



UNIVERSITÉ PARIS 1

PANTHÉON SORBONNE



UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE



1.1 Mener une recherche et une veille d'information

Sommaire

- Les sites web : ce qu'il faut savoir
- Recherche d'informations : moteurs de recherche et requêtes
- Evaluation de l'information
- Source et citation
- Veille d'information, flux et curation

Sommaire

- Les sites web
- Recherche d'informations : moteurs de recherche et requêtes
- Evaluation de l'information
- Source et citation
- Veille d'information, flux et curation

Les sites Web : WEB/Internet

Internet

Internet est la structure technique sur laquelle est construit le **WEB**. Certains composants sont essentiels : **DNS** - permettant à un humain de trouver un site et le **TCP** - protocole permettant la communication entre un navigateur et un serveur



- 1) DNS : Le Domain Name System ou DNS est un service informatique distribué utilisé qui traduit les noms de domaine Internet en adresse IP ou autres enregistrements
- 2) TCP : Transmission Control Protocol (littéralement, « protocole de contrôle de transmissions »), abrégé TCP, est un protocole de **transport fiable**, en **mode connecté**,

Les sites Web : définitions

Définitions

- Un site web est défini par son adresse URL (Uniform Resource Locator) ;
- L'adresse commence par http:// ou https:// si le site est sécurisé.
- Une page Web est écrite dans un langage de type **HTML**(HyperText Markup Language) à base de balises.
- Le contenu informationnel de la page est structure dans un langage appelé **CSS** (Cascading Style Sheet)



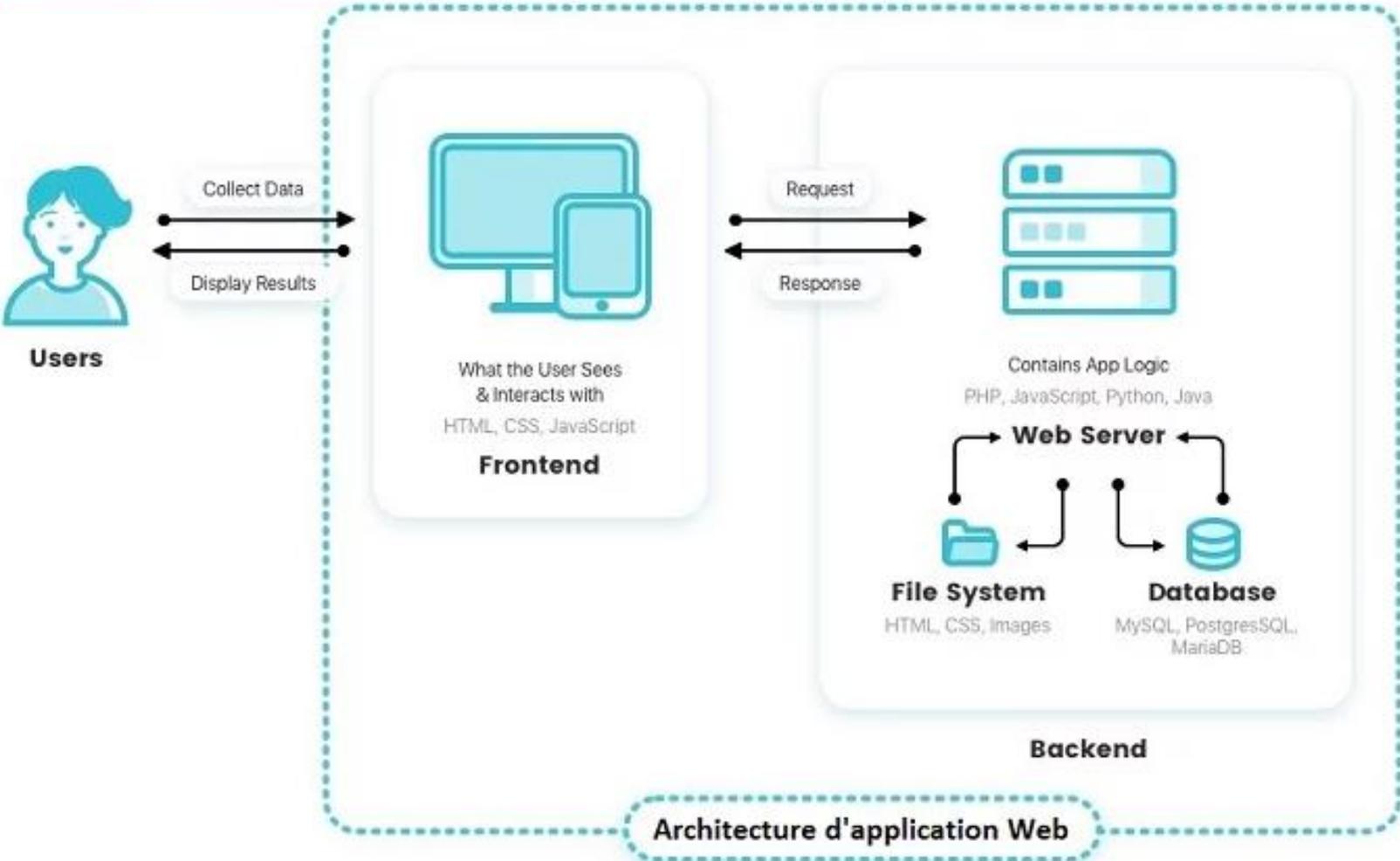
Accès libre/ accès réservé

- Accès totalement libre : **pas d'identification**, par exemple, les sites institutionnels.
- Accès soumis à une identification : accès réservé aux utilisateurs munis d'un **identifiant** et d'un **mot de passe** (par exemple, l'ENT, l'EPI).

