

Qu'est ce que le web?

SNT

Lycée Saint Sernin

Problème

Qu'est ce que le web et quels sont ses principes fondamentaux?

Thème Internet - Programme officiel

Contenus	Capacités attendues
Repères historiques ①	Connaître les étapes du développement du Web.
Notions juridiques ②	Connaître certaines notions juridiques (licence, droit d'auteur, droit d'usage, valeur d'un bien).
Hypertexte ①	Maîtriser les renvois d'un texte à différents contenus.
Langages HTML et CSS ①	Distinguer ce qui relève du contenu d'une page et de son style de présentation. Étudier et modifier une page HTML simple.
URL ①	Décomposer l'URL d'une page. Reconnaître les pages sécurisées.
Requête HTTP ①	Décomposer le contenu d'une requête HTTP et identifier les paramètres passés.
Modèle client/serveur ②	Inspecter le code d'une page hébergée par un serveur et distinguer ce qui est exécuté par le client et par le serveur.
Moteurs de recherche : principes et usages ②	Mener une analyse critique des résultats fournis par un moteur de recherche. Comprendre les enjeux de la publication d'informations.
Paramètres de sécurité d'un navigateur ① ②	Maîtriser les réglages les plus importants concernant la gestion des cookies, la sécurité et la confidentialité d'un navigateur. Sécuriser sa navigation en ligne et analyser les pages et fichiers.

Sommaire

1. Définition
2. Repères historiques
3. L'activité sur le web
4. Les piliers du Web
5. Le protocole HTTP

Définition

Géolocalisation

Définition

Définition Larousse :

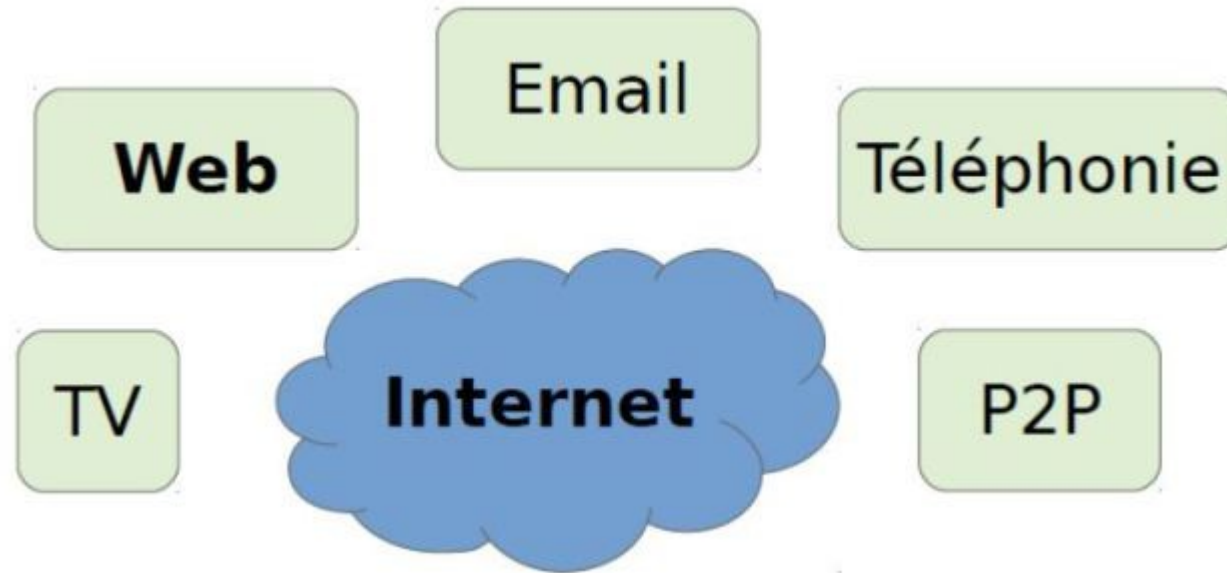
WEB [wɛb] n.m. (abrév. de l'angl. *world wide web*, réseau mondial). Système hypermédia permettant d'accéder aux ressources du réseau Internet. (S'écrit aussi sans majuscule.) SYN. : *la Toile*, *WWW*. ◇ *Page Web* → **1. page**. – *Site Web* → **site**.

