaines, dans l'exemple **www**, qui peuvent être de troisième, de quatrième niveau, etc. En fait, chaque domaine est un sous-domaine d'un domaine supérieur, y compris pour le domaine de premier niveau qui est sous domaine de l'élément nommé racine représentée par un point (.), placé à droite du TLD. Une telle représentation est nommée FQDN (*Fully Qualified Domain Name*), mais n'est généralement pas représentée dans les URL.
- **Le chemin de la ressource** : indique l'emplacement de la ressource sur le serveur en question.

Il est possible de spécifier d'autres informations au niveau de l'URL dépendant du contexte de consultation ou de la configuration du serveur.

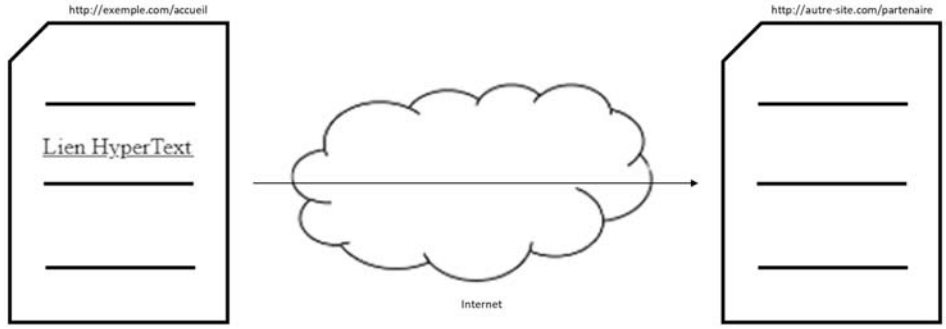
- **Un numéro de port** : il s'agit d'un nombre compris entre 1 et 65 535. Ce nombre permet à un serveur d'héberger et d'offrir plusieurs services en parallèle où chaque service a son propre numéro de port. Par convention, les serveurs web utilisant le protocole HTTP écoutent sur le port 80, et ceux utilisant la version sécurisée HTTPS écoutent sur le port 443. Les navigateurs connaissent bien cela, c'est pour cette raison qu'il n'est pas nécessaire de le préciser lors de la saisie de l'URL dans la barre d'adresse ou dans un hyperlien. En admettant un serveur web écoutant en HTTP sur le port non conventionnel 8081, l'URL sera alors :
  http://www.exemple.com:8081/chemin/de/la/ressource.
- **Des identifiants de connexion** : certains services peuvent nécessiter de renseigner un nom d'utilisateur et un mot de passe de la façon suivante :
  http://nom-d'utilisateur:mot-de-passe@www.exemple.com/chemin/de/la/ressource. Bien que cette façon de procéder n'est plus réellement utilisée de nos jours, cette possibilité existe encore et est parfois utilisée lors de l'exploitation de vulnérabilités afin de contourner certaines restrictions.
- **Des paramètres d'URL** : les paramètres d'URL sont des informations supplémentaires envoyées au serveur au format : nom du paramètre = valeur du paramètre. Le début de la liste des paramètres est marqué par le caractère ?, puis, chaque élément est séparé par le caractère & :
  http://www.exemple.com/chemin/de/la/ressource?parametre1=valeur1&parametre2=valeur2&parametre3=valeur3.

– **Des fragments d'URL** : les fragments d'URL ressemblent aux paramètres d'URL à la différence près que les valeurs spécifiées ne sont pas envoyées au serveur. Le caractère marquant le début de la liste n'est plus le signe ? mais le signe #. Ils peuvent par exemple être utilisés par le navigateur en tant qu'ancre comme le ferait un marque-page dans un livre. Une telle URL peut être la suivante : http://www.exemple.com/chemin/de/la/ressource#parametre1=valeur1&parametre2=valeur2.

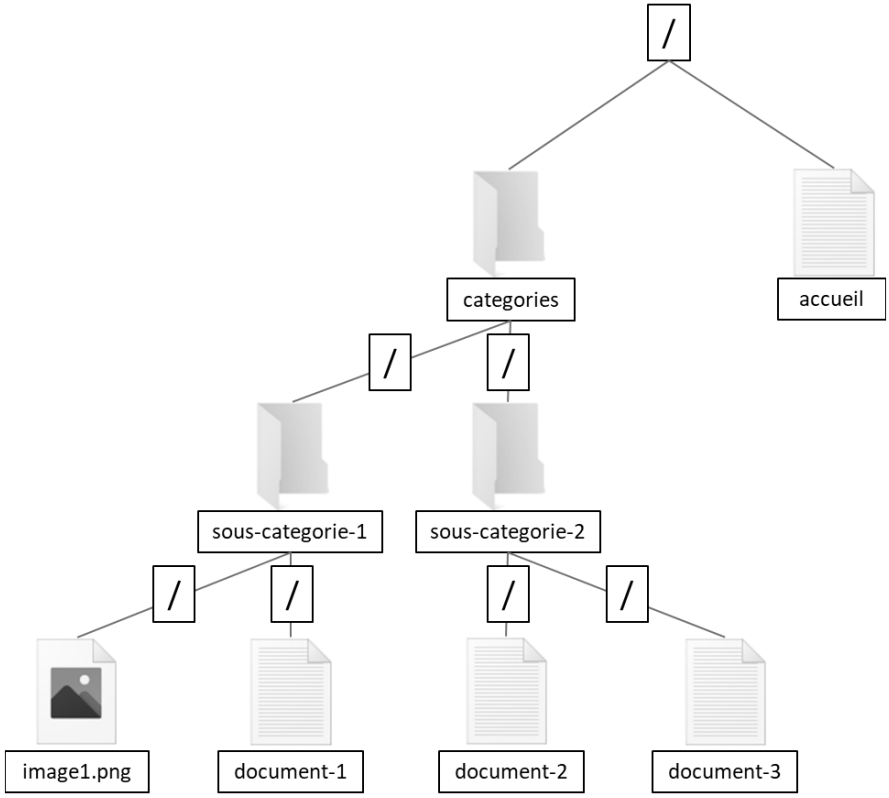
URL absolue et URL relative

Lorsque le développeur en charge du site web souhaite référencer un document à partir d'un autre document web, il lui est possible de le faire de deux manières différentes : en utilisant une URL absolue ou une URL relative.

Une URL absolue décrit complètement l'adresse permettant d'accéder à un document : l'URL http://www.exemple.com/chemin/de/la/ressource est une URL absolue. Une telle URL va permettre à un document hébergé sur un premier site, de référencer un document sur un autre site via un hyperlien. Par exemple, le document http://exemple.com.com/accueil qui inclut un lien vers le document d'un autre site http://autre-site.com/partenaire.



Si les documents sont hébergés sur le même serveur, il est alors possible (ce n'est pas une obligation), d'utiliser des URL relatives afin de les référencer. Les URL relatives prennent en compte l'emplacement actuel du document souhaitant en référencer un autre. Dans le cas du site http://exemple.com, son arborescence peut être la suivante :



Si le document nommé **accueil** souhaite référencer le document nommé **document-1**, alors l'hyperlien pourra utiliser l'URL absolue suivante : http://exemple.com/categories/sous-categorie-1/document-1. Mais puisque ce document est présent dans la même arborescence, il est également possible d'utiliser une URL relative. Le document **accueil** n'étant pas au même niveau de hiérarchie que le document **document-1**, l'URL relative à utiliser est alors **categories/sous-categorie-1/document-1**.