

Article

# Language Bias in the Google Scholar Ranking Algorithm

Cristòfol Rovira <sup>1,2</sup> , Lluís Codina <sup>1,2</sup>  and Carlos Lopezosa <sup>1,\*</sup> 

<sup>1</sup> Department of Communication, Universitat Pompeu Fabra, 08002 Barcelona, Spain; cristofol.rovira@upf.edu

<sup>2</sup> UPF Barcelona School of Management, Balmes, 134, 08008 Barcelona, Spain; lluis.codina@upf.edu

\* Correspondence: carlos.lopezosa@upf.edu; Tel.: +34-66-162-4297

**Abstract:** The visibility of academic articles or conference papers depends on their being easily found in academic search engines, above all in 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). To achieve this degree of optimization, we first 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. This study seeks to advance our knowledge in this line of research by determining whether the language in which a document is published is a positioning factor in the Google Scholar relevance ranking algorithm. Here, we employ a reverse engineering research methodology based on a statistical analysis that uses Spearman's correlation coefficient. The results obtained point to a bias in multilingual searches conducted in Google Scholar with documents published in languages other than in English being systematically relegated to positions that make them virtually invisible. This finding has important repercussions, both for conducting searches and for optimizing positioning in Google Scholar, being especially critical for articles on subjects that are expressed in the same way in English and other languages, the case, for example, of trademarks, chemical compounds, industrial products, acronyms, drugs, diseases, etc.

**Keywords:** ASEO; SEO; reverse engineering; citations; google scholar; algorithms; relevance ranking; citation databases; academic search engines; multilingual search



**Citation:** Rovira, C.; Codina, L.; Lopezosa, C. Language Bias in the Google Scholar Ranking Algorithm. *Future Internet* **2021**, *13*, 31. <https://doi.org/10.3390/fi13020031>

Academic Editor: Fatos Xhafa  
Received: 30 December 2020  
Accepted: 23 January 2021  
Published: 27 January 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

A researcher's professional career is heavily dependent on the visibility of, and the recognition afforded by, their scholarly output. The number of citations received and the corresponding indexes associated with this variable, in particular the h-index, are typically the most widely used metrics employed in official processes of accreditation before appropriate academic boards or commissions. Indeed, the need to be cited, combined with the exponential increase in world bibliographic output, means that researchers today have to promote their own articles as the final step in the complex process of publishing their research findings. This promotion of their research output usually also implies building their personal academic brand [1,2], including the creation of complementary content that extends well beyond traditional scholarly articles for specialized publications or papers for conferences.

Among the actions researchers can take to promote both their personal brand and scientific output are mentioning their articles in academic and non-academic social networks, creating a professional blog with complementary content (videos, presentations, PDFs), creating profiles on a range of platforms from ORCID, Google Scholar Citation, and ResearcherID to Mendeley, depositing documents in open access repositories and optimizing articles so that they command a good position in search engines, especially Google Scholar. Indeed, the vast majority of Internet users do not look beyond the second

or third page of results [3], which means for a document to be found easily it has to be optimized to ensure it appears towards the top of the first page.

Search results are ranked according to relevance, a value automatically calculated by search engines. Moreover, this ranking is usually established as the default order, because other forms of sorting—for example, by title or by date—are considered less significant for most search intentions (although these other forms of sorting are often available). Relevance is calculated using an algorithm that takes into account a range of factors, which means that for each search engine, relevance will differ to some degree, given that it is being defined by a distinct algorithm in each case.

Search engine optimization (SEO) [4,5] is, today, a well-established discipline, the goal of which is to highlight the quality of web pages and so improve their position in the results pages. This goal should not be achieved by fraudulent means, but depends rather on knowing how the algorithms that determine relevance operate, identifying the factors taken into account and, finally, optimizing these factors in one's documents. However, Google Scholar is not always able to detect attempts at manipulation [6,7].

Today, there is a huge community of SEO experts and companies that dedicate their efforts to analyzing and discussing Google's relevance ranking algorithm. Via blogs [8–11], online publications [12–14], and books [15–17] they advise designers and webmasters as to how they can optimize their websites so that they are easily indexed and can occupy the highest rankings in the results pages.

Google's relevance algorithm is based on more than 200 factors [18] and include the number of links received, the keywords and related terms in the title and other significant areas of the document, the download speed of the server on which the page is hosted, the length of the text, the user experience, the mobile-first design, the semantic tagging, the age of the domain, etc. Google has never released complete information about all these factors or the exact weighting attached to each; the company only provides general, incomplete information in order to avoid spam. Indeed, if all the details of how the algorithm works were known, then poor quality documents could be placed at the top of the results page.

This “black box” policy has led SEO professionals to conduct reverse engineering research in an effort to identify the specific factors involved in relevance ranking. Thus, they analyze the search results in order to infer how the algorithm works. However, it is a complicated process in which many factors intervene and it is not easy to draw any conclusive results.

In recent years, this ecosystem of research concerned with algorithms and the subsequent publication of recommendations has been extended to Google Scholar and academic articles. On a much smaller scale, reverse engineering research has been applied to Google Scholar [19–25] and blogs [26–30], university library guidelines [31–34] and the authors' services of the publishers of academic journals [35–40] offer their recommendations as to how to optimize articles so that they appear at the top of the rankings of Google Scholar's results pages. This SEO applied to academic search engines has been called academic search engine optimization or ASEO [20,22,41–43].

This research community is still in its infancy both in terms of the quantity and quality of its output; moreover, the recommendations given for Google Scholar are often contaminated by research findings for Google. In fact, they are two quite distinct algorithms that operate on two quite distinct types of document in two very different environments. Indeed, as far as the ranking algorithm is concerned, academic documents have at least four major characteristics that clearly distinguish them from web pages: most are in PDF (and not HTML) format; they contain links based on bibliographic citations with other academic documents (not hyperlinks); once published, they are not modified; and, usually, author metadata and the date of publication are clearly identified.

Promoting a personal academic brand and the visibility of web pages, blogs, videos and other complementary content depend to a large extent on Google positioning. But the visibility of academic articles or conference papers is determined by their optimization for Google Scholar. These differences need to be clarified, while it is necessary to further

our understanding of Google Scholar's relevance ranking algorithm, which is not so well known and widely analyzed as Google's general search algorithm.

The aim of this study is to do just that and, more specifically, here, because of its far-reaching implications, we seek to determine whether the language in which a document is written is a key positioning factor. In this regard, no previous study, to the best of our knowledge, has attempted to find a relationship between positioning and language, be it for Google Scholar or for the general Google search engine.

Normally, language plays no role as a ranking factor in keyword searches, given that the language of the search word itself determines the language of the documents retrieved. If the documents are written in the same language, this factor is overridden. Language only intervenes in those few cases of keywords with the same spelling in different languages that generate multilingual lists of results. In contrast, searches by author or year are conducted independently of language and always provide what we shall refer to henceforth as *multilingual* results (or searches).

When searches are multilingual, that is, when results are provided in different languages for the same search, the language of the documents can be a decisive factor if it can be shown that this conditions ranking. Thus, our primary research question here is the following: In multilingual searches, is the language in which a document is written a factor in Google Scholar's ranking algorithm?

Our hypothesis is that Google Scholar favors the English language in multilingual search results. As a result, documents in other languages have fewer possibilities of being placed at the top of the rankings for the sole reason that they are not published in English.

In the following section, we discuss related studies in the literature, before moving on to present the applied research methodology and the method used to select our sample. Next, we analyze the results obtained from our statistical data and from observation of our scatter plots. The limitations of the study are discussed and new lines of research are proposed. Finally, in the conclusions, the repercussions of our findings are highlighted, both for searches and for the optimization of positioning in Google Scholar.

## 2. Prior Studies

Google Scholar has been the subject of numerous studies, aimed, above all, at assessing its quality [44–53] and search effectiveness [45,54], determining whether it is a suitable tool for conducting bibliometric studies [55–61] and evaluating author impact using the h-index [25,56,62–67].

