
Cours 11 : Rechercher sur internet : méthodologie et outils



CHERCHER L'INFORMATION / TROUVER LES SOURCES

1. Introduction

La recherche d'information est le domaine qui étudie la manière de retrouver des informations dans un corpus. Celui-ci est composé de documents d'une ou plusieurs bases de données, qui sont décrits par un contenu ou les métadonnées associées. Les bases de données peuvent être relationnelles ou non structurées, telles celles mises en

réseau par des liens hypertexte comme dans le World Wide Web, l'internet et les intranets. Le contenu des documents peut être du texte, des sons, ses images ou des données.

La recherche d'information est historiquement liée aux sciences de l'information et à la bibliothéconomie qui visent à représenter des documents dans le but d'en récupérer des informations, au moyen de la construction d'index.

L'informatique a permis le développement d'outils pour traiter l'information et établir la représentation des documents au moment de leur indexation, ainsi que pour rechercher l'information. La recherche d'information est aujourd'hui un champ pluridisciplinaire, intéressant même les sciences cognitives. La recherche d'information sur le web à l'aide d'un moteur de recherche est une technique de l'information et de la communication, désormais massivement adoptée par les usagers.

L'Internet n'est pas une bibliothèque ordinaire :

L'Internet est un peu comme une gigantesque bibliothèque publique où l'on peut consulter tous les documents.

Attention

Mais il existe quelques différences entre une bibliothèque et l'Internet

1. La quantité d'informations disponibles est bien plus grande sur l'Internet que dans n'importe quelle bibliothèque.

-
2. Il n'y a pas de bibliothécaire et il n'existe pas de catalogue.
 3. Les documents ne sont pas rangés: lorsqu'un document arrive, il est simplement ajouté à l'ensemble des documents existants
 4. Tout le monde peut venir déposer un document ou enlever un document qu'il avait déposé.

Les ressources documentaires :

La recherche d'information est de retrouver des informations dans une bibliothèque sous forme de papier, ouvrages, périodiques papier encyclopédies dictionnaires ou bien dans l'internet en accédant à des bases de données, et des moteurs de recherche.

Base de données :

Une base de données (en anglais : database) est un conteneur permettant de stocker et de retrouver l'intégralité des informations en rapport avec une activité ; celles-ci peuvent être de natures différentes et reliées entre elles. Dans la très grande majorité des cas, les informations contenues dans la base de données sont très fortement structurées, la base est usuellement localisée dans un même lieu et sur un même support.

Ce dernier est généralement informatisé. Une base de données est la pièce centrale des dispositifs informatiques qui servent à la collecte, le stockage, le travail et l'utilisation d'informations.

Le dispositif comporte un système de gestion de base de données (abr. SGBD) : un logiciel moteur qui manipule la base de données et dirige l'accès à son contenu. De tels dispositifs — souvent appelés base de données — comportent également des

logiciels applicatifs, et un ensemble de règles relatives à l'accès et l'utilisation des informations. Moteur de recherche Un moteur de recherche est une application web permettant de retrouver des ressources (**pages web, articles de forums Usenet, images, vidéo, fichiers, etc.**) associées à des mots quelconques.

Certains sites web offrent un moteur de recherche comme principale fonctionnalité ; on appelle alors moteur de recherche le site lui-même (Google Vidéo par exemple est un moteur de recherche vidéo). Le moteur de recherche Google, qui a donné le nom à la société Google, est le moteur de recherche sur le Web le plus utilisé au monde. En 2009, 67 % des internautes l'utilisent.

La recherche d'informations sur l'Internet pose au moins trois problèmes :

- 1.La quantité d'informations disponibles est colossale
- 2.On ne sait pas sur quel site aller pour trouver une information intéressante
- 3.N'importe qui peut venir déposer des informations sur l'Internet; même des informations fausses Dans une bibliothèque classique, on peut faire appel au bibliothécaire.

Mais personne n'est chargé du rôle de bibliothécaire sur l'Internet.

Un moteur de recherche est un programme d'ordinateur capable de trouver des informations parmi un grand nombre de documents de différents types. Un certain nombre de sites web proposent un moteur de recherche. Parmi les moteurs de recherche les plus connus.

