

Utiliser des moteurs de recherche académiques gratuits, en 7 points

1. Moteurs de recherche *versus* bases de données
2. Critères pour évaluer un moteur de recherche
3. Google Scholar (2004), le plus ancien, le plus connu
4. Dimensions (2018), moteur se présentant comme une base de données
5. Semantic Scholar (2015), moteur basé sur l'intelligence artificielle
6. Lens (1998), moteur de recherche sophistiqué et expert
7. Scinapse (2019), méta-moteur académique

Liens utiles

1. Moteurs de recherche *versus* bases de données

Jusqu'en 2004, année du lancement du moteur de recherche académique [Google Scholar](#), la recherche d'information scientifique sur internet passait par les bases de données bibliographiques pour la plupart payantes (comme les bases pluridisciplinaires Web of Science ou Scopus), auxquelles s'abonnent les établissements de recherche et d'enseignement supérieur (voir les fiches CoopIST [Choisir les sources](#) et [Interroger les sources](#)).

Aujourd'hui, la gratuité des moteurs de recherche sur internet et leur accès facile ont favorisé leur usage par tous. L'adoption rapide de Google Scholar par les scientifiques a concurrencé les bases de données bibliographiques. Celles-ci ont réagi en rendant leur utilisation plus aisée.

Le développement du libre accès aux résultats de recherche (publications, jeux de données) et aux métadonnées les décrivant (voir fiche CoopIST [Agir pour le libre accès](#)) a accru la compétition entre les bases de données et les moteurs de recherche.

Les utilisateurs combinent les deux types d'outils :

- **les bases de données bibliographiques** : leurs fonctionnalités poussées permettent de répondre à des requêtes complexes. Elles ont une couverture maîtrisée, liée à une stratégie éditoriale élaborée. La création, l'enrichissement, le stockage, la qualité et la valorisation de leur contenu (références bibliographiques, index d'auteurs, index des organismes d'affiliation des auteurs, des revues sélectionnées, thésaurus, etc.) passent par des opérations manuelles de sélection, de référencement, d'indexation, de classification et de contrôle qualité ;
- **les moteurs de recherche académiques**, c'est-à-dire spécialisés dans les ressources documentaires scientifiques : ce sont plutôt des outils exploratoires, complémentaires aux bases de données scientifiques. Ils fouillent et indexent de manière automatique les contenus des bases de données librement accessibles (archives ouvertes de publications, entrepôts de données de recherche, bases de brevets gratuites...). Leur valeur ajoutée est liée à la sophistication de leurs traitements automatisés sur les contenus qu'ils indexent (métadonnées, texte intégral). Mais leur périmètre est imprécis et leur stratégie souvent peu transparente.

2. Critères pour évaluer un moteur de recherche

A la suite de Google Scholar, mais plus tardivement, vers 2015, de nouveaux moteurs de recherche académiques gratuits ont vu le jour à l'initiative de projets de recherche ou de sociétés de

développement informatique. Ces nouveaux outils mettent en avant leur proximité avec l'intelligence artificielle, le web sémantique et le traitement automatique du langage naturel.

Selon l'information recherchée (types de données, publications d'un auteur ou d'une institution dans une thématique, recherche sur un sujet précis, indicateurs de citations, etc.) et l'utilisation souhaitée (découverte d'un domaine, étude bibliométrique), l'utilisateur a intérêt à tester différents moteurs et à confronter leurs résultats (nombre de réponses, pertinence des références affichées) pour évaluer l'intérêt d'un outil par rapport à l'autre.

Des critères peuvent aider à évaluer les avantages et inconvénients :

- le périmètre couvert : sources indexées, types d'informations mises à disposition, couverture temporelle, langues des ressources indexées. Consulter les rubriques *FAQ*, *About*, *Learn more*, *Sources*, *Support*, *Blog* qui peuvent donner des indications utiles ;
- la transparence de la stratégie éditoriale, l'affichage des partenariats et des contributions, le statut juridique et la place occupée sur le marché des moteurs de recherche, étayés par des données factuelles et chiffrées. Explorer les liens d'information sur la page d'accueil, lire les mentions légales et les crédits du site ;
- la dynamique et la longévité de l'outil : année de création, engagements en termes d'offre de services, partenaires et sponsors, évolutions fonctionnelles ou techniques ;
- les listes et vocabulaires utilisés : auteurs, institutions, classification thématique, listes de revues ;
- les fonctionnalités de recherche : options avancées (filtres, limitation de résultats par langue et par date, historique, alerte, sauvegarde et reproduction d'une requête) ;
- les options d'affichage, de tri, d'export des résultats de recherche, la mention des sources et les liens vers les données sources ;
- les données de citations et autres métriques (altmétriques) avec leur méthode de calcul (voir fiches CoopIST de la rubrique *Evaluer*) ;
- la possibilité de créer son compte utilisateur et son profil ;
- la possibilité de suggérer des sources à indexer, l'interactivité avec les questionnaires du moteur.