In contrast, few studies have specifically examined how the Google Scholar ranking algorithm works [19–22]. One of the reasons for this is probably the erroneous belief that Google Scholar uses the same or a very similar algorithm to the one used by the general Google search engine. However, the few studies of ranking by relevance in Google Scholar that have been conducted [19–25] conclude that it is a specific algorithm, adapted to the particular characteristics of scholarly documents; nevertheless, some ranking factors may be the same in both cases, such as the weight attached to keywords in the title [19]. Other factors, though, are clearly specific, such as the number of citations received, which is unique to Google Scholar [21,25].

Moreover, the frequency of keywords in the text of the article does not appear to be a factor that the Google Scholar algorithm takes into account [19]. Yet, there is evidence that the date of publication is of relevance, given that according to Beel and Gipp [19] older articles have a higher ranking than more recent publications.

A further factor identified by various authors is that the search term has to coincide exactly with a term in the documents retrieved and, moreover, that Google Scholar does not expand its searches through synonyms of keywords in the same way that Google does [68,69].

Where there is greater consensus is in attributing considerable weight to the number of citations received in Google Scholar's calculation of ranking by relevance [19,23–25]. This

is hardly surprising, given that Google Scholar is a bibliographic database and traditionally tools of this type use citations received as a means for sorting their results page.

### 3. Methodology

As indicated, the aim of this study is to further our understanding of the Google Scholar relevance ranking algorithm. We are particularly interested in determining whether the language in which a document is written affects its position on the results page, in other words, if language is one of the factors that intervenes in this algorithm when results exist in several languages.

We have employed reverse engineering techniques as our research methodology. This method is frequently used to determine how a device works by undertaking an analysis of its behavior and results. The method is also used to extract the source code from the compiled files.

Reverse engineering has been used for years to study search engine ranking algorithms [70–72]. The documents that occupy the highest rankings in the results lists are analyzed and, then, by means of a statistical analysis it can be deduced whether the factors they present intervene in the ranking. The most frequently used statistical test in this context is Spearman's correlation, given that the data never form normal distributions. The native search engine ranking is compared with the rankings created by the researcher using some of the frequent characteristics presented by the highest-ranking documents. The higher the correlation coefficient, the more similar the two rankings are and, therefore, the greater the weight that can be attributed to the alternative ranking factor in the search engine algorithm under analysis.

In the case of Google's algorithm, more than 200 factors are considered [73,74]. This algorithm is highly complex and, moreover, it is subject to constant improvements, a process in which the application of artificial intelligence is involved [75]. It is for this reason that in studies of Google correlation coefficients rarely exceed 0.3. These results are very low, but sufficient to obtain some indication as to whether a particular factor under analysis is included in the computation of the ranking of documents.

Google Scholar adheres to the same (dis)information policy as Google and does not provide specific details as to how the algorithm works, officially, in an effort to prevent spam. This means it publishes only very general guidelines:

"Google Scholar aims to rank documents the way researchers do, weighing the full text of each document, where it was published, who it was written by as well as how often and how recently it has been cited in other scholarly literature." [76]

However, thanks to the findings of previous studies [57,77], it is evident that the Google Scholar algorithm is simpler than that of the general Google search engine. This has certain advantages for reverse engineering studies of Google Scholar, with correlation coefficients higher than 0.8 often being obtained, which allows more reliable conclusions to be drawn.

The sample data for this study consisted of a total of 45 searches each producing 1000 results, which means we analyzed approximately 45,000 items of information, a similar number to other studies of the same type in which reverse engineering was employed [70,71,78] to study the Google Scholar algorithm [19–23,28].

More specifically, we conducted three different types of search, obtaining 15,000 results in each case: searches by author, by year, and by keyword. By so doing, our aim was to analyze the algorithm in different contexts and to identify similar patterns of behavior that would lend more robustness to our results.

In the case of searches by keyword, we selected the most frequent terms meeting the following conditions:

1. General, non-specialized, language terms, in order to avoid any thematic biases.
2. Terms spelt the same in English and Spanish (i.e., cognates), in order to generate multilingual searches in at least these two languages. Subsequently, we also detected

results in Portuguese and other languages, as the spelling of some of the terms selected also coincided with the spellings in these languages.

Additionally, by using the search form, we forced the appearance of the keyword in the title of the documents found so that, in this way, we could neutralize this factor and prevent it from distorting our results.

In the case of searches by year, we included documents published before 2015, in order to avoid more recent documents that might not have had sufficient time to generate a significant number of citations. As we report below, citations are a central element in this methodology.

Finally, in the case of searches by author, to generate as many multilingual results as possible, we selected the most frequently used surnames of Hispanic origin in the United States. Note, surnames were used in isolation from any first names.

To avoid data bias due to low data volumes, the following conditions were also applied to all searches in the sample:

1. At least 2.5% of the documents in all the searches conducted had to be written in languages other than English.
2. All documents included in the analysis had to have at least 15 citations received. The reason for our introducing this condition was to avoid a high number of documents with zero citations received which would have made it impossible to differentiate their ranking by citations.

Our final sample was made up as follows:

- Keywords: capital, combustible, crisis, cultural, federal, festival, final, general, idea, invisible, moral, popular, social, terror, total.
- Authors: Cruz, Diaz, Flores, Garcia, Gomez, Gutierrez, Hernandez, Martinez, Morales, Ortiz, Perez, Ramirez, Ramos, Sanchez, Torres.
- Years: from 2000 to 2014.

Table 1 shows the percentages of documents written in languages other than English retrieved in each search. The data for this study were collected between 28 August 2020 and 22 September 2020 using the Publish or Perish tool [79,80]. Searches by keyword and author were performed without accents and in lower case. The language of each item was identified by employing Google's language recognition technology by primarily analyzing the title of the documents. This technology is available among the spreadsheet functions of Google Drive.

**Table 1.** Sample of searches carried out and the percentage of documents retrieved written in languages other than English.

Percentage of Documents in Languages other than English					
Keyword		Author		Year	
Search	% Non-English	Search	% Non-English	Search	% Non-English
capital	4.0	cruz	7.1	2000	7.5
combustible	7.1	diaz	6.2	2001	7.2
crisis	2.7	flores	5.8	2002	7.4
cultural	4.9	garcia	15.7	2003	8.6
federal	3.7	gomez	8.0	2004	8.7
festival	3.8	gutierrez	5.2	2005	8.6
final	2.5	hernandez	7.8	2006	8.2
general	4.6	martinez	13.5	2007	9.9
idea	5.9	morales	5.2	2008	7.9
invisible	3.7	ortiz	6.9	2009	8.1
moral	4.2	perez	6.1	2010	8.0
popular	8.6	ramirez	5.3	2011	8.8
social	3.0	ramos	10.9	2012	7.2
terror	4.2	ranches	5.6	2013	10.7
total	2.5	torres	6.3	2014	9.6%
Average	4.4%		7.7%		8.4%



For the statistical study, as the distributions were not normal according to the Kolmogorov-Smirnov test, we used Spearman's correlation.

For each of the searches analyzed, the number of citations received was transformed to an ordinal scale. In this way, we constructed an alternative ranking that could subsequently be correlated with the native Google Scholar ranking. This is a very common procedure [19,21] in this type of research as it allows the comparison to be more naturally made between two ordinal variables.

As mentioned, the number of citations a document receives is a very important factor in the Google Scholar ranking algorithm, especially in searches by date and author, when keywords do not intervene. In searches of this type, the ranking by number of citations received is very similar to the native Google Scholar ranking, with correlation coefficients higher than 0.9 [23].

To determine the role that language plays in Google Scholar's ranking, we exploited this similarity. Our aim in so doing was to determine whether documents written in English follow a similar pattern to documents written in other languages when the native ranking is compared with the ranking by citations received. We wished to see if the number of citations received is an equally determining factor regardless of the language in which the document is written.

If the correlations of the documents written in English are similar to those of documents written in other languages, it means that language does not condition the ranking algorithm; on the other hand, if the correlations are different we will have found indications that language may play a role in this ranking, be it positive or negative.