Le nom Google vient du mot Gogol, nom donné au nombre 10100. Ce nombre a été choisi pour évoquer la capacité de Google à traiter une très grande quantité de données. La construction de la stratégie de recherche, c'est la quatrième étape de cette

méthodologie, s'intéresse aux opérateurs logiques ou booléens : AND (et, tous les mots), OR (ou, au moins un de ces mots) et NOT (sauf, à l'exception de ces mots), opérateurs de proximité (pour

affiner une recherche) en utilisant des expressions exactes, termes adjacents (ADJ, NEAR, WITH, SAME), aussi les opérateurs syntaxiques tels que les troncatures * (aliment*), guillemets « » (« politique agricole commune », parenthèses () (fruit ET (guide ou manuel)), et les opérateurs par étapes qui servent à traiter la question par étapes successives en constituant des sous-ensembles de références pertinentes : La dernière étape est l'évaluation des résultats de la recherche par les éléments suivants : Fiabilité des sources interrogées Bruit / silence de l'équation de recherche.

-
- **Nombre de réponses**

 - **Affiner la recherche (étape supplémentaire, limite thématique, discipline, revue, pays, auteur, type de documents)**
-

Il existe plusieurs outils de recherche sur le web:

les moteurs de recherche

les méta-moteurs

et les annuaires de recherche.

Premier type d'outils :

Les annuaires

Principales caractéristiques

Nous retiendrons comme premier type d'outils les annuaires– qu'on appelle également guides, répertoires ou catalogues.

Les annuaires sont des guides par sujet des ressources d'Internet. Ils consistent en des classements arborescents où l'accès au

Définition : Une requête sur un moteur de recherche est une question posée à la base de données du moteur de recherche.

thème souhaité s'effectue en parcourant une série de rubriques et de sous-rubriques de plus en plus pointues. Ils incorporent également, d'ordinaire, un moteur de recherche par mot clé qui permet d'arriver directement à la bonne rubrique.

Ces listes thématiques de sites constituent en quelque sorte l'équivalent cybernétique (et moins élaboré) du plan de classification que l'on applique traditionnellement dans les bibliothèques et centres de documentation. Elles présentent également des similitudes avec les bibliographies thématiques, info-guides et autres listes imprimées de ressources que les bibliothécaires mettent à la disposition de leur clientèle, et avec ces pages Web personnelles qui proposent en compilation les « meilleures » ressources d'Internet ou, tout simplement, les sites préférés de leur auteur.

Les différentes phases du processus :

1. Un internaute envoie une requête sur un moteur de recherche
2. Les systèmes informatiques du moteur de recherche interrogent la banque de données des serveurs d'index
3. Les serveurs d'index renvoient une série de résultats pour la requête qui a été transmise
4. Le serveur web du moteur de recherche retourne les résultats à l'internaute

Les techniques de recherche

1. Les opérateurs booléens : ET, OU, SAUF Ils permettent, en établissant une connexion logique entre des termes de recherche ou mots-clés, de trouver une information ou un document.

1-Recherche très précise : seulement sur l'histoire d'Oran pas sur la ville d'Oran ni sur d'autres histoires en utilise [Oran Et Histoire].

2-Recherche très large : par exemple • sur la ville d'Oran • sur l'histoire • sur l'histoire d'Oran en utilise [Oran OU Histoire].

3-Recherche orientée : tout sur la ville d'Oran sauf ce qui touche à son histoire en utilise [Oran sauf Histoire].

Les techniques de recherche

La recherche par expression

L'utilisation des guillemets « » permet de lancer une recherche sur une « chaîne de caractères » (mêmes mots dans le même ordre). Elle est particulièrement utile lorsqu'une recherche entraîne un trop grand nombre de résultats ou pour rechercher précisément une expression. Ex. : « vitamine C » recherchera les références contenant cette expression dans l'ordre où sont saisis les termes

4. La casse (majuscules, mots vides)

Il est recommandé pour toute recherche sur une base de données ou un catalogue de bibliothèque de n'utiliser que des lettres minuscules non accentuées. Les mots vides sont les mots non significatifs tels que les articles, les prépositions, les pronoms, etc. A l'exception d'une recherche par expression, ils sont totalement inutiles.

