

Usage des moteurs de recherche pour la recherche documentaire

Module 1.3



Auteur : F. Flamerie - Ce contenu est mis à disposition selon les termes de la [Licence Creative Commons Attribution - Partage dans les Mêmes Conditions 3.0 France](https://creativecommons.org/licenses/by-sa/3.0/fr/).

Programme

Google et Google Scholar peuvent-ils être utilisés pour la recherche documentaire? Pour quel usage ? Cette session propose une présentation pratique et critique de ces 2 moteurs de recherche, ainsi que des alternatives académiques non commerciales, offrant de meilleures garanties de transparence.

- Optimiser l'usage de Google et Google Scholar
- Biais et limites de Google et Google Scholar
- Moteurs de recherche complémentaires ou alternatifs



Optimiser l'usage de Google et Google Scholar

Fonctionnement des moteurs de recherche web

Syntaxe de recherche

Intégrations : Zotero et navigateur

Accéder à l'information scientifique et technique

SUR LE WEB



Web visible

Regroupe toutes les **pages indexées** par les moteurs de recherche généralistes et que l'on retrouve dans les résultats de recherche.

Web invisible

Regroupe les **pages non indexées** : intranet, boîte mail, compte bancaire... Ainsi que des **pages indexées par les moteurs de recherche spécifiques** notamment scientifiques : bases de données académiques, entrepôts de données en ligne...

Dark Web

Regroupe des **pages non indexées** et utilisant des **protocoles d'accès cryptés** aux informations du World Wide Web. Ces pages ne sont accessibles que par des navigateurs spécifiques (par ex. Tor) et nécessitent de connaître leur URL.

Fonctionnement des moteurs de recherche web

« Les moteurs de recherche web interrogent une partie des ressources d'Internet appelée le « **web visible** ». Pour schématiser, il s'agit des différentes pages, blogs, images, vidéos... qui sont parcourus et récupérés par les robots d'indexation, puis indexés et restitués à l'aide d'algorithmes lorsque vous interrogez un moteur de recherche web. »

Source : Réseau Ascodocpsy. (2021, janvier 12). *[Allô la doc ? #6] Les moteurs de recherche - Épisode 1 : Introduction aux moteurs de recherche web*. Ascodocpsy.

<https://www.ascodocpsy.org/allo-la-doc-6-les-moteurs-de-recherche-episode-1-introduction-aux-moteurs-de-recherche-web/>

NB Google Scholar -> accords avec les éditeurs scientifiques pour l'indexation des contenus soumis à abonnement

Syntaxe de recherche Google

→ Opérateurs

AND opérateur implicite

OR

- équivaut à NOT, exclut le mot qui précède

« » recherche d'expression

→ Préciser un champ d'interrogation

site: pour préciser un nom de domaine ou de site

filetype: pour préciser un type de fichier

intitle: vérifie la présence de l'un des mots dans le titre de la page web

allintitle: vérifie la présence de TOUS les mots dans le titre de la page web

inurl: chercher seulement dans l'URL

Syntaxe de recherche Google

Formulaire de recherche avancée -

https://www.google.fr/advanced_search?hl=fr&fg=1

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"`

Syntaxe de recherche Google

A votre avis, quelle est la meilleure formule à saisir dans la boîte de recherche de Google pour trouver des documents gouvernementaux français sur les risques psychosociaux?

1. risques psychosociaux France gouvernement
2. site:.gouv.fr filetype:pdf "risques psychosociaux"
3. site:.gouv.fr "risques psychosociaux"

Syntaxe de recherche Google

A votre avis, quelle est la meilleure formule à saisir dans la boîte de recherche de Google pour trouver des documents gouvernementaux français sur les risques psychosociaux?

