Analysis of PubMed reports on agriculture over the last 50 years

Jagajjit Sahu

National Center for Cell Sciences (NCCS), India

Statement of the Problem: In the age of exponentially growing scientific literature, it is becoming increasingly difficult for manual reviews to comprehend the research scenario and understand gaps in all areas. Cutting-edge text mining tools have allowed relevant information to be extracted from Big Literature Data, but a few fields such as agriculture are still underexplored.

Methodology & Theoretical Orientation: The citation data for five decades of Agriculture reports were downloaded from PubMed. Data were analyzed using scientometrics to identify trends and patterns in research based on various bibliometric indicators. Keyword mining was carried out in order to discover the frequent research domains. Custom R scripts along with PubTator tool was also employed to get a deeper understanding of various biological concepts and their associations.

Findings: The study concludes that Agriculture publications are rising on an overall basis, although the number of publications has increased significantly since 2000. A total of 96079 records were retrieved on PubMed, with the top keyword being economic factors. The keyword association network showed a tight bond for the cluster containing aquaculture, despite its ranking in the seventh position in terms of frequency. There were 82433 articles containing bioconcepts under 10 categories (such as Gene, Disease, CellLine, etc.) with infection disease and SCC as the top diseases. Further analysis of co-occurrence networks was carried out to investigate gene-disease associations.

Conclusion & Significance: To gain a better understanding of the agricultural research landscape, this study explores the productivity and patterns of research over the last 50 years. As part of my efforts to promote text mining in agriculture and related fields, I will also point out the limitations of these endeavors compared to biomedical literature mining. In conclusion, further studies of this kind will definitely lead to more research gaps and hidden information.
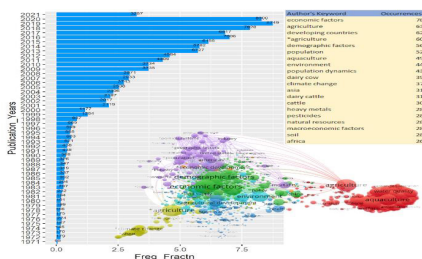

Figure 1: A summary of the information extracted from 50 years of PubMed Agriculture records. Graphs show the year-wise trend in the publication of articles, the top 20 frequent keywords for authors, and network visualization for 1000 popular keywords.

Biography

Jagajjit Sahu is currently working in the Manav - Human Atlas Initiative at the National Centre for Cell Science in Pune, India. His primary area of expertise is bioinformatics and the analysis of big data. He has received a PhD from the Assam University, India, and postdoctoral research experiences from the University of Aberdeen, UK. He has published 37 journal articles and a few book chapters and currently editing a book for CRC press. Dr. Sahu serves on the editorial boards of Archives of Phytopathology and Plant Protection and Current Genomics as well as serving as a reviewer for reputed journals such as Briefings in Bioinformatics, Genes and Sensors. Furthermore, he has been invited by several institutes as a speaker, a mentor, and a resource person as well as provided hands-on sessions and judged posters apart from several informal invitations.