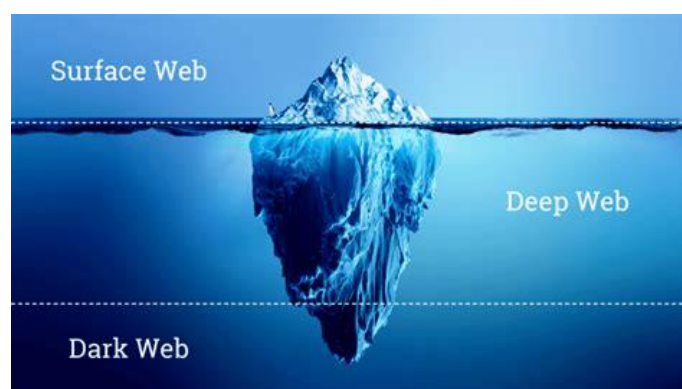


## Sciences du Numérique et Technologie



## Le Web

## TD n° 2 Deep Web et Dark Web



## Objectif de la séance :

- Connaître les différents types de Web
- Comprendre les enjeux du Web



## Mots-clés :

**Client :** logiciel envoyant des demandes à un serveur.

**CSS (Cascading Style Sheets) :** feuilles de style décrivant la présentation des documents HTML.

**HTML (HyperText Markup Language) :** langage de balisage utilisé pour la création de pages Web.

**HTTP ou HTTPS :** protocole de transmission permettant d'accéder à des pages Web via un navigateur. Le HTTPS est sécurisé.

**Indexation :** analyse du contenu des pages Web pour détecter les mots-clés.

**Lien hypertexte :** élément d'une page Web (texte, image, etc.) qui, lorsque l'on clique dessus, renvoie vers une autre zone de la même page, une autre page, ou un autre site Web.

**Moteur de recherche :** application informatique permettant de rechercher une ressource (pages Web, images, vidéos, fichiers, etc.) à partir d'une requête sous forme de mots.

**Navigateur :** logiciel permettant d'afficher des sites Internet (Chrome, Firefox, Edge, Safari, etc.).

**Serveur :** ordinateur exécutant des opérations suivant les requêtes effectuées par un ordinateur appelé « client ».

**URL (Uniform Resource Locator) :** adresse d'une page d'un site.

**Web (World Wide Web) :** système hypertexte, utilisant le protocole HTTP, permettant de visiter des pages sur Internet.



## A la fin de cette séance j'aurai acquis ...

| Des connaissances                                                                                                 | Des compétences                                                                                                                                                                                |
|-------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ul style="list-style-type: none"><li>- Web visible ou de surface</li><li>- Deep Web</li><li>- Dark Web</li></ul> | <ul style="list-style-type: none"><li>- Comparer et analyser les différents types de Web</li><li>- Reconnaître les pages sécurisées</li><li>- Connaître certaines notions juridiques</li></ul> |

*Vous écrirez les réponses aux questions sur le document*

### I. Quels sont les risques du Web : Web visible ou Web de surface

**Le Web visible ou Web de surface** est la partie du WEB indexée par les moteurs de recherche comme Google, Bing, Qwant et bien d'autres. Cette indexation se fait automatiquement par des robots ou crawlers en anglais, qui parcourent les sites internet et passent de liens en liens (cf. séance 4). Chaque page est enregistrée et indexée par le moteur de recherche qui peut ensuite restituer les résultats à travers des mots clés.

Des études ont montré que le Web visible n'est que d'environ 10% de la totalité du WEB.

Les autres éléments non indexés par les moteurs de recherche ne font pas partie du Web Visible. On distingue alors :

- **Le Web opaque** : ce sont les données du WEB que les moteurs de recherche pourraient indexer mais qu'ils ne font pas. Par exemple, les sites trop volumineux ne sont pas totalement indexés par les moteurs de recherche, c'est par exemple, le cas du site de la NASA. Les pages trop volumineuses ne sont également pas indexées.
- **Le Deep Web (Web Profond)** : C'est la partie du Web que les moteurs de recherche ne sont pas capables d'indexer  
Certains sites ont beaucoup de pages dynamiques, il est assez difficile pour un moteur de recherche de les indexer car les liens hypertextes sont aussi dynamiques et changent d'une visite à l'autre. Une page avec un formulaire de recherche dont le contenu est lié à ces recherches ne pourra être indexé (site SNCF, Air France, etc).

Un site WEB peut avoir du contenu qui n'a pas de liens hypertextes menant à ce dernier, le robot du moteur de recherche ne pourra pas explorer cette partie du site.

Beaucoup de sites sont des sites clients et ne sont pas indexables. Par exemple, le contenu client du site de votre banque est privé, Google n'y a pas accès.

