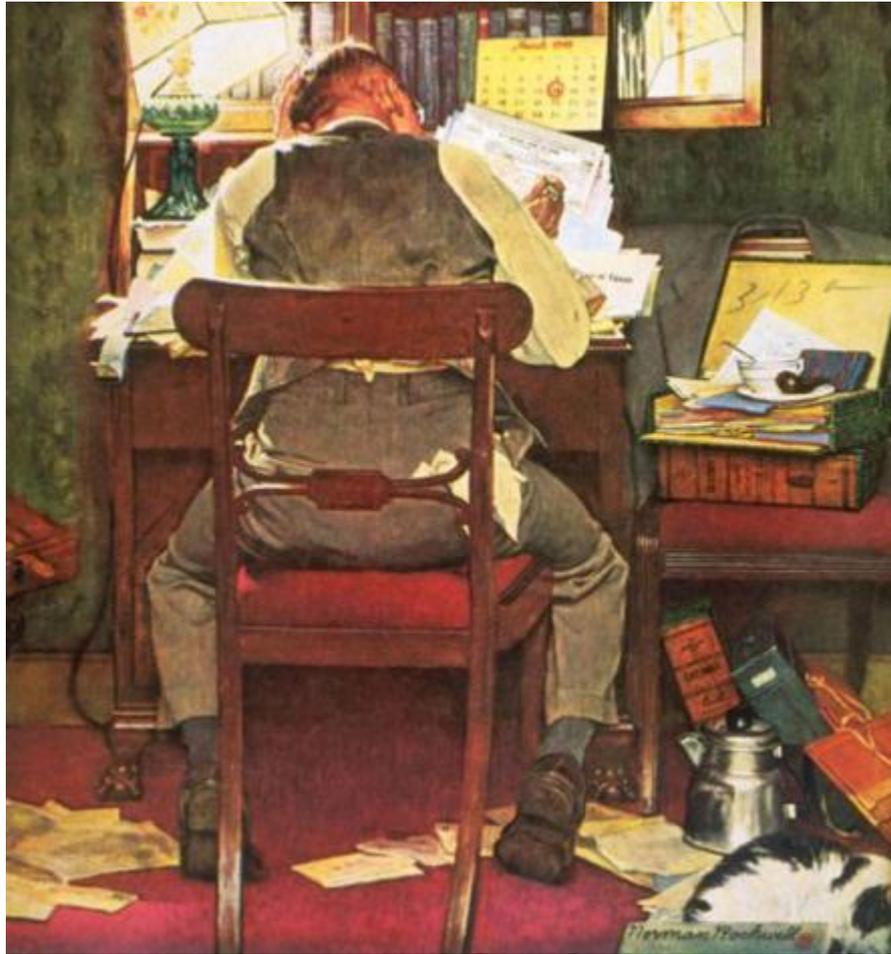


# Recherche d'informations sur Internet (perfectionnement) méthodologie et outils disponibles





Pour commencer

Principes

Moteurs de recherche

Sites internet

Bases de données bibliographiques

Autres bases de données textuelles

Images et multimédia

Web social

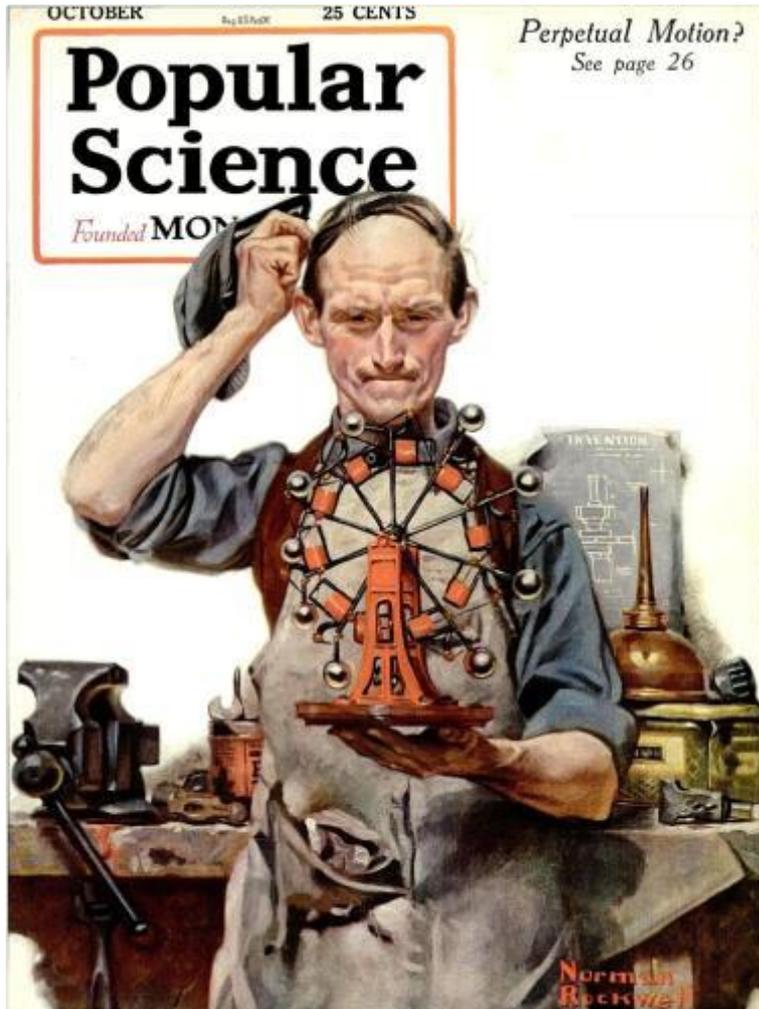
Actualités et temps réel

Quelques outils complémentaires

Veille automatisée

Exercices de synthèse

Bibliographie



Principles

# Internet ? web ?

- internet
  - réseau de réseaux
    - fin des années 1960
    - protocole TCP/IP
    - applications et services divers : courrier électronique (mail), messagerie instantanée (IM), forums de discussion, transfert de fichiers (FTP), pair à pair (P2P), web (www)...
  - données
    - utilisateurs : 4,2 MM. dans le monde ([Internet World Stats](#), 2018)
    - langues : 54 % du contenu en anglais, 4 % en français ([W³Techs](#), 2019)
    - voir également [Internet live stats](#)

# Internet ? web ?

- Web
  - World Wide Web (www)
    - milieu des années 1990
    - ensemble de pages HTML (textes, images, liens...) avec une URL et accessibles avec le protocole HTTP
  - web visible / web invisible
    - web invisible ou web profond (*deep web*) : partie du web non indexée et qui ne peut être trouvée par les moteurs de recherche (pages protégées par un mot de passe, pages générées dynamiquement à la suite d'une requête...), voire *dark web* (web illégal) : 95 % du total ?
  - taille
    - 1,8 MM. de sites ([Netcraft](#))
    - web indexé : au moins 4,5 milliards de pages ([Worldwidewebsize](#))
    - taille du web identifié (URL connues) ? web général ?



- **DEBUSQUER l'information**

Différents outils

Esprit critique

Bookmark organisé

URL significative

Syntaxe de recherche

Questions préalables

Utilisation réfléchie

Evaluation

Règles à respecter

## 1° définir le sujet (contexte de la recherche et mots-clés)

### Questions préalables

*Prendre du temps au départ pour en gagner par la suite...*

*Ne pas se lancer tout de suite dans la requête. Se poser des questions :*

*- Au minimum :*

*Qu'est-ce que je cherche exactement ? Le web est-il le moyen le plus rapide ?*

*- Pour aller plus loin :*

*Qui (profil du demandeur) ? Quoi (nature de l'information) ? Où (limites géographiques, lieux ressources, producteur de l'information) ? Quand ? Pourquoi (objectifs) ? Comment (outils, méthodologie, stratégie) ?*

d'après [Eduscol](#). *Rechercher sur internet*

→ définir en amont une stratégie de recherche, avant même d'utiliser un moteur de recherche

- dictionnaires : abréviations, synonymes, hiérarchies, associations
  - généralistes
    - Dictionnaires Larousse : <http://www.larousse.fr/dictionnaires>  
21 dictionnaires de langue (français et bilingues)
    - Portail du CNRTL : <http://www.cnrtl.fr/>  
portail du Centre national de ressources textuelles et lexicales, rassemblant un ensemble de ressources linguistiques (lexiques, dictionnaires...)
    - Trésor de la langue française (TLF) : <http://atilf.atilf.fr/tlf.htm>
    - Grand dictionnaire terminologique (GDT) : <http://gdt.oqlf.gouv.qc.ca/>  
dictionnaire québécois, propose également une traduction en anglais
    - Dictionnaire de l'Académie française : <http://www.dictionnaire-academie.fr/>
    - Orthonet : <http://orthonet.sdv.fr/>  
difficultés linguistiques de la langue française
    - FranceTerme : <http://www.culture.fr/franceterme>  
termes recommandés au *Journal officiel*  
à noter : vocabulaires spécialisés : <http://www.culture.fr/Ressources/FranceTerme/Librairie>

# Méthodologie

- dictionnaires multilingues
  - WordReference : <http://www.wordreference.com/fr/>  
intérêt des forums
  - OneLook : <http://www.onelook.com/>
  - Lexilogos : <http://www.lexilogos.com>
  - Interglot : <http://www.interglot.com>
  - Linguee : <http://www.linguee.fr/>  
contextualisation de la recherche
  - Acronym Finder : <http://www.acronymfinder.com/>
- <http://magicsearch.org> outil multilingue
- synonymes
  - Dictionnaire électronique des synonymes (DES) : <http://www.crisco.unicaen.fr/des/> (Fr)
  - Reverso synonymes : <https://synonyms.reverso.net/synonym/> (multilingue, dont Fr)
  - WordNet : <https://wordnet.princeton.edu/> (GB)
  - Synonym : <https://www.synonym.com/> (GB)
  - Thesaurus.com : <https://www.thesaurus.com/> (GB)

- thésaurus bibliothéconomiques  
abréviations, synonymes, hiérarchies, associations
  - RAMEAU : <http://catalogue.bnf.fr>  
Notices d'autorité (type de notice > Répertoire RAMEAU > Nom commun RAMEAU)  
thésaurus utilisé par un grand nombre de bibliothèques françaises et notamment le catalogue de la BnF et le SUDOC pour indexer les documents
  - MOTBIS : <http://www.cndp.fr/motbis/>  
thésaurus de l'Education nationale, utilisé dans les CDI de l'enseignement secondaire
  - Library of Congress Authorities : <http://authorities.loc.gov/>  
équivalent de RAMEAU pour la bibliothèque nationale des Etats-Unis
  - MeSH (*Medical Subject Headings*) : <http://www.ncbi.nlm.nih.gov/mesh>  
thésaurus utilisé pour indexer les articles dans la base de données PubMed (médecine)
- + Bartoc (*Basel Register of Thesauri, Ontologies and Classifications*) : <http://bartoc.org/>  
répertoire de vocabulaires contrôlés (thesauri, ontologies, glossaires...)  
2 800 ressources, 77 langues

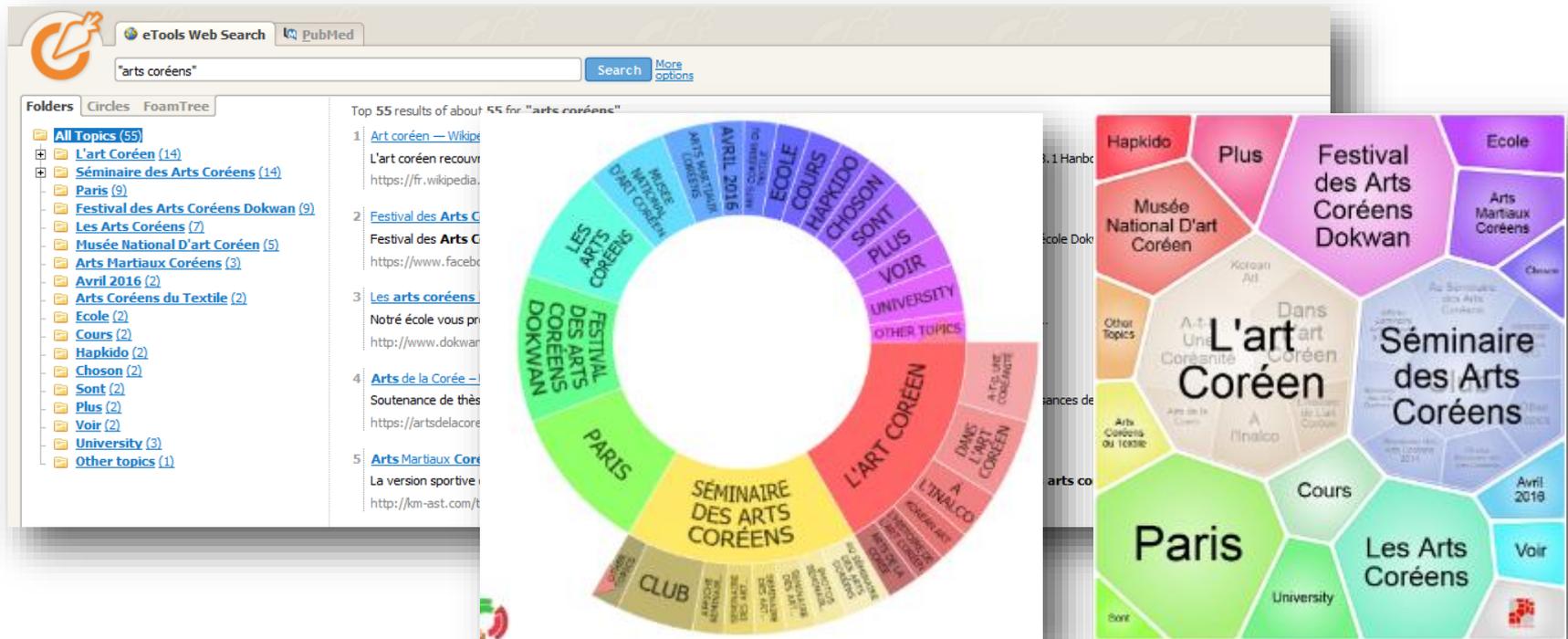
- outils multilingues
  - scientifiques

- TermSciences : <http://www.termosciences.fr/>  
portail terminologique multidisciplinaire français (CNRS, INIST, INRIA...)  
termes français, anglais, espagnols, allemands  
arbre des concepts  
+ 190 000 concepts, 650 000 termes
- EuroVoc : <http://eurovoc.europa.eu/drupal/?q=fr/>  
thésaurus multilingue et multidisciplinaire de l'Union européenne (27 langues)  
à compléter par la base IATE (Inter-Active Terminology for Europe) : <https://iate.europa.eu/home>
- AGROVOC : <http://aims.fao.org/standards/agrovoc/functionalities/search>  
thésaurus multilingue sur l'agriculture de la FAO (Organisation des Nations unies pour l'alimentation et l'agriculture)  
+ 33 000 concepts ; 23 langues : nourriture, agriculture, environnement...
- Termium Plus : <http://www.btb.termiumplus.gc.ca>  
banque de données terminologiques et linguistiques du gouvernement du Canada  
4 langues
- Exemplar : <http://springerexemplar.com/index.aspx>  
uniquement sur les ressources Springer (+ 12 M. de documents)  
terminologie scientifique utilisée dans la littérature scientifique publiée (« *words in context* »)
- Lexique Techniques de l'ingénieur : <http://www.techniques-ingenieur.fr/lexique.html>  
secteurs industriels majeurs (partiellement payant)  
45 000 concepts ; 4 langues
- WIPO Pearl : <http://www.wipo.int/wipopearl/search/home.html>  
portail multilingue des termes scientifiques et techniques issus de brevets  
OMPI (Organisation Mondiale de la Propriété Intellectuelle)  
16 000 concepts, 105 000 termes ; 10 langues ; recherche par langue ou par carte de concepts
- Thésaurus de l'UNESCO : <http://databases.unesco.org/thesfr/>  
7 000 termes en anglais et en russe ; 8 600 termes en français et en espagnol : éducation, culture...
- Atlas sémantiques du CNRS : [http://dico.isc.cnrs.fr/dico\\_html/fr/index.html](http://dico.isc.cnrs.fr/dico_html/fr/index.html)

! différences entre l'anglais et l'américain :  
Ex. : *fibre/fiber* , *flavour/flavor* , *sulphur/sulfur*...

# Méthodologie

- outils de *clustering*
  - Carrot : <http://search.carrotsearch.com> [et <http://search.carrot2.org>] métamoteur (Bing, Google, Yahoo, etc.) peu de résultats, mais affichage des résultats par groupes (*clusters*)



- Yippy : <http://yippy.com/>

- suggestions des moteurs de recherche

## Recherches associées à maladie vache folle

maladie **de la** vache folle **traitement**

vache folle **2018**

**la** maladie **de la** vache folle **c'est quoi**

**crise de la** vache folle **conséquences**

maladie **de la** vache **chez l'homme**

vache folle **2017**

vache folle **video**

maladie **creutzfeldt jakob**

Google

- ! : suggestions en fonction des recherches les plus populaires des internautes, et non les plus pertinentes par rapport à sa propre recherche

## Exemple de listes de mots-clés

### Exemple 1 : Traitements biologiques aérobies des eaux de rejets industriels pollués

Entre parenthèses les équivalents anglais des mots-clés

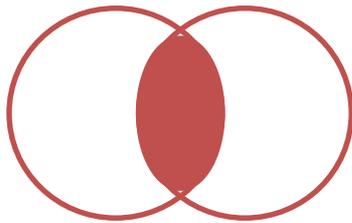
Concepts retenus	Termes génériques (termes plus larges)	Termes spécifiques (termes plus précis)	Termes associés (termes de sens voisin)
Mots trouvés par simple analyse des termes du sujet	Mots trouvés avec un dictionnaire de terminologie scientifique ou technique	Mots trouvés à l'aide d'un dictionnaire de terminologie scientifique ou technique, index "sujets" ou "mots clés" d'un catalogue ou d'une base de données	Mots trouvés à l'aide de l'index "sujets" ou "mots clés" d'un catalogue ou d'une base de données
<ul style="list-style-type: none"> <li>- traitement biologique (biological treatment)</li> <li>- aérobie* (aerobic)</li> <li>- rejet* industriel*</li> <li>(industrial waste)</li> <li>- eau* (water)</li> <li>- polluant* (pollutant)</li> </ul>	<ul style="list-style-type: none"> <li>- procaryote* (prokaryote*)</li> <li>- microorganisme*</li> <li>(microorganism*)</li> <li>- épuration des eaux usées (wastewater treatment)</li> <li>- assainissement (sanitation)</li> </ul>	<ul style="list-style-type: none"> <li>- dépollution biologique, biodépollution, bioremédiation (bioremediation)</li> <li>- effluent industriel (industrial effluent)</li> <li>- digestion aérobie (aerobic digestion)</li> </ul>	<ul style="list-style-type: none"> <li>- épuration (purification)</li> <li>- biodégradation (biodegradation)</li> <li>- bactérie* aérobie (aerobic bacterium/ bacteria)</li> <li>- bactérie* (bacterium/ bacteria, eubacterium / eubacteria)</li> </ul>

## 2° connaître les outils et la construction d'équations de recherche

→ réfléchir en terme d'outils (hors moteurs de recherche)

<b>Différents outils</b>	<p><i>Il n'y a pas que Google !</i></p> <ul style="list-style-type: none"><li>- bien comprendre les différences entre moteurs, métamoteurs et annuaires,</li><li>- distinguer les outils généralistes et les outils spécialisés,</li><li>- repérer les sites portails utiles pour vos recherches les plus fréquentes,</li><li>- essayer et comparer des outils de types différents, et d'autres du même type,</li><li>- lire l'aide en ligne, l'imprimer, mieux encore, faire sa propre fiche,</li><li>- mieux vaut maîtriser 1 ou 2 outils que sous-utiliser un grand nombre.</li></ul>
<b>Utilisation réfléchie</b>	<p><i>En fonction des réponses apportées, utiliser de façon réfléchie</i></p> <ul style="list-style-type: none"><li>- l(es) outil(s) adapté(s), avec au besoin des filtres pays/langues,</li><li>- de bons mots clés avec une syntaxe correcte (en saisissant bien la requête).</li></ul>
<b>Syntaxe de recherche</b>	<p><i>Se familiariser avec la syntaxe</i></p> <ul style="list-style-type: none"><li>- les principes d'interrogation des bases de données, (opérateurs booléens et de proximité, signes + et -, troncature...),</li><li>- la syntaxe de saisie spécifique à chaque outil (recherche simple et recherche avancée).</li></ul>

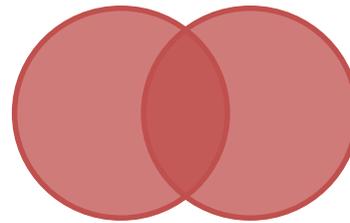
- syntaxe de recherche
  - opérateurs booléens
    - AND, OR, NOT / BUT NOT / AND NOT / - , plus rarement ET/OU/SAUF



AND  
ET

= réunion

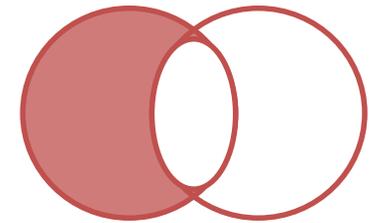
le document doit comprendre  
tous les termes  
[parc ET jardin]



OR  
OU

= union

le document peut comprendre  
tous les termes  
ou seulement certains d'entre eux  
[parc OR jardin]



NOT  
SAUF

= exclusion

le document ne doit pas  
comprendre certains termes  
[parc NOT jardin]

- parenthèses : ex. : [Versailles (parc OR jardin)], [Edo (période OR époque)]

# Méthodologie

- syntaxe
  - ordre des mots
  - mots vides : articles (le, la, du...), conjonctions de coordination (donc, ni...), prépositions (pour, sur...)
  - termes implicites dans le document (document...)
    - ex. : [travail enfants Chine] pour trouver des documents sur le travail des enfants en Chine ;
    - [droit femmes France] pour des documents sur le droit des femmes en France
  - guillemets pour expression exacte : ex. : [« base de données »]
    - notamment pour chercher des noms de personnes ; dans ce cas, penser éventuellement à une double recherche, pour ramener aussi les classements alphabétiques au nom propre : ex. : [Emmanuel Petit] et [Petit Emmanuel]
    - ! aux abréviations et initiales : ex. : [« George Bush »] : distinction George W. Bush / George H. W. Bush ?
- orthographe
  - casse (majuscules/minuscules ?)
  - +mot et -mot pour indiquer ou soustraire un terme exact, ex. [+De Wilde], [+marché] (pour exclure « marche »)
  - troncature : ex. : [jardin\*] pour jardinage, jardiner, jardinier...
  - accents (notamment pour outils non francophones) , ponctuation et signes diacritiques (@, #...)
  - pluriels : ex. : [« textiles techniques »] et [« textile technique »]
- proximité
- syntaxe de saisie
  - interface de l'outil : recherche simple et avancée
  - filtres avant ou après la saisie de la requête

→ construction de l'équation de recherche variable selon les outils  
→ consulter l'aide

## 3° analyser les résultats

### URL significative

*L'URL donne des informations importantes.*  
*- Décrypter la structure des adresses url pour identifier la nature du site,*  
*- Modifier une adresse url pour accéder à une page web.*

### Esprit critique

*Tout au long de la recherche : faire preuve d'esprit critique, prendre du recul...*  
*Est-ce fiable ? S'agit-il d'un fait réel ? d'une idée reçue ? d'une rumeur ?*  
*S'agit-il d'une opinion ? Quels autres points de vue sur le sujet ?*  
*Est-ce bien ce que je cherche ? Ma méthode de recherche est-elle adaptée ?*

### Evaluation

*Evaluer*  
*- l'information (fiabilité, intérêt),*  
*- l'intérêt du site (richesse du contenu) et son ergonomie (clarté, navigation dans le site),*  
*- les résultats (performance des outils, pertinence de la requête).*  
*Adopter au besoin une nouvelle stratégie (élargir ou restreindre la requête).*  
*Conserver éventuellement la requête (termes et opérateurs utilisés) dans le bookmark.*

- principaux critères d'évaluation

page de résultats	présence des mots recherchés URL de la page
site	qualité de la rédaction qualité des liens proposés présence de publicité actualisation et mises à jour
auteur du site ou du document	authentification et formulaire de contact compétences légitimité
contenu du document	public destinataire sources buts

→ crédibilité, qualité, pertinence et fiabilité

pour aller plus loin, [A. Serres, \*Evaluation de l'information sur internet\*](#)  
*Guide de vérification : la référence de la vérification du contenu numérique*  
*pour couverture d'événements dans l'urgence*  
voir également la méthode CRAAP (Currency [opportunité], Relevance [adéquation],  
Authority [autorité], Accuracy [fiabilité], Purpose [finalité])

- « *Keep wading to a minimum :  
Size of the search tool does not matter* »

« Si vous n'avez pas trouvé ce que vous cherchez dans les 20 à 50 premiers sites, abandonnez et n'allez pas plus loin. Soit reformulez votre recherche, soit essayez un autre outil de recherche [...].

