

Activité 1 / Qu'est-ce qu'une page web ?



30 min

Les pages des sites web sont accessibles en ligne et sont affichées grâce à un navigateur.

1. Citer différents noms de navigateurs.

Dans l'ordre de popularité : Chrome, Firefox, Safari, Edge, Opera, Internet Explorer, etc.

2. Quels sont les éléments contenus dans une page web ?

Une page web peut contenir du texte, des images, des vidéos et des liens hypertextes...

3. Grâce à un survol de souris, compter le nombre de liens hypertextes de la page d'accueil de Snapchat : <https://www.snapchat.com>.

La page comporte plus de 33 liens hypertextes (à la parution de ce cahier !)

4. En survolant un lien hypertexte, qu'affiche le navigateur en bas de la page web ?

Le navigateur affiche l'adresse de la ressource vers laquelle pointe le lien hypertexte.

5. Que se passe-t-il quand l'utilisateur clique sur un lien hypertexte ?

Un clic sur un lien hypertexte redirige l'utilisateur vers une autre page web qui traite du sujet choisi ou une autre partie de la page.

L'adresse web (aussi appelée URL) est l'invention la plus fondamentale du Web. Elle commence souvent par « www. » mais la syntaxe correcte devrait toujours commencer par le nom du protocole de communication utilisé pour nommer les adresses : « http » ou « https ». L'adresse web doit être saisie dans la barre d'adresse d'un navigateur web.

https://www.instagram.com/?hl=fr

6. Saisir l'adresse <https://www.instagram.com> dans la barre d'adresse d'un navigateur web et appuyer sur la touche Entrée. Qu'obtenez-vous à l'écran ?

La page d'accueil du site « Instagram » s'affiche dans le navigateur.

7. Rechercher la signification du cadenas situé devant l'adresse <https://www.instagram.com>, ainsi que la différence entre les protocoles HTTP et HTTPS.

Le cadenas et la mention « HTTPS » garantissent la confidentialité et l'intégrité des données échangées entre le site web et l'utilisateur.

8. Rechercher le rôle joué par chaque partie de l'adresse web de la question 7 ci-dessous :

https://	www	.	instagram	.	com
C'est le nom du protocole de communication servant à nommer les adresses web.	C'est un sous-domaine du domaine « instagram ».	Le point est un séparateur.	C'est le nom de domaine du site web de l'entreprise Instagram. C'est aussi un domaine de 2 ^e niveau ou un sous-domaine du domaine « com ».	Séparateur entre le domaine principal et le sous-domaine « instagram ».	C'est le domaine principal ou le domaine de 1 ^{er} niveau.





Arthur surfe sur le Web avec le navigateur Firefox. Lors de sa première visite sur le site web du journal *Le Monde*, un bandeau lui demande d'accepter des cookies.



1. Rechercher sur le Web ce qu'est un « cookie HTTP ». Où est-il enregistré ?

Un cookie HTTP [appelé aussi cookie web ou cookie de navigateur] est un ensemble de données envoyé par un serveur au navigateur web de l'utilisateur. Les cookies sont enregistrés dans la mémoire de l'appareil utilisé.

2. Arthur est-il obligé d'accepter les cookies de ce site ?

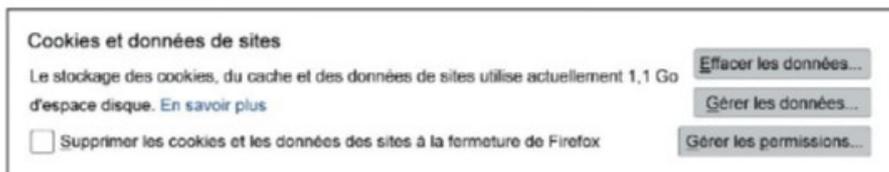
Non, Arthur peut continuer sa visite sans accepter les cookies.

Curieux, Arthur mène l'enquête. Il clique sur « Paramétrer les cookies » de la fenêtre précédente. Une nouvelle page web lui propose de paramétrer lui-même les différents cookies.

3. Dresser la liste des différentes catégories de cookies utilisés.

Le site utilise des « cookies de fonctionnement » [obligatoires], des « cookies analytics », des « cookies sociaux », des « cookies de ciblage publicitaire » et des « cookies de personnalisation du parcours lecteur ».