3. Google Scholar (2004), le plus ancien, le plus connu

Lancé par l'entreprise américaine Google en 2004, Google Scholar (GS - <https://scholar.google.fr/>) est un moteur de recherche spécialisé, dédié à la littérature scientifique.

Nombre d'enregistrements - Evalué à 389 millions en janvier 2018 (pas de chiffres officiels).

Qu'indexe-t-il ? Google Scholar indexe les références bibliographiques et le texte intégral du contenu librement accessible de sites d'éditeurs scientifiques commerciaux (Elsevier, Springer, Wiley, etc.), de certains producteurs commerciaux de bases de données ou de portails commerciaux de revues selon les accords conclus (EBSCO, Proquest, Cairn.info, etc.), de sites de revues scientifiques, d'archives ouvertes (par exemple HAL, Agritrop), d'entrepôts de preprints (comme ArXiv, BioRxiv), etc.

Mode de recherche - Par recherche simple et recherche avancée (accessible par le menu). La recherche avancée permet de limiter à un ou plusieurs mots et expressions (*Articles contenant : tous les mots suivants, cette expression exacte, au moins un des mots suivants, aucun des mots suivants, mes mots*), à un ou plusieurs auteurs (*Rechercher des articles rédigés par*), à des revues (*Rechercher des articles publiés dans*), à des années de publication (*Rechercher des articles datés de*). Les résultats peuvent être filtrés par langue ; les brevets et les citations peuvent être exclus (cases à cocher). Les résultats peuvent être triés par date de publication ou par pertinence.

Affichage - Google Scholar n'affiche que les 1 000 premiers résultats d'une recherche, qu'il sélectionne lui-même. Pour chaque référence affichée, sont mentionnés la source avec le lien au document, le nombre de citations du document avec le lien vers les documents citants, les autres articles proches

par leur sujet des références affichées, le nombre de versions du document avec le lien vers ces versions, et le lien vers le *profil utilisateur* lorsque l'auteur recherché a un compte GS.

Avantages / inconvénients - Son ancienneté et sa simplicité d'utilisation expliquent son succès. Mais ses fonctionnalités sont trop limitées pour une recherche experte. Google Scholar est une source d'analyse bibliométrique intégrée au logiciel gratuit [Harzing's Publish or Perish](#) (PoP).

4. Dimensions (2018), moteur se présentant comme une base de données

Dimensions (<https://www.dimensions.ai/>) a une version allégée gratuite accessible à tous et une version complète payante.

Lancé en janvier 2018, Dimensions est un moteur de recherche qui se présente comme une base de données bibliographique multidisciplinaire. Il est produit par [Digital Science](#) (Londres, appartenant à la société allemande [Holtzbrinck Publishing Group](#) détentrice d'une part majoritaire de [Nature Publishing Group](#)).

Nombre d'enregistrements - Au 9 février 2023, Dimensions affichait sur son site un volume total de 134 millions de publications, 151 millions de brevets, 934 000 policy documents, 759 000 essais cliniques... Dimensions souhaite concurrencer les bases de données multidisciplinaires Web of Science Core Collection de Clarivate (171 millions d'enregistrements) et Scopus d'Elsevier (plus de 87 millions de références).

Qu'indexe-t-il ? Les données traitées sont issues de plateformes d'éditeurs, de bases librement accessibles ([Crossref](#), [PubMed](#), [OpenCitations](#), bases des offices de brevets), d'archives ouvertes (comme [HAL](#), [Agritrop](#)), d'entrepôts de preprints (comme [ArXiv](#), [BioRxiv](#)), d'entrepôts de données (comme [Figshare](#), [Dryad](#), [Zenodo](#)), etc. Consulter le site [Support](#) de Dimensions.

Mode de recherche - La recherche permet de saisir tout ou partie d'une référence (*Full data, Title and Abstract, DOI, Abstract Search, Keyword Search*). Des filtres limitent la recherche courante à des auteurs (*RESEARCHER*), des champs thématiques (*FIELDS OF RESEARCH*), des types de publications (*PUBLICATION TYPE*), des sources d'information (*SOURCE TITLE*), des listes de revues (*JOURNAL LIST*), et des conditions d'accès (*OPEN ACCESS*).

Affichage - Chaque référence contient le lien vers sa source et la notice originelle. Egalement, des données de citations propres à Dimensions sont fournies (*Citations, Recent Citations, Field Citation Ratio, Relative Citation Ratio*) ainsi que des données de citations sur internet fournies par la société [Altmetric.com](#) (voir fiche CoopIST [Se familiariser avec les altmetrics](#)).