Aucun moteur de recherche n'est le meilleur. Une recherche complexe nécessite des outils de recherche nombreux. Le nombre de sites appropriés est plus important que le nombre de sites recherchés. »

*D. Vidmar. Tips to Effective Internet Searching*

## 4° exploiter les résultats

### **Bookmark organisé**

*Un bookmark bien organisé facilite les recherches et permet de retrouver les informations.*

- Avoir le réflexe (marque-pages sur [Firefox], favoris sur Internet Explorer, [outils de social bookmarking comme Diigo]),*
- Identifier ses sources pour bien structurer le bookmark,*
- Constituer des dossiers en fonction de ses centres d'intérêt et les actualiser,*
- Constituer un dossier spécifique à l'occasion d'une recherche.*

### **Règles à respecter**

*Respecter les règles*

- respecter la netiquette mais aussi naviguer en internaute averti,*
- respecter le droit d'auteur, citer la source d'un document et ses références.*

d'après [Eduscol](#). *Rechercher sur internet*

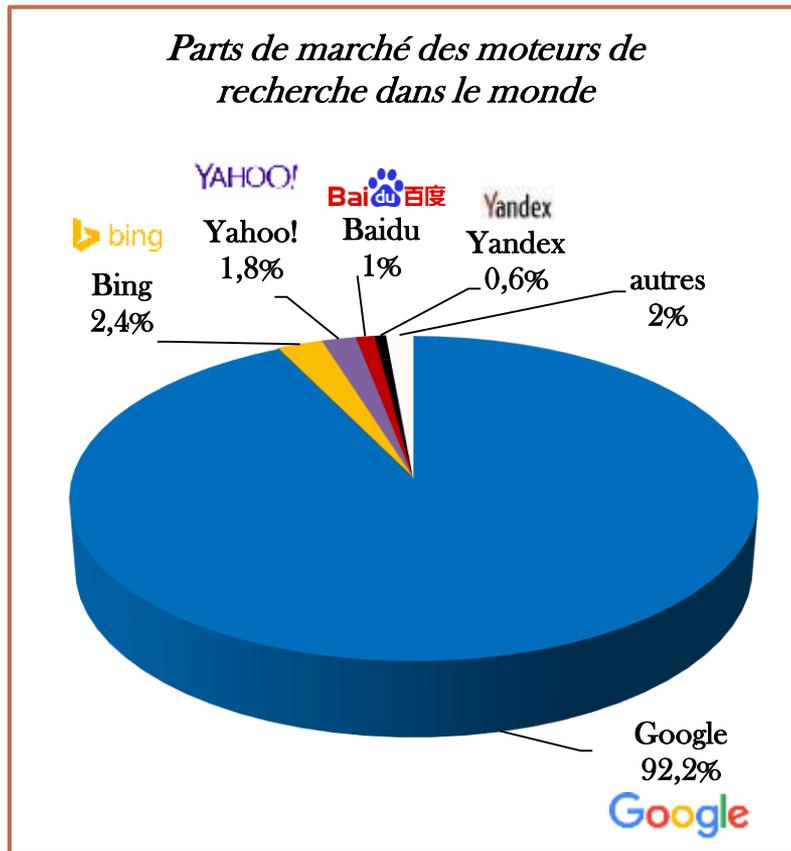


Moteurs de recherche

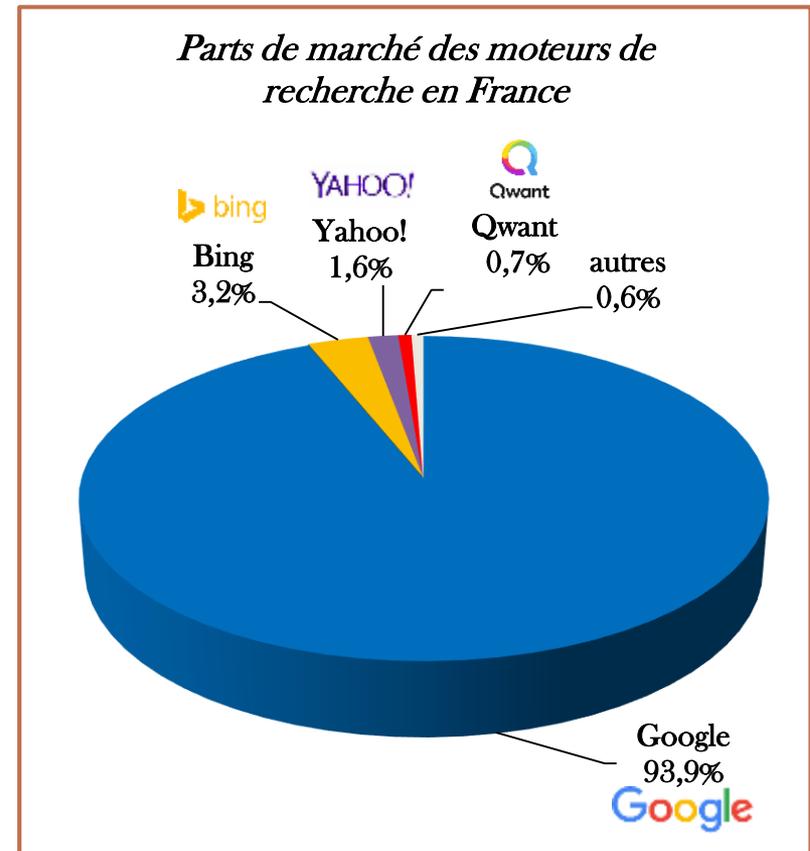
# Moteurs de recherche

<b>Documents</b>	diversité des documents et des formats : <ul style="list-style-type: none"><li>- pages web</li><li>- sites</li><li>- documents (texte, images, vidéos...)</li></ul>
<b>Recherche</b>	RECHERCHE PAR MOTS-CLES (texte intégral, titre, URL...) spécifique à chaque outil présence généralement d'une recherche avancée avec des possibilités de recherches complexes (requêtes construites) existence possible de tris et de filtres, d'une aide
<b>Périmètre</b>	généralement web visible peu d'interrogations du web invisible existence de moteurs généralistes et spécialisés mais indexation et présentation des résultats automatiques du fait du classement par « pertinence » (i.e. généralement par popularité), difficile de trouver des documents originaux et peu connus
<b>Outils</b>	moteurs de recherche généralistes métamoteurs moteurs de recherche spécialisés (scientifiques notamment) ou « verticaux » (thématiques)
<b>+ / -</b>	! pas d'exhaustivité + pour informations, données factuelles (dates, adresses...) ou questions générales ! les mots-clés recherchés doivent se trouver dans les documents (texte intégral, métadonnées...) ! présence de publicités ; validité des sources non vérifiée évolution vers une prise en compte du langage naturel et une reconnaissance sémantique des termes  pour une recherche sur un concept ou une vue d'ensemble, intérêt de faire sa recherche sur différents outils afin de profiter de l'index de chacun des services, de leurs fonctionnalités spécifiques et de classement différent des résultats

## Monde

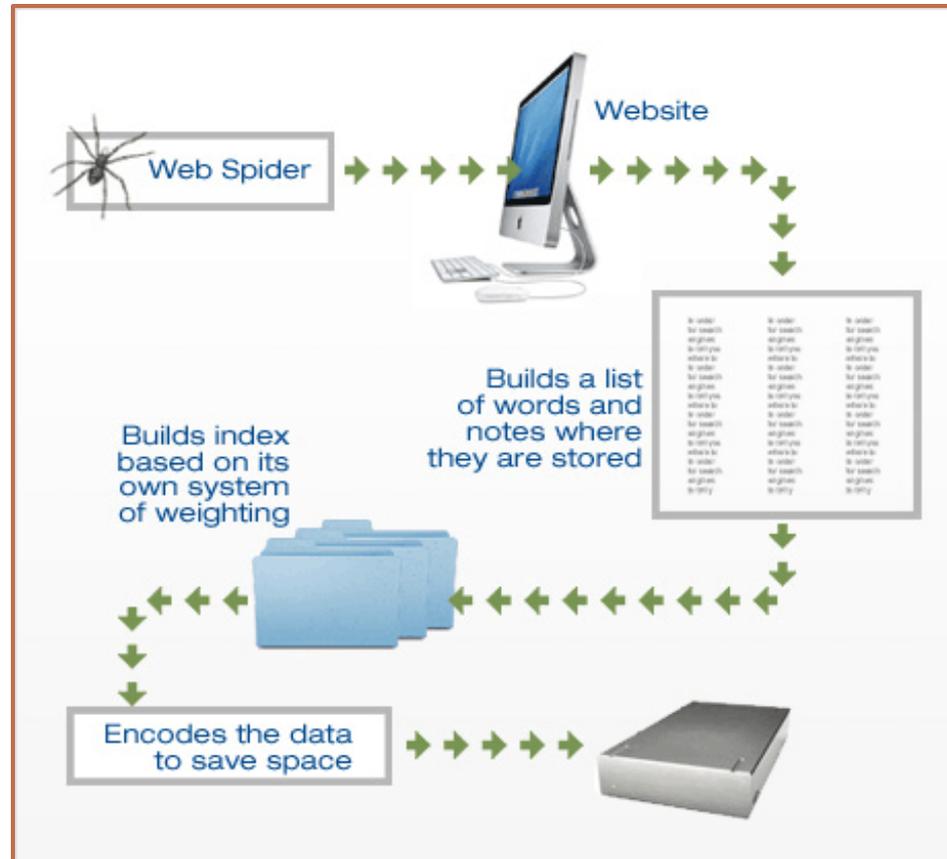


## France



# Moteurs et métamoteurs de recherche

- Comment fonctionne un moteur de recherche ?



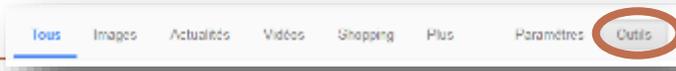
! chaque moteur a sa propre couverture du web, ses propres index, algorithmes et fonctionnalités

# Moteurs de recherche

- fonctionnalités générales possibles
  - combinaison des différents opérateurs (AND/OR..., parenthèses, guillemets...)
  - filtres de recherche
    - titres des pages
      - ! aux accents
    - nom de domaine ou site
      - + fonctionnalités plus avancées que moteur de recherche sur le site
      - ! seulement sur les pages indexées par le moteur
    - URL des pages
      - ! pas d'accent
    - type de fichier
    - date (Google, Exalead)
      - ! date de publication (1<sup>e</sup> indexation des pages) et pas toujours date de rédaction de la page
  - « backlinks » d'une page ou d'un site
  - suggestions de recherches associées
  - formulaires de recherche avancée
    - ne permettent généralement pas des combinaisons très poussées
  - préférences de recherche (langue, résultats, sécurité)

- Google.fr : <https://www.google.fr>
- Google.com :  
[https://www.google.com/webhp?hl=en&gws\\_rd=cr,ssl&ei=oY5dVsjINcXdUaqDucAH&fg=1](https://www.google.com/webhp?hl=en&gws_rd=cr,ssl&ei=oY5dVsjINcXdUaqDucAH&fg=1)
- révolutionnaire en 1998  
mode de recherche (AND par défaut)  
simplicité de l'interface (*one box*) - recherche simple
- données  
chiffres inconnus (connaissance de [130 MM. de pages web](#))  
ne concerne qu'une partie du web  
100 milliards de requêtes par mois
- [recherche avancée](#)
- classement des résultats  
algorithme tenu secret prenant en compte plusieurs éléments :
  - ordre des mots dans la requête
  - fréquence et place des mots
  - fraîcheur de l'information
  - *PageRank* (notoriété d'une page grâce aux liens)

→ pages connues et non les plus pertinentes dans les premiers résultats  
→ prise en compte des nouveaux sites pas encore populaires ?

fonctionnalités de recherche	opérateurs booléens	<p><b>AND</b> : tous les mots (implicite) : ex. [jardin jardinage]</p> <p><b>OR</b> : au moins un des mots : ex. [jardin OR jardinage]</p> <p><b>*</b> : joker : ex. [<b>*</b> de Sisyphes] pour le mythe ou le supplice de Sisyphes (opérateur de proximité)</p> <p><b>-mot</b> : mot exclu</p> <p>problème avec les parenthèses ( )</p>
	<p><a href="#">critères de recherche avancée</a></p> 	<p><b>«mot»</b> : mot exact, sans synonyme ni variante orthographique : ex. [Mickael «Jackson»] pour exclure Jackson</p> <p><b>« »</b> : expression exacte : ex. [«réchauffement climatique»]</p> <p><b>intext</b> et <b>allintext</b> : pour forcer Google à prendre en compte le terme sans le laisser de côté</p> <p><b>intitle</b> et <b>allintitle</b> : pour un mot ou une expression dans le titre de la page</p> <p><b>inurl</b> et <b>allinurl</b> : pour un mot ou une expression dans l'URL</p> <p><b>site</b> : site ou domaine : ex. [site:google.fr], [site:.edu]</p> <p><b>filetype</b> : type de fichier : ex. [filetype:ppt]</p> <p>chiffre..chiffre (+ unité de mesure au besoin) : fourchette de chiffres (comparaison, prospective..) : ex. [Michel Platini 1970..1980]</p>
	autres fonctionnalités	troncature : implicite et aléatoire
résultats	filtres	<p>variables selon le type de document recherché (web, images..) : pays, région, langue, date, localisation, réutilisation</p> <p>+ mot à mot (in Outils &gt; Tous les résultats) : chaîne exacte de caractères :</p> <p>ex. [littérature tibétaine] ne donnera ni littératures ni tibétain</p>
	rech. associées	<p>en bas de la page de résultats (! : traitement automatique)</p> 
autres fonctionnalités	cache, sites similaires	<p>en face d'un résultat, en cliquant sur le petit triangle au bout de l'URL</p> <p><b>related</b> : pages similaires (en fonction des mots-clés)</p> 
	<a href="#">préférences</a>	<p>filtre SafeSearch</p> <p>résultats par page</p> <p>blocage des résultats indésirables</p> <p><a href="#">historique web</a></p>
	indexation d'un site par Google	<p><b>info</b> : ex. [info:google.fr] (page en cache, pages similaires, liens entrants..)</p> <p><b>cache</b> : version en cache telle qu'enregistrée lors du dernier passage par les robots de Google</p>

Voir également : [Aide sur Google](#)  
opérateurs de recherche

- ! tendance de plus en plus fréquente de Google (cf. [aide](#)) à :
  - suggérer des corrections orthographiques et d'autres orthographes
  - personnaliser la recherche en utilisant les données de l'utilisateur (notamment localisation et compte Google)
  - inclure des synonymes des mots recherchés (recherches considérées comme similaires)
  - chercher sur la racine d'un mot (inclusion du pluriel quand on cherche un singulier, etc.)
  - exclure certains termes recherchés
  
- comment bien interroger Google en 2018 ? ([C. Tisserand-Barthole](#) et *Bases*, 340)
  - conseil 1 : éviter les longues requêtes booléennes, et multiplier les requêtes courtes (3-4 mots) ; inutile de saisir toutes les variantes d'un terme séparées de OR ; mieux vaut lancer des requêtes différentes en variant les mots-clés
  - conseil 2 : vérifier les termes recherchés : extension à des synonymes et des termes proches ?
  - conseil 3 : tester différentes formulations, dans cet ordre
    - 1° requête simple en langage naturel (sans opérateurs, ex. : [vente médicaments en ligne France]) → pour laisser une partie d'interprétation au moteur, notamment des pistes insoupçonnées
    - 2° requête booléenne classique (avec les opérateurs, ex. : [(vente OR achat) AND (médicaments OR médicament) AND (« en ligne » OR internet) AND France] → pour affiner
    - 3° requête sous forme de question (i.e. une phrase déjà construite, comme [La vente des médicaments en ligne en France est autorisée]) → pour cibler des éléments précis

# Moteurs de recherche

- moteurs multilingues

- Bing : <http://www.bing.com/>

Microsoft, 2009

actuellement, le meilleur complément à Google à défaut de proposer une vraie alternative

- moins de critères de recherche avancée et de filtres que Google

- Yahoo! : <https://search.yahoo.com/> et <https://fr.yahoo.com/>

technologie : Bing

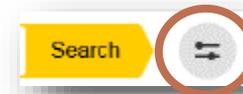
- moins de critères de recherche avancée et de filtres que Google

- Yandex : <https://www.yandex.com>

moteur de recherche russe, avec une version anglophone

[settings](#)

- moins de critères de recherche avancée et de filtres que Google  
à tester cependant pour la recherche anglophone



recherche avancée (sur la page de résultats)

- moteurs territoriaux et nationaux (listes)

! des informations souvent anciennes

- Search Engine Colossus : <http://searchenginecolossus.com/>

répertoire de moteurs de recherche et d'annuaires (315 pays et territoires)

- Search Engines of the World : <http://www.searchenginesoftheworld.com/>

# Moteurs de recherche

- moteurs francophones

- Qwant : <https://www.qwant.com/>

moteur de recherche français, ouvert en 2013

couverture de 15 langues et 25 pays

- peu de tris (date)
- peu de fonctionnalités avancées, et qui, parfois, connaissent certains dysfonctionnements (OR , - , intitle: , filetype:, site: et [Qwicks](#) pour des recherches sur certains sites spécifiques)
- dysfonctionnement sur la recherche exacte par guillemets (« »)
- un index encore limité (20 MM. de pages), d'où une [utilisation ponctuelle de Bing](#) au besoin
- fraîcheur des données ? (et pas d'indication de la date de publication dans les résultats)
- nombre de réponses limité (à 50 ?)
- + respecte les recommandations de la CNIL sur les données personnelles ([privacy](#)) : pas de traçage des utilisateurs (cookies provisoires, site sécurisé), pas de filtrage du contenu d'internet (anonymisation)
- ? une [bonne alternative](#) à Google

- Exalead : <http://www.exalead.com/search/>

Dassault Systèmes, plus de 16 MM. de pages indexées

moteur ciblant plus les professionnels que Google (grand public)

- des résultats de plus en plus décevants et grand public, à n'utiliser qu'en dernier lieu

+ opérateurs (notamment)

- OPT : terme dont la présence est optionnelle
- NEXT et NEAR : opérateurs de proximité (NEAR/n pour le nombre max. entre les deux termes)
- soundslike: recherche phonétique
- spelllike: orthographe approchée

+ filtres par termes associés que l'on peut rajouter à la requête (clic) ou exclure (shift + clic)

+ troncature par \*

+ outils d'affinage

# Moteurs de recherche

- la personnalisation des résultats (1) (C. Tisserand-Barthole, *Netsources*, 136)
  - navigation sur internet = enregistrement possible de l'adresse IP, de l'historique de navigation, des cookies, etc.
    - ex. : Google prend en compte avant tout la localisation, la langue et l'historique de recherche  
mais une pratique en baisse chez Google ?
  - résultats les plus susceptibles d'être modifiés : requêtes liées à l'actualité et à des questions locales ; peu de conséquences pour les questions factuelles (s'il y a des différences, elles se situent plutôt dans le bas de la page de résultats)
- avoir de bonnes pratiques pour favoriser une recherche plus neutre et objective
  - sur Google
    - ne pas effectuer de recherche sur Google en étant connecté à un service Google (Gmail, agenda, YouTube...)
    - détruire l'historique, les cookies et les personnalisations : [aide](#), notamment *via* [Google My activity](#)
    - possibilité d'utiliser des services qui anonymisent la recherche
      - <https://www.startpage.com/> : pour interroger Google de façon anonyme
      - <https://isearchfrom.com/> : pour *simuler* une recherche sur Google depuis un autre pays
    - ajouter &pws=0 à la fin de l'URL de la requête
  - en général
    - nettoyer régulièrement son navigateur (historique de navigation, cache et cookies)
    - utiliser des navigations en mode privé (ex. : extensions du navigateur, fenêtre privée...) ; problème : souvent maintien de la géolocalisation
    - voire utiliser des navigateurs liés à des réseaux d'anonymisation ([Brave](#), Tor browser) ou des VPN et proxy

# Moteurs de recherche

- la personnalisation des résultats (2)

→ moteurs ne personnalisant pas les résultats

! limites : plus de suggestions orthographiques ou en fonction de l'historique de navigation

- DuckDuckGo : <https://duckduckgo.com/>

le plus important moteur de recherche neutre à l'heure actuelle (syntaxe de recherche)

+ un classement éventuellement différent des résultats sur Google

- surtout pour les recherches anglophones, parfois moins pertinent pour les recherches francophones (ex. : Brexit)

- peu de fonctionnalités avancées (OR , « » , - , intitle: , filetype:, site:)

- Qwant : <https://www.qwant.com/>

cf. *supra*

+ respecte les recommandations de la CNIL sur les données personnelles (privacy) : « La philosophie de Qwant

repose sur 2 principes : ne pas tracer les utilisateurs et ne pas filtrer le contenu d'internet »

- Hulbee : <https://hulbee.com/>

version suisse : Swisscows : <https://swisscows.ch/?uiLanguage=de&region=fr-FR>

# Moteurs de recherche

- encore quelques autres
  - WolframAlpha : <https://www.wolframalpha.com/>  
n'est pas à proprement parler un moteur de recherche sur internet : interroge sa propre base de données mais alimente d'autres moteurs (Bing, DuckDuckGo, Siri...) ; ~ une encyclopédie
    - contenu essentiellement textuel
    - avant tout pour des recherches factuelles, et plutôt orientées sciences, techniques et médecine
    - pas de listes des sources utilisées
  - trouver des idées
    - Answer the public : <https://answerthepublic.com>
  - explorer la « longue traîne » des résultats (i.e. des résultats moins visibles sur les moteurs)
    - Million Short : <http://millionshort.com/>  
permet d'exclure les sites les plus populaires (100 à 1 M. de sites)
    - BananaSlug : <http://www.bananaslug.com>  
propose des termes aléatoires