Définition référentiel:

« Le web (toile ou réseau) désigne un système donnant accès à un ensemble de données (page, image, son, vidéo) reliées par des liens hypertextes et accessibles sur le réseau internet. »

Le **web** est donc une application utilisant le réseau internet, au même titre que d'autres.

Autres applications internet



Le **web** est donc une application utilisant le réseau internet, au même titre que d'autres.

Repères historiques

Géolocalisation

Activité 1 - Repères Historiques

A l'aide des vidéos suivantes : <https://www.youtube.com/watch?v=hrzXdKvG1CY>

<https://tube.nuagelibre.fr/videos/watch/7ebf14af-7686-4171-bbdb-e65df26259e6?start=36s>

Et de la frise suivante :

https://www.tiki-toki.com/timeline/embed/137139/6372410394/#vars!date=1988-09-06_12:57:55!

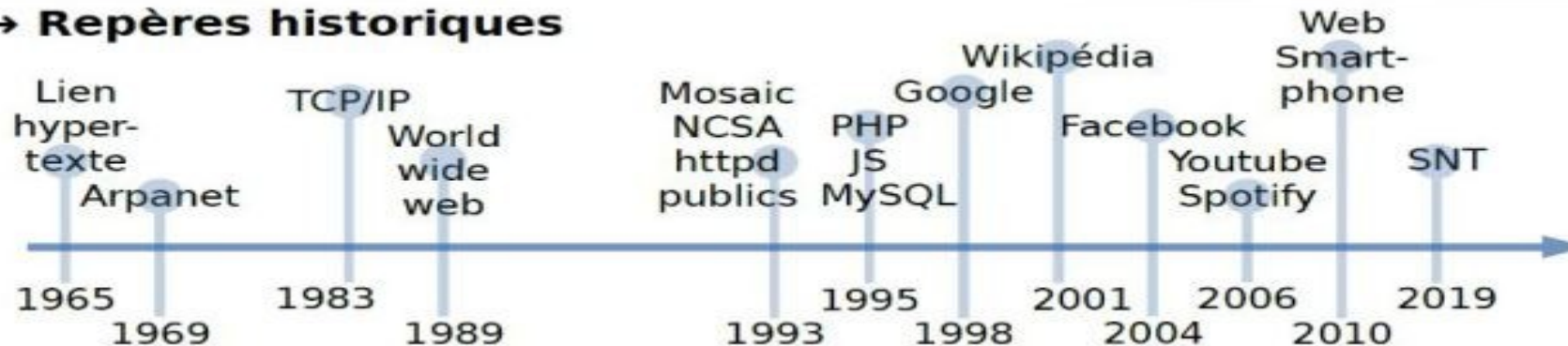
Veuillez remplir une frise des repères historiques.

Activité 1 - Correction

Repères historiques :

- **1965** : invention concept hypertexte (Ted Nelson)
- **1969** : lancement projet ARPANET
- **1983** : adoption de TCP/IP sur ARPANET
- **1989** : naissance web au CERN (Tim Berners Lee)
- **1993** : navigateur et serveur dans domaine public
- **1995** : technologies pour web 2.0 interactif (JS, PHP, MySQL)
- **1998** : lancement de Google
- **2001** : standardisation pages web (DOM) + Wikipédia
- **2004** : Facebook
- **2005** : Youtube
- **2006** : Twitter
- **2010** : technologies web pour s'adapter aux smartphones
- **2019** : SNT

→ Repères historiques



L'activité sur le web

Géolocalisation

L'activité sur le web

Utilisateurs
~4 milliards

Sites
~1,5 milliards*

* dont 75 % sont inactifs

Web public ~5% (Google, Facebook)

Deepweb ~90% (derrière pwd)

Darkweb ~5% (TOR)

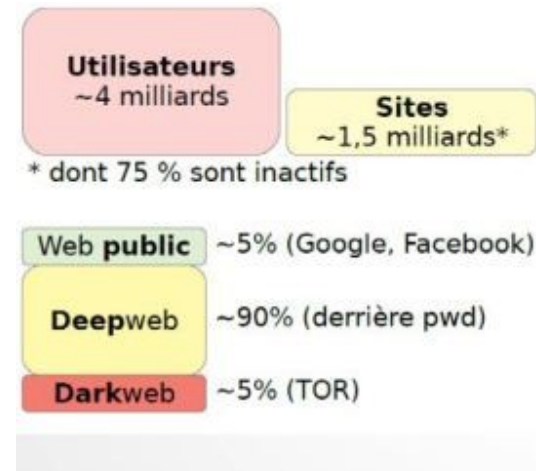
Visibilité :



Sur un Seul 5% de l'activité sur le web est du domaine public :

- réseaux sociaux
- commerces en ligne
- moteurs de recherche
- ...

