

## THE UNCONTROLLED GIANT: GOOGLE SCHOLAR & GOOGLE SCHOLAR CITATIONS

The errors that can compromise the metric portrait of an author offered by Google Scholar can be grouped into two main sections. First, the errors Google Scholar sometimes makes when it indexes a document or when it assigns citations to it. Second, the specific errors that are sometimes made during the creation of a Google Scholar Citations profile.

The former are a logical consequence of the tricky and complex task that is automatically searching the current academic papers available in the net. This task also involves merging in only one record all possible versions of the same work, and linking to it all documents in which it is cited (keeping in mind that these documents and references can be presented in the most varied formats).

The latter are the ultimate responsibility of the author, who must periodically revise his/her profile in order to eliminate misattributed documents which might have been included in the automatic weekly updates, clean the records by merging different versions of the same document when Google Scholar's algorithms are not able to detect their similarity, as well as improve and complete the bibliographic references of these documents (filling in blank fields in a document when Google Scholar hasn't been able to find that information).

### 1. Mistakes in Google Scholar by automatic indexes a document or assigns citations to it.

Next, we classify, describe, and illustrate some of the most common mistakes in Google Scholar:

#### a) Incorrect identification of the title of the document

Google Scholar always tries to extract bibliographic information from the HTML Meta tags in a webpage. When there are no Meta tags available, it parses the webpage itself (the HTML code of the page, or even PDFs themselves). Even though its spiders are able to successfully parse pages with a quite broad range of different structures, and despite the fact that they have published a very clear set of inclusion guidelines, some parsing errors occasionally arise for documents extracted from websites with unusual layouts. It is not rare in these cases that an incorrect text string is selected as the title of the document. In Figure 1 we illustrate an example in which an incorrect string ("www.redalyc.org") has been selected as the title of the document in several records, probably because it is the string that is featured with a higher font size in the first page of the PDF document from which Google Scholar has parsed the bibliographic information. Note that the authors and the source publications are correctly assigned.

Figure 1. Document titles improperly identified in Google Scholar: URLs

[PDF] <a href="http://www.redalyc.org">www. redalyc. org</a> J Rzedowski, GC de Rzedowski - Acta botánica mexicana, 1990 - core.ac.uk ABSTRACT Phoradendron perredactum, a miniature leafless and practically acaulescent stem parasite of Bursera discolor, is described as new and illustrated. Two types of individuals were observed: the more common branched ones, less than 1 cm long with a ... Citado por 626 Artículos relacionados Las 8 versiones Citar Guardar Más	[PDF] de core.ac.uk
[PDF] <a href="http://www.redalyc.org">www. redalyc. org</a> A Escobar - Tabula rasa, 2003 - core.ac.uk Abstract This article presents and examines critically the work of a group of researchers in Latin America and the United States that are crafting a novel reinterpretation of modernity, globality and difference. I label this group as «the modernity/coloniality research program» ... Citado por 263 Artículos relacionados Las 16 versiones Citar Guardar Más	[PDF] de core.ac.uk
[PDF] <a href="http://www.redalyc.org">www. redalyc. org</a> MB Salas - Acta Zoológica Mexicana (nueva serie), 2000 - core.ac.uk ABSTRACT The species richness of mammals from the region of La Cañada, in the Tehuacan-Cuicatlan Valley, in the state Oaxaca, is presented. As a result of field works, revision of scientific collections, and literature. 52 species, in seven orders, 15 families, ... Citado por 244 Artículos relacionados Las 4 versiones Citar Guardar Más	[PDF] de core.ac.uk
[PDF] <a href="http://www.redalyc.org">www. redalyc. org</a> M Montero - Revista latinoamericana de psicología, 1990 - catedralibremartinbaro.org En este trabajo se plantea la idea de que la psicología social comunitaria, la psicología social crítica y la psicología social de la liberación configuraron una forma de respuesta neoparadigmática (las dos primeras surgen en América Latina en los años 70, la tercera ... Citado por 149 Artículos relacionados Las 11 versiones Citar Guardar Más	[PDF] de catedralibremartinbaro.org
[PDF] <a href="http://www.redalyc.org">www. redalyc. org</a> R Grosfoguel - Tabula Rasa, 2006 - core.ac.uk Abstract This article discusses the epistemic implications of the de-colonial turn of the decolonization of the «global capitalism» concept, as it has been used in the paradigms of political economy and cultural studies. Other terms that are discussed include « ... Citado por 158 Artículos relacionados Las 22 versiones Citar Guardar Más	[PDF] de core.ac.uk
[PDF] <a href="http://www.redalyc.org">www. redalyc. org</a> O Velasco, JL Valdespino, R Tapia... - Salud Pública de ..., 1992 - s2 medicina uady.mx De entrada es importante reconocer que se ha avanzado algo en la salud de la gente tanto en los países desarrollados como en los menos desarrollados. Estos logros en salud entran en la gran dimensión de las causas y efectos del crecimiento poblacional como resultante ... Citado por 117 Artículos relacionados Las 10 versiones Citar Guardar Más	[PDF] de uady.mx

In many other occasions, other text strings, such as the author's name and/or the year of publication, are incorrectly selected as the title of the document. In Figure 2 we can observe how "de Solla" has been selected as the title in many records.

