

SNT - Thème 2 : LE WEB

I. L'architecture du Web

Créé en 1989 par Tim Berners-Lee, le « Web » (_____ , littéralement « la toile d'araignée à l'échelle mondiale ») est un réseau mondial d'informations constitué de milliards de documents divers (textes, images, vidéos, sons etc.), regroupés sous forme de « sites », stockés sur des millions de _____ répartis sur le globe.

Ils sont accessibles par des adresses appelées _____ , qui sont composées de trois éléments : https:// www.facebook.com /page/creation

_____ :
http:// (HyperText Transfer Protocol) ou son équivalent sécurisé https:// qui correspondent au protocole de communication client-serveur développé pour le Web. Il utilise le protocole TCP/IP

_____ :
souvent une marque, une entreprise, une association etc.

_____ :
pointe vers une ressource ou une page précise

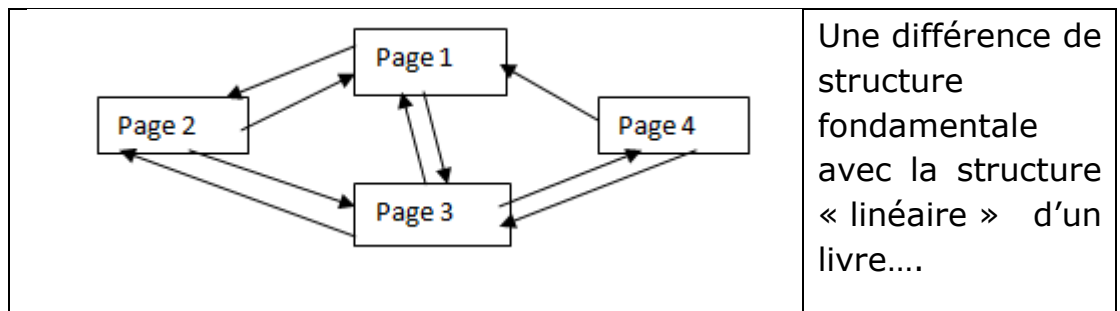
Principe de fonctionnement :

Pour consulter une page Web, le _____ , via un logiciel appelé _____ (le premier, en 1993, s'appelait Mozaic ; depuis : Edge, Firefox, Chrome, Safari, etc.), envoie une _____ au serveur avec l'adresse URL souhaitée. L'adresse est transformée en adresse IP via l'annuaire DNS. Une fois identifié, le _____ lui envoie une réponse avec le document demandé (ou un message d'erreur).



Souvent, un site WEB est un ensemble de pages = des fichiers au format HTML (hyper text markup language) (exemple page1.html) reliées (inter connectées) entre elles via des _____.

La structure en graphe via des hyperliens :



II. Langages utilisés

1) Le langage HTML

La programmation d'une page Web repose sur l'utilisation d'un langage de description (par opposition à un langage de programmation) appelé « _____ » _____ . Le langage HTML utilise des _____ pour délimiter le début et la fin de chaque instruction lue par le navigateur. La plupart des instructions HTML sont ainsi délimitées par une balise ouvrante <élément> et une balise fermante </élément>. Ces balises sont invisibles à l'écran mais elles permettent au navigateur d'interpréter ce qu'il doit afficher.

Voici le « squelette » d'une page web :

```
<html>
  <head>
    <meta charset="UTF-8">
    <title> Génial ! </title>
  </head>

  <body>
    Ceci est une page web.
  </body>
</html>
```

< ! -- tête de la page , pour les métadonnées-->
< ! -- encodage des caractères -->

< ! -- corps de la page, pour le contenu -->

La partie « head » du fichier HTML contient des métadonnées :
(Définition : une métadonnée est une donnée servant à décrire une autre donnée, par exemple la date de prise de vue d'une photo, l'auteur d'un site web...)

- encodage des caractères (suivant les langues)
- Nom de l'auteur
- Autorisations pour les robots (« crawlers » : voir plus loin) des moteurs de recherche
- Appel à des fichiers externes (mise en forme, programmation etc.)

2) Séparer le fond et la forme : le langage CSS

Pour des raisons pratiques évidentes (répétition de la mise en forme des milliers de fois dans de grands sites web), la mise en forme, le style graphique est déporté dans un fichier externe, écrit dans un autre langage : le **CSS (Cascading Style Sheet)**. On peut ainsi décider que tous les titres seront écrits en bleu, que les menus auront tous un fond turquoise et seront tous en haut à gauche etc.

3) L'interaction et la personnalisation : le web dynamique :

L'utilisateur rentre des informations sur la page web, le _____ (en PHP ou en python) une page WEB tenant compte des informations rentrées par l'utilisateur et envoie la page.



III. Les moteurs de recherche

Le Web contient des milliards de pages et il est constitué de millions de sites. Pour pouvoir trouver de l'information dans cette masse de données, les moteurs de recherche parcourent le Web (les pages qui autorisent à le faire) à l'aide de _____ (crawler) pour créer un index à partir des contenus.

L'index permet de répondre aux requêtes des utilisateurs. Par exemple, si l'internaute recherche des informations sur « informatique au lycée », le moteur recherchera dans son index toutes les pages qui contiennent ces mots.

Après avoir analysé les textes, mots-clés, balises, popularité de la page, liens entrants, etc., le moteur de recherche classera ensuite les réponses par ordre de _____.

Le **référencement naturel** ou SEO (*Search Engine Optimization*) permet d'améliorer le classement d'un site Web dans les résultats de recherche grâce notamment à des **mots-clés** placés dans le contenu rédactionnel. Le référencement naturel est à distinguer du **référencement payant** et les **liens sponsorisés**.

IV. Le Web et les données personnelles

1) La sécurisation des données

Au cours d'une connexion _____, le serveur envoie au client un certificat SSL contenant une _____. Le navigateur crypte une _____ que seul le serveur peut déchiffrer, et dès lors les échanges se font en cryptant les informations selon la clé privée. Présence de la sécurité SSL : l'URL commence par HTTPS et affichage d'un cadenas verrouillé.

Les certificats sont datés et possèdent une durée de validation. Ils ne sont en principe pas accessibles à des pirates qui reproduisent des sites

2) Le traçage

L'accès à la plupart des sites web est gratuit. Pour financer les coûts d'hébergement, mais aussi de production des contenus, beaucoup de sites font appel à la _____. Afin que les annonceurs puissent cibler ces publicités en fonction du profil de l'utilisateur, les sites Web sont amenés à **collecter des données personnelles**. Certaines données sont fournies volontairement par l'utilisateur (par exemple, lorsqu'il crée un compte sur un site Web), mais la plupart sont collectées à leur insu, lorsqu'il navigue d'une page Web à l'autre en choisissant de cliquer sur un lien plutôt que sur un autre.

La collecte de ces données utilise plusieurs mécanismes. Le plus courant est le _____, qui permet de récupérer les pages visitées par un même utilisateur. Grâce au RGPD (Règlement Général sur la Protection des Données) et aux options des navigateurs, il est possible de réduire ce traçage en réglant les paramètres de sécurité, mais il reste très difficile de l'empêcher complètement. Désormais une demande vous a été faite sur ce point lorsque vous arrivez sur un site...

3) Le droit d'auteur

Les ressources publiées sur le Web (textes, photos, images, vidéos etc.) sont la _____ de leur auteur et ne sont pas libres de droit. Leur usage est réglementé par différentes législations en fonction des pays ou encore des régimes spécifiques dont peuvent se réclamer les auteurs, comme les licences *Creative Commons* par exemple.