Le reste est du domaine privé, avec plus ou moins de légalité.



Le trafic sur le web

Les bots représentent 42 % du trafic web mondial en 2024, avec une prédominance des acteurs malveillants. D'après le dernier rapport SOTI d'Akamai, les deux tiers des bots en ligne sont malveillants et mènent en priorité des actions de Web scraping, lesquels ont un impact majeur sur le e-commerce.

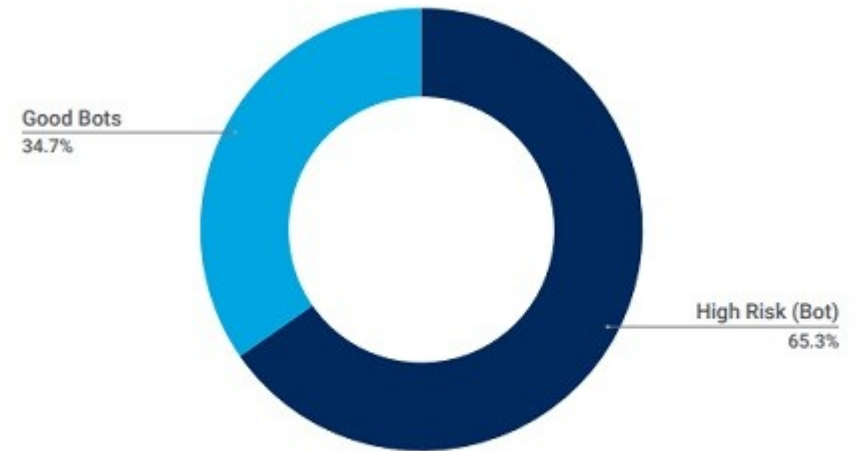
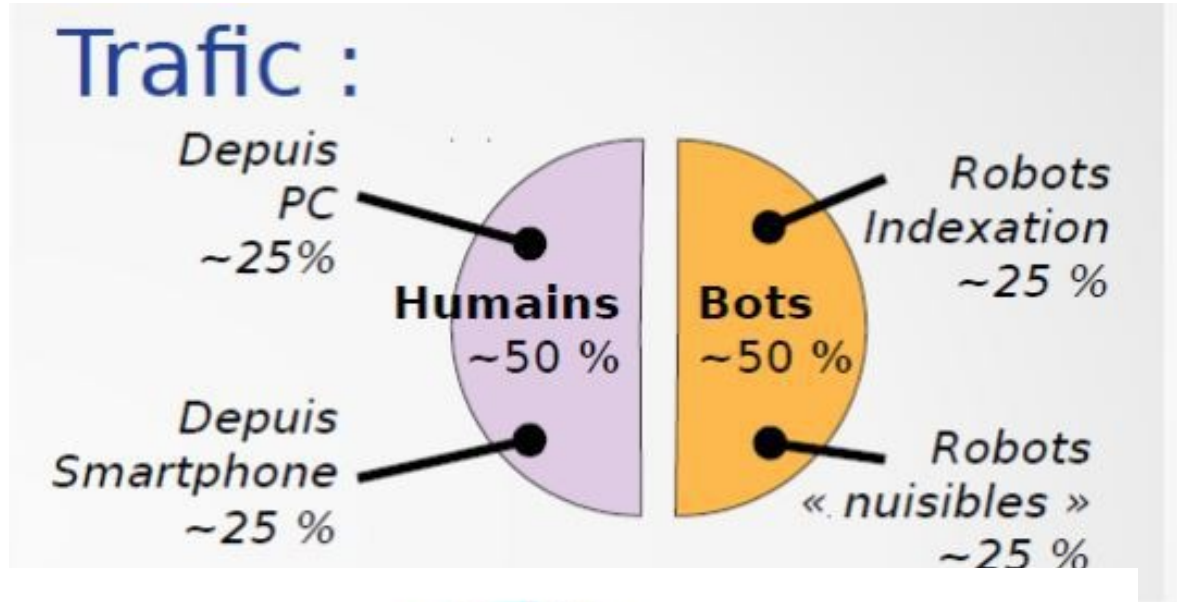