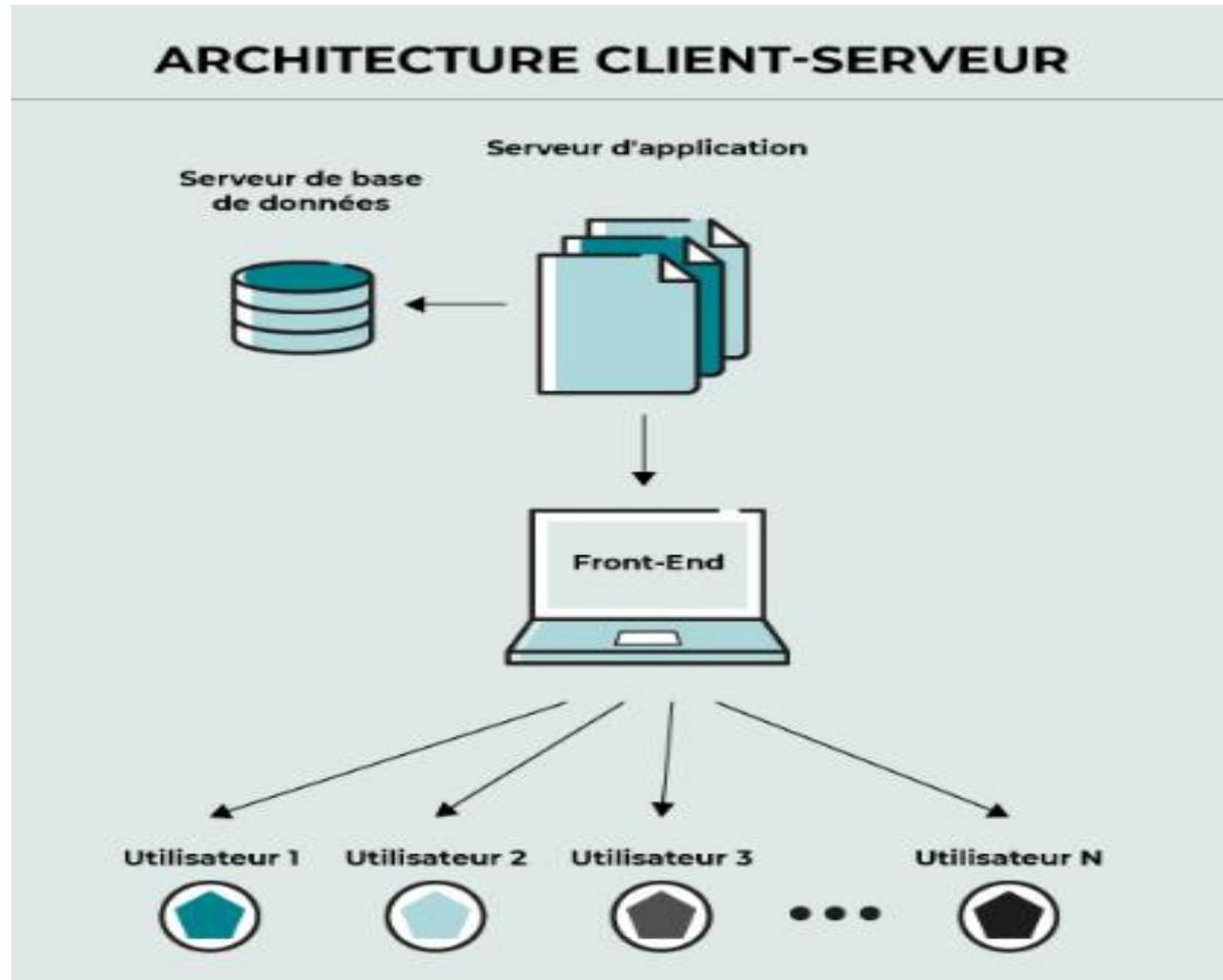
Page statique/Page dynamique

- Une **page dynamique** utilise des scripts (programmes) et elle est liée à une base de données qui alimente les page Web d'informations qui sont rafraichies au fur et à mesure (commande en ligne et gestion du stock en back end).
- Une **page statique** ne fait pas appel à une base de données, la page statique évolue peu (par exemple CV en Ligne)

Les sites Web : Architecture



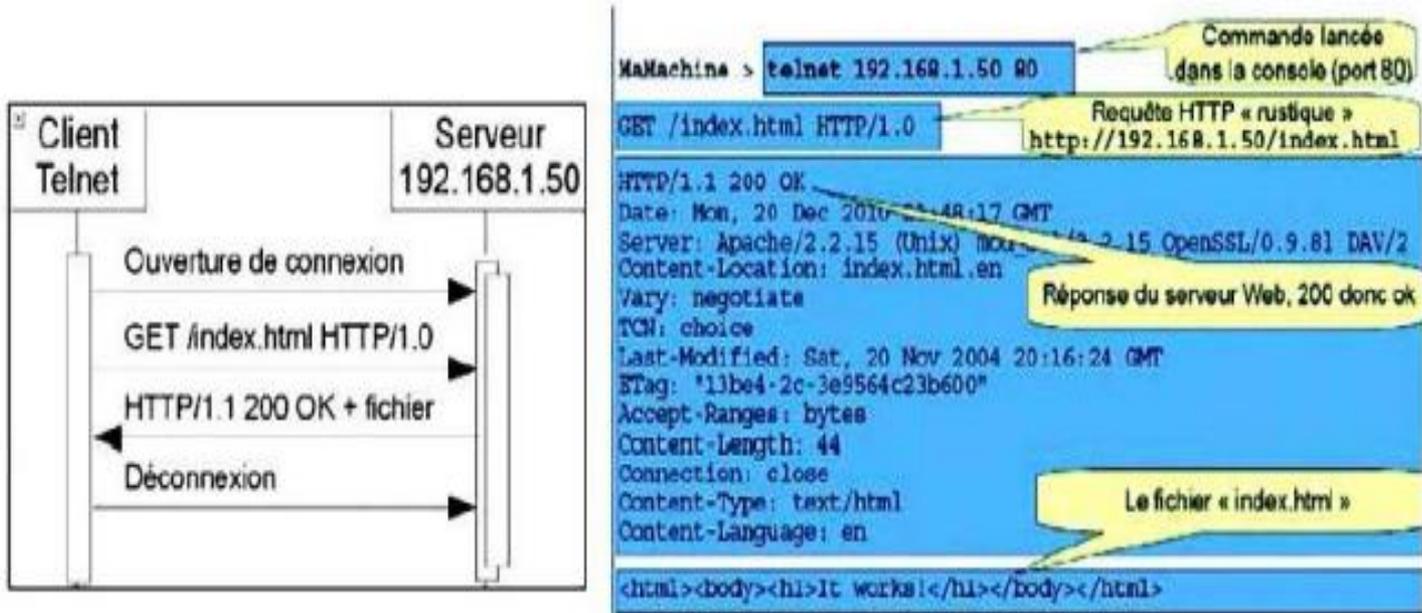
Les sites Web : Client Serveur



Les sites Web : Communication

- Un internaute connecté au réseau via son ordinateur et un navigateur web est le client, le serveur est constitué par le ou les ordinateurs contenant les applications qui délivrent les pages demandées. Dans ce cas, c'est le protocole de communication HTTP (HyperText Transfer Protocol) qui est utilisé. Les navigateurs sont les clients (Firefox, Internet Explorer, ...). Ces clients se connectent à des serveurs http tels qu'Apache http Server ou IIS (Internet Information Services).
- HTTPS (avec S pour Secured ou Sécurisé) est la variante de http sécurisée par l'usage des protocoles SSL (Secure Sockets Layer) ou nouvellement nommé TLS (Transport Layer Security) qui est un protocole de sécurisation des échanges sur internet.
- Dans le protocole http, une méthode est une commande spécifiant un type de requête, c'est-à-dire qu'elle demande au serveur d'effectuer une action. En général, l'action concerne une ressource identifiée par l'URL qui suit le nom de la méthode.
- Les méthodes les plus utilisées sont GET et POST. GET : c'est la méthode la plus courante pour demander une ressource. POST : cette méthode doit être utilisée pour ajouter une nouvelle ressource, comme un message sur un forum, un article dans un site ou encore un login et un mot de passe. Les autres méthodes sont : HEAD, OPTIONS, CONNECT, TRACE, PUT, DELETE