Arthur ouvre alors le module de paramétrage des cookies de son navigateur web et obtient la fenêtre ci-contre :



4. Quelles sont les options à la disposition d'Arthur pour contrôler les informations stockées par les sites web sur son ordinateur (cookies, cache et données des sites) ?

Arthur peut supprimer les cookies à la fermeture du navigateur ou encore n'autoriser que certaines catégories de cookies, en les paramétrant site par site. Toutes les traces laissées lors de sa navigation sont enregistrées dans le cache de l'ordinateur, qu'il peut vider régulièrement.

5. Arthur accepte tous les cookies suggérés. En utilisant le module de paramétrage du navigateur web (ci-contre), indiquer combien de cookies ont été enregistrés par le site web du journal *Le Monde* lors de la visite d'Arthur.

27 cookies ont été sauvegardés sur l'ordinateur d'Arthur.

Site	Cookies
www.lemonde.fr	10
buf.lemonde.fr	3
lemonde.fr	14

6. Arthur retourne le lendemain sur le site lemonde.fr.

Expliquer pourquoi la page s'affiche normalement sans lui demander d'accepter des cookies.

Le site a identifié la machine qui vient d'afficher la page web en consultant les cookies enregistrés localement. Le site possède déjà toutes les informations requises à son sujet, inutile de les redemander.

Activité 3 / Comment peut-on classer les pages web issues d'une recherche ?



30 min

Les moteurs de recherche permettent de trouver des informations dans des pages dont on ne connaît pas l'adresse, voire dont on ignore l'existence. On souhaite effectuer une recherche avec les deux mots clés suivants « **Apollo 11** » sur trois moteurs de recherche différents : Google, Qwant et Yahoo!.

1. Effectuer la recherche « **Apollo 11** » sur les trois moteurs proposés et compléter le tableau suivant avec vos résultats de recherche :

Moteur de recherche	Présence d'une annonce de publicité en tête des résultats	Position du premier lien vers Wikipédia	Position du premier lien vers un média culturel	Position du premier lien vers le film Apollo 11
Google	Non	lien n° 1	lien n° 4 (Futura.Sciences)	lien n° 2
Qwant	Oui (Amazon)	lien n° 3	lien n° 2 (Podcast de France Inter)	Au-delà du lien n° 30
Yahoo!	Non	lien n° 1	n°5 (Rollingstone)	lien n° 2

2. Expliquer pourquoi le classement des résultats est différent selon les moteurs de recherche.

Les algorithmes chargés de classer les résultats sont propres au moteur de recherche utilisé, le classement des résultats de recherche en dépend et ces résultats sont donc classés différemment.

3. Quel moteur de recherche conseilleriez-vous ? Justifier votre réponse.

Les résultats de la recherche sont globalement équivalents. Seul l'ordre des résultats change d'un moteur à l'autre. On peut conseiller de ne pas s'arrêter sur le premier résultat de recherche mais de bien parcourir les 10 premiers résultats en faisant preuve d'esprit critique.

Découvrir le système de ranking

Le système de « *ranking* » fait la force des moteurs de recherche. Notons que les algorithmes exacts ne sont jamais connus car ils sont considérés comme un secret industriel. Il existe au moins 3 méthodes principales pour classer les pages :

- Le **tri par pertinence**, qui évalue entre autres le nombre de fois qu'un mot ou un document apparaît dans une page web, ou encore la correspondance entre l'expression cherchée et le contenu du document.

- Le **tri par popularité**, comme le *PageRank* de Google, qui note la popularité d'une page web. Plus une page possède de *backlinks* (c'est-à-dire de liens qui pointent vers elle), plus la page est populaire.

- Le **tri par rapport à l'audience**, qui évalue le temps resté par l'internaute sur une page web, synonyme de son intérêt.

Des études ont montré que 90 % des internautes ne consultent que le premier résultat d'une recherche.

4. Pourquoi est-il important pour une page web d'être en tête d'un classement de recherche ?

Les internautes sont le plus souvent attentifs aux premiers résultats de la recherche, l'apparition en tête de classement est donc essentielle.

5. La méthode du « *spamming* » consiste à écrire du texte en blanc sur fond blanc dans une page web. Expliquer pourquoi le *tri par pertinence* (voir le document ci-contre) est considéré comme peu fiable par les moteurs de recherche actuels.