Fig. 8: Good bot traffic vs. bad bot traffic

Le trafic sur le web

Site les plus visités en France

Classement <small>i</small>	Site Web <small>i</small>
1	 google.com
2	 youtube.com
3	 facebook.com
4	 google.fr
5	 amazon.fr
6	 orange.fr
7	 wikipedia.org
8	 instagram.com
9	 leboncoin.fr
10	 chatgpt.com

Les 3 piliers du web

Géolocalisation

Comment fonctionne un site web

Surfer sur le Web, c'est un échange. Le Web est en fait un système d'échange entre un client et un serveur.


- Le client c'est nous ! Ou plutôt notre navigateur Web, qui va nous permettre d'accéder au Web depuis un ordinateur, smartphone ou tablette. Par exemple : Chrome, Firefox, Safari, Internet Explorer, etc...
- Le serveur, c'est un ordinateur puissant qui stocke et héberge des sites Web. C'est sur cet ordinateur que se trouvent les pages Web, donc tous les fichiers du site internet auquel on veut accéder. Le but du serveur web est de servir des clients, d'où le nom « serveur ».

Activité 2- Modèle client serveur

A l'aide de la vidéo suivante : <https://vimeo.com/showcase/3555227/video/138623558>

Veuillez remplir la partie de votre fiche sur le modèle client serveur.

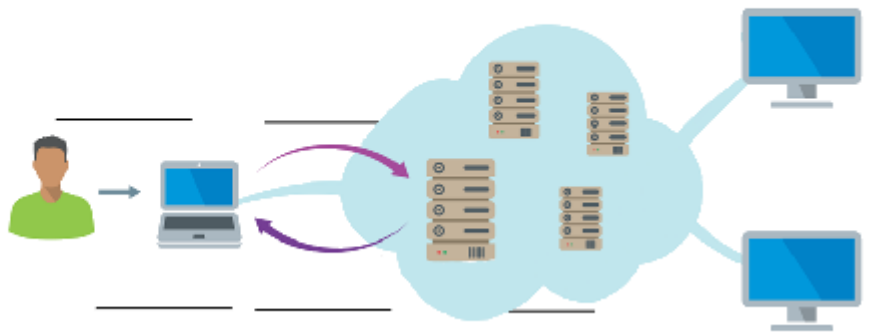
Le modèle client - serveur

 vimeo.com/showcase/3555227/video/138623558

Regarder la vidéo et Compléter le modèle Client - Serveur qui est la base du fonctionnement du Web :