1. risques psychosociaux France gouvernement
2. **site:.gouv.fr filetype:pdf "risques psychosociaux"**
3. site:.gouv.fr "risques psychosociaux"

NB : 2023-04-05 : la requête 2 retrouve 3970 résultats / la requête <site:.gouv.fr filetype:pdf risques psychosociaux> en retrouve 9400 -> quasiment aucune différence dans les 6 premières pages de résultats

Syntaxe de recherche Google Scholar

→ Opérateurs

AND opérateur implicite

OR

- équivaut à NOT, exclut le mot qui précède

« » recherche d'expression

→ Préciser un champ d'interrogation

author: chercher sur l'auteur du document

allintitle: chercher seulement dans le titre du document

Syntaxe de recherche Google Scholar

Formulaire de
recherche avancée -
https://scholar.google.fr/#d=gs_asd

×

Recherche avancée

🔍

Articles contenant

tous les mots suivants

cette **expression exacte**

au moins un des mots suivants

aucun des mots suivants

mes mots

☒ à n'importe quel endroit de l'article

☐ dans le titre de l'article

Rechercher des articles **rédigés**

par

ex. : "P Flajolet" ou J Fauvet

Rechercher des articles **publiés**

dans

ex. : *Journal de la Société de biologie* ou *Revue Romane*

Rechercher des articles **datés**

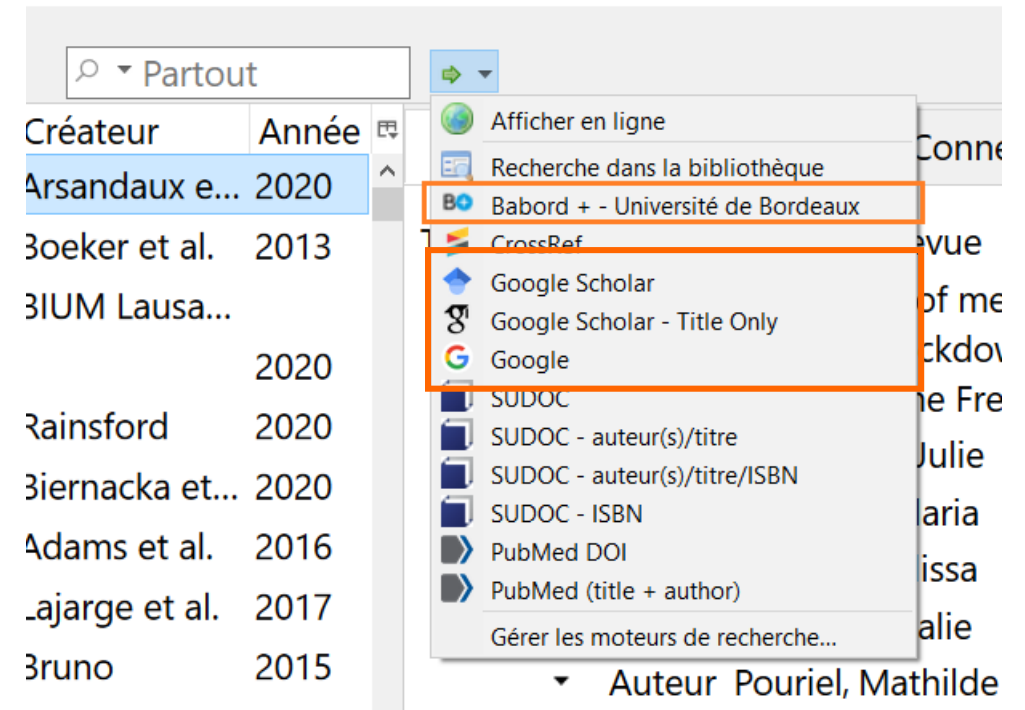
de

—

ex. : 1996

Intégration Zotero : Google et Google Scholar

- Cette intégration permet de lancer une recherche dans le moteur choisi, à partir d'un ou plusieurs documents de la bibliothèque Zotero.
- Usage : compléter des informations bibliographiques ou trouver le texte intégral
- Voir le tuto doc'Ispep [Lancer une recherche dans les ressources de votre institution depuis Zotero](#)



Intégration navigateur : Google Scholar

Extension « Bouton Google Scholar »

- › Ouvrir une fenêtre de recherche dans Google Scholar
- › Sur la page d'un article ou à partir de sa citation en bibliographie, afficher le résultat de la recherche Google Scholar correspondante

The screenshot shows a web browser window with a JSTOR article page. The article title is "Cancer epidemiology, prevention and control" by N. S. Murthy* and Aleyamma Mathew. The article is from "Current Science", Vol. 86, No. 4 (25 February 2004), pp. 518-527 (10 pages). The article is published by Current Science Association. The article is available as a PDF on researchgate.net. The article is also available on Google Scholar. The Google Scholar extension is visible in the top right corner of the browser window, showing the Google Scholar logo and a search bar. The extension is also visible in the bottom right corner of the browser window, showing the Google Scholar logo and a search bar. The extension is also visible in the bottom right corner of the browser window, showing the Google Scholar logo and a search bar.

