

Errors in DOI indexing by bibliometric
databases

Original

Errors in DOI indexing by bibliometric
databases / Franceschini, Fiorenzo; Maisano, DOMENICO AUGUSTO FRANCESCO; Mastrogiacomo, Luca. - In:
SCIENTOMETRICS. - ISSN 0138-9130. - STAMPA. - 102:3(2015), pp. 2181-2186. [10.1007/s11192-014-1503-4]

Availability:

This version is available at: 11583/2591592 since:

Publisher:

Springer

Published

DOI:10.1007/s11192-014-1503-4

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in
the repository

Publisher copyright

(Article begins on next page)

Errors in DOI indexing by bibliometric databases

Fiorenzo Franceschini¹, Domenico Maisano, Luca Mastrogiacomo

¹*fiorenzo.franceschini@polito.it*

Politecnico di Torino, DIGEP (Department of Management and Production Engineering),
Corso Duca degli Abruzzi 24, 10129, Torino (Italy)

Abstract

DOI – i.e., Digital Object Identifier – is a character string, which univocally identifies entities that are object of intellectual property. In bibliometrics, DOIs are used for univocally identifying scientific papers. The aim of this short communication is to raise the reader's awareness of bibliometric database errors in DOI indexing, in particular, the incorrect assignment of a single DOI to multiple papers. This error is quite interesting since DOI is commonly regarded as an effective means to identify scientific articles unambiguously. For the purpose of example, a short list of DOIs, which have been wrongly assigned by the Scopus database to multiple papers, is shown. Although being relatively rare, DOI indexing errors should be considered by bibliometricians when querying bibliometric databases by DOI.

Keywords: Database errors, Citation Analysis, DOI indexing error, Multiple DOI assignment, Scopus.

Introduction

Bibliometric databases – e.g., Scopus and Web of Science (WoS) – are affected by different types of errors, which may impact on citation statistics. Over the past decade, the accuracy of bibliometric databases has improved significantly, probably due to the methodical use of (semi)automatic tools for correcting errors in cited article lists by editors and database administrators (Adam 2002). However, the problem is far from being solved, as proven by the fact that bibliometric database staff constantly encourage users to report any noticed inaccuracy.

Several recent articles document the existence of database errors of different nature (Jacso 2006; Jacso 2012; Franceschini, Maisano et al. 2013; Franceschini, Maisano et al. 2014); a synthetic taxonomy of the major errors is reported in Tab.1, distinguishing between author and database mapping errors (Buchanan 2006).

While author errors are generally due to lack of care of authors when creating the cited article lists, database mapping errors may originate from poor or imperfect communication between database administrators (i.e., data receivers) and editors (i.e., data providers). In fact, the administrators' requirement to quickly update bibliometric databases, whenever new data are available from editors, may sometimes conflict with indexing accuracy.

Tab.1. Classification of bibliometric database errors according to (Buchanan 2006).

Error type	Author errors	Database mapping errors
Definition	Errors made by authors when creating the list of cited articles for their publication.	Failure to establish an electronic link between a cited article and the corresponding citing articles that can be attributed to a data-entry error.
Examples	<ul style="list-style-type: none"> - Errors in name and initials of the first author; - Errors in publication title; - Errors in publication year; - Errors in volume number; - Errors in pagination. 	<ul style="list-style-type: none"> - Transcription errors; - Target-source article record errors; - Cited article omitted from a cited-article list; - Reason unknown.

Multiple DOI assignment

In a relatively recent paper (Franceschini, Maisano et al. 2014), we analyzed a set of journals indexed by Scopus and WoS, in order to estimate the rate of omitted citations, i.e., missing links between a cited paper and the corresponding citing papers. Precisely, we examined the citations indexed by Scopus and WoS, focusing on those received from journals purportedly indexed by both databases, and we showed that the citations obtained from some publishers are more likely to be omitted than those from other ones, for several reasons. The dataset included more than 27,000 articles of interest, published by 34 different journals in the field of Manufacturing Engineering field, with more than 112,000 citing papers issued in the time-window from 2006 to 2012.

A curious outcome of the analysis, which has not been explained in our previous paper (Franceschini, Maisano et al. 2014), is the fact that bibliometric databases may sometimes mistakenly assign the same DOI to multiple papers. We recall that DOI (i.e., Digital Object Identifier) is a character string used to univocally identify entities that are object of intellectual property (Paskin 2002). Each DOI is by definition unique and must identify one and only one entity. Since several years, DOIs are used in bibliometrics for identifying and disambiguating scientific papers. Obviously, errors in the DOI character string could distort the results of bibliometric analyses, in the case DOI is used for querying bibliometric databases.

Returning to the previous discussion, Tab. 2 shows an example of multiple DOI assignments by the Scopus database, in which each of the eight DOIs reported (i.e., i to viii) is mistakenly assigned to two or more different articles. For the purpose of example, Fig. 1 shows the screenshot of the Scopus database response, relating to the first DOI.

It is worth remarking that the errors in Tab.2 represent only a subset of those found in our analysis. Also, these errors have been detected few weeks before the submission of this short communication and they might be corrected by the database in the next future.

Tab.2 List of DOIs mistakenly assigned to multiple articles by the Scopus database. Symbols “✓” and “✗” respectively denote the correct and incorrect DOI assignment. It is worth noting that the last DOI (viii), which is probably incomplete, is mistakenly assigned to the totality of the nine papers returned by Scopus.

DOI	Corresponding articles	
i) 10.1016/j.jmatprotec.2005.11.008	1) Costa, P. J. M. S., R. Q. B. Vilela, R. Cipolotti and M. S. Figueiredo (2006). "Clinical and laboratorial diversity in the bantu haplotype of sickle cell anemia." <i>Revista Brasileira de Hematologia e Hemoterapia</i> 28(1): 40-44.	✗
	2) Costa, T. H. C., M. C. Feitor, C. Alves Jr, P. B. Freire and C. M. De Bezerra (2006). "Effects of gas composition during plasma modification of polyester fabrics." <i>Journal of Materials Processing Technology</i> 173(1): 40-43.	✓
ii) 10.1016/j.jmatprotec.2005.11.033	1) Pan, W. (2006). "Incorporating gene functional annotations in detecting differential gene expression." <i>Journal of the Royal Statistical Society. Series C: Applied Statistics</i> 55(3): 301-316.	✗
	2) Pan, Z., H. Zhang, Z. Zhu and J. Wang (2006). "Chatter analysis of robotic machining process." <i>Journal of Materials Processing Technology</i> 173(3): 301-309.	✓
iii) 10.1016/j.jmatprotec.2005.11.037	1) Bao, G., K. Shinozaki, S. Iguro, M. Inkyo, M. Yamamoto, Y. Mahara and H. Watanabe (2006). "Stress corrosion cracking sealing in overlaying of Inconel 182 by laser surface melting." <i>Journal of Materials Processing Technology</i> 173(3): 330-336.	✓
	2) Bao, L. and M. C. Trachtenberg (2006). "Facilitated transport of CO ₂ across a liquid membrane: Comparing enzyme, amine, and alkaline." <i>Journal of Membrane Science</i> 280(1-2): 330-334.	✗
iv) 10.1007/s00170-004-1884-2	1) Hu, J., L. Xiao, Y. Wang and Z. Wu (2006). "An optimal feedrate model and solution algorithm for a high-speed machine of small line blocks with look-ahead." <i>International Journal of Advanced Manufacturing Technology</i> 28(9): 930-935.	✓
	2) Hu, M. C. (2006). "Erratum: Steroid deficiency syndromes in mice with targeted disruption of Cyp11a1 (<i>Molecular Endocrinology</i> (2000) 16 (1943-1950))." <i>Molecular Endocrinology</i> 20(4): 930.	✗
v) 10.1007/s00170-004-2324-z	1) Wu, D. (2006). "Detecting information technology impact on firm performance using DEA and decision tree." <i>International Journal of Information Technology and Management</i> 5(2-3): 162-174.	✗
	2) Wu, X., J. Chen, R. Li and F. Li (2006). "Web-based remote monitoring and fault diagnosis system." <i>International Journal of Advanced Manufacturing Technology</i> 28(1-2): 162-175.	✓
vi) 10.1016/j.ijmachtools.2005.08.006	1) Lee, C. Z., J. S. Yao, Y. Huang, W. Zhai, W. Liu, B. J. Guglielmo, E. Lin, G. Y. Yang and W. L. Young (2006). "Dose-response effect of tetracyclines on cerebral matrix metalloproteinase-9 after vascular endothelial growth factor hyperstimulation." <i>Journal of Cerebral Blood Flow and Metabolism</i> 26(9): 1157-1164.	✗
	2) Lee, R. T., Y. C. Hwang and Y. C. Chiou (2006). "Lapping of ultra-precision ball surfaces. Part II. Eccentric V-groove lapping system." <i>International Journal of Machine Tools and Manufacture</i> 46(10): 1157-1169.	✓
vii) 10.1016/j.rcim.2005.02.007	1) Zhao, J. S., Y. Z. Fu, K. Zhou and Z. J. Feng (2006). "Mobility properties of a Schoenflies-type parallel manipulator." <i>Robotics and Computer-Integrated Manufacturing</i> 22(2): 124-133.	✓
	2) Zhao, Y., X. Luo, Y. Ming, L. Chen and Y. Li (2006). "Rapid separation and determination of resibufogenin and cinobufagin in toad venom and Liushen tablet by β -cyclodextrin modified micellar electrokinetic chromatography." <i>Journal of Pharmaceutical and Biomedical Analysis</i> 41(1): 124-128.	✗
viii) 10.3401/poms	1) Aksin, Z., M. Armony and V. Mehrotra (2007). "The modern call center: A multi-disciplinary perspective on operations management research." <i>Production and Operations Management</i> 16(6): 665-688.	✗
	2) Bardhan, I., S. Mithas and S. Lin (2007). "Performance impacts of strategy, information technology applications, and business process outsourcing in U.S. manufacturing plants." <i>Production and Operations Management</i> 16(6): 747-762.	✗
	3) Gupta, D. (2007). "Surgical suites' operations management." <i>Production and Operations Management</i> 16(6): 689-700.	✗
	4) Queenan, C. C., M. Ferguson, J. Higbie and R. Kapoor (2007). "A comparison of unconstraining methods to improve revenue management systems." <i>Production and Operations Management</i> 16(6): 729-746.	✗
	5) Shanthikumar, J. G. and D. D. Yao (2007). "John A. Buzacott and his pioneering contributions to manufacturing and service systems." <i>Production and Operations Management</i> 16(6): 657-664.	✗
	6) Shen, Z. J. M. and X. Su (2007). "Customer behavior modeling in revenue management and auctions: A review and New Research Opportunities." <i>Production and Operations Management</i> 16(6): 713-728.	✗
	7) Smith, J. S., K. R. Karwan and R. E. Markland (2007). "A note on the growth of research in service operations management." <i>Production and Operations Management</i> 16(6): 780-790.	✗
	8) Subramanian, R., S. Gupta and B. Talbot (2007). "Compliance strategies under permits for emissions." <i>Production and Operations Management</i> 16(6): 763-779.	✗
	9) Xia, C. H. and P. Dube (2007). "Dynamic pricing in e-services under Demand Uncertainty." <i>Production and Operations Management</i> 16(6): 701-712.	✗

The screenshot shows the Scopus search results for the DOI 10.1016/j.jmatprotec.2005.11.008. The search results are displayed in a table with columns for title, authors, year, journal, and citations. The first result is 'Diversidade clinica e laboratorial no haplotipo bantu da anemia falciforme I [Clinical and laboratorial diversity in the bantu haplotype of sickle cell anemia]' by Costa, P.J.M.S., Vilela, R.O.B., Cipolotti, R., and Figueiredo, M.S., published in 2006 in 'Revista Brasileira de Hematologia e Hemoterapia' with 0 citations. A red 'X' is placed over this result. The second result is 'Effects of gas composition during plasma modification of polyester fabrics' by Costa, T.H.C., Feitor, M.C., Alves Jr., C., Froire, P.B., and De Bezerra, C.M., published in 2006 in 'Journal of Materials Processing Technology' with 30 citations. A green checkmark is placed over this result. The interface includes search filters for Year and Author Name, and options to refine, export, or download results.

Fig. 1. Example of Scopus query, returning two papers mistakenly associated with the same DOI (in the top left-hand side).

Final Remarks

This short communication reported the existence of bibliometric database errors concerning DOI indexing. These errors are quite interesting since the DOI to a scientific paper is like the “ID card” to a person and is commonly regarded as an effective means of disambiguation. The most surprising manifestation of these errors is represented by multiple DOI assignment. Although probably being rather unusual (according to a preliminary estimation, the error rate in the Scopus database is of the order of magnitude of 1/1000), the fact remains that this error type could potentially influence the results of bibliometric analyses.

Future research will go into the multiple DOI assignment issue in more detail, investigating the practical reasons behind this error and providing a more accurate estimation of the error rate, for both Scopus and WoS.

References

- Adam, D. (2002). "The counting house." Nature **415**(6873): 726-729.
- Buchanan, R. A. (2006). "Accuracy of cited references: The role of citation databases." College and Research Libraries **67**(4): 292-303.
- Franceschini, F., D. Maisano and L. Mastrogiacomo (2013). "The effect of database dirty data on h-index calculation." Scientometrics **95**(3): 1179-1188.
- Franceschini, F., D. Maisano and L. Mastrogiacomo (2014). "Scientific journal publishers and omitted citations in bibliometric databases: Any relationship?" Journal of Informetrics **8**(3): 751-765.
- Jacso, P. (2006). "Deflated, inflated and phantom citation counts." Online Information Review **30**(3): 297-309.
- Jacso, P. (2012). "Grim tales about the impact factor and the h-index in the Web of Science and the Journal Citation Reports databases: reflections on Vanclay's criticism." Scientometrics **92**(2): 325-354.
- Paskin, N. (2002). The DOI® handbook, IDF-Intern. International DOI Foundation.