Can Google Scholar measure accurate the highly cited documents?[1]

Alberto Martín-Martín[1], Enrique Orduña-Malea[2], Juan Manuel Ayllón[1], Emilio Delgado López-Cózar[1]

[1] EC3: Evaluación de la Ciencia y de la Comunicación Científica, Universidad de Granada (Spain)
[2] EC3: Evaluación de la Ciencia y de la Comunicación Científica, Universidad Politécnica de Valencia (Spain)

This article, despite explores the most-cited research of all time using Web of Science, additionally provides an alternative ranking from Google Scholar (see Noorden, Maher & Nuzzo *Nature,* 9**,** 550-553; 2014)[1]. Although we know that the main focus of this paper is on the data extracted from Web of Science, we want to point out some discrepancies in this Google Scholar league table.

This ranking shows "Protein measurement with the Folin phenol reagent" as the second most cited article (192,710 citations), contradicting the Web of Science ranking, which historically shows it as the most cited by far over all other (see Garfield's citation classics).

Investigating on the presence of highly cited documents in Google Scholar (see http://arxiv.org/abs/1410.8464), we find certain inconsistencies (instability in the allocation of citations and identification and linkage of versions).

Lowry's et al article obtained, as of May 2014, 253,671 citations. How can an article lose 60,961 citations in 5 months? Conversely, "Diagnostic and statistical manual of mental disorders", not included in the top ten despite having 185,000 citations in Google Scholar (and almost 220,000 if we merge its various versions), receives the incredible number of 55,170 citations from May.

Otherwise, "Molecular cloning" appears duplicated, separating two different editions of the same work. Adding these and other unmerged records, the citations amount to 268,834, up to the first position. Likewise, we found 164 additional unmerged records to "A mathematical theory of communication", where citations to the article and the subsequent book are mixed.

Thus, to what extent can we trust on this ranking provided by Google Scholar?

Nonetheless, even dirty, Google Scholar is capable not only to identify the highly cited papers but to provide a complementary academic landscape. And this is the portrait that should prevail.

---

[1] This letter was rejected by Nature. Submitted 31-10-2014; Rejected 5-11-2014

Bibliography

Van Noorden, R., Maher, B., Nuzzo, R. (2014). The top 100 papers. Nature, 514(7524), 550-553.

Garfield, E. (2005). The Agony and the Ecstasy—The History and Meaning of the Journal Impact Factor. International Congress on Peer Review And Biomedical Publication. Chicago, September 16, 2005. http://www.garfield.library.upenn.edu/papers/jifchicago2005.pdf

Martín-Martín, A.; Orduña-Malea, E.; Ayllón, J.M.; Delgado López-Cózar, E. (2014). Does Google Scholar contain all highly cited documents (1950-2013)? Granada: EC3 Working Papers, 19: November 3, 2014. http://arxiv.org/pdf/1410.8464