In order to obtain global values of the correlations, we calculated the median values of the ranking by citations of the 45 searches making up the sample and for each of the 1000 document positions in the results list. These one thousand medians (for each of the 45 searches) provide a measure of central tendency of the citation ranking which can then be correlated with the native Google Scholar ranking.

To conduct the statistical analysis, we used R, version 3.4.0 [71]; SPSS, version 20.0.0; Excel, and Google Sheets. Confidence intervals were constructed using the normal approximation method and applying Fisher's transformation using the R psych package [81–83]. Fisher's transformation when applied to Spearman's correlation coefficient is asymptotically normal. The scatter plots were created with Google Sheets and Tableau.

#### 4. Results

In line with previous research [23,24], an analysis of our data verified a high correlation between the ranking by the number of citations received and Google Scholar's native ranking. The high value of these coefficients indicates that a document receiving a high ranking on the basis of the number of citations received is also highly positioned in Google Scholar's native ranking.

However, when we focused our analysis solely on documents written in languages other than English, we detected an unusual ranking pattern, one that had not previously been identified. This pattern can be defined in terms of the following two characteristics:

1. In multilingual searches in Google Scholar, documents written in languages other than English fail, in the main, to appear among the first 900 results on the search page (i.e., they present a ranking >900).
2. The Google Scholar ranking algorithm works differently for documents written in languages other than English, as it does not consider the number of citations received as the main ranking factor.

This implies that while documents written in English are awarded a good position based on the number of citations received, this same factor, even when the number of citations is high, does not benefit documents written in other languages. In other words, in multilingual searches in Google Scholar, the widely recognized positioning factor based on the number of citations received is not applicable to documents that are not published in English.

Below, we present in detail the data tables and scatter plots that corroborate these conclusions.

#### 4.1. Data Analysis Identifying the Different Ranking Pattern Presented by Documents Written in Languages other than English

Tables 2–4 show the percentages of documents in languages other than in English which failed to appear among the top 900 search results. It is evident that they constitute the vast majority: 92.1% in the case of searches by year, 85.3% in those by author and 89.7% in those by keyword.

**Table 2.** Number of documents ranked above 900 in the case of searches by year.

Year	Documents Ranked above 900 in Searches by Year								
	All Results			Results for English			Results for Non-English		
	#Items	>900	%	#Items	>900	%	#Items	>900	%
2000	995	95	9.5	920	23	2.5	75	72	96.0
2001	993	93	9.4	922	26	2.8	71	67	94.4
2002	995	95	9.5	921	26	2.8	74	69	93.2
2003	993	93	9.4	908	13	1.4	85	80	94.1
2004	989	89	9.0	903	7	0.8	86	82	95.3
2005	993	93	9.4	908	16	1.8	85	77	90.6
2006	994	94	9.5	913	17	1.9	81	77	95.1
2007	985	85	8.6	888	3	0.3	97	82	84.5
2008	996	96	9.6	917	19	2.1	79	77	97.5
2009	989	89	9.0	909	12	1.3	80	77	96.3
2010	994	94	9.5	915	20	2.2	79	74	93.7
2011	987	87	8.8	900	5	0.6	87	82	94.3
2012	991	91	9.2	920	26	2.8	71	65	91.5
2013	982	82	8.4	877	1	0.1	105	81	77.1
2014	988	88	8.9	893	4	0.4	95	84	88.4
Total	14,864	1364	9.2	13,614	218	1.6	1250	1146	92.1

**Table 3.** Number of documents ranked above 900 in the case of searches by author.

Author	Documents Ranked above 900 in Searches by Author								
	All Results			Results for English			Results for Non-English		
	#Items	>900	%	#Items	>900	%	#Items	>900	%
cruz	1000	100	10.0	929	32	3.4	71	68	95.8
diaz	1000	100	10.0	938	42	4.5	62	58	93.5
Flores	1000	100	10.0	942	47	5.0	58	53	91.4
garcia	1000	100	10.0	843	5	0.6	157	95	60.5
gomez	1000	100	10.0	920	27	2.9	80	73	91.3
gutierrez	1000	100	10.0	948	50	5.3	52	50	96.2
hernandez	999	100	10.0	921	28	3.0	78	72	92.3
martinez	994	95	9.6	860	6	0.7	134	89	66.4
morales	997	98	9.8	945	52	5.5	52	46	88.5
ortiz	1000	100	10.0	931	34	3.7	69	66	95.7
perez	998	99	9.9	937	40	4.3	61	59	96.7
ramirez	1000	100	10.0	947	50	5.3	53	50	94.3
ramos	999	100	10.0	890	4	0.4	109	96	88.1
sanchez	999	100	10.0	943	50	5.3	56	50	89.3
torres	1000	100	10.0	937	40	4.3	63	60	95.2
Total	14,986	1492	10.0	13,831	507	3.7	1155	985	85.3

**Table 4.** Number of documents ranked above 900 in the case of searches by keyword.

Keyword	Documents Ranked above 900 in Searches by Keyword								
	All Results			Results for English			Results for Non-English		
	#Items	>900	%	#Items	>900	%	#Items	>900	%
capital	1000	100	10.0	960	63	6.6	40	37	92.5
combustible	990	91	9.2	924	35	3.8	66	56	84.8
crisis	1000	100	10.0	973	75	7.7	27	25	92.6
cultural	999	100	10.0	950	56	5.9	49	44	89.8
exterior	1000	100	10.0	908	17	1.9	92	83	90.2
federal	1000	100	10.0	963	65	6.7	37	35	94.6
festival	1000	100	10.0	962	76	7.9	38	24	63.2
final	1000	100	10.0	975	78	8.0	25	22	88.0
general	999	100	10.0	953	59	6.2	46	41	89.1
idea	999	100	10.0	940	43	4.6	59	57	96.6
invisible	1000	100	10.0	963	65	6.7	37	35	94.6
moral	1000	100	10.0	958	61	6.4	42	39	92.9
popular	1000	100	10.0	914	21	2.3	86	79	91.9
social	1000	100	10.0	970	71	7.3	30	29	96.7
terror	1000	100	10.0	958	64	6.7	42	36	85.7
Total	14,987	1491	9.9	14,271	849	5.9	716	642	89.7

Table 5 provides a summary of these results, indicating that, overall, 91% of documents in languages other than English are ranked above 900 (i.e., >900). This contrasts with just 3.7% of documents in English that appear in these positions. It is also striking that only 0.7% of documents in languages other than English appear in the top 100 positions. This figure falls to 0.2% when the first 20 positions are analyzed (Table 5).

**Table 5.** Number of documents by band of ranking in the results list.

Band of Ranking	Documents by Band of Ranking							
	All Results		Results for English		Results for Non-English			
	#Items	%Items	#Items	%Items	#Items	%Items	#Expected Items *	%Expected Items *
from 1 to 100	4500	10.0	4480	10.7	20	0.7	381	9.0
from 101 to 200	4500	10.0	4477	10.7	23	0.8	351	11.2
from 201 to 300	4500	10.0	4482	10.8	18	0.6	286	9.2
from 301 to 400	4500	10.0	4485	10.8	15	0.6	338	10.9
from 401 to 500	4500	10.0	4481	10.8	19	0.7	323	10.4
from 501 to 600	4500	10.0	4477	10.7	23	0.8	376	12.0
from 601 to 700	4500	10.0	4480	10.7	20	0.6	337	10.8
from 701 to 800	4500	10.0	4474	10.7	26	0.9	284	9.1
from 801 to 900	4500	10.0	4331	10.4	169	3.4	245	7.9
from 901 to 1000	4,337	9.7	1549	3.6	2,788	91.0	200	6.4
Total: from 1 to 1000	44,837	100.0	41,716	100.0	3121	100.0	3,121	100.0
from 1 to 20	900	2.0	893	2.1	7	0.2	100	3.2

\* Expected outcomes if the algorithm was applied homogeneously. Estimates based on the rank positions by number of citations of the documents in English.

