

Google Scholar, Web of Science, and Scopus: Which is best for me?



*Being able to find, assess and place new research within a field of knowledge, is integral to any research project. For social scientists this process is [increasingly likely](#) to take place on Google Scholar, closely followed by traditional scholarly databases. In this post, **Alberto Martín-Martín, Enrique Orduna-Malea, Mike Thelwall, Emilio Delgado-***

López-Cózar, analyse the relative coverage of the three main research databases, Google Scholar, Web of Science and Scopus, finding significant divergences in the social sciences and humanities and suggest that researchers face a trade-off when using different databases: between more comprehensive, but disorderly systems and orderly, but limited systems.

Researchers routinely use databases such as Google Scholar, Web of Science, and Scopus to search scholarly information and consult bibliometric indicators such as citation counts. However, although an understanding of the basic characteristics of these services is needed for effective literature searches and for deciding whether their indicators are appropriate for use in research evaluations, the differences between these databases in terms of coverage and reliability of the data are still not widely known.

A crucial aspect in which these services differ is in their approach to document inclusion. Web of Science and Scopus rely on a set of source selection criteria, applied by expert editors, to decide which journals, conference proceedings, and books the database should index. Conversely, Google Scholar follows an inclusive and automated approach, indexing any (apparently) scholarly document that its robot crawlers are able to find on the academic web.

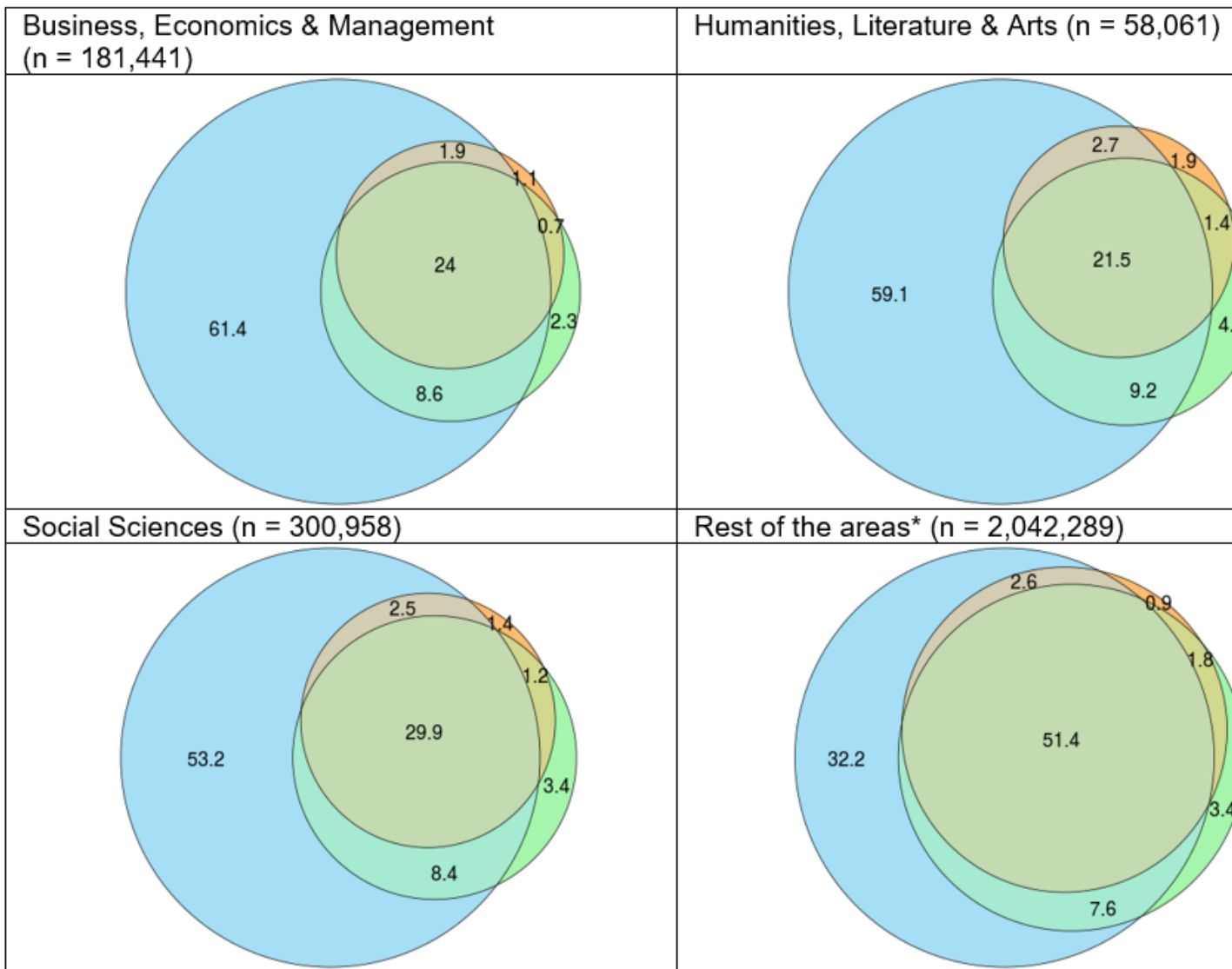
Each approach has its pros and cons. The selective approach of Web of Science and Scopus produces a curated collection of documents, but is sensitive to biases in the selection criteria. Indeed, evidence has shown that these databases have limited coverage in the areas of Social Sciences and Humanities, literature written in languages other than English, and scholarly documents other than journal articles. For its part, Google Scholar's inclusive and unsupervised approach maximises coverage, [giving each article "the chance to rise on its own merit"](#). Nevertheless, it leads to the presence of technical errors in the platform, such as duplicate entries that refer to the same document, incorrect or incomplete bibliographic information, and the inclusion of non-scholarly materials.

