

Google Scholar

Fin 2004 est apparu, de manière confidentielle, un nouvel outil de recherche, Google Scholar¹, appelé parfois « schoogle ». S'il n'a pas provoqué les mêmes foudres² que celles suscitées par le projet de numérisation des fonds des plus prestigieuses universités américaines pour Google print, cet outil peut néanmoins être un sérieux concurrent à l'offre de recherche d'informations bibliographiques proposée par les bibliothèques à travers les bases de données.

Élisabeth Noël

École nationale supérieure
des sciences de l'information
et des bibliothèques
noel@enssib.fr

De quoi s'agit-il ?

Google Scholar est disponible pour l'instant en version bêta, et est donc en train de se mettre en place. Aussi, ses fonctionnalités et sa couverture telles qu'abordées ici ne sauraient être définitives – du moins osons-nous l'espérer. Les rares éléments actuellement disponibles sur sa constitution et son fonctionnement peuvent donc être amenés à évoluer.

Tout d'abord, Google Scholar ne fait pas pour l'instant l'objet d'un affichage aussi clair que pour les options de recherche d'images ou d'actualités. Il faut donc soit connaître l'adresse, soit aller sur le site de Google.com (« Google.com in English » depuis la version française), puis, par le lien « about Google », accéder au laboratoire des idées de Google « Google Labs » (visite à conseiller pour avoir une idée des projets en cours et essayer les prototypes). La présentation de Google Scholar figure à « Search through journal articles, abstracts and

other scholarly literature », ou à « Search scholarly papers », selon l'endroit où se trouve l'utilisateur dans le site.

Seule la consultation de l'aide permet d'obtenir quelques informations plus précises : cet outil référence des livres, des thèses, des pré-publications, des articles de périodiques, des résumés et des rapports techniques, dans des domaines multidisciplinaires, et provenant d'éditeurs scientifiques, de sociétés savantes, de dépôts de preprints, d'universités, et du web en général. Un peu de recherche sur le web permet d'obtenir une information plus précise : Google Scholar est partenaire de 29 éditeurs de Crossref³, et indexe à l'heure actuelle 2 millions des 57 millions de références de Worldcat ; parmi les autres partenariats connus, citons ACM (Association for Computing Machinery), IEEE (Institute of Electrical and Electronics Engineers) et Pubmed.

Les résultats d'une recherche semblent proposer une meilleure description bibliographique de l'information récoltée que Google basic, puisqu'on voit apparaître, dans la liste des résultats, le nom de l'auteur, l'année de publication, la source ou le titre de la revue dont est tiré l'article.

1. <http://scholar.google.com>

2. Voir, bien sûr, la tribune de Jean-Noël Jeanneney, « Quand Google défie l'Europe » (*Le Monde*, 23 décembre 2005), et le débat qui s'est ensuivi. Consulter à ce titre la revue de presse d'Educnet : <http://www.educnet.education.fr/actua/html/arts/G20050207.htm> et Info.Tic n° 33, mars 2005, dans sa partie « document numérique » : <http://www.educnet.education.fr/actu/infotoc/00/doc.htm>

Voir aussi : Jean-Noël Jeanneney, *Quand Google défie l'Europe : plaidoyer pour un sursaut*, Éd. Mille et une nuits, 2005, 120 p.

3. Crossref regroupe éditeurs, bibliothèques et sociétés avec, pour objectif, de lier de façon pérenne les références bibliographiques aux contenus des articles diffusés en ligne par le DOI (Digital Object Identifier).

* Article écrit avec la contribution de Kari Desservettaz, élève conservateur.

Élisabeth Noël a travaillé dans les bibliothèques des universités de Rouen et de Jussieu (Paris). Depuis 1997, elle est enseignante et coresponsable de la cellule Formation à l'information scientifique et technique (Formist) de l'Enssib. Auteur de plusieurs articles dans le BBF, elle a collaboré à Information Literacy in Europe (Rome, Consiglio Nazionale delle Ricerche, 2003). Elle dirige le site web Repere, Ressources électroniques pour les étudiants, la recherche et l'enseignement. <http://repere.enssib.fr>

À cela s'ajoutent des « bonus », indication de la nature du document ou de l'information, selon qu'il s'agit d'une citation, d'un livre... avec les mentions [book], [citation], ou le type de fichier [PS], [PDF], et la localisation du document, à travers une recherche croisée sur Worldcat.