5. Évaluer la qualité et la pertinence des sources

Un travail de recherche doit s'appuyer sur des informations fiables. Cela est particulièrement vrai lorsque les recherches portent sur des sites web. Toute information dont on ignore la provenance devrait a priori être écartée. les principaux critères de fiabilité des sources:

- L'auteur, l'éditeur de la ressource

-
- **La date de publication du document**

- **Le domaine de la ressource (adresse URL)**

-
- **L'objectif du site**
 - **La notoriété, l'indice de popularité du site**
 - **Le contenu de l'information (structuration, argumentation, sources, etc.)**
-

La sélection des documents doit se faire non seulement en fonction de leur qualité mais aussi de leur pertinence par rapport au travail à réaliser. Une exploration rapide peut suffire pour évaluer la pertinence d'un contenu. Pour cela, il faudra analyser les éléments suivants :

- **Titre du document** : pour un livre, il faut regarder celui figurant sur la page de titre
- **Résumé (abstract)** : on le trouve dans la plupart des notices bibliographiques tirées des bases de données, au début ou à la fin des articles de périodiques et souvent au dos des livres (quatrième de couverture).
- **Table des matières** : elle permet de mieux apprécier le contenu (plan et logique de l'argumentation) et de bien repérer les chapitres qui peuvent être pertinents.
- **Tableaux, graphiques, etc.** : ils peuvent aider à la compréhension du sujet et être utiles pour le travail.

- **Nature du document** : déterminer s'il s'agit d'un document pédagogique, de recherche ou de vulgarisation

- **Introduction et conclusion** : leur consultation permet de cerner la question de départ et les conclusions que l'auteur en tire...

6-Mettre en place une veille documentaire

Au fur et à mesure des recherches, il est recommandé de mettre en place une veille informationnelle automatisée. Elle permettra d'être alerté des nouvelles publications dans un domaine particulier sans avoir à relancer manuellement la recherche sur chaque source. L'utilisateur reçoit des messages provenant des divers outils sur lesquels il a paramétré sa veille. Selon le cas, ces messages sont reçus dans son courriel, le compte d'un agrégateur de son choix, sa page Web personnalisée, etc.

Le catalogue Koha referme les notices de la collection de la bibliothèque

- Fonctionnalités simples et avancées;
- Comprend plus de 50000 titres de livres, des documentaires et films de fiction, etc.
- Transactions à distance: en vous authentifiant, vous pouvez renouveler à distance vos emprunts de la bibliothèque, ainsi que consulter des livres numériques,

des périodiques et visionner à distance des films...Les bases de données dont: Repère;
Eureka; Cairn;etc. Internet !

Je cherche un site web :

Avantages :

- La navigation parmi la taxonomie est une stratégie de recherche;
 - La personne explore les ressources web en parcourant le contenu.
-

Inconvénients :

- Il est difficile, voire impossible parfois de trouver une information vraiment spécifique; Noms des catégories –subjectifs;
 - Problèmes avec liens morts –difficile de mettre à jour.
-

Utiles pour :

- Trouver des sites ressources;
 - Rechercher des sites généraux –recherche exploratoire Répertoires
Commerciaux:
 - Yahoo!–Yahoo! Québec;oDmoz/ Open Directory Project;
 - Free Medical Journals :
-

Répertoire par grands thèmes :

oLII –Librarian’s Internet Index; Bibliothèque et Archives nationales du

Québec Répertoires Universitaires.

Infomine (Scholarly Internet Resource Collections) Les moteurs de recherche –Index

Avantages:

- Richesse de l’interrogation, recherche en texte intégral;
 - Plus d'exhaustivité, mise à jour plus rapide;
 - **Inconvénients:**
 - Pas de sélection de l’information;
 - Hétérogénéité des sources, des formats;
 - Bruit* documentaire, pertinence variable, publicités;
 - Malgré une simplicité aux premiers abords, complexité de la syntaxe d’interrogation.
 - Fonctionnalités variables selon les moteurs (utiliser supports d’aide).
-

À utiliser pour :

- Pour une première approche;
 - Trouver une information précise sur une personne, un organisme, un produit...;
 - Effectuer des recherches complexes;
 - Repérer des sites récents
-

Google; Exalead; BING; Ask; Gigablast

Les moteurs de recherche –Index spécialisés

Google Scholar

Volunia



Google Scholar indexe les écrits universitaires (thèses, articles, livres, colloques...). Mais attention! Vous n'avez pas toujours accès au texte intégral du document que vous trouvez.