# Métamoteurs de recherche

## Métamoteurs de recherche

- outils interrogeant simultanément plusieurs outils de recherche
  - peuvent retraiter les résultats de façon structurée
  - peuvent disposer de fonctionnalités propres
  - permettent d'élargir la recherche

mais

- recherche sur des critères *a minima* communs aux outils interrogés : pas d'opérateurs booléens, peu de filtres, peu d'équations complexes
- ne disposent pas de leur propre index

→ de bons outils pour commencer une recherche (mots-clés, informations, idée de « Top 50 »), ou pour compléter une recherche sur Google

cf. [retours d'expériences 2019](#)

rappel : aucun outil n'indexe la totalité du web

# Métamoteurs de recherche

- eTools : <http://www.ertools.ch/>
  - recherche sur 16 moteurs dont Google, Bing, , Yahoo, Yandex, Qwant, DuckDuckGo
  - + [préférences](#) : poids des moteurs interrogés et nombre de résultats pour chacun, recherche par HTTPS... classement des résultats par *relevance* et indication de la source
- Searx
  - logiciel libre permettant de créer un métamoteur paramétrable
  - nombreuses instances dont <https://framabee.org/> et <https://searx.laquadrature.net/> - ! bien que basés sur le même logiciel, peuvent ne pas donner les mêmes résultats
- Instya : <http://www.instya.com>
  - recherche sur de nombreux outils dont moteurs de recherche et bases de données
  - + possibilité d'indiquer les sites interrogés
  - un onglet par outil interrogé, pas de fusion
- Info.com : <http://info.com/>
  - + recherche sur Google, Bing, Yahoo et Yandex
  - pas d'indication de l'origine des résultats

- tendances actuelles
  - présentation de moins en moins de résultats (quelques centaines au maximum), alors même que la taille des index progresse : « Quelle que soit la requête, Google, Bing et les autres effectuent implicitement une sélection de pages qu'ils jugent pertinentes pour l'utilisateur au sein de leurs index » (C. Tisserand-Barthole, *Netsources*, 132)
- problèmes : - l'utilisateur n'est pas maître de sa recherche d'informations
  - on ignore les critères de ce choix et comment les dépasser *a priori*
- un web de moins en moins « cherchable » : développement de l'information payante (ex. : sites de presse) ou de sites accessibles après inscription (ex. : réseaux sociaux)
- nécessité (cf. [Carole Tisserand-Barthole](#))
  - de bien cibler ses besoins
  - d'utiliser les fonctionnalités avancées des moteurs
  - et, le cas échéant, de relancer la recherche en la réorientant, par exemple en ajoutant ou changeant les mots-clé
  - de chercher la source plutôt que l'information
  - pour les médias sociaux, les interroger directement depuis ces médias, et non pas *via* les moteurs de recherche généraliste
- moteurs de recherche intéressants pour des informations ponctuelles ; plus compliqués pour des états des lieux, des synthèses, des comparatifs
- place toujours importante des outils thématiques (recherche par type de documents, recherche par discipline) hors des moteurs de recherche
- intérêt constant des agrégateurs de presse, bases de données professionnelles, etc. souvent payantes

# Évolutions des moteurs de recherche



sur cette question, cf. [A. Bouchard, \*Évolutions des moteurs de recherche sur internet\* \(2016\)](#)

- évolutions techniques actuelles
  - recherche inversée : reconnaissances non textuelles (audio, images), cf. Shazam pour la musique
  - web sémantique / web de données (reconnaissance d'entités nommées ; synonymes)
  - intelligence artificielle et *machine learning* (cf. distinction « il lit le livre » et « il livre le lit »)
- de plus en plus des moteurs de réponse
  - constat : les moteurs de recherche sont souvent utilisés comme outil de repérage :  
sur Google : 1<sup>er</sup> résultat : + 20 % des clics, 2<sup>e</sup> : 10 %, 3<sup>e</sup> : 7,5 %
  - réponse des moteurs : de moins en moins de listes de résultats, mais une réponse  
sur Google : *featured snippets* et *Knowledge Graph*  
sur Bing : *intelligent answers*  
→ Google se déclare de plus en plus dans l'« ère de l'assistance » (personnalisation + web sémantique + recherche vocale)
  - limite pour la recherche d'information : vérification des sources ?, comparaison ?, état des lieux ?  
(cf. C. Tisserand-Barthole)
- de la recherche textuelle à la recherche conversationnelle
  - développement de la recherche vocale  
ex. : les assistants personnels (Siri, Google home, Alexa)  
ex. : « OK Google » sur Google (120 langues disponibles)
  - recherche conversationnelle : capacité de pouvoir tenir une conversation avec le moteur pour formuler correctement et clarifier la recherche  
limites : encore balbutiant

# Moteurs de recherche scientifiques

## Moteurs de recherche scientifiques

- spécialisés dans la littérature scientifique et académique
- privilégient les sources d'information des chercheurs (articles, congrès, brevets...)
- destinés à la communauté universitaire et scientifique
- gratuits
- parmi leurs fonctionnalités, peuvent proposer l'exportation de références bibliographiques, des analyses bibliométriques, des visualisations

## Evolutions en cours

- avec le développement de l'*open access* et des données ouvertes, multiplication de l'offre, même si les nouveaux entrants interrogent généralement les mêmes bases (ex. : CrossRef, Pubmed, Google scholar, archives ouvertes)
- développement du *machine learning* et du traitement automatique des langues pour mettre en relation des corpus et des données jusque-là séparés (*linked data*) et fournir ainsi des rapprochements et de nouvelles pistes de recherche (développement des moteurs « sémantiques » notamment)

# Moteurs de recherche scientifiques

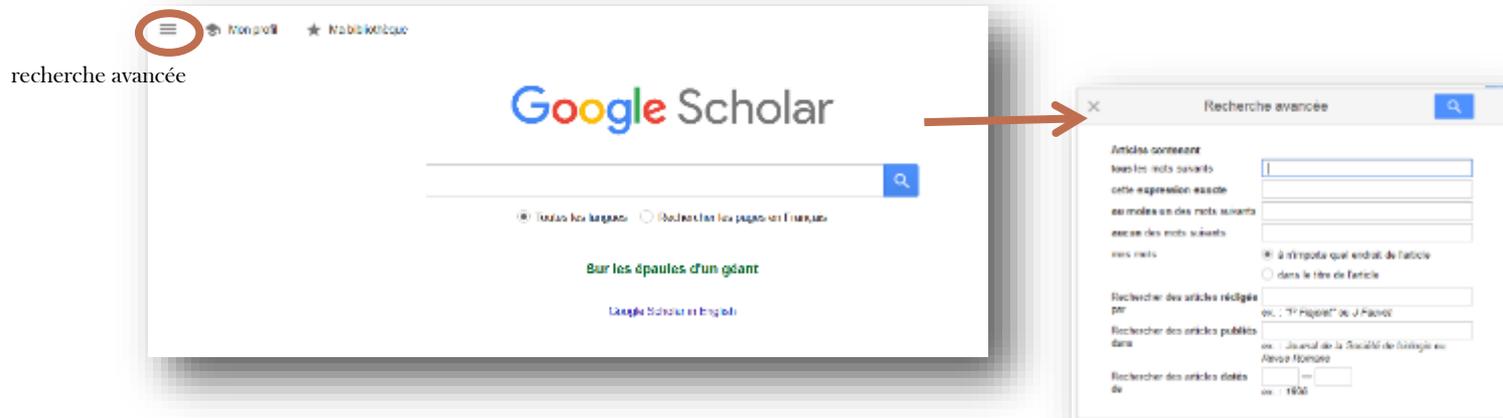
## Moteurs de recherche scientifiques : typologies

Diversité des moteurs. Plusieurs critères possibles :

- **Le contenu des ressources**
  - Moteurs multidisciplinaires
    - [Google Scholar](#), [Scirus](#), [FreeFullPDF](#)...
  - Moteurs disciplinaires
    - [GetInfo](#), [Jurn](#)...
- **L'accessibilité des ressources**
  - Moteurs spécialisés dans les ressources du Libre Accès
    - [OAISTER](#), [Scientific Commons](#)
  - Moteurs spécialisés dans la littérature scientifique commercialisée
    - [Scirus](#)...
- **Le type de documents, de ressources**
  - Moteurs de recherche :
    - de thèses
    - de brevets
    - d'articles
    - de données brutes
- **La nature de l'outil**
  - Véritable moteur de recherche indépendant
    - [Google Scholar](#), [Scirus](#)
  - Moteur personnalisable
    - [Jurn](#),
  - Portail scientifique
    - [Science.gov](#)

# Moteurs de recherche scientifiques

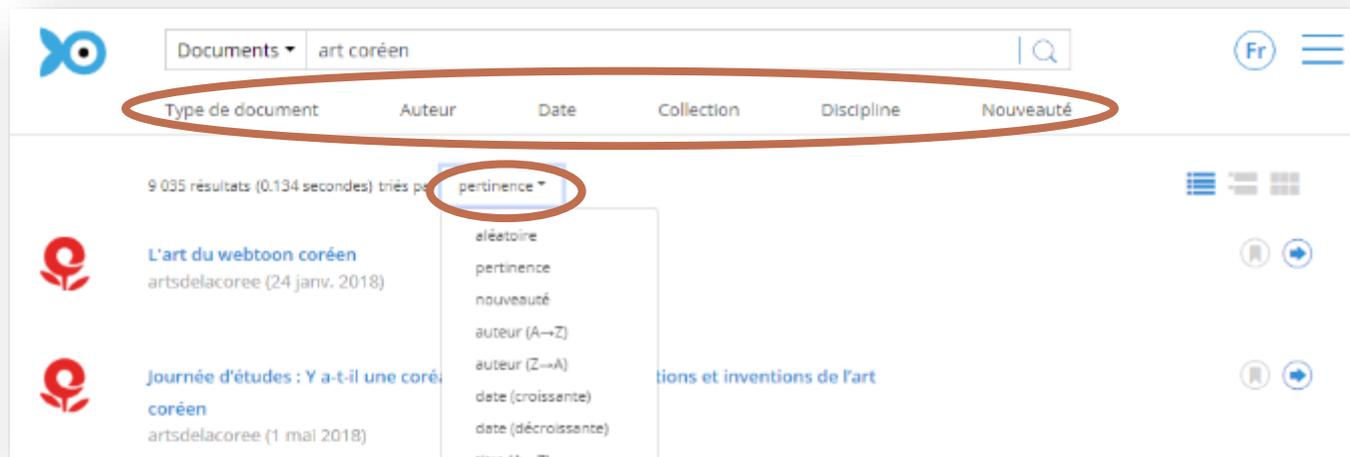
- **Google Scholar** : <https://scholar.google.fr/> (recherche avancée)  
2004, métamoteur  
articles scientifiques approuvés ou non par des comités de lecture (plupart des revues académiques en ligne, en accord avec les éditeurs), thèses, citations... ; prend également en compte les sites internet des auteurs et des laboratoires, les archives ouvertes, les réseaux sociaux académiques  
300 M. de références ?



- + versions d'un même articles (différentes versions, avec mise en avant de la version en accès libre)
  - + système d'alertes (utile aussi pour les alertes sur les citations de ses propres articles *via* le Google scholar *profile*)
  - + recommandation d'« autres articles » (articles associés)
  - surtout pour des références d'articles (peu d'ouvrages), anglophones (peu d'autres langues)
  - couverture pas connue précisément (liste des journaux, disciplines, types de documents, mises à jour) → exhaustivité et fraîcheur de la recherche ?
  - premiers résultats : souvent articles très cités (popularité cf. PageRank) → place des nouveaux articles encore peu cités ?, et limiter à 1 000 réponses ?
  - indexation automatique (auteurs, date, mots-clés) → recherche en texte libre, sans thésaurus
  - pas toujours accès à l'ensemble des résultats trouvés (1 000 premiers)
  - accès au texte intégral pas systématique (indexation notamment de portails d'éditeurs accessibles sur abonnement)  
→ voir avec les bibliothèques et les laboratoires pour les abonnements
  - nombre de citations peu fiable, cf. l'exemple de [Ike Antkare \(C. Labbe\)](#) → ne peut servir d'indicateur
- un outil moins efficace que les bases de données payantes (interface d'interrogation, contenu, [H. Basset](#) et [C. Tisserand-Barthole](#))

# Moteurs de recherche scientifiques

- **Isidore** : <https://isidore.science/>  
2009, TGE Adonis - CNRS (acronyme pour Intégration de services, Interconnexion de DOnnées de la Recherche et de l'Enseignement) ; nouvelle version en 12/2018, TGIR Huma-Num  
moteur de données francophones en SHS prioritairement en accès ouvert (bases de données, fonds numérisés, carnets de recherche, actualités, sites internet..., notamment HAL, OpenEdition Journals, Calenda et Hypotheses.org)  
5,3 M. de ressources et 6 600 collections, classées selon plusieurs référentiels (dont RAMEAU)  
recherche par documents, collections et sujets  
opérateurs booléens : ET (AND), OU (OR), SAUF (NOT), PROCHE(n.) (NEAR(n.)), mais fonctionne mal  
recherche avancée  
compte personnel avec services complémentaires (alertes par mail ou flux RSS)  
+ nombreux filtres (type de document, auteur, date, collection, discipline, nouveauté) et classements *a posteriori*



- + indexation (fiches auteurs et sujets) avec possibilité de rebonds ex. : [art chinois](#)
- des recherches par auteur pas fiables
- des possibilités de rebonds, sur chaque fiche de référence bibliographiques (par les mêmes auteurs, sur les mêmes sujets, sur les mêmes disciplines, sujets proches) pas toujours pertinentes

# Moteurs de recherche scientifiques

- Microsoft Academic : <https://academic.microsoft.com/>  
nouvelle version : 2017 ; 210 M. de références
  - + plus de filtres que sur Google scholar
  - peu d'explications sur le fonctionnement (pas de moteur de recherche avancée, pas d'opérateurs booléens, filtres *a posteriori*)
  - indexation automatique (auteurs, date, mots-clés), recherche en texte libre (recherche « sémantique »), sans thésaurus seulement en complément de Google scholar (des résultats moins nombreux et sans plus d'explications)
- WorldWideScience : <http://worldwidescience.org/>  
2007, réalisé par un consortium de 17 organismes de l'IST et réseaux
  - + portail mondial scientifique multilingue (recherche fédérée et technologie de traduction)
  - 80 bases de données et portails de 70 pays
  - + nombreux filtres, y compris un filtre visuel
- Science.gov : <http://www.science.gov/>  
portail des ressources officielles scientifiques américaines (annuaire et moteur de recherche)  
60 bases de données, 2 200 sites, 200 M. de pages
  - + nombreux filtres, y compris un filtre visuel
- Semantic scholar : <https://www.semanticscholar.org/>  
42 M. de références, plutôt informatique et médecine  
indexation sémantique automatique (auteurs, références, figures et sujets)
- CiteSeerX : <http://citeseerx.ist.psu.edu>  
portail de ressources en informatique et IST
- ScienceResearch : <http://scienceresearch.com/scienceresearch/>  
métamoteur (portails comme Science.gov, WorldWideScience et autres sources comme PubMed...)
  - + clusterisation des résultats
- FreeFullPDF : <http://www.freefullpdf.com/>  
80 M. de publications scientifiques en PDF (articles, brevets, posters, thèses), 15 000 sources
  - limité aux 100 premières réponses (Google CSE)

pour aller plus loin, notamment des moteurs de recherche disciplinaires,

cf. [M.-L. Malingre, A. Serres et Florence Thiault, \*Rechercher l'information scientifique : plateformes et outils spécialisés\* \(2018\)](#)

# Moteurs de recherche scientifiques

des nouveaux venus (2018) : les « méga-index »

- **Dimensions** : <https://app.dimensions.ai/discover/publication>  
2018, société Digital Science  
dans une logique de données liées (*machine learning* et traitement automatique des langues), moteur rassemblant : 140 M. de publications, financements, essais cliniques, brevets, etc.  
citations, métriques alternatives  
contenu partiellement réservé aux comptes institutionnels  
des fonctionnalités plus importantes que Google scholar (filtres, visualisations)
- **The Lens** : <https://www.lens.org>  
2018, Queensland University Technology et société Cambia  
dans une logique de données liées (*machine learning* et traitement automatique des langues), moteur rassemblant 114 M. brevets et 188 M. articles scientifiques  
+ nombreux filtres de recherche
- **1findr** : <https://1findr.1science.com/home>  
société 1science (Elsevier)  
95 M. de références d'articles *peer-reviewed*, dont 27 M. d'articles en *open access*  
métriques alternatives  
deux versions : *free edition* et *institutional edition*
- **Scilit** : <https://www.scilit.net/>  
MDPI (éditeur suisse d'open access)  
120 M. de références, dont 24 M. en texte intégral (à partir de CrossRef, PubMed, etc.)

# Moteurs de recherche scientifiques

## pour les personnels de l'ESR

- ISTEEX : <https://www.istex.fr>  
développé et géré par l'INIST-CNRS (CNRS, ABES, Couperin et CPU *via* l'université de Lorraine)  
Initiative d'excellence de l'Information Scientifique et Technique (projet ANR)  
! réservé à la communauté de l'enseignement supérieur et de la recherche française (ayant-droits)  
accès pérenne à des corpus scientifiques rétrospectifs de [plus de 20 éditeurs](#) (Elsevier, Springer, Wiley, Brill, etc.)  
21 M. de documents, 9 000 revues, multidisciplinaire  
nombreux services associés (extensions de navigateurs, outils de TDM - *text and data mining*, API, etc.)

### ISTEX

extension ISTEEX : <https://addons.istex.fr/>

détecte les ressources ISTEEX sur une page web et affiche un lien direct vers la plateforme

! pour une reconnaissance sur Google scholar : besoin d'activer ISTEEX *via* les options de l'extension



[CITATION] Les bons 'tuyaux' du marché de l'emploi  
N Jean-Philippe - Internet peut-il faire de l'économie de la qualité'un ..., 2001  
☆ 96 Cité 2 fois Autres articles

[HTML] Les bons «tuyaux» du marché de l'emploi: Internet peut-il faire de «l'économie de la qualité» un marché? scienencedirect.com  
JP Neuville - Sociologie du travail, 2001 - Elsevier ISTEX

Résumé Il est fréquent de lire que l'arrivée du marché de l'emploi sur internet permet d'optimiser le processus de rencontre entre offreurs et demandeurs. Prenant appui sur des enregistrements réalisés au cours d'un salon de l'emploi, l'auteur propose des conclusions plus nuancées. Il montre que malgré la pluralité des dispositifs techniques, on observe un même processus de rationalisation qui consiste à accumuler et standardiser l'information, puis à effectuer des opérations automatisées de tri. Il constate de plus que pour «échapper» ...  
☆ 96 Cité 15 fois Autres articles Les 3 versions

Résultats de recherche les plus pertinents Voir tous les résultats

exemple sur Google scholar

# Moteurs de recherche scientifiques

- trouver des chercheurs / des experts
    - bases de données bibliographiques
    - plateformes de dépôts (ex. : archives ouvertes)
    - sites institutionnels (annuaires)
    - réseaux sociaux (cf. *infra*)
    - moteurs de recherche *ad hoc*
      - ScanR : <https://scanr.enseignementsup-recherche.gouv.fr/>  
MESR : moteur de la recherche et de l'innovation en France : + 15 000 acteurs publics et privés (équipes et laboratoires)  
nombreux filtres (structures, tutelles, thématiques, projets, etc.)
      - AuthorMapper : <https://www.authormapper.com/>  
Springer (et données uniquement Springer)  
localisation d'équipes, de thèmes, d'experts....
      - AMiner : <https://aminer.org/>  
sources ?, mise à jour ?
      - Scholar Universe : <http://www.scholaruniverse.com/>  
RefWorks et COS (données ProQuest)  
mise à jour ? ( © 2011)
- pas toujours satisfaisants  
- pas exhaustifs  
- pas de dédoublonnage des profils

# Moteurs de recherche scientifiques



Unité Régionale de Formation à l'Information Scientifique et Technique  
de Bretagne et des Pays de la Loire

Stages ●●● Formations

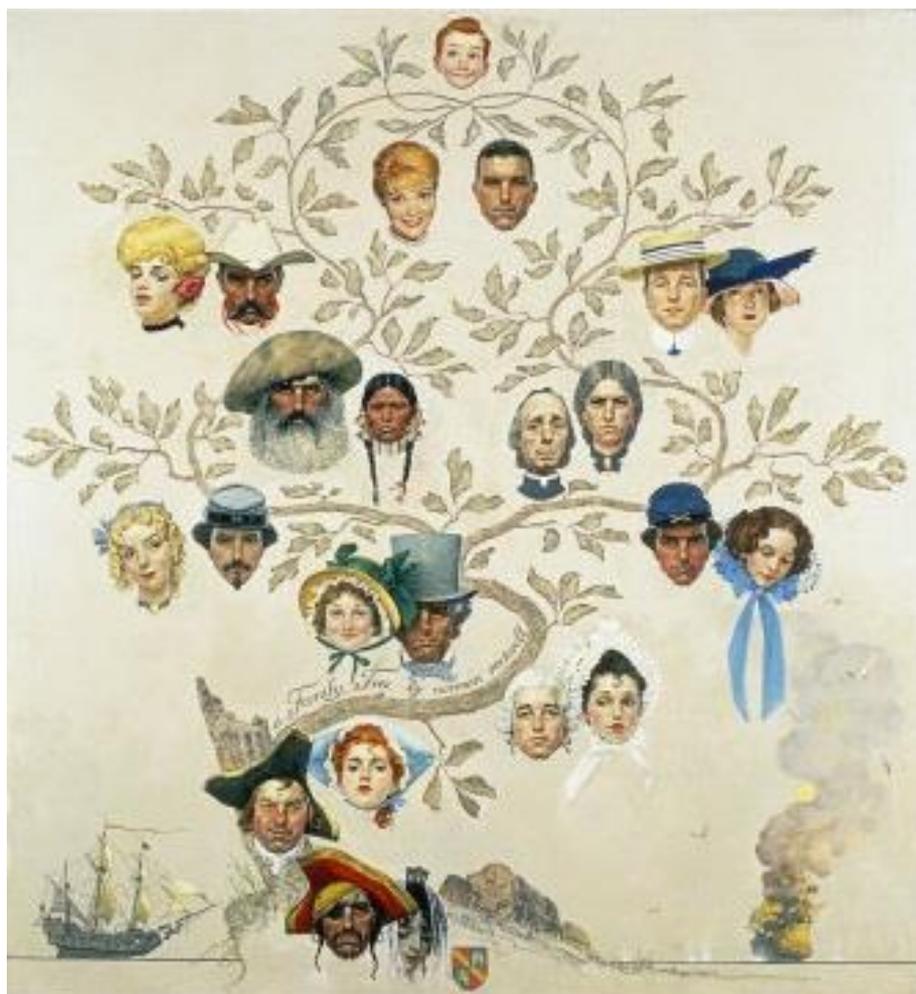
**Rechercher l'information  
scientifique : plateformes et  
outils spécialisés**

Marie-Laure Malingre, Alexandre Serre,  
Florence Thiault  
URFIST de Rennes

 creative commons

30 novembre 2018

pour aller plus loin : cf. [M.-L. Malingre, A. Serres et Florence Thiault, \*Rechercher l'information scientifique : plateformes et outils spécialisés\* \(2018\) et \*carte associée\*](#)