Tables 6–8 show the correlation coefficients between the ranking by the number of citations received and the Google Scholar native ranking. The results in English are differentiated from those of the other languages in order to highlight the different ranking pattern identified. In all cases, the correlations are higher for the documents published in English. These differences are especially notable in searches by years and by authors with values always higher than 0.82 in the case of documents in English and values always lower than 0.44 in documents published in other languages.



**Table 6.** Correlations between ranking by citations and the GS native ranking for searches by year.

Correlations for Searches by Year									
Year	All Results			Results for English			Results for Non-English		
	#Items	R	p-Value	#Items	R	p-Value	#Items	R	p-Value
2000	995	0.8882	0.00000	920	0.9951	0.00000	75	0.3107	0.00666
2001	993	0.8858	0.00000	922	0.9773	0.00000	71	0.3454	0.00318
2002	995	0.8653	0.00000	921	0.9856	0.00000	74	0.2660	0.02200
2003	993	0.8501	0.00000	908	0.9792	0.00000	85	0.2704	0.01231
2004	989	0.8259	0.00000	903	0.9844	0.00000	86	0.3170	0.00294
2005	993	0.8522	0.00000	908	0.9864	0.00000	85	0.2098	0.05394 *
2006	994	0.8391	0.00000	913	0.9907	0.00000	81	0.3894	0.00033
2007	985	0.8158	0.00000	888	0.9687	0.00000	97	0.3751	0.00015
2008	996	0.8529	0.00000	917	0.9867	0.00000	79	0.3370	0.00239
2009	989	0.8528	0.00000	909	0.9885	0.00000	80	0.3810	0.00049
2010	994	0.8557	0.00000	915	0.9928	0.00000	79	0.3291	0.00306
2011	987	0.8592	0.00000	900	0.9926	0.00000	87	0.2599	0.01506
2012	991	0.8640	0.00000	920	0.9897	0.00000	71	0.2347	0.04885
2013	982	0.8658	0.00000	877	0.9790	0.00000	105	0.4430	0.00000
2014	988	0.8431	0.00000	893	0.9817	0.00000	95	0.3374	0.00083
Total	14,864			13,614			1,250		
Median	996	0.8767	0.00000	962	0.9322	0.00000	182	0.3498	0.00000

\* Not statistically significant.

**Table 7.** Correlations between ranking by citations and the GS native ranking for searches by author.

Correlations for Searches by Author									
Author	All Results			Results for English			Results for Non-English		
	#Items	R	<i>p</i> -Value	#Items	R	<i>p</i> -Value	#Items	R	<i>p</i> -Value
cruz	1000	0.6729	0.00000	929	0.8262	0.00000	71	0.3679	0.00159
diaz	1000	0.7821	0.00000	938	0.8954	0.00000	62	0.1047	0.41809 *
Flores	1000	0.7796	0.00000	942	0.9160	0.00000	58	0.1390	0.29822 *
garcia	1000	0.5487	0.00000	843	0.8325	0.00000	157	0.4265	0.00000
gomez	1000	0.7042	0.00000	920	0.8545	0.00000	80	0.0452	0.69062 *
gutierrez	1000	0.7928	0.00000	948	0.8978	0.00000	52	0.2654	0.05726 *
hernandez	999	0.7216	0.00000	921	0.8813	0.00000	78	0.3070	0.00625
martinez	994	0.6207	0.00000	860	0.8885	0.00000	134	0.4415	0.00000
morales	997	0.8578	0.00000	945	0.9286	0.00000	52	0.2754	0.04813
ortiz	1000	0.7781	0.00000	931	0.8873	0.00000	69	0.0645	0.59849 *
perez	998	0.7537	0.00000	937	0.8862	0.00000	61	0.1784	0.16898 *
ramirez	1000	0.8082	0.00000	947	0.8973	0.00000	53	0.2030	0.14494 *
ramos	999	0.6948	0.00000	890	0.9140	0.00000	109	0.1556	0.10608 *
sanchez	999	0.7556	0.00000	943	0.8682	0.00000	56	0.0680	0.61837 *
torres	1000	0.7624	0.00000	937	0.82	0.00000	63	0.2043	0.10821*
Total	14,986			13,831			1155		
Median	1000	0.8932	0.00000	979	0.9509	0.00000	211	0.1096	0.11232 *

\* Not statistically significant.

**Table 8.** Correlations between ranking by citations and the GS native ranking for searches by keyword.

Keyword	Correlations for Searches by Keywords								
	All Results			Results for English			Results for Non-English		
	#Items	R	p-Value	#Items	R	p-Value	#Items	R	p-Value
capital	1000	0.4418	0.00000	960	0.4939	0.00000	40	0.0908	0.57740 *
combustible	990	0.7532	0.00000	924	0.8568	0.00000	66	0.3047	0.01287
crisis	1000	0.4950	0.00000	973	0.5199	0.00000	27	−0.1713	0.39291 *
cultural	999	0.3792	0.00000	950	0.4178	0.00000	49	−0.0201	0.89107 *
exterior	1000	0.6286	0.00000	908	0.7293	0.00000	92	−0.0705	0.50417 *
federal	1000	0.5872	0.00000	963	0.5736	0.00000	37	−0.0445	0.79361 *
festival	1000	0.8446	0.00000	962	0.8661	0.00000	38	0.3535	0.02948
final	1000	0.5770	0.00000	975	0.5874	0.00000	25	0.0514	0.80732 *
general	999	0.4012	0.00000	953	0.4161	0.00000	46	0.1698	0.25934 *
idea	999	0.7966	0.00000	940	0.8829	0.00000	59	0.2875	0.02725
invisible	1000	0.6922	0.00000	963	0.7492	0.00000	37	0.0934	0.58233 *
moral	1000	0.5502	0.00000	958	0.5868	0.00000	42	−0.0086	0.95691 *
popular	1000	0.6121	0.00000	914	0.6763	0.00000	86	−0.0022	0.98379 *
social	1000	0.2335	0.00000	970	0.2561	0.00000	30	0.6076	0.00037
terror	1000	0.7719	0.00000	958	0.7872	0.00000	42	0.0189	0.90535 *
Total	14,987			14,271			716		
Median	1000	0.9180	0.00000	996	0.9412	0.00000	151	0.0549	0.50301 *

\* Not statistically significant.

In the case of searches by keyword, some outcomes point to more moderate differences, but if we observe the correlations between the median values, the differences are again very high with values of 0.91 for documents in English and 0.05 for those in other languages, although this last value is not statistically significant (Table 8). In fact, most of the correlations of the results for documents in other languages are not statistically significant, another indication that we are dealing with two different distributions.

If we take the overall data (Table 9) for all the searches and compare the correlations of the medians of the two rankings, we obtain values of 0.97 for the documents in English and of just 0.18 for those in other languages.

**Table 9.** Correlations between median rank positions by citations and the GS native ranking for all searches conducted.

	Correlations for All Searches								
	All Results			Results for English			Results for Non-English		
	#Items	R	p-Value	#Items	R	p-Value	#Items	R	p-Value
All searches	1000	0.9171	0.0000	998	0.9797	0.0000	321	0.1803	0.0012

The value of the correlation for all the data is, likewise, very high at 0.91. As discussed previously, this indicates that while the number of citations is a very important factor in the ranking algorithm, the low values recorded in the case of documents written in languages other than English shows that this factor is not being applied and, as a result, the behavior of the ranking algorithm is unusual in these cases.

#### 4.2. Analysis of Scatter Plots to Show the Differ Ranking Pattern of Documents Written in Languages other than English

The different ranking pattern presented by documents written in languages other than English is also evident in the scatter plots (Figures 1–4), where the documents published in English are represented by gray dots and those corresponding to other languages are represented by red dots. The blue dots represent the median values of the ranking by citations of the documents listed among the top thousand in the results page. As is evident, the vast majority of the results corresponding to documents written in languages other than English (red dots) do not appear in the top 900 positions. There are, in fact, two-

point clouds: One essentially for documents published in English and one for documents published in other languages.

