

Guide du professeur - Chapitre 1

L'objectif de ce chapitre est de présenter les grands principes de **fonctionnement d'Internet**, les mécanismes utilisés et les services fondamentaux, ainsi qu'un type d'utilisation particulier (le pair à pair). Il vise à faire comprendre aux élèves qu'Internet est un ensemble de techniques diverses.

Histoire, enjeux, débats

Ce chapitre permet d'aborder les principes des réseaux qui se cachent derrière Internet, des protocoles type TCP et IP, du routage, des serveurs DNS et des réseaux pair-à-pair. La double-page controversée permet d'aborder un sujet délicat qui continue à opposer différentes communautés, la neutralité du Net et le monde de demain permet de prendre un premier contact avec l'Internet des objets (qui sera étudié de manière plus précise dans le chapitre 6).

- ❖ Activité 1 Trafic Internet et indépendance des réseaux
- ❖ Activité 2 Le protocole TCP/IP et le routage
- ❖ Activité 3 Adresses symboliques et serveurs DNS
- ❖ Activité 4 Les réseaux pair-à-pair

Bande dessinée d'ouverture

Cette bande dessinée illustre la différence entre Internet en tant que réseau de réseau et les différents services qui reposent sur Internet. Un problème chez un utilisateur pour un service donné ne veut pas dire qu'Internet « est en panne », ce qui n'a pas de sens, puisqu'Internet est d'une part un réseau de différents réseaux, et que d'autre part l'utilisation habituelle d'Internet fait appel à plusieurs systèmes différent (généralement, au minimum le transport TCP/IP et le DNS.) Le problème peut venir par exemple, comme illustré dans cette bande dessinée, d'un problème du service de nom (DNS, activité 3), et pour seulement une partie des utilisateurs d'Internet. D'autres utilisateurs, connectés différemment à Internet, au même moment peuvent ne rien voir de ce problème.

Note : Les pages **Controverse** et **Monde de demain** ne sont pas corrigées. En effet, dans les premières, il est plus question d'ouvrir un débat avec les élèves, les questions n'étant là que pour initier une discussion, et pour les secondes, le but est de prendre connaissance des pistes de recherche actuelles sur ce sujet.

Les activités

ACTIVITÉ 1

Les compétences travaillées

Cette activité permet de prendre conscience de la diversité des moyens d'accès à Internet, ainsi que de réaliser que, contrairement à l'idée couramment répandue, Internet repose principalement sur des câbles, en particulier sous-marins.

Cette activité peut donner lieu à une recherche d'information sur l'infrastructure d'Internet (document 1), la quantité de trafic sur Internet (document 2) et les moyens d'accès (documents 3 et 4). Elle vise à démystifier le côté « magique » et « sans fil » d'Internet, et à fixer les idées sur les ordres de grandeur de données véhiculées.

Les savoir-faire travaillés

- Caractériser quelques types des réseaux physiques.
- Caractériser l'ordre de grandeur de trafic Internet et son évolution.

Question 1

Utilisation du document 1

À l'aide de la densité de connexion entre régions, identifier les grands flux de trafic entre : Amérique du Nord, Europe, Asie (Chine et Japon).

L'accès au site « Submarine Cable Map » permet de compléter cette analyse et aussi de se rendre compte de l'omniprésence des câbles de connexion utilisés par Internet, y compris depuis le territoire français.

Question 2

Utilisation du document 2

L'évaluation d'un trafic de 201 milliards de Gigaoctets par mois correspond à près de 50 milliards de films haute définition transférés par mois.

Le trafic ayant été multiplié par près de 130 millions entre 1992 et l'estimation de 2022, il est préférable de tracer une courbe avec une échelle logarithmique.

On peut faire calculer un taux de croissance annuel de 86 %.

Question 3

Utilisation du document 3

a. Le trafic 3G reste inférieur à 10 exaoctets soit 10 milliards de Gigaoctets (ou environ 2,5 milliards de films) par mois.

b. Le trafic 5G évolue de manière exponentielle, alors que le trafic 3G reste assez stable.

c. La 5G n'est mise en place à large échelle qu'à partir de 2020, l'évaluation du trafic faible les premières années correspond au nombre restreint d'utilisateurs. On peut anticiper que le trafic 5G évoluera de manière exponentielle également, avec des débits de l'ordre de 10 fois ceux de la 4G.

Question 4

Utilisation des documents 3 et 4

Exemples de type d'accès possibles :

Haut débit	Très haut débit
ADSL	Fibre optique
3G	4G
	WiFi

Le débit constaté des technologies peut varier selon l'endroit où l'on se trouve par rapport au central téléphonique (ADSL), à l'antenne relais (3G, 4G), ou au point d'accès (WiFi). Pour les technologies radio, cela dépend également du nombre de personnes utilisant le même accès radio. Seule la fibre permet d'avoir un débit à peu près fixe.

ACTIVITÉ 2

Les compétences travaillées

Cette activité présente les principes mis en œuvre dans Internet : la notion de paquets de données, l'adressage IP et le routage permettant aux paquets de données d'atteindre leurs destinations.

Les savoir-faire travaillés

- Comprendre la notion de paquet de données.
- Faire la différence entre les paquets IP et TCP/IP, et leur utilité.
- Appréhender la notion de route entre machines sur Internet.

Question 1

Utilisation du document 1

Les principaux éléments essentiels pour le fonctionnement d'Internet sont :

- le paquet de donnée, qui permet de découper une information en éléments de transport (paquet) ;
- les protocoles TCP/IP qui permettent d'identifier les paquets (grâce aux adresses IP) et leur bonne transmission (c'est le rôle principal du protocole TCP) ;
- les routeurs qui organisent l'envoi des paquets.

On peut également citer le système de nom (ou DNS) qui est vu dans l'activité 3, qui rend l'utilisation d'Internet beaucoup plus facile, bien que non indispensable. On peut utiliser Internet avec uniquement les adresses IP, c'est ce qui est fait dans l'activité FabLab

Question 2

Utilisation du document 4 (et non 2)

Deux paquets envoyés à la même destination ne mettent pas forcément le même temps pour arriver à leurs destinations, ils peuvent passer par des routes différentes, en fonction des configurations des routeurs, des pannes de réseau, etc.

Question 3

Utilisation du document 3

Une adresse IPv4 est codée sur 32 bits, soit 4 octets. On représente une adresse IPv4 avec 4 nombres entre 0 et 255 (un octet permet de coder 256 nombres) séparés par des points. Par exemple : 192.168.0.1 ou 10.0.10.3

Lorsque des machines sont connectées physiquement entre elles, elles sont généralement sur le même **réseau** dont le numéro est la partie commune des adresses IP des machines concernées. Par exemple, si les machines qui ont les adresses 192.168.12.23 et 192.168.100.45 sont connectées entre elles, elles appartiennent toutes deux au réseau 192.168/16. Le « /16 » indique que seuls les 16 premiers bits (soit 2 octets) sont significants pour le numéro de réseau. On a choisi pour une raison de simplicité de ne garder que le numéro de réseau, mais la convention habituelle est de noter une adresse sur 32 bits, soit par exemple 192.168.0.0/16. Dans notre cas, toutes les adresses entre 192.168.0.0 et 192.168.255.255 appartiennent au réseau 192.168/16. Les valeurs 255 ont une signification particulière et ne sont pas utilisables pour une machine.

Question 4

Utilisation du document 4

Un paquet TCP contient les mêmes informations qu'un paquet IP ainsi que des informations complémentaires. Dans les deux cas, ils contiennent principalement :

- les adresses IP de l'émetteur et du destinataire du paquet ;
- une taille de paquet, ce qui permet d'avoir des paquets de taille variable, en fonction des données à transporter (NB : le programme officiel parle de paquets de taille fixe ce qui est incorrect dans IP) ;
- un compteur de durée de vie qui est décrémentée à chaque fois que le paquet est reçu par routeur. Lorsque le compteur tombe à 0, le paquet est supprimé de ceux à transmettre ;
- un numéro de protocole qui est l'identifiant du protocole chargé de traiter les données contenues dans le paquet. Pour un paquet TCP, ce numéro est toujours le même et est égal à 6 ;
- un code de vérification (ou « somme de contrôle » ou « checksum » en anglais) qui permet de vérifier que le paquet reçu est bien le même que celui envoyé (plus précisément que le code de vérification est bien le même. La construction du code de vérification fait que la probabilité que 2 paquets différents aient le même code est extrêmement faible.) Ce code est nécessaire car lors de l'envoi des données sur le lien physique (et plus particulièrement par radio), il peut y avoir des erreurs de transmission (un 0 qui se change en 1 et inversement) ;
- les données à transmettre.

Un paquet TCP contient en plus un numéro qui est le numéro du paquet dans la suite de paquets à envoyer. Cela permet de vérifier qu'un paquet numéroté n est bien arrivé, et dans le cas contraire de demander la retransmission de ce paquet, en l'identifiant par son numéro.

C'est ce type de paquet qu'on doit utiliser pour vérifier la bonne transmission d'une donnée.

Question 5

Utilisation des document 3 et 5

Une autre route possible entre le PC1 et PC2 est :

PC1	Réseau 10.0.1/24	Routeur 1	Réseau 10.0.3/24	Routeur 3	Réseau 10.0.4/24	Routeur 2	Réseau 10.0.5/24	PC2
-----	---------------------	-----------	---------------------	-----------	---------------------	-----------	---------------------	-----

On choisira de préférence la première route car elle est plus courte en nombre de routeurs traversés.

En cas de panne du réseau 10.0.2/24, un paquet allant de PC2 à PC1 devra suivre la route passant par les réseaux 10.0.4/24 et 10.0.3/24 :

PC2	Réseau 10.0.5/24	Routeur 2	Réseau 10.0.4/24	Routeur 3	Réseau 10.0.3/24	Routeur 1	Réseau 10.0.1/24	PC2
-----	---------------------	-----------	---------------------	-----------	---------------------	-----------	---------------------	-----

Le PC4 étant connecté au réseau 10.0.4/24, son adresse IP doit commencer par les mêmes 24 bits. Son adresse doit être entre 10.0.4.0 et 10.0.4.254 (la valeur 10.0.4.255 est réservée), par exemple 10.0.4.10

Question 6

Utilisation du document 5

Lorsqu'un paquet reçu par un routeur, son compteur de durée de vie est décrémenté. S'il atteint 0 le paquet est supprimé. Si le paquet est renvoyé de routeur 1 vers routeur 2 puis de routeur 2 vers routeur 1, ce dernier va de nouveau le renvoyer vers routeur 2 (sauf un changement de configuration de route intervient à ce moment-là). Le compteur sera décrémenté à chaque fois jusqu'à atteindre à 0 et être supprimé. Le paquet n'aura pas atteint sa destination. L'application qui en a besoin sera en erreur.

ACTIVITÉ 3

Les compétences travaillées

Cette activité permet de comprendre comment les associations entre les noms de machine et leurs adresses IP sont gérés dans Internet grâce au service de nom, ou « DNS »

Les savoir-faire travaillés

- Comprendre l'organisation des noms et l'architecture du DNS.
- Comprendre le principe du protocole DNS.

Question 1

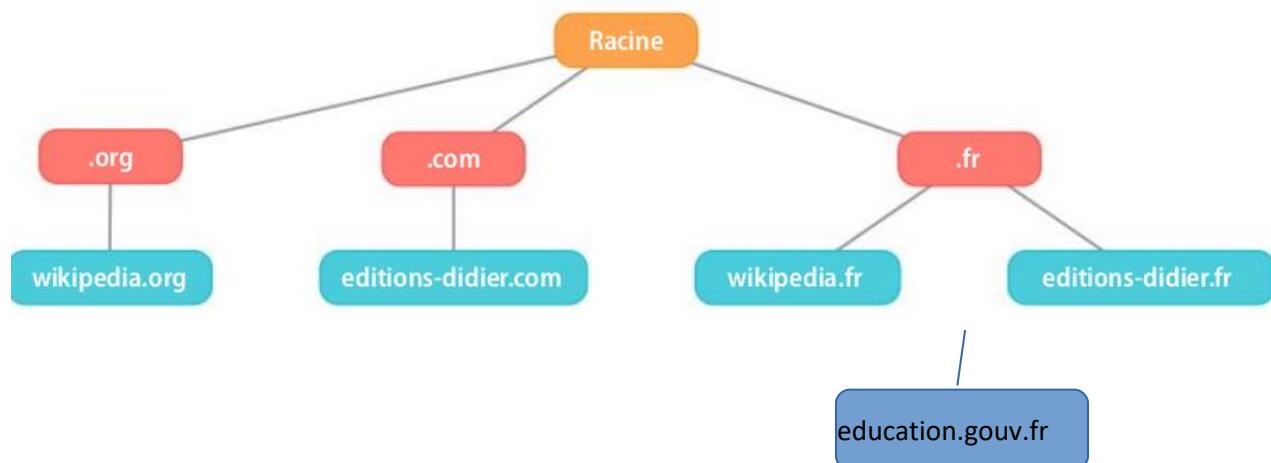
Utilisation du document 1

a. Les deux commandes ping sont adressées à la même adresse IP 192.134.5.25 : c'est la même machine qui a répondu bien qu'on ait utilisé 2 noms différents. Une adresse IP peut être associé à plusieurs noms différents.

b. Une machine peut être identifiée par son adresse IP ou par un de ses noms (une chaîne de caractères).

Question 2

Utilisation du document 2



Le nom `education.gouv.fr` appartient au sous-domaine `gouv.fr` qui lui-même appartient au domaine `.fr`.

Question 3

Utilisation des documents 3 et 4

- a. Il y a 3 serveurs DNS associés au nom de domaine `education.gouv.fr`
- b. Ils sont utilisés pour trouver l'adresse IP d'une machine dont le nom appartient au domaine `education.gouv.fr`, par exemple `www.education.gouv.fr`
- c. Une première requête est faite au serveur racine, celui indique de demander à un serveur responsable du domaine `.fr`. Une deuxième requête est faite vers ce serveur, qui indique d'aller demander à un serveur pour `gouv.fr`. La troisième requête faite à ce serveur indique de demander

à serveur pour education.gouv.fr, par exemple ns1.ate.tm.fr. Une requête à ce serveur renvoie l'adresse IP de www.education.gouv.fr. Il y a eu 4 requêtes.

Question 4

Exercice à faire sur ordinateur

L'outil *nslookup* permet de récupérer une adresse à partir d'un nom et l'inverse.

```
> nslookup education.gouv.fr
```

Non-authoritative answer:

Name: education.gouv.fr

Address: 185.75.143.24

L'adresse IP de education.gouv.fr est donc 185.75.143.24

```
> nslookup www.editions-didier.fr
```

Non-authoritative answer:

Name: www.editions-didier.fr

Address: 195.81.225.138

```
> nslookup 194.199.8.10
```

```
10.8.199.194.in-addr.arpa name = www.bnf.fr.
```

Adresse symbolique	Adresse IP
education.gouv.fr	185.75.143.24
www.editions-didier.fr	195.81.225.138
www.bnf.fr	194.199.8.10

NB : ces valeurs sont celles disponibles dans le DNS à la date de l'édition de ce manuel, et depuis un lieu de connexion donné. Les mêmes requêtes peuvent au cours du temps ou en fonction réseaux utilisés donner d'autres résultats.

ACTIVITE 4

Les compétences travaillées

Cette activité permet de donner les clefs du fonctionnement des réseaux pair-à-pair qui sont utilisés en particulier pour le téléchargement (utilisation de fichiers « torrent », logiciels comme « emule »). Il permet d'illustrer qu'Internet n'est pas une architecture centralisée sur quelques serveurs (Google, Facebook, etc.) mais permet bien une communication directe entre toutes les machines connectées à Internet. Ce fonctionnement en mode pair-à-pair est utilisé dans bien d'autres services sur Internet, par exemple les « blockchains ».

Les savoir-faire travaillés

- Comprendre le principe d'un fonctionnement en pair-à-pair.
- Sensibiliser aux risques de l'utilisation majoritaire du pair-à-pair, le téléchargement illégal.

Question 1

Utilisation du document 1

Le but de cette activité est de prendre conscience des différentes méthodes d'accès aux données sur Internet et d'introduire l'approche pair-à-pair.

Les logiciels de téléchargement reposent habituellement sur un mécanisme de ratio, qui permet de télécharger en fonction de ce qu'on a mis à disposition des autres utilisateurs. Cette fonctionnalité illustre parfaitement la notion de pair-à-pair : chaque utilisateur est consommateur et en même temps fournisseur, il n'y a plus la distinction avec les serveurs centralisés ou seul le serveur fournit les données.

Question 2

Utilisation du document 2

- ① Mon ordinateur
- ② Nombre de pairs
- ③ Adresse d'un des pairs
- ④ Quantité téléchargée
- ⑤ Taille totale de la vidéo

Question 3

Utilisation du document 3

Lors de l'utilisation du logiciel, la machine 1 est initialement « client » uniquement, puis elle devient « pair » une fois qu'elle a récupéré de fragments de données qu'elle met à disposition. La machine 2 est un « tracker ». Les machines 3 et 4 sont des pairs mais peuvent aussi être des clients si elles téléchargent d'autres données (elles l'ont été par le passé, pour pouvoir récupérer les fragments).

Question 4

Utilisation du document 4

- a. Le texte fait référence au téléchargement illicite de biens culturels (films, musique, etc.) Un contrevenant risque une recommandation de la part de l'Hadopi, pouvant aller d'un simple avertissement à la suspension du l'accès à Internet.
- b. Le P2P permet de télécharger des contenus qui ne sont pas disponibles sur des serveurs commerciaux. Du fait de son architecture, il permet un téléchargement qui peut être plus rapide que depuis un seul serveur, car plusieurs fragments sont téléchargés en même temps. Cependant,

il permet également télécharger des contenus sans respecter les droits d'auteurs. Du fait de l'absence de contrôle des données échangées, il augmente également le risque de télécharger des virus informatiques qu'une personne malveillante peut avoir introduit dans les données mise à disposition.

Je vérifie mes connaissances

Solutions détaillées

1. QCM

1. Une adresse IP valide est :

A 124.45.67.89 : il y a 4 parties toutes comprises entre 0 et 255

Les autres réponses :

B 124.45.672 : NON, il n'y a que 3 parties (3 octets, soit 24 bits).

C 990.340.465.888 : NON, les valeurs ne sont pas entre 0 et 255.

2. Dans le nom www.editions-didier.fr, quel est le domaine de premier niveau ?

C .fr

Le domaine de premier niveau est celui le plus à droite du nom de la machine.

Les autres réponses :

www est le nom (sans nom de domaine) de la machine.

editions-didier.fr est le nom de domaine auquel appartient la machine www.

3. Avec quel outil pouvez-vous trouver l'adresse IP d'une machine à partir de son nom ?

Réponse : C les deux

Les deux outils recherchent l'adresse IP d'une machine dont on donne le nom et l'affichent. Plus généralement, tous les outils sur Internet font dès le début cette recherche d'adresse IP depuis un nom, puis qu'Internet ne fonctionne qu'avec des adresses IP (c'est la seule information contenue dans les paquets) et pas avec les noms.

4. Dans quel cas le serveur de nom du domaine editions-didier.fr est-il utilisé ?

Réponse : A

Il est utilisé pour trouver l'adresse IP de la machine www.editions-didier.fr. Comme ce nom appartient au domaine editions-didier.fr c'est un de ses serveurs qui connaît l'adresse IP.

2. Vrai/Faux

a. **VRAI.** Les 2 machines ont des adresses qui commencent par la même partie 12.34.56, elles peuvent donc toutes les deux appartenir au réseau 12.34.56/24.

Pour aller plus loin : On peut remarquer que la dernière partie de la première adresse (78) s'écrit en binaire sur 8 bits : 01011000 et celle de la deuxième adresse (99) s'écrit : 01100011. Les deux premiers bits sont égaux. Les deux adresses peuvent donc appartenir également au réseau 12.34.56.64/26

b. **FAUX.** Il existe plusieurs technologies telles que le WiFi, la 3G ou encore, bientôt, la 5G.

c. **FAUX.** Il continue à augmenter, encore aujourd'hui.

d. **VRAI.**

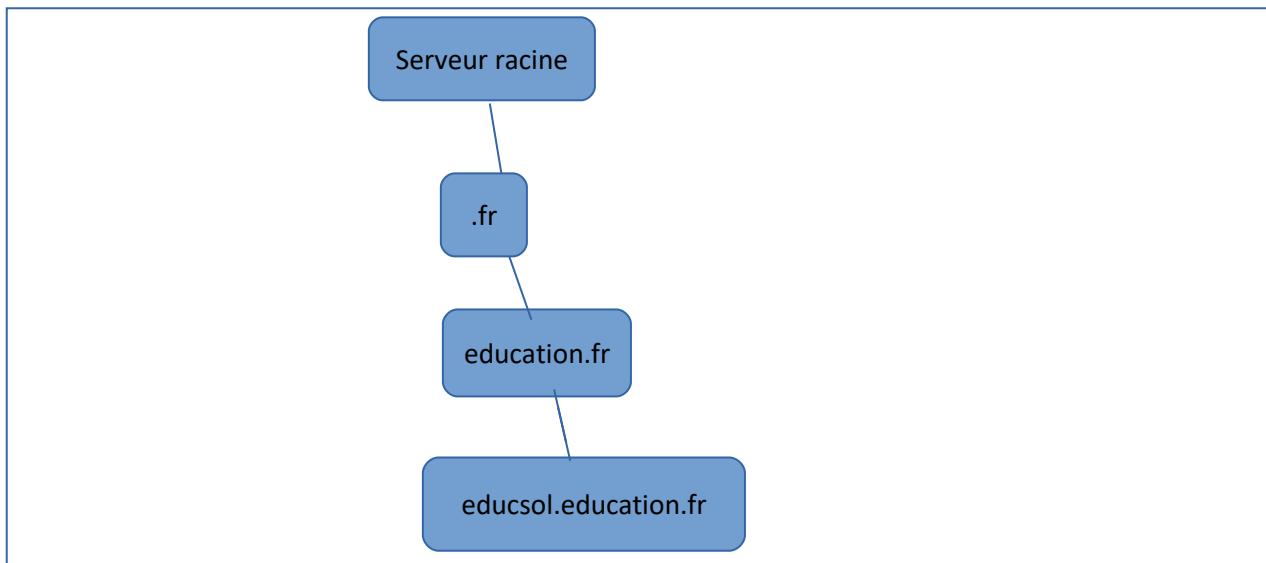
- e. **FAUX.** Les chiffres montrent que l'accès à Internet mobile commence à prendre une place importante.
- f. **FAUX.** On peut y accéder par son adresse IP ou son nom (ou adresse) symbolique.
- g. **FAUX.** Le tracker maintient la liste des machines qui possèdent les fragments de donnée.
- h. **VRAI.**
- i. **FAUX.** C'est un paquet TCP.

3. Vers une définition

Un **tracker** maintient la liste des pairs et leurs **adresses** IP. Un pair stocke des **fragments** de données. Un client se connecte d'abord à un **tracker** puis aux **pairs**.

4. Réaliser un schéma

Le schéma ci-dessous représente l'organisation de l'adresse edusol.education.fr dans l'arborescence du DNS



5. Organisation des connaissances

Un paquet devra passer par au moins un routeur pour atteindre une machine qui n'est pas sur le même réseau que l'émetteur. S'ils sont sur le même réseau, le paquet peut atteindre l'émetteur directement.

JE M'ENTRAÎNE

6. L'accès à Internet

- a. L'accès mobile représente $6,5/14,8 = 44\%$ des accès en 2017 et 51 % en 2018, soit une croissance annuelle de 7 %. On peut donc estimer que les accès mobiles atteindront $51 + 3 * 7 = 72\%$ en 2021 et qu'ils dépasseront la barre des 75 % en 2022.
- b. Pour télécharger un film de 4 Goctets, soit $4000 * 8$ Mbits, avec un accès haut débit de 12 Mbit/s il faudra $4000 * 8 / 12 = 2\ 667$ secondes, soit 45 minutes. Avec un très haut débit de 42 Mbits/s, il faudra $4000 * 8 / 42 = 762$ secondes ou 12 minutes et demi.

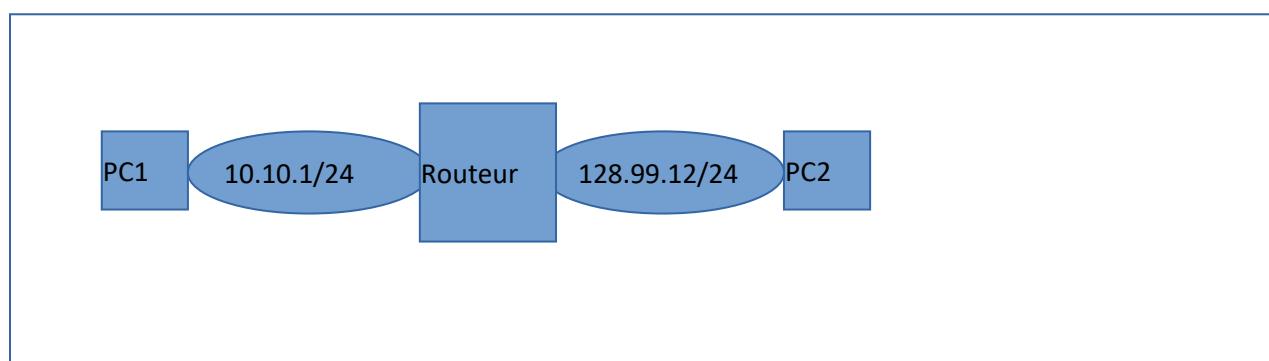
7. IP ou TCP

- a. Un paquet IP de taille totale 1 500 octets peut contenir $1500 - 24$ octets qui composent l'en-tête (voir Activité 2, document 4, il y a 6 lignes de 4 octets chacune dans l'en-tête). Il faudra donc $10000 / (1500 - 24) = 6,7$, ce qui donne 6 paquets pleins et un paquet contenant 1 144 octets.
- b. Avec un paquet TCP dont l'en-tête fait 48 octets, il faudra 6 paquets pleins et un paquet contenant 1288 octets.
- c. On peut utiliser TCP dont les en-têtes sont plus importantes car ce protocole permet de vérifier que les données ont été reçues.