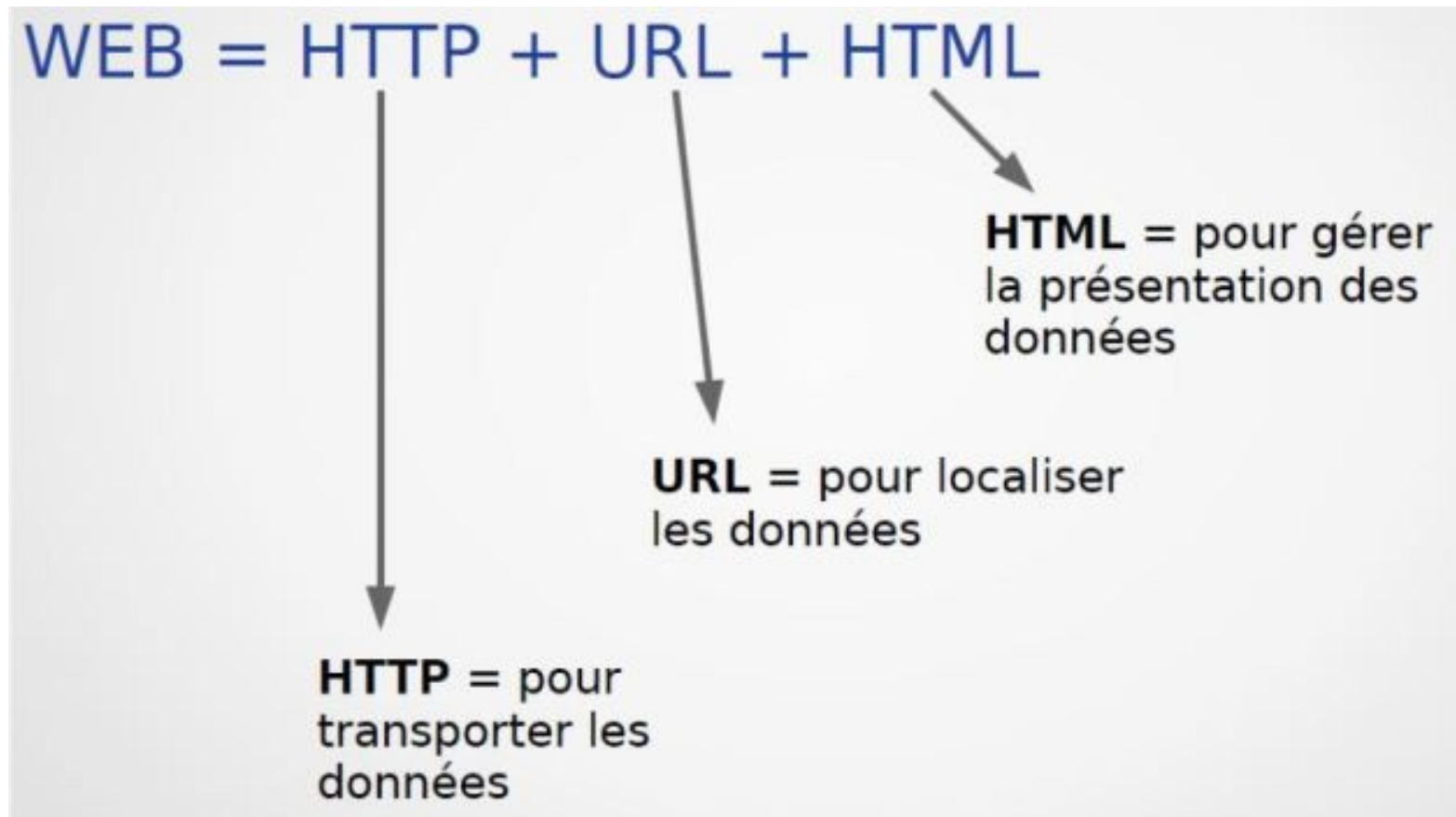
Quel est le rôle du serveur ?

Quel est le rôle du navigateur du client ?



Comment fonctionne un site web

Ces échanges reposent sur 3 grands piliers indispensables à son bon fonctionnement.



Le protocole HTTP

Géolocalisation

Le protocole HTTP

Surfer sur le Web, c'est un échange. Le Web est en fait un système d'échange entre un client et un serveur.

L'acronyme **HTTP** signifie "**H**ypertext **T**ransfert **P**rotocol". Il a été inventé par Tim-Berner Lee dans les années 1990.