Avantages / inconvénients - La version gratuite ne couvre que certains types de contenus (publications, datasets). Malgré cela, l'outil reste intéressant par ses fonctionnalités avancées proches de celles des bases de données scientifiques classiques. Dimensions est une source d'analyse bibliométrique intégrée au logiciel gratuit [VOSViewer](#) de l'Université de Leiden (Pays-Bas).

5. Semantic Scholar (2015), moteur basé sur l'intelligence artificielle

Lancé en 2015, Semantic Scholar (<https://www.semanticscholar.org/>) est le moteur de recherche universitaire d'[Allen Institute for Art Intelligence](#) (AI2) créé en 2013 par Paul Allen, cofondateur de Microsoft (USA). Outre son partenaire clé Microsoft, AI2 s'est associé à des éditeurs universitaires et à l'organisme américain à but non lucratif [Impactstory](#) pour son application [Unpaywall](#) qui identifie les textes librement accessibles sur internet.

Nombre d'enregistrements - Au 9 février 2023, Semantic Scholar affichait plus de 210 millions d'enregistrements dans tous les domaines scientifiques.

Qu'indexe-t-il ? Semantic Scholar indexe des documents scientifiques en libre accès (en bas de page du site, le lien [Sources](#) liste les sources indexées) et fournit des liens vers des reportages, articles de blog et vidéos suscités par les documents.

Mode de recherche - C'est une recherche simple, à partir de tout ou partie d'un mot ou d'une expression. Seuls les opérateurs AND et OR sont utilisables. On peut affiner par l'un des 15 champs thématiques proposés.

Affichage - Les résultats s'affichent sous forme de listes de références bibliographiques courtes. Chaque référence, qui peut être visualisée sous sa forme longue dans une nouvelle page, est accompagnée d'indicateurs de citations (*Highly Influential Citations, Citations Velocity*). L'internaute peut la sauvegarder s'il a créé son compte Semantic Scholar. La référence peut être mise en forme selon des styles bibliographiques prédéfinis (*BibTex, EndNote...*) et donne accès au texte intégral. L'aide en ligne est une foire aux questions ([FAQ](#)).

Avantages / inconvénients - Ses fonctions de recherche étant limitées, Semantic Scholar ne rivalise pas avec une base de données bibliographique, même si l'affichage des résultats s'en rapproche. Il est une source d'analyse bibliométrique intégrée au logiciel gratuit [VOSViewer](#) de l'Université de Leiden (Pays-Bas).

6. Lens (1998), moteur de recherche sophistiqué et expert

Lancé en 1998, Lens (<http://lens.org/>) est un projet de [Cambia](#), institut australien indépendant à but non lucratif. Initialement centré sur les brevets, le périmètre et les fonctionnalités de Lens se sont enrichies. Lens se définit aujourd'hui comme une base de données gratuite pour les particuliers et les organisations à but non lucratif donnant accès aux brevets et aux travaux académiques mondiaux. Lens propose aussi une offre payante pour un usage commercial et pour les entreprises commerciales.

Nombre d'enregistrements - Au 9 février 2023, Lens affichait plus de 252 millions de publications et près de 146 millions de données de brevets.

Qu'indexe-t-il ? Lens compile et harmonise les données issues des bases Microsoft Academic (arrêtée en 2021), [PubMed](#) et [Crossref](#). Ces données sont enrichies d'informations sur le libre accès (sources [OpenAlex](#) et [UnPaywall](#)), de liens vers le texte intégral (base de données [CORE](#) moissonnant elle-même de nombreuses archives ouvertes dont celle du Cirad, Agritrop) et des identifiants chercheurs [ORCID](#).

Mode de recherche - C'est une recherche simple sur les brevets (onglet *Patents*), les publications (onglet *Scholarly Works*), ou par le nom d'auteur de publications ou d'inventeur de brevets (onglet *Profiles*). Des filtres tels que la date, le type de publications, l'identifiant ORCID de l'auteur sont proposés selon l'objet recherché : brevets, publications, auteurs.

Pour chaque onglet *Patents* et *Scholarly Works*, le lien *Structured Search* donne accès à une recherche avancée spécifique. Les termes sont à saisir dans la barre en haut de la page (*Refine your Search*). En haut à droite du formulaire, le bouton *Search* permet de passer d'un objet de recherche à l'autre. Les fonctions de recherche avancée ont de nombreux champs d'interrogation en central (*All Fields* par défaut) et de nombreux filtres à gauche (*Date Range, Author, Institution, Funding, Journal, Open Access...*). A droite de la liste des références, des graphes visualisent certains résultats comme l'évolution annuelle du nombre de publications, les profils des auteurs. De nombreux indicateurs sont proposés au-dessus de la liste (*Works Cited by Patents, Works Cited by Scholarly, Scholarly Citations...*)

Affichage - Les résultats d'une recherche s'affichent en central sous forme d'une liste de références. La liste et les références peuvent être personnalisées, étendues, partagées, exportées, triées, via les menus *Customize List, Expand, Share, Export, Sort by Relevance...*

L'aide en ligne (<https://support.lens.org/>) est accessible par le lien *Support* en haut à droite de l'écran. Une courte démo en ligne (*Feature tour*) est proposée.