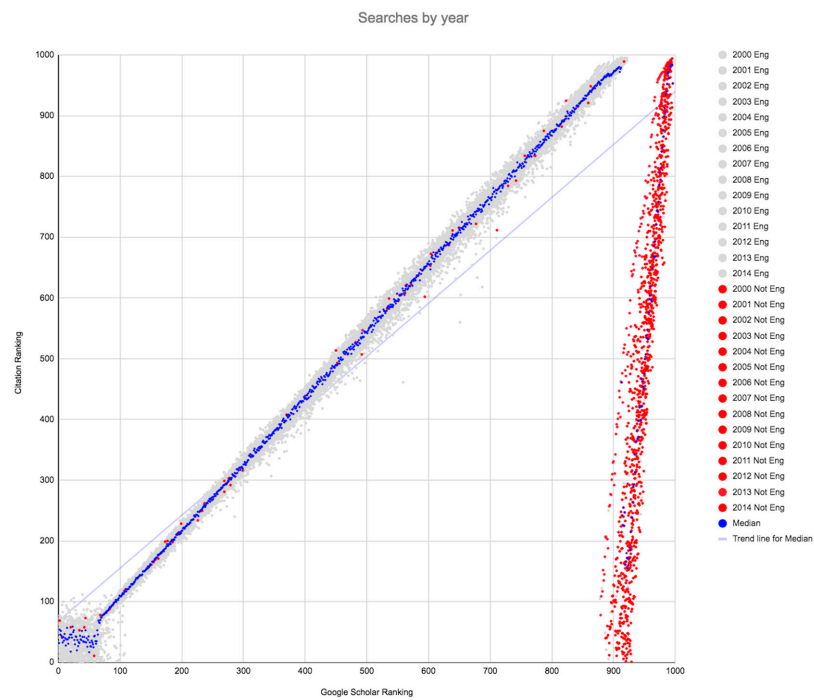


Figure 1. Scatter plot of searches by year.

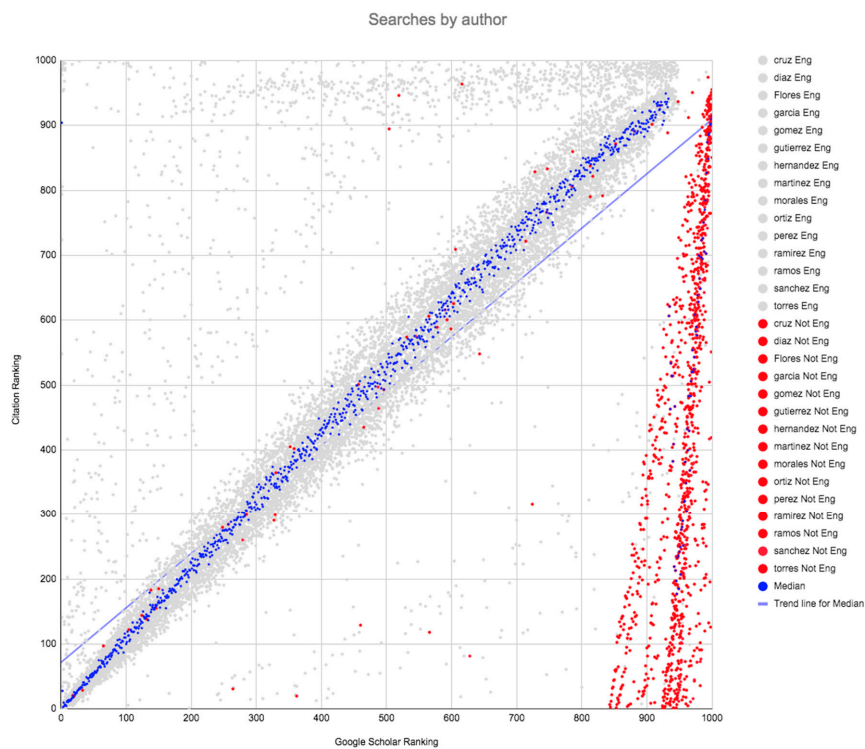


Figure 2. Scatter plot of searches by author.

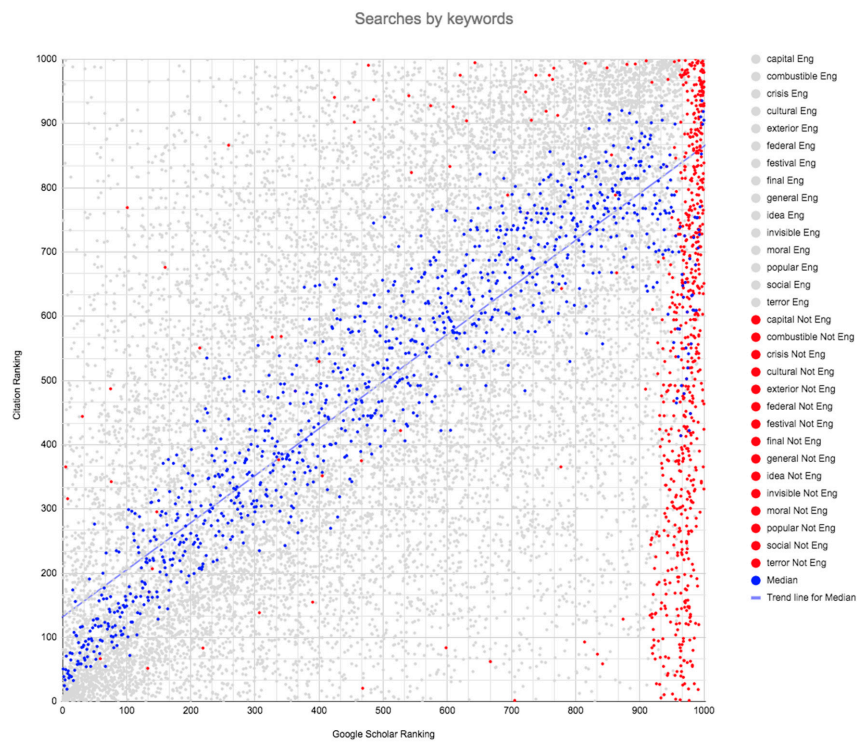


Figure 3. Scatter plot of searches by keyword.

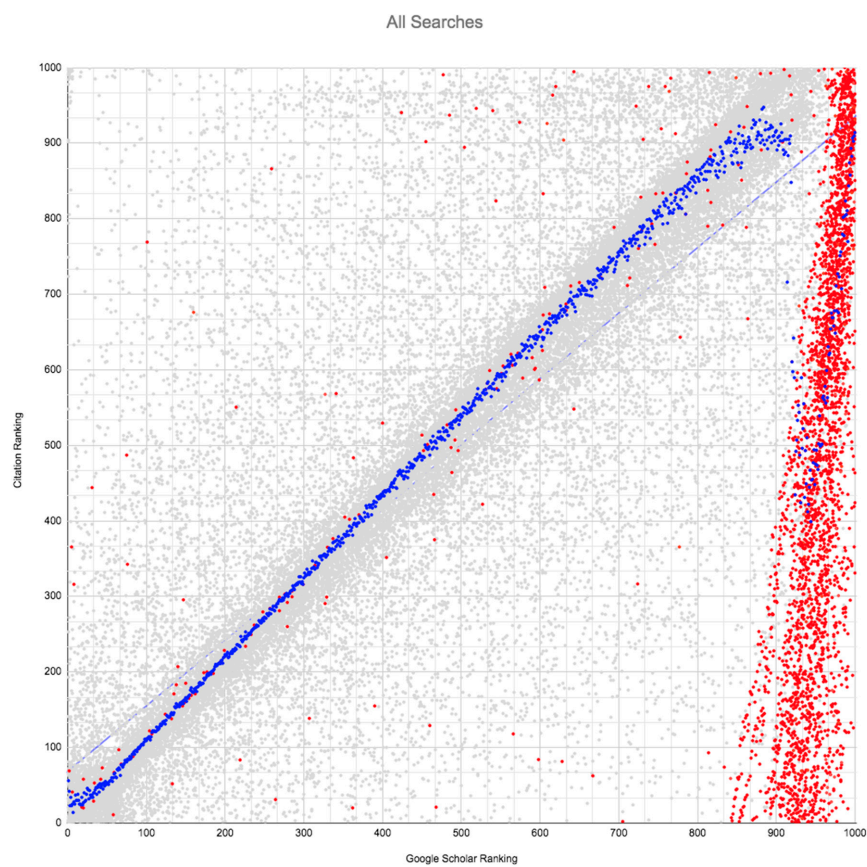


Figure 4. Scatter plot of all searches.