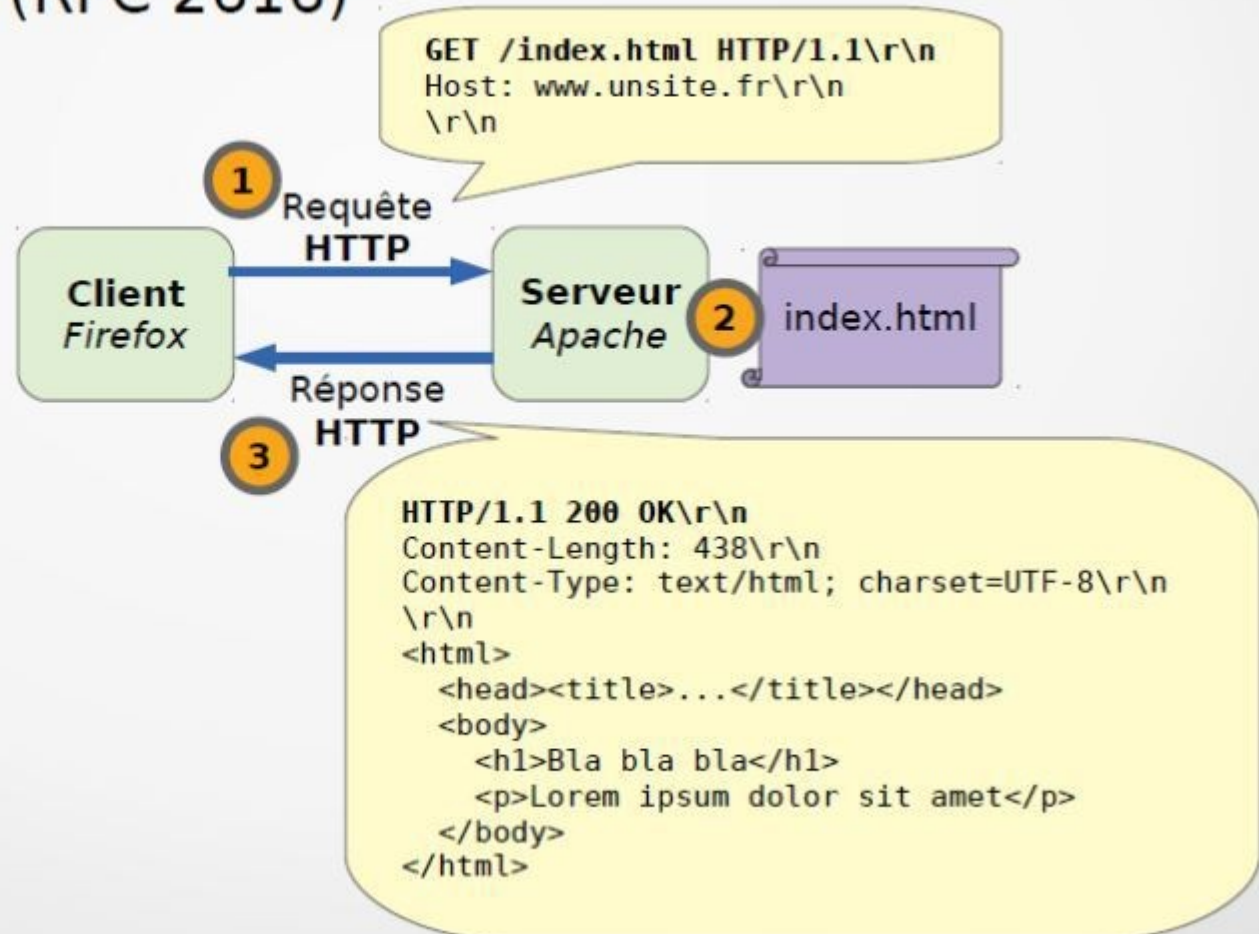
C'est le protocole le plus utilisé sur Internet.

Le but du protocole HTTP est de permettre un transfert de fichiers (essentiellement au format HTML) localisés grâce à une chaîne de caractères appelée URL entre un navigateur (le client) et un serveur Web.

Le navigateur envoie une requête HTTP au serveur qui la décode et renvoie une réponse HTTP au navigateur. Cette requête demande un document (une page HTML, une image, etc...). Le serveur cherche les informations, peut être amené à interpréter les résultats et envoie la réponse. Cette réponse contient les entêtes du protocole HTTP et le contenu demandé.

Le protocole HTTP

- **HTTP = HyperText Transfer Protocol**
- Protocole normalisé (RFC 2616)



Activité 3- la requête HTTP

Faites l'activité sur la requête HTTP sur votre fiche élève

La requête HTTP

Que signifie HTTP ? _____

Visualisez et commentez l'échange entre le client et le serveur lors d'une requête HTTP.

Pour cela ouvrez votre navigateur, puis ouvrez la fenêtre « outil de développement ».

Cliquez sur l'onglet « Réseau ». Il reste maintenant à saisir l'URL dans le navigateur : <http://www.onisep.fr>

Complétez le schéma ci-dessous avec vos observations :

The screenshot shows the Network tab in Chrome DevTools. It lists several requests made by the browser to the domain www.onisep.fr. The first request is a GET for /document.html, which is highlighted with a red circle. Other requests include POST for login, GET for log_recherche_treuve.js, GET for log_orisey.js, GET for ctmap73-CTMAK6W9DZ, GET for ic-grille-mat-0-4-exercitation_carrousel.png, GET for ic-coursupiles-notes-devel-reue-carrousel-image_circuits.png, GET for ic-images-202-2021-cours-upiles-composer-et-reviser-ense, and GET for /... . Each row shows details like status (200), size, and time taken.

