

Google Scholar service matches Thomson ISI citation index

by Léo Charbonneau

The free Google Scholar service does as good a job as Thomson ISI's science citation index for performing citation counts and could be used as a cheap substitute to the costly Thomson service, says a University of British Columbia professor. Thomson's citation databases are accessible through the company's Web of Science portal only by subscription, which can cost a university tens of thousands of dollars a year.

Daniel Pauly, director of the Fisheries Centre at UBC, and Konstantinos Stergiou, of Aristotle University of Thessaloniki in Greece, compared the two methods using 114 papers from 11 disciplines published between 1925 and 2004. For papers published before 1990, the authors found that the citation counts were proportional. In other words, if Thomson ISI found that a particular paper was cited 10 times as often as another, Google Scholar found the same ratio. However, for these older papers, the actual citation counts with Google were about half that of Thomson.

But, for papers published from 1990 on, not only were the citation counts proportional, the actual number of citations was nearly the same. The result is "very surprising," said Dr. Pauly. "I didn't expect Google to perform so well. . . . I expected some vague proportionality, but I did not expect that it would be roughly one to one."

Thomson ISI, which recently changed its name to Thomson Scientific, counts all the references of articles in several thousand pre-selected journals; Google Scholar searches only scientific sources available on the web. Many of the journals indexed by Thomson ISI are not available electronically, or are available only by subscription, and so would not necessarily be picked up by Google Scholar. Dr. Pauly suspects that Google Scholar compensates for this by finding citations in the so-called "grey literature," such as reports, books, conference proceedings and other items found

on the web. The result is "you can use Google essentially for the same purpose as you use Thomson ISI."

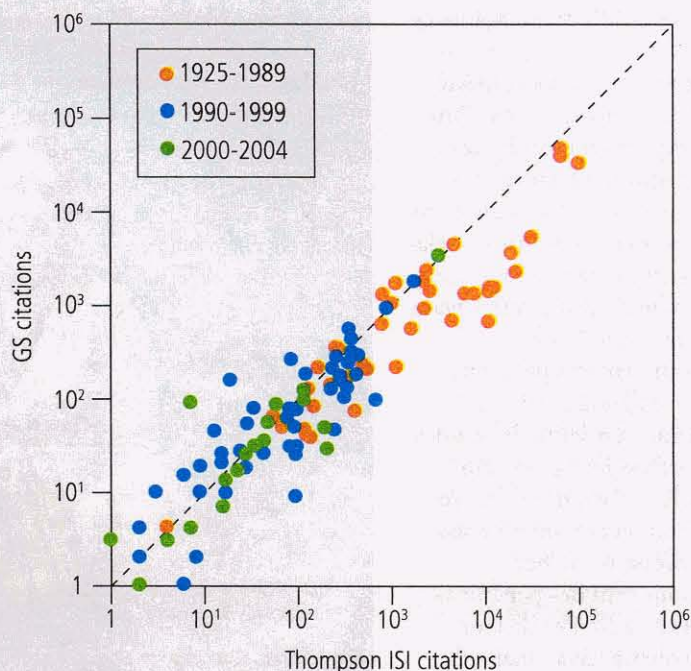
This has many implications for science policy, says Dr. Pauly, due to the price differential between the costly Thomson ISI products and Google Scholar, which he presumes will continue to be free. For example, research and academic institutions in developing countries, and even modestly endowed institutions in developed countries, will be able to assess and document their scientific progress through Google Scholar at minimal cost.

In addition, free access to Google Scholar "offers an avenue for more transparency in tenure reviews, funding and other science policy issues," since it allows citation counts and analyses "to be performed and duplicated by anyone." This might also lead Thomson ISI to reduce its prices, which would

create "a better situation for everybody," said Dr. Pauly.

Not everyone agrees on Google Scholar's utility. In a recent article published in the journal *Current Science*, University of Hawaii computer science professor Peter Jacso says the service "gives a bad name to autonomous citation indexing." Google Scholar, he writes, "shows lack of competence and understanding of basic issues of citation indexing" and fails in implementing even the most basic Boolean commands.