Les sites Web : Structure d'une requête HTTP



Pour plus de détails, **A LIRE**

<https://www.maxicours.com/se/cours/la-requete-http/>

Les sites Web : FTP

- Le protocole de transfert de fichiers, ou FTP (**File Transfer Protocol**), est un protocole de communication destiné à l'échange informatique de fichiers sur un réseau TCP/IP.
- Il permet, depuis un ordinateur, de copier des **fichiers** vers un autre ordinateur du réseau, d'alimenter un site web, ou encore de supprimer ou de modifier des fichiers sur cet ordinateur.
- FTP obéit à un modèle client/serveur, c'est-à-dire qu'une des deux parties, le client, envoie des requêtes et le serveur répond.
- Pour accéder à un serveur FTP, on utilise un logiciel client FTP (possédant une interface graphique comme FileZilla par exemple ou en ligne de commande).



Les sites web :

typologie des ressources web

- Publiés par « tout le monde » :
 - sites commerciaux, institutionnels, personnels, etc. ;
 - ils sont recensés par les **moteurs de recherche** ou les **annuaires**.
- Publiés par des organismes reconnus :
 - bibliothèques, articles de presse ou de revue scientifique, etc. ;
 - ils sont recensés par les **catalogues** des bibliothèques ou par des **bases de données**.

Sommaire

- Les sites web
- Recherche d'informations : moteurs de recherche et requêtes
- Evaluation de l'information
- Source et citation
- Veille d'information, flux et curation

Recherche d'information

Les moteurs de recherche : Définition

- Un moteur de recherche est un site Web qui permet d'apporter aux utilisateurs les **ressources d'information** demandées (pages Web, forums, images, documents, vidéos...etc)
- Il existe sur le Web des centaines de moteurs de recherche,
- Un moteur de recherche est continuellement mis à jour par des **robots** (bots).
- Un robot appelé aussi **spiders** est un logiciel qui collecte automatiquement les nouvelles ressources d'information et les sauvegarde dans l'index du moteur de recherche .

Recherche d'information

Les moteurs de recherche : Définition et fonctionnement

Un moteur de recherche est un site qui **recense et référence** des pages internet de manière organisée : pour cela, il utilise un logiciel, le "**robot** d'indexation« (ou **spider**), qui parcourt l'ensemble du web.

Le moteur de recherche travaille en trois étapes :

- L'**exploration** (en anglais "**crawling**") : le robot d'indexation explore régulièrement le web en suivant tous les liens qu'il trouve et analyse les pages jugées intéressantes.
- L'**indexation** : il extrait les mots clés de chacune des pages visitées (mots choisis par le webmestre ou apparaissant sur la page) et conserve une copie de ces pages : le cache.
- Le **classement** ("**ranking**") : le moteur de recherche classe les pages selon leur pertinence par rapport à une série de mots clés. Cela lui permet, quand vous tapez les mots de votre recherche, de retrouver le contenu recherché. Ce classement est lié notamment à la "popularité" de chaque page et au nombre de liens qui dirigent vers celle-ci

Recherche d'information

Les moteurs de recherche : tâches effectuées

- Un moteur de recherche réalise les tâches suivantes :
 - **moissonnage** : les pages du web sont parcourues automatiquement par un robot. Les robots suivent les liens hypertextes (qui relient les pages les unes aux autres) rencontrés sur chaque page.
 - **indexation automatique** : les pages sont analysées pour en extraire les mots significatifs. Chaque page visitée est indexée dans une base de données, accessible ensuite par les internautes à partir de mots clés.
 - **interrogation en fonction d'une requête** : les mots clés de la requête de l'internaute sont comparés aux mots extraits par l'indexation et une liste de pages web sélectionnées est affichée par ordre de pertinence.
- Ex de moteurs de recherche : Google, Yahoo!, Bing, Yandex (Russe), Baidu (Chinois), Exalead :
 - avec 60% des recherches sur le Web, Google est le principal moteur de recherche.

Recherche d'information

Les moteurs de recherche : indexation des pages

- **Pour chaque page web indexée :**
 - une copie de la page est conservée au moment de son indexation ;
 - conséquence : les informations supprimées resteront encore accessibles un certain temps en cache.
- **Pour optimiser l'indexation des pages web :**
 - les moteurs de recherche doivent parcourir chaque page d'un site. Pour faciliter le travail des robots d'indexation, les concepteurs du site doivent s'assurer que chaque page est accessible par un lien hypertexte explicite ;
 - les concepteurs de sites placent des **mots clés** dans l'en-tête de leur code : **les métadonnées**.
 - les mots clés doivent être bien choisis de manière à optimiser la visibilité du site sur les moteurs de recherche.

Recherche d'information

Les moteurs de recherche : pertinence des résultats affichés

- Chaque moteur a ses propres règles pour établir la pertinence d'une page par rapport à une requête :
 - **algorithme de recherche** : le poids des mots recherchés est mis en relation avec ceux contenus dans les pages web ;
 - l'« **indice de popularité** » de chaque page web est calculé en fonction du nombre et de l'importance des sites référençant cette page (Page Rank de Google).
 - certains liens apparaissant en haut ou sur le côté sont des liens commerciaux :
 - Les moteurs de recherche se financent par la vente de publicité. Ils présentent des publicités correspondant aux mots clefs indiqués par l'internaute. Ex : pour une recherche d'hôtels dans une ville, il vous présente juste une sélection d'hôtels.
 - Vous reconnaitrez ce lien car il est indiqué annonce ou lien commercial. Il pointe toutefois vers l'adresse réelle du site.

Recherche d'information

moteurs de recherche : le langage de requête

- Le principe des langages de requête est de combiner les mots clés de façon à exprimer le plus de sens : on peut préciser le rôle que chaque mot joue, par rapport aux autres mots ou par rapport à la requête dans son ensemble.
- Il est possible d'affiner la recherche en utilisant des **opérateurs logiques ou booléens** :
 - en mettant le signe - (ou **SAUF** ou **AND NOT**) devant les mots ou URL à exclure ;
 - en mettant les expressions exactes entre **guillemets** ;
 - en proposant une liste de mots au choix séparés par **OU** ou **OR**, il suffit alors que l'un des critères soit vérifié ;
 - en utilisant **ET** ou **AND**, ce qui rend tous les critères obligatoires ;
 - en regroupant les opérateurs booléens, on peut construire des équations de recherche ;
 - Ces termes sont communs à tous les outils de recherche sur l'internet.
- Au-delà de ces opérateurs booléens, il existe l'opérateur de troncature * qui permet de ne donner que le début d'un mot-clé pour inclure virtuellement toutes les terminaisons possibles. Ex : certificat* signifie qu'on s'intéresse tout aussi bien au mot certificat qu'aux mots certificats, certification, etc.