Sites internet

# Sites internet

<b>Documents</b>	sites et portails internet (y compris catalogues en lignes, revues en lignes, corpus, bases de données) liste de sites plus rarement sections ou pages de site, dossiers internet
<b>Recherche</b>	<b>RECHERCHE THEMATIQUE</b> alphabétique/arborescente/moteurs de recherche portant généralement sur les notices descriptives et non sur les sites sélectionnés « sourcing » (identification de sources)
<b>Périmètre</b>	information sélectionnée, organisée et hiérarchisée : identification de sites spécialisés ou localisés dans un pays donné (validée voire commentée : répertoires) généralistes thématiques et spécialisés
<b>Outils</b>	annuaires généralistes répertoires sélectifs (signets) outils thématiques pages de liens
<b>+/-</b>	<b>+</b> identification d'une liste de sites (domaines, pays...) <b>!</b> mises à jour ? (fermeture de Dmoz, Signets de la BnF et signets des universités début 2017) <b>!</b> liens commerciaux

# Annuaire généraliste

- international

- Curlie : <http://curlie.org/>  
anciennement Dmoz (*open directory project*)  
annuaire de sites maintenu par une communauté de rédacteurs bénévoles  
3,5 M. de sites, 1 M. de catégories, 92 000 rédacteurs bénévoles, multilingue  
Français : <http://curlie.org/World/Fran%C3%A7ais/> (227 000 sites, 20 000 catégories en 2017)
- Virtual Library (Vlib) : <http://vlib.org/>  
premier répertoire du web (1991), association à but non lucratif
- Best of the web (BOTW) : <http://botw.org/>  
soumission partiellement payante  
+ indication de blogs et forums

- France

- Les annuaires : <http://www.lesannuaires.com/>
  - + annuaires régionaux et des pays d'Europe
  - + repérage d'outils spécialisés
- Annuaire 1000 sites utiles pour l'entreprise : [http://www.lyon.cci.fr/jcms/s-informer/reperer-l-information-utile-sur-internet-1-000-sites-d\\_8944.html](http://www.lyon.cci.fr/jcms/s-informer/reperer-l-information-utile-sur-internet-1-000-sites-d_8944.html)  
CCI de Lyon

# Répertoires de signets universitaires

- France

- Signets de la BnF [ † ] : <http://signets.bnf.fr/accueil.html>  
7 300 sites, 16 000 liens, 1 150 rubriques  
classement selon le classement des salles de lecture du site François-Mitterrand  
accès alphabétique, thématique , moteur de recherche, nouveautés  
fermé fin 2016 ; dernière version accessible sur la [Wayback machine](#)
- Signets des universités [ † ] : <http://www.signets-universites.fr/>  
fermé début 2017 ; dernière version accessible sur la [Wayback machine](#)
- pour des signets disciplinaires, voir par exemple
  - [BIU Cujas](#) (droit)
  - [BIU Santé](#) (santé)

- international - universités

- Academicinfo : <http://www.academicinfo.net/subject-guides>  
25 000 ressources, nombreuses ressources sur les Etats-Unis, mais pas exclusivement

# Informations sur les sites

- informations factuelles

- Alexa : <http://www.alexa.com/topsites>  
Amazon  
classement des sites selon leur *traffic rank* (usagers + pages vues) : top 500 mondial et par pays, catégories
- SimilarWeb : <http://www.similarweb.com>

- bases « Whois »

- ! mise à jour des données

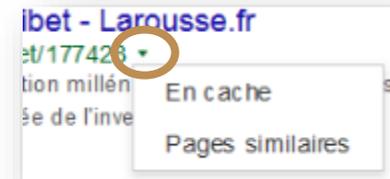
- registres : organismes gérant les noms de domaines
    - AFNIC (Association française pour le nommage internet en coopération) <http://www.afnic.fr/>  
registre des noms de domaines .fr et .re (La Réunion) ([Whois](#))
  - interfaces de recherche
    - WHOIS : <http://whois.domaintools.com/>  
base de noms de domaines

# Informations sur les sites

- date de publication d'un site / d'une page web [B. Foenix-Riou, 2012]
  - bases « Whois » (cf. *supra*)
  - « archives du web » (cf. *infra*)
  - recherche par date des moteurs : date de publication estimée / d'indexation par les robots
    - ! recouper les différentes informations

- historique

- cache des moteurs de recherche (notamment Google)
  - instantané de la page telle qu'elle était affichée lors de l'indexation par le moteur (équation : cache: )
  - + peut permettre de résoudre une erreur 404 (*page not found*)

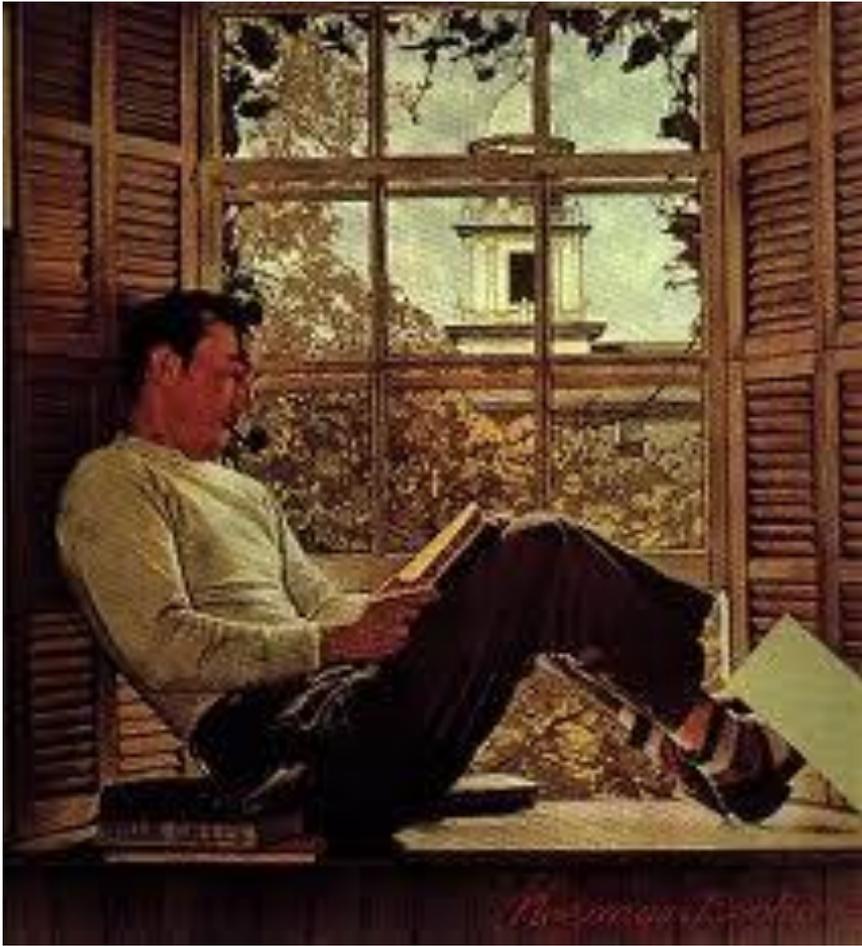


- archives du web
  - Internet Archive – Wayback Machine : <http://archive.org/web/web.php>
    - Internet Archive (organisme à but non lucratif)
    - « archives du web »
    - archivage de 650 MM. de pages web visibles depuis 1996
    - recherche de préférence par l'URL (n'est pas un moteur de recherche dans le temps)
    - + peut permettre de retrouver des pages supprimées, d'anciennes versions de site (ex. : [BnF](#))
    - ! liens brisés, images supprimées et délai d'archivage
    - peu intéressant pour les données du web 2.0 (réseaux sociaux, etc.)
    - pour la France, penser également au [dépôt légal du web / archives de l'internet à la BnF](#) ([où le consulter ?](#))
  - Time travel : <http://timetravel.mementoweb.org>
    - métamoteur sur une vingtaine d'archives du web

Voir également l'extension de navigateur *Resurrect pages*

# Informations sur les sites

- sites similaires
  - Google : related:url ou « pages similaires » de la liste de résultats (  à côté du lien vert)
    - ! attention construire sur le modèle : `[related:google.fr]`
    - visible également via l'outil TouchGraph : <http://www.touchgraph.com/seo> (10 premiers sites similaires)
  - SimilarSites : <http://www.similarsites.com/>
  - SimilarSiteSearch : <http://www.similarsitesearch.com/>
- identifier des pages de liens sélectionnés
  - intérêt des pages de liens (liens / bookmarks / signets) et des portails (portails / annuaires / répertoires) avec une requête de type : intitle:, inurl:
    - ex. [`« réchauffement climatique » intitle:liens OR intitle:sites OR inurl:liens OR inurl:sites`] permet de trouver des pages contenant soit « liens » soit « sites » dans le titre de la page ou l'URL et portant sur le « réchauffement climatique »
  - chercher également des listes (de sources, d'experts), ex. Top 10, etc.
- liens entrants et sortants
  - [aléatoire selon les moteurs] link:url : liste des pages web qui proposent un lien vers une URL
  - outils spécifiques webmasters : Bing webmaster tools, Google Analytics



# Bases de données bibliographiques

# Bases de données bibliographiques

<b>Documents</b>	références bibliographiques de livres, périodiques, articles, rapports, congrès, thèses... informations sur un auteur, un directeur de thèse informations sur les travaux d'un établissement, laboratoire... réseaux de co-auteurs éléments de bibliométrie (localisation et/ou accès au texte intégral)	
<b>Recherche</b>	recherche par mots-clés, parfois par index spécifique à chaque outil présence généralement d'une recherche avancée et de filtres ( <i>a priori</i> et/ou <i>a posteriori</i> )	
<b>Périmètre</b>	disciplinaire pluridisciplinaire littérature blanche (édition commerciale : livres, revues, bases de données...) littérature grise (édition non commerciale : thèses, mémoires, rapports, articles non publiés...)	
<b>Outils</b>	catalogues de bibliothèques (ressources possédées par un/des organismes documentaires) bibliothèques numériques bases de données bibliographiques (signalement de l'information produite, indépendamment de son lieu de conservation ; souvent dépouillement d'articles)	} avec ou sans accès au texte intégral
<b>+ / -</b>	➤ recherche documentaire (références de documents ou documents eux-mêmes) ! pas d'exhaustivité ! mises à jour ? penser également aux catalogues d'éditeurs et aux sites commerciaux (comme Amazon), notamment pour les nouveautés et les annonces de sortie	

# Catalogues de bibliothèques

- France

- SUDOC : <http://www.sudoc.abes.fr>

ABES

catalogue collectif des bibliothèques et centres de documentations de l'enseignement supérieur et de la recherche

12 M. de références

+ localisation

- pas d'exhaustivité des collections des établissements concernés, concerne peu d'établissements de recherche (CNRS...)

- BnF catalogue général : <http://catalogue.bnf.fr>

catalogue général de la BnF, à compléter par d'autres catalogues pour certains documents spécialisés (<https://www.bnf.fr> > Explorer > Chercher un document)

12 M. de références bibliographiques et 5 M. de notices d'autorité

+ dépôt légal français

- catalogage parfois plus tardif et moins de langues étrangères que SUDOC

- CCFr : <http://ccfr.bnf.fr>

catalogue collectif de France, permettant de localiser + 30 M. de documents

répertoire décrivant + 5 000 bibliothèques et 3 000 fonds

+ fonds anciens des bibliothèques françaises rétroconverties

- peu de critères de recherche, lenteurs

! : pas/peu de dépouillement de périodiques (→ autres outils)

indexation aléatoire (périodes, type de documents...)

ressources numériques peu prises en compte (→ autres outils)

# Catalogues de bibliothèques

- international
  - WorldCat : <http://www.worldcat.org/>  
OCLC
    - + plus grand catalogue mondial (2 MM. de références, dont 71 M. de bib. françaises dont SUDOC, BnF...)
    - + localisation des documents
    - références majoritairement anglo-saxonnes
    - pas exhaustif (démarche volontaire des établissements, erreurs et mises à jour)
  - KVK : <http://kvk.bibliothek.kit.edu> (version anglaise)
    - + et - métacatalogue (WorldCat, catalogues collectifs et/ou de bibliothèques nationales français, anglais, espagnols, allemands, italiens..., bibliothèques numériques, archives ouvertes) : peu de critères de recherche
    - + possibilité de choisir les moteurs interrogés
  - Library of Congress : <http://catalog.loc.gov/>

! : pas/peu de dépouillement de périodiques (→ autres outils)  
indexation aléatoire (périodes, type de documents...)  
ressources numériques peu prises en compte (→ autres outils)

# Catalogues de bibliothèques et archives

- archives et manuscrits
  - Calames : <http://www.calames.abes.fr/pub/>  
archives et manuscrits de l'enseignement supérieur  
BU, établissements de recherche, sociétés savantes
  - CCFR manuscrits :  
[http://ccfr.bnf.fr/portailccfr/jsp/portal/index.jsp?action=jsp/ccfr/search/search\\_form\\_ccfr\\_manuscripts.jsp?menuSelected=search\\_simple&portalMenu=ccfr\\_menu\\_manuscripts](http://ccfr.bnf.fr/portailccfr/jsp/portal/index.jsp?action=jsp/ccfr/search/search_form_ccfr_manuscripts.jsp?menuSelected=search_simple&portalMenu=ccfr_menu_manuscripts)  
catalogue collectif de France comprenant CGM (catalogue général des manuscrits),  
BnF archives et manuscrits et PALME (manuscrits littéraires français du XX<sup>e</sup> s.)
  - voir également le portail national : <https://francearchives.fr/>  
et le portail européen des archives : <http://www.archivesportaleurope.net/>

**! : catalogues non exhaustifs**

# Bases de données bibliographiques

- pluridisciplinaires
  - JSTOR : <http://www.jstor.org/>  
archives de revues de sciences humaines et sociales  
12 millions de références, 2 300 revues et 5 000 ouvrages  
principalement en anglais et partiellement payant
  - Ingenta : <http://www.ingentaconnect.com/>  
5,9 M. de références, 13 000 publications
  - EZB (*Elektronische Zeitschriftenbibliothek*) : <http://rzblx1.uni-regensburg.de/ezeit/about.phtml?bibid=AAAAA&colors=7&lang=en>  
répertoire de revues électroniques en ligne, 130 000 titres (accès libre ou non)

**! : pas/peu de références de documents autres qu'articles  
accès au texte intégral pas systématique et souvent payant**

**→ voir avec les bibliothèques et les laboratoires**

# Portails de revues

- bases en accès libre
  - Persée : <http://www.persee.fr>  
portail de diffusion de publications scientifiques (revues scientifiques, publications et séries) en sciences humaines et sociales  
vocation notamment patrimoniale *via* la numérisation des archives de revues savantes françaises  
300 collections, 755 000 documents en texte intégral et en accès libre (revues, livres, colloques, sources primaires, etc.)  
participe à la plateforme canadienne Érudit : <http://www.erudit.org/revue/>  
recherche par moteur, discipline et feuilletage  
+ nombreux filtres sur les 3 types de pages de résultats (documents, auteurs, illustrations)
  - OpenEdition Journals (anciennement Revues.org) : <http://journals.openedition.org/>  
plateforme de revues en sciences humaines et sociales, notamment de revues nativement numériques  
500 revues scientifiques  
recherche par moteur et par feuilletage
- base en accès restreint
  - Cairn : <https://www.cairn.info/>  
plateforme de publications de revues en langue française  
500 revues, mais également magazines, ouvrages collectifs et encyclopédies et ouvrages de références  
recherche par moteur, discipline et feuilletage

**! quelle complémentarité entre les programmes ?**

à consulter : Mir@bel : <http://www.reseau-mirabel.info/>

+ réservoir d'informations sur les revues (dont accès au texte intégral), mais pas exhaustif (5 400 revues, notamment sur Cairn, Érudit, Persée, OpenEdition Journals, Sign@l)

# Bibliothèques numériques

- France

- Gallica : <http://gallica.bnf.fr/>

- réalisée par la BnF mais rassemble des collections d'institutions partenaires (bibliothèques, centres de recherche : BIUS, BIU Cujas, CNAM, INHA...) et des partenaires commerciaux

- 5 M. documents (textes, images, presse...)

- accès par types de documents et par thèmes

- ! interrogation du texte intégral pas possible pour l'ensemble des fonds

- Europe

- Europeana collections : <http://www.europeana.eu/portal/>

- projet européen réunissant des institutions européennes

- 58 M. documents (textes, images, presse...)

**! : ouvrages en texte intégral = essentiellement libres de droit  
numérisation de documents anciens sans édition critique  
accès possible à différents formats (PDF, texte, EPUB...)**

# Bibliothèques numériques

- Google books (Google livres) : <https://books.google.fr/?hl=fr>
  - projet de numérisation et d'OCRisation de Google, dont « programme partenaires » et « projet bibliothèque »
  - 20 M. d'ouvrages scannés (03/2012)
  - interrogation plein texte
  - + [recherche avancée](#)
  - surreprésentation des ouvrages anglo-saxons
  - pour les ressources académiques et notamment des articles scientifiques, voir Google Scholar (deux programmes de mise en ligne différents)

**! : accès au texte intégral partiel (aperçu), voire inexistant**

# Bibliothèques numériques

- World Digital Library / Bibliothèque numérique mondiale : <http://www.wdl.org/fr/>  
projet porté par l'UNESCO  
plutôt une vitrine  
navigation par lieu, période, thème, type d'élément et institution
- Archive.org : <https://archive.org/details/texts>  
20 M. de textes  
  
à compléter par le projet Open Library : <http://openlibrary.org/>  
projet de créer une page web pour chaque livre publié dans le monde  
1 M. de livres gratuits
- Project Gutenberg : <http://www.gutenberg.org/>  
plus ancien projet de numérisation (e-books gratuits)  
58 000 ouvrages (! livres de droit aux États-Unis)  
+ plusieurs formats de téléchargement
- HathiTrust : <http://www.hathitrust.org/>  
bibliothèque numérique de bibliothèques universitaires américaines (120 partenaires)  
17 M. de livres dont 40 % dans le domaine public (50 % en anglais) et 450 000 titres de périodiques

à compléter par [liste Wikipédia](#)

- plateformes de contenus  
livres dans le domaine public ou œuvres originales
    - Numilog : [https://www.numilog.com/Pages/Livres/EbookGratuit.aspx?grid\\_page=0](https://www.numilog.com/Pages/Livres/EbookGratuit.aspx?grid_page=0)
    - In libro veritas : <http://www.inlibroveritas.net/>
    - Many Books : <https://manybooks.net/>
    - des libraires proposent parfois aussi une offre gratuite : [Decitre](#), [ePage](#), [Feedbooks](#) ([domaine public](#), [œuvres originales](#))

pour aller plus loin : [\*20 Best Sites to Download Free Books in 2019\*](#)
  - moteurs de recherche
    - The online books page : <http://digital.library.upenn.edu/books/search.html>
    - Nos livres : <http://www.noslivres.net/>  
Bibliothèque électronique du Québec, Bibliothèque numérique romande, ÉFÉLÉ, Projet Gutenberg, Bibliothèque Russe et Slave
- + penser également à une recherche sur un moteur généraliste filetype: (pdf, mobi, etc.)

# Parutions et critiques

- **JournalTOCs (*Journal Tables of Contents*):** <http://www.journaltoCs.ac.uk/>  
Heriot-Watt University d'Edimbourg  
plus grande collection de tables de matières de revues scientifiques  
33 000 revues, dont 15 000 en *open access*  
alertes lors des nouveaux numéros
- **Sign@l :** <http://doc.sciencespo-lyon.fr/Signal/>  
service de la documentation de SciencesPo Lyon  
signalement des contenus de 433 périodiques en sciences humaines et sociales
- **Parutions.com :** <http://parutions.com/>  
site animé par un collectif de chercheurs, de journalistes  
classement par discipline  
comptes rendus
- **Nonfiction.com :** <http://www.nonfiction.fr/>  
site animé par un collectif de chercheurs, de journalistes  
classement par discipline  
comptes rendus
- **Recensio.net :** <http://www.recensio.net/front-page-fr>  
plateforme de recensions d'ouvrages pour la recherche historique européenne
- voir également les sites internet des éditeurs et libraires ([Le comptoir des presses d'universités](#), [Amazon](#), [GibertJoseph](#), [Le choix des libraires...](#)), notamment les nouveautés et suggestions automatiques



Autres  
bases de données textuelles

# Autres bases de données textuelles

<b>Documents</b>	données factuelles et/ou bibliographiques tout type de documents, dont articles scientifiques, pré et post-publications, littérature grise... calendriers et offres d'emploi
<b>Recherche</b>	spécifique à chaque outil présence généralement d'une recherche avancée
<b>Périmètre</b>	disciplinaire pluridisciplinaire littérature blanche (édition commerciale : livres, revues, bases de données...) littérature grise (édition non commerciale : thèses, mémoires, rapports, articles non publiés...) jeux de données (statistiques, <i>open data</i> , données de la recherche)
<b>Outils</b>	moteurs de recherche spécialisés moissonneurs OAI plateformes de dépôt et publication (archives ouvertes, réseaux sociaux académiques)
<b>+ / -</b>	mise en ligne et accessibilité parfois dès la publication et/ou la soumission à des éditeurs peut être la version <i>pre-print</i> et non la version éditeurs encore plutôt orientées articles / STM / anglophones et non ouvrages / SHS / francophones

# ALTERNATIVE ACCESS TO ELSEVIER ARTICLES



## Install browser plug-ins

[Open Access Button](#)

[Unpaywall](#)



## Search for open access copies

[Google Scholar](#)

[PubMed Central](#)



## Get it from the Library

[UC-eLinks](#)

[Contact your librarian](#)



## Use your network

[Contact the author](#)

[ResearchGate](#)

**Note: The UC Libraries do not endorse using Sci-Hub for article access.**

# Open access ? archives ouvertes ?

+ 20 % des articles publiés par an : *open access* (en français : accès « ouvert »)

- 8 % : revues scientifiques ouvertes (*gold open access*)

publication d'articles dans des journaux électroniques à comités de lecture (*peer-reviewed*) immédiatement et librement accessibles, sans restriction d'utilisation, c'est-à-dire sans barrière financière (ex. : pas d'abonnement, pas d'embargo), légale (ex. : pas de licences restrictives) ou technique (ex. : possibilité d'utiliser le protocole d'interopérabilité OAI)

- 12 % : archives ouvertes (1,5 M. d'articles ?) (*green open access*)

archives ouvertes : réservoirs de dépôt de documents créés par des réseaux de chercheurs ou des institutions pour favoriser l'auto-archivage (sans contrôle) (*open archives*) tout en en assurant la pérennité

+ 2 000 réservoirs dans le monde (France : 65)

**! : de plus en plus de chercheurs déposent leurs productions non pas sur des plateformes d'archives ouvertes mais sur leurs blogs et/ou leurs profils de réseaux sociaux académiques ([Academia](#), [ResearchGate](#)...)**

- types de documents possibles

*preprints* et prépublications (manuscripts soumis pour publication, avant évaluation par un comité de lecture)

*postprints* ou postpublications (manuscripts révisés par les pairs et acceptés pour publication, mais avant leur mise en page par l'éditeur)

articles en version éditeur (manuscripts révisés et validés par les pairs et mis en page par l'éditeur)

littérature grise (thèses et HDR, colloques, manuscrits non publiés ailleurs...)

données de la recherche

production multimédia

pour aller plus loin sur l'*open access* : [Marie Lebert \(2016\)](#)

sur la recherche d'informations en *open access* :