» Voir notre tuto doc'Ispep : [Trouver rapidement le texte intégral d'une publication scientifique](#)



Biais et limites de Google et Google Scholar

Couverture

Fonctionnalités

Biais

Quel usage? Publications scientifiques et littérature grise

Couverture

→ Indéfinie

- › Pas de liste de sources indexées
- › NB Google Scholar indexe à l'unité des documents, des articles, et non des revues

→ Mouvante

- › Reflet de ce qui est disponible en ligne
-> ce qui ne l'est plus est retiré de l'index
- › Cela est particulièrement problématique pour un usage de recherche bibliographique ou de compte de citations

"In a bibliographic database that is intended for literature discovery, coverage stability is a desirable property, if we assume that users intuitively expect a system to retrieve the same documents over time given the same query (in addition to new documents that also meet the search criteria). [...]. In a citation index, the disappearance of a document would affect the citation counts of all its cited documents, impeding some types of citation analysis."

Source : Martín-Martín, A., & López-Cózar, E. D. (2021). Large coverage fluctuations in Google Scholar : A case study. arXiv, 102.07571 [cs.DL]. <http://arxiv.org/abs/2102.07571>

Fonctionnalités

- Syntaxe de recherche pauvre
 - › Pas de troncature ni de caractères de remplacement
 - › Requêtes booléennes avec parenthèses non prises en charge
 - › Google Scholar : limite 256 caractères / Google : 100
- Pas ou peu de filtres
- Pas d'historique de recherche

“In terms of functionality, two broad types of search systems exist at present: the traditional “transparent-comprehensive” (eg, ProQuest, PubMed, Web of Science) and the newer “efficient-slick” (eg, Google Scholar, Semantic Scholar).”

Source : Gusenbauer, M., & Haddaway, N. (2021). What every Researcher should know about Searching – Clarified Concepts, Search Advice, and an Agenda to improve Finding in Academia. *Research Synthesis Methods*, 12(2), 136-147.
<https://doi.org/10.1002/jrsm.1457>

Fonctionnalités

→ Pas d'accès à tous les résultats

- › Affichage des 1000 premiers résultats seulement
- › Par tranche de 20 résultats (non modifiable) dans Google Scholar
- › Interruption de service possible par Google Scholar si usage détecté comme trop intensif

→ Fonctionnalités d'export quasiment nulles

- › En utilisant le connecteur Zotero, export par lot possible depuis Google Scholar, limité aux 20 résultats affichés sur la page
- › Logiciel [Publish or Perish](#) pour exporter jusqu'aux 1000 résultats autorisés

Fonctionnalités

PubMed
GS

Reproducible search	●	◉
Export of all search results	●	◉
Search history: edit/download	●	○
Search strategy documentation	●	◉
Advanced Search	●	◉
Forward citation search	◉	●

Source : Klopfenstein, D. V., & Dampier, W. (2021).
 Commentary to Gusenbauer and Haddaway 2020 : Evaluating
 retrieval qualities of Google Scholar and PubMed. *Research
 Synthesis Methods*, 12(2), 126-135.
<https://doi.org/10.1002/jrsm.1456>

- Better
- ◉ Good
- ◉ Minimal
- Missing

Biais

→ Absence de transparence et de reproductibilité

- › Couverture (cf. supra)
- › Algorithmes opaques - NB Google prend en compte l'historique de recherche du navigateur pour le classement des résultats

→ Biais de confirmation

→ Biais de publication

- › Google Scholar favorise les articles publiés vs. la littérature grise
- › ... alors que la recherche de littérature grise vise à minimiser ce biais
- › Voir : Haddaway, N., Collins, A. M., Coughlin, D., & Kirk, S. (2015). The Role of Google Scholar in Evidence Reviews and Its Applicability to Grey Literature Searching. *PLoS ONE*, 10(9), e0138237. <https://doi.org/10.1371/journal.pone.0138237>

“Second, we must stay alert as these efficient-slick systems aim at transforming ‘inefficient’ exploratory searching into ‘efficient’ lookup searching (eg, through presentation of pre-selected cues). This means exploratory searching (and thus learning) might be more and more crippled toward quick, unconsciously biased lookup searching (cherry picking) that users more and more expect when engaging with online systems”