Enfin, certains administrateurs peuvent interdire l'indexation de certains contenus grâce à des directives sur le fichier robots.txt.

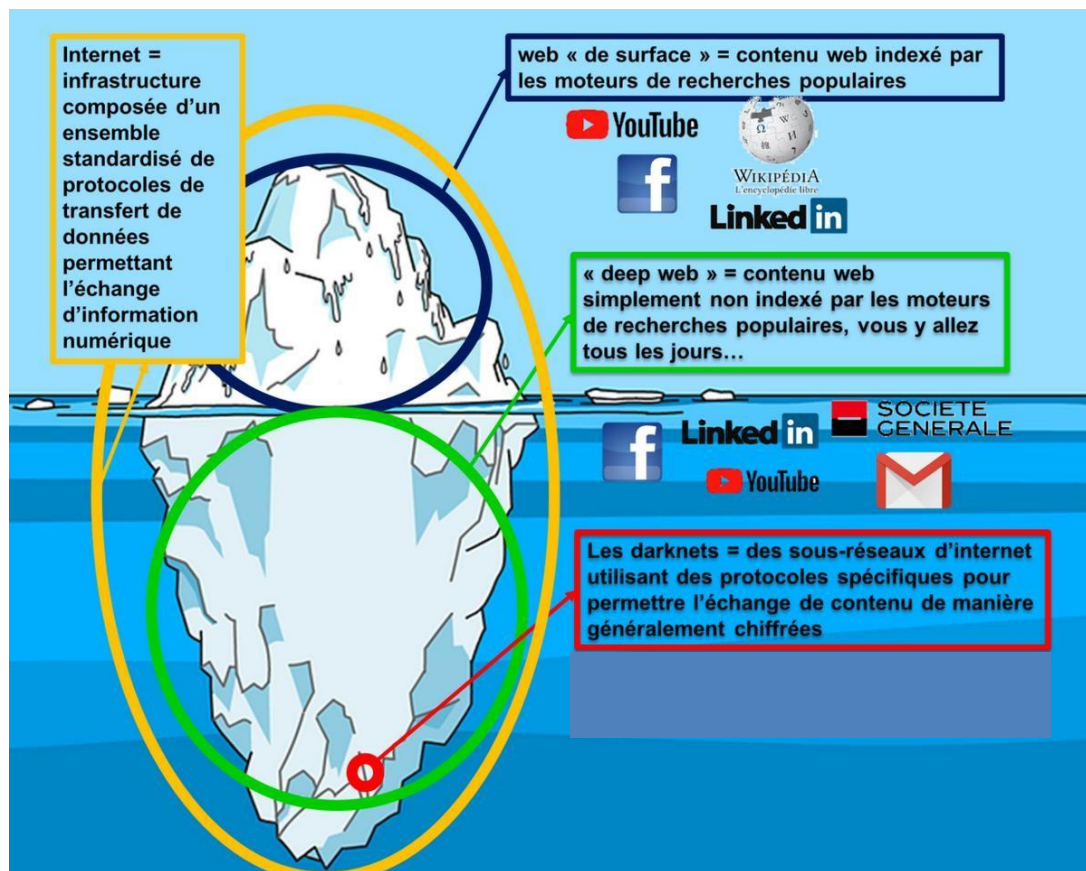
On peut aussi aller plus loin en protégeant l'accès par une authentification. Il faudra alors un mot de passe pour accéder au contenu. Dans ce cas-là, on peut assimiler cela à du Web privé.

- **Le Dark Web ou Web Sombre** est le Web contenu dans les darknet auxquels les moteurs de recherche n'ont pas accès.  
Un darknet est un réseau superposé dans internet, seulement accessible à l'aide d'un logiciel ou d'une configuration spéciale.

Le Dark Web est assez connu des internautes par la seule représentation qui en est faite : un supermarché pour criminels.

Mais l'existence des darknet est motivée par l'anonymisation, par exemple :

- Les opposants aux dictatures ou régime autoritaire,
- Les journalistes peuvent utiliser les darknet pour communiquer ou mettre en ligne des informations.
- Des groupes persécutés (religion, groupes ethniques, ...) au sein d'un pays peuvent communiquer via le Dark web de manière plus sécurisée.



### Exercice 1 :

A l'aide du document et des vidéos sur le site, répondre aux questions suivantes :

1. Quels sont les différents types de web ?
2. Quelle proportion du WEB le web visible représente-t-il ?
3. Qu'est-ce que le web opaque ?
4. De quels sites le Deep Web est-il constitué ?
5. Pour quel motif le Dark Web a-t-il été créé ?
6. Quelles personnes peuvent avoir intérêt à utiliser le Dark Web ?