8. Adresse de réseau

- a. Le réseau qui relie les machines 12.34.76.89, 12.24.76.223 et 12.34.76.1 et 12.34.76/24
- b. Le réseau qui relie les machines 10.1.2.3 et 10.1.128.3 est 10.1/16 puisque seuls les 16 premiers bits sont les mêmes.

9. Des réseaux



10. Les serveurs DNS

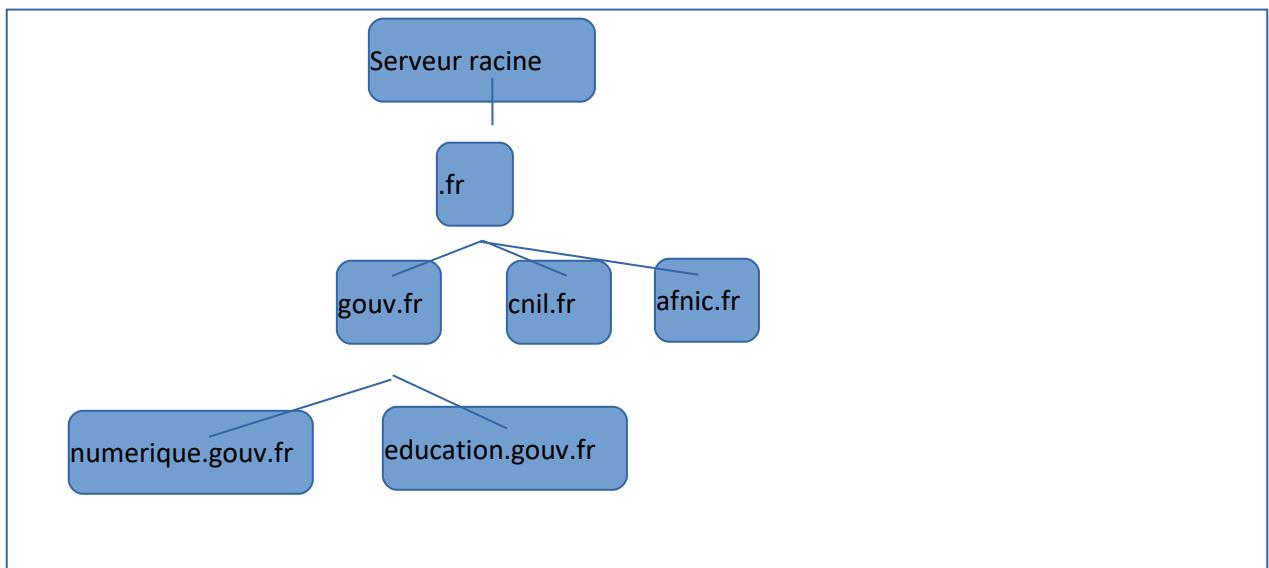
a. C'est le serveur (ou un des serveurs) du domaine wikipedia.org qui connaît l'adresse de la machine fr.wikipedia.org. Attention, dans cette adresse symbolique, le domaine de premier niveau est bien .org, le plus à droite.

b. C'est un des serveurs de lemonde.fr qui connaît l'adresse de www.lemonde.fr

11. Durée de vie d'un paquet

Le compteur « durée de vie » est décrémentée à chaque fois qu'il passe par un routeur et est détruit s'il atteint 0. Un paquet avec une durée de vie de 3 pourra passer 2 routeurs : le 3eme routeur qui le recevra le détruira.

12. Structure hiérarchique



13. L'adresse IP

Exercice à faire sur ordinateur.

En étant connecté à Internet, taper par exemple :

> ping (nom du lycée sur Internet)

> ping www.iana.org

Vous pouvez obtenir :

PING www.iana.org(www.iana.org (2620:0:2d0:200::8)) 56 data bytes

L'adresse 2620:0:2d0:200::8 est une adresse IPv6

Pour obtenir l'adresse IPv4, taper : > ping -4 www.iana.org

PING ianawww.vip.icann.org (192.0.32.8) 56(84) bytes of data.

> ping -4 www.afnic.fr

PING lb01-1.nic.fr (192.134.5.24) 56(84) bytes of data.

14. Les requêtes DNS

Exercice à faire sur ordinateur.

Il faut 3 requêtes DNS pour trouver l'adresse de la machine www.afnic.fr : une vers le serveur racine pour trouver le serveur du domaine .fr, puis une 2eme pour trouver le serveur du domaine afnic.fr, puis une dernière pour trouver l'adresse de www.afnic.fr

15. WebTorrent

Exercice à faire sur ordinateur.

Lorsqu'on se connecte on voit le graphe se mettre à jour en fonction des pairs auquel se connecte sa machine pour récupérer les fragments de fichier. Les points jaunes sont les pairs, les chiffres sont leurs adresses (qui peuvent être en IPv6).

16. Le téléchargement en P2P

Avec des connexions 12 Mbit/s, on pourra télécharger depuis les 23 pairs avec ce même débit. Il faudra donc $4000 * 8 / (12 * 23) = 116$ secondes, soit moins de 2 minutes.

17. Utilisation du P2P

Pour : organisation décentralisée, mise à disposition par tout un chacun de données téléchargeables, téléchargement pouvant être plus rapide lorsqu'il y a plusieurs pairs
Contre : risque de téléchargement illégal, risque de diffusion de contenus illicites.

FAB LAB

Le but de ce TP est réaliser un Internet privé, c'est-à-dire faire communiquer 2 machines qui se trouvent sur 2 réseaux différents en utilisant les technologies d'Internet. Pour cela on a besoin d'une 3^e machine qui fera fonction de routeur entre les 2 réseaux.

Pour réaliser ce TP, on se base sur l'outil VirtualBox qui permet de faire fonctionner plusieurs machines virtuelles sur une seule machine réelle. Le TP se déroulera sur les machines virtuelles disponibles sur l'espace numérique editions-didier (voir installation en annexe).

1. Ping

En utilisant l'outil `ping`, vérifier si la `machine1` et la `machine2` sont connectées entre elles.

Utiliser `ping machine1` et `ping machine2` sur chacune des 2 machines, et remplir le tableau suivant :

Nom symbolique	Adresse IP	Réseau
machine1		/24
machine2		/24

Pourquoi les deux machines ne sont pas connectées ? Que faut-il pour réaliser la connexion ?

Explication

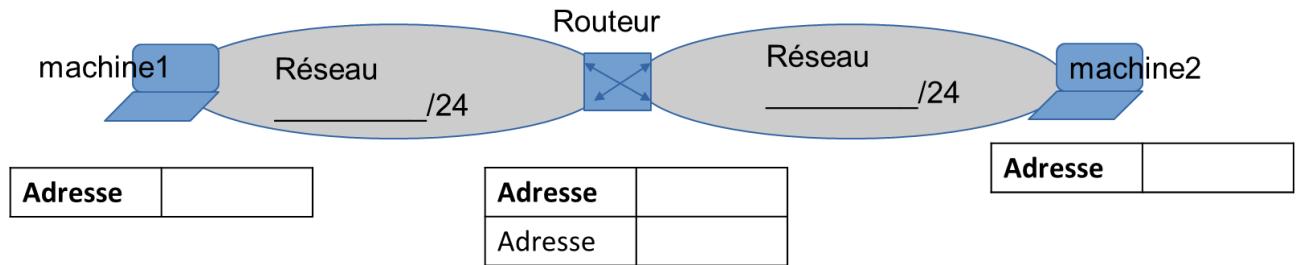
Les deux machines ne sont pas sur le même **réseau IP** : la partie de leurs adresses IP correspondant au 24 premiers bits (soit 3 octets, soit les 3 premiers termes de l'adresse sous forme A.B.C.D) ne sont pas les mêmes. Les deux machines ont besoin d'une 3^e machine qui va **router** les paquets d'un réseau vers l'autre.

Pour aller plus loin :

Pourquoi le ping fonctionne-t-il alors qu'on lui donne une "adresse symbolique" et pas l'adresse IP de la machine ? Normalement, une requête DNS est nécessaire pour traduire "l'adresse symbolique" en adresse IP. Mais il existe des cas où il est plus pratique d'avoir une association "adresse symbolique"/adresse IP sans support du DNS (dans le cas présent, il n'y a aucun serveur DNS d'installé). Pour cela on utilise un fichier appelé "host" qui se trouve dans le répertoire "/etc" qui contient cette information. En général, l'ordinateur regarde en priorité si l'adresse symbolique existe dans ce fichier et ne fait une requête DNS que s'il ne l'a pas trouvée.

2. Installation du routeur

Lancer la machine virtuelle routeur. En utilisant l'outil `ping` sur `machine1` et `machine2`, compléter le schéma ci-dessous.



Explication

Selon la machine utilisée, la machine routeur renvoie 2 adresses différentes. C'est normal puisque son rôle est de router les paquets d'un réseau vers l'autre. Il a donc besoin d'avoir une connexion sur chacun des réseaux. Dans notre exemple, le routeur possède 2 adresses qui se terminent toutes les deux par la même valeur (3), pour des raisons pédagogiques uniquement.

Tester sur chacune des machines si la connexion est effective entre la machine et `routeur`.

Explication

La connexion est effective entre `machine1` et `routeur`, et entre `machine2` et `routeur` puisque les commandes `ping` répondent. Mais ce sont les 2 "côtés" (plus précisément, les 2 interfaces) du routeur qui répondent chacun de leurs côtés, chacun sur son réseau propre (le réseau `10.0.1/24` vers `machine1`, le réseau `10.0.2/24` vers `machine2`). La connexion est effective entre toutes les machines sur un même réseau. A condition bien sûr qu'elles soient physiquement connectées ! Dans notre cas, VirtualBox simule cette connexion physique. Vous pouvez le vérifier en allant sur "Configuration" de chacune des machines virtuelles dans l'interface de VirtualBox, puis dans l'onglet "réseau". Vous pourrez vérifier que `machine1` est connecté sur le réseau physique (simulé) "reseau1", `machine2` sur reseau2, et que `routeur` est connecté sur reseau1 et sur reseau2, grâce à ses deux "cartes réseaux" (ou interfaces réseaux).

Vérifier avec `ping machine1` et `ping machine2` sur chacune des 2 machines si la connexion entre `machine1` et `machine2` est effective.

Explication

Les machines ne sont toujours pas accessibles l'une à l'autre. Pourtant, `routeur` est démarré et on vient de le vérifier, bien connecté. Alors pourquoi ? C'est parce qu'on n'a pas encore indiqué comment router les paquets d'un réseau à l'autre.

3. Analyse du trafic IP

La commande `tcpdump` permet de visualiser les paquets vus par la machine sur laquelle la commande est exécutée.

Lancer la commande `tcpdump icmp` sur `routeur`, et lancer `ping routeur` sur `machine1`. Que constatez-vous ?

Note : ICMP est le protocole utilisé par l'outil `ping`

Explication À l'écran, vous devriez voir des messages de ce type :

```
root@routeur:~# tcpdump icmp
tcpdump: verbose output suppressed, use -v or -vv for full protocol decode
listening on enp0s3, link-type EN10MB (Ethernet), capture size 262144 bytes
14:55:43.511728 IP machine1 > routeur: ICMP echo request, id 775, seq 1, length 64
14:55:43.511802 IP routeur > machine1: ICMP echo reply, id 775, seq 1, length 64
14:55:44.515778 IP machine1 > routeur: ICMP echo request, id 775, seq 2, length 64
14:55:44.515854 IP routeur > machine1: ICMP echo reply, id 775, seq 2, length 64
14:55:45.517759 IP machine1 > routeur: ICMP echo request, id 775, seq 3, length 64
14:55:45.517844 IP routeur > machine1: ICMP echo reply, id 775, seq 3, length 64
```

On peut y voir l'envoie du message permettant de faire le ping ("machine1 > routeur : ICMP echo request", le signe ">" indique que le message va de machine1 vers routeur) et sa réponse ("routeur > machine1 : ICMP echo reply"). Lorsqu'on fait un ping c'est toujours cet échange qui est fait.

On constate donc que les messages passent bien entre machine1 et routeur.

4. Routage

Sur la machine `machine1`, taper la commande :

```
ip route add default via 10.0.1.3
```

Cette commande indique à la machine d'envoyer par défaut vers l'adresse 10.0.1.3 tous les paquets qui ne lui sont pas destinés.

Puis tester à nouveau `ping machine2`. Quelle différence y a-t-il avec le point 2 ?

Explication Le message renvoyé par ping est différent : il indique que le message demandant le ping est bien envoyé, mais la réponse n'est pas reçue. On voit le message passer sur routeur ("machine1 > machine2 : ICMP echo request").

Le message arrive bien maintenant à machine2 car machine1 a été configurée pour envoyer vers l'adresse 10.0.1.3, qui est une adresse de routeur, sur le même réseau que machine1, tous les paquets qui ne sont pas à destination de son réseau (10.0.1/24). En d'autres termes, lorsque machine1 a un paquet à envoyer, elle l'envoie directement à l'adresse de destination si celle-ci est sur le réseau 10.0.1/24, ou bien l'envoie à l'adresse 10.0.1.3 si ce n'est pas le cas.

Pour aller plus loin :

La commande tapée sur machine1 ne permet que à machine1 de savoir ou envoyer les paquets, c'est à dire vers routeur (sur son interface qui a l'adresse 10.0.1.3). Cela ne suffit pas à envoyer les paquets vers machine2. Cela fonctionne dans notre cas car routeur est configuré pour renvoyer les paquets à destination d'une adresse du réseau 10.0.2/24 qu'il reçoit vers l'adresse qui se trouve sur ce réseau (et pareillement pour les paquets à destination d'une adresse du réseau 10.0.1/24). On peut le voir en tapant la commande "ip route" sur routeur : chacune des 2 lignes indique quoi faire pour un paquet allant soit vers le réseau 10.0.1/24 soit vers 10.0.2/24.

Lancer maintenant également la commande `tcpdump icmp` sur `machine2`. Que constatez-vous ?

Envoi de messages vers	Résultat
routeur	
machine2	

Expliquer ce qu'il se passe.

Explication On voit bien le message "echo request" passer également sur machine2. Mais on ne voit pas le message de réponse passer. Pourquoi ? C'est parce que machine2 ne sait pas comment renvoyer la réponse vers machine1

5. Routage suite

Sur la machine `machine2`, taper la commande :

```
ip route add default via 10.0.2.3
```

Que donne `ping machine2` sur `machine1` et `tcpdump icmp` sur `machine2` et routeur ?

Expliquer la différence.

Explication Maintenant que machine2 est configurée pour renvoyer les messages vers machine1, et que routeur est également configuré (voir plus haut), les réponses peuvent être reçues par machine1 : le ping fonctionne entre les 2 machines, les paquets sont routés d'un réseau vers l'autre. Nous avons construit un réseau Internet (un réseau de réseau), privé car il n'est pas connecté à d'autres réseaux.

6. Transport de données

La commande `nc` permet d'envoyer ou de recevoir des données.

- Sur `machine2` lancer `nc` en mode réception : `nc -l -p 4242`.
- Sur `routeur` lancer : `tcpdump -A tcp`.
- Sur `machine1` lancer `nc` en mode émission : `nc machine2 4242`, puis taper un texte et appuyez sur Entrée. Que constatez-vous sur `routeur` et `machine2` ?

Explication Sur machine2, on voit le texte tapé sur machine1 qui s'affiche : le texte a été envoyé depuis machine1, vers machine2, au travers de routeur, en utilisant le protocole TCP.

Sur routeur on voit les différents paquets passer entre machine1 et machine2. On peut constater les choses suivantes :

- à chaque fois qu'on envoie un texte, il y a au moins un paquet TCP vers machine2 et un paquet en retour; en effet TCP s'assure de la fiabilité de la transmission en renvoyant une information sur les données reçues ;
- on voit apparaître le texte tapé dans ce qui est affiché pour le paquet allant de machine1 vers machine2 : l'information envoyée d'une machine à une autre est transportée en clair.

La page du codeur

1 Lecture de code

Le code affiche la partie commune entre 2 adresses IPv4 exprimées sous forme de liste de 4 octets, soit l'adresse d'un réseau sur lequel les machines possédant ces adresses peuvent être connectées.

Le résultat du programme est [128,192,23] pour le test entre les adresses IP1 et IP2 et [128,192] pour les tests avec IP3.

NB : c'est une version simplifiée du calcul du réseau en faisant l'hypothèse que les réseaux ne sont exprimés que sur 8,16 ou 24 bits (1, 2 ou 3 octets entier). Un calcul exact nécessite de considérer les valeurs sous forme binaire.

2 Correction de code

Erratum 1 : le code doit donner l'affichage :

« de :128.192.1.2 vers : 10.0.1.3

hello »

Erratum 2 : la ligne 22 est à supprimer

Pour faire cet exercice, il faut analyser le paquet IP (Activité 2, document 4)

La variable p contient un paquet IP codé sous la forme d'une suite d'octets.

L'adresse source se situe sur 4 octets, de 12 à 15 inclus. La ligne 6 doit donc être :

for j in range(4)

L'adresse destination se situe sur les 4 octets suivants, la ligne 10 doit donc itérer sur le contenu du paquet et être :

dst+=str(p[i+4+j])+ "."

Pour afficher le contenu du paquet, qui est de taille définie par les 3ème et 4ème octets, soit p[2] et p[3], il faut lire le contenu à partir du 25ème octet, ce contenu étant de taille : taille du paquet – taille de l'entête (24 octets pour un paquet IP non TCP).

La ligne 13 doit donc être :

l=p[2]*256+p[3]-24

Et la ligne 16 doit donc être :

d+=p[25+j]

(ou bien par exemple modifier le range de la ligne 15:range(25, 25+l))

3 Conception de programme

```
routes=[("Paris","Londres",200),("Paris","Amsterdam",300),("Londres","New York",6000),("Amsterdam","Oslo",300),("Oslo","New York",5000)]
```

```
# Fonction récursive permettant de calculer les chemins et leur longueur
def route(routes,src,dst):
    chemins=[]
    # On itère chaque route connue
    for r in routes :
        # si la route est déjà connue on la met dans la liste des chemins
        if r[0]==src and r[1]==dst :
            chemins.append(r)
            continue
        # si on a une route depuis la source (src) :
        if r[0]==src :
            # on cherche récursivement les chemins depuis la destination de cette route
            ch=route(routes,r[1],dst)

            # On ajoute à chaque résultat la longueur du chemin déjà parcouru
            ch2=[]
            for c in ch:
                ch2.append((r[0],c[1],c[2]+r[2]))

            # On ajoute les chemins aux chemins à retourner
            chemins+=ch2

    return chemins

def routage(routes,src,dst):
    chemins=route(routes,src,dst)
    if chemins==[]:
        print "Pas de route entre %s et %s" % (src,dst)
        return

    # On cherche le chemin le plus court
    pluscourt=chemins[0]
    for ch in chemins[1:]:
        if ch[2]<chemin[2]:
            pluscourt=ch

    print pluscourt

routage(routes,"Paris", "New York")
```

4

Affichage du contenu d'un paquet (1)

```
f=open("paquet1.txt","rb")
pkt=f.read()

pkt=bytearray(pkt)

dst=""
for i in range(16,20):
    dst+=str(pkt[i])+"."

print "Destinataire : ",dst

src=""
for i in range(12,16):
    src+=str(pkt[i])+"."

print "Emetteur : ",src

print "Duree de vie : %d "%pkt[8]

print "Protocole : %d "%pkt[9]

print "Longueur totale : %d" % (pkt[2]*256+pkt[3])
```

5 Affichage du contenu d'un paquet (2)

```
f=open("paquet2.txt","rb")
pkt=f.read()

pkt=bytearray(pkt)

dst=""
for i in range(16,20):
    dst+=str(pkt[i])+"."

print "Destinataire : ",dst

src=""
for i in range(12,16):
    src+=str(pkt[i])+"."

print "Emetteur : ",src

print "Duree de vie : %d "%pkt[8]

print "Protocole : %d "%pkt[9]

l=pkt[2]*256+pkt[3]
print "Longueur totale : ",l

if pkt[9]==6 :
    # Protocole TCP
    num=0
    for i in range(28,32):
        num=num*256+pkt[i]

    print "Numero de sequence : ",num

# L'entete TCP fait 48 octets (en l'absence d'options)
print "taille de la donnee : ",l-48
```

6 Contenu d'une suite de paquets

```
def printPkt(pkt):
    dst=""
    for i in range(16,20):
        dst+=str(pkt[i])+"."
    print "Destinataire : ",dst

    src=""
    for i in range(12,16):
        src+=str(pkt[i])+"."
    print "Emetteur : ",src

    print "Duree de vie : %d " %pkt[8]

    print "Protocole : %d " %pkt[9]

    l=pkt[2]*256+pkt[3]
    print "Longueur totale : ",l

    if pkt[9]==6 :
        # Protocole TCP
        num=0
        for i in range(28,32):
            num=num*256+pkt[i]

        print "Numero de sequence : ",num

    # L'entete TCP fait 48 octets (en l'absence d'options)
    print "taille de la donnee : ",l-48

f=open("paquets.txt","rb")
pkt=f.read()
pkt=bytearray(pkt)

i=0
while i < len(pkt):
    taille=pkt[i+2]*256+pkt[i+3]
    print "taille=",taille
    printPkt(pkt[i:i+taille])
    i+=taille
```

7 Affichage des données transportées

```
f=open("paquets.txt","rb")
pkt=f.read()
pkt=bytearray(pkt)

data=[]
i=0
while i < len(pkt):
    taille=pkt[i+2]*256+pkt[i+3]

    # Lis les donnees contenues dans le paquet
    st=""
    for j in range(48,taille):
        st+="%" % pkt[i+j]

    # Lis le numero de sequence
    num=0
    for j in range(28,32):
        num=num*256+pkt[i+j]

    # Stock le couple (numero de sequence, donnees)
    data.append((num,st))
    i+=taille

# tri les couples par numero de sequence
data=sorted(data,cmp=lambda x,y: x[0]-y[0])

# cree la donnee dans sa totalite
st=""
for d in data:
    st+=d[1]

print st
```

Guide du professeur - Chapitre 2

L'objectif de ce chapitre est de présenter les grands principes de fonctionnement du **Web**, en tant que système donnant accès à un **ensemble de données** (page, image, son, vidéo) reliées par des **liens hypertextes** et accessibles sur le **réseau Internet**.

Histoire, enjeux, débats

Ce chapitre permet d'aborder les principes de l'HTML, du CSS et du protocole http, de comprendre comment fonctionnent les moteurs de recherche, d'appréhender la notion de trace numérique, intimement liée à la question de l'anonymat sur Internet (Controverse) et découvrir le concept de Web sémantique, c'est-à-dire le Web de demain.

- ❖ Activité 1 Les ingrédients d'une page Web : HTML et CSS
- ❖ Activité 2 HTTP : les clients parlent aux serveurs
- ❖ Activité 3 Sous le capot des moteurs de recherche
- ❖ Activité 4 Passer inaperçu sur le Web
- ❖ Activité 5 Publier dans les règles

Bande dessinée d'ouverture

Cette bande dessinée permet de poser la question des traces numériques laissées sur le Web par l'utilisateur, ces traces permettant la création d'un profil utilisateur utilisé par les applications pour des recommandations ou des publicités ciblées. L'activité 4 permet de comprendre la manière dont nous laissons des traces sur le Web et donne quelques pistes afin d'en laisser le moins possible. On pourra également se reporter au guide d'hygiène numérique en fin de manuel, ce guide approfondissant les moyens de protéger son identité numérique.

Note : Les pages **Controverse** et **Monde de demain** ne sont pas corrigées. En effet, dans les premières, il est plus question d'ouvrir un débat avec les élèves, les questions n'étant là que pour initier une discussion, et pour les secondes, le but est de prendre connaissance des pistes de recherche actuelles sur ce sujet.

Les activités

ACTIVITE 1

Les compétences travaillées

Cette activité apporte des connaissances techniques sur des **langages** qui, contrairement aux langages de programmation, ne sont pas destinés à décrire des algorithmes :

- HTML est un **langage de données** permettant de décrire la structure et le contenu d'un document ;
- CSS est un **langage de règles** permettant de décrire des contraintes sur la présentation, sans expliquer comment les réaliser.

HTML et CSS illustrent la diversité des langages informatiques. Dans les deux cas il s'agit de coder des informations avec une syntaxe rigoureuse, qui permet à la machine d'interpréter le code.