Source : Gusenbauer, M., & Haddaway, N. (2021). What every Researcher should know about Searching – Clarified Concepts, Search Advice, and an Agenda to improve Finding in Academia. *Research Synthesis Methods*, 12(2), 136-147. <https://doi.org/10.1002/jrsm.1457>

Quel usage? - publications scientifiques

Conclusion de l'article de Klopfenstein & Dampier + cf. supra intégrations -> localisation rapide du texte intégral

“GS excels for simple lookup tasks, like finding a paper by entering its title in the query box. Both GS and PubMed can be used for exploratory searches, but we urge biomedical researchers to use PubMed rather than GS, because PubMed is one of the top recommended primary sources for literature searches of peer-reviewed research in the biomedical sciences and has search feature criteria that GS has lacked since its inception.”

Source: Klopfenstein, D. V., & Dampier, W. (2021). Commentary to Gusenbauer and Haddaway 2020 : Evaluating retrieval qualities of Google Scholar and PubMed. *Research Synthesis Methods*, 12(2), 126-135. <https://doi.org/10.1002/jrsm.1456>

Quel usage? - littérature grise

Les moteurs de recherche (Google, Google Scholar) sont souvent utilisés pour la recherche de **littérature grise**.

Définition dite « de Luxembourg » de la littérature grise

« ce qui est produit par toutes les instances du gouvernement, de l'enseignement et la recherche publique, du commerce et de l'industrie, sous un format papier ou numérique, et qui n'est pas contrôlé par l'édition commerciale »

Source : Schöpfel, Joachim. « Comprendre la littérature grise ». *I2D Information, données documents* 52, n° 1 (3 avril 2015): 30-32.
<https://doi.org/10.3917/i2d.151.0030>.

Méthodes de recherche de littérature grise

Le « Googling »
n'est toutefois
qu'une méthode
parmi d'autres.
La recherche de
littérature grise
peut aussi rendre
nécessaire
l'exploration
manuelle d'un site.



- Database searching (including specialized databases and search portals)
- Searching in obscure or small library catalogues
- Hand-searching of journals
- Personal communication (i.e. telephone, email, etc.)
- Scanning reference lists; snowballing, bibliographies and academic CVs [tip: read more about [snowballing techniques](#)]
- Googling ([Google](#) or [Google Scholar](#))
- Other search engines, such as [Bing](#), [Yahoo](#), [Microsoft Academic Search](#), [DuckDuckGo](#)
- Blogging (finding the experts)
- Blogsearch, podsearch, specialized directories

Source : Schoonmade, L. (2022). *Grey Literature for Health Sciences*. LibGuides Vrije Universiteit Amsterdam.

<https://libguides.vu.nl/greylit>

Organiser la recherche de littérature grise

Modèle au format Word pour documenter une recherche de littérature grise en 2 étapes et 4 stratégies de recherche

→ Etape 1 : identifier les autorités pertinentes

→ Etape 2 : stratégies de recherche et documentation de ces recherches

- › Stratégie 1 : consultation/recherche dans des sites ciblés : 1 site à la fois, l'un après l'autre
- › Stratégie 2 : recherche dans des bases de données de littérature grise
- › Stratégie 3 : recherche au moyen de moteurs de recherche
- › Stratégie 4 : contact des experts du domaine

Voir : Nekolaichuk, E. (2023). *Searching the Literature : A Guide to Comprehensive Searching in the Health Sciences*. University of Toronto Library Research Guides. <https://guides.library.utoronto.ca/comprehensivesearching>, rubrique *Grey Literature*

Littérature grise : ressources monde 1/2

→ [Guidelines International Network](https://g-i-n.net/international-guidelines-library) - recommandations

 <https://g-i-n.net/international-guidelines-library>

Filtres par pays d'application, statut de publication, *endorsing member organization*, *publication scope*, etc.

→ [Iris \[OMS\]](https://apps.who.int/iris) - Institutional Repository for Information Sharing

 <https://apps.who.int/iris>

Publications de l'OMS, organisées par communautés (bureaux régionaux) et collections (ex : documents techniques, publications, articles de revue)

→ [Nations Unies : Bibliothèque numérique](https://digitallibrary.un.org)

 <https://digitallibrary.un.org>

Publications de l'ONU, organisées par organes de l'ONU (ex: fonds et programmes) et types de ressource (ex : documents et publications, discours, etc.)