We have recently tested the differences in coverage in these three data sources across subject categories. For a sample of over 2,500 very highly-cited documents across 252 subject categories that [Google Scholar released in 2017](#), we checked whether the documents were also covered by Web of Science and Scopus. This comparison favours Google Scholar, since it is the original source of the documents, but is nevertheless a reasonable test since it seems that any scholarly database ought to have quite comprehensive coverage of highly cited documents. The [results](#) showed that, even within this highly-selective set of documents (all published in English), a significant amount in the Social Sciences and Humanities were not covered by the selective databases. In most cases, the cause was that the database did not cover the journal at the time the article was published.

We later decided to [dig deeper](#) into this issue, and for all the highly-cited documents in the sample, we collected the complete list of citations that each of the three databases provided, and identified the overlapping and unique citations. This new sample, which amounted to just below 2.5 million citations, gave us a more detailed picture of the relative differences in coverage across the three databases, not only at the level of broad areas, but also for each of the 252 subject categories.

The results by broad areas showed that Google Scholar was able to find most of the citations to Social Sciences articles (94%), while Web of Science and Scopus found 35% and 43%, respectively. Moreover, Google Scholar appeared to be a superset of Web of Science and Scopus, as it was able to find 93% of the citations found by Web of Science, and 89% of the citations found by Scopus. Last but not least, over 50% of all the citations to Social Science articles were only found by Google Scholar. The same analysis was applied to the 252 specific subject categories, and can be viewed in this [interactive web application](#).

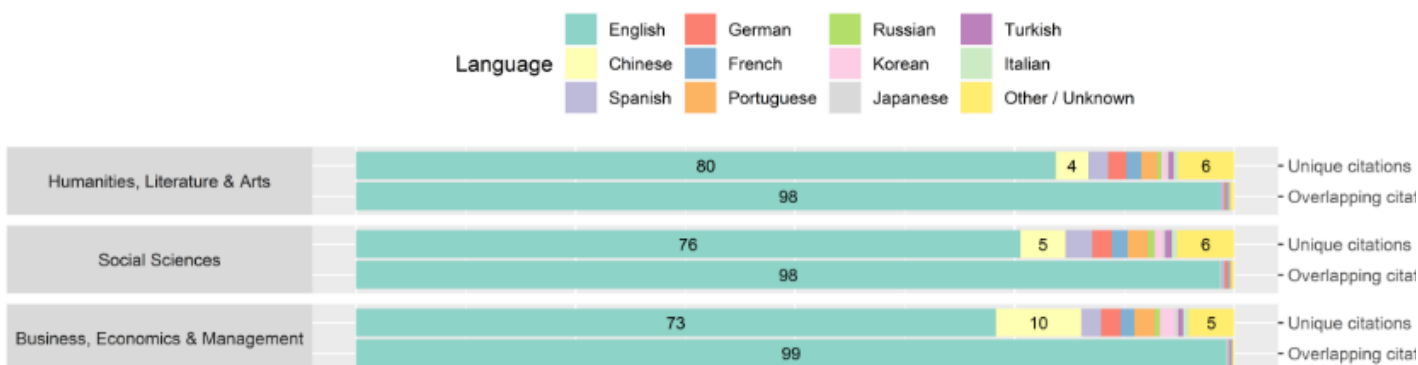
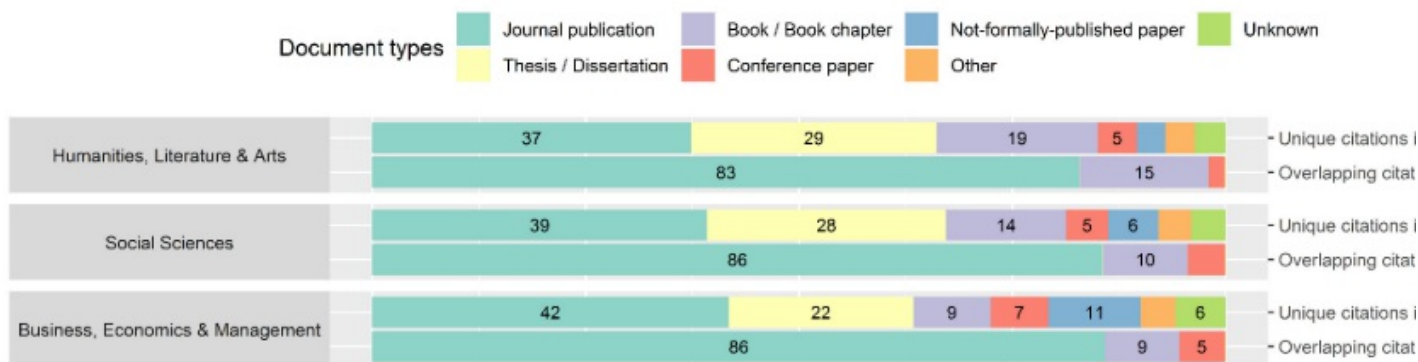
● Google Scholar
 ● Web of Science
 ● Scopus