Ces connaissances peuvent être mises en œuvre :

- Par un **exposé oral** expliquant les bases des langages HTML et CSS, ou bien des caractéristiques plus avancées (à partir de ressources trouvées en ligne) : ajout de tableaux ou de vidéos dans un document HTML, utilisation de CSS pour créer un menu fixe, etc. ;
- Par un **projet de groupe** consistant à créer un petit site Web avec plusieurs pages liées entre elles, en se répartissant les rôles : recherche d'information, structuration du site, écriture des pages, amélioration du formatage par des feuilles CSS.

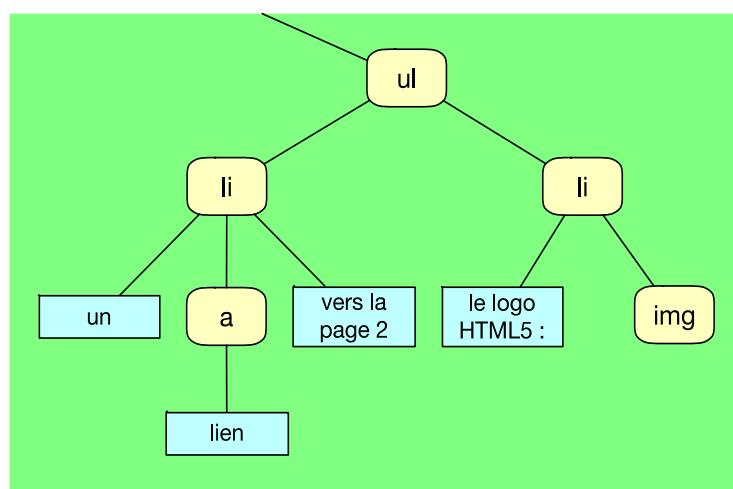
Les savoir-faire travaillés

- Analyser une page Web et faire le lien entre le code HTML et le contenu affiché.
- Analyser une feuille de style CSS et faire le lien entre les règles et leur effet sur l'affichage.
- Créer ou modifier une page Web simple.
- Créer ou modifier une feuille de style simple.

Question 1

Utilisation du document 1

L'exploitation du code HTML du document 1 permet de compléter l'arbre, à partir de la balise .



Question 2

Utilisation du document 1

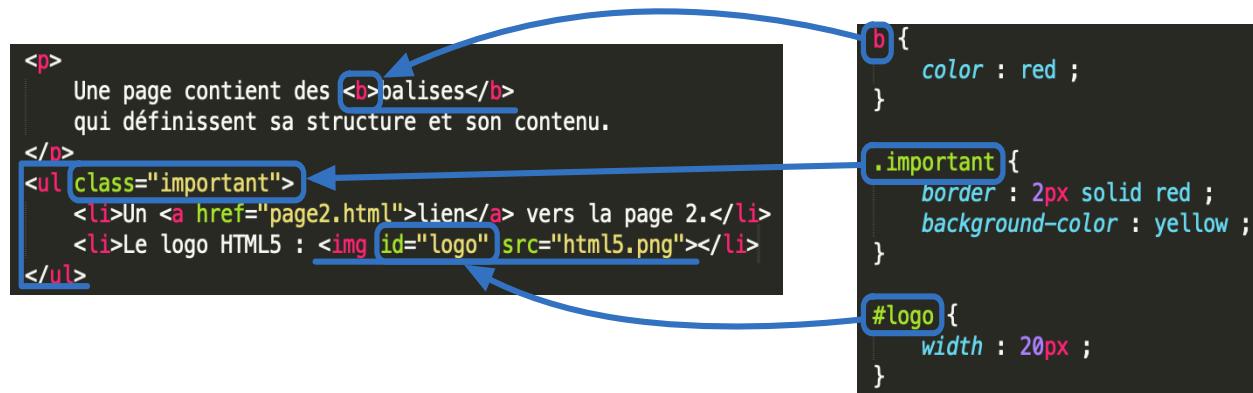
En s'inspirant du document 1, on crée le document suivant. La principale différence est que on a des listes imbriquées : dans l'élément `` de chaque pays, il y a un élément `` avec une liste de villes.

```
1 <html>
2   <head>
3     <meta charset="utf-8">
4     <title>Géographie</title>
5   </head>
6   <body>
7     <h1>Principales villes de quelques pays</h1>
8     <ul>
9       <li>France
10         <ul><li>Paris</li><li>Lyon</li><li>Marseille</li></ul>
11       </li>
12       <li>Italie
13         <ul><li>Rome</li><li>Milan</li><li>Venise</li></ul>
14       </li>
15     </ul>
16   </body>
17 </html>
```

Question 3

Utilisation du document 2

Il s'agit de faire correspondre les sélecteurs de la feuille de style CSS (`« b »`, `« .important »`, `« #logo »`) avec les éléments du document HTML.



Question 4

Utilisation du document 2

Le style suivant colore le paragraphe (balise `<p>`) en bleu : `p { background-color : blue ; }`

Tous les paragraphes seront alors en bleu. Pour éviter cela, on peut ajouter un attribut `id="enbleu"` au premier paragraphe et changer la règle de la feuille de style en :

```
#enbleu { background-color : blue ; }
```

Question 5

Utilisation du document 2

On va utiliser deux classes : pays et ville, et leur associer une couleur :

```
.pays { color : red ; }
.ville { color : blue ; }
```

Pour que ces styles soient pris en compte, il faut ajouter les attributs class="pays" et class="ville" dans le code HTML :

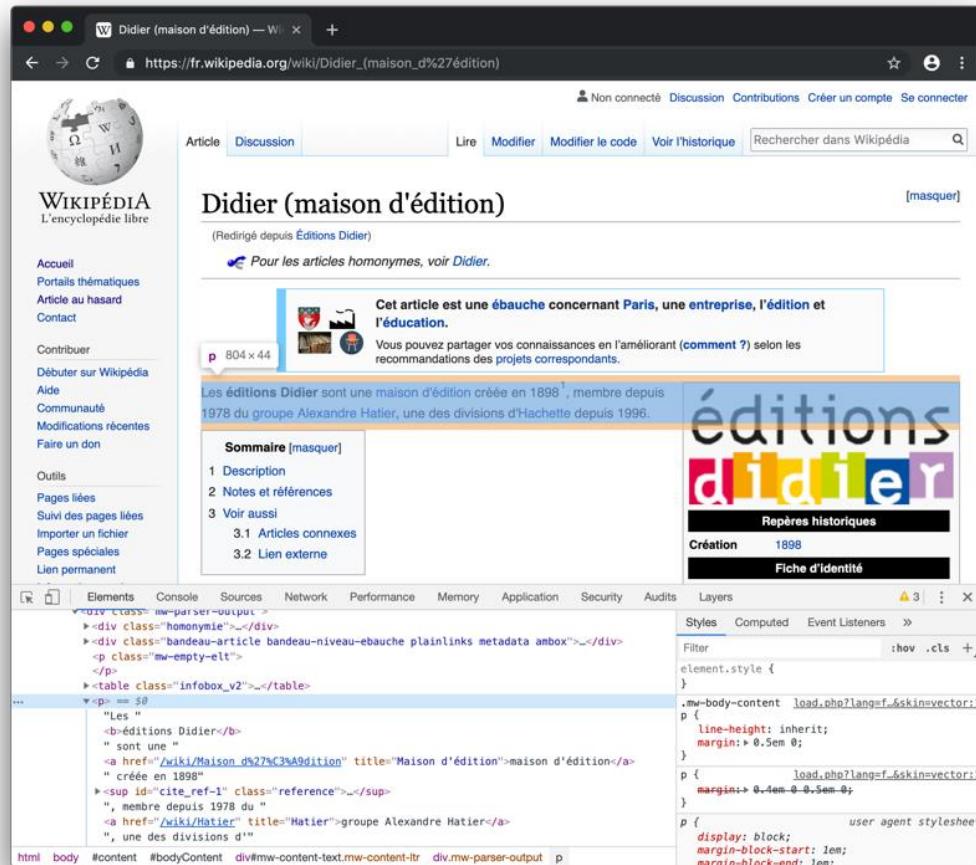
```
8      <ul class="pays">
9          <li>France
10             <ul class="ville"><li>Paris</li><li>Lyon</li><li>Marseille</li></ul>
11         </li>
12         <li>Italie
13             <ul class="ville"><li>Rome</li><li>Milan</li><li>Venise</li></ul>
14     </li>
```

Question 6

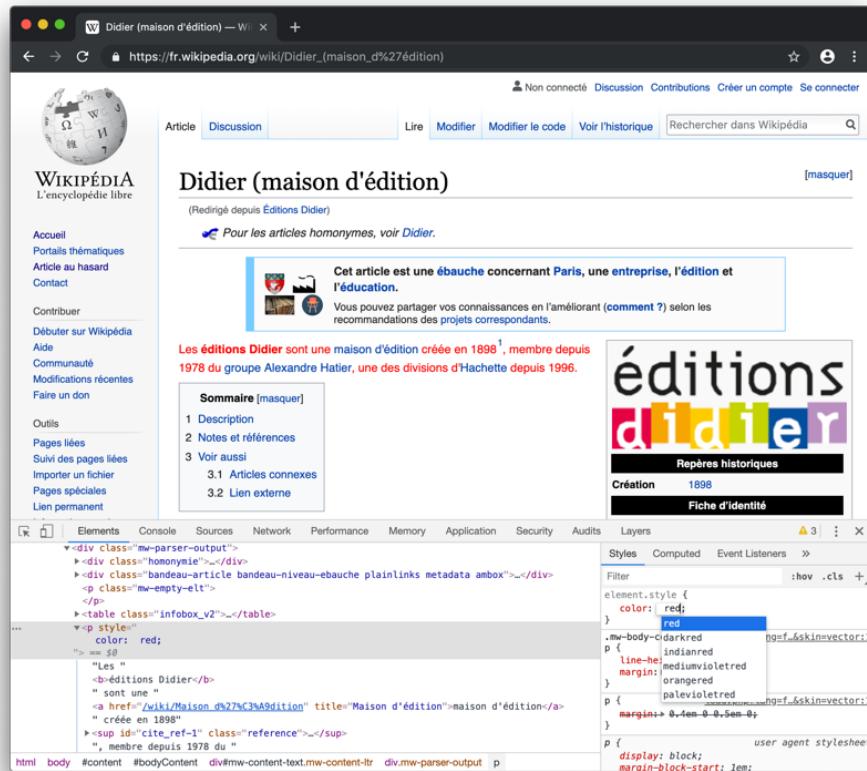
Utilisation du document 3

L'exemple ci-dessous utilise le navigateur Chrome. Les autres navigateurs, comme Firefox, ont les mêmes fonctionnalités avec une interface un peu différente.

Une fois l'inspecteur ouvert et l'onglet « Elements » actif, lorsque l'on passe le curseur sur le code HTML de l'onglet « Elements », la partie correspondante de la page Web est mise en évidence :

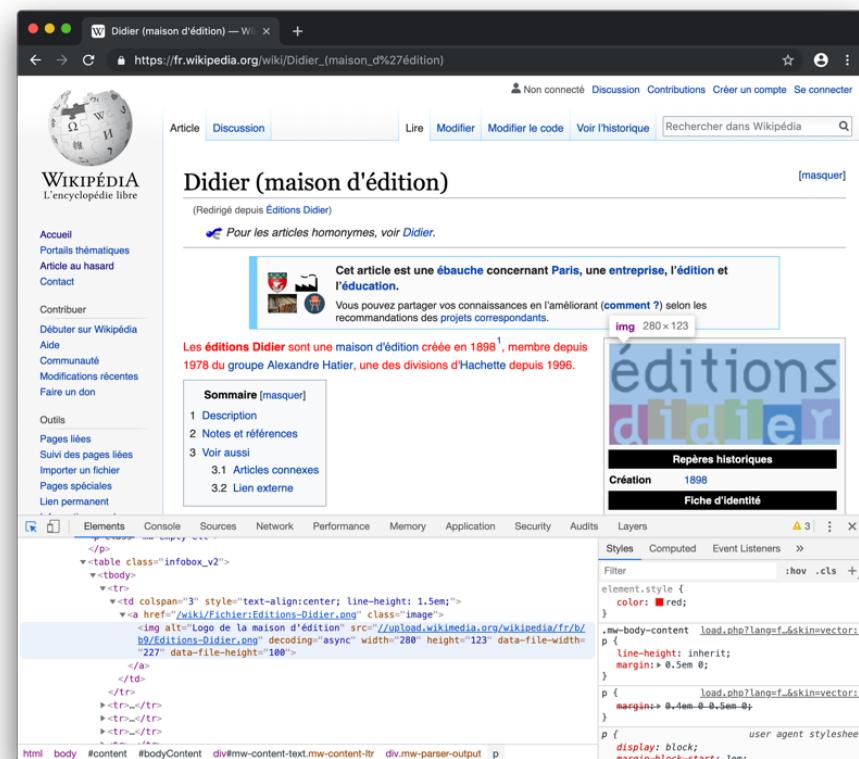


Dans la partie droite de l'inspecteur, on peut ajouter des propriétés CSS, comme ici changer la couleur du texte en rouge :



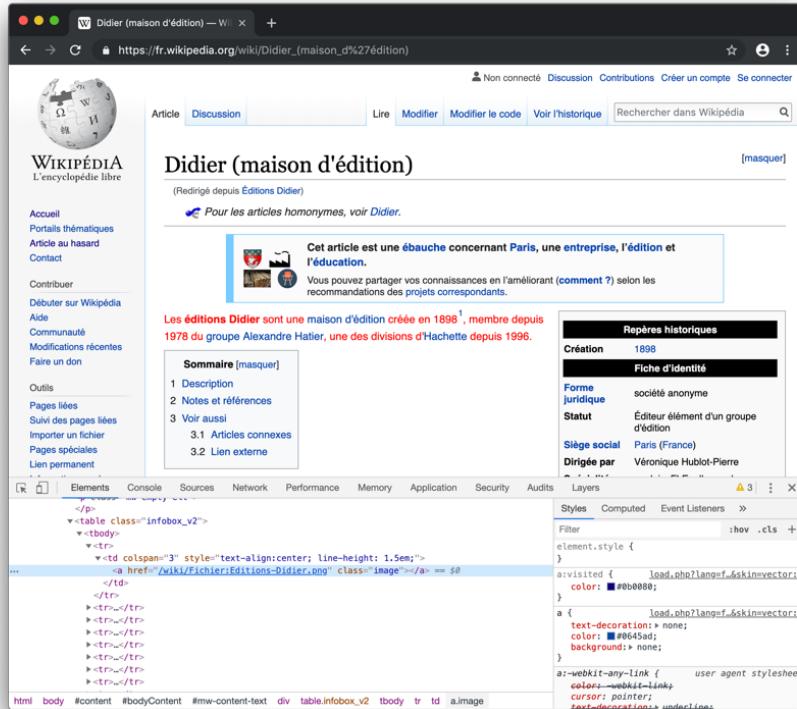
The screenshot shows a browser window displaying the Wikipedia article "Didier (maison d'édition)". On the right, the developer tools' style inspector is open, showing a list of CSS rules. One rule, specifically for the color property, is highlighted in blue, indicating it's currently being edited. The rule is: `color: red;`. This change has been applied to the text of the article, making it appear red.

Après avoir cliqué sur l'icône à gauche de la barre de menus de l'inspecteur , on peut déplacer le curseur dans le corps de la page Web et voir le code HTML correspondant dans l'inspecteur :



The screenshot shows the same Wikipedia article with the developer tools open. The 'Elements' tab is selected, and the cursor is hovering over the image of the Didier logo. The corresponding HTML code for the image tag is visible in the bottom-left panel of the developer tools, showing attributes like `alt="Logo de la maison d'édition"` and `src="//upload.wikimedia.org/wikipedia/fr/b/b/Editions-Didier.png"`.

Si l'on clique sur l'élément, la sélection correspondante dans le code HTML peut être modifiée. Par exemple, on peut enlever le code (ici le logo des éditions Didier) simplement avec la touche « Effacement » :



ACTIVITE 2

Les compétences travaillées :

Cette activité apporte des connaissances techniques sur le fonctionnement en **réseau** du Web. Elle fait le lien avec le chapitre 1 sur le réseau Internet.

Cette activité peut donner lieu à une **recherche d'information** concernant par exemple le trafic internet lié au Web : estimations du nombre de sites Web et de pages Web disponibles, du nombre de requêtes Web effectuée chaque jour (ou chaque minute), de l'impact environnemental des serveurs, des réseaux et des terminaux qui servent à faire marcher et à consulter le Web.

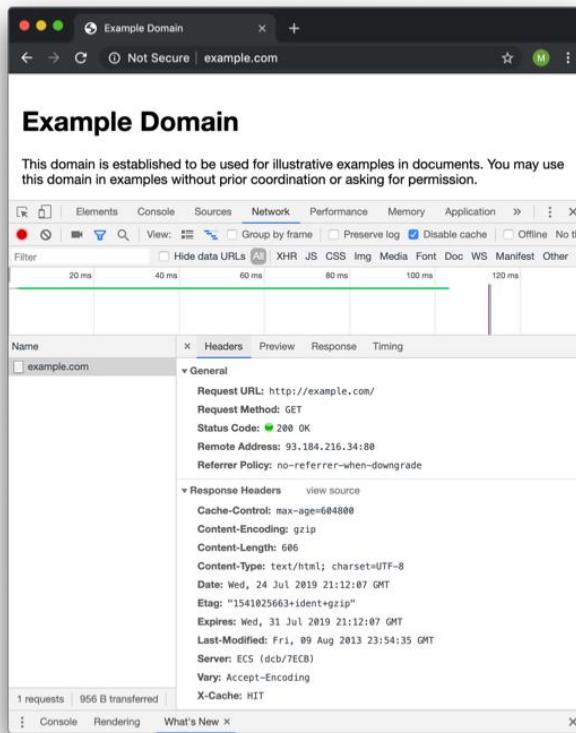
Les savoir-faire travaillés :

- Analyser une adresse URL.
- Comprendre le rôle des clients et des serveurs Web.
- Créer un formulaire simple pour interroger un site dynamique.

Complément au contenu du manuel :

Il existe plusieurs moyens de visualiser les requêtes et réponses du protocole http (ou https) :

- le site Web <https://websniffer.cc> : entrer l'URL dans le champ de saisie (par exemple `http://www.example.com`) et cliquer « Submit » ;
- l'inspecteur du navigateur : dans Chrome, ouvrir l'inspecteur et cliquer l'onglet « Network » puis entrer une URL (par exemple `http://www.example.com`) ou recharger la page courante. Cliquer ensuite sur le nom de la page dans la liste de gauche et sur l'onglet « Headers » (voir image ci-dessous) ;



- la commande « curl », disponible sur tous les systèmes d'exploitation, permet de visualiser les requêtes et réponses du protocole http depuis le terminal. Par exemple, en tapant la commande suivante (sans le %) dans le terminal :
`% curl -v http://www.example.com`
celle-ci affiche la requête envoyée et la réponse reçue pour l'exemple du document 3.

En utilisant l'une de ces méthodes pour afficher la requête et la réponse pour l'URL suivante qui correspond à la requête envoyée par le formulaire du document 4 :

<https://fr.wikipedia.org/w/index.php?search=Web>
on trouvera dans la réponse la ligne

Location: <https://fr.wikipedia.org/wiki/Web>
qui indique une redirection vers la page cherchée.

Question 1

Utilisation du document 1

- a. Les pages Web sont stockées dans les serveurs. Les navigateurs sont des clients qui obtiennent les pages Web en les demandant aux serveurs à l'aide du protocole http (ou https).
- b. Le navigateur envoie des requêtes http (ou https) au serveur en spécifiant l'URL souhaitée. Le navigateur affiche ensuite la réponse du serveur : soit la page demandée, soit un message d'erreur.

Question 2

Utilisation du document 2

- Protocole : <https://www.larousse.fr/dictionnaires/anglais-francais/request>
Chemin : /dictionnaires/anglais-francais/request
- Protocole : http://www.youtube.com/watch?v=7z_hEZABUs4
Chemin : /watch Paramètre : v=7z_hEZABUs4
- Protocole : <https://developer.mozilla.org/fr/docs/Web/HTML/Element/Form> Ancre : Exemples

La première et la troisième adresse sont sécurisées (https), la seconde ne l'est pas (http).

Question 3

Utilisation du document 3

Le navigateur effectuera quatre requêtes : une pour récupérer la page, puis une pour chacune des trois images.

Question 4

Utilisation du document 4

Exercice à faire sur ordinateur (code du formulaire dans le fichier SNT2101.html).

Sur un ordinateur connecté à internet, ouvrir la page <https://jsfiddle.net/>. Fermer le panneau bleu en haut de la page s'il apparaît. La page est divisée en quatre panneaux principaux : HTML, CSS, Javascript, Result.

Taper le code HTML dans le panneau HTML (ou le copier depuis le fichier **SNT2101.html**), puis cliquez sur le bouton « Run ». Le résultat s'affiche dans le panneau Result.

Taper un mot (par exemple « avocat ») dans le champ de saisie et cliquer sur le bouton ou taper la touche « Entrée » pour vérifier qu'on va bien sur la page Avocat de Wikipedia.

Si vous avez un éditeur de texte (pas un traitement de texte tel que Word), vous pouvez aussi taper le texte HTML dans cet éditeur et l'enregistrer dans un fichier, par exemple **wikipedia.html** et ouvrir ce fichier dans votre navigateur en double-cliquant dessus ou, si cela ne fonctionne pas, en le faisant glisser dans une fenêtre ouverte du navigateur.

Question 5

Utilisation du document 4

Exercice à faire sur ordinateur.

Il faut changer dans le formulaire l'URL de l'action du formulaire : <https://google.fr/search> et le nom du champ de saisie : **q**.

```
<h3>Rechercher sur Google : </h3>
<form action="https://google.fr/search">
    <input name="q">
```

Voir le fichier **google.html** pour la solution complète.

Note : il semble que le formulaire ne fonctionne pas depuis jsfiddle.net car Google se rend compte que la requête n'est pas faite depuis une page Web « normale » mais depuis une page qui est à l'intérieur d'une autre page.

Question 6

Utilisation du document 4

Si le serveur devait retourner le contenu de la page recherchée et que celle-ci est hébergée sur un autre serveur, il devrait la récupérer depuis ce serveur puis la retourner au client. Cela alourdit le trafic sur le réseau internet.

De plus, dans l'architecture du Web, les serveurs ne communiquent pas avec d'autres serveurs, seulement avec des clients. Le serveur retourne donc l'URL de la page recherchée au client (ici le navigateur), et celui-ci envoie une requête au serveur qui héberge cette page pour la récupérer.

ACTIVITE 3

Les compétences travaillées

Cette activité apporte des connaissances sur les **algorithmes de base** des moteurs de recherche. Elle doit encourager un **usage critique des moteurs de recherche**, en ne prenant pas leurs résultats pour argent comptant.

Ces connaissances peuvent être mises en œuvre par un **travail d'équipe ou personnel** de comparaison des résultats de différents moteurs à des requêtes identiques, ou d'un même moteur à des requêtes proches (en changeant par exemple l'ordre des mots).

Les savoir-faire travaillés

- Maîtriser un algorithme de parcours de graphe, technique utilisée dans de nombreux domaines de l'informatique.
- Varier les requêtes à un moteur de recherche pour obtenir des réponses plus adaptées.
- Parcourir la liste de réponses d'un moteur de recherche (au-delà de la première page) pour trouver des réponses peut-être moins populaires mais plus pertinentes.

Question 1

Utilisation du document 1

Action	Pages visitées	Page à visiter
On visite E, lié à A et H	E	A H
On visite A, lié à B	E A	H B
On visite H, lié à J	E A H	B J
On visite B, lié à A et C	E A H B	J C [A a déjà été visité]
On visite J, lié à F	E A H B J	C F
On visite C, lié à A	E A H B J	C F [A a déjà été visité]
On visite F, lié à C D G H	E A H B J C F	D G [C et H ont déjà été visités]
On visite D (pas de liens)	E A H B J C F D	G
On visite G (pas de liens)	E A H B J C F D G	

La visite est terminée en 9 étapes.

Question 2

Utilisation du document 1

Action	Pages visitées	Page à visiter
On visite A, lié à B	A	B
On visite B, lié à A et C	A B	C [A a déjà été visité]
On visite C, lié à A	A B C	[A a déjà été visité]

La visite est terminée, mais on a « raté » une grande partie du graphe !

Il faut donc visiter le graphe à partir de plusieurs points d'entrée, que l'on choisit en espérant qu'ils donnent accès chacun à une grande partie du graphe.

Question 3

Utilisation du document 2

Les ajouts sont indiqués en rouge ci-dessous. Les scores de 6 sont dus au fait que le mot apparaît à la fois dans le titre de la page, et dans la page elle-même.

...	
avocat	(justice, 6) (avocat, 6)
...	
droit	(justice, 2)
fruit	(avocat, 1) (fruit, 6)
juriste	(justice.html, 1)
...	

Question 4

Utilisation du document 3

Pour la page p2, l'importance devient $5 \times 2 + 25 = 35$.

Pour la page p3, l'importance devient $15 \times 2 + 10 = 40$.

Le classement s'inverse et la page p3 est donnée avant la page p2 dans la liste des résultats.

Question 5

Utilisation du document 3

« avocat fruit » : seule la page **p1** contient chacun des mots-clés, avec une importance de $20+10 = 30$.

« avocat droit » : le résultat est **p2** et **p3** (voir Doc 3).