Littérature grise : ressources monde 2/2

→ [Science.gov](http://www.science.gov) - production des agences gouvernementales [Etats-Unis]



<http://www.science.gov>

60 bases de données et plus de 2,200 sites web tous domaines scientifiques confondus - domaine Health & Medicine -> ClinicalTrials.gov, MedlinePlus, Cancer.gov, etc.

→ [Catalogue du Réseau Santécom](https://catalogue.santecom.qc.ca) - base documentaire de l'Institut national de santé publique au Québec[Canada]



<https://catalogue.santecom.qc.ca>

Publications de 50 organismes et groupes de recherche + ressources de 37 bibliothèques spécialisées - recherche avancée - typologie documentaire incluant "Veille informationnelle", "Rapport d'évaluation des technologies et modes d'intervention en santé et services sociaux"

Littérature grise : ressources France 1/2

→ [BNSP](https://www.bnsp.insee.fr/bnsp) - Bibliothèque nationale de la statistique publique - Insee

 <https://www.bnsp.insee.fr/bnsp>

Toutes les publications de la statistique publique française depuis le 19^e siècle - Navigation par thématique, région & pays, période, partenaire (ex: SSM Santé et solidarités - DREES)

→ [DREES > Publications](https://drees.solidarites-sante.gouv.fr)

 <https://drees.solidarites-sante.gouv.fr>

Toutes les productions de la DREES (publications, données, etc.), avec des services supplémentaires tels que la visualisation de données

→ [SCORE-Santé](https://www.scoresante.org/index.html) - données des observatoires régionaux de la santé

 <https://www.scoresante.org/index.html>

Accès aux données STATISS, à la base d'indicateurs SCORE-Santé (niveau territorial/national), à des publications comme les profils de territoire

Littérature grise - ressources France 2/2

→ [Portail documentaire de l'EHESP](https://documentation.ehesp.fr/index.php?search_type_asked=simple_search) - Ecole des hautes études en santé publique



https://documentation.ehesp.fr/index.php?search_type_asked=simple_search

Recensement des productions de l'EHESP (mémoires, etc.) et d'autres sources de littérature grise ; dépouillement de revues.

→ Bibliographies commentées, pouvant comporter des références de littérature grise

› EHESP : fiches synthétiques et dossiers documentaires



https://documentation.ehesp.fr/index.php?lvl=cmspage&pageid=6&id_rubrique=311

› IRDES : synthèses documentaires et bibliographies thématiques



<https://www.irdes.fr/documentation/syntheses-et-dossiers-bibliographiques.html>



Moteurs de recherche complémentaires ou alternatifs

Google Scholar -> BASE - Bielefeld Academic Search Engine et CORE

Google -> Doc'CISMeF et HONcode

BASE - Bielefeld Academic Search Engine

<https://base.ub.uni-bielefeld.de>

- Indexation des métadonnées seulement (et non du texte intégral)
- NB PubMed = une des sources de BASE
- Critère « Type de publication » pour cibler la littérature grise
- [Syntaxe de recherche](#)

Vue partielle du formulaire

 **BASE**

Se connecter Frang

Recherche simple **Recherche avancée** Navigation Historique de recherche

Recherche avancée

Document entier	<input checked="" type="checkbox"/>	
Titre	<input checked="" type="checkbox"/>	
Auteur	<input checked="" type="checkbox"/>	
ORCID iD	<input checked="" type="checkbox"/>	
Entrées matière	<input checked="" type="checkbox"/>	
DOI	<input checked="" type="checkbox"/>	
(Partie de l')URL	<input checked="" type="checkbox"/>	
10 Résultats par page	<input checked="" type="checkbox"/>	<input type="checkbox"/> Privilégier des documents en accès libre

Type de document

☒ **Tous**

☒ **Text**

- ☐ Book
 - ☐ Book part
- ☐ Journal/Newspaper
 - ☐ Article contribution
 - ☐ Other non-article
- ☐ Patent
- ☐ Thesis
 - ☐ Bachelor thesis
 - ☐ Master thesis
 - ☐ Doctoral and postdoctoral thesis
- ☐ Conference object
- ☐ Report
- ☐ Review
- ☐ Course material
- ☐ Lecture
- ☐ Manuscript

Accès

BASE / Google Scholar ?