Les moteurs de recherche :

le langage de requête - exemples

- Exclure un **mot ou URL** :
 - la requête *japon -wikipedia.fr* cherchera les sites parlant du Japon en excluant les articles du site Wikipédia France.
- Chercher une expression précise avec les **guillemets** :
 - la requête « *I have a dream* » cherchera les sites comportant l'expression « I have a dream » dans son ensemble.
- Proposer une liste de mots séparés avec **OU** ou **OR** :
 - la requête *jeux olympiques 2016 OU 2021* cherchera les sites mentionnant d'une part les jeux olympiques de 2016 et d'autre part les sites mentionnant les jeux olympiques de 2021.
- Lier des concepts avec **ET** ou **AND** :
 - la requête *sociologie ET médias* cherchera les sites mentionnant la sociologie des médias de façon plus spécifique.
- Regrouper des opérateurs booléens :
 - la requête « *révolution française* » **OR** « *French révolution* » - *wikipedia.fr* recherchera d'une part les sites comportant l'expression « révolution française », d'autre part les sites comportant l'expression « French révolution » en excluant les articles des sites Wikipédia .fr ainsi que .org

Recherche d'information

Les moteurs de recherche : formaliser les requêtes

- Conseils avant de lancer une requête :
 - être **précis** ;
 - utiliser des **mots concrets** ;
 - mettre les mots les plus importants en tête ;
 - pour être sûr qu'un mot soit cherché mettre + devant ;
 - utiliser au moins 3 mots-clés dans la requête ;
 - quand c'est possible, rechercher des phrases ;
 - éviter les mots trop communs, sauf dans des phrases ;
 - ex : eau dans « eau gazeuse » ;
 - penser à tous les mots probables dans le corps de la page.

Recherche d'information

Moteurs de recherche: Filtrer les pages

- Lors d'une recherche internet, des millions de pages et de résultats sont triés par les différents moteurs de recherche.
- Afin de faciliter une recherche lorsque certains résultats précis sont requis, il est possible de **filtrer ces pages**.
- Les filtres de recherche sont de différents ordres : on peut borner des **dates**, se limiter à certaines **sources**, ou encore certaines **langues**.

Recherche d'information

Moteurs de recherche: Référencement d'une page

- Afin de modifier une page/un site et toutes les informations le concernant, ce dernier dispose d'un **code source**.
- Le code source correspond à l'ensemble des informations et consignes d'un **programme informatique**.
- Il s'agit généralement d'un **ensemble de mots** ou de caractères **ponctués**, formant des instructions à mettre en œuvre par le programme, dans un langage de programmation spécifique. Il est possible, via ce code source et la fonction « **noindex** », d'empêcher le référencement d'une page.

Recherche d'information

Moteurs de recherche: Masquer une page

- Outil très puissant, le moteur de recherche peut éventuellement permettre de **masquer** une page.
- Cela signifie qu'en cas de recherche, elle n'apparaîtra pas avec l'ensemble des **résultats**.
- Grâce à la **fonction « robots.txt »**, il est possible d'indiquer sur son site les pages à ne pas divulguer via les moteurs de recherche (ex : les paniers d'achats sur certains sites, propres à chaque individu). Les robots seront alors incapables de parcourir le site et de l'indexer.

Recherche d'information

Les alternatives aux moteurs de recherche

- Les premiers résultats que Google renvoie tiennent fortement compte de la popularité de la page web dans les réseaux sociaux et dans les autres pages du web. Selon le sujet auquel on s'intéresse, ce facteur n'est pas toujours intéressant et son importance risque de masquer de très bonnes pages qui, étant moins populaires, apparaîtront sur la 10ème ou 20ème page de résultats, alors même qu'elles sont plus pertinentes pour notre besoin.
- C'est pourquoi il est essentiel de connaître les alternatives aux moteurs, qui présentent des atouts que les moteurs n'ont pas :
 - ils sont extrêmement **sélectifs** (on n'est pas noyé dans des milliers de pages de résultats) ;
 - ils favorisent le **critère du contenu** (nos mots clés !) ;
 - ils s'appuient sur une autre forme d'intelligence : celle des **experts humains** !

Sommaire

- Les sites web
- Recherche d'informations : moteurs de recherche et requêtes
- **Evaluation de l'information**
- Source et citation
- Veille d'information, flux et curation

Evaluation de l'information: résultats d'une recherche

- Typologie des sites internet
- Pourquoi évaluer les résultats d'une recherche ?
- Les critères de qualité des résultats :
 - Qui ? Quoi ? Où ? Quand ? Comment ? Pourquoi ?
- Les indices de fiabilité : exemples concrets

Evaluation de l'information: Pourquoi évaluer les résultats d'une recherche ?

- Avec :
 - 40 millions d'internautes en France ;
 - 77 % des français membres d'un réseau social ;
 - 100 heures de vidéos téléchargées sur YouTube chaque minute ;
 - 64 % des internautes qui partagent leur opinion sur Internet ;
 - Plus de 22 millions d'articles proposé par Wikipédia (ce qui, transformé en livre, représenterait 2,25 millions de pages et prendrait 123 ans à lire) :
 - Ce n'est pas l'information qui manque sur la toile !
- Dans une recherche sur Internet, nombreuses sont les personnes qui ont tendance, entre autres, à :
 - n'utiliser qu'un seul outil de recherche (et dans 90 % des cas Google) ;
 - ne rechercher qu'avec très peu de mots clés (deux en moyenne) ;
 - se limiter à l'interface de recherche simple ;
 - restreindre la lecture des résultats aux toutes premières suggestions ;
 - et ceci les conduit à des résultats à la fois incertains et non probants !
 - Source : Jacques PERRIAULT « L'accès au savoir en ligne »

Evaluation de l'information: Pourquoi évaluer les résultats d'une recherche ? (suite)

- N'importe qui peut maintenant publier n'importe quoi sur le Web : la limite n'est pas technique mais seulement parfois celle de la légalité.
- Les premiers résultats d'une recherche **sont rarement les meilleurs**.
- Les informations disponibles sur les pages web sont rédigées **sous la responsabilité de leurs auteurs**, mais **sans garantie** pour le lecteur.
- Le contenu d'une page web n'est **pas toujours exact et exhaustif**.
- Les documents doivent donc être évalués en prenant compte :
 - de la pertinence : le document répond-il à la question que vous vous posez ?
 - de la fiabilité.

Evaluation de l'information: Typologie des sites internet

- **Sites institutionnels** : gouvernement, éducation, université, administratif, etc.
 - <http://impots.gouv.fr>
- **Sites commerciaux** : informations, vente en ligne, etc.
 - <http://www.laredoute.fr> <http://www.lemonde.fr>
- **Sites collaboratifs** : forums, wikis, etc. qui contiennent des contributions d'internautes qui ne sont pas forcément des experts du domaine
- **Sites personnels** : pages perso, blogs, etc.
 - leur contenu n'est soumis à aucun contrôle.