Documents in languages other than English do not adhere to the pattern set by the overall results, a pattern that places the points along the diagonal of the scatter plot, corresponding to a high correlation coefficient. A pattern that gives a good position in the ranking according to the number of citations received also gives a good position in the Google Scholar native ranking. The breakdown in this pattern is especially notable in the case of the dots located in the lower right corner of the four graphs, corresponding to documents that appear in the top 100 positions in the ranking by number of citations, but which are located above rank position 900 in the Google Scholar native ranking.

This second point cloud that forms above rank position 900 has been previously detected by Beel [19] and Rovira [23], but these authors did not attribute its origin to the language in which the documents were published.

The scatter plots illustrate quite clearly the same inconsistency in the ranking algorithm that we report above in our data analysis. This inconsistency implies that documents written in languages other than English are treated in a discriminatory manner.

## 5. Discussion

Our data are insufficient to be able to conclude whether the ranking of documents in languages other than English is due simply to an accidental bias, caused by an unintentional error in the software program, or whether it is an intentional outcome of the design of the algorithm, based, for example, on a belief that the English language has what might be considered an internationally superior dimension to that of other languages. This would explain (although we cannot state with any certainty) why in multilingual searches documents written in other languages become virtually invisible.

Neither do we have sufficient data to determine whether, in fact, two different algorithms are being used, one for English and one for all other languages, or whether the documents in English are ranked first and only then are the documents in all other languages ranked employing the same algorithm, to mention just a few possibilities.

However, what we have been able to demonstrate is that the Google Scholar relevance ranking algorithm favors documents in English and handicaps documents written in other languages in searches that produce multilingual results. Moreover, the vast majority of documents written in languages other than English have no chance at all of being found in searches of this type, even if they are high quality articles with hundreds or, even, thousands of citations received. Indeed, for 94% of documents of this type their possibilities of being found are absolutely zero, given that almost nobody accesses items located above a rank position of 700 on a list of search results (Table 5).

More specifically, only 0.2% of these non-English documents have a chance of being found in a multilingual search in Google Scholar, given that they would occupy a place in the top 20 rankings. However, if the ranking algorithm were to be applied homogeneously across all documents, giving the same weight to citations received in all of them, the percentage of non-English documents among these top 20 positions would be 3.2%, i.e., much higher than the value of 0.2% actually obtained (Table 5).

This anomalous behavior is detected in multilingual searches that return results in different languages, that is, searches conducted, for example, by year, by author, or by keywords with identical spelling in more than one language. Clearly, in typical keyword searches, in which all the results are in the same language, this bias does not occur.

We can, moreover, rule out the possibility that other variables, in addition to the language in which a document is written, might be intervening, to produce this anomalous ranking. We have been able to verify that these results only occur in documents written in a language other than English and at such high percentages that the possibility of the participation of any other factors can be discarded.

Despite the restrictions we imposed on our data collection, the number of results in languages other than English is relatively low compared to the whole sample of data analyzed: 4.4% in the case of keywords, 7.7% in that of authors and 8.4% in the case of years (Table 1). These low percentages are also a consequence of a biased ranking. Future studies



need to include searches with a more balanced number of results between documents in English and those published in other languages in order to corroborate that the ranking is also discriminatory in these cases. However, the only way to obtain percentages close to 50% is by conducting unnatural searches, such as combining years and keywords of one or two letters.

Finally, it is worth highlighting two additional aspects that might also be the object of future research. First, we have reported that the point cloud for searches by keyword is much more dispersed than those for searches by year or by author. This dispersion is a clear indication that in searches by keyword other factors, in addition to the number of citations received, are intervening. For this reason, more research is needed to identify the other factors that contribute to this dispersion. Second, it would be of interest to determine whether the bias we report differs according to the non-English language used for publication, in other words, whether documents written in Spanish, French or German, for example, present a specific ranking pattern in Google Scholar searches.

## 6. Conclusions

Both standard SEO and ASEO have a primary concern with determining the main factors employed in ranking algorithms. The aim of these two fields of enquiry is to make more visible the characteristics that articles present and which, in turn, serve as criteria for ranking. This is the case of so-called “white hat SEO”, that is, positioning actions that seek to optimize the visibility of information by ethical means.

This need to identify positioning factors has motivated and guided our research here, which reports an unexpected result with major repercussions for researchers. We have demonstrated that in Google Scholar, in the case of multilingual results, documents in languages other than English are systematically relegated to positions that make them virtually invisible. These positions are not the result of applying the standard ranking algorithm, as is done for documents in English. Moreover, in practical terms, it is of little importance whether this bias is accidental or the result of an algorithm designed specifically to rank differently according to the language in which a document is written.

However, it is important to draw attention to this major bias so that it might be limited or compensated for when a search is conducted. 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.

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. This recommendation is especially critical for researchers working in the areas mentioned in the previous paragraph whose names are the same in English and in their national language.

**Author Contributions:** Conceptualization, C.R.; Methodology, C.R.; Validation L.C. and C.L.; Investigation C.R., and L.C.; Resources, C.R. and C.L.; Data Curation, C.R.; Writing—Original Draft Preparation, C.R.; Writing—Review and Editing, L.C. and C.L.; Supervision L.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work is part of the Project “Interactive storytelling and digital visibility in interactive documentary and structured journalism”. RTI2018-095714-B-C21 (Micinn/Feder) Ministry of Science, Innovation and Universities of Spain.