The diagram illustrates the HTTP process flow. On the left, there are four empty boxes connected by arrows pointing right towards a central server icon. Below the server icon, there are three numbered steps in French:

- 1 Envoi d'une requête HTTP
- 2 Interprétation des éléments de la page demandée (Textes, images, scripts, ...)
- 3 Renvoi des résultats sous forme HTML / CSS

To the right of the server icon, there is a yellow oval containing icons for various devices (tablet, smartphone, laptop) representing clients. Arrows indicate the flow of data between the server and the clients.

La requête HTTP

Une **requête HTTP** est un ensemble de lignes envoyé au serveur par le navigateur. Elle comprend :

1) **une ligne de requête** : c'est une ligne précisant le type de document demandé, la méthode qui doit être appliquée, et la version du protocole utilisée.

La ligne comprend trois éléments devant être séparés par un espace :

- la méthode
- l'URL
- la version du protocole utilisé par le client (généralement HTTP/1.1)

2) **les champs d'en-tête de la requête** : c'est un ensemble de lignes facultatives permettant de donner des informations supplémentaires sur la requête et/ou le client (Navigateur, système d'exploitation, ...).

Chacune de ces lignes est composée d'un nom qualifiant le type d'en-tête, suivi de deux points (:) et de la valeur de l'en-tête.

Voici un exemple d'en-tête de requête envoyé par un navigateur :

La requête HTTP

Voici un exemple d'en-tête de requête envoyé par un navigateur :

```
GET /index.html HTTP/1.1\r\n
Host: www.unsite.fr\r\n
\r\n
```

La commande **GET** sert à obtenir une ressource. C'est la requête la plus utilisée sur Internet. Il s'agit de la méthode qui doit être appliquée. La commande **POST** est une méthode qui permet d'envoyer le contenu d'un formulaire HTML.

La ligne suivante est un champs HTTP, toujours formatée de cette manière "**NomChamps: Valeur**".

Tout ceci constitue l'en-tête de la requête.

La réponse HTTP

Une réponse HTTP est un ensemble de lignes envoyées au navigateur par le serveur.

Elle comprend:

1) **une ligne de statut** précisant la version du protocole utilisé et l'état du traitement de la requête à l'aide d'un code et d'un texte explicatif.

Cette ligne comprend trois éléments devant être séparés par un espace :

- La version du protocole utilisé
- le code de statut
- la signification du code

2) les **champs d'en-tête de la réponse** : il s'agit d'un ensemble de lignes facultatives permettant de donner des informations supplémentaires sur la réponse et/ou le serveur. Chacune de ces lignes est composée d'un nom qualifiant le type d'en-tête, suivi de deux points (:) et de la valeur de l'en-tête

3) **le corps de la réponse** : il contient le document demandé.

La réponse HTTP

Voici un exemple d'en-tête de réponse envoyé par un serveur :

```
HTTP/1.1 200 OK\r\n
Content-Length: 438\r\n
Content-Type: text/html; charset=UTF-8\r\n
\r\n
<html>
  <head><title>...</title></head>
  <body>
    <h1>Bla bla bla</h1>
    <p>Lorem ipsum dolor sit amet</p>
  </body>
</html>
```

La première ligne de la réponse contient toujours le 'code' HTTP indiquant si la requête a réussi ou pas.

Puis, comme pour la requête, on trouve les lignes des champs HTTP. Tout ceci constitue l'en-tête de la réponse.

Ensuite se trouve le corps de la réponse, qui dans le cas d'un GET d'un fichier HTML contient par exemple le code HTML de la page visée.

La réponse HTTP

Voici un exemple d'en-tête de réponse envoyé par un serveur :

```
HTTP/1.1 200 OK\r\n
Content-Length: 438\r\n
Content-Type: text/html; charset=UTF-8\r\n
\r\n
<html>
  <head><title>...</title></head>
  <body>
    <h1>Bla bla bla</h1>
    <p>Lorem ipsum dolor sit amet</p>
  </body>
</html>
```

La première ligne de la réponse contient toujours le 'code' HTTP indiquant si la requête a réussi ou pas.