« avocat » : les pages **p1, p2 et p3** contiennent ce mot-clé. Le classement est p1 (20), puis p3 (15) et enfin p2 (5).

On voit comment l'ajout de mots-clés permet d'avoir des réponses plus précises.

Question 6

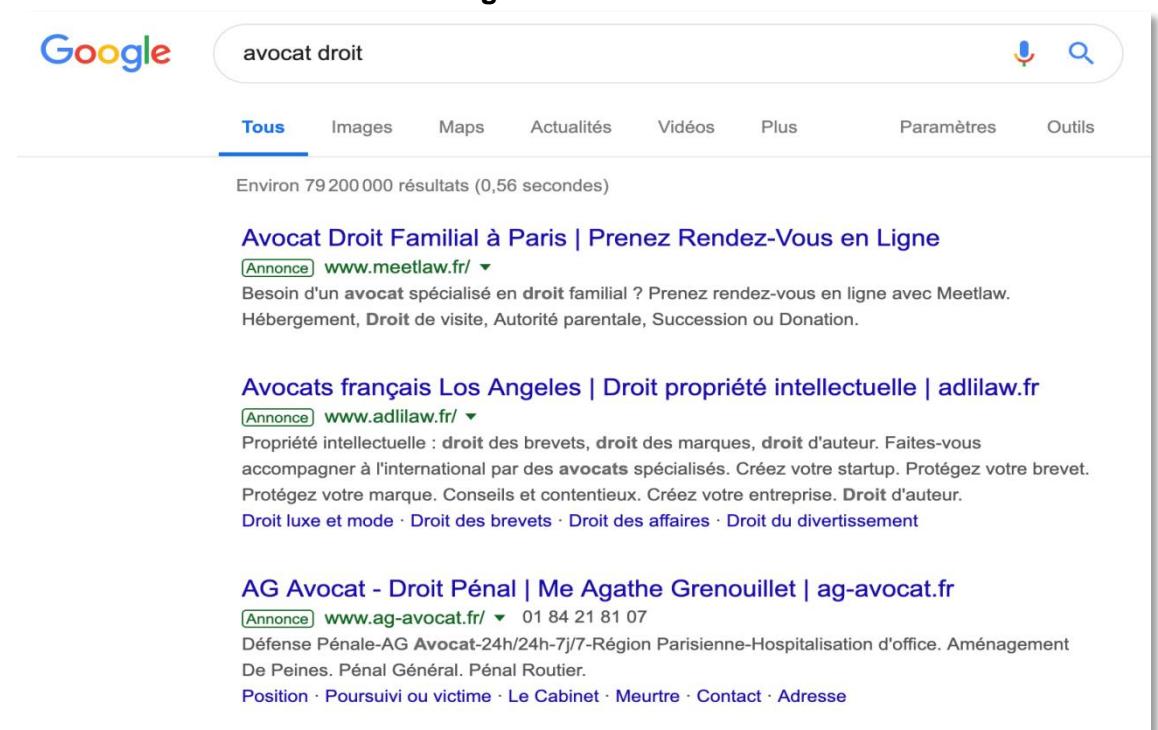
Utilisation du document 4

Exercice à faire sur ordinateur

Voici les résultats pour la même requête « avocat droit » sur <https://google.fr/>, <https://bing.com/> et <https://duckduckgo.com> : on observe que Google met en avant des liens sponsorisés, Bing la page Wikipédia et DuckDuckGo un site spécialisé.

Google et Bing fournissent également une carte (non visible sur les images) avec des cabinets d'avocat proches du lieu où a été faite la requête (ce qui indique que le navigateur dispose de cette information et l'a transmise au moteur de recherche).

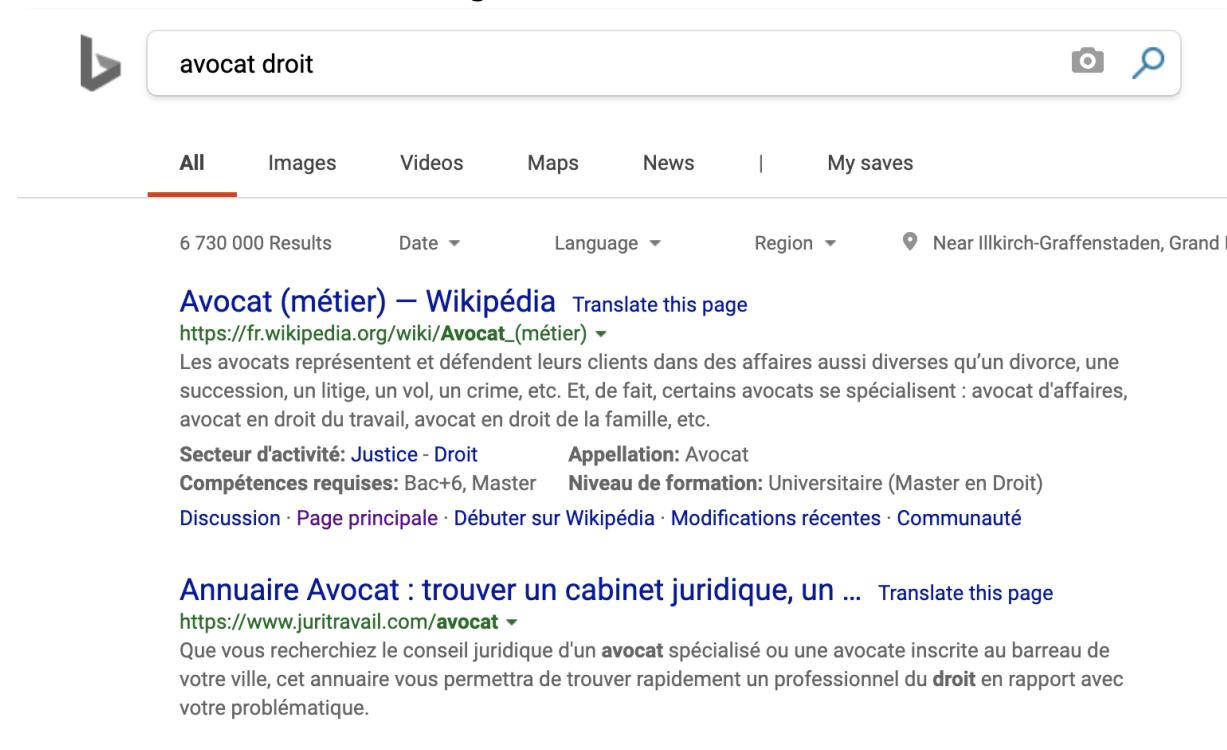
Résultats de la recherche avec Google



Google search results for "avocat droit". The results include:

- Avocat Droit Familial à Paris | Prenez Rendez-Vous en Ligne**
[Annonce] www.meetlaw.fr/ ▾
Besoin d'un avocat spécialisé en droit familial ? Prenez rendez-vous en ligne avec Meetlaw.
Hébergement, Droit de visite, Autorité parentale, Succession ou Donation.
- Avocats français Los Angeles | Droit propriété intellectuelle | adlilaw.fr**
[Annonce] www.adlilaw.fr/ ▾
Propriété intellectuelle : droit des brevets, droit des marques, droit d'auteur. Faites-vous accompagner à l'international par des avocats spécialisés. Créez votre startup. Protégez votre brevet. Protégez votre marque. Conseils et contentieux. Créez votre entreprise. Droit d'auteur.
Droit luxe et mode · Droit des brevets · Droit des affaires · Droit du divertissement
- AG Avocat - Droit Pénal | Me Agathe Grenouillet | ag-avocat.fr**
[Annonce] www.ag-avocat.fr/ ▾ 01 84 21 81 07
Défense Pénale-AG Avocat-24h/24h-7j/7-Région Parisienne-Hospitalisation d'office. Aménagement De Peines. Pénal Général. Pénal Routier.
Position · Poursuivi ou victime · Le Cabinet · Meurtre · Contact · Adresse

Résultats de la recherche avec Bing



Bing search results for "avocat droit". The results include:

- Avocat (métier) – Wikipédia** [Translate this page](https://fr.wikipedia.org/wiki/Avocat_(métier))
[https://fr.wikipedia.org/wiki/Avocat_\(métier\)](https://fr.wikipedia.org/wiki/Avocat_(métier)) ▾
Les avocats représentent et défendent leurs clients dans des affaires aussi diverses qu'un divorce, une succession, un litige, un vol, un crime, etc. Et, de fait, certains avocats se spécialisent : avocat d'affaires, avocat en droit du travail, avocat en droit de la famille, etc.
Secteur d'activité: Justice - Droit Appellation: Avocat
Compétences requises: Bac+6, Master Niveau de formation: Universitaire (Master en Droit)
Discussion · Page principale · Débuter sur Wikipédia · Modifications récentes · Communauté
- Annuaire Avocat : trouver un cabinet juridique, un ...** [Translate this page](https://www.juritravail.com/avocat)
<https://www.juritravail.com/avocat> ▾
Que vous recherchez le conseil juridique d'un avocat spécialisé ou une avocate inscrite au barreau de votre ville, cet annuaire vous permettra de trouver rapidement un professionnel du droit en rapport avec votre problématique.

Résultats de la recherche avec DuckDuckGo

The screenshot shows the DuckDuckGo search interface. The search bar at the top contains the query "avocat droit". Below the search bar are navigation links for "Web", "Images", "Videos", and "News", with "Web" being underlined. There are also filters for location ("France"), safe search ("Moderate"), and time ("Any Time").

Avocat spécialiste en droit immobilier
www.avocatdroitimmobilier.com

Un **avocat** expert en **droit immobilier** peut intervenir afin de résoudre différents conflits concernant les baux commerciaux, permis de construire, préavis, contrats en viager, l'usufruit, travaux de rénovation, saisie immobilière... L'expert en **droit** est disponible en ligne pour apporter des conseils avisés aux internautes.

Avocat Droit Public (@avocatpublic) | Twitter
https://twitter.com/avocatpublic

The latest Tweets from **Avocat Droit Public (@avocatpublic)**. Qui me contredit m'instruit (Montaigne). 06.74.13.44.59

Avocat (métier) — Wikipédia
https://fr.wikipedia.org/wiki/Avocat_(métier)

En **droit**, l'avocat est un juriste dont les fonctions traditionnelles sont de conseiller, représenter, d'assister et de défendre ses clients, personnes physiques ou morales, en justice, en plaidant pour faire valoir leurs **droits** et, plus généralement, pour les représenter.

ACTIVITE 4

Les compétences travaillées

Cette activité apporte des connaissances sur les mécanismes utilisés par les sites Web pour enregistrer des traces numériques. Elle doit encourager un **usage responsable** du Web et nourrir un **débat argumenté** sur l'équilibre entre les services apportés par ces traces et leurs risques.

Dans le prolongement de l'activité précédente sur les moteurs de recherche, on peut ainsi envisager un **travail d'équipe** pour analyser les différences entre les réponses du même moteur de recherche à la même requête sur des ordinateurs personnels de différents utilisateurs. On pourra aussi faire un **exposé** sur la façon dont différents sites mettent en œuvre le contrôle de l'utilisateur sur ses données en application du RGPD.

Les savoir-faire travaillés

- Connaitre les mécanismes de base de collecte de données personnelles, notamment les cookies.
- Comprendre et répondre de manière informée aux demandes d'autorisation de collecte de données personnelles par les sites Web.
- Gérer les réglages du navigateur (détails dans le chapitre Hygiène Numérique).

Complément au contenu du manuel :

Les trois méthodes introduites précédemment (site <https://websniffer.cc>, inspecteur du navigateur, ou commande « curl ») pour visualiser les requêtes et réponses du protocole http permettent de voir les cookies dans les échanges de messages entre client et serveur.

Ainsi, si on visualise la requête et la réponse pour l'URL

<https://fr.wikipedia.org>

on trouve dans la réponse plusieurs lignes « set-cookie » dont une pour le cookie de nom « GeolP » qui contient la localisation géographique du client. D'après la politique de confidentialité de Wikipedia (https://meta.wikimedia.org/wiki/Privacy_policy/FAQ/fr), ce cookie est utilisé pour que l'utilisateur puisse « recevoir les contenus régionaux pertinents ».

Question 1

Utilisation du document 1

a. Un cookie est une information déposée par un serveur Web dans le navigateur en réponse à une requête, et qui est transmis à nouveau au serveur lorsque l'utilisateur utilise le même navigateur pour aller sur le même site.

Les cookies permettent de reconnaître qu'un même utilisateur (plus précisément le même navigateur) visite le même site à plusieurs reprises. Le serveur peut ainsi garder trace des pages visitées par cet utilisateur.

D'autres traces sont laissées simplement en affichant une page Web (images traîtres) ou en cliquant un lien (liens trompeurs). Les images traîtres permettent de savoir par exemple lorsqu'une publicité a été affichée, et les liens trompeurs permettent de savoir quel lien a été cliqué.

b. Le cookie est créé à l'étape 4 : le cookie n'existe vraiment que lorsqu'il est stocké dans le navigateur

Si l'on efface le cookie et qu'on visite à nouveau le site, tout se passe comme lors de la première visite, mais le cookie aura un numéro différent car le serveur a déjà utilisé le numéro 123. Le serveur nous considérera alors comme un nouvel utilisateur.

Question 2

Utilisation du document 2

Seul le lien A est un lien trompeur : il envoie sur le site tracker.com alors que le texte du lien semble pointer vers Wikipédia. Si on clique sur le lien, la requête est envoyée au site tracker.com, avec un paramètre qui est la page Wikipédia sur les cookies.

Le lien B est un lien vers la page de recherche de Wikipedia.

Question 3

Utilisation du document 3

Le moteur peut ainsi savoir quelle réponse a choisi l'utilisateur : c'est un lien trompeur qui semble aller vers la Page A, mais qui en réalité envoie une requête au moteur de recherche avec comme paramètre l'URL de la page visée. Lorsqu'il reçoit cette requête, le moteur note quel lien a été choisi, ce qui peut l'aider à améliorer son algorithme de classement des réponses. Puis il envoie une réponse de type « Found » (voir Activité 2, document 4, page 37) avec l'adresse de la Page A, de telle sorte que le navigateur aille chercher cette page. Pour l'utilisateur, tout se passe comme si le lien pointait directement vers la Page A.

Si le lien allait directement sur pageA.html, le moteur de recherche ne saurait pas quelle réponse à la requête de recherche l'internaute a choisi.

Question 4

Utilisation du document 3

Exercice à faire sur ordinateur Windows.

Le logiciel CookieViz fonctionne seulement sur machines Windows. Il peut être téléchargé à l'adresse suivante : <https://github.com/LINCnil/CookieViz/releases>.

Une vidéo montrant son fonctionnement est disponible ici :

<https://www.dailymotion.com/video/x18gigz>.

Question 5

Utilisation du document 4

La page de gauche permet de voir tous les sites concernés, et de désactiver tous les cookies avec un seul clic, grâce au bouton « Tout interdire ».

La page de droite est moins explicite. Il faut cliquer sur les marques à droite de chaque réglage pour avoir des détails, et elle a seulement un bouton pour tout accepter mais pas tout refuser. Si l'on clique sur le bouton « Partenaires potentiels du site » en bas, on obtient une liste de plus de 370 sites, qu'il faut désactiver un par un ! Même si cette page est conforme au RGPD (règlement général de protection des données), elle est clairement conçue pour encourager l'utilisateur à accepter tous les cookies du site, afin de collecter le maximum d'information.

ACTIVITE 5

Les compétences travaillées

Cette activité apporte des connaissances sur les règles de publication sur le Web, dans le but de promouvoir un **usage responsable** et conforme à la loi.

Elle se prête à l'organisation de **débats contradictoires** sur l'utilisation des ressources disponibles en ligne et les bonnes pratiques, à partir de cas d'usages concrets mettant en jeu le droit à l'image, le plagiat ou la cession de droits patrimoniaux. Elle se prête aussi à une **analyse critique et comparative** des conditions d'utilisation de différents sites Web. Elle peut aussi donner lieu à un **travail collectif** pour la création d'un petit site Web donnant des conseils du type « Il faut » / « Il ne faut pas ».

Les savoir-faire travaillés

- Connaître ses droits et devoirs en tant qu'auteur de contenus.
- Comprendre les conditions d'utilisations des sites Web vis-à-vis des droits d'auteur ;
- Choisir une licence lors de la publication de contenus.
- Connaître ses droits vis-à-vis des données et informations personnelles qui viendraient à être publiées.

Question 1

Utilisation du document 1

- Assumer la responsabilité de la publication, même si celle-ci est faite de manière anonyme ou sous pseudonyme.
- Respecter les limites de la liberté d'expression (diffamation, incitation à la haine, etc.).
- Respecter les droits d'auteurs si l'on utilise des contenus dont on n'est pas l'auteur, et le droit à l'image si l'on poste des images ou des vidéos d'autres personnes.

Question 2

Utilisation du document 2

a. « Perpétuel » signifie que l'auteur conserve le droit moral même après sa mort. Il peut être exercé par ses ayants droits après la mort de l'auteur, même si l'œuvre tombe dans le domaine public (c'est-à-dire que les droits patrimoniaux se sont éteints).

« Inaliénable » signifie que l'auteur ne peut renoncer à son droit moral ni le céder, même s'il le souhaite.

« Imprescriptible » signifie que le droit moral ne s'éteint pas même si l'œuvre n'est pas exploitée, et même s'il n'y a plus d'ayants droits.

b. Le plagiat est une violation du droit moral puisqu'il consiste à s'attribuer la paternité d'un œuvre qui n'est pas la sienne.

Mettre un morceau de musique ou un film sur sa page Web sans autorisation des auteurs viole en général les droits patrimoniaux, sauf si ces contenus ont été publiés avec une licence gratuite.

c. Oui, mais l'internaute autorise Facebook à en faire à peu près ce qu'il veut, y compris à autoriser d'autres à en faire ce qu'ils veulent (car la licence est transférable et Facebook peut accorder des sous-licences à des tiers).

Question 3

Utilisation du document 3

a. On peut publier des images d'une manifestation si celle-ci est un événement d'actualité ou à valeur historique, à condition de respecter la dignité de la personne humaine.

b. Des propos diffamatoires ou discriminatoires (racistes, sexistes, etc.) portent atteinte à la personne et engagent la responsabilité de son auteur. La discrimination raciale, par exemple, est passible de deux ans de prison et/ou 30 000 euros d'amende.

Complément : l'article 225-1 du Code pénal énumère 20 critères de discrimination :

« Constitue une discrimination toute distinction opérée entre les personnes physiques sur le fondement de leur origine, de leur sexe, de leur situation de famille, de leur grossesse, de leur apparence physique, de la particulière vulnérabilité résultant de leur situation économique, apparente ou connue de son auteur, de leur patronyme, de leur lieu de résidence, de leur état de santé, de leur perte d'autonomie, de leur handicap, de leurs caractéristiques génétiques, de leurs mœurs, de leur orientation sexuelle, de leur identité de genre, de leur âge, de leurs opinions politiques, de leurs activités syndicales, de leur capacité à s'exprimer dans une langue autre que le français, de leur appartenance ou de leur non-appartenance, vraie ou supposée, à une ethnie, une Nation, une prétendue race ou une religion déterminée ».

Question 4

Utilisation du document 4

« Partage dans les mêmes conditions » autorise des modifications, mais impose que celles-ci soient diffusées avec les mêmes conditions que l'original.

« Pas de modification » n'autorise pas les modifications de l'œuvre.

Cela n'a donc pas de sens de simultanément autoriser et ne pas autoriser les modifications.

Question 5

Utilisation du document 5

- Des articles de presse ont relaté des accusations de malversations ou de violences à l'encontre d'une personne. La justice a acquitté cette personne, mais les articles continuent d'apparaître en tête de liste lorsque l'on tape le nom de cette personne dans un moteur de recherche : la personne peut faire valoir son droit au déréférencement pour que ces résultats n'apparaissent plus (mais les pages contenant les informations, elles, ne sont pas retirées).

- Un internaute a publié sur son réseau social des contenus ou des images qu'il regrette car ils pourraient le pénaliser dans une recherche d'emploi. Il peut demander au réseau social le retrait de ces données.

Dans les deux cas, le moteur de recherche ou le réseau social peuvent refuser l'effacement ou le déréférencement. La jurisprudence est encore peu fournie sur ces droits récents, et il est nécessaire d'équilibrer le droit à l'information avec le droit à l'oubli.

Je vérifie mes connaissances

Solutions détaillées

1. QCM

- 1. b.
- 2. c.
- 3. a.
- 4. a. et c.
- 5. b.

2. Vrai/Faux

- a. **FAUX.** Cette notion apparaît dès les années 1960 dans le projet Xanadu du sociologue Ted Nelson.
- b. **FAUX.** Seul le protocole HTTPS est sécurisé.
- c. **VRAI.**
- d. **FAUX.** Elles sont contenues dans le <head> et non dans le <body>.
- e. **VRAI.**
- f. **VRAI.**
- g. **FAUX.** Cela peut se faire à l'aide de liens trompeurs ou d'images traîtres.
- h. **VRAI.**

3. Organisation des connaissances

- a. Un index qui indique, pour chaque mot, les pages qui le contiennent avec l'importance du mot dans la page. Éventuellement la popularité de chaque page (calculée par exemple avec l'algorithme PageRank).
L'index est calculé en parcourant les pages Web et en analysant leur contenu. Le classement des réponses utilise l'importance des mots dans chaque page et la popularité des pages.

- b.** Les cookies servent à enregistrer des informations à caractère non personnel comme des préférences ou le contenu d'un panier.

Le RGPD impose d'informer les utilisateurs sur la collecte de données à caractère personnel et d'obtenir leur consentement avant d'effectuer cette collecte.

4. Vers une définition

- a. Les **balises** d'une page HTML définissent sa **structure**, qui peut être représentée par un **arbre**.
- b. Dans le protocole http, le **client** envoie des requêtes au **serveur** et celui-ci retourne des contenus Web.

- c. Un **cookie** enregistre des informations qui seront transmises au serveur lors d'une prochaine requête.
- d. Certains droits **patrimoniaux** sont cédés par les internautes lorsqu'ils postent des contenus sur un réseau social.

5. Trouver l'intrus

<head> est un élément pour les métadonnées.

 : gras

 : liste

<p> : paragraphe

 : élément de liste

L'intrus est donc « titre » dont la balise est <title>.