Evaluation de l'information: Les critères de qualité des résultats : qui ?

- L'auteur d'une ressource est le **1^{er} responsable** de celle-ci.

Questions à se poser	Indices
L'auteur du document est-il clairement identifié ?	Prénom NOM de l'auteur
L'auteur peut-il être contacté ?	Adresse email, téléphone
Quelles sont les compétences de l'auteur ?	Fonction, titres, appartenance à une institution, etc.
Qui est l' éditeur de la ressource : organisme ou webmestre responsable du site ?	Vérifier la présence du nom de l'organisme dans l'URL Vérifier le nom du responsable du site dans les mentions légales

Vous trouvez le nom du directeur de la publication, du responsable éditorial et de l'hébergeur dans les **mention légales du site**.

Evaluation de l'information: Les critères de qualité des résultats : qui est l'auteur de la ressource ?

- L'auteur d'une information est le premier responsable de celle-ci. Il est donc important de pouvoir l'identifier et déterminer ses compétences.
- **Les questions à se poser :**
 - l'auteur du document est-il clairement identifié ?
 - dans le cas d'un article publié dans une revue en ligne, le nom de l'auteur (ou des auteurs) est clairement indiqué dans le sommaire des articles de la revue ainsi qu'en tête de l'article lui-même ;
 - dans le cas d'un document faisant partie du site d'un organisme officiel ou commercial, le nom de l'auteur ou du rédacteur n'est pas toujours mentionné ;
 - dans le cas d'un site personnel, le seul renseignement mentionné est parfois une adresse de courriel ;
 - quelles sont les compétences de l'auteur ou des auteurs ? Sont-elles clairement énoncées ?
 - l'auteur fait-il référence à des publications antérieures, parues dans des périodiques imprimés ou publiés sur le Web et faisant autorité ?

Evaluation de l'information: Les critères de qualité des résultats :qui est l'auteur et l'éditeur de la ressource ?

- Pour trouver l'auteur, s'il n'est pas clairement identifié :
 - dans **Firefox**, allez dans **Outils**, puis **Développement web : Code source de la page**.
 - affiche le code source HTML de la page consultée ;
 - permet de repérer éventuellement les balises de métadonnées, notamment la balise Author ;
 - ensuite, une simple recherche sur le Web sur le nom de l'auteur vous donne des renseignements sur ses compétences sur le sujet.

Evaluation de l'information: Les critères de qualité des résultats : qui est l'auteur de la ressource ?

- Y a-t il un moyen de vérifier le sérieux d'une organisation si elle n'est pas connue ?
 - vous pouvez considérer que, si un organisme ou une source qui vous est inconnue est citée par plusieurs autres sources que vous savez être sérieuses, la source inconnue sera certainement sérieuse elle aussi ;
 - en dehors des sites personnels, l'éditeur d'un document ou d'un site Web est l'organisme ou le webmestre responsable du site. Le nom du serveur peut vous donner une indication sur la nature de l'organisme éditeur du site. Ce nom se trouve habituellement au début de l'adresse URL, entre http:// et le premier /.

Evaluation de l'information: Les critères de qualité des résultats : quoi ?

Questions à se poser	Indices
Quel est le thème annoncé dans le résumé du site ?	Vérifier la cohérence du résumé affiché dans les résultats du moteur
Quelle est la pertinence de l'information ? - L'information fournie est-elle unique ? - Le traitement du sujet est-il simple ou approfondi ?	Regarder si elle est présente dans d'autres sites, si elle est identique, s'il existe des liens entre ces sites. Si le sujet est approfondi, il y a des chances pour que l'information soit plus fiable

- **La pertinence d'une source s'évalue en premier lieu en fonction de vos besoins :**
 - vérifiez que les documents trouvés traitent bien du sujet qui vous intéresse. Pour cela, regardez le titre et les mots-clés ;
 - vérifiez que le sujet est bien traité sous l'angle qui vous convient. Pour cela, consultez le résumé, le chapeau, la table des matières, l'introduction et la conclusion etc.

Evaluation de l'information: Les critères de qualité des résultats : quoi ?

- Elle s'évalue aussi sur la pertinence par rapport au contexte :
 - le contexte s'entend comme le but dans lequel vous effectuez une recherche d'information :
 - Exemple : vous recherchez de l'information sur un pays :
 - s'il s'agit de préparer vos prochaines vacances, vous avez besoin d'informations pratiques et culturelles récentes : vous consultez des guides touristiques, les sites des offices de tourisme, des forums, etc ;
 - si vous devez préparer un exposé, peu vous importe les adresses des hôtels ou des meilleurs restaurants : vous recherchez plutôt des informations historiques et vous consultez des ouvrages, des articles de presse, des encyclopédies, etc.

Evaluation de l'information: Les critères de qualité des résultats : quoi ?

- La pertinence s'également en fonction de votre niveau de compétence dans le domaine :
 - vous devez tenir compte de votre niveau de compréhension, tant du point de vue du langage que des prérequis ;
 - exemple : un article médical à destination des médecins peut s'avérer incompréhensible pour un néophyte qui veut se renseigner sur une pathologie.

Evaluation de l'information: Les critères de qualité des résultats : quoi ?

- La recherche d'une information à l'aide d'un moteur de recherche ou dans un annuaire peut amener un nombre élevé de résultats :
 - lors de la phase de “tri” de ces résultats, vous devez parcourir les descriptifs et résumés du contenu des sites et documents trouvés en vous posant des questions vous permettant d'éliminer les résultats inintéressants pour votre travail de recherche.
- Question lors de la lecture des résumés :
 - quel est le sujet ou le thème général annoncé dans le résumé ?
Ce sujet correspond-il bien à votre thème de recherche ?
- Questions à propos d'un site :
 - l'information fournie est-elle unique ?
 - existe-t-elle sur plusieurs sites ? Existe-t-il des liens entre ces sites ?
 - à quel public s'adresse-t-on (spécialiste, général, initié, ...) ?
 - le traitement du sujet est-il exhaustif ?

Evaluation de l'information: Les critères de qualité des résultats : où ?

- En règle générale, l'URL d'un site est intéressante, car elle apporte de précieux renseignements :

Questions à se poser	Indices
Quelle est la nature du site : site institutionnel, commercial, collaboratif, personnel ?	Examiner le nom de domaine et son extension : gouv.fr, gov, com, edu, ...
De quelle zone géographique provient l'information : site français, francophone, européen, autre ?	Examiner l'extension du nom de domaine : fr, be, au, ca, eu, ...
Qui est le propriétaire du site ? Par quel serveur est-il hébergé ?	Taper le nom de domaine du site dans www.whois.net ou recherchez le dans Afnic (français) ou ICANN (international)._
Le site pointe-t-il vers des sites fiables ? Et inversement ?	Examiner les liens sortants et entrants

Evaluation de l'information: Les critères de qualité des résultats: où ?