**Figure 2. Author names incorrectly selected as document titles in Google Scholar**

The screenshot shows a list of search results for 'de Solla'. Each entry includes a citation title, a link to the document, and citation statistics. The titles are:
 

- [CITATION] de Solla (1951) DJ Price - Quantitative Measures of the Development of Science...
- [CITATION] de Solla (1961) Science since babylon DJ Price - New Haven, 1961
- [CITATION] de Solla (1961) DJ Price - Science since Babylon, 1975
- [CITATION] de Solla (1961): Science since Babylon. Yale DJ Price - Yale University Press
- [CITATION] de Solla (1961) 1975. Science since Babylon DJ Price - New Haven, CT: Yale University ...
- [CITATION] de Solla (1963) Little science, big science DJ Price - 1963 - New York: Columbia University ...
- [CITATION] de Solla (1963) DJ Price - Little science, big science, 1973
- [CITATION] de Solla (1963). Little science DJ Price - Big Science
- [CITATION] de Solla (1963, 1986) DJ Price - Little Science, Big Science... and Beyond
- [CITATION] de Solla (1963) Big Science, Little Science DJ Price - New York: Columbia University Press)
- [CITATION] de Solla (1972) DJ Price - ... and technology: Distinctions and interrelationships. En ...
- [CITATION] de Solla (1974) Little science, Big science DJ Price - Suhrkamp, Frankfurt/Main
- [CITATION] de Solla (1974) DJ Price - Little Science, Big Science
- [CITATION] de Solla (1976a) J Price - Little science, big science... and beyond
- [CITATION] de Solla (1984) DJ Price - The science/technology relationship, the craft of ...
- [CITATION] de Solla 1984a. DJ Price - Notes towards a Philosophy of the
- [CITATION] de Solla 1986 DJ Price - Little science, big science... and beyond
- [CITATION] de Solla (1986 [1963]), Little Science, Big Science... DJ Price - New York
- [CITATION] de Solla 1986 [1963] DJ Price - Little Science, Big Science and Beyond

[https://scholar.google.com/scholar?start=0&q=allintitle:%22de+solla%22+-Moravcsik+-gulls+-comments+-1922+-foreword+-Toward+-tribute+-space+-pensamento+-address+-appreciation&hl=en&as\\_sdt=0,5](https://scholar.google.com/scholar?start=0&q=allintitle:%22de+solla%22+-Moravcsik+-gulls+-comments+-1922+-foreword+-Toward+-tribute+-space+-pensamento+-address+-appreciation&hl=en&as_sdt=0,5)

### b) Ghost authors

The topic of ghost authors, citations, and documents was approached by Jacsó in numerous works, mostly before Google Scholar Citations was launched. Although profiles have served to filter and correct many mistakes, some of them still persist, especially if authors do not clean their personal profiles. In Figure 3 we can see one such example. In this case, the record only displays one person as the author of the article (Carmen Martín Moreno), when in fact the article was written by two authors (Elías Sanz-Casado and Carmen Martín Moreno). In this case, Google Scholar extracted the bibliographic information from the HTML Meta tags in the website of the journal where the article was published, but, as we can see in Figure 3 (bottom image), these metadata were already incorrect (the title should read "Técnicas bibliométricas aplicadas a los estudios de usuarios"), and incomplete (Elías Sanz-Casado is missing from the record). Nonetheless, thanks to Google Scholar Citations, Elías was able to add the document to his profile, even if his name is still missing from the authors field (Figure 3, top left).

**Figure 3. Missing authors in primary versions of documents in Google Scholar**

The screenshot shows search results for 'Técnicas bibliométricas aplicadas a los estudios de usuarios'. The top result is by Carmen Martín Moreno, with a red box around the author name. Below it, a snippet shows the title and authors: 'Técnicas bibliométricas aplicadas a los estudios de usuarios' by 'E Sanz Casado, C Martín Moreno'. To the right, a document preview shows the title and a red box around the author field containing only 'Carmen Martín Moreno'. Below the preview, there is a small profile picture of Elías Sanz-Casado and a list of metadata including 'Journal: Revista general de información y documentación', 'Volume: 7', 'Issue: 2', and 'Pages: 41'. The description at the bottom repeats the abstract text from the snippet.

Revista General de Información y Documentación

UNIVERSIDAD COMPLUTENSE MADRID

Acerca de... Buscar Último número Números anteriores Noticias Normas Indexación

Inicio » 1987, No 2 (1987) » Martín Moreno

**Técnicas bibliométricas aplicadas a los estudios de usuarios.**  
Carmen Martín Moreno

**Resumen**  
Sin resumen  
Texto completo: PDF

**Rebacks**  
No hay Rebacks actualmente.

COMPARTIR

Revista General de Información y Documentación  
ISSN 1132-1973  
ISSN-e 1989-2958

© 2016, Universidad Complutense de Madrid  
Instituto Complutense de Gestión de Publicaciones

Si está interesado en publicar la edición electrónica de una revista de la UCM envíenos un correo a [servicio\\_publicaciones@net.ucm.es](mailto:servicio_publicaciones@net.ucm.es) y nos pondremos en contacto.