* Chemical & Material Sc
 Engineering & Computer Sc
 Health & Medical Sc
 Life Sciences & Earth Sc
 Physics & Mathe

The large proportion of citations that are only found by Google Scholar, especially in the Social Sciences, the Humanities, and Business, Economics & Management, raises the question of which types of sources Google Scholar covers that the other databases do not. To provide an answer, we identified the document types and the languages of the citations in our sample, and compared the proportions of document types and languages of citations only found by Google Scholar on one side (unique citations in Google Scholar), and citations found by two or more databases on the other (overlapping citations). The results were aggregated at the level of broad areas.

The majority (~60%) of the citations found only by Google Scholar come from non-journal sources: among these we find theses and dissertations, books and book chapters, not-formally-published papers such as preprints and working papers (especially important in Business and Economics), and conference papers. Nevertheless, there is still a large proportion of citations to Social Sciences and Humanities articles from journals that are not indexed in Web of Science or Scopus. There is also a significant minority of citations to Social Sciences and Humanities articles that only Google Scholar can find, that come from documents published in languages other than English, which are not covered in the selective databases.



Interestingly, despite the significant differences in coverage, and despite the known errors that may be present in the data from Google Scholar, which we did not attempt to eliminate (e.g. inflated citation counts caused by duplicate entries), Spearman correlations between citation counts are very strong across all areas and databases (in most cases over .90, although sometimes lower in some fields of the Humanities). Thus, if Google Scholar citation counts were used for research evaluations then its data would be unlikely to produce large changes in the results. It would be particularly useful when there is a reason to believe that documents not covered by Web of Science or Scopus are important for an evaluation.

In conclusion, the inclusive paradigm of document indexing popularised by Google Scholar facilitates discovery of not only the most well-known sources, but also of sectors of scholarly communication that were previously hidden from view. This can be useful in literature searches, as well as for those who need to compile evidences of research impact for a collection of outputs, but at the same time it has created some problems of its own. The question, as our colleague Professor Harzing put it, is whether we are ready to [accept a trade-off](#): going beyond the comfortable and orderly borders of curated databases in exchange for more diverse coverage. Our hope is that these results can help researchers and other stakeholders make informed decisions in this regard.

This post draws on the authors' co-authored article, [Google Scholar, Web of Science, and Scopus: a systematic comparison of citations in 252 subject categories](#) available on SocArXiv.

About the authors

Alberto Martín-Martín is a lecturer in the department of communication and information science at the Universidad de Granada, Spain.

Enrique Orduna-Malea an assistant professor at the Universitat Politècnica de València, Spain.

Mike Thelwall is a professor of information science at the University of Wolverhampton, UK.

Emilio Delgado-López-Cózar is Professor of Research Methods at the Universidad de Granada, Spain.

Note: This article gives the views of the authors, and not the position of the LSE Impact Blog, nor of the London School of Economics. Please review our [comments policy](#) if you have any concerns on posting a comment below.

Featured Image Credit adapted from, [Fernand de Canne](#) via Unsplash (Licensed under a [CC0 1.0](#) licence)