Avantages / inconvénients - Les fonctions très avancées de Lens, aussi riches que celles de bases de données bibliographiques comme Web of Science ou Scopus, en font un moteur de recherche académique sophistiqué, requérant un apprentissage et s'adressant plutôt à un public de spécialistes.

7. Scinapse (2019), méta-moteur académique

Lancé en 2019, Scinapse (<https://scinapse.io/>) est un moteur de recherche académique développé par Pluto, un réseau de développeurs et de chercheurs de Corée du Sud.

Nombre d'enregistrements - En février 2023, Scinapse (dans la rubrique [FAQ](#)) affichait toujours les chiffres datant d'avril 2019, soit plus de 200 millions d'articles de plus de 40 000 revues.

Qu'indexe-t-il ? Scinapse utilise les données liées (publications, projets, citations, financements) qu'il collecte et agrège de quatre sources : [Microsoft Academic Graph](#) (source principale de Scinapse mais arrêtée en 2021), [Semantic Scholar \[Open Research Corpus\]](#), [PubMed](#), [Springer Nature \[SciGraph\]](#).

Mode de recherche - La recherche est simple, par titre, auteur, DOI ou mot-clé. Deux onglets permettent de mener la recherche sur l'ensemble des données (*All*) ou sur les auteurs (*Authors*) identifiés dans les sources interrogées par Scinapse.

Affichage - La page de résultats permet de filtrer les résultats par date (*Any time* par défaut), type de document (*Any Journal or conference* par défaut), discipline (*Any topic* par défaut) et de les trier (*Pertinence* par défaut). Chaque référence bibliographique peut être visualisée sous sa forme complète avec le lien vers la source, les références des auteurs, les articles de la revue, etc. Une référence peut être sauvegardée (*Save*), affichée selon un format bibliographique prédéfini (*Cite*).

Des graphes sont proposés (onglet *Analysis* en fin de notice) relatifs aux citations reçues par la publication référencée, ainsi qu'aux citations reçues par la revue et ses autres articles.

Une foire aux questions ([FAQ](#)) explique les principes et les objectifs de Scinapse.

Avantages / inconvénients - Les résultats agrégés et emboîtés, par juxtaposition des résultats extraits d'autres moteurs, ne donnent pas de véritable valeur ajoutée à Scinapse.

Liens utiles

Bode C., Herzog C., Hook D., McGrath R. 2019. A guide to the Dimensions data approach. Digital Science, Dimensions. <https://www.digital-science.com/resources/portfolio-reports/a-guide-to-the-dimensions-data-approach/>

Gusenbauer M. 2019. Google Scholar to overshadow them all? Comparing the sizes of 12 academic search engines and bibliographic databases. *Scientometrics*, 118, 177–214. <https://doi.org/10.1007/s11192-018-2958-5>

Semantic Scholar, le moteur de recherche qui explore toute la science. 2019. Newstrotteur. <https://newstrotteur.fr/2019/10/23/semantic-scholar-le-moteur-de-recherche-qui-explore-toute-la-science-newstrotteur/>

Tay A., Martín-Martín A., Hug S.E. 2021. Goodbye, Microsoft Academic – Hello, open research infrastructure? LSE blog. <https://blogs.lse.ac.uk/impactofsocialsciences/2021/05/27/goodbye-microsoft-academic-hello-open-research-infrastructure/>

Marie-Claude Deboin

Délégation à l'information scientifique et technique, Cirad

10 février 2020, mise à jour 10 février 2023

Comment citer ce document :

Deboin M. C. 2023. Utiliser des moteurs de recherche académiques gratuits, en 7 points. Montpellier (FRA) : CIRAD, 5 p. <https://doi.org/10.18167/coopist/0068>

Cette œuvre est mise à disposition selon les termes de la Licence Creative Commons : Attribution - Pas d'Utilisation Commerciale - Partage dans les Mêmes Conditions 4.0 International, disponible en ligne : <http://creativecommons.org/licenses/by-nc-sa/4.0/deed.fr>

ou par courrier postal à : Creative Commons, 171 Second Street, Suite 300, San Francisco, California 94105, USA.

Cette licence vous permet de remixer, arranger, et adapter cette œuvre à des fins non commerciales tant que vous créditez l'auteur en citant son nom et que les nouvelles œuvres sont diffusées selon les mêmes conditions.