- **Structure** d'une adresse internet (ou URL) :
 - exemple : `http://www.pantheonsorbonne.fr/pix/` ;
 - **http://** est l'indicatif qui signale le protocole utilisé : ici, HyperText Transfert Protocol ;
 - **Pantheonsorbonne** est la dénomination de l'hébergeur (le service propriétaire du serveur qui héberge le site) et aussi le nom du site ;
 - **.fr** est le code pays ou l'indicatif de la nationalité de l'hébergeur :
 - Attention : les codes-pays ne désignent pas la nationalité du site, mais la nationalité de l'hébergeur ;
 - si aucune page n'apparaît, c'est la page par défaut, `index.htm` ou `index.html` qui sera affiché :
 - effacez dans l'adresse le nom de la page affichée, et aussi les sous-répertoires du site, puis lancer le navigateur. De la sorte, vous pouvez consulter la page d'index du site ;
 - si cette opération ne donne pas de résultat, le site n'est pas bien fait.

Evaluation de l'information: Les critères de qualité des résultats : quand ?

- Dans certains domaines, la date des documents trouvés ainsi que la date de dernière mise à jour des sites consultés sont des informations à prendre en considération.

Questions à se poser	Indices
Quelle est la date de la ressource ? - Est-elle clairement indiquée ? - La ressource a-t-elle besoin d'être actualisée ?	Si la date n'est pas indiquée ou si elle est ancienne, la ressource est peut-être obsolète
Les liens vers d'autres sites sont-ils obsolètes ou morts ?	Une page web non mise à jour peut comporter des liens vers des pages qui ont disparu

Evaluation de l'information: Les critères de qualité des résultats : quand ?

- Un des avantages d'Internet est de fournir des informations presque en temps réel. L'inconvénient est le manque du recul nécessaire pour confronter l'information trouvée avec d'autres sources. Dans le domaine des lettres et de la littérature, les documents les plus récents ne sont pas nécessairement les plus pertinents. En revanche, dans les domaines de la civilisation (contemporaine), de la linguistique et de la didactique, la date des documents trouvés ainsi que la date de dernière mise à jour des sites consultés sont importantes à prendre en considération.
- **Questions:**
 - quelle est la date de création du site? Quelle est celle de sa dernière mise à jour ? Cette date est-elle affichée dans la page ?
 - les liens vers d'autres sites sont-ils à jour ? Existe-t-il beaucoup de liens brisés ? Dans un site Web régulièrement mis à jour, le responsable du site vérifie périodiquement la validité des liens vers d'autres sites. Si vous constatez la présence de nombreux liens brisés sur un site, ce peut être un indice que le responsable du site n'est pas sérieux ou que le site n'est plus mis à jour.

Evaluation de l'information: Les critères de qualité des résultats : pourquoi ?

- On s'intéresse ici aux **motivations de l'auteur** d'un site Web. Mettre des informations sur le Web coûte du temps et de l'énergie ; on suppose donc que l'on en retire un bénéfice.

Questions à se poser	Indices
Quels sont les objectifs du site ? Dans quel but cette ressource a-t-elle été réalisée ?	Les motivations diffèrent suivant que le site est un site institutionnel, commercial, personnel, etc.
L'information est-elle gratuite ? Y a-t-il des publicités sur le site ?	Les ressources peuvent être en vente...

Evaluation de l'information: Les critères de qualité des résultats : pourquoi ?

- Quelles sont les motivations de l'auteur d'un site Web?
 - En ce qui concerne les sites personnels, les motivations peuvent être :
 - la satisfaction d'un ego, le contributeur sachant qu'il sera lu ;
 - le plaisir altruiste de rendre service à la collectivité ;
 - la volonté prosélyte de diffuser des idées.
 - Les sites des organismes officiels ont une vocation naturelle à diffuser de l'information et leur présence sur le Web n'est que le prolongement naturel de leurs canaux d'information traditionnels (papier, médias).
 - Il faudra être particulièrement méfiant envers les sites commerciaux et plus précisément envers ceux qui n'annoncent pas clairement la couleur. Autant il est évident que les sites marchands de la FNAC, de Microsoft ou de Amazon affichent clairement leur but de vous vendre leur marchandise, autant de nombreux sites se présentent comme un service, une association, et déguisent habilement leur intention de vendre.

Evaluation de l'information: Les critères de qualité des résultats : comment ?

- Bien que l'ergonomie et l'esthétique d'un site ne soient pas des indicateurs de la validité de leur contenu, méfiez-vous des sites brouillons ou mal organisés : une présentation peu soignée est souvent la marque d'un contenu peu valable.

Questions à se poser	Indices
Comment se présente la ressource ?	Clarté de l'information, structuration du contenu, citation des sources
Comment accède-t-on à l'information ?	Navigation aisée, ergonomie du site agréable, liens hypertexte cohérents
Le type de langage employé est-il familier ou scientifique ?	Clarté de l'information, références pertinentes, liens hypertextes cohérents

Evaluation de l'information: Les critères de qualité des résultats : comment ?

Questions

- les ressources sont-elles bien organisées et présentées de façon logique ?
- s'il s'agit d'un site multimédia, quelle est la qualité visuelle et sonore des documents proposés ?
- les liens sont-ils complets, pertinents et appropriés? Les liens sont un peu comparables aux références dans les travaux scientifiques et universitaires. L'auteur d'un site ou d'un document Web qui vous invite à visiter d'autres sites ou documents semblables aux siens confirme la validité de ses affirmations en citant ses sources ;
- sont-ils évalués ou commentés ?
- sont-ils relativement exhaustifs sur le sujet, ou servent-ils d'illustration ?
- ont-ils tous un rapport avec le sujet traité ?
- le type de langage employé est-il familier ou scientifique ?
- dans le cadre d'un travail de recherche universitaire, vous devez vous assurer que le document trouvé sur le Web offre toutes les garanties d'un article de type scientifique : article publié dans une revue en ligne, extrait d'ouvrage (thèse, etc.), extrait de cours d'un universitaire ;
- les documents sont-ils rédigés dans une langue correcte (orthographe et grammaire) ? La correction de la langue d'un document trouvé sur le Web sera un élément déterminant quant à sa validité. Un texte rédigé dans une langue approximative et incorrecte ne peut évidemment pas être cité tel quel. En outre, cette mauvaise qualité de langage est très probablement l'indice d'une mauvaise qualité du contenu du document lui-même ;
- le site propose-t-il un corpus cohérent et complet, ou bien est-il "en construction" ?
- un site peut être "en construction" et présenter néanmoins des documents tout à fait valables. Il faut juger les documents présents sur le site au cas par cas.