Couverture

In comparison to commercial search engines, BASE is characterised by the following features:

- Content providers are indexed only after check by qualified personnel of Bielefeld University Library
- Only document servers and journals that comply with the specific requirements of academic quality and relevance are included
- Our [list of content providers](#) provides transparency in the searches
- Discloses web resources of the "Deep Web", which are ignored by commercial search engines or get lost in the vast quantity of hits
- Correction, normalization and enrichment of metadata by means of automated methods
- Multilingual search (find search terms in more than 20 translated languages)
- The display of search results includes precise bibliographic data
- Several options for sorting the result list (by author, title, date)
- "Refine your search result" options (by author, subject, DDC, year of publication, content provider, language, document type, access and terms of re-use)
- [Browsing](#) by DDC (Dewey Decimal Classification), document type, access and terms of re-use / licence.

Fonctionnalités de recherche

CORE - <https://core.ac.uk>



The world's largest collection of open access research papers

- Moteur sémantique
- Indexe uniquement des documents en libre accès - types : *research, thesis, unknown, slides*
- [Liste des fournisseurs de données](#)
- Recherche y compris dans le texte intégral des documents
- Filtres - !\ fiabilité aléatoire, dépend des informations fournies par les sources
- Prend en charge les requêtes complexes - cf. diapo suiv.

The screenshot displays the CORE search interface. At the top, there is a search bar with the text "Search 207,255,818 papers around the world" and a "SEARCH" button. Below the search bar, the search query "title:'obesity AND depression'" is entered. The interface shows 443 research outputs found. A filter menu is open, showing options for "Research" (27), "Thesis" (3), "Slides" (2), and "Unknown" (1). A sample result is visible, titled "Childhood Obesity" by Horton Leslie N, with a "Get PDF" button. The result snippet mentions "Obesity is a cor... suffer from obe...". The Marshall University logo is also visible.

CORE

You are able to search across all fields including title, abstract, authors and full text (to find out about all the fields that you can search, please refer to the [API documentation](#)).



To refine your searches you can use:

`" "` - to specify exact matching

`AND, OR` - to add logic to to your search

`()` - to group your search terms

`property_name:"value"` - a value for a specific property, e.g. *title:"Connecting Repositories", doi:"10..."*

`><=` - to query on numeric fields, e.g. *year>2020*

`_exists_:"field_name"` - to get all the records that have a certain field e.g. `_exists_:issn`

To get you started quickly follow this links to prefill your search box:

SEARCH BY DOI

SEARCH BY TITLE

SEARCH BY OAI

SEARCH BY ISSN

Core / Google Scholar ?

“Google Scholar is a search engine containing scholarly research papers but it is not designed to aggregate repository and journal systems. More specifically:

- Google Scholar crawls and indexes research papers that can be found on the web and links to the original source, while CORE harvests and caches the full text.
- Google Scholar limits its access only at the [granularity level](#), i.e. its search engine, whereas CORE is in position to extend access to raw data and, apart from the CORE search engine.
- Google Scholar offers both closed access and open access resources, while CORE offers open access resources only, enabling immediate access to full text.
- The audience is different. Even though CORE has a search engine where users can retrieve scientific literature, it focuses also on other types of research stakeholders, such as text miners and repository managers, and offers services designed for them, such as the [CORE API](#) and [Dataset](#).”

Source : Page du site CORE [How is CORE different from Google Scholar?](#)

Doc'CisMeF et HONCode

Ces moteurs de recherche spécialisés ont une **couverture définie et contrôlée**.

→ **Doc'CiSMeF** - <https://doccismef.chu-rouen.fr>

- › Produit par le CiSMeF -> interrogation en rebond depuis [LiSSa](#) - ressources en français
- › Interroge aussi des sources bibliographiques : catalogue Sudoc, mémoires sur DUMAS, etc.

→ **HONcode Search** - <https://www.hon.ch/en/search.html>

- › Produit par l'ONG Health On the Net - couverture mondiale
- › [Infos sur la certification HONcode](#)
- › Disponible sous forme d'**extension** pour votre navigateur : [HONcode Toolbar](#)



Merci pour votre attention

Des questions?

Contact : doc.isped@u-bordeaux.fr