**Herramientas del artículo**  
Imprimir resumen  
Información de educación  
Información bibliográfica  
Buscar referencias  
Política de Revisión  
Enviar por correo este artículo (a nuevos entes)  
Mandar correo a autor (a nuevos entes)  
Enviar un comentario de reseña (a new entes)

**Items relacionados**  
Cruce datos

**Idioma**  
Español

**Contenido de la Revista**  
Por número  
Por autor  
Por título  
Índice de revistas

**Alertas**  
Ver  
Suscribirse / Desuscribirse  
Ayuda de la revista

c) Book reviews indexed as books

Among the most common mistakes in document identification is mistaking the review of a book for the book itself. In Figure 4 we show two different records which correspond with book reviews of the work “Introduction to informetrics. Quantitative methods in Library, Documentation and Information Science” by Egghe and Rousseau. At a first glance the first record (Figure 4; top) looks like a normal record, since the title and authors of the book have been correctly identified. However, the record actually points to a review of the book published in Revista Española de Documentación Científica. The second record (Figure 4; bottom), is also a review of the book which was published in Aslib Proceedings. In this case, the author of the review is the one who appears in the GS record (Brookes).

Figure 4. Authorship and attribution of book reviews

**Introduction to Informetrics. Quantitative methods in Library, Documentation and Information Science**  
L Egghe, R Rousseau... - Revista Española de ..., 1991 - search.proquest.com  
REVISION DE LIBROS INTRODUCTION TO INFORMETRICS. QUANTITATIVE METHODS IN LIBRARY, DOCUMENTATION AND INFORMATION SCIENCE Egghe L. y R. Rousseau Edit. Elsevier Science Publishers. Amsterdam, 1990 450 pags., 24 cm. Indices. Precio: ...  
Cite Save

[CITATION] **INTRODUCTION TO INFORMETRICS-QUANTITATIVE METHODS IN LIBRARY, DOCUMENTATION AND INFORMATION-SCIENCE-EGGHE, L, ...**  
BC BROOKES - 1991 - ASLIB 20-24 OLD ST, LONDON, ...  
Cite Save

d) Incorrect attribution of documents to authors

Somewhat related to the previous error is the attribution of a document to the wrong authors. In Figure 5 we observe a special case: the book “Introduction to informetrics. Quantitative methods in Library, Documentation and Information Science” by Egghe and Rousseau, is wrongly attributed to Tague-Sutcliffe, probably because this author has a short publication in the journal Information Processing & Management (Figure 5; bottom) with a similar title (“An introduction to informetrics”).

Figure 5. Authorship improperly assigned in Google Scholar

[CITATION] **INTRODUCTION TO INFORMETRICS-QUANTITATIVE METHODS IN LIBRARY, DOCUMENTATION AND INFORMATION-SCIENCE-EGGHE, L, ...**  
J TAGUE - 1991 - UNIV CHICAGO PRESS 5720 S ...  
Cite Save

[HTML] **An introduction to informetrics**  
J Tague-Sutcliffe - Information processing & management, 1992 - Elsevier  
Abstract The scope and significance of the field of **informetrics** is defined and related to the earlier fields of bibliometrics and scientometrics. The phenomena studied by informetricians are identified. The major contributors to the field in the past are described and current ...  
Cited by 318 Related articles All 5 versions Cite Save

e) Failing to merge all versions of a same document into one record

Although the algorithms for grouping versions work well in most cases, Google Scholar sometimes fails to realize that two or more records it has indexed actually represent the same document. This happens when there are enough formal differences between the metadata of the two versions (differences in the way the name of the authors have been stored, in the title, the year of publication...), that Google Scholar judges they're not similar enough to be the same document. This issue mostly affects document types other than journal articles (books, book chapters, reports), but duplicate articles also exist. Articles translated into one or more languages are an extreme example: in those cases, the title of the original version is completely different to that of the translated version, so it is understandable that Google Scholar doesn't realize they are the same document. From a bibliometric perspective, however, their citation counts shouldn't be split.

This issue obviously affects the citation count of some documents. In Figure 6 we can observe how this phenomenon affects a book chapter: "Measuring science", by Van Raan.

*Figure 6. Versions of book chapters improperly tied in Google Scholar*

The screenshot shows a list of search results for the query "Measuring science". The results are as follows:

- Measuring science**  
AFJ Van Raan - Handbook of quantitative science and technology ..., 2005 - Springer  
Abstract After a review of developments in the quantitative study of science, particularly since the early 1970s, I focus on two current main lines of 'measuring science' based on bibliometric analysis. With the developments in the Leiden group as an example of daily ...  
Cited by 299 Related articles All 8 versions Cite Save
- Measuring science**  
A Raan - Handbook of Quantitative Science and Technology ..., 2005 - Springer  
After a review of developments in the quantitative study of science, particularly since the early 1970s, I focus on two current main lines of 'measuring science' based on bibliometric analysis. With the developments in the Leiden group as an example of daily practice, the ...  
Cited by 19 Related articles Cite Save
- [CITATION] **Measuring Science**. Handbook of Quantitative Science and Technology Research  
A van Raan - 2004 - HF Moed, W. Glänzel, U. Schmoch, ...  
Cited by 10 Related articles Cite Save
- [CITATION] **Measuring science** In Moed H, Glänzel W, & Schmoch U  
AFJ Van Raan - Handbook of Quantitative Science and Technology ..., 2004  
Cited by 4 Related articles Cite Save
- [CITATION] J (2004), " **Measuring Science**,"  
AF Van Raan - Handbook of Quantitative Science and Technology ...  
Cited by 3 Related articles Cite Save
- measuring Science**  
AFJ Van Raan - Handbook of Quantitative Science and ..., 2006 - books.google.com  
After a review of developments in the quantitative study of science, particularly since the early 1970s, I focus on two current main lines of 'measuring science' based on bibliometric analysis. With the developments in the Leiden group as an example of daily practice, the ...  
Cited by 8 Related articles All 4 versions Cite Save

f) Grouping different editions of the same book in a single record

Conversely to the previous error, Google Scholar sometimes groups together records that should stay separate, for example in the cases when there are different editions of the same book (a new book edition provides new content, contrary to a reprinting of a book, which is identical to the previous printing).

Figure 7 illustrates the case of "Little Science, big Science", written by Price. This book was first published in 1963 by Columbia University Press, and reedited in 1986 under the title "Little science, big science... and beyond", an edition that contained the original text of the book, as well as seven of his most famous articles.

Figure 7. Different book editions tied in Google Scholar

**[BOOK] Little science, big science... and beyond**  
DJ de Solta Price, DJ de Solta Price, DJ de Solta Price... - 1986 - garfield.library.upenn.edu  
On its first appearance, this book crystallized a new element in the historiography and sociology of science. It did so in the course of examining the major transformation in the structure of science prefigured in its title: from little to big science. As is often the case with ...  
Cited by 4130 Related articles All 5 versions Cite Saved More

**[CITATION] Little Science, Big Science... and Beyond**  
DDS Price - 1963 - citeulike.org  
Search all the public and authenticated articles in CiteULike. Include unauthenticated results too (may include "spam") Enter a search phrase. You can also specify a CiteULike article id (123456)... a DOI (doi:10.1234/12345678), or a PubMed ID (pmid:12345678). ...  
Cite

**[CITATION] Little Science, Big Science--And Beyond**  
DJDS Price - 1986 - philpapers.org  
Sign in | Create an account. PhilPapers PhilEvents PhilJobs. PhilPapers home. philosophical research online. Entries: 1,733,955 New this week: 655. General search Category finder. syntax | advanced search. Type words to match in category names. ...  
Cite

**[CITATION] Little Science, Big Science**  
DJS Price - 1963 - citeulike.org  
Search all the public and authenticated articles in CiteULike. Include unauthenticated results too (may include "spam") Enter a search phrase. You can also specify a CiteULike article id (123456)... a DOI (doi:10.1234/12345678), or a PubMed ID (pmid:12345678). ...  
Cite

**[CITATION] Little Science, Big Science**  
DJDS Price - 1963 - philpapers.org  
Sign in | Create an account. PhilPapers PhilEvents PhilJobs. PhilPapers home. philosophical research online. Entries: 1,663,534 New this week: 758. General search Category finder. syntax | advanced search. Type words to match in category names. ...  
Cite

The primary version (which has received 4,130 citations) is the edition from 1986, but among its versions are several records pointing to the version from 1963. Different editions of the same book should be treated as separate documents when computing citations because their content may be very different. Of course, automatically detecting and managing these details is a very complex task, and only a very tiny fraction of the documents indexed in Google Scholar (the most influential manuals and seminal works) would benefit from this thorough treatment. We must not forget that Google Scholar is, first of all, a search tool devoted to helping researchers find academic information. A great percentage of users probably don't care about the different editions of a book, and those who do probably just want the most recent one. That may be the reason why Google Scholar usually displays the most recent edition of a book as the primary version. The use of separate entries for different editions is something just a few people, like librarians, would be interested in. In any case, this may have an important effect in citation counts because citations to different editions (providing different content) are added together. In Figure 8 we can see how the 1986 edition of the book is receiving citations that were actually made to the original work published in 1963.

Figure 8. Citations to different book editions tied in Google Scholar

Little science, big science... and beyond  
 Search within citing articles

[PDF] The Matthew effect in science [PDF] from unc.edu  
 RK Merton - Science, 1968 - unc.edu  
 This paper develops a conception of ways in which certain psychosocial processes affect the allocation of rewards to scientists for their contributions—an allocation which in turn affects the flow of ideas and findings through the communication networks of science. The ...  
 Cited by 4274 Related articles All 14 versions Cite Save

[BOOK] The higher education system: Academic organization in cross-national perspective [HTML] from google.com  
 BR Clark - 1986 - books.google.com  
 How can we compare national systems of higher education, since their organization varies from country to country? Clark identifies the basic elements common to all such systems, and proceeds to thematic comparisons among a number of countries.  
 Cited by 3349 Related articles All 5 versions Cite Save More

