

# De la recherche sur Internet

Ecole nationale supérieure d'Architecture Paris Val-de-Seine

Bibliothèque - matériauthèque, mise à jour 9/09/2022

# Sur internet, tout sur tout et aussi... n'importe quoi !

- ❑ Des buzz (nouvelles sans fondement)
- ❑ Des Hoax (rumeurs)...

<http://www.e-scio.net/noire.html>

<http://www.hoaxbuster.com/index.php>

- ❑ Une majorité d'informations commerciales (.com) au détriment de la documentation.
- ❑ Des informations erronées mais bien présentes...

# Les moteurs de recherche

## □ Comment fonctionnent-ils ?

- l'exploration : exploration des hyperliens par des robots.
- l'indexation : copie, extraction des mots significatifs des pages, non des sites.
- la recherche : restitution des résultats à l'utilisateur après application d'un algorithme visant à les classer par pertinence (place du mot dans le document, nombre d'occurrences...) et par popularité.
- Quand on fait une recherche via Google, on recherche dans son index (la dernière copie de la page). Quand on clique sur le titre dans la liste de résultats en revanche on accède à la page du site en direct. Google liste des pages, non des sites.

# Web visible / invisible

- ❑ Les moteurs de recherche généralistes n'indexent qu'une partie du web (10%).
- ❑ Ils n'indexent pas ou peu le web invisible : les bases de données, les bibliothèques en ligne, les pré-prints, certains formats ou des pages trop volumineuses, toute page accessible via un formulaire, derrière un mot de passe, un système d'authentification....
- ❑ Pour interroger le web invisible, se demander dans quel type de base de données on peut trouver tel résultat.

# GOOGLE

- ❑ Google : un moteur de recherche très puissant, quasi hégémonique (actuellement 92,3 % du trafic en France)
- ❑ Mais comment fonctionne t-il réellement ?

<https://www.abondance.com/20160129-16083-infographie-comment-fonctionne-google-2.html>

- 👍 Le catalogue de pages indexées le plus vaste et le plus réactif qui soit, des algorithmes continuellement améliorés pour classer les résultats avec efficacité et pertinence (module d'intelligence artificielle Rankbrain, prise en compte du nombre de liens signalant un document (PageRank) etc.
- 👎 Google exploite allègrement les données privées, personnalise de plus en plus et tend à interpréter la requête. Page nouvelle ou très spécialisée peu signalée.

# Recherche simple / avancée

- ❑ La recherche simple est souvent décevante...

Analyser les 100 premiers résultats pour voir comment Google reformule la requête.

- ❑ Privilégiez la recherche avancée qui permet d'élaborer une recherche plus pertinente, d'effectuer un tri par date, par langue, par format de fichier, par droits d'usages, de choisir l'emplacement des mots-clefs interrogés (titre, corps du texte, url, hyperliens...)
- ❑ Evitez les termes polysémiques ou larges.

# Recherche avancée

Accès par l'icône paramètres en haut à droite : 

## Réglages rapides

[Afficher tous les paramètres](#)

## Confidentialité

Historique des recherches

Enregistrement activé

Vos données dans la recherche Google

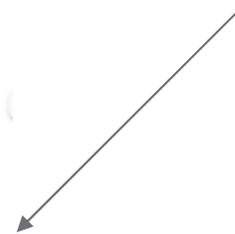
## Utilisation de la recherche

Filtre de résultats explicites

[En savoir plus sur SafeSearch](#)

Langues (Langues)

Recherche avancée



# Recherche avancée



## Recherche avancée

Trouvez des pages avec...

tous les mots suivants :

Pour effectuer cette opération dans le champ de recherche

Saisissez les mots importants : `terrier tricolore`

ce mot ou cette expression exact(e) :

Ajoutez des guillemets autour des mots exacts : `"terrier"`

l'un des mots suivants :

Saisissez OR entre tous les mots à inclure : `miniature OR standard`

aucun des mots suivants :

Placez un signe - (moins) devant les mots à exclure : `-rongeur, -"Jack Russell"`

nombres compris entre :

 et 

Placez deux points entre les nombres, et ajoutez une unité de mesure : `10..35 kilos, 300..500 USD, 2010..2011`

Affinez ensuite la recherche par...

langue :

Rechercher des pages dans la langue sélectionnée

région :

Rechercher des pages publiées dans une région précise

dernière mise à jour :

Rechercher des pages mises à jour durant la période spécifiée

site ou domaine :