Puis, comme pour la réponse, on trouve les lignes des champs HTTP. Tout ceci constitue l'en-tête de la réponse.

Ensuite se trouve le corps de la réponse, qui dans le cas d'un GET d'un fichier HTML contient par exemple le code HTML de la page visée.

L'URL

URL signifie **Uniform Resource Locator** soit en français, « **localisateur uniforme de ressource** ».

C'est l'adresse d'une ressource donnée, unique sur le Web qui permet d'identifier une ressource comme une page web, un fichier pdf, une image...

Les URL sont utilisées notamment dans :

- la barre d'adresse des navigateurs
- les liens hypertextes dans les pages web
- les marques-pages des navigateurs
- les cartes de visites, publicités, QR code

La ressource représentée par l'URL et l'URL elle-même sont gérées par le serveur web.

L'URL

Le découpage d'une URL doit être la suivante :

- **URL = Uniform Resource Locator**
- **Découpage d'une URL :**

http://www.unsite.fr:1337/intro/page.html#chap1

- **Protocole (http, https, ftp...)**
- **Adresse du serveur**
- **Numéro de port TCP**
- **Répertoire sur le serveur**
- **Nom du fichier demandé**
- **Signet pour désigner un emplacement**

- **Utilisation des URL :**
 - Barre d'adresse des navigateurs
 - Liens hypertexte dans code HTML
 - Marques-pages des navigateurs

http://www.unsite.fr|

Uniform Resource Locator — Wikipédia
https://fr.wikipedia.org/wiki/Uniform_Resource_Locator

☆ Marque-pages

L'URL

http:// correspond au protocole. Cela indique au navigateur le protocole qui doit être utilisé pour récupérer le contenu.

Généralement, ce protocole sera HTTP ou sa version sécurisée : HTTPS.

Le « Web » fonctionne autour de ces deux protocoles mais le navigateur peut parfois gérer d'autres protocoles comme mailto qui permet d'ouvrir un client de messagerie électronique ou ftp: qui permet de transférer des fichiers.

www.unsite.fr correspond au nom de domaine. Il indique le serveur web auquel le navigateur s'adresse pour échanger le contenu. À la place du nom de domaine, on peut utiliser une adresse IP, ce qui est moins pratique et donc moins utilisé sur le Web.

:1337 correspond au port utilisé sur le serveur web. Il indique la « porte » technique à utiliser pour accéder aux ressources du serveur. Généralement, ce fragment est absent car le navigateur utilise les ports standards associés aux protocoles (80 pour HTTP, 443 pour HTTPS). Si le port utilisé par le serveur n'est pas celui par défaut, il faudra l'indiquer.

L'URL

/intro/page.html est le chemin, sur le serveur web, vers la ressource. Aux débuts du Web, ce chemin correspondait souvent à un chemin « physique » existant sur le serveur. De nos jours, ce chemin n'est qu'une abstraction qui est gérée par le serveur web, il ne correspond plus à une réalité « physique ».

#chap1 correspond à une ancre. Celle-ci désigne un endroit donné de la ressource. Une ancre représente, en quelque sorte, un marque-page à l'intérieur de la ressource.

Ajouter une ancre à une URL permet au navigateur d'afficher la ressource à l'endroit de ce marque page. Pour un document HTML, par exemple, le navigateur défilera la page jusqu'au niveau de l'ancre. Pour un document audio ou vidéo, le navigateur ira se placer à l'instant représenté par l'ancre.

Dans les URL on ne peut pas utiliser n'importe quel caractère. La norme RFC1378 définit les 3 possibilités suivantes :

- des caractères alphanumériques : 0→9, a→z, A→Z
- des caractères spéciaux : \$-_.+ !*()
- des caractères encodés : %C3%A0 pour à, %C3%A9 pour é...

Résumé

Le Web expliqué avec des emojis

<https://ladigitale.dev/digiview/#/v/676e9de2a5fbb>

<https://www.youtube.com/watch?v=RHIjpE7pZh8>