On peut noter que Google Scholar repère les références bibliographiques à l'intérieur des ressources indexées, et les propose donc, en réponse à une question, en tant que citations, sans lien vers le texte cité. Par ailleurs, il utilise cette fonctionnalité pour indiquer le nombre de fois où un article proposé en réponse est signalé à l'intérieur de sa propre base (*cited by*). Ce calcul ne se fait que sur sa propre base de données et n'est donc pas exhaustif.

Certains résultats proposent un lien « *library search* », qui renvoie vers la notice de l'ouvrage dans Worldcat, afin de le localiser dans une bibliothèque participante. Enfin, Google Scholar peut renvoyer aux différentes versions d'un même article, localisées sur des serveurs variés.

En ce qui concerne la recherche, le formulaire est tout aussi simple que celui qui a fait le succès de Google basic. Le formulaire de recherche avancée de l'outil de base a aussi été adapté et permet de faire la recherche par auteur et dans les titres de périodiques, et depuis peu, par date, mais ni par langue, ni par domaine de connaissance ou par discipline.

La syntaxe classique est bien sûr utilisable : opérateurs « + » et « - » pour imposer ou exclure un terme, guillemets pour la recherche d'expressions.

Exemple de résultats de recherche

Une recherche sur « intelligence artificielle » donnera par exemple, parmi les 4 720 réponses, des résultats présentés ainsi :

[BOOK] Artificial Intelligence
PH Winston, BKP Horn – [Cited by 650](#) – [Library Search](#) – [Web Search](#)
Reading, Mass.: Addison-Wesley, 1984

[CITATION] L'intelligence artificielle et le langage
G Sabah – [Cited by 32](#) – [Web Search](#)
Hermès, Paris, 1990

A symbolic and surgical acquisition of terms through variation
C Jacquemin – [View as HTML](#) – [Cited by 17](#) – [Web Search](#)
... This work has also benefited from rich discussions in the research group Terminologie et Intelligence Artificielle of the PRC Intelligence Artificielle. ... Connectionist, Statistical, and Symbolic Approaches to..., 1996 – [arxiv.org](#) – [univ-tlse2.fr](#) – [limsi.fr](#) – [csi.uottawa.ca](#) – [all 9 versions](#) »

[PS] Comparaison de diverses méthodes d'intelligence artificielle pour la résolution de conflit en ...
H Gruber, CAPS Regions – [View as HTML](#) – [Cited by 2](#) – [Web Search](#)
Page 1. Comparaison de diverses méthodes d'intelligence artificielle. pour la résolution de conflits en contrôle de trafic aérien. 1. Hervé Gruber 2. Juin 1992. ...
Rapport de stage, Centre d'Etudes de la Navigation..., 1992 – [recherche.enac.fr](#)

sions. Les équations expertes permettent un peu plus de précisions, avec juste un petit nombre des attributs de recherche que propose Google basic : *Author:nomdelauteur*, *intitle:titre*, *site:url*. Une recherche posée ainsi : *artificial site.stanford.edu*, permet de limiter la recherche sur « artificial » aux sites dont le domaine est *stanford.edu*.

Les résultats de la recherche

Enfin, l'ordre des résultats est conforme à la technique habituelle de Google qui propose d'abord les références les plus populaires, c'est-à-dire les plus citées par d'autres pages.

Alors, faut-il finalement recommander Google Scholar aux étudiants et aux chercheurs ?