PubMed Central(PMC) (Free biomedical and life sciences journal literature);

Les métamoteurs :

Avantages :

- Gain de temps et élargissement du champ de recherche.
-

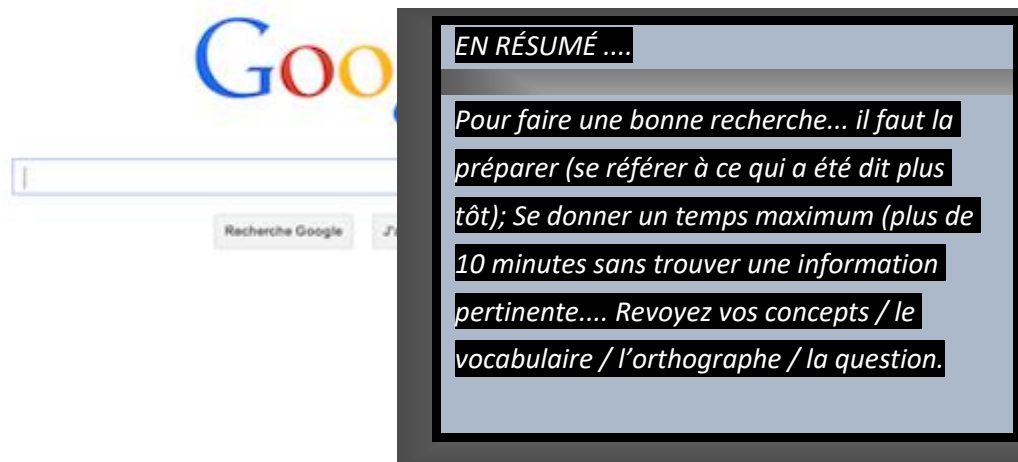
Inconvénients :

- Pas de syntaxe commune, impossible d'utiliser les fonctionnalités avancées des moteurs de recherche.
-

À utiliser pour :

-
- Débroussailler rapidement un sujet
 - Repérer les outils qui répondent le mieux à un sujet. □ Yippi; □ Zuula;
□ Ixquick; □ iZito; □ Metacrawler;.
-

Chercher l'information



Vous connaissez ? Eh oui, il s'agit du moteur de recherche le plus connu et utilisé au monde : Google. Il y en a plein d'autres (Yahoo Search, See, Bing), mais prenons celui-ci !

Contrairement à ce que pensent certains, il ne suffit pas de taper la question, le mot ou l'idée qu'on vous demande de chercher. Si on vous demande par exemple "*Quels sont les impacts du réchauffement climatique sur notre environnement et notre santé ?*", vous tapez quoi ?

Si vous tapez Réchauffement climatique, vous aurez 654 000 résultats! Google, dans ce cas, cherche toutes les pages, images, vidéos... contenant ces mots

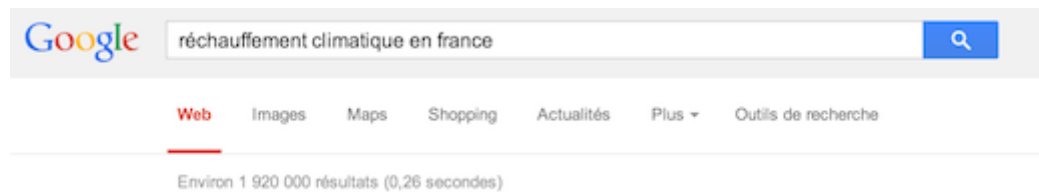
individuellement Réchauffement, Climatique, mais aussi ensemble Réchauffement climatique.

Je vais devoir regarder toutes ces pages ? Comment affiner les réponses ?

Étape 1 : Faites un remue-méninges sur la question posée.

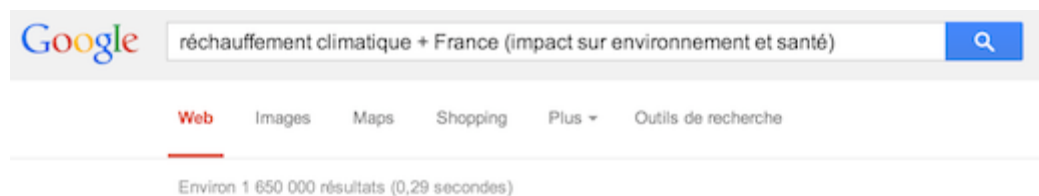
Posez-vous les 6 questions de base : **3QOCP** ? (Qui ? Quoi ? Quand ? Où ? Comment ? Pourquoi ?). Cela vous permettra de trouver de nouveaux mots-clés.