Dr. Pauly says Dr. Jacso's review is "very harsh" and points out that the author is on the advisory board of Thomson ISI. Dr. Pauly also notes that Thomson ISI is 40 years old, while Google Scholar "is one and a half years old, and will improve." The findings of Drs. Pauly and Stergiou are published in a recent issue of the online journal *Ethics in Science and Environmental Politics*.



The relationship between the citations in Thomson ISI's citation index and Google Scholar for 114 selected articles, September 2005.

Google Scholar talonne Thomson ISI

par Léo Charbonneau

Selon un professeur de la University of British Columbia, la base de citations gratuites Google Scholar recenserait pratiquement le même nombre de citations que celle de Thomson ISI, ce qui en fait un substitut intéressant à l'onéreux service de Thomson. Comme la base de citations de l'éditeur de publications érudites peut être consultée à partir du portail Web of Science de l'entreprise sur abonnement seulement, il peut en coûter des dizaines de milliers de dollars annuellement pour une université.

Daniel Pauly, directeur du Centre des pêches à la UBC, et Konstantinos Stergiou, de l'Université Aristotle, située à Thessaloniki, en Grèce, ont comparé les deux services à partir de 114 articles issus de 11 disciplines et publiés entre 1925 et 2004. Ils ont ainsi découvert que le nombre de citations était proportionnel dans les deux bases pour les articles antérieurs à 1990. En d'autres mots, si Thomson ISI indiquait qu'un article était cité dix fois plus souvent qu'un autre, Google Scholar donnait le même ratio, mais le nombre réel de citations calculées par Google était environ deux fois moindre.

Par contre, pour les articles ultérieurs à 1990, non seulement le nombre de citations était-il proportionnel, mais leur nombre réel était pratiquement le

même. Les résultats ont grandement surpris M. Pauly, qui ne s'attendait pas à un si bon rendement de la part de Google : « J'envisageais une certaine proportionnalité, mais certainement pas une quasi-équivalence. »

Thomson ISI, rebaptisé Thomson Scientific depuis peu, dénombre toutes les citations d'articles dans plusieurs milliers de revues scientifiques présélectionnées, tandis que Google Scholar cherche uniquement dans les sources scientifiques accessibles sur Internet. Bon nombre de revues indexées par Thomson ISI ne sont pas offertes en format électronique et ne sont donc pas nécessairement recensées par Google Scholar. M. Pauly est d'avis que le moteur de recherche compense ce désavantage en fouillant dans ce qu'on appelle la « littérature grise », soit les rapports, les livres, les comptes rendus de conférences et d'autres documents publiés sur le Web. Il conclut donc « que l'outil de Google peut essentiellement être utilisé aux mêmes fins que celui de Thomson ISI ».

Selon M. Pauly, ce constat pourrait avoir d'importantes conséquences pour la politique scientifique en raison de la différence de prix entre le coûteux service de Thomson et Google Scholar, accessible gratuitement. Par exemple,

les établissements d'enseignement et de recherche des pays en développement ainsi que les établissements moins bien nantis des pays développés pourront, grâce à Google Scholar, évaluer et documenter leurs progrès sur le plan scientifique à un coût minime. En outre, l'accès gratuit au moteur de recherche assure une meilleure transparence au chapitre du financement et de l'évaluation des permanences, car il permet à tous de faire le dénombrement et l'analyse de citations.

Cependant, tous ne sont pas convaincus de l'utilité de Google Scholar. Dans un article publié récemment dans la revue *Current Science*, Peter Jasco, professeur d'informatique à l'Université d'Hawaï, affirme que le service trahit un manque de compétence et de compréhension des principes de base de l'indexage de citations et n'offre même pas les commandes booléennes les plus élémentaires.

M. Pauly répond que l'auteur fait partie du conseil consultatif de Thomson ISI. Il souligne du même coup que Thomson ISI existe depuis 40 ans, tandis que Google Scholar a vu le jour il y a à peine un an et demi « et s'améliorera au fil du temps ». Les résultats de l'étude de M. Pauly et Stergiou ont été publiés récemment dans la revue électronique *Ethics in Science and Environmental Politics*.