Networks of scientific papers  
 P Yu, H Van de Sompel - Science, 1965 - eigenfactor.org  
 Borrowing methods from network theory, eigenfactor.org ranks the influence of journals much as Google's PageRank algorithm ranks the influence of web pages [2]. By this approach, journals are considered to be influential if they are cited often by other influential journals. Iterative ...  
 Cited by 2204 Related articles All 10 versions Cite Save More

[CITATION] Evaluation as illumination: A new approach to the study of inovatory programs  
 MR Parlett, D Hamilton - 1972 - ... of Edinburgh, Centre for Research in ...  
 Cited by 1476 Related articles Cite Save

[PDF] A General Theory of Bibliometric and Other Cumulative Advantage Processes [PDF] from asis.org  
 CAD CAD - Journal of the American society for Information science, 1976 - asis.org  
 It is common in bibliometric matters and in many diverse social phenomena, that success seems to breed success. A paper which has been cited many times is more likely to be cited again than one which has been little cited. An author of many papers is more likely to ...  
 Cited by 1180 Related articles All 10 versions Cite Save

Relevance: A review of and a framework for the thinking on the notion in information science [PDF] from rutgers.edu  
 I Saracevic - Journal of the American Society for Information ... , 1975 - Wiley Online Library  
 Abstract Information science emerged as the third subject, along with logic and philosophy, to deal with relevance—an elusive, human notion. The concern with relevance, as a key notion in information science, is traced to the problems of scientific communication. ...  
 Cited by 802 Related articles All 8 versions Cite Save

g) Improper attribution of citations to a document

Document citation counts in Google Scholar are also affected by the attribution of “ghost” citations to documents, that is, citations that aren’t actually there when we examine the citing document. Figure 9 shows an example of this issue: the work “Le transfert de l’information scientifique et technique: le rôle des nouvelles technologies de l’information face à la crise du modèle actuel de communication écrite” has allegedly received eight citations, but if we manually examine the second document in the list (marked in red), we can’t find any mention of the cited work. This phenomenon has been frequently observed in documents stored in the E-LIS repository.

Figure 9. Appearance of false citations

Le transfert de l'information scientifique et technique: le rôle des nouvelles technologies...  
 Search within citing articles

Políticas editoriales en la documentación española e implantación de E-LIS (E-Prints in Library and Information Science) [PDF] from rclis.org  
 F Páez, I Subirats-Coll, JM Barriocano, N Noverges - 2003 - eprints.rclis.org  
 The communication's objective is to establish a network of collaboration between the authors of scientific works of information science, through the implantation of a computer science tool. Eprints. It raises the problematic of the scientific communication in electronic ...  
 Cited by 8 Related articles All 3 versions Cite Save

A tiresome exercise, or a first step in library leadership? **Incorrect citation** [PDF] from rclis.org  
 JM Johnson - 2010 - eprints.rclis.org  
 This paper reviews the challenges facing librarians and other information specialists in terms of employers' expectations, and how the necessary skills could be developed in professional education. It focuses on the role of writing in professional education and in ...  
 Cited by 1 Related articles All 3 versions Cite Save

Open Access: la información científica al alcance de la sociedad [PDF] from rclis.org  
 L Barriocano - 2009 - eprints.rclis.org  
 The concept of scholarly information and the importance of Open Access is analysed along with the importance that initiative is giving in the academic-scientific field and in society. The model of scholarly communication is studied, as well as, the concerns and problems that ...  
 Related articles Cite Save

[PDF] Open Access: la información científica al alcance de la sociedad [PDF] from core.ac.uk  
 LB Almuzaraa - ¿Qué es información?: Actas del primer Encuentro ... , 2008 - core.ac.uk  
 Resumen. Se analiza el concepto de información científica y la importancia que la Iniciativa Open Access o de Acceso abierto, está teniendo en el ámbito académicocientífico y en la sociedad en general. Se estudia el modelo tradicional de comunicación científica, a la vez ...  
 Related articles All 2 versions Cite Save More

[PDF] Red española de trabajos científicos. Estudio de viabilidad de la implantación de una biblioteca digital y análisis de sus derechos de [PDF] from researchgate.net  
 MFP Mancoabo, JMB Cruz, IS Coll, NN Doménech - - researchgate.net  
 Actualmente existen múltiples maneras formales—publicaciones periódicas, actas de congresos y reuniones—e informales—listas de distribución o redes personales—de comunicación entre los creadores de información STM.  
 Related articles Cite Save More

[HTML] Information Architecture= Enformasyon Mimarisi **Incorrect citation** [HTML] from rclis.org  
 A Gilchrist - Bilgi Dünyası, 2004 - eprints.rclis.org  
 This report is the result of a funded project of the Spanish Ministry of Education and Culture given in 2002 to the Department of Audiovisual Communication, Information Science and History of Art (Polytechnic University of Valencia-Spain). The aim of the project was ...  
 Related articles All 2 versions Cite Save More

## h) Duplicate citations

This phenomenon is a consequence of an issue previously discussed. When Google Scholar fails to realise that two records are actually versions of the same document, these versions are stored as if they were different documents. Therefore, each of them provides its own set of citations to the citation pool. Since the two sets of citations are probably identical, each cited document will receive two citations from what is actually only one document, thus falsely inflating their citation counts. In Figure 10 we observe a double example of this phenomenon. In the first case (first red rectangle), there are three versions of the same document. Note the differences in the way the authors' names are stored, since this is probably the reason why the records weren't merged into one. In the second case (second red rectangle), the two records refer to the same document (the first one is the English version of the article, and the second one is the Spanish version).