Oui, car, de toute façon, cet outil risque de connaître le même engouement que Google Basic, parce que sa simplicité le rend plus attrayant que nos bases de données. Mais attention, nous devons rester vigilants et alerter sur les limites de l'outil. En effet, si Google Scholar est simple d'utilisation, quelle qualité propose-t-il ? Tout d'abord, l'usage de Google Scholar permet de circonscrire la recherche à un certain type et un certain niveau d'informations. Ainsi une recherche sur « l'indexation du web » par Google Scholar permettra d'échapper à tous les sites commerciaux proposant des offres de référencement.

Cependant, un tel outil ne peut rivaliser avec la qualité d'interrogation d'une base de données, dans laquelle la structuration de l'information permettra toujours une recherche plus fine. Ainsi, la recherche par date, même si elle est possible, n'est pas fiable, comme cela est indiqué dans l'aide de la recherche avancée⁴. De la même manière, il n'est pas possible de rechercher des propriétés chimiques spécifiées, ni d'utiliser des champs spécifiques tels que ceux qu'une base de données spécialisée peut offrir.

Dans le forum de discussion sur cet outil, un chercheur du nom de Blanc déplore la difficulté qu'il a à trouver ses propres publications⁵, en raison de la fréquence de ce terme comme substantif et de la difficulté qu'il y a à chercher un auteur parmi de nombreux homonymes sur Google Scholar.

Enfin, les résultats affichés sont obtenus « à la volée », à la suite d'une indexation automatique de pages web, ce qui peut amener, à la suite de mauvaises interprétations par les systèmes, à des résultats parfois cocasses ou aberrants : une recherche sur l'ou-

4. « Bear in mind, however, that some web sources don't include publication dates, and a date-restricted search will not return articles for which Google Scholar was unable to determine a date of publication. »

5. Google Scholar is a failure
<http://groups-beta.google.com/group/Google-Labs-Google-Scholar>

vrage de Bertrand Calenge, « Accueillir, orienter, informer... » propose deux citations, de deux auteurs différents, l'un étant « B Calenge », l'autre « C Bertrand ». Les bibliothécaires y retrouveront le leur, peut-être pas les usagers !

Les résultats, quant à eux, présentent certaines limites : Google Scholar ne permet pas d'aller au-delà des 1 000 premières réponses, même s'il en annonce un nombre plus important. Certes, on espère ne pas devoir aller aussi loin dans la liste des résultats. Il y a deux impossibilités plus gênantes : celle de trier les réponses selon des critères choisis (par année, par titre de périodiques, par type de documents) et surtout, celle d'utiliser un vocabulaire contrôlé, puisque toute la recherche repose sur de l'indexation en texte intégral.

En ce qui concerne les réponses, l'outil n'est pas très performant dans le domaine des sciences humaines, et propose des résultats incomplets, même dans ceux proposés avec des partenaires. Google Scholar a, par exemple, oublié un an de références à Pubmed⁶. Il est aussi possible de comparer les résultats d'une recherche sur Google Scholar et sur l'outil de recherche natif de l'éditeur, et la comparaison est souvent au détriment de Google Scholar⁷. De plus, des acteurs majeurs comme Biological abstracts, Inspec, Georef, ne sont pas partenaires de Google Scholar. De toute façon, les résultats ne sont pas exhaustifs, ce qui renvoie à une des grosses lacunes de l'outil : quelle est la taille de la base de ressources indexées, combien de ressources sont référencées ?

Et surtout, comment Google définit-il qu'une publication relève du « Scholar » : là-dessus, rien ne filtre. Si on peut imaginer que c'est la nature

du site qui le détermine (par exemple, les titres de revues ou les sites universitaires), comment se fait-il que l'on puisse trouver des weblogs⁸ en résultats ? Le concept de « scholarly » est-il alors défini par un algorithme informatique ?

Enfin, Google Scholar prétend fournir l'accès au texte intégral des publications signalées. Effectivement, s'il indexe bien ce texte intégral, *via* les partenariats avec les éditeurs, il ne permet d'y accéder que si le texte est en libre accès, ou, sinon, avec des accès payants. Si la bibliothèque propose un accès gratuit *via* ses abonnements, celui-ci n'est alors pas signalé, et l'utilisateur, mal orienté, va payer ce qu'il pourrait obtenir sans bourse délier par son établissement documentaire⁹.