[A. Bouchard \(2017\)](#) et [T. Hameau \(2017\)](#)

# Open access et archives ouvertes

- moteurs de recherche
  - pluridisciplinaires (moissonneurs OAI) - multiressources
    - BASE (Bielefeld Academic Search Engine) : <https://www.base-search.net/>  
bibliothèque de l'université de Bielefeld  
145 M. de documents (75 % accessibles en texte intégral), 7 000 sources
      - + [recherche avancée](#)
      - + intègre une recherche « synonymes multilingues »
      - 60 % des références sont en texte intégralà compléter par l'[Open knowledge maps](#) pour une représentation graphique des résultats
    - CORE (*C*Onnecting *R*Epositories) : <http://core.kmi.open.ac.uk/search>  
projet JISC / The Open Library  
135 M. notices bibliographiques
      - + suggestions d'articles
    - OAIster : <http://oaister.worldcat.org/>  
premier moissonneur OAI (moissonneur historique)  
université du Michigan et OCLC, accessible dans WorldCat  
50 M. de références, 2 000 participants (archives ouvertes essentiellement)  
n'indexe pas le texte intégral
    - OpenAIRE (*O*pen *A*ccess *I*nfrastructure for *R*esearch in *E*urope) : <https://explore.openaire.eu/>  
projet européen de dépôt de résultats de recherches financées par l'Union européenne  
25 M. de publications, 1 M. de données de recherche de 14 fournisseurs
    - SHARE : <https://share.osf.io/>  
projet du Center for open science  
180 sources agrégées (dont HAL, arXiv, Zenodo, etc.)

voir également Google scholar <https://scholar.google.fr/> qui présente en priorité les versions accessibles d'un article

# Open access et archives ouvertes

- en cas d'article derrière un *paywall*, essayer de trouver une version en *open access* via
  - des résolveurs de DOI
    - DOAI : <http://doai.io>
  - des extensions de navigateur (! de qualité inégale)
    -  • Google Scholar button  
extension liée au moteur Google Scholar (pour lancer la recherche, sélectionner son titre sur la page)
    -  • Open Access Button : <https://openaccessbutton.org/>  
soutenu par SPARC (coalition de bibliothèques académiques) : moteur de recherche sur le site ; extension de navigateur (pour lancer la recherche, cliquer sur l'extension quand un *paywall* données issues de bases de données spécialisées en *open access* ; exclut les publications hébergées sur les réseaux sociaux académiques ResearchGate et Academia)
    -  • Unpaywall : <http://unpaywall.org/>  
créé par Impactstory (base de données : 50 000 éditeurs et répertoires, 23 M. de références)  
recherche automatiquement des versions gratuites des documents (si présence de l'icône, cliquer dessus pour accéder au document)
    -  • Lazy Scholar : <http://www.lazyscholar.org/>  
interroge Google scholar et des bases spécialisées en *open access*  
recherche automatiquement des versions gratuites des documents (si c'est le cas, apparition d'une barre supérieure dans la page où lien vers la version gratuite, et d'autres fonctionnalités comme métriques, outil de citation, recommandations...)
    -  • Kopernio : <https://kopernio.com>  
Clarivate analytics (la maison du Web of science)  
lié à un compte

# Open access et archives ouvertes

- politique des éditeurs face à l'*open access* (*copyright* et archives ouvertes)

- SHERPA/RoMEO : <http://www.sherpa.ac.uk/romeo/search.php> (anglo-saxons)

- Héloïse : <http://heloise.ccsd.cnrs.fr/> (France)

- DULCINEA : <http://www.accesoabierto.net/dulcinea/> (Espagne)

! ne sont pas exhaustifs, et ne portent pas sur le dépôt sur les réseaux sociaux académiques

**NB** : pour la France : [art. 30 de la loi pour une République numérique](#) (2016, [guide d'application](#))

- politique des financeurs vis-à-vis de l'*open access*

- SHERPA/JULIET : <http://v2.sherpa.ac.uk/juliet/>

- ROARMAP (Registry of Open Access Repository Mandates and Policies) : <https://roarmap.eprints.org/>

! ne sont pas exhaustifs

# Open access et archives ouvertes

- aide au choix d'une revue en *open access*
  - choix d'une revue
    - Quality Open Access Market (QOAM) : <https://www.qoam.eu/>  
aide les chercheurs à choisir une revue en libre accès de haute qualité, à des coûts raisonnables
    - How open is it ? : <https://www.accesoabierto.net/howopenisit/?idioma=en>
    - Open access spectrum evaluation tool : <http://www.oaspectrum.org>  
calcule du degré d'ouverture des revues
  - lutter contre les éditeurs prédateurs
    - Beall's list :  
liste d'éditeurs prédateurs ; liste de référence, arrêtée début 2017, mais archivée sur la [Wayback machine](#)
    - solutions alternatives : Stop predatory journals : <https://predatoryjournals.com/about/> ou le DOAJ : <http://www.doaj.org/> et ROAD : <http://road.issn.org/>
    - Think Check Summit : <http://thinkchecksubmit.org/> (version française : <http://thinkchecksubmit.org/translations/french/>)
- aide au chercheur
  - Dissem.in : <https://dissem.in/>  
2015, association CAPSHS (Comité pour l'accessibilité aux publications en sciences et humanités)  
but : favoriser le dépôt en archives ouvertes, en détectant l'accessibilité des publications  
détermine la présence ou non d'un article sur des plateformes ouvertes et propose au besoin le dépôt sur Zenodo ou HAL
    - ! pas exhaustif
    - ! prend mal en compte les articles publiés dans des revues en *open access*

# Revue et ouvrages ouverts

- répertoires

- DOAJ (*Directory of Open Access Journals*) : <http://www.doaj.org/>  
université de Lünd  
répertoire de journaux en *open access* et en texte intégral, 13 000 journaux dont 10 000 interrogeables au niveau de l'article, 4 M. d'articles
- DOAB (*Directory of Open Access Books*) : <http://www.doabooks.org/>  
OAPEN Foundation  
répertoire de livres *peer-reviewed* en *open access*, 16 000 livres, 300 éditeurs déclarés
- ROAD (*Registry of Open Access scholarly Resources*) : <http://road.issn.org/>  
centre ISSN et UNESCO  
ressources librement accessibles comportant un ISSN (revues, collections, actes de conférences...)  
10 000 ressources

# Revue et ouvrages ouverts

- plateformes et moteurs spécialisés
  - OAPEN Library : <http://oapen.org/home>  
OAPEN Foundation  
plateforme de livres *peer-reviewed* en *open access* - surtout en SHS
  - OpenEdition books : <https://books.openedition.org/>  
CLEO  
plateforme de livres SHS (+ 2/3 en *open access*) : 7 000 livres (objectif : 16 000 en 2020), 90 éditeurs  
services complémentaires pour les bibliothèques et institutions abonnées
  - Paperity : <http://paperity.org/>  
moteur interrogeant 8 000 revues en *gold open access* et en modèle hybride (multidisciplinaire)  
2,3 M. articles
  - JURN : <http://www.jurn.org/>  
Google CSE  
indexe + 5 000 revues en *open access* (arts et humanités principalement)  
répertoire de 3 000 revues
  - Open Science Directory : <http://www.opensciencedirectory.net/>  
EBSCO, bibliothèque de l'université de Hasselt et UNESCO, + 20 000 journaux  
autour de programmes spéciaux pour les pays en développement

# Archives ouvertes

- répertoires
  - OpenDOAR (*Directory of Open Access Repositories*) : <http://v2.sherpa.ac.uk/opensoar/>  
université de Nottingham  
[annuaire de réservoirs d'archives ouvertes](#) (+ 2 600 références)  
[moteur de recherche sur leur contenu](#) (Google CSE, n'affiche que les 100 premières réponses)
  - ROAR (*Registry of Open Access Repositories*) : <http://roar.eprints.org/>  
université de Southampton  
mêmes principes qu'OpenDOAR : répertoire et [moteur de recherche sur leur contenu](#) (Google CSE, n'affiche que les 100 premières réponses)

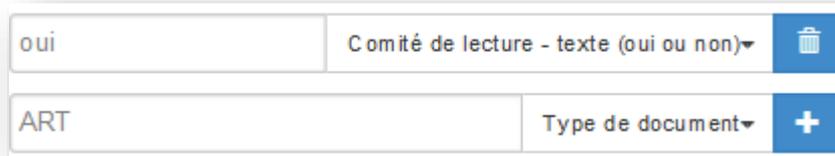
# Archives ouvertes

- réservoirs d'archives ouvertes (disciplinaires et thématiques)
  - arXiv : <http://arxiv.org/>  
première archive ouverte au monde (1991)  
Cornell University, physique, mathématiques, informatique, statistiques  
1,5 M. de *papers*, notamment *preprints*
  - SSRN (Social Science Research Network) : <http://www.ssrn.com/en/>  
Elsevier, sciences sociales  
850 000 articles
  - PMC (PubMed Central) : <http://www.ncbi.nlm.nih.gov/pmc/>  
NIH/NLM (U.S. National Institutes of Health's National Library of Medicine), biomédical et sciences de la vie  
5,3 M. d'articles (notamment postpublications déposées par les revues elles-mêmes)
  - RePEc (*Research Papers in Economics*) : <http://repec.org/>  
sciences économiques - se veut exhaustif (collecte d'autres archives ouvertes)  
2,6 M. de documents, 99 pays
  - CogPrints : <http://cogprints.org/>  
notamment psychologie, neurosciences, linguistique
  - e-Lis (*E-prints in Library and Information Science*) : <http://eprints.rclis.org/>  
sciences de l'information et des bibliothèques  
21 000 documents

pour aller plus loin, voir *Disciplinary repositories* : [http://oad.simmons.edu/oadwiki/Disciplinary\\_repositories](http://oad.simmons.edu/oadwiki/Disciplinary_repositories)

- France

- HAL (Hyperarticles en ligne) : <https://hal.archives-ouvertes.fr/>  
CCSD-CNRS, pluridisciplinaire  
composée de plusieurs collections (génériques, thématiques, institutionnelles) :  
1,7 M. de références, 570 000 documents en texte intégral, de niveau recherche  
*preprints*, articles (47 %), communications (27 %), thèses, HDR...  
! beaucoup de références sans accès au texte intégral  
recherche avancée : <https://hal.archives-ouvertes.fr/search/index>



The image shows a screenshot of the HAL search interface. It features two filter boxes. The first box contains the text 'oui' and a dropdown menu labeled 'Comité de lecture - texte (oui ou non)' with a trash icon to its right. The second box contains the text 'ART' and a dropdown menu labeled 'Type de document' with a plus icon to its right.

Ex. pour repérer uniquement les articles  
des revues à comité de lecture (via [F. Bordignon](#))  
- formulaire de recherche avancée

autres plateformes institutionnelles (exemples) :

Archimer (Ifremer) : <http://archimer.ifremer.fr/>

ProdINRA (INRA) : <http://prodinra.inra.fr/?locale=fr>

OATAO (Open archive Toulouse archive ouverte) : <http://oatao.univ-toulouse.fr/>

→ voir pages France d'[OpenDOAR](#) et [ROAR](#)

- France
  - thèses en préparation voire soutenues
    - Theses.fr : <http://www.theses.fr/>  
application STEP (Signalement des Thèses En Préparation, qui succède au Fichier central des thèses de Nanterre) doit signaler l'ensemble des thèses de doctorat soutenues en France depuis 1985, y compris des thèses de la CGE (confédération des Grandes Écoles, après la fermeture de Thesa en 2013)  
460 000 thèses (dont 380 000 thèses soutenues, 100 000 thèses accessibles en ligne et 76 000 en préparation)  
! informations parfois incomplètes et pas toujours à jour  
! outil non exhaustif : pour les thèses en préparation, notamment, le signalement est facultatif et dépend de chaque établissement  
à compléter notamment par TEL (<https://tel.archives-ouvertes.fr/>, cf. *infra*) pour le texte intégral de thèses qui ne seraient pas concernées par l'application STAR (Signalement des Thèses électroniques, Archivage et Recherche)
  - thèses soutenues (documents)
    - à regarder en priorité : SUDOC (version de soutenance) : <http://www.sudoc.abes.fr>  
référence des thèses françaises toutes disciplines depuis 1972 (et 1982 pour les thèses de médecine et chirurgie dentaire)  
! toutes les données du SUDOC n'ont pas été versées dans theses.fr, même pour les thèses après 1985 ([source](#))
    - à compléter éventuellement par  
BIUS (médecine-odontologie) : [http://www2.biusante.parisdescartes.fr/theses/theses\\_rech.htm](http://www2.biusante.parisdescartes.fr/theses/theses_rech.htm)  
IFE (ex-INRP, éducation et formation) : <http://ife.ens-lyon.fr/vst/Recherches/AccueilTheses.php>  
ANRT : <http://www.diffusiontheses.fr/anrt-catalogue-theses-doctorat.html>
    - guides de la BnF  
*Comment trouver les thèses de doctorat à la BnF*: [http://bnf.libguides.com/theses\\_doctorat?hs=a](http://bnf.libguides.com/theses_doctorat?hs=a)

! : HDR : dépôt facultatif ; signalement non exhaustif

- France

- texte intégral

- TEL (Thèses En Ligne) : <https://tel.archives-ouvertes.fr/>

CCSD

archive ouverte de thèses, déposées soit en auto-archivage par les auteurs (sous-partie de HAL) soit par les établissements passés au dépôt électronique (application STAR)

90 000 thèses et HDR

peut compléter Theses.fr pour trouver des thèses en texte intégral ; les thèses déposées sur TEL sont [progressivement signalées](#) dans le SUDOC

**! : voir également le site et /ou les archives ouvertes de l'établissement de soutenance ou de sa bibliothèque**

notamment Pastel (ParisTech) : <http://pastel.archives-ouvertes.fr/>

Petale (Lorraine) : <http://petale.univ-lorraine.fr/index.html>

cf. [« Trouver des thèses françaises en texte intégral » \(ABES\)](#)

- DUMAS (Dépôt universitaire de mémoires après soutenance) :

<https://dumas.ccsd.cnrs.fr/>

CCSD

15 000 mémoires de M1 et M2 déposés par les établissements

- international
  - texte intégral
    - DART-Europe E-theses Portal : <http://www.dart-europe.eu/>  
LIBER (Ligue des bibliothèques européennes de recherche)  
28 pays européens, 620 universités, 800 000 thèses (dont TEL)
    - OATD (*Open Access Theses and Dissertations*) : <http://oatd.org/>  
moissonneur de thèses et de mémoires (universités et institutions de recherche)  
4,8 M. de références indexées de + 1 100 établissements
    - NDLTD (*Networked Digital Library of Theses and Dissertations*) : <http://search.ndltd.org/>  
100 établissements, 5,5 M. de références de documents (en *open access* ou non)
      - pas toujours très performant (lien à la source...)
    - par pays : voir « [Trouver des thèses étrangères en texte intégral](#) » (ABES)
  - pour aller plus loin
    - Google Scholar <http://scholar.google.fr/>
    - Microsoft Academic : <https://academic.microsoft.com/>
    - la base de données *ProQuest Theses and Dissertations* (PQDT) est payante  
→ voir avec les bibliothèques et les laboratoires

- annonces de colloques, journées d'étude, etc.
  - Calenda : <https://calenda.org/>  
OpenEdition  
calendrier scientifique SHS (majoritairement francophone)  
annonce de colloques, programmes de séminaires et ateliers, appels à contribution, etc.  
41 000 événements signalés
  - Sciencesconf : <https://www.sciencesconf.org/>  
CCSD  
plateforme de gestion de conférence  
7 000 événements signalés  
le cas échéant, propose outre le programme et l'inscription, les communications et autres documents
- littérature grise (non publiée) : communication, poster...
  - plateformes de dépôt d'archives ouvertes  
HAL : <https://hal.archives-ouvertes.fr/>  
recherche avancée : conférences invitées, communications avec ou sans actes, projets européens...
  - web social (blogs, réseaux sociaux académiques...) - cf. *infra*
- publication des actes
  - bases de données bibliographiques (bases de données d'articles, catalogues de bibliothèques...)

- littérature grise

« Tout document dactylographié ou imprimé, produit à l'intention d'un public restreint, en dehors des circuits commerciaux de l'édition et de la diffusion et en marge des dispositifs de contrôle bibliographiques »

(AFNOR)

- GreySource : <http://www.greynet.org/greysourceindex.html>  
sélection de ressources en littérature grise  
classification par thèmes
- OpenGrey Repository : <http://www.opengrey.eu>  
SIGLE (système pour l'information en littérature grise en Europe) : base de données multidisciplinaire européenne, gérée et maintenue par l'INIST-CNRS  
signale les ressources des grands centres européens d'information et de fourniture de documents  
accès à 1 M. de références bibliographiques produites en Europe
- Lara : <http://lara.inist.fr/>  
accès au texte intégral des rapports scientifiques et techniques français (INSERM, INRP...)
- Think Tank search : [http://guides.library.harvard.edu/hks/think\\_tank\\_search](http://guides.library.harvard.edu/hks/think_tank_search)  
Google CSE  
recherche sur les sites de 1 200 *think tanks*, partisans et non partisans
- annuaire de Wordpress : <https://www.worldpress.org/library/ngo.cfm>  
index de *think tanks* et d'organismes de recherche

# Données et statistiques

- moteurs de recherche et répertoires
  - Zanran : <http://www.zanran.com/q/>  
moteur de données et statistiques (PDF, Excel, HTML)  
recherche uniquement en anglais
    - + copies de pages internet et documents qui ont pu être supprimés depuis
  - Knoema : <http://knoema.fr/>  
atlas mondial de données : cartes, infographies, classements...  
accès par catégories, pays...

# Données et statistiques

- institutions

- INSEE : <http://www.insee.fr/fr/>  
pour les autres pays du monde, voir la page [National Statistical Offices](#) de l'ONU
- EUROSTAT : <http://ec.europa.eu/eurostat>  
site de la Commission européenne  
voir également la base de données macroéconomiques AMECO :  
[http://ec.europa.eu/economy\\_finance/ameco/user/serie/SelectSerie.cfm](http://ec.europa.eu/economy_finance/ameco/user/serie/SelectSerie.cfm)
- OCDE : <http://stats.oecd.org/>  
pays de l'OCDE et quelques pays non membres  
arborescence de 22 thématiques, français et anglais
- UNdata : <http://data.un.org/>  
site de l'ONU (UNICEF, UNESCO...)
- RechStat : <http://www.kyxar.fr/~jalac/>  
répertoire des sites des institutions fournissant des statistiques officielles mondiales  
accès par moteur, pays et thème

# Jeux de données (*datasets*)

- jeux de données (*datasets*) [C. Tisserand-Barthole, 2018]
  - données ouvertes de l'*open data* : souvent mises à disposition par des organismes publics, mais pas exclusivement (listes, recensements, etc.)
  - données de la recherche (*research data*) : produites à l'occasion de recherches scientifiques et académiques, et de plus en plus mises à disposition dans une optique de science ouverte (*open science*)
  - des ressources hétérogènes, dispersées et encore très mal indexées par les moteurs de recherche classiques (métadonnées pauvres et non standardisées)
- moteurs de recherche
  - Google dataset search : <https://toolbox.google.com/datasetsearch>  
Google, 2018, encore en version bêta
    - + le seul outil actuellement permettant de chercher à la fois *open data* et données de la recherche
    - une couverture restreinte (uniquement les *datasets* utilisant schema.org)
    - interroge uniquement les métadonnées et non le texte intégral
    - des fonctionnalités de recherche moindre que Google (opérateurs, filtres)

# Données de la recherche

- moteurs de recherche scientifiques généralistes
  - BASE : <https://www.base-search.net/> (type de document : *dataset*)
  - OpenAIRE (*Open Access Infrastructure for Research in Europe*) :  
<https://explore.openaire.eu/search/find/datasets>
  - Mendeley data : <https://data.mendeley.com/>  
Elsevier, 38 sources différentes y compris non Elsevier, 10 M. de références  
plateforme de dépôt de jeux de données et moteur de recherche scientifique (ScienceDirect, Zenodo, etc.)  
opérateurs booléens et filtres
- plateformes de dépôt
  - Zenodo : <http://zenodo.org/>  
2013, soutenu par la communauté européenne et le CERN
  - Figshare : <http://figshare.com/>  
2012, partage de données, de résultats négatifs et de schémas  
2 M. de *datasets*  
[fonctionnalités de recherche](#)
  - Dryad : <http://datadryad.org/>  
2008, à l'origine, plutôt tourné sur la biologie et l'écologie
  - 4TU.Centre for Research Data : <http://researchdata.4tu.nl/en/home/>
  - myExperiment : <http://www.myexperiment.org/>  
partage de schémas de *workflows*, 10 500 membres, 40 groupes, 3 800 *workflows*

# Données de la recherche

- répertoires de plateformes

- re3data (*Registry of research data repositories*) : <http://www.re3data.org/>  
proposé par DataCite, est la version *data* d'OpenDOAR et ROAR  
2 000 entrepôts pluridisciplinaires ou thématiques  
navigation par sujet, type de contenu, pays  
moteur de recherche
- Datacite search : <http://www.re3data.org/>  
proposé par DataCite, est la version *data* d'OpenDOAR et ROAR  
13 M. de *datasets*
- Scholxplorer : <https://dliservice.research-infrastructures.eu/#/>  
2,6 M. de *datasets*  
liens entre les *datasets* et les publications (mais peu satisfaisant)

pour des références de répertoires thématiques,  
cf. [Open access directory](#)

- moteurs de recherche

- Apertio : <https://apert.io/>  
2 000 sites en *open data*  
ressources plutôt anglophones

- portails institutionnels

- Data.gouv.fr : <http://www.data.gouv.fr/>  
plateforme française d'ouverture des données publiques (*open data*)  
30 000 jeux de données  
nombreux tris et filtres (thématique, couverture, licence...)  
pour l'ESR, voir #dataESR : [https://data.esr.gouv.fr/FR/?pk\\_campaign=EmailFermetureDataESR](https://data.esr.gouv.fr/FR/?pk_campaign=EmailFermetureDataESR)
- European union open data portal : <https://open-data.europa.eu/en/data>  
site de l'*open data* de la Commission européenne  
8 500 jeux de données  
voir également le portail européen des données : <https://www.europeandataportal.eu>

- cartes

- acteurs de l'*open data* en France : <http://www.opendatafrance.net/lassociation/les-membres/>
- liste des portails *open data* dans le monde : <http://www.opendatasoft.fr/ressource-liste-portails-open-data-dans-le-monde/>
- Data portals : <http://dataportals.org/>
- Open data inception : <https://opendatainception.io/>
  - à compléter par des recherches sur les sites producteurs directement (filtres ou visualisations plus poussées, etc.)
  - à compléter par des recherches sur des ressources payantes (plus complètes, plus récentes et avec plus d'analyse)

# Brevets, marques...

- moteurs de recherche
  - Google search patents : <http://www.google.fr/webhp?tbm=pts>
    - peu de possibilités de recherche
    - + permet de trouver des documents non brevets issus de Google scholar et Google book
- sites spécialisés
  - INPI : <http://www.inpi.fr/fr/services-et-prestations/bases-de-donnees-gratuites.html>  
marques, brevets, dessins et modèles, jurisprudence
  - OEB (Office européen des brevets) : Espacenet : <http://www.epo.org/searching/free/espacenet.html>  
90 M. de documents, 1836 →  
constitue la base de nombreuses bases de données brevets comme *The lens*, <https://www.lens.org/>
  - USPTO (*US Patent and Trademark Office*) : <http://www.uspto.gov/>  
*patents, trademarks*

pour aller plus loin, [tutoriel sur les brevets d'invention \(URFIST de Paris\)](#)

# Presse en ligne

- presse francophone
  - Giga Presse : <http://www.giga-presse.com/>  
annuaire de 860 journaux francophones et 10 000 journaux internationaux
  - Guide Presse : <http://www.press-directory.com/>  
annuaire francophone de la presse en ligne  
notamment [presse quotidienne et hebdomadaire régionale](#)
  - Retronews : <https://www.retronews.fr/>  
archives de la presse française (1631-1945) ; projet BnF en cours
- presse internationale
  - Newspaper Map : <http://newspapermap.com/>  
annuaire, accessible *via* une carte
  - Online Newspapers.com : <http://www.onlinenewspapers.com/>  
annuaire de journaux en ligne ; possibilité d'interroger par pays, catégorie
  - Paperboy : <http://www.thepaperboy.com/>  
12 000 journaux en ligne  
possibilité de voir les unes
  - Planète Presse : [http://www.courrierinternational.com/sources\\_overview](http://www.courrierinternational.com/sources_overview)  
sources de *Courrier international* ; possibilité d'interroger par pays

# Type de documents

- généralistes (PDF, PPT, Excel...)  
penser en premier lieu aux filtres de recherche sur certains moteurs de recherche généralistes  
sur Google, opérateur filetype: (cf. [recherche avancée](#))
- cartes heuristiques
  - BiggerPlate : <http://www.biggerplate.com/> ([Library](#))
  - voir également les galeries des plateformes de logiciels : [Mindmeister](#), [Mindomo](#), [XMind](#)...

