
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. Microsoft Academic (2016), moteur sémantique
5. Dimensions (2018), moteur se présentant comme une base de données
6. Semantic Scholar (2015), moteur basé sur l'intelligence artificielle
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 en ligne d'information scientifique passait par la consultation de bases de données bibliographiques multidisciplinaires pour la plupart payantes (comme Web of Science ou Scopus), acquises sur abonnement par les établissements de recherche et d'enseignement supérieur (voir fiches CoopIST [Choisir les sources d'information scientifique sur internet](#) et [Interroger les sources d'information sur internet](#)).

Avec la généralisation d'internet, la gratuité des moteurs de recherche, leur accès facile et leur utilisation simple ont favorisé leur usage par tous. L'engouement des scientifiques pour Google Scholar dès 2004 et sa large adoption ont créé une concurrence avec les bases de données bibliographiques. Celles-ci ont réagi en améliorant leur ergonomie et en rendant leur utilisation plus aisée.

En 2013, l'intérêt public porté à la science ouverte, c'est-à-dire le libre accès aux résultats de recherche (publications, jeux de données) et aux métadonnées les décrivant, a accru la compétition entre les bases de données et les moteurs de recherche.

Aujourd'hui, les utilisateurs combinent les deux types d'outils, selon leurs besoins, leurs connaissances et leur maîtrise de ces outils :

- les **bases de données bibliographiques**. Elles ont une couverture circonscrite et 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. Leurs fonctionnalités élaborées permettent de répondre à des requêtes complexes d'utilisateurs.
- Les **moteurs de recherche « académiques »**, c'est-à-dire spécialisés dans les ressources documentaires scientifiques. Ils parcourent, fouillent et indexent de manière automatique les contenus des sites web. Ils créent leurs propres index, qui sont uniquement basés sur les contenus accessibles des ressources web qu'ils trouvent et collectent. Leur valeur ajoutée est essentiellement liée à la sophistication des traitements automatisés qu'ils développent. Néanmoins, leur périmètre est imprécis et leur stratégie manque souvent de transparence. De ce fait, ils doivent plutôt être considérés comme des outils de découverte ou exploratoires, complémentaires aux bases de données scientifiques.

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.

Des critères peuvent aider l'utilisateur à évaluer les avantages et inconvénients d'un moteur de recherche académique :

- 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* qui peuvent donner des indications utiles sur les ressources indexées ;
- la transparence de la stratégie éditoriale, l'affichage des partenariats et des contributions, le positionnement sur le marché, étayés par des données factuelles et chiffrées. Explorer les différents liens d'information sur la page d'accueil, lire les mentions légales et les crédits du site ;
- 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.

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.

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 - Estimé à 160 ou 200 millions (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, etc.), de sites de revues scientifiques, d'archives ouvertes (par exemple HAL, Agritrop), d'entrepôts de preprints (comme ArXiv, BioRxiv), etc.

Mode de recherche - Google Scholar propose un formulaire de recherche simple et un formulaire de recherche avancée (accessible par le menu). La recherche avancée permet de limiter une recherche 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*) ou à des titres de revues (*Rechercher des articles publiés dans*) ou 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 1000 premiers résultats d'une recherche, qu'il sélectionne lui-même. Pour chaque référence affichée, les informations suivantes sont mentionnées : 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 limitées sont insatisfaisantes pour une recherche experte ou pour l'internaute qui a l'habitude des bases de données scientifiques. Google Scholar est une source d'analyse bibliométrique intégrée au logiciel gratuit [Harzing's Publish or Perish](#) (PoP).

4. Microsoft Academic (2016), moteur sémantique

Relancé en 2016, Microsoft Academic (MA - <https://academic.microsoft.com/>) est un moteur de recherche sémantique développé par la société américaine Microsoft (USA).

Nombre d'enregistrements - Au 29 janvier 2020, Microsoft Academic affichait plus de 231 millions d'enregistrements de publications, dont 36 % d'articles de revues (pour 48 800 revues), 22 % de brevets, 1 % de chapitres d'ouvrages et moins de 1 % d'ouvrages.

Qu'indexe-t-il ? Microsoft Academic, qui s'appuie sur le moteur de recherche Bing (<https://www.bing.com/>) de Microsoft, indexe des données issues d'éditeurs scientifiques et de pages indexées sur le web (revues ; archives ouvertes comme HAL, Agritrop ; entrepôts de preprints comme ArXiv, BioRxiv). A la différence de Google Scholar, Microsoft Academic interprète le sens des termes d'une requête en exploitant les différentes entités associées à une publication : sujets de recherche (*topics*), titres de revues (*journals*), titres de conférences, noms d'auteurs, institutions d'affiliation des auteurs. Par exemple, le moteur interprète le mot « Microsoft » comme Microsoft en tant qu'institution, et affiche les références des publications dont les auteurs sont affiliés à Microsoft.