JE M'ENTRAÎNE

6. Le code HTML

Les erreurs sont surlignées en jaune ci-dessous :

```
Il manque la balise <html>
Il manque la balise <head>
<title>Cherchez l'erreur</title>
<meta charset="utf-8"> Cet élément doit apparaître de préférence en premier car
il spécifie le jeu de caractères utilisé dans la suite
Il manque la balise de fin </head>
Il manque la balise <body>
<h2>Ouvrez l'oeil</h3> La balise de fin </h3> ne correspond pas à celle de début
<p>Ce paragraphe contient <b>du texte <i>très</b> important</i>.
L'élément <b> se termine au milieu de l'élément <i>
Il manque la balise de fin </p>
<p>Courses à faire : /p> Il manque le caractère < au début de la balise de fin
Il manque la balise de début de liste <ul>
<li>Acheter du pain Il manque la balise de fin </li>
<li>Passer au pressing Il manque la balise de fin </li>
Il manque la balise de fin de liste </ul>
</body>
Il manque la balise de fin </html>
```

La version corrigée est la suivante :

```
<html>
<head>
    <meta charset="utf-8">
    <title>Cherchez l'erreur</title>
</head>
<body>
    <h2>Ouvrez l'oeil</h2>
    <p>Ce paragraphe contient <b>du texte <i>très</i></b><i> important</i>. </p>
    <p>Courses à faire : </p>
    <ul>
        <li>Acheter du pain</li>
        <li>Passer au pressing</li>
    </ul>
</body>
</html>
```

Il est à noter que malgré toutes ces erreurs, le document initial sera affiché dans un navigateur de façon à peu près correcte. Ci-dessous la version initiale à gauche et corrigée à droite :

Ouvrez l'oeil

Ce paragraphe contient **du texte très important.**

Courses à faire : /p>

- Acheter du pain
- Passer au pressing

Ouvrez l'oeil

Ce paragraphe contient **du texte très important.**

Courses à faire :

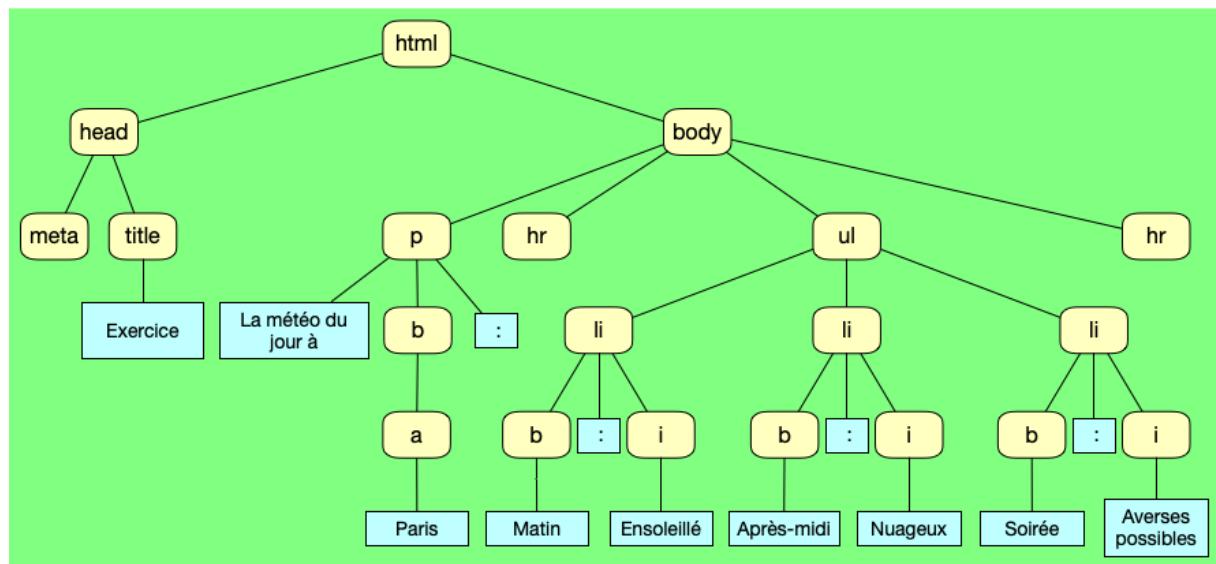
- Acheter du pain
- Passer au pressing

7. De la page au code

a. Code HTML de la solution:

```
<html>
<head>
    <meta charset="utf-8">
    <title>Exercice</title>
</head>
<body>
    <p>La météo du jour à <b><a href="">Paris</a></b> :</p>
    <hr>
    <ul>
        <li><b>Matin</b> : <i>Ensoleillé</i></li>
        <li><b>Après-midi</b> : <i>Nuageux</i></li>
        <li><b>Soirée</b> : <i>Averses possibles</i></li>
    </ul>
    <hr>
</body>
</html>
```

b. Voici l'arbre complet :



8. Les feuilles de style

a. On définit trois classes (repérées par le fait que leur nom commence par un point) pour chacun des trois types d'affichage :

```
.soleil { background-color: yellow; }
.nuage { background-color: lightgrey; }
.pluie { background-color: darkgrey; }
```

b. Le fichier HTML modifié est le suivant. On importe la feuille de style et on ajoute l'attribut « class » aux trois éléments de la liste pour que les règles correspondantes de la feuille de style leur soient appliquées. Les modifications sont surlignées en jaune :

```
<html>
<head>
    <meta charset="utf-8">
    <!-- importation de la feuille de style externe : -->
    <link rel="stylesheet" href="Q8-style.css">
    <title>Exercice</title>
</head>
<body>
    <p>La météo du jour à <b><a href="">Paris</a></b> :</p>
    <hr>
    <ul>
        <li class="soleil"><b>Matin</b> : <i>Ensoleillé</i></li>
        <li class="nuage"><b>Après-midi</b> : <i>Nuageux</i></li>
        <li class="pluie"><b>Soirée</b> : <i>Averses possibles</i></li>
    </ul>
    <hr>
</body>
</html>
```

Dans l'exemple ci-dessus on a utilisé une feuille de style externe, référencée par la balise `<link>` dans l'en-tête du document. C'est la méthode recommandée car ainsi une même feuille de style peut être utilisée par plusieurs pages. On peut aussi intégrer directement la feuille de style dans le fichier HTML. La partie `<head>` devient alors (le reste est inchangé) :

```
<head>
    <meta charset="utf-8">
    <style>
        .soleil { background-color: yellow; }
        .nuage { background-color: lightgrey; }
        .pluie { background-color: darkgrey; }
    </style>
    <title>Exercice</title>
</head>
```

Dans les deux cas le résultat est le suivant :

La météo du jour à Paris :

- **Matin** : Ensoleillé
- **Après-midi** : Nuageux
- **Soirée** : Averses possibles

9. Les moteurs de recherche

Exercice à faire sur ordinateur.

a. Une recherche avec la requête « moteurs de recherche » sur Google donne comme premier résultat une page Wikipédia (https://fr.wikipedia.org/wiki/Liste_des_moteurs_de_recherche) qui liste les moteurs de recherche. On trouve aussi, bien placé dans la liste des réponses, le moteur de recherche (développé en France) Qwant (<https://www.qwant.com/?l=fr>).

b. Lorsque l'on fait une requête sur le moteur Qwant indiqué ci-dessus, l'URL est de la forme <https://www.qwant.com/?q=moteurs%20de%20recherche&t=web>

On peut donc aisément adapter le formulaire vu dans l'Activité 2 (document 4, page 37) (fichier Q9-qwant.html):

```
<html>
  <head>
    <meta charset="utf-8">
    <title>Interroger Qwant</title>
  </head>
  <body>
    <h3>Rechercher sur Qwant :</h3>
    <form action="https://www.qwant.com">
      <input name="q">
      <input type="submit">
    </form>
  </body>
</html>
```

10. L'algorithme

En partant de A :

Action	Visitées	À visiter
On visite A, lié à B et E	A	B E
On visite B, lié à A, C et D	A B	E C D
On visite E, lié à A et D	A B E	C D
On visite C, lié à A et E	A B E C	D
On visite D, lié à C	A B E C D	
Visite terminée		

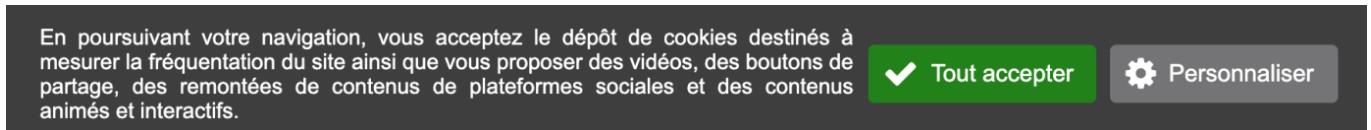
En partant de D :

Action	Visitées	À visiter
On visite D, lié à C	D	C
On visite C, lié à A et E	D C	A E
On visite A, lié à B et E	D C A	E B
On visite E, lié à A et D	D C A E	B
On visite B, lié à A, C et D	D C A E B	
Visite terminée		

11. Les cookies

Exercice à faire sur ordinateur.

- a. Lors de la première visite le bandeau ci-dessous s'affiche :



Si l'on clique sur « Personnaliser », la fenêtre suivante s'affiche :

A screenshot of the "Gestion des cookies" (Cookie Management) window. It includes sections for "Mesure d'audience", "Campagnes d'information", and "Boutons de partage de contenus", each with checkboxes for various tracking services like Google Analytics, ClickTale, Doubleclick, Facebook, and Twitter. At the bottom is a large black button labeled "APPLIQUER MES CHOIX" (Apply my choices).

Après avoir complété les choix (par exemple « Tout interdire » puis « Appliquer mes choix »), on peut voir que si l'on va à nouveau sur le site www.education.gouv.fr, le bandeau d'acceptation des cookies ne s'affiche plus.

- b. Pour trouver les cookies enregistrés dans le navigateur Chrome, on peut aller directement sur la page d'URL <chrome://settings/siteData>. On peut aussi trouver cette page comme suit :

- Aller dans les réglages du navigateur (commande « Préférences » du menu « Chrome »)
- Cliquer sur « Paramètres avancés » en bas de la page, puis sur « Paramètres du contenu » en bas du panneau « Confidentialité et sécurité », puis sur « Cookies », puis sur « Afficher l'ensemble des cookies et données de sites ».

Une fois sur la page qui liste tous les cookies, on peut rechercher un cookie en haut à droite de la page. En tapant « education » on voit apparaître deux sites et 6 cookies :

A screenshot of the Chrome "Site Data" settings page. A search bar at the top right contains the text "education". Below it, a list shows two entries: "education.gouv.fr" with 3 cookies and "www.education.gouv.fr" with 3 cookies. Each entry has a trash can icon to its right.

Si on clique sur « www.education.gouv.fr », on voit 3 cookies, dont un s'appelle « CONSENT_RGPD », avec le contenu suivant :

The screenshot shows the Firefox cookie manager interface. At the top, there's a search bar and a button labeled "Tout supprimer". Below that, a list of cookies is displayed:

- CONSENT_RGPD**:
 - Nom: CONSENT_RGPD
 - Contenu: versionDate:1548263081|googleanalytics_visite:false|xiti_visite:false|clicktale_campagne:false|doubleclick_campaigne:false|facebookpixel_campaigne:false|facebook:false|twitter:false
 - Domaine: www.education.gouv.fr
 - Chemin d'accès: /
 - Envoyer pour: Toutes sortes de connexions
 - Accessible aux scripts: Oui
 - Date de création: lundi 15 avril 2019 à 14:31:58
 - Date d'expiration: mardi 14 avril 2020 à 14:31:58
- PHPSESSID**:
 - Domaine: www.education.gouv.fr
- style**:
 - Domaine: www.education.gouv.fr

Le contenu du cookie indique les options que l'on a choisies. Ici, toutes les options ont été refusées et sont mentionnées comme « false » (faux).

c. Depuis l'étape précédente, cliquer sur la croix à droite du cookie CONSENT_RGPD. Aller à nouveau sur www.education.gouv.fr et constater qu'il affiche le bandeau d'acceptation des cookies.

Avec le navigateur Firefox, la liste des cookies est accessible depuis le menu « Préférences », cliquer sur « Vie privée et sécurité » dans le menu à gauche, puis défiler jusqu'à « Cookies et données de sites » et cliquer sur « Gérer les données » :

The screenshot shows the Firefox preferences window under the "Vie privée et sécurité" tab. On the left, there are several options:

- Général
- Accueil
- Recherche
- Vie privée et sécurité
- Compte Firefox

In the main area, there's a message: "avez confiance. Découvrez comment". Below it, there are two settings for tracking:

- Envoyer aux sites web un signal « Ne pas me pister » indiquant que vous ne souhaitez pas être pisté
 - Toujours
 - Seulement quand Firefox est paramtré pour bloquer les traqueurs connus

At the bottom, there's a section titled "Cookies et données de sites" with the following content:

Le stockage des cookies, du cache et des données de sites utilise actuellement 243 Mo d'espace disque. [En savoir plus](#)

Supprimer les cookies et les données des sites à la fermeture de Firefox

[Effacer les données...](#) [Gérer les données...](#) [Gérer les permissions...](#)

12. Les licences

Exercice à faire sur ordinateur.

Sur la page Wikipédia du Web (https://fr.wikipedia.org/wiki/World_Wide_Web) figure une photo de la plaque commémorant la création du Web. En cliquant dessus, la photo s'affiche en grand avec, en dessous, sa légende et la licence d'utilisation :

Où la toile est née, plaque commémorant la création du *World Wide Web* dans les locaux du CERN.

 Plus de détails

 Max Braun — originally posted to [Flickr](#) as WHERE THE WEB WAS BORN

 CC BY-SA 2.0 afficher les

La mention CC BY-SA 2.0 en bas à droite signifie que l'image est publiée sous licence Creative Commons 2.0, avec comme conditions l'attribution et le partage dans les mêmes conditions :   Les détails apparaissent en cliquant à côté de la mention ou sur le bouton « Plus de détails ».

Si, dans la même page Wikipédia, on clique sur le logo du Web, on voit que celui-ci a été mis dans le domaine public, c'est-à-dire qu'il peut être utilisée sans restriction.

13. Les données personnelles

Exercice à faire sur ordinateur.

a. Le réseau social Twitter a de nombreuses règles et politiques d'utilisation (voir <https://help.twitter.com/fr/rules-and-policies>). Concernant les données personnelles, un long document (plus de 6000 mots, soit plus de 20 pages de texte) détaille les informations collectées et leurs usages (<https://twitter.com/fr/privacy#update>). Dans les conditions d'utilisation (<https://twitter.com/fr/tos#update>) on trouve une cession de droits similaire à celle du réseau Facebook qui figurent dans le manuel (Acrtivité 5, page 42) :

« En soumettant, en publiant ou en affichant un Contenu sur ou via les Services, vous nous accordez une licence mondiale, non exclusive et libre de redevances (incluant le droit de sous-licencier), nous autorisant à utiliser, copier, reproduire, traiter, adapter, modifier, publier, transmettre, afficher et distribuer ce Contenu sur tout support et selon toute méthode de distribution (actuellement connus ou développés dans le futur). Cette licence nous autorise à mettre votre Contenu à disposition du reste du monde et autorise les autres à en faire de même. »

L'utilisateur de Twitter autorise donc le service à faire à peu près ce qu'il veut des données que l'utilisateur poste sur le réseau.

b. Toujours sur le réseau social Twitter, il suffit d'aller sur le profil de l'utilisateur, de cliquer « Paramètres et confidentialité » puis « Vos données Twitter » sur la gauche. En bas de la page qui s'affiche figure un panneau pour télécharger les données. Lorsque l'on clique sur le bouton « Demander les données » la récupération est lancée et lorsque les données sont prêtes un message est envoyé à l'utilisateur par courrier électronique.

Téléchargez vos données Twitter

Vous pouvez demander un fichier contenant les informations dont nous pensons qu'elles sont pertinentes et utiles pour vous. Lorsque ce fichier sera prêt à être téléchargé, vous recevrez une notification et un email envoyé à nicolas.taffin@gmail.com, tous deux dotés d'un lien.

Twitter

Récupération des données

Periscope

Demander les données

FAB LAB

Télécharger le fichier fablab.zip et extraire son contenu dans un dossier de nom fablab :

```
fablab /  
    fablab.py  
    fablab2.py  
    html5.png  
    index.html  
    page1-css.html  
    page1.html  
    page2.html  
    serveur.py  
    style.css
```

Les fichiers destinés aux élèves sont dans le fichier SNT2023.zip.

Les commandes ci-dessous (indiquées par %) correspondent à une utilisation avec un terminal Linux.

1 Créer un serveur Web de pages statiques

Créer un dossier www au même niveau que le dossier fablab et y copier le fichier index.html :

```
% mkdir www  
% cd www  
% cp ../fablab /index.html .  
% python -m SimpleHTTPServer 8000
```

Note : lancer si possible le serveur python dans une autre fenêtre terminal, car ainsi il pourra rester actif alors que l'on fait des modifications.

Entrer l'URL <http://localhost:8000/index.html> dans un navigateur ou cliquer directement sur le lien.

La page index.html s'affiche, mais les liens de fonctionnent pas car les pages page1.html et page2.html n'existent pas.

Pour arrêter le serveur, taper Control-C dans la fenêtre où il a été lancé.

2 Ajouter du contenu au serveur

Copier les fichiers page1.html et page2.html dans le dossier www :

```
% cp ../fablab /page1.html ../fablab /page2.html .
```

Relancer le serveur python si besoin :

```
% python -m SimpleHTTPServer 8000
```

Entrer l'URL <http://localhost:8000/index.html> ou cliquer directement sur le lien.

Cette fois-ci les liens vers page1.html et page2.html doivent fonctionner.

Cependant le logo HTML5 ne s'affiche pas dans la page1 car le fichier html5.png n'existe pas.

Copier ce fichier :

```
% cp ../fablab /html5.png .
```

Aller à nouveau sur la page1 et vérifier que le logo s'affiche

Copier la feuille de style style.css et la version modifiée de la page1, page1-css.html :

```
% cp ../fablab /style.css .  
% cp ../fablab /page1-css.html page1.html
```

Aller à nouveau sur la page1 et vérifier que le style est pris en compte.

Arrêter le serveur python car on va lancer un serveur différent dans la prochaine étape.

3 Créer un serveur avec une page dynamique qui donne l'heure

Copier les fichiers `serveur.py` et `fablab.py` dans le dossier `www` :

```
% cp .../fablab /serveur.py .../fablab /fablab.py .
```

Lancer (si possible dans une autre fenêtre terminal) le serveur :

```
% python fablab.py
```

Entrer l'URL <http://localhost:8000/date> dans un navigateur ou cliquer le lien. Le navigateur devrait afficher la date et l'heure. Recharger la page plusieurs fois pour vérifier qu'elle est bien recalculée à chaque fois.

Dans le fichier `fablab.py`, la fonction `date`, qui est appelée lorsque l'on clique ce lien, est la suivante :

```
def date(path):
    date = datetime.now()
    jour = date.strftime("%d %B %Y")
    heure = date.strftime("%H%M")
    return pageHTML('Date', '<p>Nous sommes le %s, il est %s.</p>' % (jour, heure))
```

Cette fonction calcule la date et l'heure, puis retourne une page HTML qui a pour titre « Date » et donc le contenu est un paragraphe contenant la date et l'heure.

4 Ajouter une page dynamique qui liste les fichiers

Copier le fichier `fablab2.py` pour remplacer `fablab.py` :

```
% cp .../fablab /fablab2.py fablab.py
```

Arrêter le serveur (Control-C) et le relancer afin qu'il prenne en compte cette nouvelle version :

```
% python fablab.py
```

Entrer l'URL <http://localhost:8000/> dans le navigateur ou cliquer sur le lien. La liste des fichiers qui sont dans le dossier `www` devrait s'afficher.

Pour cela, on a ajouté dans `fablab.py` une fonction `dir` qui liste l'ensemble des fichiers et crée une chaîne de caractères avec le code HTML de la page :

```
def dir(path):
    titre = 'Liste des fichiers de '+path
    liste = '<h2>' + titre + '</h2><ul>'
    for name in listdir(curdirec + path):
        liste += '<li><a href="' + name + '">' + name + '</a></li>'
    return pageHTML(titre, liste)
```

Cette fonction utilise le module « `listdir` » qui permet d'énumérer les fichiers du dossier « `path` ». On construit dans la variable `liste` une liste HTML avec un élément pour chaque fichier, qui est aussi un lien vers le fichier, de la forme :

```
<li><a href="nom-du-fichier">nom-du-fichier</a></li>
```

Enfin, la ligne

```
pageDynamique('/', dir)
```

à la fin de `fablab.py` indique au serveur d'appeler la fonction `dir` lorsque le chemin « `/` » du serveur est chargé, ce qui correspond à l'URL <http://localhost:8000/>.

Note : On peut lancer la commande

```
% python fablab2.py
```

depuis le dossier `fablab` pour tester directement la dernière étape.

La page du codeur

Télécharger le fichier moteur.zip et extraire son contenu dans un dossier de nom moteur. Il contient les fichiers de départ donnés aux étudiants et les solutions des différentes étapes (fichiers avec le suffixe « -sol »).

```
moteur /  
    codeur.py  
    index.html  
    moteur.py  
    moteur1.py  
    moteur1-sol.py  
    moteur2.py  
    moteur2-sol.py  
    moteur3.py  
    moteur3-sol.py  
    moteur4.py  
    moteur4-sol.py  
    page1.html  
    page2.html  
    search.html  
    serveur.py
```

Les fichiers destinés aux élèves sont dans le fichier SNT2024.zip.

Pour pouvoir travailler, il faut installer le module python « BeautifulSoup » qui permet d'analyser le contenu des pages Web :

```
% pip install beautifulsoup4
```

1 Construire l'index

Lancer la commande

```
% python moteur1.py
```

qui appelle trois fois la fonction indexerPage (qui ne fait rien) et affiche l'index (qui est vide).

La fonction indexerPage doit ajouter le contenu d'une page à l'index. Pour cela elle doit lire une page, extraire ses mots, et les ajouter un par un à l'index. L'index est un dictionnaire qui associe à chaque mot une liste des pages qui contiennent ce mot. Si le mot est déjà dans l'index, on ajoute la page à la liste, sinon on crée une nouvelle entrée dans l'index avec cette seule page :

```
def indexerPage(page):  
    global index  
    contenu = lirePage(page)      # lire la page  
    for mot in listeMots(contenu): # pour chaque mot de la page  
        if mot:                  # ajouter le mot à l'index  
            if mot in index:  
                index[mot].append(page)  
            else:  
                index[mot] = [page]
```

Lancer la commande

```
% python moteur1-sol.py
```

qui contient la fonction complétée et vérifier que l'index est correct.

2 Parcourir les pages par leurs liens

Lancer la commande

```
% python moteur2.py
```

qui appelle la fonction `parcourir` avec comme page de départ la page `index.html`. Tant qu'il reste des pages à visiter, cette fonction prend une page à visiter, lit son contenu pour en extraire les liens, et appelle la fonction `visiterPage` avec le nom de la page et la liste des liens. Comme la fonction `visiterPage` ne fait rien, le programme boucle indéfiniment sur la page `index.html`. Taper Control-C pour l'interrompre.

La fonction `visiterPage` doit transférer la page des pages à visiter aux pages visitées puis, pour chaque page dans la liste des liens, l'ajouter aux pages à visiter sauf si elle a déjà été visitée et qu'elle n'est pas déjà dans les pages à visiter. La compléter comme suit :

```
def visiterPage(page, listeLiens):
    visitees.append(page)      # ajouter 'page' aux pages visitées
    aVisiter.remove(page)      # retirer 'page' des pages déjà visitées
    for page in listeLiens:   # pour chaque page dans la liste de liens
        if not page in visitees and not page in aVisiter:
            aVisiter.append(page)    # l'ajouter à la liste des pages à visiter
```

Lancer la commande

```
% python moteur2-sol.py
```

qui contient la fonction complétée et vérifier que les pages sont visitées une et une seule fois.

3 Construire le graphe du Web

Lancer la commande

```
% python moteur3.py
```

qui appelle la fonction `parcourir` avec comme page de départ la page `index.html`, puis la fonction `afficherGraphe` qui affiche le graphe du Web. Comme la fonction `ajouterLiens`, qui est appelée par la fonction `parcourir` à chaque page visitée pour construire le graphe, ne fait rien, le graphe affiché est vide.

Le graphe est un dictionnaire Python qui associe à chaque page la liste des pages vers lesquelles elle a des liens. La fonction `ajouterLiens` doit donc ajouter la destination de chaque lien à l'entrée de la page dans le graphe. La compléter comme suit :

```
def ajouterLiens(page, listeLiens):
    global graphe
    for destination in listeLiens:      # pour chaque destination
        if page in graphe:              # si la page est dans le graphe
            graphe[page].append(destination)  # ajouter la destination
        else:
            graphe[page] = [destination]    # créer la page et la destination
```

Lancer la commande

```
% python moteur3-sol.py
```

qui contient la fonction complétée et vérifier que le graphe des pages est correct.

4 Calculer l'importance de chaque page avec PageRank

Lancer la commande

```
% python moteur3.py
```

qui appelle la fonction parcourir avec comme page de départ la page index.html, puis la fonction calculerPageRank qui calcule la PageRank de chaque page et la fonction afficherPageRank qui affiche celui-ci. La fonction calculerPageRank initialise le PageRank de chaque page à $1/n$ s'il y a n pages, puis appelle la fonction itererPageRank 10 fois pour que celui-ci se stabilise. Comme la fonction itererPageRank ne fait rien, le PageRank de chaque page est 0.

La fonction itererPageRank doit calculer un nouveau PageRank de chaque page à partir de son PageRank précédent. Pour cela elle doit initialiser le nouveau PageRank de chaque page à 0, puis répartir le PageRank précédent de chaque page de manière équitable entre les pages liées à cette page, en utilisant le graphe calculé précédemment. Compléter la fonction comme suit :

```
def itererPageRank():
    global pageRank

    # initialiser le nouveau PageRank à 0
    PR = {}
    for page in graphe:
        PR[page] = 0

    # pour chaque page du graphe
    for page in graphe:
        # calculer pr = PageRank de cette page divisé
        # par le nombre de pages auxquelles elle est liée
        pr = pageRank[page] / len(graphe[page])
        # pour chacune des pages liées
        for destination in graphe[page]:
            # ajouter pr au PageRank de cette page
            if destination not in PR:
                PR[destination] = 0
            PR[destination] += pr

    return PR
```

Lancer la commande

```
% python moteur4-sol.py
```

qui contient la fonction complétée et vérifier que le PageRank n'est plus nul. L'affichage du PageRank est fait à chacune des 10 répétitions, ce qui permet de voir son évolution et de vérifier qu'il se stabilise.

5 Répondre à une requête de recherche

Lancer la commande

```
% python moteur.py
```

qui contient la version finale du moteur d'indexation. Cette commande affiche l'index, le graphe et le PageRank de chaque page, et enregistre ces informations dans un fichier moteur.json.

Il reste à exploiter ce fichier dans le serveur Web créé dans la page du FabLab. Le fichier codeur.py contient une requête dynamique pour l'URL <http://localhost:8000/rechercher?search=motsclés>. Cette requête dynamique appelle la fonction rechercher dans le fichier codeur.py.

Lancer la commande

```
% python codeur.py
```

et entrer l'URL <http://localhost:8000/search.html> ou cliquer sur le lien. La page search.html affiche un formulaire similaire à celui de l'activité 2 (document 4, page 37). Comme la fonction rechercher du formulaire n'est pas complète, la réponse à la requête est toujours vide.

La fonction rechercher extrait les mots-clés de la requête et appelle la fonction trouverPages qui utilise l'index pour trouver les pages qui contiennent tous les mots-clés. Si aucune page ne répond à la requête, elle retourne une page avec le message correspondant. Sinon, elle doit retourner le code HTML de la liste des pages. Compléter le code de la fonction ainsi :

```
html += "<ol>"          # début de liste numérotée
for page in pages:      # pour chaque page du graphe
    # ajouter un élément de liste avec un lien vers la page et son PageRank
    html += '<li>(PageRank=%f) <a href="%s">%s</a></li>' % (pageRank[page], page, page)
html += '</ol>'          # fin de la liste
```

Arrêter le serveur en tapant Control-C et le relancer :

```
% python codeur-sol.py
```

Aller à nouveau sur la page <http://localhost:8000/search.html> dans le navigateur et faire une recherche, par exemple du mot clé « page ». La réponse doit lister chaque page résultat avec son PageRank.

