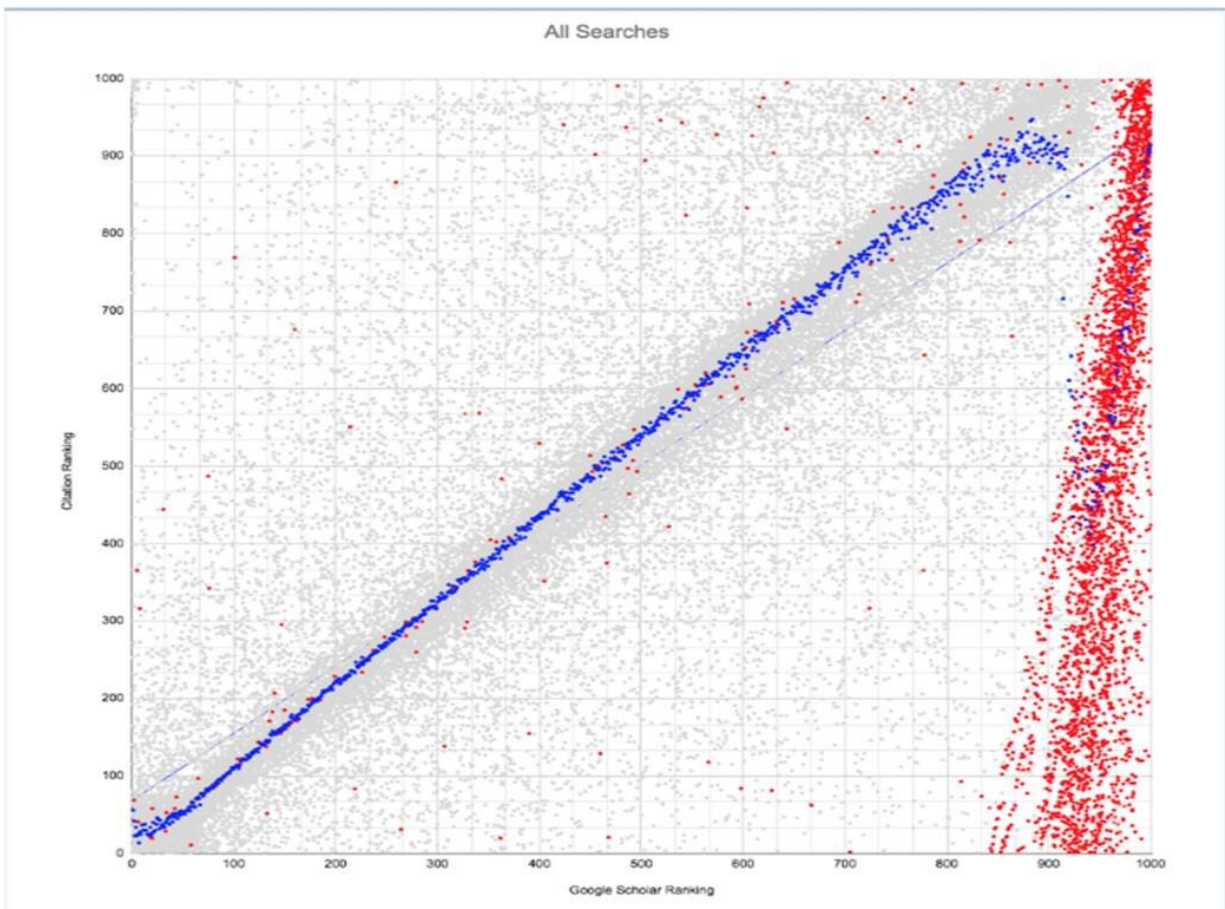# Google Scholar renders documents not in English invisible

February 10 2021

Study Results: The graph shows how articles written in languages other than English appear above position 900 in the Google Scholar ranking. Credit: Universitat Pompeu Fabra - Barcelona

The visibility of scientific articles and conference papers is conditional upon being easily found in academic search engines, especially Google Scholar. To enhance this visibility, search engine optimization (SEO) has been applied in recent years to academic search engines in order to optimize documents and, thereby, ensure they are better ranked in search pages (i.e., academic search engine optimization or ASEO).

Recent research, published in *Future Internet*, has found out whether the language of the document is a factor involved in the sorting algorithm of search results on Google Scholar. The study authors are Cristòfol Rovira, Lluís Codina and Carlos Lopezosa, members of the Department of Communication at UPF.

"To implement this optimization we need to further our understanding of Google Scholar's relevance ranking algorithm, so that, based on this knowledge, we can highlight or improve those characteristics that academic documents already present and which are taken into account by the algorithm," says Rovira, first author of the study. To prevent fraudulent practices, Google Scholar does not explain this algorithm and, therefore, this kind of research becomes necessary.

For the study, the authors applied an inverse engineering research methodology based on statistical analysis using Spearman's correlation coefficient. Three different types of search were conducted yielding a sample of 45 searches each with 1,000 results (45,000 documents): by author, by year, and by keyword.

Quality articles with hundreds of citations are treated in a discriminatory manner

The results show that when a search is performed on Google Scholar with results in various languages, the vast majority (90%) of documents

in languages other than English are systematically relegated to positions that render them totally invisible. These documents are almost always placed in positions above rank position 900, even though they are quality articles with hundreds of citations. Thus, it can be stated that Google Scholar discriminates against documents not written in English in searches with multilingual results.

A lack of awareness of this factor could be detrimental to researchers from all over the non-English-speaking world, making them believe that there is no literature in their national language when they conduct searches with multilingual results.

"This is particularly the case in the most frequent searches, that is, those conducted by year. Nevertheless, it can also occur in searches using certain keywords that are the same in languages around the world, including trademarks, chemical compounds, industrial products, acronyms, drugs, and diseases, with COVID-19 being the most recent example," the study authors reveal.

And they add "moreover, if we consider the results of this study from the perspective of ASEO, it is more than evident that until this bias is addressed, the chances of being ranked in a multilingual Google Scholar search increase remarkably if the researchers opt for publication in English."

Graph of the results of the study

The scatter plot above summarizes the research results. There are 45,000 dots, one per document. The gray dots represent documents written in English, other languages in red, and blue shows the median positions.

The graph shows how articles written in languages other than English appear above 900th position in the Google Scholar ranking. This is so

even for quality documents that have hundreds of citations and are well placed in the ranking for number of citations.

The most striking cases are the red dots located in the bottom-right corner. They correspond to documents written in languages other than English that are ranked by number of citations below 100 and have a Google Scholar ranking over 900. This means that all of them receive over a thousand citations and appear in Google Scholar in the same positions as documents in English cited just a few dozen times.

More information: Cristòfol Rovira et al, Language Bias in the Google Scholar Ranking Algorithm, *Future Internet* (2021). DOI: 10.3390/fi13020031

Provided by Universitat Pompeu Fabra - Barcelona