# Information circulante

- forums et listes de discussion
  - Google groups : <https://groups.google.com/forum/?fromgroups#!overview>  
archive des groupes de discussion Usenet et d'autres forums depuis 1981  
recherche dans le texte (1 MM. de messages)
  - BoardReader : <http://boardreader.com/>  
moteur de recherche  
graphiques de tendances
- listes de diffusion
  - CataList : <http://www.lsoft.com/lists/listref.html>  
catalogue des listes LISTSERV, 58 000 listes publiques  
+ recherche par pays, par nombre d'inscrits...
  - JISCMail - *National Academic Mailing List Service* : <http://www.jiscmail.ac.uk/>  
annuaire et moteur de recherche de listes du Royaume-Uni (enseignement et recherche), 8 000 listes
  - Universalistes (listes Renater) : <https://groupes.renater.fr/sympa/home>  
[catégories](#) et [moteur de recherche](#)



Images et multimédia

# Images et multimédia

<b>Documents</b>	images, dessins et graphiques son podcasts vidéos (colloques, conférences...) et screencasts
<b>Recherche</b>	recherche par mots-clés recherche inversée (par images, par sons) [ <i>reverse search</i> ] recherche textuelle
<b>Périmètre</b>	documents du domaine public et libres de droits documents sous droit
<b>Outils</b>	recherches avancées des moteurs de recherche moteurs de recherche spécialisés plateformes de dépôt et publication
<b>+/-</b>	<b>+</b> moteurs spécifiques <b>-</b> peu de recherches non textuelles, mais développement de la recherche inversée

- recherche d'images
  - moteurs de recherche classiques
    - Google images : <http://www.google.fr/imghp?hl=fr&tab=wi>
    - Bing images : <http://www.bing.com/?scope=images&FORM=Z9LH>
    - Yahoo! images : [http://images.search.yahoo.com/images;\\_ylt=A0oG7nlVwJdPkTMADFCI87UF?ei=UTF-8&fr2=tab-web&p=](http://images.search.yahoo.com/images;_ylt=A0oG7nlVwJdPkTMADFCI87UF?ei=UTF-8&fr2=tab-web&p=)

→ proposent généralement les mêmes filtres : taille, couleur, style, contenu (visage...)
  - réseaux sociaux centrés sur l'image
    - Flickr : <http://www.flickr.com/> (nombreuses images en CC0)
    - Pinterest : <https://www.pinterest.fr/>
  - moteurs d'images libres de droit ou sous licence *Creative commons*
    - CC search : <http://search.creativecommons.org/>
    - Pixabay : <https://pixabay.com/>
    - Everystockphoto : <http://www.everystockphoto.com/>
  - photos de stocks (images réutilisables, mais payantes)
    - Fotolia / Adobe Stock : <https://fr.fotolia.com/>
    - iStockphoto (Getty images) : <https://www.istockphoto.com/fr>
- + AllTheFreeStock : <https://allthefreestock.com/>  
recherche d'images, par palettes, etc.

! révéfier les droits indiqués sur l'image

! attention aux fausses images / *fake* ([comment les détecter ?](#))

- bases spécialisées (art, histoire de l'art)

Culture.fr : moteur Collections : <http://www.culture.fr/Ressources/Moteur-Collections>

métamoteur du patrimoine culturel français numérisé

1 900 collections numérisées et productions multimédia associées (bases de données, expositions virtuelles, sites internet), 660 institutions (Gallica, Musées nationaux, Direction des archives de France...)

tous sujets, tous types de documents

7,5 M. de documents, 5,7 M. d'images

à noter : [Joconde](#) (collections des musées de France), bases [Architecture et patrimoine](#) et [Histoire des arts](#)

patrimoine numérique : <http://www.numerique.culture.fr/pub-fr/index.html> - catalogue des collections numérisées

pour des ressources muséales, voir également

- [Images d'art](#) : les œuvres des musées français (RMN-GP)

- [Panorama de l'art](#) (RMN, 550 000 images, site pédagogique)

Google Arts & culture : <https://artsandculture.google.com/>

projet Google, en partenariat notamment avec des institutions culturelles : visites virtuelles, dossiers, etc.

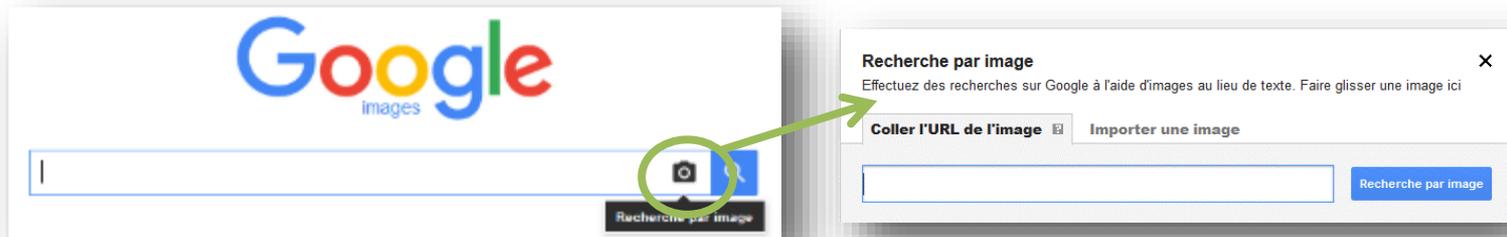
recherche par navigation ou par mot-clé, recherche inversée par [palette chromatique](#)

MédiHAL : <https://medihal.archives-ouvertes.fr/>

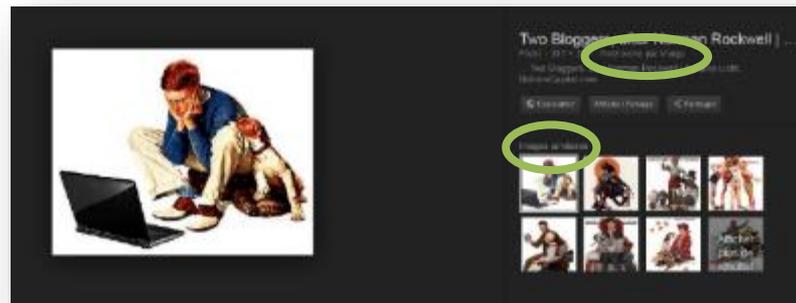
archive ouverte de photographies et d'images scientifiques

41 000 documents

- la recherche inversée par image
  - sur Google : <https://www.google.fr/imghp?hl=fr&tab=wi>  
depuis le moteur



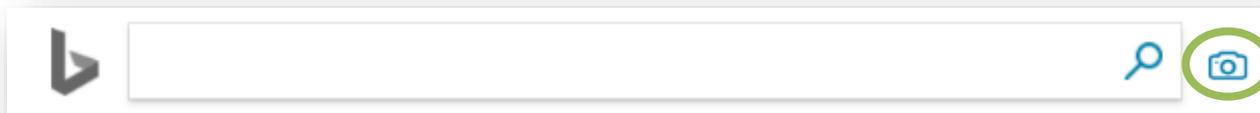
depuis la page de résultats (détails de l'image) : recherche par image, images similaires



- la recherche inversée par image

- sur Bing

depuis le moteur : <http://www.bing.com/?scope=images&FORM=HDRSC2>



depuis la page de résultats (détails de l'image) : recherche par image, images similaires



☺ possibilité de rechercher sur un détail de l'image



- autres moteurs de recherche inversée
  - TinEye : <https://www.tineye.com>  
recherche des pages où l'image est utilisée sur le web
  - Flickr : <http://flickr.com/>



- Gallica similitudes : <http://gallicastudio.bnf.fr/gallicaimages/>
- la recherche par couleur
  - MulticolorEngine (TinEye) : <http://labs.tineye.com/multicolr> (images Flickr en CC)
  - Chromatik (Exalead) : <http://chromatik.labs.exalead.com/> (images Flickr)
- informations sur la photo (données EXIF, métadonnées)
  - FotoForensics : <https://fotoforensics.com/>
  - Findexif : <http://www.findexif.com/>

- généralistes
  - Google maps : <https://maps.google.com/>
  - Bing maps : <http://www.bing.com/maps/>
  - Yahoo! maps : <http://maps.yahoo.com/>

→ proposent généralement les mêmes fonctionnalités :  
cartes, images aériennes et satellites, voire *streetview*
- spécialisés
  - Géoportail : <http://www.geoportail.gouv.fr/accueil>  
IGN et BRGM  
portail des territoires et des citoyens qui a pour but de « faciliter l'accès à l'information géographique de référence »  
+ cartes thématiques
  - penser également aux bibliothèques numériques comme les cartes dans [Gallica](#)

- podcasts
  - Digital Podcast : <http://www.digitalpodcast.com/>  
répertoire (25 000 podcasts en 97 catégories)
  - Listen Notes : <https://www.listennotes.com/>  
650 000 podcasts
- SONS
  - FindSounds : <http://www.findsounds.com/>  
moteur de recherche d'effets sonores (animaux, instruments de musique, TV et cinéma...)
  - Freesound : <http://www.freesound.org>  
moteur de recherche d'effets sonores
  - Soundcloud : <http://soundcloud.com/>  
plateforme de partage avec moteur de recherche
  - Midomi : <http://www.midomi.com/>  
recherche de musique  
+ recherche vocale
  - Music Smasher : <http://www.musicmasher.net>  
métamoteur de recherche de chansons

- généralistes

- Google vidéos : <https://www.google.fr/videohp?hl=fr>
- Bing vidéos : <http://www.bing.com/?scope=video&FORM=Z9LH2>  
possibilité de visionner les vidéos dès la page de résultats
- Yahoo! vidéo :  
[http://video.search.yahoo.com/video;\\_ylt=A0oG7IT8wZdPxTUAhZel87UF?ei=UTF-8&fr2=tab-web&p=](http://video.search.yahoo.com/video;_ylt=A0oG7IT8wZdPxTUAhZel87UF?ei=UTF-8&fr2=tab-web&p=)

→ proposent généralement les mêmes filtres : durée, date, résolution, source

- voir également les moteurs de recherche de plateformes de dépôt et publication : [YouTube](#) (Google), [Dailymotion](#), [Vimeo](#)...

- thématiques
  - IMDb (*Internet Movie Database*) : <http://www.imdb.com/>  
plus importante base sur le cinéma  
900 000 titres
  - INA : <http://www.ina.fr/>  
recherche vidéo, audio, publicités  
accès par moteur, thèmes, personnalités, émissions, époques
  - EUScreen : <http://www.euscreen.eu/>  
projet européen (22 pays), lié à Europeana  
accès par moteur, sujet, genre, langue et fournisseur

# Plateformes multimédia

- universitaires et académiques

- France

- Moteur des ressources pédagogiques : <http://www.sup-numerique.gouv.fr/pid33288/moteur-des-ressources-pedagogiques.html>
  - 32 000 ressources pédagogiques numériques de l'enseignement supérieur, notamment les MOOC de la plateforme FUN
  - peu de nouveautés
- UNT Universités numériques thématiques : <http://univ-numerique.fr/>
  - ressources pédagogiques rassemblées par domaines et disciplines ([sciences fondamentales – Unisciel](#), [sciences juridiques et politiques – UNJF / Univ-droit](#), [santé et sport – UNESS](#), [sciences de l'ingénieur et technologie – UNIT](#), [environnement – UVED](#), [économie, gestion, marketing – AUNEGE](#), [humanités – UOH](#), [enseignement technologique – IUT en ligne](#))
  - grande disparité de contenus selon les domaines
- Canal-U : <http://www.canal-u.tv/>
  - vidéothèque de l'enseignement supérieur (conférences, cours)
  - à compléter par [Canalc2](#) (colloques...)
- Eduthèque : <http://www.edutheque.fr/accueil.html>
  - Ministère de l'éducation nationale
  - ressources pédagogiques, culturelles et scientifiques, pour les enseignants (création de compte nécessaire)

# Plateformes multimédia

- international
  - YouTube
    - dont YouTube EDU : <http://www.youtube.com/education>
    - chaînes d'établissements ou d'événements (ex. : [Collège de France](#), [TEDTalks](#))
  - Dailymotion
    - chaînes d'établissements ou d'événements (ex. : [Les Ernest](#) de l'ENS)
  - Khan Academy : <https://www.khanacademy.org/>
    - classement thématique

→ voir également les sites des établissements
- MOOC (*Massive open online courses*)
  - plateformes de MOOC
    - FUN (France université numérique) : <https://www.fun-mooc.fr/>
    - Coursera : <https://www.coursera.org>
    - EdX : <https://www.edx.org>
    - OpenupEd : <https://www.openuped.eu/>
  - annuaires et listes
    - [annuaires de MOOC francophones](#)
    - MOOC List : <https://www.mooc-list.com>



Web social

<b>Documents</b>	<p>1° informations</p> <ul style="list-style-type: none"><li>outils 2.0 (blogs, wikis)</li><li>références bibliographiques</li><li>sites et pages internet</li><li>documents multimédia</li><li>articles</li><li>supports d'interventions</li><li>offres d'emploi</li><li>actualités (colloques, recensions...)</li></ul> <p>2° personnes</p> <ul style="list-style-type: none"><li>individus : professionnels, experts, veilleurs...</li><li>groupes d'intérêts et de discussions</li></ul>
<b>Recherche</b>	<p>peu d'indexation dans les moteurs de recherche « classiques » comme Google</p> <p>outils spécifiques</p> <p>modalités variables selon les outils</p> <p>sérendipité</p>
<b>Périmètre</b>	<p>web visible et partie du web invisible</p> <p>web 2.0 / social / collaboratif</p>
<b>Outils</b>	<p>moteurs de recherche généralistes</p> <p>annuaires</p> <p>plateformes de bookmarking</p> <p>plateformes de partage</p> <p>moteurs de recherche sociale</p>
<b>+/-</b>	<ul style="list-style-type: none"><li>+ pour des documents originaux et peu connus (signaux faibles) : complément aux bases de données professionnelles [C. Tisserand-Barthole, 2012] et aux moteurs de recherche « traditionnels »</li><li>+ pour des personnes ressources dont activité peut être suivie</li><li>- pas d'outil unique spécifique à tout le web social / web 2.0</li><li>- peut nécessiter d'avoir un abonnement pour accéder aux fonctionnalités de recherche et/ou aux résultats</li><li>- information non validée</li></ul>

## Wikis

1995 : mode de publication rapide (« wiki » = rapide)

- présentation

  - architecture et contenu modifiables

  - structuration en rubriques reliées par des liens hypertexte

  - présence d'un historique

} ex. de fonctionnement : [Wikipédia](#)

- recherche de wikis

  - [Qwika](#)

    - moteur de recherche 1 200 wikis, 22 M. d'articles en 12 langues (notamment les différentes Wikipedia)

  - [WikiIndex](#)

  - [Liste Wikipédia](#)

  - voir également les listes des hébergeurs, comme [Sites MediaWiki](#) , [Wikia](#)

## Blogs

mi-1990's : mode de publication rapide (blog = *web* + *log*)

- présentation
  - architecture et contenu modifiables
  - structuration en billets (présentation ante-chronologique) et catégories
  - liens permanents
  - possibilité de commentaires
- cas particulier du blogging scientifique
  - carnets de recherche et d'expérimentations, ex. : plateforme [Hypothèses](#)

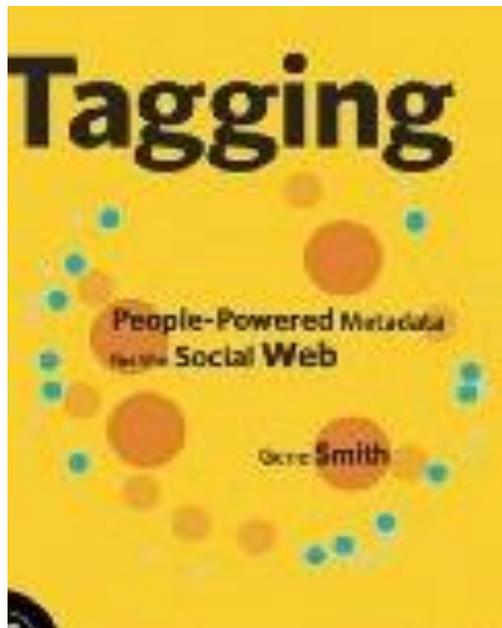
- recherche de blogs
  - annuaires
    - EatonWeb : <http://portal.eatonweb.com/>  
annuaire de blogs sur soumission  
+ *metrics* : *overall*, *strength* et *momentum*
    - voir également les signets de bibliothèques
    - voir également les annuaires et répertoires ([BOTW](#))
  - blogrolls des blogs
  - voir également les plateformes de blogs, comme [OverBlog](#)

- moteurs de recherche sur le contenu de blogs
    - académiques
      - Isidore : <https://isidore.science/>  
pour rechercher dans le contenu de la plateforme Hypothèses (plutôt sciences humaines francophones)  
filtre *a posteriori* sur les résultats d'une recherche (type de document > billets de blog)
- + recherche sur un moteur de recherche avec une équation de type  
inurl:blog OR intitle:blog



# La folksonomie

- intérêt des tags : visibilité de concepts qui ne sont pas encore entrés dans le domaine bibliothéconomique



Library of Congress



Metadata.

Internet --Social aspects.

Library Thing



classification **folksonomies folksonomy** information information architecture information  
retrieval information technology internet IT libraries library 2.0 librarything **metadata** non-fiction read  
social **social bookmarking social media** social networking social software **social tagging social**  
**web tag tagging tags** taxonomy to read WEB **web 2.0**

! pour des termes associés, construire sa requête ainsi : eau\_potable (avec \_),  
« eau potable » (avec « »), voire eaupotable

## Plateformes de « *social bookmarking* »

- outils de stockage et de gestion de contenus numériques  
URL, références bibliographiques, images, texte...
- présentation de chaque élément  
titre ou références bibliographiques  
(tags)
- outil de mise en commun et de réseautage  
groupes, abonnements à des comptes...

Ex. de fonctionnement : [Diigo](#)

intérêt de ces plateformes pour la recherche :

Diigo : top 10 des outils de recherche ([CNET, 2006](#))

# Bookmarking

- bibliothèques et avis
  - LibraryThing : <http://www.librarything.com/>  
majoritairement anglo-saxon  
2,4 M. membres ; 130 M. de livres catalogués ; 150 M. de tags
  - Babelio : <http://www.babelio.com/>  
majoritairement francophone
  - Goodreads : <https://www.goodreads.com/>
  - Sens critique : <https://www.senscritique.com/>
- bookmarks
  - généralistes
    - Diigo : <https://www.diigo.com/>  
plateforme de favoris internet  
[formulaire de recherche](#)  
[formulaire de recherche de personne](#) : champs, opérateurs
    - Pearltrees : <http://www.pearltrees.com>  
plateforme de favoris internet  
possibilité d'interroger via Google avec une requête `site:pearltrees.com`

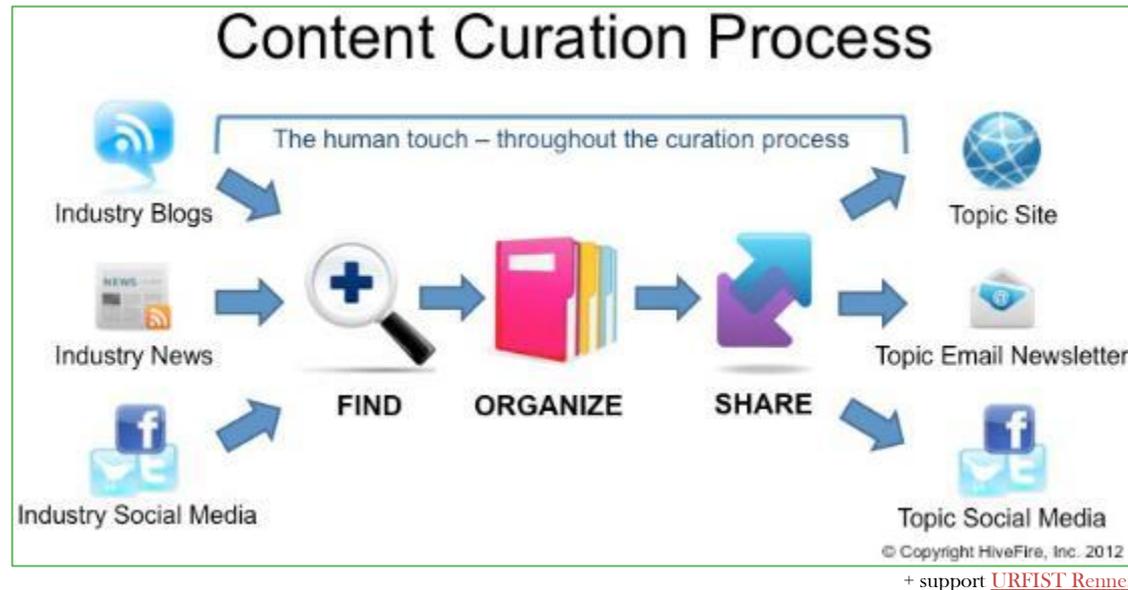
# Bookmarking

- références bibliographiques
  - BibSonomy : <http://www.bibsonomy.org/>
    - + bookmarks et références bibliographiques
  - CiteULike : <http://www.citeulike.org>
    - 8,6 M. de références
    - [critères de recherche avancée](#) (titre, revue, auteur, numéros internationaux, tags, utilisateur...)
  - Mendeley : <https://www.mendeley.com/>
    - Elsevier
    - 3 M. de membres, 264 000 groupes, 520 M. de références
    - est également un véritable gestionnaire de références bibliographiques
    - [formulaire de recherche avancée](#) (critères : titre, auteur, résumé, MeSH, type de document, date, disciplines)
  - Zotero : <https://www.zotero.org/>
    - est avant tout un gestionnaire de références bibliographiques local (installation sur l'ordinateur)
    - peu de possibilités de recherche (via [formulaire](#) : utilisateur, groupe ; [forum](#) : sujet, commentaire et utilisateur)