Note : il est bien sûr possible de créer d'autres pages Web (à condition d'ajouter des liens depuis les pages existantes) et de relancer le moteur d'indexation pour créer une nouvelle version du fichier moteur.json. On peut alors relancer le serveur comme ci-dessus pour utiliser ce nouvel index.

Guide du professeur - Chapitre 4

L'objectif de ce chapitre est de montrer comment les données sont stockées et manipulées par les outils numériques.

Histoire, enjeux, débats

Ce chapitre permet d'aborder la modélisation des collections de données structurées, et de découvrir comment l'on stocke et consulte ces collections. Le chapitre s'intéresse particulièrement aux enjeux sociaux liés au stockage et traitement de grands volumes de données, et présente le stockage de données dans le Cloud.

- ❖ Activité 1 Des données non structurées aux collections
- ❖ Activité 2 Les principaux formats
- ❖ Activité 3 Où sont stockées les données ?
- ❖ Activité 4 Utiliser le cloud

Bande dessinée d'ouverture

Cette bande dessinée permet de poser la question de l'organisation des données. L'ajout d'étiquettes ou la répartition des fichiers (en l'occurrence musicaux) dans des répertoires ajoutent des métadonnées - informations externes sur les données - qui facilitent les recherches ultérieures.

Cette histoire illustre aussi la difficulté à organiser de façon pertinente les données.

Note : Les pages **Controverse** et **Monde de demain** ne sont pas corrigées. En effet, dans les premières, il est plus question d'ouvrir un débat avec les élèves, les questions n'étant là que pour initier une discussion, et pour les secondes, le but est de prendre connaissance des pistes de recherche actuelles sur ce sujet.

Les activités

ACTIVITÉ 1

Les compétences travaillées

- Cette activité apporte des connaissances techniques sur les données structurées, la modélisation, et les premiers algorithmes de recherche associés.
- Ces connaissances peuvent être mises en œuvre par le FabLab.

Les savoir-faire travaillés

- Modéliser des informations banales du quotidien par des données structurées.
- Vocabulaire : valeur, donnée, descripteur, donnée structurée, collection, etc.
- Identifier les différents descripteurs d'un objet. Distinguer la valeur d'un objet de son descripteur. Réaliser des opérations de filtre sur une table.

Exercices liés à l'activité :

Activité 2 (documents 1 et 4, question 1). Cours. Je vérifie : 2.a, 3, 4. FabLab.

Vue d'ensemble de l'activité :

Cette activité introduit, à partir d'exemples quotidiens de la vie réelle (YouTube), *les données structurées et leurs traitements* les plus simples, comme l'indique le titre du chapitre. Ces premiers traitements sont, dans cette activité, les premiers algorithmes de manipulations « internes » aux systèmes de gestion de bases de données (SGBD, qui sont les logiciels permettant de gérer et utiliser une base de données, comme SQLite dans le FabLab). Les manipulations « externes », par un utilisateur expert ou apprenant, faites depuis l'extérieur de la base, étant faites pour leur part avec SQL (comme dans le FabLab).

On part dans le document 1 d'un objet de la vie réelle, dans lequel on identifie des paquets d'informations similaires, avec des organisations identiques. Dans le document 2, on modélise cet objet par une collection structurée, et on introduit soigneusement le vocabulaire correspondant. Dans le document 3, on aborde la recherche (le filtrage) dans une telle collection d'objets réels, et comment un moteur de base de données peut la faire en interne. Les questions font pratiquer le vocabulaire, la modélisation, puis une variante de l'algorithme de filtrage.

Contexte scientifique de l'activité :

Il est détaillé et approfondi dans l'introduction de l'Activité 2.

Introduction :

Dans la mesure où le chapitre introduit directement les données structurées, on pourrait se demander au préalable ce que seraient des données *non structurées*, comme le suggère le titre de l'activité.

Imaginons un chapeau rempli de petits carrés de papier blanc, sur chacun étant écrit une *valeur*, par exemple *Stromae*, sur un autre carré *Papaoutai*, sur un autre encore *500*, etc. Bien sûr les carrés ne sont pas attachés entre eux et ne sont pas rangés dans le chapeau.

Si l'on cherche à obtenir des informations à partir de ces valeurs, c'est-à-dire à les *exploiter*, cela sera en général impossible : comment deviner à partir de rien les liens entre les valeurs *Stromae* et *Papaoutai* ? Ou encore entre les trois valeurs, chacun sur un carré différent, *François*, *Alexandra* et *Claude* ? Exercice : Dire quels objets de la vie réelle décrivent ces valeurs isolées dans le chapeau ; ou simplement quelles sont toutes les chansons de Claude François dans le chapeau ? Réponse : c'est impossible ; à part inventer, on ne peut rien dire, *on ne peut pas exploiter ces données*.

La structuration de ces valeurs dans le chapeau *de facto* isolées et indépendantes passe par deux concepts. D'une part, transformer chaque valeur en *donnée* en lui associant un *descripteur* (ex : nom : *Stromae* ; et sur un autre papier, prénom : *François* ; et encore, titre : *Papaoutai*, etc.) ; et d'autre part regrouper les couples descripteur-données qui décrivent un seul et même objet de la vie réelle. Un tel ensemble de couples « donnée-descripteur » est une donnée structurée. On reconnaît alors l'information présente dans le triplet (prénom : *Claude*, nom : *François*, titre : *Alexandra*) ; et, en particulier, que ces trois prénoms ne sont pas ceux d'une fratrie, mais ceux de la chanson d'un artiste.

On a structuré les données originelles isolées, on va pouvoir les exploiter. Par exemple les filtrer pour retrouver toutes les chansons de Claude François.

Cette structuration paraît très naturelle - et elle l'est. Toutefois d'autres structurations sont possibles, comme on le verra dans l'Activité 2.

Ce passage originel des données non structurées aux collections ne fait pas l'objet d'autre exercice dans l'activité ou le chapitre, mais la présente introduction pourrait faire fonction de démarrage du chapitre, avec un vrai chapeau et de vrais carrés sur lesquels sont écrites des valeurs, et en demandant aux élèves d'en extraire la réponse à la question ci-dessus (quelles sont toutes les chansons du chanteur Claude François). Dans cet état, ils ne trouveront pas (sauf à faire usage de connaissances personnelles extérieures au chapeau). Si on scotche physiquement ensemble les carrés, on ne peut toujours rien dire. Mais si on ajoute maintenant en plus les descripteurs, alors on peut obtenir une réponse parfaite.

(On pourra préférer un tel chanteur historique, plutôt qu'un plus connu par eux dont ils pourraient deviner par cœur les chansons.)

Question 1

Utilisation des documents 1 et 2

a. Chacun des trois objets de la table possède les mêmes quatre descripteurs : Artiste, Titre, Vues, Ancienneté.

Pour l'objet de la première ligne, la valeur pour le descripteur Artiste est Stromae, Papaoutai pour Titre, 500 M pour Vues, et 5 ans pour Ancienneté, et il correspond à la première suggestion YouTube du doc. 1.

Pour l'objet 2 : (Artiste : Stromae, Titre : Papaoutai Live, Vues : 4 M, Ancienneté : 4 ans), et il correspond au deuxième objet du doc. 1.

Pour l'objet 3 : (Artiste : Stromae, Titre : Tous les mêmes, Vues: 220 M, Ancienneté : 6 ans), et il correspond au troisième objet du doc. 1.

b.

Artiste	Titre	Vues	Ancienneté
Stromae	Papaoutai	500 M	5 ans
Stromae	Papaoutai Live	4 M	4 ans
Stromae	Tous les mêmes	220 M	6 ans
Stromae	Racine carrée Live	15 M	3 ans

c. *Clip,Lyrics* pour le descripteur : Nature de la vidéo ; *Racine carrée* pour le descripteur Titre de l'album ; etc.

On voit d'habitude la valeur de la durée (sur la vidéo elle-même), et qui l'a mise en ligne. Mettre une vidéo en ligne sur YouTube est très simple, mais attention à ce que vous mettez, et paramétrez bien la confidentialité.

Question 2

Utilisation des documents1 et 2

Jour	Créneau	Salle	Matière	Professeur
Lundi	8 :30-10 :00	45	Mathématiques	Mme Plessis C.
Lundi	10 :00-12 :00	Gymnase	Ed. Physique et Sport	M. Boulanger H.
Lundi	14 :30-15 :30	44	Mathématiques	Mme Plessis C.
Mardi	8 :30-9 :30	42	Espagnol LV2	M. Valencia H.
Mardi	9 :30-10 :30	44	Langues & culture Antiquité	M. César J.

S'il y a certains créneaux une semaine sur deux, on ajoutera une colonne avec le nom de la semaine (valeurs A ou B par exemple) ; de même si des cours ont lieu en demi-classe, on ajoutera une colonne groupe (avec les valeurs classe entière, groupe 1, groupe 2).

Table extraite contenant uniquement les jours, créneaux et salles des cours de mathématiques.

Jour	Créneau	Salle
Lundi	8 :30-10 :00	45
Lundi	14 :30-15 :30	44

Question 3

Utilisation des documents3 et 4

L'algorithme suivant parcourt une collection de clips et affiche la sous-collection de ceux dont l'artiste est celui donné en entrée (ici dans la variable n pour « nom »). Comme dans l'exemple du doc. 4, il manipule une table et ses lignes, et crée une nouvelle table pour le résultat.

Variables : c, résultat sont des tables avec les colonnes nom, titre, vues

l est une ligne de table avec nom, titre, vues

n est un mot

Entrée : saisir n

Initialisation : c prend comme valeur la table de tous les clips de YouTube

Traitement : Pour chaque ligne l de c faire

```
| Si l.nom = n  
|   | alors ajouter l à résultat  
| Fin Si
```

Fin Pour

Sortie : Pour chaque ligne l de résultat faire

```
| afficher l.nom, l.titre, l.vues
```

Fin Pour

ACTIVITÉ 2

Les compétences travaillées :

Cette activité apporte des connaissances techniques pour enregistrer les données dans des collections, et montre qu'une même information peut être représentée de différentes manières:

- Pour stocker les collections de façon « standardisée » dans des fichiers, on a le choix entre au moins deux formats: CSV et JSON
- Pour faciliter l'exploitation et l'interprétation des données, l'ordinateur et/ou l'utilisateur leur attachent des métadonnées qui enregistrent en général des informations de « contexte ».
- Différentes collections peuvent être combinées pour « croiser » leurs informations, ce qui amène parfois à considérer des ensembles de collections plutôt que chaque collection séparément.

Ces connaissances peuvent être mises en oeuvre :

- Par une brève séance sur machine : ouverture de fichiers, consultation des métadonnées.
- Par la page du codeur, en ce qui concerne la manipulation d'un fichier csv correspondant à une seule collection. On peut facilement étendre le travail en ajoutant d'autres collections pour illustrer l'opération consistant à croiser des collections.

Les savoir-faire travaillés :

- Reconnaître les principaux types de fichiers.
- Consulter le contenu de fichiers CSV (identifier le séparateur, en-tête éventuel, etc.).
- Consulter les métadonnées d'un fichier.
- Prendre conscience que représenter des informations du monde réel implique un travail de modélisation, et peut amener à organiser les données en plusieurs collections.
- Comparer des données structurées et des données qui ne sont pas organisées sous forme de table.

Vue d'ensemble de l'activité

Cette activité complète l'introduction au modèle relationnel de l'activité 1. Deux répartitions possibles des données sont présentées par les documents 1 et 4, et la question 1 a pour but de réfléchir aux avantages et inconvénients des deux organisations.

Une fois les données réparties en collections, chaque collection sera enregistrée dans un fichier. Les documents 2 et 3 présentent donc deux des principaux formats de fichiers utilisés pour stocker des données structurées.

Contexte scientifique de l'activité :

Plus généralement, la gestion de données sous forme de tables a été formalisée par le modèle relationnel, mentionné dans les repères historiques, et pour lequel E.F. Codd a obtenu l'équivalent du prix Nobel en informatique ; le prix Turing. La question 1 introduit deux aspects clés du modèle relationnel :

- comment répartir les données dans des tables selon les règles de l'art.

Les informaticiens appellent « *normalisation* » le procédé consistant à organiser une base de données de façon à éviter les redondances. Les redondances à éviter ont été codifiées par ce qu'on appelle les « *formes normales* ». La répartition en 3 tables suggérée en question 1b pour éliminer toute forme de redondance est une version « normalisée » de la base de donnée.

- comment interroger une base de données.

Le modèle relationnel définit pour cela un ensemble d'opérations qui forment ce qu'on appelle l'*algèbre relationnelle*. Dans l'activité 1 doc 4, on a vu une opération qui consiste à sélectionner les lignes d'une table satisfaisant certaines conditions. Cette opération de filtrage est appelée « *sélection* » dans le modèle relationnel. Le croisement de collections introduit pour le doc 4 (question 1.c) est une autre opération du modèle relationnel appellée « *jointure* ». Ici l'on a

présenté un cas très simple de jointure (seulement deux tables, et la jointure se fait sur un descripteur commun aux deux tables, qui en plus est un identifiant dans une des tables).

Il n'y a pas vraiment de concept clé en ce qui concerne les métadonnées ou les formats de fichiers: l'activité vise avant tout à familiariser les élèves par la pratique.

Les métadonnées peuvent être très variées (il y a en fait une grande variété de standards souvent spécifiques à des domaines particuliers : photos, audio, etc.). En fait la distinction entre données et métadonnées n'est même pas claire dans certains cas.

Les métadonnées peuvent être ajoutées soit par l'utilisateur soit automatiquement par les outils.

Des métadonnées sont souvent attachées aux collections, en particulier pour expliquer la signification des descripteurs (c'est extrêmement important pour que les données soient utilisables : un exemple ; la définition de « segment drg » dans l'onglet « information » pour la page SNCF sus-mentionnée).

Les métadonnées automatiquement ajoutées à tous les fichiers de l'ordinateur sont utiles pour effectuer des recherches : on peut par exemple les utiliser pour afficher les fichiers par ordre de modification, rechercher les plus gros fichiers, etc.

Voir aussi : format EXIF de métadonnées pour les photos numériques p151 (ou aussi p107).

Les deux formats, CSV et JSON, sont des formats « texte », c'est-à-dire qu'ils sont lisibles par l'humain dans un éditeur de texte. À la différence du format pdf, par exemple, dont le code source est à peu près illisible pour un humain : il faut utiliser un programme dédié à l'affichage de pdf pour pouvoir afficher un tel document.

Le format CSV est très simple et surtout utilisé pour stocker des collections simples. Un des soucis est qu'il n'est pas vraiment standardisé et il faut donc les manipuler avec prudence pour identifier les en-tête, séparateurs, etc.

Le format JSON est aussi adapté pour stocker des données plus riches que de simples tables :

Les valeurs associées aux descripteurs pouvant elles-mêmes être un objet JSON, un tableau de valeur, etc. Le format est donc souvent utilisé sur le web pour échanger des données entre applications.

Exercices liés à l'activité:

Je vérifie : 1,2,3,4 et 5. Je m'entraîne : 8 et 9. FabLab. Page du codeur.

Question 1

Utilisation des documents 1 à 4

a. Lorsque de nombreuses gares ont le même code postal, l'organisation en 2 tables du doc. 4 évite de répéter à chaque fois le code postal et la région. Elle est donc généralement considérée comme une meilleure organisation.

Lorsque les gares ont toutes des codes postaux distincts, alors l'organisation du document 1 est préférable.

b. Table1 : (Intitulé gare, code postal)

Table2 : (code postal, département)

Table3 : (Département, région)

c. La première requête sera *a priori* plus efficace avec l'organisation du doc. 1, car avec le doc. 4 il faut croiser les tables, ce qui demande plus de travail. La seconde requête sera plus efficace sur le doc. 4 parce qu'il suffira de consulter que la table des départements/régions, qui comporte moins de lignes et moins de descripteurs, donc on perdra moins de temps à lire des informations qui ne servent à rien pour cette requête.

Remarques complémentaires

Il est bon de nuancer notre comparaison entre les organisations des doc 1 et 4 en fonction de l'usage qui va être fait des données. Dans la réalité, on utilise parfois quand même la représentation du doc 1 au lieu de celle du doc 4, parce qu'il est plus simple de n'avoir qu'une seule collection à gérer. C'est probablement pour cette raison que le fichier récupéré sur le site SNCF ne contient qu'une seule collection : le site a préféré organiser les données comme dans le doc. 1 plutôt que de préparer une archive contenant 2 fichiers (voire plus).

Inversement, il arrive parfois (mais c'est beaucoup plus rare) que l'on rencontre l'organisation en plusieurs tables du document 4, même lorsque les codes postaux sont uniques : cela arrive par exemple lorsque les collections que l'on veut joindre sont gérées par des organismes (ex : des services web) différents, ou lorsqu'une grande majorité des utilisateurs ne s'intéresse pas aux départements et aux régions (cf. question c).

Question 2

Utilisation des documents 1 et 2

Exercice à faire sur ordinateur.

Il est ici plus simple d'utiliser le point-virgule comme séparateur de colonne : on ne risque pas de confondre une séparation de colonne et une donnée car aucune valeur de ce jeu de donnée ne comporte de point-virgule, alors que certaines colonnes ont des valeurs contenant des virgules (par exemple les coordonnées géographiques WGS84).

Question 3

Utilisation du document 1

Exercice à faire sur ordinateur.

Les instructions permettant d'afficher le code-source dépendent bien entendu de l'interface. Sous Thunderbird : Ctrl+u ou Affichage → code source du message.

Sur le site Gmail : « afficher l'original ».

On observe que l'en-tête du message est une liste de paires descripteur : valeur.

La format de cet en-tête est principalement spécifié par le standard RFC 5322.

Quelques descripteurs courants : Delivered-To, Received, etc.

Pour une page Web, le format sera probablement HTML. C'est un langage à balise. Une manière de représenter une paire (descripteur, valeur) en HTML est d'entourer la valeur par deux balises : une ouvrante et une fermante. La balise est le descripteur.

Par exemple, pour la page view-source:<https://fr.wikipedia.org/wiki/France>

Le titre a pour valeur : France — Wikipédia

On a aussi d'autres descripteurs css : class, id, etc. qui permettent de contrôler la manière dont s'affiche la page.

Cf chapitre sur le web.

Remarque : si l'on peut encore parler de descripteur et valeur pour HTML, on ne considère pas les pages webs comme des données structurées : comme vu dans le chapitre sur le Web, une page web se modélise naturellement par un arbre et non par une table. On dit parfois que les pages web sont des données *semi-structurées* : on a des éléments de structures, mais les éléments de cette structure arborescente ne partagent pas nécessairement les mêmes descripteurs.

Question 4

Exercice à faire sur ordinateur.

On peut consulter les métadonnées stockées par le système de fichier.

Sous Ubuntu : Clic droit → « Lire les informations » (Get Info)

Ou en ligne de commande : stat chemin_du_fichier

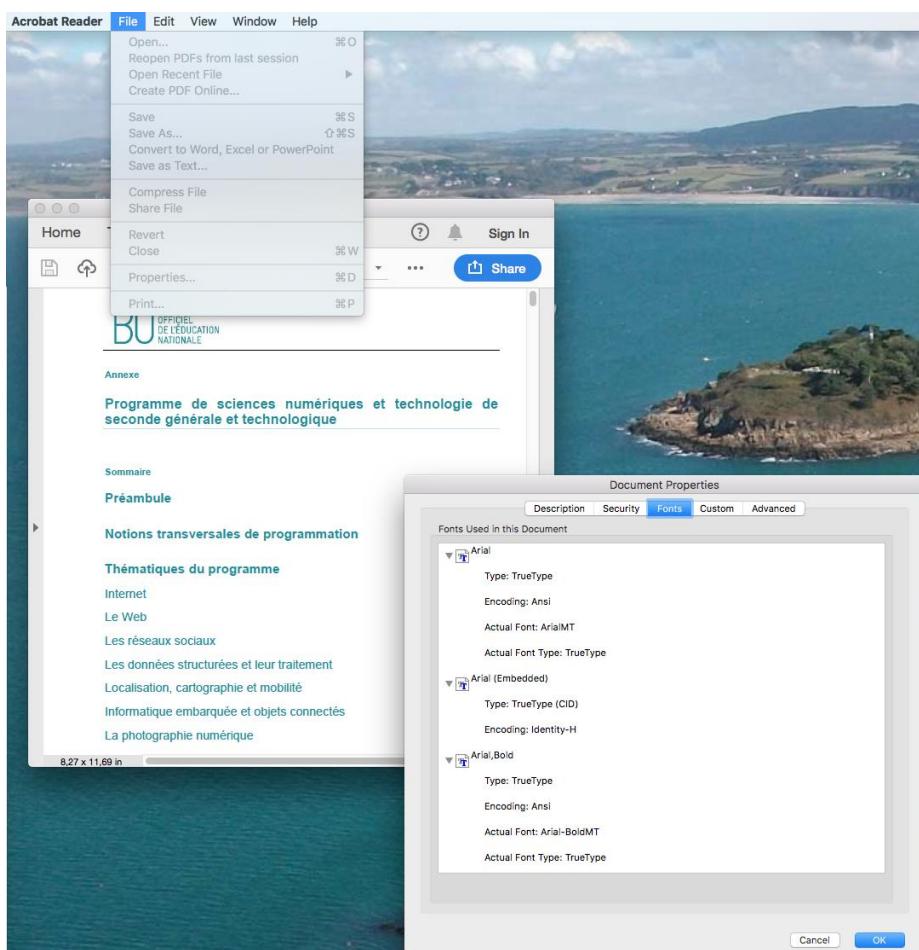
Sous Mac : Clic droit → « Lire les informations » (Get Info)

Ou en ligne de commande : stat -x chemin_du_fichier

Sous Windows : Clic droit → « Propriétés »

On notera que quel que soit le type de fichier, on va avoir des métadonnées similaires à celles du doc. 2 : type de fichier, taille du fichier... Mais on pourra aussi avoir des informations spécifiques pour certains types de fichiers, par exemple pour un fichier pdf : « version du format pdf, pour une image jpeg : dimension, résolution, etc.

Pour certains types de fichiers on obtiendra plus de métadonnées en ouvrant le fichier avec un logiciel dédié (qui ne se contentera pas des métadonnées stockées par le système de fichier mais y ajoutera des informations obtenues par analyse du fichier). Par exemple, pour un pdf ouvert sous Acrobat: avec « Fichiers » → « Propriétés » on peut entre autres observer la liste des polices utilisées dans le document.



ACTIVITÉ 3

Les compétences travaillées

Cette activité cherche à donner un aperçu global de nos activités numériques, en présentant les supports matériels sur lesquels sont stockées les données. L'objectif est de prendre conscience en tant que « citoyen du monde numérique » des infrastructures permettant ces activités. Il s'agit en particulier d'appréhender le coût écologique de nos activités numériques.

La mise en œuvre en pratique du transfert et partage de fichiers sur le Cloud est détaillée dans l'activité suivante. Pour cette activité, les vidéos suggérées donnent un panorama assez complet des usages du cloud pour les particuliers, et des techniques mises en œuvre dans les data centers pour permettre ce stockage.

Les savoir-faire travaillés

- Appréhender les ordres de grandeurs traités lorsque l'on parle de Big Data.
- Comprendre ce qu'est le cloud. Identifier ses avantages et inconvénients.
- Prendre conscience du coût écologique du numérique.

Approches complémentaires qui ne sont pas proposées par le programme:

Il peut être intéressant :

- de compléter par quelques aspects économiques concernant l'économie des données. On peut par exemple donner des idées des montants investis dans les infrastructures, des emplois ou capitalisations des géants du numérique, mentionner que parmi les plus grandes capitalisations boursières on trouve Oracle et IBM (qui se sont initialement développées en vendant respectivement des machines et logiciels pour gérer les données) ;
- de comparer différents supports de stockage en terme de capacité et de performance : la frise en début de chapitre mentionne carte perforées et disques durs, mais il existe d'autres supports de stockage : il est important de noter que les disques magnétiques sont progressivement remplacés par des mémoires flash, non seulement dans les smartphones mais aussi sur les ordinateurs. Des mémoires basées sur des semi-conducteurs (SSD) donc ;
- de mettre en regard la capacité de stockage de différents appareils en fonction du type de fichier (ex : image au format « sans perte », photos compressées, films HD, etc.

Exercices liés à l'activité:

Je vérifie : 2. Je m'entraîne: 6,10 et 11.

Question 1

Utilisation des documents 1, 2 et 3

Définition du cloud : voir doc. 2.

Avantages et désavantages : voir aussi doc.1 de l'activité 4.

Avantages potentiels (pour du stockage mais c'est globalement vrai aussi pour d'autres types de services cloud : calcul, logiciel, etc.) :

- les données sont facilement accessibles à travers internet ;
 - on n'a pas à gérer la maintenance (au moins des machines, et généralement des logiciels).
- Cela peut se traduire par un gain de temps, des économies (par rapport à l'achat d'une machine

avec les logiciels, et par rapport à d'éventuels achats de services de réparations matérielles ou logicielles) ;

- on bénéficie de l'expertise de l'entreprise du cloud ;
- souvent très sécurisé contre pannes et piratages ;
- possibilité d'économie d'échelle à travers le partage des ressources ;
- facilité de passer à l'échelle : on peut facilement augmenter son espace de stockage, sans véritable limite du moment que l'on paye en conséquence.