- **Recherche** : réchauffement climatique en France



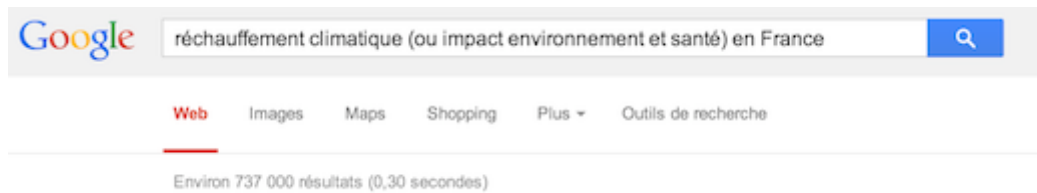
Résultats Google : 1 920 000 résultats

- **Recherche** : réchauffement climatique + France (impact sur environnement et santé)



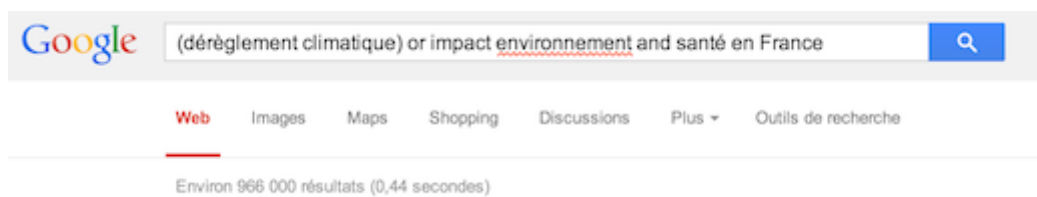
Résultats Google : 1 650 000 résultats

- **Recherche** : réchauffement climatique (ou impact environnement et santé) en France



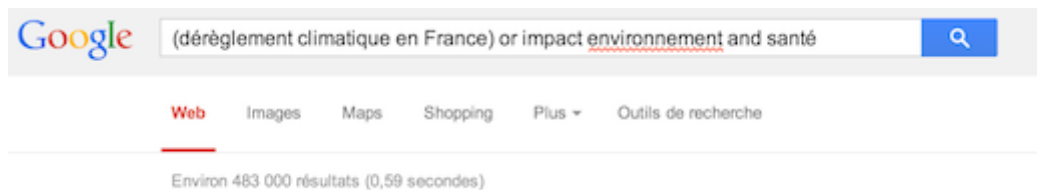
Résultats Google : 737 000 résultats

- **Recherche** : (dérèglement climatique) or impact environnement and santé en France



Résultats Google : 966 000 résultats.

- **Recherche** : (dérèglement climatique en France) or impact environnement and santé



Résultats Google : 483 000 résultats.

Etc.

Regardez à nouveau les formulations précédentes. Vous remarquez quelque chose ?

Étape 2 : Utilisez des opérateurs de recherche

Des quoi ?

Et oui, dans les formulations précédentes, vous avez sûrement remarqué les parenthèses () , les mots **et / ou / or**, les signes + / - ...

Ils servent à affiner votre recherche selon vos besoins. Il en existe beaucoup et varient selon le moteur de recherche!

Les opérateurs	Pour quoi faire?	Exemples
Entre guillemets “...”	Pour chercher la phrase telle quelle	Retrouver l’auteur d’une citation par exemple
L’étoile *	Pour chercher des mots avec la même racine	Je veux trouver tous les mots de la famille comme polluer, pollution, pollution de l’air, etc. Au lieu de taper tous ces mots, il faudra taper : pollu*
Numériques > / < / = / :	Permet de faire une recherche en précisant une période	Chercher toutes les infos sur le réchauffement

chronologique

climatique entre 1950

et 1955 :

Réchauffement

climatique 1950 :

1955

ET ou + ou **AND**

Permet de mettre en
relation deux mots

Je souhaite avoir des
informations sur le
chômage uniquement
en France :

chômage + France

chômage ET France

OU ou - ou **OR**

Permet de chercher
les deux mots
indépendamment ou
ensemble

Je souhaite des
informations sur le
chômage en général,
la France en générale
ou le chômage en
France :

chômage - France

chômage OU France

Les parenthèses ()