**Data Availability Statement:** [dataset] Cristòfol Rovira; Lluís Codina; Carlos Lopezosa. 2021. Data set of the article: Language Bias in the Google Scholar Ranking Algorithm; Zenodo; Version 1; <http://doi.org/10.5281/zenodo.4463201>.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- González-Solar, L. Marca personal en entornos académicos: Una perspectiva institucional. *An. Doc.* **2018**, *21*. [CrossRef]
- Harzing, A.W. Building Your Academic Brand through Engagement with Social Media. 2018. Available online: <https://harzing.com/blog/2018/05/building-your-academic-brand-through-engagement-with-social-media> (accessed on 17 October 2020).
- Marcos, M.-C.; González-Caro, C. Comportamiento de los usuarios en la página de resultados de los buscadores. Un estudio basado en eye tracking. *El Prof. de la Inf.* **2010**, *19*, 348–358. [CrossRef]
- Yalçın, N.; Köse, U. What is search engine optimization: SEO? *Procedia Soc. Behav. Sci.* **2010**, *9*, 487–493. [CrossRef]
- Ziakis, C.; Vlachopoulou, M.; Kyrkoudis, T.; Karagkiozidou, M. Important factors for improving Google search rank. *Future Internet* **2019**, *11*, 32. [CrossRef]
- López-Cózar, D.E.; Robinson-García, N.; Torres-Salinas, D. Manipulating Google Scholar citations and Google Scholar metrics: Simple, easy and tempting. *arXiv* **2012**, arXiv:1212.0638.
- López-Cózar, D.E.; Robinson-García, N.; Torres-Salinas, D. The Google Scholar experiment: How to index false papers and manipulate bibliometric indicators. *J. Assoc. Inf. Sci. Technol.* **2014**, *65*, 446–454. [CrossRef]
- SEMrush Blog. Available online: <https://www.semrush.com/blog/> (accessed on 1 July 2020).
- The Moz Blog. Available online: <https://moz.com/blog> (accessed on 1 July 2020).
- Yoast SEO Blog. Available online: <https://yoast.com/seo-blog/> (accessed on 1 July 2020).
- Developers Google. Available online: <https://developers.google.com/search/blog#blog-principal-de-la-busqueda-de-google> (accessed on 1 July 2020).
- Search Engine Journal. Available online: <https://www.searchenginejournal.com/> (accessed on 1 July 2020).
- Search Engine Land. Available online: <https://searchengineland.com/> (accessed on 1 July 2020).
- Search Engine Watch. Available online: <https://www.searchenginewatch.com/> (accessed on 1 July 2020).
- Clarke, A. *SEO 2020 Learn Search Engine Optimization with Smart Internet Marketing Strategies: Learn SEO with Smart Internet Marketing Strategies*; Newest, Ed.; EEUU: Trenton, NJ, USA, 2019; pp. 1–244.
- Kent, P. *SEO For Dummies*; EEUU: Trenton, NJ, USA, 2019; pp. 1–478.
- Maciá, F. *SEO Avanzado. Casi todo lo Que sé Sobre Posicionamiento Web*; Anaya: Barcelona, Spain, 2020; pp. 1–416.
- Google. How Google Search Works. *Learn How Google Discovers, Crawls, and Serves Web Pages*. Available online: <https://support.google.com/webmasters/answer/70897?hl=en> (accessed on 1 July 2020).
- Beel, J.; Gipp, B. Google scholar's ranking algorithm: An introductory overview. In Proceedings of the 12th International Conference on Scientometrics and Informetrics, ISSI'09, Istanbul, Turkey, 14–17 July 2009; pp. 230–241. Available online: <https://goo.gl/c8a6YU> (accessed on 23 January 2021).
- Beel, J.; Gipp, B. Google Scholar's ranking algorithm: The impact of articles' age (an empirical study). In Proceedings of the Sixth International Conference on Information Technology: New Generations, ITNG'09, Las Vegas, NA, USA, 27–29 April 2009; pp. 160–164. [CrossRef]
- Beel, J.; Gipp, B. Google scholar's ranking algorithm: The impact of citation counts (an empirical study). In Proceedings of the Third International Conference on Research Challenges in Information Science, RCIS 2009c, Nice, France, 22–24 April 2009; pp. 439–446. [CrossRef]
- Beel, J.; Gipp, B.; Wilde, E. Academic search engine optimization (ASEO). Optimizing scholarly literature for Google Scholar & co. *J. Sch. Publ.* **2010**, *41*, 176–190. [CrossRef]
- Rovira, C.; Guerrero-Solé, F.; Codina, L. Received citations as a main SEO factor of Google Scholar results ranking. *El Prof. de la Inf.* **2018**, *27*, 559–569. [CrossRef]
- Rovira, C.; Codina, L.; Guerrero-Solé, F.; Lopezosa, C. Ranking by relevance and citation counts, a comparative study: Google Scholar, Microsoft Academic, WoS and Scopus. *Future Internet* **2019**, *11*, 202. [CrossRef]
- Martín-Martín, A.; Orduña-Malea, E.; Ayllón, J.M.; López-Cózar, E.D. Does Google Scholar contain all highly cited documents (1950–2013). *arXiv* **2014**, arXiv:1410.8464.
- Sparks, A. 8 Winning Hacks to Use Google Scholar for Your Research Paper, Editage. 2018. Available online: <https://www.editage.com/insights/8-winning-hacks-to-use-google-scholar-for-your-research-paper> (accessed on 1 July 2020).
- Cordero, J.J. 8 Estrategias Para Promocionar los Artículos Profesionales de tu Blog. 2018. Available online: <https://www.javiercordero.com/como-promocionar-articulos-blog/> (accessed on 1 July 2020).
- Florido, M. Google Académico—7 Consejos para mejorar el posicionamiento. 2015. Available online: <https://www.marketingandweb.es/marketing/google-academico/> (accessed on 1 July 2020).
- Drew, C. 11 Best Tips on How to use Google Scholar. 2020. Available online: <https://helpfulprofessor.com/google-scholar/> (accessed on 1 July 2020).
- Miles, S. 12 Tips for Increasing Your Visibility in Google Search. 2020. Available online: <https://webpublisherpro.com/12-tips-for-increasing-your-visibility-in-google-search/> (accessed on 1 July 2020).

31. UCA Library. Research Visibility. SEO for Authors: A How-to Guide. Available online: <https://guides.library.ucla.edu/seo/author> (accessed on 2 July 2020).
32. UNED. Investiga UNED. 10 Consejos Para Difundir tu Investigación y Conseguir más Impacto. 2017. Available online: <http://investigauned.uned.es/10-consejos-para-difundir-tu-investigacion-y-conseguir-mas-impacto/> (accessed on 2 July 2020).
33. University of Pittsburgh. How to Increase the Visibility of Your Research? 2020. Available online: <https://pitt.libguides.com/researchvisibility> (accessed on 2 July 2020).
34. University of Montana. Google Scholar at the University of Montana. 2020. Available online: <https://libguides.lib.umt.edu/c.php?g=854135&p=6115806> (accessed on 2 July 2020).
35. Wiley. Search Engine Optimization (SEO) for Your Article. Available online: <https://authorservices.wiley.com/author-resources/Journal-Authors/Prepare/writing-for-seo.html> (accessed on 2 July 2020).
36. Emerald Publishing. How to . . . Make Your Research Easy to Find with SEO. Available online: <https://www.emeraldgroupublishing.com/services/authors/author-how-guides/make-your-research-easy-find-seo> (accessed on 2 July 2020).
37. Sage Publishing. Promote Your Article. Available online: <https://uk.sagepub.com/en-gb/eur/promote-your-article> (accessed on 2 July 2020).
38. Plos. Spreading the Word about Your Research. Available online: <https://plos.org/article-promotion/> (accessed on 2 July 2020).
39. Taylor and Francis. Search Engine Optimization for Academic Articles. Available online: <https://authorservices.taylorandfrancis.com/research-impact/search-engine-optimization-for-academic-articles/> (accessed on 2 July 2020).
40. Elsevier. Get Found—Optimize Your Research Articles for Search Engines. 2012. Available online: <https://www.elsevier.com/connect/get-found-optimize-your-research-articles-for-search-engines> (accessed on 2 July 2020).
41. Codina, L. SEO Académico: Definición, Componentes y Guía de Herramientas. 2019. Available online: <https://www.lluiscodina.com/seo-academico-guia> (accessed on 9 July 2020).
42. Martín-Martín, A.; Ayllón, J.M.; Orduña-Malea, E.; López-Cózar, E.D. Google Scholar metrics released: A matter of languages and something else. *arXiv* **2016**, arXiv:1607.06260v1.
43. Muñoz-Martín, B. Incrementa el impacto de tus artículos y blogs: De la invisibilidad a la visibilidad. *Rev. Soc. Otorrinolaringol. Castilla León Cantab. Rioja* **2015**, *6*, 6–32. Available online: <http://hdl.handle.net/10366/126907> (accessed on 23 January 2021).
44. Giustini, D.; Boulos, M.N.K. Google Scholar is not enough to be used alone for systematic reviews. *Online J. Public Health Inform.* **2013**, *5*, 214. [CrossRef] [PubMed]
45. Walters, W.H. Google Scholar search performance: Comparative recall and precision. *Portal-Libr. Acad.* **2008**, *9*, 5–24. [CrossRef]
46. De-Winter, J.; Zadpoor, A.; Dodou, D. The expansion of Google Scholar versus Web of Science: A longitudinal study. *Scientometrics* **2014**, *98*, 1547–1565. [CrossRef]
47. Harzing, A.W. A preliminary test of Google Scholar as a source for citation data: A longitudinal study of Nobel prize winners. *Scientometrics* **2013**, *94*, 1057–1075. [CrossRef]
48. Harzing, A.W. A longitudinal study of Google Scholar coverage between 2012 and 2013. *Scientometrics* **2014**, *98*, 565–575. [CrossRef]
49. De-Groote, S.L.; Raszewski, R. Coverage of Google Scholar, Scopus, and Web of Science: A case study of the h-index in nursing. *Nurs Outlook* **2012**, *60*, 391–400. [CrossRef]
50. Orduña-Malea, E.; Ayllón, J.M.; Martín-Martín, A.; Delgado-López-Cózar, E. About the size of Google Scholar: Playing the numbers. *arXiv* **2014**, arXiv:1407.6239.
51. Orduña-Malea, E.; Ayllón, J.M.; Martín-Martín, A.; Delgado-López-Cózar, E. Methods for estimating the size of Google Scholar. *Scientometrics* **2015**, *104*, 931–949. [CrossRef]
52. Pedersen, L.A.; Arendt, J. Decrease in free computer science papers found through Google Scholar. *Online Inf. Rev.* **2014**, *38*, 348–361. [CrossRef]
53. Jamali, H.R.; Nabavi, M. Open access and sources of full-text articles in Google Scholar in different subject fields. *Scientometrics* **2015**, *105*, 1635–1651. [CrossRef]
54. Jamali, H.R.; Asadi, S. Google and the scholar: The role of Google in scientists' information-seeking behaviour. *Online Inf. Rev.* **2010**, *34*, 282–294. [CrossRef]
55. Aguillo, I.F. Is Google Scholar useful for bibliometrics? A webometric analysis. *Scientometrics* **2012**, *91*, 343–351. [CrossRef]
56. Jacsó, P. Calculating the h-index and other bibliometric and scientometric indicators from Google Scholar with the Publish or Perish software. *Online Inf. Rev.* **2009**, *33*, 1189–1200. [CrossRef]
57. Torres-Salinas, D.; Ruiz-Pérez, R.; Delgado-López-Cózar, E. Google scholar como herramienta para la evaluación científica. *El Prof. de la Inf.* **2009**, *18*, 501–510. [CrossRef]
58. Beel, J.; Gipp, B. Academic search engine spam and Google Scholar's resilience against it. *J. Electron. Publ.* **2010**, *13*. [CrossRef]
59. Delgado-López-Cózar, E.; Robinson-García, N.; Torres-Salinas, D. Manipular Google Scholar citations y Google Scholar metrics: Simple, sencillo y tentador. In *EC3 Working Papers*; Universidad De Granada: Granada, Spain, 2012. Available online: <http://hdl.handle.net/10481/20469> (accessed on 1 July 2019).
60. Meho, L.; Yang, K. Impact of Data Sources on Citation Counts and Rankings of LIS Faculty: Web of Science Versus Scopus and Google Scholar. *J. Assoc. Inf. Sci. Technol.* **2006**, *58*, 2105–2125. [CrossRef]
61. Martín-Martín, A.; Orduña-Malea, E.; Ayllón, J.M.; Delgado-López-Cózar, E. Back to the past: On the shoulders of an academic search engine giant. *Scientometrics* **2016**, *107*, 1477–1487. [CrossRef]