Désavantages potentiels :

- nécessité d'être connecté pour accéder aux données (en général, sauf si on a conservé une copie locale « hors ligne ») ;
- cela génère beaucoup de trafic Internet ;
- les données inutilisées coûtent plus d'énergie vu qu'elles restent constamment ;
- le risque de piratage augmente lorsque l'on permet l'accès aux données depuis Internet ;
- risque éventuel d'espionnage, en particulier lorsque les données ne sont pas cryptées (lors du transfert des données vers le cloud, ou même de façon légale pour les données à l'intérieur du cloud, certains pays autorisant des agences gouvernementales à demander l'accès aux données).

Question 2

Utilisation du document 3

Pour un data center ayant un taux de disponibilité de 99.6% :

$365 \times 24 \times 60 \times (1 - 0,996) =$ à peu près 2100minutes par an, un peu moins de 6 minutes par jour.
Avec un taux de 99,995 %: 26 minutes par an.

Pour garantir une telle disponibilité, il faut de la redondance : on copie les données pour qu'un ordinateur puisse en remplacer un autre qui serait indisponible (panne ou maintenance), on double les circuits d'alimentations électrique, et on installe des générateurs pour s'assurer qu'on puisse résister à une panne de courant.

Question 3

Utilisation du document 4

Raison 1 : ce sont des énergies intermittentes. Les data centers sont alimentés en permanence.

Raison 2 : les data centers ne sont qu'une petite partie du numérique.

Raison 3 : le plus gros coût écologique du numérique est sans doute la construction des appareils électroniques (ordinateurs, smartphones, etc.).

Question 4

Utilisation du document 4

Limiter la consommation électrique, ne pas changer trop souvent les appareils, choisir la sobriété (ne pas prendre d'appareils inutiles ne correspondant pas à nos besoins).

On peut se reporter aux recommandations de l'ADEME.

ACTIVITE 4

Les compétences travaillées

Ce chapitre a pour but de présenter des solutions pour le stockage de données personnelles dans le cloud. Il s'agit de comparer les services cloud proposés, en attirant l'attention sur la protection des données, et aussi de se familiariser avec un outil cloud pour partager des dossiers.

Les savoir-faire travaillés

- Appréhender la diversité des services cloud.
- Comprendre les enjeux de protection des données lorsque l'on utilise des services logiciels, et a fortiori les services cloud.
- Et surtout : transférer, synchroniser, et partager des documents sur le cloud.

Contexte de l'activité :

Le programme semble plutôt considérer le cloud du point de vue du stockage et partage de données personnelles. Le doc 1 vise à faire prendre conscience que le Cloud ce n'est pas que le stockage de données personnelles : une grande part du marché du cloud c'est des services aux entreprises, soit pour déployer des applications (par ex : nécessitant une forte disponibilité comme le commerce en ligne), pour louer des capacités de calcul importantes, pour utiliser des services proposés par certaines compagnies au sein de plateformes logicielles (ex : outils d'analyse, intelligence artificielle etc), etc. Il y a donc un marché du cloud à destination des entreprises (B2B) qui se chiffre au minimum en centaines de milliards d'euros : Amazon, par exemple, réalise la majorité de ses profits dans les services Cloud !

Le document complète aussi l'activité précédente par rapport au stockage de donnée personnelle.

Précision concernant le doc 1 :

Dans le tableau du doc 1, le nombre de Gigaoctets indiqués correspondent à la capacité maximale de stockage (en Go) mise à disposition gratuitement, au moment de la rédaction du manuel. Cette capacité peut évidemment être augmentée en choisissant une version payante du service. De plus, par manque d'espace, nous avons été obligés de limiter le nombre de services comparés et surtout les caractéristiques mises en avant. Un des critères de performance importants outre la capacité de stockage est la fiabilité et la vitesse (et coûts pour les versions payantes) de transfert (upload/download). On trouve plusieurs comparatifs en ligne, mais il faut se référer à un comparatif récent car les offres évoluent.

Indications concernant le choix d'un service cloud pour l'illustration et la mise en pratique :

Nous avons eu à cœur de mettre en avant des acteurs locaux (français) du cloud, qui à défaut d'avoir la célébrité des géants du web, proposent des services tout à fait comparables dans leur domaine de compétence. Pour les services aux entreprises, nous pourrions encore mentionner outre OVH : Online (Iliad), Orange Business Services, et Outscale (Dassault). Pour les particuliers, à part Cozy (qui en fait est hébergé par défaut dans des datacenters d'OVH), on peut mentionner Ikoula (qui a ses propres datacenters) pour des solutions payantes, et les FAI (Orange, SFR, etc) qui proposent souvent un stockage cloud gratuit. À l'international il existe une pléthore d'offres cloud dont beaucoup ont une version freemium.

Pour le doc 2 (et en question 2), nous avons néanmoins choisi d'illustrer le Cloud avec Google car il s'agit probablement du service avec la plus large communauté d'utilisateurs, et est donc le plus à même d'être déjà connu des élèves. De plus, cette plateforme dispose d'outils performants pour l'édition collaborative de documents (texte, feuille de calculs, etc.), et est déjà installé sur les smartphones Android. Néanmoins des opérations similaires peuvent tout à fait être réalisées avec d'autres applications moins « propriétaires », plus open source, plus françaises, etc.

Quelques exemples parmi d'autres : Cozy Drive pour la synchronisation et le partage de fichiers, Framapad (et autres services listés sur Framasoft) pour l'édition collaborative, etc. Il faudra probablement modifier un peu l'énoncé des questions en ce cas, car les options de synchronisation proposées varient selon les services cloud.

Question 1

Utilisation du document 1

Cette question « naturelle » soulève en fait une petite difficulté: « Cloud » est plus un buzzword marketing qu'un concept scientifique défini précisément.

La définition que l'on cite généralement est celle du NIST :

<https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf>

Peut-être l'élève loue-t-il une machine chez OVH pour héberger son propre site?

Il y a une petite subtilité ici. Si l'élève réserve une seule machine physique chez l'hébergeur alors on pourrait considérer que ce n'est pas du Cloud mais plus de l'hébergement Web.

Mais en pratique on réserve désormais plutôt une machine virtuelle avec certaines capacités de calcul ou stockage ou même juste un « service d'hébergement » (IaaS, PaaS ou SaaS, etc.) en laissant à OVH le soin de déployer cette machine virtuelle sur ses clusters, auquel cas on s'accordera à considérer qu'il s'agit de cloud.

Ou même (le cas est probablement rare) certains auront développé leur propre application Web sur Google Engine ou Microsoft Azure (PaaS) ?

Peut-être l'élève a-t-il utilisé un des outils de stockage cloud tels que Google Drive, Dropbox, pCloud, ou une application de Google Apps ou de framasoft (formulaire, tableur, traitement de texte, etc.) (SaaS) ?

En tout cas, même un élève qui n'a pas utilisé de stockage cloud (directement) a probablement utilisé des applications cloud, par exemple en consultant ses courriels depuis une interface Web : Gmail, yahoo mail, hotmail, etc. sont des services cloud puisque les messages sont stockés sur un serveur distant accessible en ligne.

Youtube, Deezer (qui a ses propres serveurs), Spotify (hébergé par Google), Netflix (hébergé par Amazon) sont des services cloud (SaaS).

La différence entre PaaS, IaaS, SaaS (hors programme) est parfois confuse selon l'usage fait du service et n'est pas cruciale même si on la trouve dans le document du NIST ci-dessus. Il n'est sans doute pas utile que les élèves retiennent ou comprennent ces distinctions. Je l'indique ici uniquement pour illustrer le concept, et montrer la diversité des services proposés sur le cloud. Il n'est pas non plus nécessaire de distinguer l'hébergement Web du cloud.

Question 2

Utilisation du document 2

Question à faire sur ordinateur par groupe d'élèves.

Question 3

Utilisation du document 2

Question à faire sur ordinateur par groupe d'élèves.

Question 4

Utilisation du document 3

Clairement, les entreprises ne sont pas égales : beaucoup se permettent d'utiliser les données pour du ciblage publicitaire. De plus les entreprises américaines sont soumises au Cloud Actaméricain leur imposant de fournir aux autorités américaines l'accès aux informations stockées sur leurs serveurs si celles-ci leur en font la demande.

Je vérifie mes connaissances

Solutions détaillées

1. QCM

1. a. et b.

2. a. et b.

3. a.,b. etc.

Remarque complémentaire :

Techniquement 2c n'est pas faux : tout document texte peut être vu comme un cas particulier (et assez inutile) de csv à une seule colonne, mais il est peu vraisemblable que ce soit un csv ici alors que 2.a et 2b, eux, sont vraisemblablement des fichiers csv. On pourrait envisager de changer la question de l'énoncé en « lesquels de ces fichiers sont probablement des fichiers csv » mais alors un esprit pinailleur pourrait discuter le sens de « probablement ».

2. Vrai/Faux

a. **FAUX.** Au contraire, on parle de données structurées lorsqu'elles forment une collection et partagent les mêmes descripteurs.

b. **FAUX.** Les descripteurs se trouvent au début du fichier.

c. **VRAI.**

d. **VRAI.**

e. **VRAI.**

f. **FAUX.** La consommation en ressources énergétiques se situe plutôt autour de 10 %.

Remarques complémentaires :

Il est toujours difficile d'avoir des chiffres fiables en matière d'énergie sur l'ensemble du cycle de vie. Mais les 10% sont une surestimation très pessimiste ; la vraie consommation était proche de 3% en 2017 et les projections prévoient qu'elle reste entre 3 et 6% jusque 2025:

https://theshiftproject.org/wp-content/uploads/2019/03/Lean-ICT-Report_The-Shift-Project_2019.pdf(p.16)

3. Associer des termes

1 Données

2 Descripteurs

3 Collection

4. Vers une définition

a.Une **collection** regroupe des objets partageant les mêmes descripteurs et peut être représentée sous la forme de **table**.

b.Une **base de données** est un ensemble de collections de données reliées entre elles.

c.Les **métadonnées** fournissent des informations sur le contenu du fichier.

5. Commenter un schéma

Sur la machine

Descripteur : dernière visite

Valeur : 10/01/2019, 16:14

Métadonnée : fichier sauvegardant l'historique de navigation enregistré (ou consulté) le 22/01/19

Pour l'utilisateur

Information : « à la date du 22/01/19, la dernière fois que l'utilisateur a visité l'url correspondante était le 10/01 à 16 h 14. »

JE M'ENTRAÎNE

6. Le cloud et l'accès à Internet

1 Po = 8×10^{15} bits. Donc $8 \times 10^{15} / (100 \times 10^6) = 8 \times 10^7$ s.

Donc il faut $8 \times 10^7 / (3600 \times 24) \approx 926$ jours par Pétaoctet, même avec une connexion très haut débit.
Une solution plus efficace est de remplir un camion avec les disques durset envoyer le camion déposer les données dans le datacenter.

Remarque complémentaire : cette solution est en fait réellement utilisée en pratique. Amazon propose effectivement un (voire deux) service de ce genre : Amazon snowmobile, capable de charger 100 Po (et Amazon snowball pour des volumes de données plus faibles)
<https://aws.amazon.com/fr/snowmobile/>

L'idée est effectivement que l'entreprise charge les données sur des machines spéciales (que l'on peut voir comme un gros disque dur) qui sont ensuite transportées par camion jusqu'aux datacenters d'Amazon.

7. La publication des données

Les données ouvertes, même si en l'occurrence ce procès aux USA concernait le FOIA de 1966.

8. Les données structurées

On observe le nombre d'avis, la meilleure et la pire évaluation, ainsi que le classement moyen. Google incorpore le nombre d'avis et le classement dans son résultat. Si l'étudiant est curieux et essaie aussi un autre moteur de recherche, il peut voir que d'autres moteurs de recherche (Yahoo, Qwant, DuckDuckgo) ne le font pas (Bing le fait mais n'indique que 2 avis au lieu de 3 (il n'a pas mis à jour suffisamment récemment ses informations).

9. Des opérations sur des tables

Appelons T_ALBUM la table p92 et T_NAISSANCE celle p95.

a. Il faut croiser T_ALBUM et T_NAISSANCE de sorte que T_ALBUM.artiste = T_NAISSANCE.nom (on complète les données de T_ALBUM avec la date et le lieu de naissance des artistes). Puis on ne conserve que les lignes pour lesquelles le pays est « Belgique » et la date de naissance supérieure à 1990.

Enfin on ne garde que les colonnes « T_ALBUM.titre » et « T_ALBUM.ancienneté ».
Donc on a 3 opérations : croisement des 2 tables (= jointure si on veut utiliser le terme technique), filtrer les lignes (sélection) et élimination des colonnes superflues (projection).

Remarques complémentaires

- Il n'est pas attendu de l'élève qu'il connaisse le nom des opérateurs de l'algèbre relationnelle : termes jointure, sélection, projection.
Par contre on s'attend à ce qu'il identifie la nécessité de croiser les tables, et de filtrer les éléments qui nous intéressent.

- on dispose d'une certaine liberté sur l'ordre de ces opérations. On peut tout à fait faire la sélection sur T_NAISSANCE avant de la joindre avec T_ALBUM. Par contre la projection est à faire en dernier : si on commençait par éliminer la colonne T_ALBUM.artiste et T_ALBUM.vues, comment ferait-on pour joindre avec T_NAISSANCE?
- Lorsque l'on veut réellement implémenter ces opérations dans un programme, on a aussi le choix entre plusieurs algorithmes. Par exemple, pour joindre les deux tables, on peut utiliser des boucles « For » (= « Pour tout ») ; pour chaque ligne de T_ALBUM on va parcourir toutes les lignes de T_NAISSANCE pour trouver celle qui correspond (il pourrait en théorie il y en avoir plusieurs pour des opérations de jointures compliquées mais ici il n'y en a qu'une : un artiste n'a qu'une date de naissance). On pourrait aussi envisager de commencer par trier la table T_NAISSANCE par nom d'artiste pour faciliter les recherches. Certains logiciels dédiés aux bases de données sont conçus pour choisir la meilleure façon d'effectuer ce genre de calcul. Mais cela nous emmène bien loin de la question de départ.

10. Des quantités de données

On indique dans Activité 3, doc1 qu'il y a des dizaines de zettaoctets de données numériques, comptons 10 Zо pour simplifier : le nombre de tablettes est $8*10*10^{21}/(256*10^9)=31*10^{12}$. Ça fait à peu près 1 800 000 kilomètres : presque 5 fois la distance à la Lune et 47 fois la circonférence terrestre.

Il peut être bon de rappeler la distance terre-lune : 384 400 km, et, si l'on veut, la circonférence terrestre : 40 075 km. De toute façon, les chiffres exacts importent peu ; ce qui est intéressant c'est de comprendre que la quantité de donnée est ... astronomique.

Lorsque la comparaison a été mentionnée pour la première fois en 2014, l'on parlait d'une pile d'iPad qui faisait les 2/3 de la distance à la lune.

11. La consommation énergétique

a. Le document indique 416 TWH pour les data center. On peut deviner que la part de Google est de quelques TWH. Attention, divisé par les heures d'une année, cela fait juste quelques Gigawatts (3 en réalité mais ça on ne peut pas le deviner).

b. Cela fait à peu près 1 million d'ordinateurs.

Les 10 % d'énergie utilisée à refroidir (même si c'était sans doute un peu plus en 2010) ne changent pas l'ordre de grandeur.

12. Le choix d'un cloud

- Prix pour la quantité de données et l'usage prévu.
- Fiabilité et performance du service.
- Services proposés par l'hébergeur et compatibilité avec le terminal de l'utilisateur.
- Conditions de sécurité : chiffrement des données, possibilité par l'hébergeur de consulter et utiliser les données.
- Localisation et engagements écologiques de l'hébergeur.

13. Les données personnelles

a. À partir des informations restantes, on a de fortes chances de pouvoir identifier la personne à laquelle elles font référence, donc ces données restent personnelles mais si la personne n'est pas explicitement désignée.

b. Ces métadonnées peuvent être des informations personnelles si on sait établir un lien avec la personne à l'origine des photos. Par exemple les données de géolocalisation de toutes les photos prises par l'utilisateur d'un téléphone portable sont des données personnelles.

Remarques complémentaires

Il n'est pas toujours facile de savoir si une donnée est personnelle à partir de la définition légale (art. 4 du RGPD). On pourra se référer à :

<https://www.cnil.fr/fr/cnil-direct/question/une-donnee-caractere-personnel-cest-quoi>

ou à la page wikipedia qui est assez riche :

https://fr.wikipedia.org/wiki/Donn%C3%A9es_personnelles

FAB LAB

Dans l'Activité 1, on a introduit et travaillé la modélisation des données réelles en une table relationnelle, et vu la structure d'un algorithme simple de filtrage. Dans l'Activité 2, on a poursuivi l'étude de la modélisation, et aussi introduit deux formats de stockage logique (organisation) des tables relationnelles dans des fichiers. Puis les 3 et 4 ont abordé l'aspect du stockage physique (disques, etc.).

Enfin la page du codeur illustre comment, depuis un langage de programmation généraliste (ici Python), on peut accéder à, et manipuler, des tables relationnelles, au moyen d'une bibliothèque du langage.

Dans le FabLab, on pratique la situation « reine », la manipulation des tables relationnelles dans une vraie base de données relationnelle d'ampleur industrielle, avec le langage SQL, standard mondial depuis plusieurs décennies, conçu et parfaitement calibré pour cette situation (et bien d'autres). Très détaillée, la fiche SQLite à télécharger (SNT2039) présente très pragmatiquement SQL, SQLite, et leur application concrète dans le FabLab.

Le cahier des charges de l'application considérée dans le FabLab est un extrait de taille très modeste, mais parfaitement représentatif du cœur de quantité d'applications réelles qui tournent aujourd'hui sur la planète et gèrent tous sortes de données.

Une fois SQLite installé à l'avance, et après avoir fini les Activités 1 et 2, on peut sans doute compter une heure pour faire le FabLab avec des élèves ayant une certaine familiarité avec la programmation (Python ou autre), se répartissant par exemple en environ 20 à 30 minutes de cours sur SQL (incluant une démonstration) en suivant simplement la progression de la fiche SQLite (SN2039 à télécharger) qui est aussi conçue pour cela ; et environ 30 à 40 minutes pour les élèves accompagnés pour finir le FabLab.

L'installation proprement dite de SQLite, comme pour tout logiciel à destination des programmeurs serait plutôt destinée à des élèves très motivés du club informatique et programmation par exemple.

Le choix pour le FabLab de SQLite se base sur sa popularité, et sa dimension « légère » (*light* en anglais, qui a donné « lite », pour « moteur SQLite »), facile à obtenir et installer, et nécessitant très peu de ressource machine, tout en offrant une robustesse industrielle.

Question 1

Il n'y a pas de question explicite ici, mais il est pertinent de se donner un exemple représentatif des différentes lignes et valeurs possibles dans la table (on ignore à ce stade le type des valeurs, qui fait partie de la nature des valeurs, mais pas vraiment de leur organisation en collection). C'est-à-dire qu'on modélise logiquement les données du cahier des charges avant de mettre en œuvre la modélisation choisie dans un système/logiciel particulier (ici SQLite). Pour simplifier un jour est modélisé par un entier entre 1 et 365. Le cahier des charges fait l'hypothèse que la ville de départ est celle du lycée.

On va nommer cette table de manière assez naturelle « Billet », car la structure d'une ligne (descripteurs) est conçue pour être celle d'un billet.

Le cahier des charges à ce stade ne parle pas de quels sont les billets, mais ici on en met quand même deux pour se convaincre que la structure de la table (colonnes et descripteurs) est correcte. Noter que pour éviter d'éventuelles difficultés sur machine, on utilise les valeurs *payé* (billet payé) et *effectué* (voyage effectué) sans accent.

Prénom	Destination	Transport	Jour	Prix	Statut
Rita	New York	avion	365	450	paye
Riton	Amsterdam	train	210	59	effectue

Question 2

Comme rappelé dans les paragraphes ci-dessus, SQL et SQLite sont présentés très rigoureusement et soigneusement dans la fiche SQLite à télécharger (SN2039). L'enseignant peut souhaiter s'assurer à 'avance que la commande `sqlite3` tapée dans une console fonctionne bien sur chaque machine de la salle prévue pour le Fab Lab.

A nouveau pas de question à proprement parler ici, ce paragraphe introduit SQL, et dirige sur la fiche SQLite.

Question 3

a. Lancer SQLite comme indiqué dans la question 2 ci-dessus.

b. On suit et utilise toujours étroitement le matériel présenté dans la fiche SQLite. Taper dans SQLite :

```
create table billet (prenom varchar(15), destination varchar(20), transport varchar(9), jour smallint,
prix smallint, statut varchar(8));
```

Taper dans SQLite :

```
select * from billet;
```

La table est encore vide, rien n'est affiché.

c.

```
insert into billet values ('Rita', 'New York', 'avion', 364, 450, 'paye');
insert into billet values ('Riton', 'New York', 'avion', 365, 550, 'paye');
select * from billet where prenom = 'Rita';
```

On obtient une seule ligne(et d'ailleurs dans la question 1 il a été supposé que tous les prénoms sont différents et que chaque élève n'a qu'un billet).

```
select * from billet where destination = 'New York';
```

On obtient deux lignes, et plus généralement tous les billets pour New York.

```
insert into billet values('Ada', 'Amsterdam', 'train', 210, 49, 'paye');
```

d.

```
update billet set statut = 'effectue' where nom = 'Rita';
```

```
select * from billet;
```

Noter que dans cette proposition de solution à la question d, on a essayé d'être réaliste en faisant voyager l'élève qui a le billet à la date la plus ancienne.

La page du codeur

1 Observer les premières lignes

- a. Pas de question donc pas besoin de correction particulière : ouvrir avec la souris ou une commande unix : `xdg-open nomfichier.csv`, ou `votre-editeur-de-texte-prefere nomfichier.csv`, ou `head nomfichier.csv` si on ne veut que les premières lignes ... Excel est envisageable mais déconseillé car il risque de mettre en forme automatiquement, on ne verra alors pas le délimiteur.
- b. La première ligne est ici un en-tête indiquant ce qu'on trouve dans le document. De manière anecdotique, on pourra observer que ce « titre » a été complété par une série de virgules, parce que le créateur du fichier voulait que toutes les lignes aient le même nombre de colonnes y compris la ligne de titre.
- c. La seconde ligne est très importante, c'est le véritable en-tête contenant les descripteurs de chaque colonne.
- d. Les lignes à partir de la ligne 3 contiennent les données.

2 Lire un programme

On commence par importer le module (i.e., charger la librairie) pandas qui, comme indiqué ci-dessus, sert à manipuler les collections.

La ligne suivante nous permet d'utiliser les fonctions DataFrame et read_csv (sans avoir à les préfixer par le nom de module). Le nom indique assez clairement ce que fait chaque fonction.

Enfin la dernière ligne charge le module matplotlib qui sert à tracer une figure.

3 Charger les données

Si on voulait récupérer automatiquement les descripteurs à partir de la 2^e ligne du fichier :

```
elec =  
pd.read_csv('/home/tphome002/bgroz/electricite.csv',sep=',',skiprows=1,header=0,names=None)
```

Mais pour avoir des noms de descripteurs plus pratique on vous a demandé de plutôt créer le dataframe avec vos propres noms:

```
elec = pd.read_csv('/home/tphome002/bgroz/electricite.csv',\  
sep=',',\  
names=['mois','electr. totale',' primaire','nucleaire','hydraulique','eolien','thermique'],\  
skiprows=2)
```

C'est une bonne idée de vérifier le résultat :

```
print(elec.to_string())
```

4 Énergie primaire

a.

```
elec[elec['primaire'] != elec['nucleaire'] + elec['hydraulique'] +  
elec['eolien']]
```

b. On observe qu'il n'y a qu'une ligne pour laquelle ce n'est pas le cas : sur cette ligne l'énergie primaire est 34 450, alors que la somme des autres est 34 451. Cette incohérence est donc très probablement due à des arrondis (les valeurs en GWh de chaque énergie ont probablement été obtenues en arrondissant les valeurs exactes de consommation).

5 Répondre à une requête de recherche

La figure obtenue représente pour l'évolution de la consommation au cours du temps. En fait, on voit que sur l'axe des x c'est le numéro de ligne (attribué automatiquement par pandas) qui a été utilisé. La colonne contenant la date n'a pas été exploitée, et par contre une ligne a été tracée pour chaque autre colonne de la table.

Guide du professeur - Chapitre 5

L'objectif de ce chapitre est de présenter le principe de la géolocalisation, de montrer le lien entre les coordonnées GPS et les coordonnées géographiques, de faire manipuler des services de cartographie en ligne, par exemple en calculant des itinéraires. L'idée principale qui sous-tend ce chapitre est celle de la mobilité, prise au sens large. Comment se repère-t-on à la surface de la Terre, comment s'oriente-t-on et quels sont les futurs de la mobilité, sous l'effet des progrès technologiques ?

Histoire, enjeux, débats

Ce chapitre permet d'aborder les principes de la mobilité à l'ère du numérique, à travers la numérisation des cartes et des services de cartographie en ligne. Les progrès en la matière posent de nombreuses questions, notamment celle d'une possibilité de pistage généralisée (controverse) ou celle des futurs possibles de la mobilité (Monde de demain).