Rechercher sur un site (tel que `wikipedia.org`) ou limitez vos résultats à un domaine tel que `.edu`, `.org` ou `.gov`

termes apparaissant :

Rechercher des termes dans la page entière, dans le titre d'une page, dans une adresse Web ou dans des liens vers la page recherchée

SafeSearch :

Configurez **SafeSearch** de manière à filtrer ou non les contenus à caractère sexuel explicite.

type de fichier :

Rechercher des pages dans le format que vous préférez

droits d'usage :

Rechercher des pages que vous êtes libre d'utiliser

Recherche avancée

# Pour préciser une recherche

- ❑ Anonymiser : <https://www.startpage.com/do/search>
- ❑ Rechercher sur une expression exacte : «... »
- ❑ Ouvrir une recherche aux synonymes d'un terme : ~ (tilde devant un mot)
- ❑\* sur Google, remplace un mot, pas une lettre
- ❑ Saisir plusieurs termes de recherche pour lever les homonymies.

# Recherche booléenne

- ❑ '+' correspond à l'opérateur booléen « ET » ou au champ *tous les mots suivants* en recherche avancée Google (le 'et' s'applique par défaut).
- ❑ '|' ou OR correspond à l'opérateur booléen « OU » et au champ *au moins un des mots suivants* en recherche avancée Google.
- ❑ '-' (devant le mot à exclure) correspond à l'opérateur booléen « SAUF » ou à *aucun des mots suivants* en recherche avancée Google.

# Autres clefs de recherche -1-

- ❑ « **Site:** » : restreint la recherche à un site  
Ex : [Tadao Ando site:centrepompidou.fr](#)
- ❑ « **filetype:** » (ou « **ext:pdf** ») : restreint la recherche à un type de fichier  
Ex : [construction terre crue filetype:pdf](#)
- ❑ Les requêtes **intext:**, **intitle:**, **inurl:**, **inanchor:** permettent de limiter la recherche du mot-clef respectivement dans le corps du texte, le titre, l'url, les liens.

# Autres clefs de recherche -2-

- ❑ **Define:** permet de rechercher la définition d'un terme

Ex : Define: « économie circulaire »

- ❑ **Author:** permet de restreindre la recherche par auteur.

Ex : esthétique author: « Emmanuel Kant »

- Pour recherche à l'international : saisir les termes voulus en anglais. En revanche, interroger google.fr ou google.co.uk ne change pas les résultats, seulement l'ordre d'affichage.

# Recherches spécifiques

La plupart des moteurs de recherche permettent aussi des recherches ciblées sur :

- Les images (ex : <https://images.google.com/>)
- Les livres (ex : <https://books.google.fr/>)
- Les actualités (ex : <http://news.google.fr/>)

(ne sonde pas que les sites de presse)

- Des travaux universitaires (articles, thèses etc. : <http://scholar.google.fr/>)

# La recherche d'images

- ❑ **Getty images**

<https://www.gettyimages.fr/>

- ❑ **Google images**

<https://images.google.com/>

L'indexation des images est plus complexe que celle des textes (polysémie de l'image, éléments secondaires, contexte historico-social, subjectivité, etc.) ou alors le moteur n'indexe que le texte associé à l'image, à la vidéo, non le contenu de l'image.

# La recherche avancée d'images

[https://www.google.fr/advanced\\_image\\_search?](https://www.google.fr/advanced_image_search?)

La recherche avancée vous permet de préciser le format, la taille de l'image, les droits sur l'image, le type d'images (dessin, photographie etc.).

Pensez à analyser l'url, privilégiez les résultats référencés avec titre, date, technique, format, copyright et localisation...

On peut chercher une image en saisissant des mots-clés (cherche alors sur la légende) ou en déposant une image dans le champ de recherche directement ou en la chargeant via 

# POP -1-

<https://www.pop.culture.gouv.fr/>

La Plateforme Ouverte du Patrimoine interroge notamment :

- ❑ la base de données '**Joconde**' des musées de France ;
- ❑ la base de données '**Mérimée**', dédiée au patrimoine architectural français.

# POP -2-

Ex : une recherche sur Auguste Perret en sélectionnant les bases Joconde et Mérimée donne accès à la coupe du Mobilier National.

<https://www.pop.culture.gouv.fr/notice/merimee/PA00086602>

= des références précises, un auteur, un titre, une datation, des données techniques, une description....

Une information réutilisable et validée par des professionnels...

# Culture numérique

❑ Maintes institutions (bibliothèques, musées etc.) ont engagé des programmes de numérisation, souvent sources de documents numériques de qualité.

