

Executive Summary

This NSTMIS supported project entitled *Design of a Computational Framework for Discipline-wise and Thematic Mapping of Research Performance of Indian Higher Educational Institutions*, has been carried out during the period of 2020-2022. The project has been undertaken at the Department of Computer Science, Banaras Hindu University, Varanasi, Uttar Pradesh.

It had been motivated by the observation that the existing university ranking frameworks provide a broad indicator-based overview of higher education institutions, formulated with a commercial perspective. The requirement of a detailed information platform for research, development and education institutions has been felt by the scientometrics community in India. This study has attempted to fill this gap by developing a framework and a platform for collecting, computing and providing information on some of the specific indicators of research activity. Using the publication data of Indian institutions, a web-based portal for documenting and disseminating results from the computation of indicators on the research performance of selected institutions has been developed and hosted.

The outcomes from the study can be beneficial for various stakeholders, namely, researchers, policy makers, ranking agencies, students and the general public. It will augment existing ranking platforms, and Higher Education Institutions (HEI) listings (such as NIRF, DST India Science and Technology Portal), by providing institution specific indicators of scientific performance. The portal developed under this project will provide assistance to prospective research students in finding best research places on specific topics of interest; ranking agencies (national and international) as it will enable them to find top performing institutions on a specific research topic, as the research performance parameters computed on the detailed topics can be used to identify best research places/ institutions in a specific area and also produce discipline-wise rankings of Indian Institutions.

This project aims to meet the gap of design of "an accurate article-level fine-grained thematic area classification system" and propose applications by utilizing this knowledge. The project has led to the formulation of: (a) an article-level classification methodology, (b) an automated web-based computational system for research performance assessment, (c) Research Papers and technical reports, and (d) Research Performance assessment reports of 1000 Indian institutions. The project has produced technical reports, a website and other documentary output. The project team also aspires to utilise available computational tools for translation and make all project outputs available in Hindi language too in future.

The project has been completed in two stages. The first stage involved the development of a new computational framework for assessment of thematic performance of institutions on the basis of sample data from the Web of Science Database. Two indices to rank the institutions, named x-index and x(g)-index were developed for this purpose. The working of this framework was demonstrated using data of 195 Indian institutions in the area of Computer Science. IIT Kharagpur and Thapar institute of engineering and technology had highest x-index and x(g)-index, and were placed on the top most position. Thematic research strengths and core competencies of the 195 institutions were also identified.

In the second stage of the project, the study covered data from all the disciplines indexed on the Dimensions database for a period of 10 years. The access to the database was acquired by using Dimension's API. A total of 1000 Indian organisations (Higher Education Institutions, Government hospitals and laboratories, and R&D firms) were assessed under the project. Detailed reports on scientific productivity of India and the performance of selected 1000 Indian Institutions were developed using the conceptual framework. These reports were thereafter used to show visualisations on the web portal.

India's research output was also mapped and compared with some of the leading countries in S&T research. Indian Research Output in the 2010 to 2019 period grew from 60,250 to 148,724 publications and it moved from 10th to 6th position in terms of global productivity. The publications during this period had received a total of 11,805,480 citations. It was observed that India's contribution to the top 1% of highly cited papers has grown from 1.85% in 2010 to 4.3% in 2019. Comparatively, the top contributions were from China (33.2%) and U.S. (36.96%). The subject wise output was also estimated for 22 major subjects. The overall research output is dominated by publications in subjects like Medical & Health Sciences, Engineering, Chemical Sciences and Information & Computing Sciences. Approximately 20% of the research output involved collaboration at international level. United States of America (33.09%), United Kingdom (12.42%), Germany (9.19%), China (8.87%), South Korea (7.72%) and Australia (7.06%) were the major contributing partner countries. These publications also received higher citations at 18.65 (average citations per paper) compared to domestic papers at 7.95. The volume of research funding to Indian Institutions was estimated as 4.6 billion USD, placing India at the 15th spot in terms of research grants. It is about 100 orders of magnitude lesser as compared to the United States, and 27 times less as compared to the United Kingdom and Germany.

It was observed that only 29.3% of Indian papers are Female 1st authored, as compared to 70.7% Male 1st authored papers. Subject areas like History & Archaeology, Built Environment & Design and Law & Legal studies have a slightly higher percentage of Female 1st authored papers. India is placed on the Lower side of Open Access availability level, with only 33.61% papers available in Open Access. Out of the total funded research output during 2010-2019, 71% is Closed Access, 12% Gold Open Access, 9% Green Open Access, 5% Bronze Open Access and 3% Hybrid Open Access. The social media coverage of a research output in different subject areas vary significantly, ranging from coverage percentage of 40.18% for Biological Sciences, to 9.47% in Information & Computing Sciences. It is observed that out of the total research output in 2010, 13.7% of the research output gets some social media attention. This level has continuously increased during the period, reaching to 25.2% in 2017 and settling down to 23.3% in 2019.

The study has developed a thematic-area based article classification system; a set of detailed reports on research performance of Indian Higher Education Institutions in different detailed topics; and a fully automated web-based portal for documenting and disseminating results from the computation of indicators on the research performance of selected institutions. These outputs in addition to being useful research and consultation tools for the various stakeholders also provide a proof of concept for the possibility of a larger national information platform with granular information about the research and development. It has the following direct benefits.

1. A fully automated Web-based portal for disseminating results about computed indicators of different institutions.
2. The results generated from the study can be of great interest to ranking agencies, national and International, as it would allow them to find top performing institutions on a specific research topic.
3. Help prospective research students in finding best research places on specific topics of interest.
4. The results also provide a mechanism for institutional collaboration recommendation.
5. Produced reports and documents on research performance of Indian Institutions, something which was largely missing till date.

This project is novel in many respects with many interesting outputs and benefits. However, at the moment, the project is limited to data of Indian research Institutions only. Though the developed system has the capability of seamless extension to data for the whole world, currently only data for Indian institutions has been populated. Collection of full data for the whole world will take a lot of effort and huge memory requirements will be there. However, it is proposed to cover the data for the whole world as a future extension of the project.