

THE TOP 100 INSTITUTIONS IN INDIA AND IN THE WORLD

Rajender Kumar Bathla¹, Gurjeet Singh Pandher²

Department of Computer Science & Applications,
Desh Bhagat University, Mandi Gobindgarh

ABSTRACT

This study undertakes a comprehensive scientometric and text-oriented evaluation of computer science research outputs from the 100 leading institutions in India and around the world. The analysis draws on data indexed in the Scopus database over a 25-year span (1989–2013). Utilizing a dual-method framework, it incorporates both conventional scientometric techniques and computational text analysis.

The scientometric component focuses on comparing Indian computer science research with that of the world's top institutions by examining metrics such as total publications, citation impact, co-authorship trends, and levels of international collaboration. Meanwhile, the text analysis identifies prominent research themes and their evolution over time within both groups. This unique investigation offers one of the first comparative insights into the research characteristics of Indian versus global institutions in the field of computer science, delivering meaningful findings and valuable conclusions.

Keywords: Computer Science Research, India, Information Technology, Informetrics, Scientometrics.

1. INTRODUCTION

Information Technology (IT) plays a vital role in the progress of any nation in the modern era. The Information and Communication Technologies (ICT), to a large extent, are based on research and development in Computer Science (CS). The present century has witnessed the rise of knowledge economies, and nations worldwide are investing increasingly in science and technology, particularly ICT.

Though a late starter in CS research, India has focused on this domain in the last two decades and has attained a respectable presence in the IT development sector. However, Indian institutions are yet to reach the forefront of knowledge creation in the CS domain.

This paper aims to measure and assess the research competence of the most productive Indian institutions vis-à-vis the most productive global institutions. Our analysis involves a large set of 100 productive institutions in India as well as the world. The data covers CS research output for a 25-year period (1989–2013) indexed in Scopus.

We aim to answer the following research questions:

- What are characteristic similarities and differences in CS research of the most productive Indian institutions compared to the world's most productive institutions?
- What is India's contribution to global CS research, and what is its relative impact?
- What major research themes are pursued by Indian institutions, and how do they relate to global research themes?

- What inferences can be drawn for improving Indian CS research?

The paper is organized as follows: Section 2 presents related work. Section 3 gives an overview of CS research in India. Section 4 describes data collection, Section 5 details the methodology, Section 6 presents scientometric results, Section 7 presents trajectory analysis, Section 8 discusses research themes, and Section 9 concludes with a summary and key inferences.

2. RELATED WORK

While no previous study has performed the kind of analytical characterization we attempted, several works have focused on scientometric profiling for specific countries, regions, or subjects. Most of these are limited to standard scientometric analysis.

Key studies include:

- Fu (2013) analyzed China's research compared with seven major industrialized countries.
- Kao (2012) studied management research improvements in Taiwan.
- Gupta (2011) assessed Indian CS research output, citation impact, and international collaboration (1999–2008).
- Matthews (2013) analyzed individual productivity in physics in South African universities.
- De Souza (2012) profiled researchers and collaboration patterns in information science in Brazil.
- Uddin (2014a, 2014b) mapped CS research in South

Asian countries and SAARC regions.

- Tang (2011) explored China-US collaboration in nanotechnology.
- Several studies (Ozel, Onyancha, Costa, Prathap, Teodorescu, Abramo, Viana, Ortega, Liu, 2011–2013) analyzed collaboration, growth, and research output metrics.

While these works informed our methodology, none combined standard scientometric techniques with computational text analysis for CS research in India vis-à-vis the world.

3. OVERVIEW OF CS RESEARCH IN INDIA

CS as an academic discipline began systematically in the 1950s. The Cambridge Diploma in Computer Science started in 1953 at the University of Cambridge, followed by Purdue University in the USA (1962) and Stanford University (1965). In India, CS courses began at IIT Kanpur in 1963, with independent CS programs starting in 1971. Other key institutions include IIT Madras (1973), IIT Kharagpur (1980), IIT Bombay (1982), and the Department of Computer Science and Automation at IISc Bangalore (1969). Jadavpur University offered a postgraduate diploma in CS in 1968, and ISI Calcutta introduced an M.Tech in CS in 1981.