Figure 10. Duplicate citations in Google Scholar

H Index Communication Journals according to Google Scholar Metrics (2008-2012)

Search within citing articles

[Perfiles de comunicación en Google Scholar Metrics, índice h y nuevas estrategias de difusión de la investigación](#) [PDF] from ucm.es  
MT López - Historia y Comunicación Social, 2014 - revistas.ucm.es  
Resumen Google Scholar Metrics permite, desde 2012, que los investigadores creen un perfil con su producción científica y académica enlazada en la red y evaluada con indicadores bibliométricos ih, i10 e ih5 personalizados, que referencian la correlación ...  
Cited by 2 Related articles All 3 versions Cite Save

[in Organizational Communication: contexts, paradigms and comprehensive conceptualization](#) [PDF] from usp.br  
E Honig, SM Brown - revistas.usp.br  
conceptualization exemplifies how the theoretical and conceptual support is the generator of renovating looks on objects and known realities. Simple dichotomous vision of the media and the mediatized world as a place of commercialization of culture or further of being ...  
Related articles All 4 versions Cite Save More

[LA INVESTIGACIÓN EN COMUNICACIÓN Y PERIODISMO AMBIENTAL EN ESPAÑA: Estado de la cuestión y revisión bibliométrica de las principales revistas ...](#)  
AB Camero, BM Garoia - Revista Prisma Social, 2014 - search.elsevier.com  
RESUMEN El siguiente estudio propone un acercamiento exploratorio a la producción académica española en el campo de la comunicación y el periodismo ambiental con el objeto de delimitar cuáles son sus principales rasgos, potencialidades y limitaciones. A ...  
Related articles Cite Save

[La investigación en comunicación y periodismo ambiental en España: Estado de la cuestión y revisión bibliométrica de las principales revistas académicas en ...](#) [PDF] from uc3m.es  
A Baranquero, B Marín Garoia - 2014 - e-archivo.uc3m.es  
El siguiente estudio propone un acercamiento exploratorio a la producción académica española en el campo de la comunicación y el periodismo ambiental con el objeto de delimitar cuáles son sus principales rasgos, potencialidades y limitaciones. A partir de ...  
Cited by 2 Related articles All 2 versions Cite Save

[La investigación en comunicación y periodismo ambiental en España: Estado de la cuestión y revisión bibliométrica de las principales revistas académicas en ...](#) [PDF] from ucm.es  
A Baranquero Carretero, B Marín Garoia - Prisma Social, 2014 - eprints.ucm.es  
El siguiente estudio propone un acercamiento exploratorio a la producción académica española en el campo de la comunicación y el periodismo ambiental con el objeto de delimitar cuáles son sus principales rasgos, potencialidades y limitaciones. A partir de ...  
Related articles Cite Save

[Analysis of the productivity, impact, and collective h-index of the communication research carried out in Spain based on the information shared by researchers ...](#) [HTML] from revistalatinacs.org  
MT López - 2014 - revistalatinacs.org  
Abstract Introduction. This article examines the productivity, collective and individual h and h5 indexes, dissemination platforms (mainly books and journals), and visibility of the communication research carried out by Spanish scholars in the last four decades, based ...  
Related articles All 4 versions Cite Save More

[Análisis de productividad, impacto e índice h de la investigación en Comunicación a través de los perfiles personales en Google Académico](#) [HTML] from revistalatinacs.org  
MT López - 2014 - revistalatinacs.org

## i) Missing citations

There are cases when Google Scholar's parser fails to match a cited reference inside document, with the record of the document it is citing. When Google Scholar parses the reference section within an article, it tries to find a match for these references in its records, but if for some reason the reference hasn't been correctly recorded (authors of the citing article may have made a mistake when citing it or used an uncommon reference format Google Scholar doesn't understand) the system will be unable to make the connection between the two documents. However, we also find examples in which no apparent mistake has been made in the citing document, but still the citation isn't attributed to the cited document. In order to illustrate this issue, in Figure 11 we show how a document ("How to cook the university rankings") is citing in its reference section other document (a doctoral thesis). However, this citation doesn't appear as one of the 13 citations that the thesis has received according to Google Scholar. The reason is unknown. At the time the citing

document was first indexed, the connection wasn't made for some reason, and this error hasn't been solved since. Typos in the PDF can also generate this kind of error.