Elles permettent de

Je souhaite des

	combiner plusieurs signes ensemble afin de proposer une formule de recherche quasi mathématique !	informations sur la pollution, l'air, la mer, mais aussi la pollution de l'air ou la pollution de la mer : pollution ET (air OU mer)
--	---	--

TD /Activités

Exercice n°1

Pour réduire le nombre de résultats, il faut donc :

- Sélectionner les mots-clés.
- Rédiger la formule de recherche possible que vous taperiez sur Google.

-
- « Aucune réponse ne correspond à vos critères » et vous voulez plus de résultats ? Il faudra alors supprimer des mots, des opérateurs ou trouver des synonymes!

Valider l'information

Grâce à Internet, tout le monde peut chercher mais aussi mettre en ligne des infos sur la toile. À vous donc de savoir reconnaître une information donnée par un particulier, un professionnel, ou par un organisme officiel, et juger de la crédibilité des données.

En résumé, on peut valider la source en vérifiant l'un des 4 points suivants :

- Qui est l'auteur ? Un particulier ? Un professionnel ? Un organisme ?
- Quel est l'objectif du site ? Vendre ? Informer ?
- Quel est le type du site ? Personnel ? Blog ?
- Comment est présenté le contenu ? Comment est l'architecture du site ? Le contenu est-il mis à jour ? Quel est le registre de langue (formel, familier...) ? Y-a t-il des fautes d'orthographe ?

Exercice final

Après avoir vu tous les éléments nécessaires pour chercher et valider vos sources sur Internet, je vous propose d'effectuer la recherche suivante :

Quelles sont les différentes sources d'énergies renouvelables en France? Et quels sont leurs impacts sur l'environnement ?

1. Soulignez les mots-clés de cette problématique.
2. Rédigez la formule de recherche possible que vous taperiez sur Google.
3. Sélectionnez un site et analysez-le selon le schéma suivant :