Mode de recherche - Pour affiner les résultats, Microsoft Academic propose plusieurs filtres (auteurs, sujets, revues, conférences, institutions) et options de tri (pertinence, importance, date de publication, citation).

Affichage - Chaque référence listée est accompagnée du nombre de citations reçues et mentionne les sources (*Websites*) où elle est présente, avec le lien d'accès. Une foire aux questions ([FAQ](#)) informe sur les contenus et les fonctionnalités du moteur.

Avantages / inconvénients - Microsoft Academic est intéressant par son interface de recherche et sa navigation intuitive et par la clarté de l'affichage de ses résultats de recherche. Il est une source d'analyse bibliométrique intégrée aux logiciels gratuits [VOSViewer](#) de l'Université de Leiden (Pays-Bas), et [Harzing's Publish or Perish](#) (PoP).

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

Lancée en janvier 2018, Dimensions (<https://www.dimensions.ai/>) est un moteur de recherche qui se présente comme une base de données bibliographique multidisciplinaire. Il est produit par la société commerciale [Digital Science](#) (Londres) qui appartient à l'entreprise de média privée allemande [Holtzbrinck Publishing Group](#) détentrice d'une part majoritaire de [Nature Publishing Group](#).

Nombre d'enregistrements - Dimensions propose une version allégée gratuite accessible à tous et une version complète payante. En janvier 2020, la version complète affichait 154 millions d'enregistrements de publications, citations, brevets, financements, policy documents, essais cliniques. Dimensions ambitionne de concurrencer les bases de données multidisciplinaires Web of Science Core Collection de Clarivate Analytics (77 millions d'enregistrements) et Scopus d'Elsevier (70 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 [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 - Le formulaire de recherche permet de saisir des requêtes sur tout ou partie d'une référence (*Full data, Title and Abstract, DOI, Abstract Search, Keyword Search*). Des filtres permettent de limiter 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 issue d'une recherche contient le lien vers sa source et la notice originelle. En regard d'une liste de références et de chaque référence affichée, Dimensions fournit des données de citations propres à Dimensions (*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 - Le fait que la version gratuite ne couvre qu'une partie des contenus (publications, datasets) est un frein à son utilisation. L'outil reste néanmoins 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).

6. 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 29 janvier 2020, Semantic Scholar affichait plus de 180 millions d'enregistrements dans tous les domaines scientifiques.

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

Mode de recherche - Le formulaire de recherche est simple. Il permet de chercher des publications sur 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 donner accès au texte intégral. L'aide en ligne est sous la forme d'une foire aux questions ([FAQ](#)).

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

7. Scinapse (2019), méta-moteur académique

Lancé en 2019, Scinapse (<https://scinapse.io/>) est le dernier né des moteurs de recherche académiques, développé par [Pluto](#), un réseau de développeurs et de chercheurs de Corée du Sud.

Nombre d'enregistrements - En avril 2019, Scinapse (dans la rubrique [FAQ](#)) déclarait indexer 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 distinctes : [Microsoft Academic Graph](#) (source principale de Scinapse), [Semantic Scholar \[Open Research Corpus\]](#), [PubMed](#), [Springer Nature \[SciGraph\]](#).

Mode de recherche - Le formulaire de recherche est simple. Il permet de saisir les termes d'une requête. 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 - Au sein de la page de résultats qui s'affichent, une nouvelle recherche peut être conduite via un formulaire simple. Chaque référence affichée peut être visualisée sous sa forme complète avec le lien vers la source, vers les références des auteurs, les articles de la revue, etc. Une référence peut être sauvegardée (*Save*), affichée selon un format bibliographique prédéfini (*Cite*). En regard du nom d'auteur, un bouton (*Suggest change*) donne accès à un formulaire sous Google Doc permettant à un auteur de demander des corrections sur sa fiche (*Suggestion for Changes to Author information*). Une Foire aux questions ([FAQ](#)) explique les principes et les objectifs de Scinapse.

Avantages / inconvénients - Les résultats agrégés, par juxtaposition des résultats extraits d'autres moteurs, ne donnent pas de véritable valeur ajoutée à Scinapse, du moins à ce stade de son développement.

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/>

Google Scholar Digest: Research on Google Scholar Empirical evidences [newsletter]. <http://googlescholar Digest.blogspot.com/>

Hug S. E., Brändle M. P., 2017. Microsoft Academic is on the verge of becoming a bibliometric superpower. LSE (London School of Economic and Political Science) Impact Blog. <https://blogs.lse.ac.uk/impactofsocialsciences/2017/06/19/microsoft-academic-is-on-the-verge-of-becoming-a-bibliometric-superpower/>

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/>

Marie-Claude Deboin

Délégation à l'information scientifique et technique, Cirad

10 février 2020

Comment citer ce document :

Deboin M. C. 2020. Utiliser des moteurs de recherche académiques gratuits, en 7 points. Montpellier (FRA) : CIRAD, 5p. <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.