# Plateformes de contenu

- généralistes
  - Slideshare : <http://www.slideshare.net/>  
LinkedIn  
essentiellement des présentations (supports de formations et d'interventions...)  
[vu par un maître de conférences \(O. Ertzscheid\)](#)
    - moteur de recherche peu satisfaisant, tris et filtres *a posteriori*
  - Calameo : <http://en.calameo.com/>
  - Issuu : <http://issuu.com/>

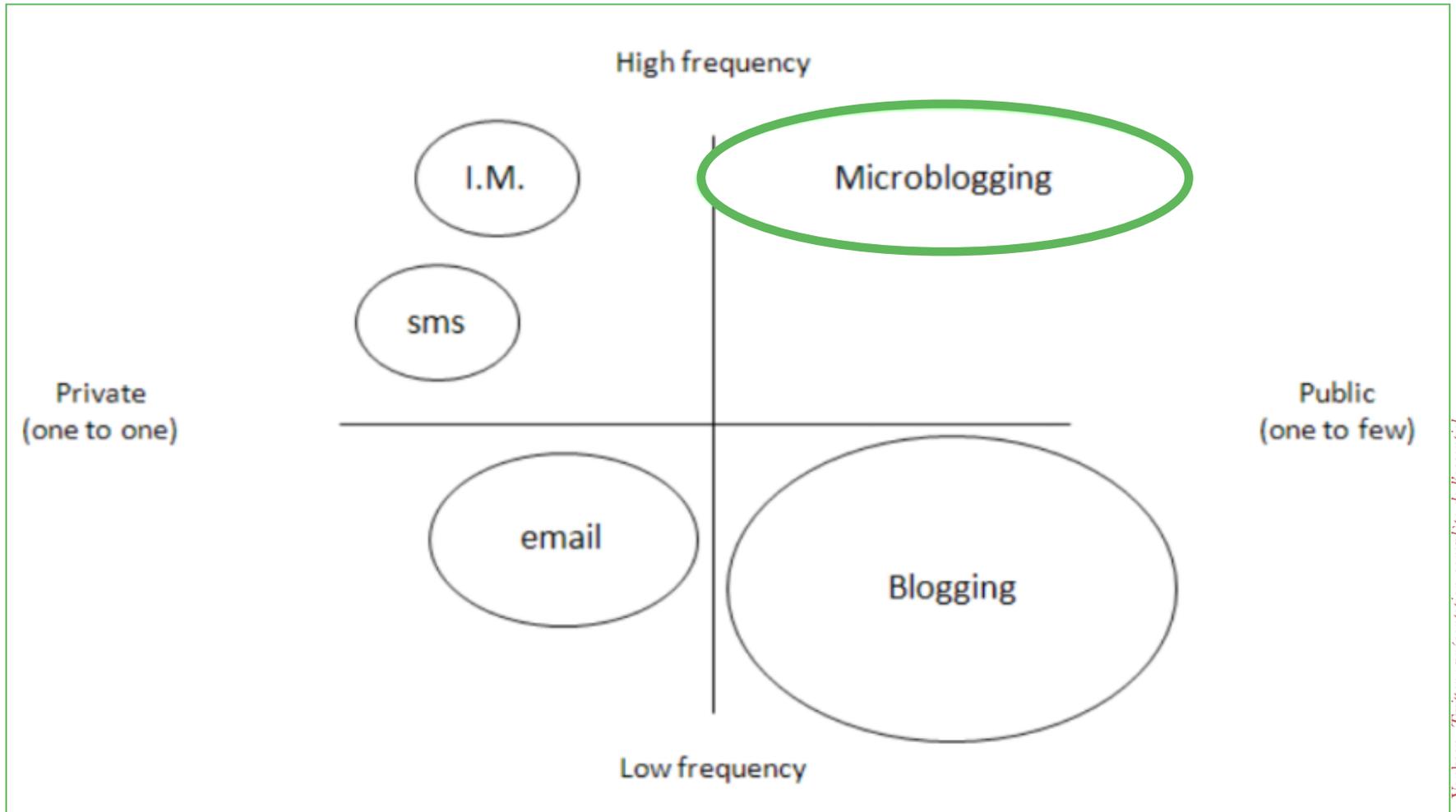
# Outils de curation



- Scoop.it : <http://www.scoop.it/>  
veille thématique  
→ diffusion de la veille  
→ suivi de l'actualité d'un domaine
- Storify : <http://storify.com/>  
dossiers éditorialisés
- Pinterest : <https://www.pinterest.com/>  
veille thématique, plutôt orientée images
- Instagram : <https://www.instagram.com/>  
images

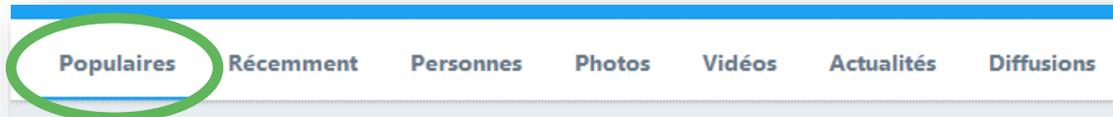
à interroger *via* un moteur de recherche  
avec une requête de type *site:*

# Microblogging



# Microblogging

- **Twitter** : <https://twitter.com/>
  - 271 M. de comptes mensuels actifs
  - tweets : messages limités à 140 puis (à partir de novembre 2017) 280 caractères (d'où syntaxe et vocabulaire spécifique : @, RT, tinyurl...)
  - utilisation du # (*hashtag*) pour ajouter des mots-clés
  - réseau d'information en temps réel (cf. pratique du live-tweet - LT)
  - recherche simple : <https://twitter.com/search-home> et recherche avancée : <https://twitter.com/search-advanced> (mots, personnes, lieux, autre avec croisements possibles) sans avoir besoin d'un compte
  - exemple d'usage dans le domaine académique : [#IcanhazPDF](#)  
[vu par un maître de conférences \(O. Ertzscheid\)](#) : filtre humain et non algorithmes d'un moteur
  - ! par défaut, classement des résultats par « Populaires » (pertinence), possibilité de classer par fraîcheur (« Récemment »), mais ordre antéchronologique
  - exhaustivité des résultats ? (ex. : mode d'affichage des RT)



à compléter par une recherche sur Qwant (<https://www.qwant.com> > Social) pour afficher plus facilement les RT

- autres moteurs pour Twitter
  - Keyhole : <http://keyhole.co>, Social Bearing : <https://socialbearing.com>  
analyses graphiques
  - Trendsmap : <http://trendsmap.com/>  
carte de tendances temps réel géolocalisées

## Réseaux sociaux

- *social networking* (« service de réseautage social en ligne ») : liens bilatéraux ou via des groupes
  - réseaux d'affinités, de pratiques et / ou d'intérêts
  - proposent notamment des espaces de discussions/forums, offres d'emploi, manifestations voire plateformes de dépôt de documents
- ✚ possible indexation dès les premières pages des moteurs de recherche généralistes (Google)
- ! recherche parfois possible seulement après inscription  
→ essayer cependant une recherche de type *site:* sur un moteur comme Google, Bing ou DuckDuckGo
- ! résultats de la recherche pouvant varier selon la taille du réseau et les abonnements (notamment payants) souscrits

# Réseaux sociaux

- généralistes
  - Facebook : <http://www.facebook.com/>  
3,5 MM. de comptes (2<sup>e</sup> site mondial, 80 langues), + 3,5 MM. de contenus partagés/sem.  
à l'origine de 20% de l'ensemble des pages web consultées aux États-Unis
- professionnels
  - LinkedIn : <http://www.linkedin.com/>  
550 M. de comptes  
à prédominance anglophone  
! résultats de recherche variables selon la taille du réseau et l'abonnement souscrit

- académiques

- ResearchGate : <https://www.researchgate.net/>
  - 15 M. de comptes, 100 M. de documents
  - réseau à prédominance scientifique (sciences, techniques, médecine)
  - centré sur le partage de papiers de recherche (*open science*)
  - ! profils d'institutions créés automatiquement
  - ! pas/peu de valeur scientifique de certaines métriques proposés (RG score)
- Academia : <https://www.academia.edu>
  - 80 M. de comptes, 22 M. d'articles
  - plutôt sciences humaines et sociales
  - centré sur le partage de papiers de recherche (*open science*)
  - ! pas de vérification des *research interests* (2 M. de tags)

+ à compléter par une recherche sur les moteurs de recherche avec *site:* et par Google Scholar

# Moteurs de recherche sociale

- Moteurs portant sur le web 2.0 (réseaux sociaux, microblogging...) et la présence en ligne en général
- peuvent interroger le web invisible
- assez inégaux
  
- moteurs de personnes
  - Youseemii : <http://www.youseemii.fr/>
  - Pipl : <http://pipl.com>
  - WebMii : <http://www.webmii.com>
  - Yasni : <http://www.yasni.com>
  
- ! penser à utiliser également :
  - les moteurs de recherche (ex. avec une recherche de type [fédération OR association] pour trouver un groupe d'expert)
  - des annuaires généralisés ou non
    - ex. : [annuaire de l'administration](#) pour le service public
    - [Cismef](#) pour les sites médicaux
    - [Indexa](#) pour les entreprises françaises



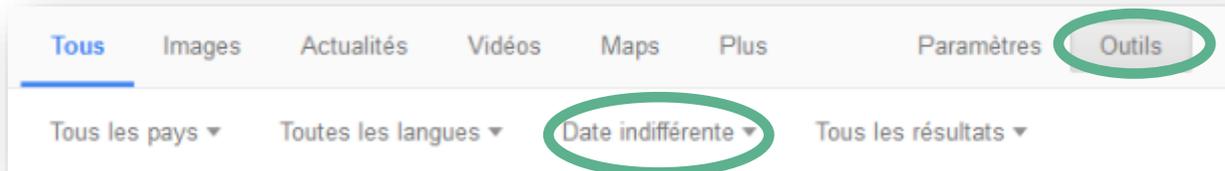
Actualités et temps réel

# Actualités et temps réel

<b>Documents</b>	actualités et informations (presse, dépêches, communiqués...) tendances ( <i>trends, buzz</i> ) événements dont colloques
<b>Recherche</b>	! sur les moteurs de recherche « traditionnels », de type Google ou Bing, bien distinguer de ce qui relève des filtres sur les dates (= tous les types d'informations web indexées par le moteur de recherche) de ce qui relève spécifiquement du filtre « actualités » (= informations issus des médias de presse) spécifique à chaque outil présence généralement d'une recherche avancée possibilité de recherches géolocalisées
<b>Périmètre</b>	généralistes thématiques
<b>Outils</b>	annuaires, répertoires... outils de recherche sur les médias d'actualités agrégateurs et bases de données de presse moteurs de recherche en temps réel réseaux sociaux (Twitter...)
<b>+/-</b>	<b>+</b> possibilité de fonctionnalités supplémentaires (visualisations, répartitions thématiques, alertes...) ! mises à jour ? (vrai temps réel ?) ! classement des résultats ? ! informations pas (encore) toujours validées

# Temps réel

- moteurs de recherche généralistes
  - Google



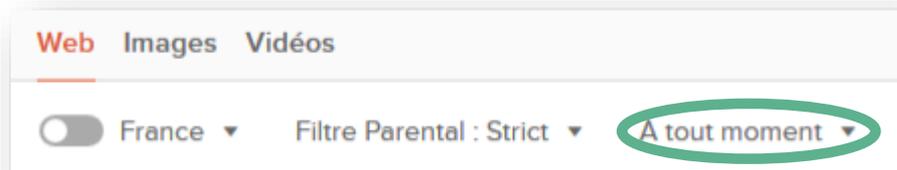
date indifférente  
moins d'une heure  
moins de 24 heures  
moins d'une semaine  
moins d'un mois  
moins d'un an  
période personnalisée

- Bing



toutes  
dernières 24 heures  
dernière semaine  
dernier mois  
plage personnalisée

- DuckDuckGo



à tout moment  
hier  
la semaine dernière  
le mois dernier

- moteurs de recherche généralistes  
contenus provenant principalement de sites d'agences de presse, de journaux  
(dont *pure-players*)
  - Google Actualités : <https://news.google.com>
    - + différentes éditions nationales (+ 72 éditions, 30 langues, 55 000 sources)
    - + possibilités de personnalisation
    - pas de recherche avancée dont filtres ou source → des possibilités de recherche et de filtres plus larges en passant directement par Google > Actualités
    - des résultats erratiques
  - Bing actualités : <http://www.bing.com/news>
    - peu de résultats
  - Yahoo ! actualités France : <http://fr.news.yahoo.com/> : agences de presse et journaux dont *pure-players*
    - peu de résultats

voir également du côté du web social

mais d'une manière générale, regarder du côté des agrégateurs de presse payants (Factiva, Press'edd, Europress, Nexis/Newsdesk) pour leur contenu et leurs fonctionnalités

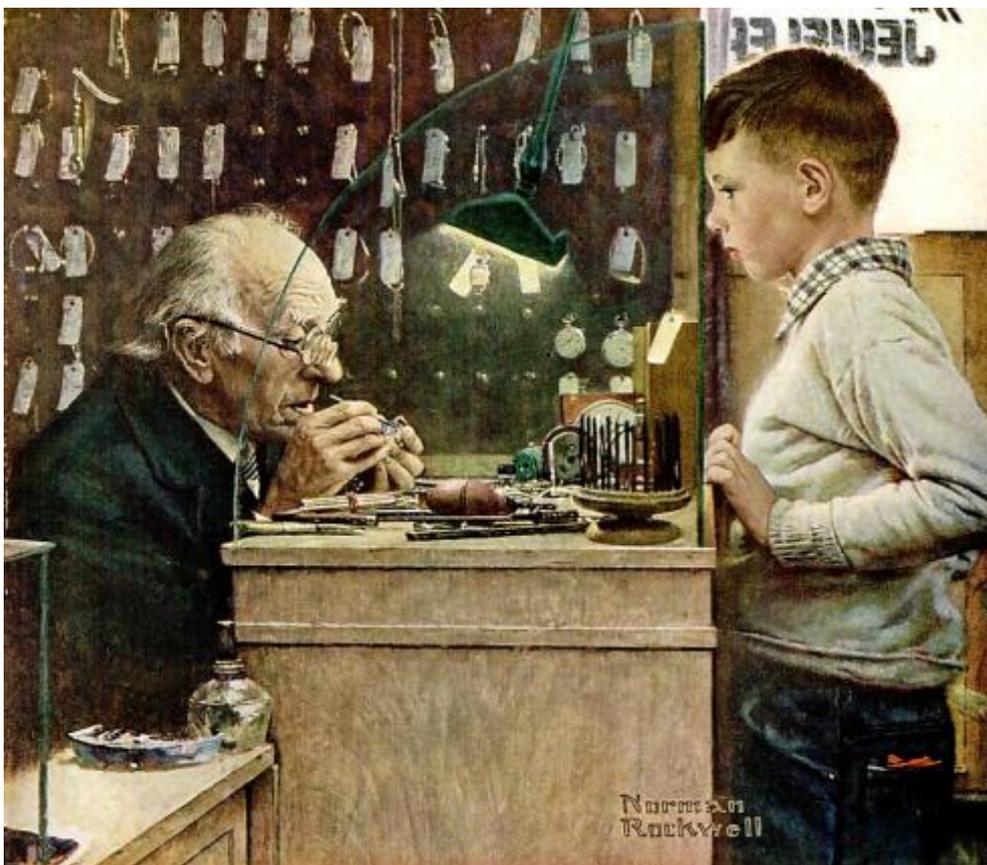
- NewsBrief : <http://emm.newsbrief.eu/NewsBrief/clusteredition/fr/latest.html>  
agrégateur de journaux européens, mis à jour toutes les 10 minutes  
+ 7 000 sites, 1 600 portails, 70 langues  
graphique de tendances, détection d'événements, clusterisation  
à compléter par NewsExplorer : <http://emm.newsexplorer.eu/NewsExplorer/home/fr/latest.html> (un des autres portails Europe Media Monitor du Joint Research Centre de la Commission européenne)
  - NewsMap : <http://newsmap.jp/>  
agrégateur mondial  
visualisation graphique des résultats (thématiques en couleurs)
  - Unfiltered news : <http://unfiltered.news>  
sélectionne les 100 articles les plus populaires dans un pays (Google actualités), les traduit (Google Translate) et les compile au niveau mondial
  - WebPlanete news : <http://news.webplanete.net/>  
930 sources d'informations francophones  
grandes thématiques
- + ces outils proposent souvent des fonctionnalités supplémentaires (graphiques, thématiques, etc.)

à compléter par [M.-L. Malingre et A. Serres. Plus de 80 outils...](#)  
et [Outils de recherche et de veille sur l'actualité](#)

# Tendances

- tendances
  - Google tendances de recherches :
    - interroge le volume de recherches de mots-clés sur Google depuis 2004
    - <https://www.google.com/trends/> : recherches du moment
    - <https://www.google.com/trends/explore#q> : possibilité de choisir un ou plusieurs termes pour voir/comparer leur évolution
      - + recherche par pays, période, catégories





Quelques outils  
complémentaires

# Appel à la communauté

- « *Ask a librarian* »

**Pour commencer :** métamoteur de différents services de questions/réponses : [sqrpro.fr](http://sqrpro.fr)

- SINDBAD : <https://www.bnf.fr/fr/une-question-pensez-sindbad>  
BnF  
[archives sélectionnées](#)
- Eurêkoi : <http://www.eurekoi.org/>  
issu du partenariat entre l'ancien BiblioSésame (réseau d'une vingtaine de bibliothèques publiques françaises, lancé par la BPI) et la Fédération Wallonie-Bruxelles  
[archives](#)
- Guichet du savoir : <http://www.guichetdusavoir.org/>  
service de la bibliothèque municipale de Lyon
- Rue des facs : fermé en 12/2016 : [archives](#)  
25 établissements documentaires d'Ile-de-France ; accès par discipline  
destiné en priorité aux étudiants et aux enseignants-chercheurs
- InterroGE : <http://www.ville-geneve.ch/themes/culture/offre-culturelle/bibliotheques/interroge/>  
service du réseau des bibliothèques genevoises
  
- forums et listes de discussion
  - Ex. : Reddit scholar : <https://www.reddit.com/r/Scholar/>  
StackExchange : <http://stackexchange.com/>
  
- réseaux sociaux
  - Ex. : Quora : <https://www.quora.com/>  
ResearchGate : section Q&A : <http://www.researchgate.net>

- ! question des droits d'auteur
- les licences Creative Commons

Désignation complète du contrat	Terme abrégé désignant la licence	Symboles désignant la licence				Type de licence
Paternité	CC-BY					Licence libre non copyleft
Paternité Partage des conditions initiales à l'identique	CC-BY-SA					Licence libre copyleft
Paternité Pas de modification	CC-BY-ND					Licence de libre diffusion
Paternité Pas d'utilisation commerciale	CC-BY-NC					Licence de libre diffusion
Paternité Pas d'utilisation commerciale Partage des conditions initiales à l'identique	CC-BY-NC-SA					Licence de libre diffusion
Paternité Pas d'utilisation commerciale Pas de modification	CC-BY-NC-ND					Licence de libre diffusion

# Les droits

- CC search : <http://search.creativecommons.org/>  
recherche sur les sites Google, Flickr, YouTube...  
! vérifier cependant au cas par cas pour chaque document obtenu
- images
  - CC search : <https://ccsearch.creativecommons.org/> (encore en version bêta)  
photos (500px, Flickr) et œuvres culturelles (Europeana, MET, New York Public Library, Rijksmuseum)
  - Visual Hunt : <https://visualhunt.com>  
images de stock en CC
  - LibreStock : <http://librestock.com/>  
métamoteur de photos de stock en CC0
  - Google images



## Bing images



- autres types de documents
  - trouver de la musique en CC

# Les extensions de navigateur

- insertion du moteur de recherche dans la barre du navigateur
- analyse de la recherche
  - *Bettersearch*
  - *ClearForest Gnosis* : repérage d'entités nommées
  - *Fast search by Surf Canyon* : réorganisation des résultats
- capture de pages
  - *ScrapBook, Screengrab !*
- gestion des favoris en ligne
  - *Diigo*
  - *Zotero*
- sécurité
  - *AdbockPlus* : stop pub
  - *BetterPrivacy, Self-Destructing Cookies* : suppression des cookies
  - *Disconnet, DoNoTrackMe/Blur, Ghostery* : navigation anonyme
  - *WOT* : évaluation des sites

Ex. [Christophe Deschamps](#)

- traduction
  - (en priorité) DeepL : <https://www.deepl.com/translator>
  - Google traduction : <https://translate.google.fr/>
- moteurs de recherche personnalisables
  - Google recherche personnalisée (Google CSE) : <https://www.google.com/cse/>
  - Bing Custom Search : <https://www.customsearch.ai/>



Veille automatisée

# Outils de veille

- newsletters et lettres d'informations
  - répertoire de newsletters : Ezine Directory : <http://www.ezine-dir.com/>, Email Tuna : <https://emailtuna.com/>
    - généralement peu satisfaisants
  - regarder directement sur les sites concernés (sites internet, médias, blogs, etc.)
    - ! comparer les informations données *via* les newsletters, les flux RSS et la présence des sites/organismes sur les réseaux sociaux (Facebook, LinkedIn, Twitter, etc.) ; les informations peuvent varier, être éditorialisées, etc.
- surveillance de mots-clés
  - Talwalker Alerts : <http://www.talkwalker.com/fr/alerts>  
paramétrage de la fréquence et du volume d'envoi et des sources  
? sources concernées (pas le web en entier)
  - Google alertes : <https://www.google.com/alerts#1:0>  
paramétrage de la fréquence et du volume d'envoi et des sources  
possibilité de s'abonner à des alertes en plusieurs langues
    - nombreux dysfonctionnements
  - voir également [autres moteurs de recherche](#) ou directement sur les ressources concernées (bases de données, catalogues de bibliothèques...)
  - outils pour les médias sociaux : Tweetdeck : <https://tweetdeck.twitter.com/> (pour Twitter), Mention : <https://mention.com/fr/>