Figure 11. Citations unrevealed in Google Scholar

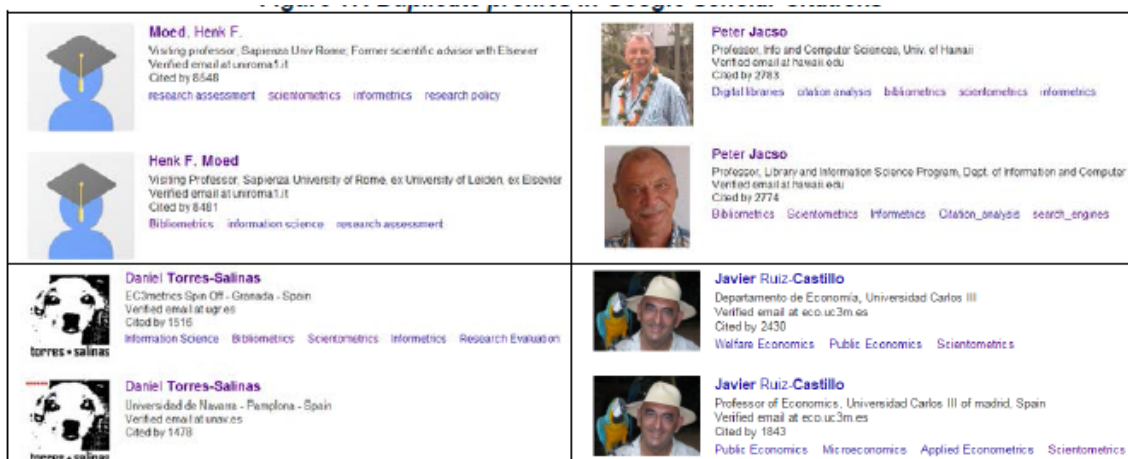


## 2. Mistakes identified in the elaboration of bibliographic profiles

All the errors previously described are related directly with the Google Scholar database (and are concerned with how the automatic parser works). Next we show some of the mistakes identified in the elaboration of bibliographic profiles through Google Scholar Citations:

a) Duplicate profiles Since the only restriction to create a public academic profile in Google Scholar Citations is to provide a valid email, an author (or anyone really) may create as many profiles as he/she wants. This opens the door to the existence of duplicate profiles, that is, different profiles about the same person. In Figure 12 we present some examples of duplicate profiles of authors related to the field of Bibliometrics. The differences in citation counts between profiles are sometimes quite high (for example, one of the profiles belonging to Ruiz-Castillo achieves 1,843 citations whereas in the second profile the figure goes up to 2,430).

Figure 12. Duplicate profiles in Google Scholar Citations



A real problem can arise when one of the profiles has been created by someone other than the author the profile is about. The author may send a request to Google Scholar to delete the profile, but this kind of requests might take a while to be processed, generating a feeling of helplessness in the author.

b) Variety of document types (including non-academic documents), one of the main criticisms to the profiles in Google Scholar Citations (when considering whether they're suited for evaluation purposes) is the inclusion of a wide variety of document types: from peer-reviewed articles to posters. An author can add any kind of work to his profile, and sometimes they aren't



even academic works: teaching materials, software, online resources, etc. (Figure 13). While this is a true shortcoming from the research evaluation perspective, these profiles are designed to showcase any material that the author considers appropriate, especially if these materials could potentially generate some kind of impact through citations. The possibility to select the document typology (as ResearchGate does) may help solve this problem. However, the selection of document type is only an internal mechanism not reflected in the public profile.

**Figure 13. Teaching materials in Google Scholar Citations**

<p><b>1. MONOGRÁFICO: RIRLIOMETRÍA</b>          ES <b>Casado</b>, CM Moreno, G Portal, ES Quesada</p>
<p><b>Nº CRÉDITOS/ECTS: 3</b>          ES <b>CASADO</b>, CQUEELED ADQUIRIR</p>
<p><b>What does publication output tell us: indicators from the institutional databases of higher education institutions</b>          E Reale, D De Filippo, I Gomez, B Lepori, B Poti, E Primeri, C Probst, ...</p>
<p><b>Diagnóstico sobre la docencia e investigación en Ciencias de la Documentación en España</b>          ES <b>Casado</b>, MLL Sánchez</p>
<p><b>CUATRIMESTRE: 2º CRÉDITOS: 3</b>          CYP DE TEORÍA, ES <b>Casado</b>, P DE PRÁCTICAS, G Ballenato</p>

c) Inclusion of missattributed documents in the profile

The Google Scholar team doesn't oversee the validity of all the information available in Google Scholar Citations. Therefore, it is the sole responsibility of the author that the information visible in his/her profile is accurate. Profiles can be set to be updated automatically (when the system finds an article that it's reasonably sure it's yours, it is automatically added to your profile), or by asking the author for confirmation first when the system thinks an addition or a change should be made. If the user selects the automatic updates, there is a risk that the system will add documents to the profile that the author hasn't actually written, thus falsely increasing the author's bibliometric indicators. The author will probably be completely oblivious to this issue if he or she doesn't check the profile regularly. If that is the case, it shouldn't be considered an active attempt to fake one's bibliometric indicators, but it is still a matter that should be fixed as soon as it comes to the author's knowledge. In Figure 14 we can see an example: the third document (marked in red), which has received 40 citations, hasn't been written by the owner of the profile (Imma Subirats-Coll).