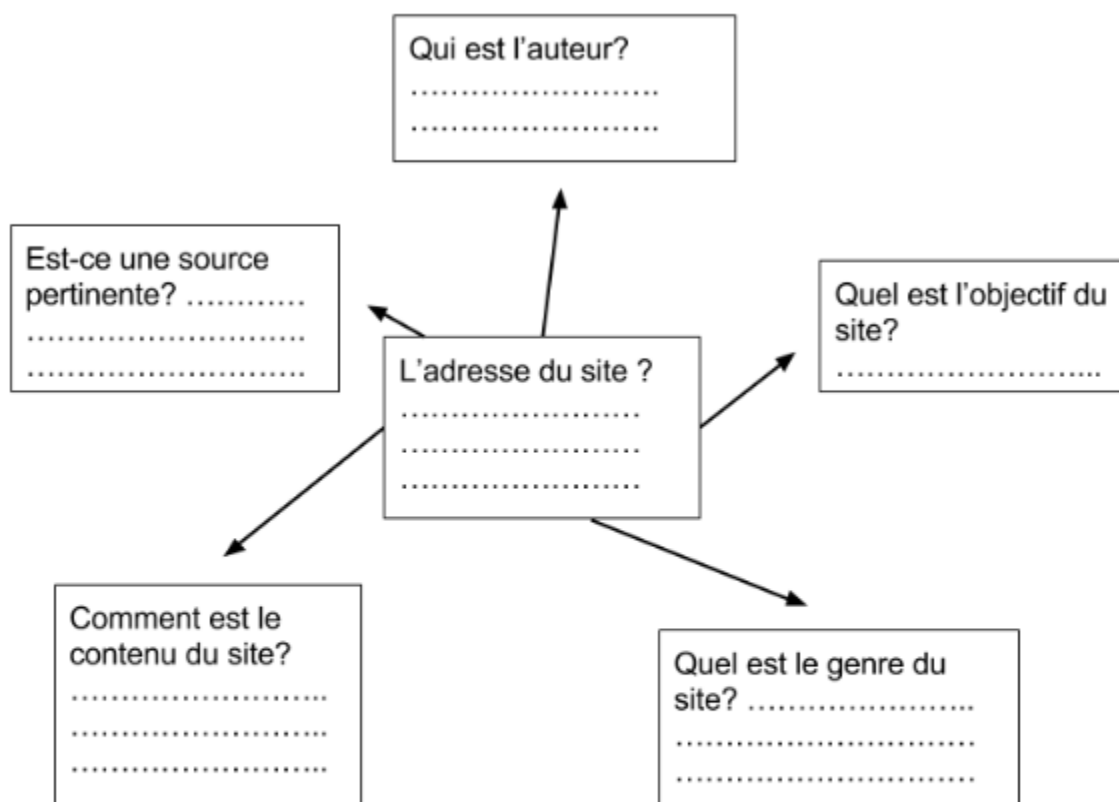


Schéma d'analyse de site Internet

Exploiter l'information trouvée

Une fois les informations trouvées, qu'en fait-on ? À vous de les exploiter selon vos besoins! Attention toutefois, ce n'est pas parce que c'est sur Internet que vous ne devez pas respecter le travail des auteurs selon la propriété intellectuelle et le droit à l'image.

Le droit d'auteur et la propriété intellectuelle, kezako ?

Il s'agit de respecter le travail d'un auteur en citant la source (le site Internet) et le nom de l'auteur. Par exemple, si vous reprenez une citation trouvée sur le site d'OpenClassrooms, vous devrez :

- Nommer le site avec la date (en effet, sur Internet, les informations peuvent disparaître très vite!)
- Nommer l'auteur s'il est précisé
- Mettre la citation entre guillemets

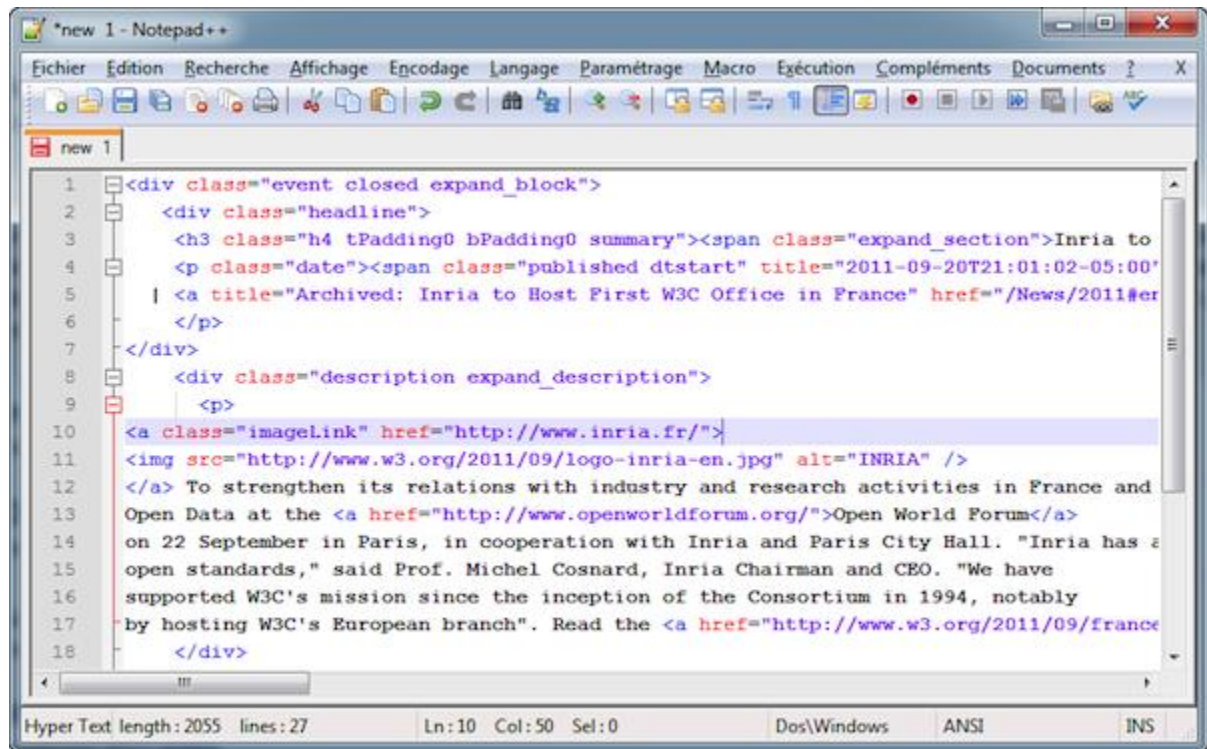
Par exemple : “ HTML (HyperText Markup Language) : il a fait son apparition dès 1991 lors du lancement du Web. Son rôle est de gérer et organiser le contenu. C'est donc en HTML que vous écrirez ce qui doit être affiché sur la page : du texte, des liens, des images...” de Mathieu Nebra

Et le droit à l'image, kezako ?