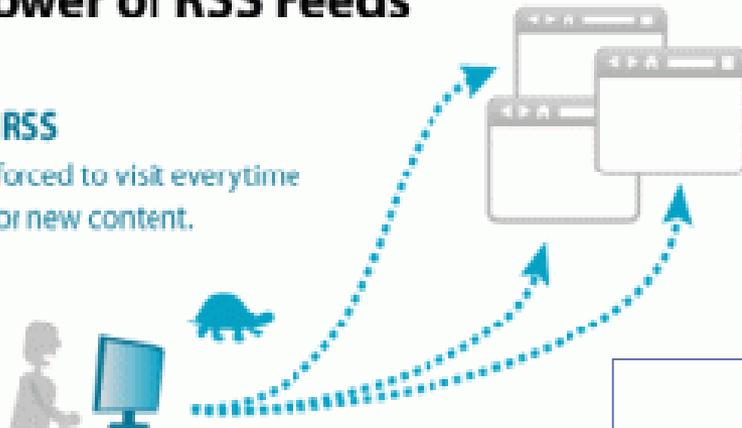
# Outils de veille

- outils de collecte (! : des outils de plus en plus payants – modèle *freemium*)
  - surveillance de pages et de sites web (modifications...)
    - extensions de navigateurs : ex. : *Update Scanner*, *Distill Web monitor*
    - services gratuits : ChangeDetection : <https://www.changedetection.com/>
    - services professionnels : ex. : WebSite Watcher : <http://www.website-watcher.fr/>
  - surveillance /découverte des actualités
    - lecteurs de flux RSS (voir principes ci-dessous)
      - marque-pages dynamique du navigateur
      - outils sur l'ordinateur (Zotero) ou outils en ligne (Inoreader : <http://www.inoreader.com/>, Feedly : <http://feedly.com/i/welcome>)
      - outils pour les réseaux sociaux (Queryfeed : <http://www.queryfeed.net/>)
    - outils de *social media monitoring*
      - Hootsuite : <https://hootsuite.com/>, Mention : <https://mention.com/fr/>, Brandwatch : <https://www.brandwatch.com/>, Visibrain : <https://www.visibrain.com/fr/> - surtout *corporate*
    - outils de curation automatisé *content discovery* / *news discovery* / magazines intelligents / personnalisés, souvent liés aux comptes sur les réseaux sociaux
      - Nuzzel : <https://nuzzel.com/>, Paper.li : <https://paper.li/>, Flipboard : <https://flipboard.com/>
- + plateformes de veille intégrées (l'ensemble du processus de veille, y compris la diffusion) : KB Crawl : <http://www.kbcrawl.net/>, Digimind : <https://www.digimind.com>

## The Power of RSS Feeds

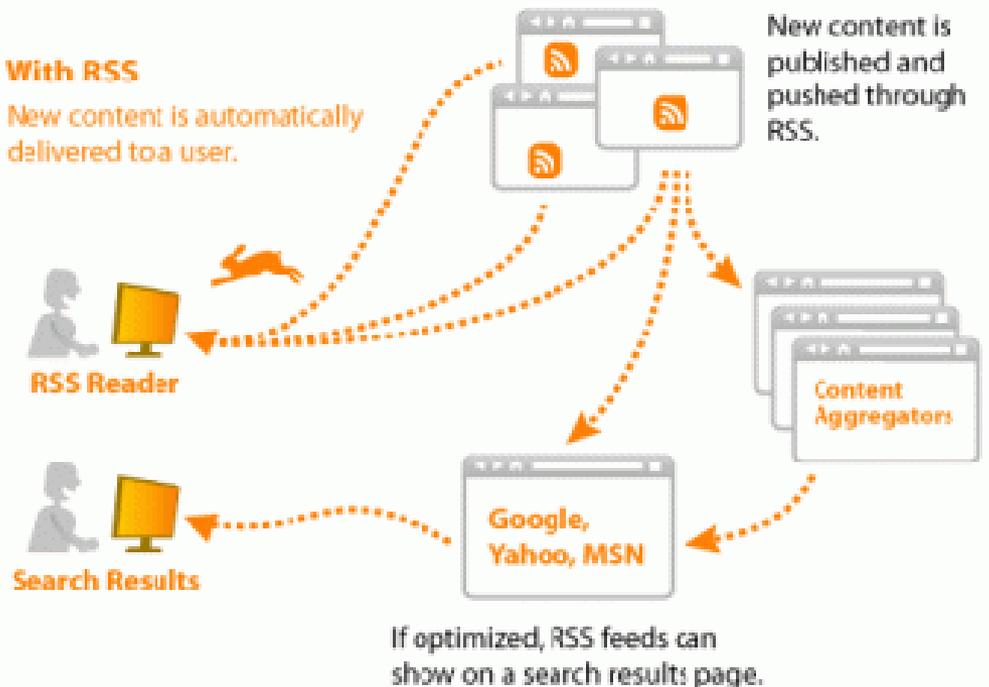
### Without RSS

Users are forced to visit everytime to check for new content.



### With RSS

New content is automatically delivered to a user.



# Les flux RSS

- suivi de l'actualité

Ex. [Le Monde](#)



Ex. le [blog de B. Foenix-Riou](#)



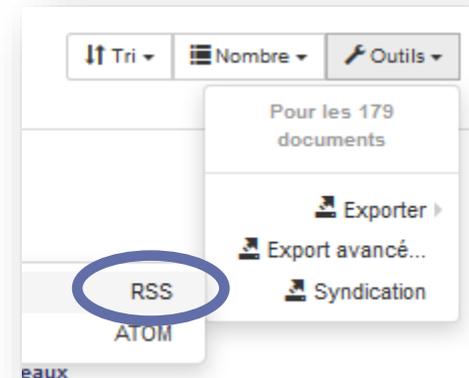
- veille  
nouvelles publications

Ex. [Cairn](#)



bases de données bibliographiques

Ex. [HAL](#)





Exercices de synthèse



# Bibliographie

# Bibliographie

## Pour commencer

- Marie-France Andral. *Recherche documentaire et recherche d'information*. URFIST de Bordeaux, 2013. [en ligne]. Disponible sur : <http://weburfist.univ-bordeaux.fr/wp-content/uploads/2013/02/2013.02.18-Recherche-doc-info2.pdf>.
- Béatrice Foenix-Riou. *Recherche éveillée sur internet : mode d'emploi*. 12/2011. 40 p. [en ligne]. Disponible sur : <http://fr.slideshare.net/bfoenix/recherche-veille-bfr>.
- . *Recherche éveillée sur Internet : mode d'emploi : outils et méthodes pour explorer le Web, Web visible, Web invisible, Web social, Web temps réel*. Paris : Lavoisier : Bases publications, DL 2011. 367 p.
- Marie-Laure Malingre et Alexandre Serres. *Plus de 80 outils spécialisés pour différentes recherches sur le web*. [en ligne]. Disponible sur : <http://www.sites.univ-remes2.fr/urfist/ressources/des-outils-specialises-pour-divers-types-de-recherches-sur-le-web>.
- . *Quels outils pour quelles recherches ? Panorama des outils de recherche du web*. URFIST de Rennes, 13 /03/2012. Présentation, 88 f. [en ligne]. Disponible sur : <http://fr.slideshare.net/UrfistRennes/quels-outils-pour-queelles-recherches-panorama-des-outils-du-web>.
- Véronique Mesguich et Armelle Thomas. *Net recherche 2013. Surveiller le web et trouver l'information utile*. Préface d'Olivier Andrieu. 5<sup>e</sup> éd. refondue et mise à jour. Paris-Bruxelles : ADBS éd.-De Boeck, 2013. 263 p. (« Information et stratégie »).
- Université européenne de Bretagne. *Form@doct. Formation à distance en information documentation pour les doctorants*. [en ligne]. Disponible sur : <http://guides-formadoct.ueb.eu/>.

## Pour aller plus loin

### Généralités sur internet

- Blog du modérateur*. [en ligne]. Disponible sur : <https://www.blogdumoderateur.com/dossier/chiffres-internet/> - état des lieux réguliers de l'utilisation d'internet (monde, France).
- CartwrightKing sollicitors. *The dark side of the internet*. Infographie. 2016. [en ligne]. Disponible sur : <http://cartwrightking.co.uk/news/the-dark-web>.
- Internet World Stats. Usage and population statistics*. [en ligne]. Disponible sur : <http://www.internetworldstats.com/stats.htm>.
- StatCounter Global Stats*. [en ligne]. Disponible sur : <http://gs.statcounter.com/>.
- « Usages of content languages for websites ». *W<sup>3</sup>Techs*. [en ligne]. Disponible sur : [http://w3techs.com/technologies/overview/content\\_language/all](http://w3techs.com/technologies/overview/content_language/all).

### Méthodologie de recherche documentaire

- Eduscol. *Rechercher sur internet : méthodologie et outils*. [en ligne]. Disponible sur : <http://eduscol.education.fr/numerique/dossier/competences/rechercher>.
- Victorine Porte. *Méthodologie et astuces de recherche sur internet*. 2010. [en ligne]. 31 p. Disponible sur : <http://fr.slideshare.net/vporte/guide-rechercheenligne>.
- URFIST de Paris. *CERISE. Conseils aux étudiants en recherche d'informations efficace*. [en ligne]. Disponible sur : <http://urfist.chartes.psl.eu/cerise/>.
- Dale Vidmar. « Tips to Effective Internet Searching ». *Hannon Library*. 2012. [en ligne]. Disponible sur : <http://hanlib.sou.edu/searchtools/searchtips.html>.

# Bibliographie

## Supports de formation généralistes

- Manuel Durand-Barthez. *Recherche d'information en sciences exactes et appliquées*. 2017. Présentation. 58 f. [en ligne]. Disponible sur : <http://fr.slideshare.net/URFISTParis/physk-131125084847phpapp01>.
- Frédérique Flamerie. *Recherche d'information bibliographique*. UPMC. 01/2015. Présentation. 39 f. [en ligne]. Disponible sur : <http://fr.slideshare.net/BUPMCformD/doctorat-sciences-recherche-dinformation-mthode-sources-et-outils>.
- Marie-Laure Malingre et Alexandre Serres. *Outils de recherche et de veille sur l'actualité (news, médias sociaux...)*. URFIST Rennes, 10/2014. 108 f. [en ligne]. Disponible sur : <http://fr.slideshare.net/UrFistRennes/stage-outils-rechercheactualit20121211>.

## Moteurs de recherche

- Hervé Basset. « Une veille professionnelle avec Google ????!!!! ». *Intelligence scientifique et veille*. 03/01/2012. [en ligne]. Disponible sur : <http://intelligencescientifique.wordpress.com/2012/01/03/une-veille-professionnelle-avec-google/>.
- Aline Bouchard. *Evolutions des moteurs de recherche sur internet*. Présentation, 06/2016. 184 p. [en ligne]. Disponible sur : <http://urfist.enc-sorbonne.fr/ressources/recherche-documentaire/evolutions-des-moteurs-de-recherche-sur-internet>.
- Phil Bradley. *Anything but Google - alternative search engines*. 10/2017. 89 p. [en ligne]. Disponible sur : <https://www.slideshare.net/Philbradley/anything-but-google-alternative-search-engines>.
- . *Evaluating search engines*. 10/2017. 32 p. [en ligne]. Disponible sur : <https://www.slideshare.net/Philbradley/evaluating-search-engines-80980536>.
- Fabrice Deprez. « Géolocalisation : peut-on faire une recherche « neutre » avec Google et par quels moyens ? ». *Netsources*, n°123, 07-08/2016. p. 7-10.
- Olivier Ertzscheid. « (My) state of search 2012 ». *Affordance.info*. 27/11/2012. [en ligne]. Disponible sur : [http://affordance.typepad.com/mon\\_weblog/2012/11/my-state-of-search-2012.html](http://affordance.typepad.com/mon_weblog/2012/11/my-state-of-search-2012.html).
- Béatrice Foenix-Riou. *Google : trucs et astuces pour les professionnels de l'infodoc*. 01/2013. 35 p. [en ligne]. Disponible sur : <http://fr.slideshare.net/bfoenix/bfrabd-2013-google-trucs-et-astuces-pour-les-professionnels-de-linfodoc-16028591>.
- . *Rechercher sur internet : y a-t-il une vie sans Google ?* Présentation, 10/12/2013. 60 p. [en ligne]. Disponible sur : <http://fr.slideshare.net/bfoenix/recherches-sur-internet-y-at-il-une-vie-sans-google-bfr-consultants>.
- . « Retrouver des informations grâce aux images ». *Netsources*, n°97, 03-04/2012. p. 1-3. [en ligne]. Disponible sur : <http://www.recherche-veillee.com/methodologies-pour-retrouver-informations-images>.
- Hackcollege. *Get more out of Google. Tips & tricks for students conducting online research*. 23/11/2011. [en ligne]. Disponible sur : <http://www.hackcollege.com/blog/2011/11/23/infographic-get-more-out-of-google.html>.
- « List of search engines ». *Wikipedia*. [en ligne]. Disponible sur : [http://en.wikipedia.org/wiki/List\\_of\\_search\\_engines](http://en.wikipedia.org/wiki/List_of_search_engines).
- Marie-Laure Malingre et Alexandre Serres. *Rechercher l'information scientifique : plateformes et outils spécialisés*. URFIST Rennes, 11/2018. Présentation, 98 f. [en ligne]. Disponible sur : <https://www.sites.univ-rennes2.fr/urfist/ressources/connaitre-les-moteurs-de-recherche-de-linformation-scientifique>.
- Carole Tisserand-Barthole. « Veille et recherche scientifique 2.0 : au-delà des ressources classiques ». *Bases*, n°206, 09/2012. p. 1-6.
- Aurélien Vathonne. « Repérer les retombées presse d'une exposition grâce aux images ». *Netsources*, n°97, 03-04/2012. p. 4-5.

## Autres outils de recherche

- Phil Bradley. *The good, the bad and the totally incorrect. Fake news*. Présentation *Search solutions 2017*. Londres, 11/2017. 83 p. [en ligne]. Disponible sur : <https://www.slideshare.net/Philbradley/fake-news-82897778>.
- Aline Bouchard. *Exploiter l'open access en recherches d'information*. 2017. [en ligne]. Disponible sur : <http://urfist.chartes.psl.eu/ressources/exploiter-l-open-access-en-recherche-d-informations>.
- Béatrice Foenix-Riou. « Identifier la date de publication d'une page ou d'un site web ». *Netsources*, n°99, 07-08/2012. p. 1-5.
- « List of academic databases and search engines ». *Wikipedia*. [en ligne]. Disponible sur : [http://en.wikipedia.org/wiki/Academic\\_databases\\_and\\_search\\_engines](http://en.wikipedia.org/wiki/Academic_databases_and_search_engines).
- « List of digital library projects ». *Wikipedia*. [en ligne]. Disponible sur : [http://en.wikipedia.org/wiki/List\\_of\\_digital\\_library\\_projects](http://en.wikipedia.org/wiki/List_of_digital_library_projects).
- Magali Guibert. « Trucs et astuces pour optimiser sa recherche dans LinkedIn ». *Netsources*, n°101, 11-12/2012. p. 10-13.
- Perrine Signoret. « Comment savoir si une image qui circule en ligne est un fake ? ». *Numérana*. 02/03/2019. [en ligne]. Disponible sur : <https://www.numerama.com/tech/468086-comment-savoir-si-une-image-qui-circule-en-ligne-est-un-fake.html>.
- Carole Tisserand-Barthole. « Comment effectuer une recherche d'information efficace sur Google News ? ». *Le blog de FLA consultants*. 26/06/2018. [en ligne]. Disponible sur : <https://www.fla-consultants.com/fr/blog-actualites/recherche-information-efficace-google-news>.
- . « Les *datasets* : un nouvel enjeu pour la recherche d'information et la veille ? ». *Bases*. n°364. 11/2018. p. 1-2. A compléter par les autres articles du même numéro.
- . « L'information sur le web est éphémère : quel impact et quelles solutions pour la recherche d'information ? ». *Bases*. n°363. 10/2018. p. 1-6.
- . « *Open data* : une mine d'or brut pour les professionnels de l'information ». *Bases*, n°333, 01/2016. p. 5-10.
- . « Revues et articles scientifiques en *open access* : où chercher ? ». *Bases*, n°304, 05/2013. p. 1-5.
- . « Thèses françaises en ligne : un joyeux bazar ». *Bases*, n°305, 06/2013. p. 1-5.
- . « Tirer parti de la recherche visuelle et d'images pour la veille ». *Netsources*. n°130, 09-10/2017. p. 1-7.

## Evolutions des moteurs de recherche (focus 2018)

- Carole Tisserand-Barthole. « Comment bien interroger Google en 2018 ? ». *Netsources*. n°133, 03-04/2018. p. 12-13. [en ligne]. Disponible sur : <https://www.fla-consultants.com/fr/blog-actualites/bien-rechercher-sur-google-2018>.
- . « De la recherche classique à la recherche conversationnelle ». *Bases*. n°354. 12/2017. p. 7-9.
- . « "*Featured snippets*" et "*intelligent answers*": Google et Bing vont-ils vraiment répondre à tout ? ». *Le blog FLA consultants*. 09/02/2018. [en ligne]. Disponible sur : <https://www.fla-consultants.com/fr/blog-actualites/featured-snippets-google-veille>.
- . « Google et la « recherche personnalisée » : quel impact pour les professionnels de l'information ? ». *Netsources*. n° 136. 09-10/2018. p. 1-10.
- . « L'avenir de la recherche web sera-t-il thématique ? ». *Bases*. n°351. p. 7-9.
- . « La mort annoncée de la recherche booléenne ». *Bases*. n°340, 09/2016, p. 1-5.
- . « La veille et la recherche d'information à l'heure de la recommandation de contenus ». *Netsources*. n°132, 01-02/2018. p. 1-11.
- . « Recherche sur les réseaux sociaux : quelles solutions gratuites ? ». *Netsources*. n°134, 05-06/2018. p. 6-10.

## Outils de veille (focus 2018-2019)

Anne-Wil Harzing. « How to keep up-to-date with the literature, but avoid information overload ? ». *Harzing.com*. 14/05/2018. [en ligne]. Disponible sur : <https://harzing.com/blog/2018/05/how-to-keep-up-to-date-with-the-literature-but-avoid-information-overload>.

Marie-Laure Malingre. Automatiser sa veille documentaire scientifique. 123 p. 14/03/2018. [en ligne]. Disponible sur : <https://www.sites.univ-rennes2.fr/urfist/ressources/automatiser-sa-veille-documentaire-scientifique>.

Carole Tisserand-Barthole. « Google Alertes est-il encore efficace pour la veille ? ». *Le blog de FLA consultants*. 09/03/2018. [en ligne]. Disponible sur : <https://www.fla-consultants.com/fr/blog-actualites/google-alertes-veille>.

« La veille et la recherche d'information à l'heure de la recommandation de contenus ». *Netsources*. n°132, 01-02/2018. p. 1-11.

-. « Le RSS est-il encore un atout pour la veille ? ». *Bases*. n°351, 09/2017. p. 1-5.

-. « Peut-on encore réaliser une veille avec des outils gratuits en 2017 ? ». *Netsources*. n°131, 11-12/2017. p. 1-8.

-. « RSS, Twitter, Facebook, LinkedIn et les autres : surveille-t-on vraiment la même chose ? ». *Netsources*. n°134. mai-juin 2018. p. 11-12. [en ligne]. Disponible sur : <https://www.fla-consultants.com/fr/blog-actualites/rss-twitter-facebook-linkedin-veille>.

Aurélien Vathonne. « Le retour en grâce de la *newsletter* ». *Netsources*. n°138, 01-02/2019. p. 8-9.

« Veille : les nouveaux fondamentaux ». *Archimag*. Guide pratique n°63. 11/2018. 76 p.

# Suivre l'actualité de ces questions

## Francophones

Olivier Andrieu. *Réacteur*. [en ligne]. Disponible sur : <https://www.reacteur.com> ; lettre *Actu Moteurs*, <https://www.abondance.com/archives-de-la-lettre-actu-moteurs> .

★ *Bases*. François Libman, dir. publ. Paris : Bases publications, 1985-.

Centre de documentation et d'information interuniversitaire en sciences sociales et infopro. *EchosDoc*. [en ligne]. Disponible sur : <http://www.echosdoc.net>.

Christophe Deschamps. *Outils froids*. [en ligne]. Disponible sur : <http://www.outilsfroids.net/>. Compte Twitter : <https://twitter.com/crid>.

Olivier Duffez. *WebRankInfo*. [en ligne]. Disponible sur : <http://www.webrankinfo.com/>.

★ FLA consultants. *Le blog de FLA consultants*. [en ligne]. Disponible sur : <https://www.fla-consultants.com/fr/blog-actualites>.

Béatrice Foenix-Riou. *Le blog de recherche-veillee.com*. [en ligne]. Disponible sur : <http://www.recherche-veillee.com/blog-recherche-veillee> (plus alimenté).  
Compte Twitter : <https://twitter.com/BFoenix>.

*Les infostratèges*. Didier Frochot et Fabrice Molinaro, dir. publ. [en ligne]. Disponible sur : <http://www.les-infostrateges.com/>.

*InternetActu*. [en ligne]. Disponible sur : <http://www.internetactu.net>.

*Le journal du net*. [en ligne]. Disponible sur : <http://www.journaldunet.com/>.

★ *Netsources*. François Libman, dir. publ. Paris : Bases publications, 1996-.

*Presse citron*. [en ligne]. Disponible sur : <http://www.presse-citron.net/>.

*ZDNet*. [en ligne]. Disponible sur : <http://www.zdnet.fr/>.

Zorgloob. Compte Twitter. [en ligne]. Disponible sur : <https://twitter.com/Zorgloob>.

## Anglophones

Karen Blakeman. *Karen Blakeman's blog*. [en ligne]. Disponible sur : <http://www.rba.co.uk/wordpress/>.

Phil Bradley. *Phil Bradley's weblog*. [en ligne]. Disponible sur : <http://philbradlev.typepad.com/>.

Gary Price. *Infodocket*. [en ligne]. Disponible sur : <http://www.infodocket.com/>.

*Research Buzz*. [en ligne]. Disponible sur : <https://researchbuzz.me/>.

Dan Russell. *SearchReSearch*. [en ligne]. Disponible sur : <http://searchresearch1.blogspot.com/>. Pour s'amuser avec Google. Voir aussi <http://www.agoogleadav.com/>.

*Search Engine Journal*. [en ligne]. Disponible sur : <http://www.searchenginejournal.com/>.

*Search Engine Land*. [en ligne]. Disponible sur : <http://searchengineland.com/>.

*Search Engine Watch*. Jonathan Allen, dir. publ. [en ligne]. Disponible sur : <http://searchenginewatch.com>.

Aaron Tay. *Musing about librarianship*. [en ligne]. Disponible sur : <https://musingsaboutlibrarianship.blogspot.com>.

« *Web & Search Tips & Strategies* ». *Lifewire*. [en ligne]. Disponible sur : <https://www.lifewire.com/web-and-search-4102742>.

Marcus P. Zillman. *Scholar Search Engine*. [en ligne]. Disponible sur : <http://www.zillman.us/white-papers/scholar-search-engines/>. nombreux *white papers* mis à jour régulièrement dont moteurs de recherche scientifiques, web invisible, recherche sociale...

# Crédits photographiques

Toutes images : Norman Rockwell, tous droits réservés  
avec, par ordre d'apparition,



Mike Licht, *Two bloggers after Norman Rockwell*. CC BY 2.0  
d'après *Boy gazing at cover girls - Saturday Evening Post*,  
22/09/1934



*It's income tax time again ! - Saturday Evening Post*,  
17/03/1945



*Perpetual motion - Popular Science*, 10/1920



*Boy with stereoscope - Saturday Evening Post*, 14/01/1922



*Family tree - Saturday Evening Post*, 24/10/1959



*Willie Gillis in college - Saturday Evening Post*, 5/10/1946



*Graduate - Saturday Evening Post*, 6/06/1959



*Movie starlet and reporters - Saturday Evening Post*,  
7/03/1936



*Four sporting boys, oh yeah - Brown and Bigelow 1951*  
*Four Seasons calendar, winter*



*The gossips - Saturday Evening Post*, 6/03/1948



*Watchmaker - Publicité pour les Watchmakers of*  
*Switzerland*, 1950



*The art critic - Saturday Evening Post*, 16/04/1955



*Little boy writing a letter - Saturday Evening Post*,  
17/01/1920



*Double take - Saturday Evening Post*, 1<sup>er</sup>/03/1941