Ajouter des mots « invisibles » permet d'augmenter artificiellement le nombre d'occurrences d'un mot clé dans une page web et donc de le faire remonter dans les classements par pertinence. Le tri par pertinence peut donc être pris en défaut par cette pratique de *spamming*.

EXERCICE 1 / Acheter un billet d'avion sur le Web !

15 min

Lucie a effectué des recherches sur le Web dans le but d'acheter un billet d'avion. Le lendemain, alors qu'elle surfe sur le Web, des publicités pour des billets d'avion apparaissent sur les pages web qu'elle consulte.

1. S'agit-il d'une publicité ciblée ? Proposer une explication à la présence de ces publicités.



Les sites web que Lucie a consultés ont très certainement stocké des cookies sur sa machine et les reconnaissent. Ils peuvent alors lui proposer une publicité ciblée en relation avec ce qu'elle a cherché.

Lucie met plusieurs jours à se décider et observe que les prix des billets augmentent à chacune de ses visites.

2. Selon vous, est-ce le fruit du hasard ou l'évolution à la hausse des prix du marché ? Proposer une explication.

Non, les cookies laissés sur sa machine peuvent détecter que Lucie est intéressée par des billets d'avion et le site marchand peut être tenté d'augmenter ses prix en ligne pour déclencher un acte d'achat.

3. Proposer à Lucie plusieurs solutions à tester pour résoudre son problème.

Lucie pourrait supprimer tous les cookies de son ordinateur ou naviguer en ouvrant une fenêtre de navigation privée.

EXERCICE 2 / Droits d'auteur

20 min

Quentin doit faire un exposé sur le *big data*. Il a fait un grand nombre de copier-coller et il a inséré des images et des textes provenant de différents sites web.

Extrait du Code de la propriété intellectuelle

D'après la loi française, les créateurs (livres, films, musique, œuvre artistique) intellectuels sont les propriétaires de leurs créations. Il est donc interdit de les utiliser (tout ou partie) sans avoir d'autorisation ou avoir acquitté un droit payant et ceci même si ces œuvres se trouvent sur le Web.

Le droit d'auteur protège « toutes les œuvres de l'esprit ». Leur créateur en est le propriétaire exclusif.

La notion d'« œuvre de l'esprit » concerne toutes les productions qui ont nécessité un travail intellectuel.

Toute reproduction, représentation ou diffusion, par quelque moyen que ce soit, d'une œuvre de l'esprit est un **délit de contrefaçon**. (article L. 335-3)

1. Une page web est-elle soumise au code de la propriété intellectuelle ?

Oui car une page web est un contenu original et a été écrite par un auteur qui possède donc des droits.

2. Quentin a-t-il respecté le droit d'auteur ?

En reproduisant tout ou partie du texte ou des images de différents sites, Quentin n'a pas respecté le droit d'auteur.

3. Quelles solutions s'offrent à lui pour réaliser son exposé ?

L'usage est toléré dans le cadre d'une utilisation pédagogique, mais il faut citer les sources (nom de l'auteur, références de l'œuvre, etc.).

Les définitions à connaître

- Le **World Wide Web** (communément appelé le **Web**) est un système utilisant le protocole HTTP permettant de visiter des pages sur le réseau Internet.
- Le **protocole HTTP** (*HyperText Transfer Protocol*) est un protocole de communication qui permet à un navigateur d'échanger avec la machine sur laquelle le site web est stocké.
- Un **lien hypertexte** est un élément d'une page web qui, lorsque l'internaute clique dessus, dirige celui-ci vers une autre page web ou une autre partie de la page.
- Une **URL** (*Uniform Resource Locator*), aussi appelée adresse web, est une chaîne de caractères qui permet d'identifier une ressource du Web. L'URL doit débuter par le protocole utilisé (HTTP ou HTTPS).
- Un **cookie** est une suite d'informations, souvent de petite taille et identifiée par un nom, qui peut être transmis à votre navigateur par un site web sur lequel vous vous connectez.
- Un **moteur de recherche** est une application qui permet de rechercher des informations sur le Web.
- Le **droit d'auteur** est l'ensemble des droits dont dispose un auteur ou ses ayants droit sur des œuvres originales de l'esprit.

Carte mentale