C'est pareil, même si une image ou une photo apparaît sur Internet, elle est la propriété de quelqu'un. Vous ne pouvez donc pas les utiliser dans vos devoirs sans

citer vos sources, et il faut parfois même demander à l'auteur l'autorisation de l'utiliser.

Par exemple :



```
1 <div class="event closed expand_block">
2   <div class="headline">
3     <h3 class="h4 tPadding0 bPadding0 summary"><span class="expand_section">Inria to
4     <p class="date"><span class="published dtstart" title="2011-09-20T21:01:02-05:00"
5     | <a title="Archived: Inria to Host First W3C Office in France" href="/News/2011#er
6     </p>
7   </div>
8   <div class="description expand_description">
9     <p>
10    <a class="imagelink" href="http://www.inria.fr/">
11    
12    </a> To strengthen its relations with industry and research activities in France and
13    Open Data at the <a href="http://www.openworldforum.org/">Open World Forum</a>
14    on 22 September in Paris, in cooperation with Inria and Paris City Hall. "Inria has s
15    open standards," said Prof. Michel Cosnard, Inria Chairman and CEO. "We have
16    supported W3C's mission since the inception of the Consortium in 1994, notably
17    by hosting W3C's European branch". Read the <a href="http://www.w3.org/2011/09/france
18    </div>
```

Coloration sous Notepad (Source : Créer un site Web avec HTML5 et CSS3 de Mathieu Nebra sur le site OpenClassrooms.com)

Je vous conseille donc d'aller directement sur des sites où les images sont libres de droit ou sur des sites qui vous autorisent à utiliser leurs images (avec des licences Creative Commons ou des licences propres au site en question). Regardez par exemple du côté du site www.sxc.hu !

Activité 2:

Test

Question 1

Choisissez la bonne formulation pour la recherche suivante: *La pollution des mers en France.*

- Pollution France Pollu* + France (pollution ET mer) ET France

Question 2

Choisissez la bonne formulation pour la recherche suivante: *L'impact des OGM sur la santé de l'homme.*

- OGM ET santé humaine OGM (santé humaine) impact + OGM + santé + homme

Question 3

Choisissez la bonne formulation pour la recherche suivante: *Architecture, agriculture en Europe.*

-

Architecture + agriculture + Europe

-

(Architecture OU agriculture) ET France

-

Architecture + Agriculture en Europe

Question 4

Choisissez la bonne formulation pour la recherche suivante: *Le réchauffement climatique et ses conséquences sur l'environnement dans le monde. .*

- Réchauffement climatique + monde (Réchauffement + climatique + conséquences) ET environnement dans le monde (Réchauffement climatique ET environnement) ET mondeValidation

Correction:

Question 1 -

Choisissez la bonne formulation pour la recherche suivante: *La pollution des mers en France.*

Pollution France Pollu* + France **[La bonne réponse est...]**

(pollution ET mer) ET France

Question 2 -

Choisissez la bonne formulation pour la recherche suivante: *L'impact des OGM sur la santé de l'homme.*

[La bonne réponse est..] OGM ET santé humaine OGM (santé humaine) impact + OGM + santé + homme

Question 3 -

Choisissez la bonne formulation pour la recherche suivante: *Architecture, agriculture en Europe.*

[La bonne réponse est

Architecture + agriculture + Europe

(Architecture OU agriculture) ET France

Architecture + Agriculture en Europe

Question 4 -

Choisissez la bonne formulation pour la recherche suivante: *Le réchauffement climatique et ses conséquences sur l'environnement dans le monde. .*

○ Réchauffement climatique + monde ○ (Réchauffement + climatique + conséquences) ET environnement dans le monde ○ [La bonne réponse est..] (Réchauffement climatique ET environnement) ET monde