Evaluation de l'information: Les indices de fiabilité exemple : site Ministère de l'éducation nationale

Logo du ministère

The screenshot shows the homepage of education.gouv.fr. At the top, there is a navigation bar with links for PROFESSIONNELS, PARENTS, PRESSE, VIE LYCÉENNE, ENSEIGNEMENT SUPÉRIEUR ET RECHERCHE, and ENGLISH. Below this is the site logo and a search bar with the text 'education.gouv.fr' and a search input field containing 'Entrez votre recherche ici'. A secondary navigation bar lists various educational levels and topics like LE SYSTÈME ÉDUCATIF, ÉCOLE, COLLÈGE, LYCÉE, etc. A main banner features a large blue and yellow graphic with the text 'Téléchargez les horaires de l'école de votre enfant' and 'Horaires Rentrée 2016'. A callout box points to this banner with the text 'Les informations sont datées et récentes'. Below the banner, there are several news snippets under the heading 'L'ACTU+'. One snippet is titled 'L'aide à la recherche du premier emploi (ARPE)' and another mentions 'Najat Vallaud-Belkacem annonce l'extension du Pass Éducation'. On the right side, there is a section 'C'EST OFFICIEL' with links to 'CALENDRIER SCOLAIRE' and 'LE BULLETIN OFFICIEL DE L'ÉDUCATION NATIONALE'.

Les informations sont datées et récentes

Evaluation de l'information: Les indices de fiabilité : exemple Encyclopédie Universalis



Il s'agit une encyclopédie française de référence. Les articles sont rédigés par des spécialistes.

L'article est bien rédigé et structuré

On connaît les auteurs de l'article et leurs fonctions

DSIUN – SERVICE DES USAGES NUMÉRIQUES

Evaluation de l'information: Les indices de fiabilité : exemple Wikipédia

The image shows a screenshot of the French Wikipedia page for 'Informatique' in a Mozilla Firefox browser. The page title is 'Informatique' and the URL is 'http://fr.wikipedia.org/wiki/Informatique'. The page content includes a definition of 'Informatique' as a domain of scientific, technical, and industrial activity, and a quote by Edsger Dijkstra: '« La science informatique n'est pas plus la science des ordinateurs que la... »'. A callout bubble on the left states: 'Wikipédia est une encyclopédie collaborative en ligne : tout le monde peut collaborer ...'. A callout bubble on the right, pointing to the 'Afficher l'historique' button, states: 'En affichant l'historique, on peut voir les contributions successives mais impossible de connaître les compétences des contributeurs !'. Below this, a list of recent contributions is visible, showing dates and usernames like 'Bublelegun' and 'EmausBot'.

Evaluation de l'information: Le cas de Wikipedia

- Wikipédia est une encyclopédie universelle géré par wiki.
- Elle est en libre accès, en lecture comme en écriture
 - n'importe qui peut modifier les articles.
- Dans l'historique, vous pouvez voir qui a modifié l'article, à quelle date et quelle action, il a effectué.
- Il est possible de comparer avec la version précédente afin de voir les modifications (« diff »).
- On considère que plus l'article a été modifié, plus il est fiable.

The screenshot shows the Wikipedia history page for the article 'Guitare'. The URL is <https://fr.wikipedia.org/w/index.php?title=Guitare&action=history>. The page title is 'Historique des versions de « Guitare »'. The page includes a search bar, a navigation menu, and a list of versions. The 'Voir l'historique' button is highlighted with a red box. The 'Filtrer les versions' section is also highlighted with a red box. The 'Comparer les versions sélectionnées' section is highlighted with a red box, showing a comparison between the current version and a previous version from June 6, 2021.

WIKIPÉDIA
L'encyclopédie libre

Rechercher dans Wikipédia

Non connecté Discussion Contributions Créer un compte Se connecter

Article Discussion Lire Voir le texte source Voir l'historique Aide

Historique des versions de « Guitare »

Voir les opérations sur cette page (voir le journal des filtrages)

▼ Filtrer les versions

Outils externes et statistiques
Auteurs et statistiques - Rechercher l'auteur d'un passage de l'article - Statistiques de consultation - contributeurs suivant cette page - Modifications par utilisateur

Autres discussions [liste]
Suppression - Neutralité - Droit d'auteur - Article de qualité - Bon article - Lumière sur - À faire - Archives

Légende : (actu) = différence avec la version actuelle - (diff) = différence avec la version précédente - m = modification mineure

(les plus récentes | les plus anciennes) Voir (50 plus récentes | 50 plus anciennes) (20 | 50 | 100 | 250 | 500)

Comparer les versions sélectionnées

- (actu | diff) 6 juin 2021 à 12:12 Thierry613 (discuter | contributions) . . (57 764 octets) (-11) . . (—Accords modaux) (Baises : Modification par mobile, Modification par le web mobile)
- (actu | diff) 28 mai 2021 à 12:13 Bob08 (discuter | contributions) m . . (57 775 octets) (+12) . . (—Liens externes)

Evaluation de l'information: Coordonnées GPS d'une photo

- Lorsqu'une photo est ajoutée à un document (word, powerpoint, pdf), il est possible d'obtenir de nombreuses informations (les **métadonnées**) via les propriétés de cette dernière.
- Ces données sont **enregistrées** par l'appareil qu'il s'agisse d'un smartphone, d'une tablette ou d'une caméra.
- Il est possible d'obtenir la **localisation** exacte du lieu où la photo a été prise grâce à la présence des coordonnées GPS.
- Ces métadonnées, ici les coordonnées GPS, peuvent ensuite être convertie en un lieu grâce à divers sites internet comme par exemple Google Map.

Evaluation de l'information: Rechercher un lieu

- Rechercher un lieu à l'aide de ses coordonnées GPS :
 - Ouvrez Google map et saisissez les coordonnées GPS dans le champ de recherche situé en haut de l'écran de la forme Latitude, Longitude
- Rechercher une rue dans une ville :
 - Entrez le nom de la ville et le nom de la rue sur un moteur de recherche, vous accédez à la carte interactive de la ville.
- Remarque : le site <https://www.gps-longitude-latitude.net/adresse-de-la-longitude-et-de-la-latitude> permet de trouver facilement un lieu à l'aide de ses coordonnées GPS.

Evaluation de l'information: Recherche d'images Google

- Sur un document, il est possible de faire apparaître différentes **informations** pouvant prendre la forme d'images ou de vidéos.
- Lorsque l'on trouve une image sur internet, celle-ci peut être examinée et vérifiée à partir d'un **traçage** réalisé notamment à partir de sites comme [Google images](#), TinEye ou Youtube Dataviewer.
- Dans Google images, cliquez sur l'appareil photo et importez la photo.
- Ces techniques permettent d'effectuer une **recherche inversée** afin de trouver la provenance d'une image, ou par exemple une information sur un objet présent sur l'image.