**Figure 14. Misattributed documents in Google Scholar Citations**

Title 1-20	Cited by	Year
<p><b>Open archives initiative. Protocol for metadata harvesting (OAI-PMH): descripción, funciones y aplicaciones de un protocolo</b>            IS Coll, JMB Cruz            El profesional de la información 12 (2), 99-106</p>	79	2003
<p><b>Open data y Linked open data: su impacto en el área de bibliotecas y documentación</b>            MF Peset Mancebo, A Ferrer Sapena, I Subirats-Coll            El profesional de la información 20 (2), 165-173</p>	48	2011
<p><b>Sistemas integrales para la automatización de bibliotecas basados en software libre</b>            O Arriola Navarrete, K Butrón Yáñez            Acimed 18 (6), 0-0</p>	40	2008
<p><b>E-LIS: an international open archive towards building open digital libraries</b>            A De Robbio, I Subirats Coll            High energy physics libraries webzine 11, 2005</p>	29	2005

We can find examples where the owner of the profile has participated as a translator or editor of a work (Figure 25). The assignment of the citation counts of a work to the people who have

fulfilled this kind of roles is controversial. At the very least, they should make sure that their role is clearly stated and visible in the profile.

Figure 15. Edition and translation roles in Google Scholar Citations



d) Deliberate manipulation of documents and citations in Google Scholar

Another issue is that of the conscious manipulation of profiles by their owners. The fact that anyone, without advanced technical skills, can manipulate his/her own bibliometric indicators, or other people's (Delgado López-Cózar, Robinson-García & Torres-Salinas, 2014) may affect the credibility of GSC academic profiles if no action to control this issue is taken by the Google Scholar team. In Figure 16 we observe how uploading a set of fake documents to a repository (with nonsensical text, and a list of references which include the set of documents whose impact you want to boost) will, in just a few days, cause the desired adulteration of citation scores in the profiles of the authors of the referenced documents.

Figure 16. Effect of data manipulation in Google Scholar Citations



e) Duplicate documents in profiles

This is also a side effect of the cases when Google Scholar fails to group together different versions of the same document. The consequence for the profiles is that the different versions will also be added as different records in the profile, which might affect (positively or negatively) indicators like the h-index and the i-index, which are computed automatically. Fortunately, profile users can manually merge records in their profile, which will solve this issue (Figure 17). This merge only affects the author's profile. It doesn't alter Google Scholar search query results in any way, that is, there will still be two (or more) records for that document in Google Scholar's index, at least until the error gets fixed in a future update.

**Figure 17. Versions not tied in Google Scholar Citations**

<a href="#">Measuring science</a> AFJ Van Raan Handbook of quantitative science and technology research, 19-50	299	2005
<a href="#">Measuring science</a> A Raan Handbook of Quantitative Science and Technology Research, 19-50	19	2005
<a href="#">measuring Science</a> AFJ Van Raan Handbook of Quantitative Science and Technology Research: The Use of ...	8	2006

f) Unclean document titles

This error is also inherited from Google Scholar's metadata parsing errors. Google Scholar Citations allows authors to modify almost all aspects of a record in their profile, including the title of the documents. Unfortunately, not all authors pay attention to such details, and so these errors persist (Figure 18).

**Figure 18. Parse errors in identifying document titles in Google Scholar Citations**

<a href="#">The problems with the subject categories schema in the EigenFactor database from the perspective of ranking journals by their prestige and impact (pre-print version). In print: Online Information Review, 36 (5), 2012, p. 758-766.</a> P Jacso
<a href="#">THE CONTEXT Usually, I review the progress and regress of Google Scholar once a year. This year is an exception simply because Google Scholar developers have introduced two new, polished services in the past 9 months, Google Scholar Author Citation Tracker in 2011 (Jacso, 2012a) and Google Scholar Metrics for Publications in April, 2012. They represent much needed progress for considering Google Scholar for bibliometric purposes, paying attention to quality not just quantity.</a> P Jacso
<a href="#">Open Access Scholarly Databases—A Bird's Eye View of the Landscape Keynote address of the Asia &amp; Oceania Section of</a> P Jacso

g) Missing or uncommon areas of interest

One last limitation that may affect the results of this Working Paper is related to the areas of interest declared by the authors in their profiles (a maximum of five areas can be provided). Researchers in bibliometrics with a public profile in Google Scholar Citations, but haven't declared any area of interest (Figure 19, top), those who use uncommon keywords, or keywords in a language other than English may have been overlooked.

**Figure 19. Missing (top) and uncommon (bottom) areas of interest in Google Scholar Citations**



NOTA: este texto es una adaptación de:

- Martin-Martin, A., Orduna-Malea, E., Ayllon, J. M., & Lopez-Cozar, E. D. (2016). The counting house: measuring those who count. Presence of Bibliometrics, Scientometrics, Informetrics, Webometrics and Altmetrics in the Google Scholar Citations, ResearcherID, ResearchGate, Mendeley & Twitter. arXiv preprint arXiv:1602.02412.

Unidad de Bibliometría

Vicerrectorado de Investigación, Desarrollo e Innovación

Teléfono: 928 451 030

Fax: 928 459 699

Correo: [vidi\\_bibliometria@ulpgc.es](mailto:vidi_bibliometria@ulpgc.es)

Ubicación: Sede Institucional