- ❖ Activité 1 La géolocalisation
- ❖ Activité 2 Du GPS aux coordonnées géographiques
- ❖ Activité 3 Les cartes numériques
- ❖ Activité 4 Calculs d'itinéraires

Bande dessinée d'ouverture

Cette bande dessinée permet de poser la question de la pertinence des systèmes GPS embarqués, censés être une aide fiable à la mobilité. On verra, dans l'activité 4, que les calculs d'itinéraires sont assez compliqués (s'appuyant sur la théorie des graphes) et que leur actualisation en temps réel n'est pas toujours évidente. On verra aussi que deux systèmes de cartographie peuvent des itinéraires différents pour un même trajet, voire des temps de transport différents pour un même itinéraire.

Note : Les pages **Controverse** et **Monde de demain** ne sont pas corrigées. En effet, dans les premières, il est plus question d'ouvrir un débat avec les élèves, les questions n'étant là que pour initier une discussion, et pour les secondes, le but est de prendre connaissance des pistes de recherche actuelles sur ce sujet.

Les activités

ACTIVITÉ 1

Les compétences travaillées

Cette activité apporte des connaissances techniques sur le principe de géolocalisation, lui-même s'appuyant sur le principe de trilateration en trois dimensions.

Les savoir-faire travaillés

- Analyser un texte pour comprendre le fonctionnement du système GPS.
- Représenter mathématiquement le principe du GPS.
- Comparer les avantages et inconvénients de deux systèmes.

Question 1

Utilisation du document 1

Le système GPS est un système de positionnement (ou géolocalisation) par satellites. Il est constitué de 24 satellites. On pourra détailler cette réponse avec les éléments du texte ou en cherchant des informations sur Internet (notamment le site du CNES).

Question 2

Utilisation du document 2

- Lorsque le satellite 2 émet un signal, cela crée, mathématiquement, un cercle dont ce satellite est le centre. L'intersection de deux cercles peut aboutir à 0, 1 ou 2 points. Ici, on fait l'hypothèse que l'on obtient 2 points d'intersection.
- De la même manière qu'à la question précédente, on obtient deux points, dont l'un commun (le réglage des signaux est fait pour). Ainsi, on obtient un lieu précis.

Question 3

Utilisation du document 3

a. Question à faire sur ordinateur.

On pourra introduire ici la notion de sphères tangentes extérieurement ou intérieurement.

b. Le quatrième satellite permet de synchroniser l'horloge du récepteur, compenser l'incertitude liée aux horloges et réduire l'incertitude.

Question 4

Utilisation du document 4

L'idée était de rendre l'Europe indépendante des États-Unis du point de vue du positionnement pas satellite, essentiellement au départ pour les militaires. Les avantages : système civil, plus précis que le système américain.

ACTIVITÉ 2

Les compétences travaillées :

Cette activité apporte des connaissances techniques sur les protocoles de transmission de données, ainsi que sur la manière de retrouver les coordonnées géographiques à partir de données fournies par un système de géolocalisation.

Les savoir-faire travaillés :

- Analyser un protocole NMEA.
- Extraire des coordonnées géographiques.
- Utiliser des données de géolocalisation.

Question 1

Utilisation du document 1

Cette création avait pour but d'harmoniser et de standardiser les équipements de la marine.

Question 2

Utilisation du document 2

Il faut ici extraire les coordonées géographiques et l'heure indiquée.

- a. Latitude 67° 15,12 minutes Nord, Longitude 32° 45,12 minutes Ouest, 15 h 34 min 21 s.
- b. Latitude 115° 43,65 minutes Sud, Longitude 23° 56,76 minutes Ouest, 4 h 54 min 32 s.

Question 3

Utilisation du document 3

On y trouve les coordonnées géographiques en trois dimensions : latitude, longitude et altitude (appelé côte en mathématique).

Pour localiser sur une carte, on peut ouvrir GoogleMaps par exemple et rentrer les coordonnées géographiques.

Question 4

Utilisation du document 4

On laissera ici les élèves donner leur avis sur le travail d'Eric Fischer.

On pourra les guider vers une des réponses possibles : les lieux où les touristes vont, dans une ville comme Paris (ce qui est vrai aussi pour les grandes capitales dans le monde) ne sont pas là les gens habitent. Mais il y a d'autres phénomènes sociaux visibles naturellement.

ACTIVITÉ 3

Les compétences travaillées

Cette activité apporte des connaissances sur les **applications ou sites Internet de cartographie**.

Ces connaissances peuvent être mises en œuvre par un **travail d'équipe ou personnel** de comparaison des résultats des différentes applications ou sites, que ce soit pour localiser des lieux ou bien pour effectuer des calculs d'itinéraires.

Les savoir-faire travaillés

- Utiliser des outils de cartographie.
- Calculer des itinéraires et comprendre le fonctionnement des algorithmes de calculs.
- Contribuer à des sites collaboratifs.

Question 1

Utilisation du document 1

Elles ont permis de manipuler de nombreuses couches d'informations sur la même carte, mais également de faciliter les changements d'échelle.

Question 2

Utilisation du document 2

C'est sur la variété des sources d'informations publiques que c'est crée le Géoportail. Les informations proviennent de l'IGN naturellement, mais aussi de très nombreuses structures publiques donnant des informations sur les sites protégés, les zones à risques, etc.

Question 3

Utilisation du document 3

Question à faire sur ordinateur.

- a. En cliquant sur l'icône de la légende, on obtient une image du lieu choisi. Par défaut, c'est une photographie aérienne, mais cela peut être autre chose (plan, carte IGN, différentes informations) car cela dépend des utilisateurs précédents. En effet, le Géoportail grade en mémoire les derniers réglages effectués sur un ordinateur.

En choisissant « Carte IGN » dans le menu, on obtient différentes informations dans la légende. Exemple ci-contre.

- b. On les obtient en cliquant sur l'onglet : . Cela dépend des réglages.

RÉSEAU ROUTIER	
	Autoroute, section à péage
	Autoroute, section gratuite
	Autoroute en construction
	Échangeur partiel ou complet
	Barrière de péage, Aire de repos et de service
	Liaison principale
	Liaison régionale
	Liaison locale
	Route en construction
	Autre route
RÉSEAU FERRÉ	
	Voie ferrée, Gare
	Téléphérique
HYDROGRAPHIE ET INFRASTRUCTURES HYDROGRAPHIQUES	
	Surface d'eau ou cours d'eau large
	Marais
	Cours d'eau large permanent

c. Afin d'obtenir une carte du relief, il faut choisir « Carte du relief » dans « Tous les fonds de carte ».

On obtient quelque chose comme :



Question 4

Utilisation du document 4

Question à faire sur ordinateur.

a. On trouve ces informations dans le Wiki d'OpenStreetMap :

https://wiki.openstreetmap.org/wiki/FR:Contribuer_aux_donn%C3%A9es_cartographiques

b. À faire par chaque élève.

ACTIVITÉ 4

Les compétences travaillées

Cette activité apporte des connaissances sur la notion d'itinéraire et le lien que l'on peut faire avec la théorie des graphes (déjà abordée dans le chapitre 3 sur les réseaux sociaux).

Les savoir-faire travaillés

- Connaitre les paramètres qui entrent en jeu dans les calculs d'itinéraires.
- Utiliser un site de cartographie en ligne pour calculer des itinéraires.
- Calculer des itinéraires en utilisant la relation entre vitesse, distance et durée.

Question 1

Utilisation du document 1

a. Les deux dimensions de la notion d'itinéraire présentées dans le texte sont d'une part celle de l'espace physique (routes, moyens de transports, etc.) et d'autre part celle du territoire social et culturel dans lequel se situe cet espace physique. Ce dernier aspect est particulièrement important dans la mesure où il structure l'espace physique. Le choix de construire une autoroute à un certain endroit dépend du contexte social et culturel de l'époque, des enjeux politiques sous-jacents, etc.

b. Le calcul d'un itinéraire dépend de nombreux paramètres, mais il paraît évident que le point de départ est le paramètre qui donne naissance à l'itinéraire lui-même. La définition donnée montre en effet qu'un itinéraire est un chemin à suivre pour aller d'un lieu à un autre, si bien que le chemin sera différent pour chaque point de départ. Plus la connaissance des coordonnées du point de départ seront précises, plus le calcul de l'itinéraire pourra être précis.

Question 2

Utilisation du document 2

L'itinéraire de gauche propose d'effectuer les 5 kilomètres qui relient le point de départ au point d'arrivée en 20 minutes. Le second propose un autre itinéraire pour relier les mêmes points de départ et d'arrivée, pour une distance de 7,1 km effectuée en 32 minutes. Sur le plan de droite deux autres parcours alternatifs sont proposés, avec des durées moindres.

De nombreux éléments peuvent expliquer ces différences. On montera, tout d'abord que le calcul d'itinéraire a été demandé pour une circulation à vélo (plan de droite). L'algorithme peut prendre en compte la densité du trafic, les travaux éventuels sur le trajet ainsi que la présence ou non de pistes cyclables. GoogleMaps, par exemple, privilégie la présence de pistes cyclables (pour une question de sécurité) même si le trajet est finalement plus long. On peut aussi faire l'hypothèse que les deux calculs d'itinéraires n'ont pas été effectués à la même heure, qui pourrait expliquer une densité du trafic différente.

Question 3

Utilisation du document 3

Exercice à faire sur ordinateur.

Question 4

Utilisation du document 4

a. Afin de déterminer le plus court chemin entre les villes A et H, il est nécessaire de passer en revue tous les chemins possibles et de calculer toutes les distances (c'est ce que fait un

algorithme afin de déterminer l'itinéraire le plus court). On trouve finalement que le chemin le plus court est le chemin qui passe par les villes A – B – F – H, qui fait 32 kilomètres.

Si l'on peut rouler à la même vitesse sur toutes routes, alors, oui, ce chemin est le plus rapide. Cependant comme les limitations de vitesse ne sont pas indiquées, on ne peut rien conclure en l'état quant à la rapidité des itinéraires.

b. La trajet le plus court, déterminé dans la question précédente, prendra, en considérant que l'on roule au maximum de la vitesse autorisée, 24 minutes.

Mais, étant donné que l'on peut rouler à 130 km/h sur les routes DE, EH et FH, il faut vérifier que les itinéraires empruntant ces routes ne sont pas plus rapides.

Itinéraire A – D – E – H : 34 kilomètres en environ 20 minutes.

Ce dernier est donc plus long mais plus rapide avec les vitesses données.

Je vérifie mes connaissances

Solutions détaillées

1. QCM

1. b. 2. b. 3. a. 4. a. et b.

2. Vrai/Faux

- a. **FAUX.** Au contraire, il s'appuie sur des mesures de distances.
b. **VRAI.** c. **VRAI.**
- d. **FAUX.** Une trame peut se terminer par différentes lettres, le « A », par exemple, signifie « Données valides ».
e. **VRAI.**
- f. **FAUX.** Les applications de calculs d'itinéraires n'utilisent pas les mêmes algorithmes de calculs, d'où des résultats différents.
g. **FAUX.** Le chemin **1 3 5 6** donne 8, alors que le chemin **1 3 4 6** donne 7.

3. Vers une définition

La géolocalisation est l'ensemble des techniques permettant de **localiser** en continu un objet, comme un téléphone portable ou un véhicule, sur lequel un **récepteur GPS** a préalablement été posé, en le positionnant sur une **carte** à l'aide de ses **coordonnées géographiques**.

4. Compléter un schéma

Voir manuel papier.

5. Restituer des connaissances

GP : récepteur GPS

GLL : positionnement géographique longitude-latitude

2311.12,W : 23° 11' 12" Ouest

6. Organiser ses connaissances

- a. Un système GPS utilise une trilateration en trois dimensions. On utilise donc des sphères à la place des cercles : l'intersection de deux sphères correspond à un cercle, et l'intersection de 3 sphères correspond à deux points. En théorie, il serait donc nécessaire d'utiliser un quatrième satellite pour savoir où on se trouve. En pratique, on élimine l'un des deux points car il ne se trouve pas sur Terre mais à une position en altitude. Cependant, pour la synchronisation de l'horloge du récepteur, il faut la précision d'une horloge atomique. Un récepteur n'en a évidemment pas. Il faut donc un quatrième satellite pour compenser cette imprécision et réduire l'incertitude.
- b. Les différents types de données (agricoles, culturelles, économiques, sociales etc.) s'appliquant sur le lieu choisi apparaissent sous forme de couches d'information superposées au fond cartographique. Ces couches sont accessibles via l'onglet situé sur la droite de l'écran, accompagnées d'informations sur la provenance de ces données.

JE M'ENTRAÎNE

7. Le calcul d'une distance

a. La formule permettant de calculer la distance, notée d , en fonction du temps (noté t) fait intervenir la vitesse (que l'on note v). On peut ici renvoyer les élèves aux compléments mathématiques (page 178 du manuel).

Cette formule s'écrit : $d = v \times t$,

avec d en mètres (m), v en mètres par secondes (m.s^{-1}) et t en secondes (s).

b. On applique la formule précédente avec $t = 0,006$ s et $v = 3.10^8$ m.s^{-1} .

Ainsi, $d = 3.10^8 \times 0,006 = 1,8.10^6$ m = 1 800 km.

Le récepteur GPS et le satellite sont séparés d'une distance de 1 800 km.

Compléments : À titre de comparaison, on peut rappeler que les satellites du système GPS se situent sur une orbite dont le rayon est légèrement supérieur à 20 000 km. Et que les avions volent, en moyenne, à 10 km au-dessus du niveau de la mer.

8. Coordonnées géographiques

On écrit les coordonnées (longitude, latitude).

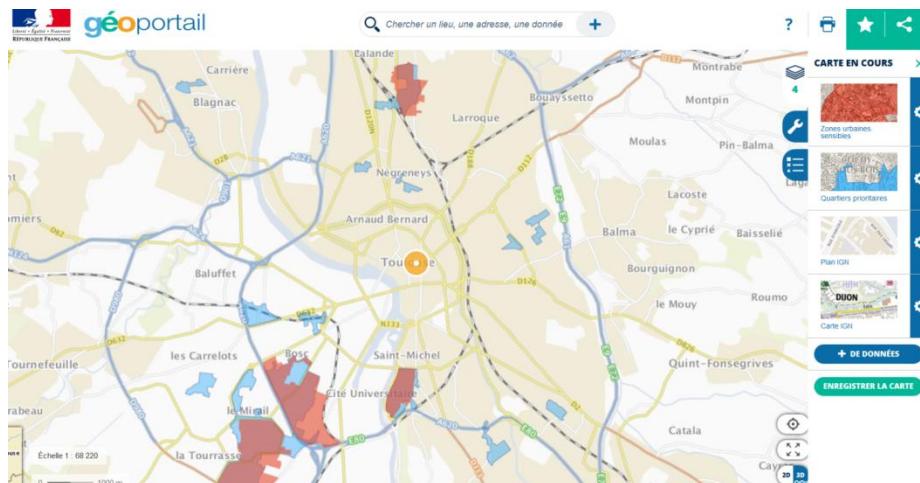
N (90° N, 0°) ; V (60° N, 40° O) ; U (20° N, 0°) ; P (0°, 40° O) ; Y (20° N, 60° E) ; Q (40° S, 10° E) et S (90° S, 0°).

9. Les couches d'informations

Exercice à faire sur ordinateur.

a. b. c. Afin de trouver ces couches d'information, il faut ouvrir le menu déroulant « Données thématiques ». Puis, ouvrir le sous-menu « Société et loisirs », et enfin, dans le menu « Zones de gestion », on trouve les couches d'information « Quartiers prioritaires » et « Zones urbaines sensibles ».

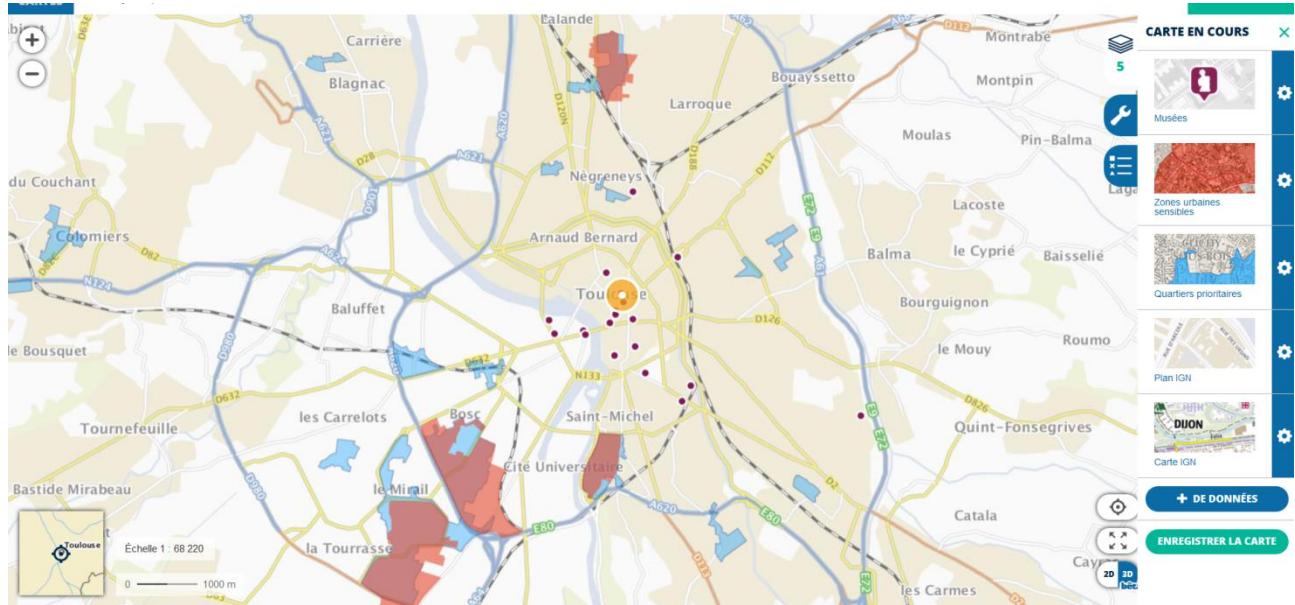
On doit obtenir quelque chose qui ressemble à l'image ci-dessous :



d. On constate qu'il y a superposition des deux types de zones, ce qui tendrait à montrer que les c'est dans les zones urbaines sensibles que les pouvoirs publics ont décidé de mener une politique prioritaires dans les quartiers. Ce qui semble assez logique.

e. La couche d'information « Musées » se trouve dans le menu « Loisirs ».

On obtient la représentation ci-dessous dans laquelle les musées apparaissent sous la forme de points violetts :



La grande majorité des musées sont situés dans le centre ville de Toulouse (sauf deux, un peu à la périphérie).

Ce que l'on peut constater avec ces trois couches d'information, c'est que les musées ne sont pas situés dans les zones urbaines sensibles, ni dans les quartiers prioritaires.

Compléments : On pourrait demander aux élèves de tester cette hypothèse (pas de musées dans les zones sensibles) dans d'autres grandes villes françaises (Paris, Lyon, Marseille, Rennes, etc.), voire des grandes métropoles dans le monde. Puis les faire réfléchir, en groupe de préférence, sur ce lien qui existe entre les lieux culturels dans les villes et la question sociale.

10. Des modifications

Exercice à faire sur ordinateur.

Non corrigé dans la mesure où les réponses dépendent du lycée choisi.

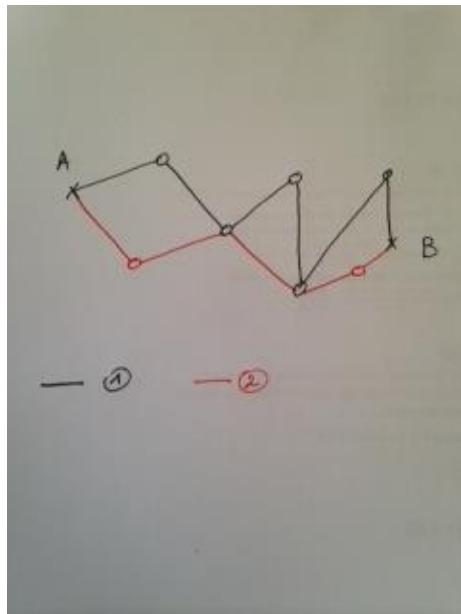
11. Comparaison de cartes

Exercice à faire sur ordinateur.

Non corrigé.

12. Un graphe

a. Exemple d'itinéraires. Le schéma peut être différent selon les élèves.



b. Il faut bien penser à ajouter les distances entre chaque ville également afin de pouvoir calculer l'itinéraire le plus rapide.

13. La prise en compte du trafic

Exercice à faire sur ordinateur.

a. Dans des conditions normales de circulation, ce trajet prend 4 heures et 21 minutes (le trajet le plus rapide) en voiture, 1 h 57 en train, 26 h en vélo et 91 h à pieds.

b. La durée peut varier selon les jours (le samedi, par exemple, présente une plus grande circulation, ce qui fait que les durées de trajet sont plus longs).

c. Oui, globalement, c'est le même itinéraire qui est le plus rapide, mais on constate bien que ce trajet le plus rapide n'a pas la même durée selon les heures de la journée. GoogleMaps prend en compte les difficultés de circulation qui pourraient exister selon les horaires.

14. Le dispositif eCall

a. Un accident se produit. Le véhicule, équipé de capteurs qui, en fonction de la gravité du choc, peuvent déclencher automatiquement l'appel vers la plateforme à l'aide d'une carte SIM embarquée. Cet appel par vers un centre de secours, tel que 112. Le dispositif est conçu pour transmettre au 112 des informations importantes telles que le type de carburant utilisé par le véhicule ou encore le nombre de passagers.

b. Ce système est basé sur la géolocalisation type GPS, qui est peut-être très précis afin de localiser le lieu où l'accident s'est produit. Plus la précision est grande, plus les secours peuvent arriver rapidement sur le lieu.

FAB LAB

Ce FabLab ne peut pas présenter un corrigé dans la mesure où l'œuvre réalisé dépend du choix des lieux et du lycée.

La question **1** ne pose pas de difficultés majeures, il suffit de retrouver les informations demandées, puis des photographies en fonction des lieux.

Les questions **2** et **3** font appel à des manipulations d'images, soit physique soit à l'aide du logiciel Gimp. On pourra, pour ce dernier, se reporter à la fiche de prise en main de Gimp.

La page du codeur

1 Utilisation d'une chaîne

a.Une trame NMEA-0183 se présente sous la forme d'une chaîne de caractères, contenant un certain nombre d'informations.

b.Les données de ce fichier sont des chaînes de caractères.

2 Utilisation d'une expression régulière

Voici un programme possible qui répond aux questions.

```
# coding=utf8
# Python 3.6+
# NMEA to OpenStreetMap and Google

import re

defdecode_nmea(trame):
    print(trame)
    regex = r"\$GPGLL,(\[d]{1,2})(\[d]{2})\.(\[d]{2}),(\[w]{1}),(\[d]{1,3})(\[d]{2})\.(\[d]{2}),(\[w]{1})"
    try:
        groups=re.search(regex, trame)
        if groups:
            lat=(-1,-1)[groups[4]== "S"]*(int(groups[1])+(int(groups[2])+int(groups[3])/100)/60)
            long=(-1,-1)[groups[8]== "W"]*(int(groups[5])+(int(groups[6])+int(groups[7])/100)/60)
            returnlat,long
        else:
            print("Les coordonnées ne sont pas bien formatées")
            return 0,0
    except:
        print("Il y a eu une erreur inconnue dans le décodage de la trame")
        return 0,0

    def display(lat,long):
        print(f"Latitude : {lat}, Longitude: {long}")
        print(f"OSM URL : https://www.openstreetmap.org/#map=18/{lat}/{long}")
        print(f"GoogleMap URL : https://www.google.com/maps/search/?api=1&query={lat},{long}")

    defstart_nmea(trame):
        if trame.startswith('$') and trame.endswith("A"): # Vérifier que la trame est valide
        if trame.startswith("$GP"): # Vérifier que la trame est une trame GPS
        if trame.startswith("$GPGLL"): # Vérifier que la trame est au format GLL
            lat,long=decode_nmea(trame) # Appel de la fonction decode_nmea
            if lat and long:
                display(lat,long) # Affichage du décodage
            else:
```

```

print("La trame n'a pas au format GLL")
else:
print("La trame n'est pas une trame GPS")
else:
print("La trame n'est pas valide")

trame="$GPGLL,4631.35,N,634.44,E,205532,A"
start_nmea(trame)

```

Et une version simplifiée, qui n'utilise pas de fonctions :

```

# coding=utf8
# Python 3.6+
# Decode NMEA

import re

trame="$GPGLL,4631.35,N,634.44,E,205532,A"
regex = r"\$GPGLL,(\d{1,2})(\d{2}).(\d{2}),(\w{1}),(\d{1,3})(\d{2}).(\d{2}),(\w{1})"
groups=re.search(regex, trame)
if groups:
lat=(1,-1)[groups[4]=="S"]*(int(groups[1])+(int(groups[2])+int(groups[3])/100)/60)
long=(1,-1)[groups[8]=="W"]*(int(groups[5])+(int(groups[6])+int(groups[7])/100)/60)
print(trame)
print(f"Latitude : {lat}, Longitude: {long}")
else:
print("Les coordonnées ne sont pas bien formatées")

```