❑ En France :

<http://www.numerique.culture.fr/pub-fr/index.html> (site désactivé)

❑ A l'échelle européenne :

<http://www.michael-culture.eu/>

# Les catalogues de bibliothèques

Recherche d'une information bibliographique : un catalogue de bibliothèque répond de manière objective, pour peu qu'on sache l'utiliser....

- ❑ Ex : Catalogue des bibliothèques d'ENSA

<https://www.archires.archi.fr/>

- ❑ Catalogue des bibliothèques universitaires françaises

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

- ❑ Catalogue de la Bibliothèque nationale de France

<http://www.bnf.fr>

# Les sites collaboratifs

□ <https://fr.wikipedia.org/>

Des informations justes et utiles existent :

[http://fr.wikipedia.org/wiki/Pierre\\_Bourdieu](http://fr.wikipedia.org/wiki/Pierre_Bourdieu)

D'autres ne sont pas abouties, voire  
erronées, sans citation des sources...

[https://fr.wikipedia.org/wiki/Tadao\\_And%C5%8D](https://fr.wikipedia.org/wiki/Tadao_And%C5%8D)

□ <https://fr.wikisource.org/>

De nombreux textes mais publiés par  
n'importe qui. Pour une édition validée,  
optez une bibliothèque institutionnelle.

# D'autres moteurs de recherche existent...

## □ Quelques exemples :

<http://www.exalead.fr/>, qui permet de relancer une recherche sur des termes associés, par type de media, type de sites etc.

<https://duckduckgo.com/>,

<https://www.qwant.com/?l=fr>, qui ne collectent pas les données utilisateurs.

<https://www.yippy.com/>, qui propose ses résultats en *clustering*. Chaque page indexée est analysée et reliée à d'autres en fonction de son contenu.

# Des moteurs à orientation universitaire

- ❑ <https://scholar.google.fr/>, indexe les sites universitaires et les éditeurs universitaires
- ❑ <https://www.base-search.net/>, capable de trouver des ressources universitaires non indexées par les moteurs commerciaux.
- ❑ <http://www.freefullpdf.com/#gsc.tab=0>, qui permet d'accéder à des publications scientifiques en libre accès.
- ❑ <https://www.scienceresearch.com/scienceresearch/desktop/en/search.html>
- ❑ Des méta-moteurs, comme Copernic : <http://www.copernic.com/fr>, qui propose une veille sur une équation de recherche, élimine les doublons...

# Gardez l'esprit critique !

- ❑ Mettez les résultats de recherche sur Internet en perspective...
- ❑ Sachez comparer et juger l'information.
- ❑ Les catalogues et bases de données sont des sources d'information fiables et validées par des professionnels de l'information.
- ❑ Citez vos sources au plus précis : le droit d'auteur existe, non au plagiat !

# Evaluer l'information sur Internet -1-

*Faute de filtres (éditeur, bibliothèque) et sachant que le numérique aplanit les différences entre documents, c'est à vous de juger de l'information trouvée sur Internet.*

## **❑ Définissez le type d'information cherchée et l'objectif de la recherche :**

Identifier l'auteur ? (réputation, crédibilité, compétences dans le domaine?), l'éditeur ?

Décrypter l'url.

Si vous accédez à la page via un moteur, remontez à la page d'accueil du site.

# Evaluer l'information sur Internet -2-

- ❑ **Identifiez l'objectif du site :**

Commercial ? Altruiste ? Égoïste ? Prosélyte ?

Cet objectif est-il clairement énoncé ?

- ❑ **Nature et limites de l'information ?**

Type d'information proposée : opinion, polémique, résultat d'une recherche, annonce ?

# Evaluer l'information sur Internet -3-

❑ **Jugez de la pertinence de la ressource par rapport à votre recherche.**

Profondeur de l'information ?

Documentation (références bibliographiques, citation des sources utilisées) ?

Stabilité de l'information ?

Qualité de l'expression et ton ? Présentation des informations (clarté, précision, exhaustivité, logique et argumentation ? Qualité visuelle et sonore ?

Ergonomie ?)

Qualité, cohérence et fraîcheur des liens proposés ?

# Evaluer l'information sur Internet -4-

- ❑ **Jugez de la fraîcheur de l'information** (mise à jour du site).
- ❑ **Ecartez les sites présentant une information approximative.**
- ❑ **Croisez les sources :**  
Confrontez des sources voisines ; vérifiez sur des documents publiés.