62. Van-Aalst, J. Using Google Scholar to estimate the impact of journal articles in education. *Educ. Res.* **2010**, *39*, 387–400. [CrossRef]
63. Jacsó, P. Testing the calculation of a realistic h-index in Google Scholar, Scopus, and Web of Science for FW Lancaster. *Libr. Trends* **2008**, *56*, 784–815. [CrossRef]
64. Jacsó, P. The pros and cons of computing the h-index using Google Scholar. *Online Inf. Rev.* **2008**, *32*, 437–452. [CrossRef]
65. Jacsó, P. Using Google Scholar for journal impact factors and the h-index in nationwide publishing assessments in academia –siren songs and air-raid sirens. *Online Inf. Rev.* **2012**, *36*, 462–478. [CrossRef]
66. Martín-Martín, A.; Orduña-Malea, E.; Harzing, A.W.; Delgado-López-Cózar, E. Can we use Google Scholar to identify highly-cited documents? *J. Informetr.* **2017**, *11*, 152–163. [CrossRef]
67. Farhadi, H.; Salehi, H.; Yunus, M.; Aghaei-Chadegani, A.; Farhadi, M.; Fooladi, M.; Ale-Ebra-him, N. Does it matter which citation tool is used to compare the h-index of a group of highly cited researchers? *Aust. J. Basic Appl. Sci.* **2013**, *7*, 198–202. Available online: <https://ssrn.com/abstract=2259614> (accessed on 23 January 2021).
68. Marks, T.; Le, A. Increasing article findability online: The four Cs of search engine optimization. *Law Libr. J.* **2017**, *109*, 83. [CrossRef]
69. Kearl, M.; Noteboom, C.; Tech, D. A proposed improvement to Google Scholar algorithms through broad topic search emergent research forum paper. *Fac. Res. Publ.* **2017**, *6*, 1–5. Available online: <https://scholar.dsu.edu/bispapers/6> (accessed on 23 January 2021).
70. Localseoguide. How Have Local Ranking Factors Changed? 2019. Available online: <https://www.localseoguide.com/guides/local-seo-ranking-factors/> (accessed on 1 July 2020).
71. Searchmetrics. Rebooting for Relevance. 2016. Available online: <https://www.searchmetrics.com/knowledge-hub/studies/ranking-factors-2016/> (accessed on 1 July 2020).
72. MOZ. Search Engine Ranking Factors 2015. Available online: <https://moz.com/search-ranking-factors/correlations> (accessed on 1 July 2019).
73. Dave, D. 11 Things You Must Know About Google's 200 Ranking Factors. 2018. Available online: <https://www.searchenginejournal.com/google-200-ranking-factors-facts/265085/> (accessed on 10 September 2019).
74. Chariton, R. Google Algorithm—What Are the 200 Variables? 2004. Available online: <https://www.webmasterworld.com/google/4030020.htm> (accessed on 10 September 2019).
75. Wiggers, K. Google Details How It's Using AI and Machine Learning to Improve Search. 2020. Available online: <https://venturebeat.com/2020/10/15/google-details-how-its-using-ai-and-machine-learning-to-improve-search/> (accessed on 10 July 2020).
76. Google. About Google Scholar. Available online: <http://scholar.google.com/intl/en/scholar/about.html> (accessed on 1 July 2019).
77. Mayr, P.; Walter, A.-K. An exploratory study of google scholar. *Online Inf. Rev.* **2007**, *31*, 814–830. [CrossRef]
78. Gielen, M.; Rosen, J. Reverse Engineering the YouTube, tubefilter.com. 2016. Available online: <http://www.tubefilter.com/2016/06/23/reverse-engineering-youtube-algorithm/> (accessed on 1 July 2019).
79. Harzing, A.-W. Publish or Perish. 2016. Available online: <https://harzing.com/resources/publish-or-perish> (accessed on 1 July 2019).
80. Harzing, A.W. *The Publish Or Perish Book: Your Guide to Effective and Responsible Citation Analysis*; Tarma Software Research Pty Ltd.: Melbourne, Australia, 2011; pp. 39–342. Available online: [https://EconPapers.repec.org/RePEc:spr:scient:v:88:y:2011:i:1:d:10.1007\\_s11192-011-0388-8](https://EconPapers.repec.org/RePEc:spr:scient:v:88:y:2011:i:1:d:10.1007_s11192-011-0388-8) (accessed on 1 July 2019).
81. R Core Team. R: A Language and Environment for Statistical Computing. Available online: <https://www.R-project.org> (accessed on 1 July 2019).
82. Revelle, W. Psych: Procedures for Personality and Psychological Research, Northwestern University. 2017. Available online: <https://www.scholars.northwestern.edu/en/publications/psych-procedures-for-personality-and-psychological-research> (accessed on 1 July 2019).
83. Lemon, J. Plotrix: A package in the red light district of R. *R News* **2006**, *6*, 8–12.