Google Scholar et les bibliothèques

Google Scholar est en cours de réalisation. On peut donc prédire que certains des défauts actuels disparaîtront rapidement, et que les partenariats avec les éditeurs et les sites universitaires, pour la diffusion et l'accès à l'information, ne pourront que s'améliorer. J'avoue avoir été séduite par sa simplicité d'usage et par la qualité de ses résultats, encore relative il est vrai. Néanmoins, il soulève déjà certaines questions :

- si son usage se généralise en bibliométrie, à l'instar de l'*impact factor* du Science Citation Index, alors l'importance des publications et leur visibilité dépendraient de l'algorithme, inconnu, d'une société privée. Ce qui ne peut qu'être choquant ;
- par ailleurs, cet outil peut favoriser une surveillance par les éditeurs des différentes versions des articles qu'ils

8. Comme par exemple Information Literacy Weblog <http://ciquest.shef.ac.uk/infolit/>

9. À noter cependant que l'utilisateur peut, dans ses « *scholar preferences* », indiquer 3 institutions parmi 28, pour lesquelles un lien sera ensuite visible dans le résultat de recherche, vers les outils locaux de localisation. Il s'agit surtout d'accès *via* des bibliothèques américaines, même si on trouve, par exemple, des universités de Bruxelles, Amsterdam ou Munich.

publient, puisque Google Scholar permet de retrouver toutes les versions d'un même article sur des sites différents. Ce qui nous entraîne bien loin de l'idée de l'*open access*.

Aussi peut-on voir en Google Scholar un nouveau péril pour les bibliothèques, car s'il risque de renforcer la recherche « à la Google », il peut aussi conduire à une hégémonie de Google, qui, par-delà la fourniture des documents *via* la numérisation, deviendrait l'outil incontournable, et commercial, de recherche de l'information purement scientifique. Google Scholar serait alors une des deux branches de l'étau qui nous étoufferait, l'autre étant Google Print !

Je proposerais plutôt de saisir autrement le problème, et de « monter sur les épaules du géant¹⁰ », pour en sortir plus grand.

Google Scholar doit inciter le monde universitaire à standardiser ses protocoles d'accès, pour rendre plus lisible l'information que nous rendons disponible. Les bases de données possèdent une forte valeur ajoutée par rapport à de tels outils, mais les services que les bibliothèques proposent (accès à l'information, formations à la recherche d'informations) doivent être plus visibles, quitte à utiliser le filtre Google pour se présenter aux usagers. Ainsi, cet outil, bien utilisé, peut nous permettre d'élargir l'accès aux ressources que nous proposons en bibliothèque¹¹.

Avril 2005

10. « *Stand on the shoulders of giants* » est la devise de Google Scholar.

11. Pour aller plus loin : des outils de marquage apparaissent déjà, pour signaler, parmi les résultats de Google Scholar, ceux disponibles localement ou *via* un agrégateur : *IngentaConnect Power Tools* (http://www.ingentaconnect.com/about/researchers/power_tools) pour Ingenta, ou bien *Google Scholar OpenURLs - Firefox Extension* (<http://www.ualberta.ca/%7Epbinkley/gso/>) ou *OpenURL Referrer* (<http://www.openurl.com/openurlref>). Voir aussi la barre d'outil « OCLC - Yahoo! Toolbar » (<http://www.oclc.org/toolbar>), qui permet de localiser les ouvrages dans les bibliothèques membres de l'OCLC, avec déjà plus de 4 millions de références versées dans l'index de Yahoo!

6. *Google Scholar is a Full Year Late Indexing PubMed Content* http://www.workingfaster.com/sitelines/archives/2005_02.html#000282

7. « Side by Side. Native Search Engines vs Google Scholar », Peter Jacso <http://www2.hawaii.edu/~jacso/scholarly/side-by-side2.htm>