Evaluation de l'information: Données EXIF d'une photo

- Les données EXIF d'une photo correspondent aux « Exchangeable Image File ». Ce sont des **métadonnées** rassemblant un grand nombre d'informations sur une image.
- Ces données peuvent concerner le **lieu** de la prise de photo, **l'heure**, la taille, ou encore **l'auteur**. Certains appareils permettent aussi d'afficher le type d'appareil ou encore ses paramètres.
- Dans la majorité des cas, ces informations sont recensées dans les **propriétés** de l'image.
 - Clic droit sur l'image/propriétés
 - Ou sur Windows, ouvrez l'image et clic droit/informations sur le fichier

Sommaire

- Recherche d'informations : moteurs de recherche et requêtes
- Evaluation de l'information
- **Source et citation**
- Veille d'information, flux et curation

Source et citation, La récupération sur le Web : les aspects juridiques

- S'assurer en consultant les **mentions légales** du site :
 - si on est autorisé à **utiliser texte, images, fichiers**, etc. pour un usage privé et/ou à des fins d'enseignement, de recherche, commerciales, etc.
 - si on peut **télécharger les ressources** figurant sur le site
ou s'il convient simplement d'effectuer un lien vers la page.
 - si, par exemple dans le cas d'une vidéo, il est autorisé de la **partager** (lien simplifié, code HTML iframe, etc.).
- Attention au **plagiat** !

Source et citation, ses sources numériques : règles élémentaires

- Respecter le **droit des auteurs** des ressources :
 - en respectant les normes en vigueur en matière de citations.
- **Se prémunir contre le caractère éphémère** des ressources du web :
 - en faisant une copie de la ressource (récupération de la ressource, copie d'écran, ...) ;
 - en précisant la date de consultation de la ressource.
- Contribuer à la **reconnaissance** de la ressource :
 - les travaux scientifiques publiés sur le web sont évalués par le nombre de références qui pointent vers le document numérique.

Source et citation : Citer ses sources numériques

- **Site Web :**
 - auteur. Titre [en ligne]. Disponible sur : <URL> (consulté le + date)
 - indiquez l'URL précise, pas la page d'accueil du site.
- **Page Web :**
 - auteur ou organisme. Titre de la page du document. In : Titre de la page d'accueil du site [en ligne]. Disponible sur : <URL> (consulté le + date).
- **Article d'encyclopédie en ligne :**
 - auteur. Titre de l'article. In : Titre de l'Encyclopédie [en ligne]. Disponible sur : <URL> (consulté le + date).

Source et citation : Citer ses sources numériques

- **Article de périodique :**
 - auteur. Titre de l'article. In : *Titre du périodique [en ligne]. Année de publication*, volume, numéro, pagination. Disponible sur : <URL> (consulté le + date).
- **Texte réglementaire :**
 - titre complet du texte réglementaire. In Intitulé du document officiel (ex : Journal officiel, Bulletin officiel...). Disponible sur <URL> (consulté le + date).
- **Dessin/ photo :**
 - auteur. Titre de l'image. In : Titre du site [en ligne]. Disponible sur : « URL » (consulté le + date).

Sommaire

- Recherche d'informations : moteurs de recherche et requêtes
- Evaluation de l'information
- Source et citation
- Veille d'information, flux et curation

Veille d'information, flux et curation: La veille informationnelle

- Que signifie le terme « **veille** » :
 - c'est le fait de se tenir au courant des informations pertinentes à un domaine défini ;
 - il s'agit d'une activité continue (à la différence de la recherche d'informations).
- Il existe plusieurs types de veille :
 - stratégique, économique, juridique, technologique, ludique, etc.
- Elle peut se heurter à 2 risques :
 - un risque de « **bruit** » : surinformation, dispersion
 - un risque de « **silence** » : sites moins dynamiques, moins réactifs.

Veille d'information, flux et curation: Construire sa veille informationnelle

- Poser ses **objectifs** de « veille »:
 - thématique, cibles.
- Identifier les sources pertinentes.
- Choisir ses **outils de veille** :
 - En fonction de leur efficacité et des sources :
 - listes de diffusion et lettres d'information ;
 - alertes ;
 - notificateurs de mises à jour de pages web ;
 - lecteurs / agrégateurs de flux de syndication.
- **Mémoriser** sa veille.

Veille d'information, flux et curation: Construire sa veille informationnelle

- Il existe 2 techniques :
 - La **méthode Pull** :
 - elle consiste à faire des requêtes manuelles sur les moteurs de recherche ;
 - c'est une méthode chronophage et lourde ;
 - elle ne nécessite pas d'abonnement ;
 - elle a l'avantage de ramener plus de résultats pertinents (moins de bruit).
 - La **méthode Push** :
 - Le processus de recherche est automatisé :
 - la veille est régulière car automatisée. On gagne du temps. Cependant mal paramétrée, elle peut s'avérer non pertinente ou trop bruyante ;
 - elle nécessite des outils de recherche :
 - » Alertes;
 - » Listes de diffusion;
 - » Fils RSS.

Veille d'information, flux et curation

Les techniques de veille : listes de diffusion et lettres d'information

- Pour recevoir une lettre d'information (newsletter), il faut être abonné à une **liste de diffusion** :
 - liste de diffusion : liste d'adresses électroniques établie dans le but de faire du publipostage (envoi en masse) ;
 - elle est gérée de façon automatique par un logiciel spécifique, appelé gestionnaire de listes ;
 - l'inscription à une liste de diffusion peut être automatique ou volontaire ;
 - il est possible de se désabonner à tout moment.

Veille d'information, flux et curation

Les outils de veille : la syndication (RSS)

- Syndication **RSS (Really Simple Syndication 2.0)** :
 - c'est un procédé qui consiste à publier partiellement ou en totalité le contenu d'un site web sur d'autres sites ;
 - cela permet d'amener sur un site de référence de l'information glanée à différents endroits du web et ainsi l'utilisateur peut prendre connaissance des mises à jour en un lieu unique.
- Elle se fait par l'abonnement à un « **flux de syndication** » proposé par le site d'origine du contenu.



Veille d'information, flux et curation

Mémoriser et partager sa veille

- le **social bookmarking** permet a minima de gérer vos signets (les références de sites web dont vous souhaitez conserver la trace).
- Ces signets peuvent être commentés et « taggués » à l'aide de mots clés.
- Ce sont aussi des réseaux sociaux utilisables pour la veille collaborative puisqu'ils vous permettent de partager votre veille et de vous abonner à des personnes intéressées par les mêmes sujets que vous ou à des hashtags.
 - exemple : Delicious, Diigo, Scoop.it, Pearltrees

DSIUN – Service des usages numériques

- **Conception – Réalisation**

- Catherine Loire – Alexa Gallo – Lalya Yatera – Mélanie Mauvoisin - Service des usages numériques - Université Paris 1 Panthéon-Sorbonne
- 27/08/22 Version 3 Valérie Monfort, UFR 27, université de Paris 1 Panthéon Sorbonne

- **Sources**

- Patricia Cavallo
- Martine Fontaine

Version du support : 3.0
Juillet 2021



[Licence Creative Commons :](#)
[Paternité - Pas d'Utilisation Commerciale](#)
[Partage des Conditions Initiales à l'Identique](#)



UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE
