Evidence of a Negativity Bias in the Keywords of Scientific Papers

Authors : Kseniia Zviagintseva, Brett Buttliere

Abstract : Science is fundamentally a problem-solving enterprise, and scientists pay more attention to the negative things, that cause them dissonance and negative affective state of uncertainty or contradiction. While this is agreed upon by philosophers of science, there are few empirical demonstrations. Here we examine the keywords from those papers published by PLoS in 2014 and show with several sentiment analyzers that negative keywords are studied more than positive keywords. Our dataset is the 927,406 keywords of 32,870 scientific articles in all fields published in 2014 by the journal PLOS ONE (collected from Altmetric.com). Counting how often the 47,415 unique keywords are used, we can examine whether those negative topics are studied more than positive. In order to find the sentiment of the keywords, we utilized two sentiment analysis tools, Hu and Liu (2004) and SentiStrength (2014). The results below are for Hu and Liu as these are the less convincing results. The average keyword was utilized 19.56 times, with half of the keywords being utilized only 1 time and the maximum number of uses being 18,589 times. The keywords identified as negative were utilized 37.39 times, on average, with the positive keywords being utilized 14.72 times and the neutral keywords - 19.29, on average. This difference is only marginally significant, with an F value of 2.82, with a p of .05, but one must keep in mind that more than half of the keywords are utilized only 1 time, artificially increasing the variance and driving the effect size down. To examine more closely, we looked at those top 25 most utilized keywords that have a sentiment. Among the top 25, there are only two positive words, 'care' and 'dynamics', in position numbers 5 and 13 respectively, with all the rest being identified as negative. 'Diseases' is the most studied keyword with 8,790 uses, with 'cancer' and 'infectious' being the second and fourth most utilized sentiment-laden keywords. The sentiment analysis is not perfect though, as the words 'diseases' and 'disease' are split by taking 1st and 3rd positions. Combining them, they remain as the most common sentiment-laden keyword, being utilized 13,236 times. More than just splitting the words, the sentiment analyzer logs 'regression' and 'rat' as negative, and these should probably be considered false positives. Despite these potential problems, the effect is apparent, as even the positive keywords like 'care' could or should be considered negative, since this word is most commonly utilized as a part of 'health care', 'critical care' or 'quality of care' and generally associated with how to improve it. All in all, the results suggest that negative concepts are studied more, also providing support for the notion that science is most generally a problem-solving enterprise. The results also provide evidence that negativity and contradiction are related to greater productivity and positive outcomes.

Keywords : bibliometrics, keywords analysis, negativity bias, positive and negative words, scientific papers, scientometrics
Conference Title : ICCSIB 2017 : International Conference on Cybermetrics, Scientometrics, Informetrics and Bibliometrics
Conference Location : Barcelona, Spain
Conference Dates : October 30-31, 2017