Initially focused on education and manpower development, research in CS gained momentum in the 1990s. Major IT companies, such as IBM (1988) and Microsoft Research (2005), established laboratories in India, alongside government research labs (CSIR, DRDO, ISRO).

Figure 1 shows India's contribution to global CS research, increasing from 1.88% (1989) to 4.95% (2013), with accelerated growth from 2000 onwards.

4. DATA COLLECTION

Data was collected from Scopus for 1989–2013. Globally, 2,876,512 records were found; India contributed 84,385 papers (~2.93%). We collected institution-wise data for the 100 most productive Indian institutions (I100) and 100 most productive global institutions (W100).

After removing duplicates and records with missing years, 59,619 unique records remained for I100, and 846,526 unique records for W100. Related statistics such as GERD and researchers per million were also collected (Table 1).

5. METHODOLOGY

Our methodology combines standard scientometric analysis

and computational text analysis.

Scientometric Analysis

Six indicators were used:

- Total Publications (TP) – primary
- Total Citations (TC)
- Average Citation per Paper (ACPP)
- Highly Cited Papers (HiCP)
- Internationally Collaborated Papers (ICP)
- Cited Percentage

Text Analysis

We classified each research paper into 11 major CS thematic areas based on author keywords, titles, and abstracts (Uddin & Singh, 2015). Thematic trends were measured for I100 and W100.

6. SCIENTOMETRIC INDICATOR COMPUTATIONS

6.1 Research Output

No Indian institution appears in the W100 set. For example, University College London produced 5,747 papers over 25 years; IIT Kharagpur and IISc Bangalore produced 3,514 and 3,506, respectively.

Key indicators show differences: ACPP (I100: 2.66, W100: 9.37), HiCP (I100: 60, W100: 855), ICP (I100: 110, W100: 2,419), cited percentage (I100: 39.33%, W100: 58.91%).

6.2 Co-authorship Patterns

About 39.4% of I100 papers involve two authors, 32% involve three, and 5% are single-authored. W100 shows higher collaboration (average co-authorship 3.41 vs. 2.92). Figures 2(a) and 2(b) illustrate trends.

6.3 International Collaboration

25% of I100 papers involve international collaboration vs. 30% for W100. Top collaborators for Indian authors include the USA (4,316 papers), Singapore (849), and Canada (801). Figure 3 visualizes the trends.

6.4 Citation-Based Impact

Citation count measures research quality and influence. Figures 4(a–c) show highly cited papers, contribution of Indian institutions, and percentage of uncited papers. Indian institutions show lower citation impact than W100, though more institutions contribute to high-impact research over time.

7. PERFORMANCE BASED ON TRAJECTORY ANALYSIS

Exergy-based analysis combines quantity (papers) and

quality (impact) of research: $X = \frac{C^2}{P}$, where C is citations and P is papers. Figures 8(a–c) compare top Indian institutions (I100) with W100 institutions. Indian institutions lag in quantity-quality composite performance.

8. IDENTIFYING RESEARCH THEMES

Text-based analysis using burst detection (Kleinberg, 2003) identified leading themes. Figures 6(a–b) and supplementary figures 6(c–g) visualize keyword density and tag clouds.

Key findings:

- I100 emphasizes “Computer Networks”, “Information Technology”, “Computational Methods”.
- W100 emphasizes “Algorithms”, “Approximation Theory”, “Network Protocols”, “Computational Complexity”, and “Optimization”.

Both I100 and W100 share “Algorithm”, “Computer Simulation”, and “Mathematical Models” as prominent research themes.

9. SUMMARY AND CONCLUSION

This study presents scientometric, network-theoretic, and text-based analyses of CS research in India (I100) and globally (W100) for 1989–2013.

Key conclusions:

- Indian institutions have improved research output but lag in quality compared to global peers.
- Exergy-based performance measures highlight gaps in quantity-quality research output.
- Research themes show similarities and differences between Indian and global institutions.

The results can guide policy formulation, student decisions, and academic planning to enhance CS research in India.

Acknowledgements

Supported by research grants from the Department of Science and Technology, Government of India (Grant: INT/MEXICO/P-13/2012) and University Grants Commission of India (Grant: F.No. 41-624/2012(SR)).

REFERENCES

1. Abramo, G., D'Angelo, C.A. & Murgia, G. (2013). The collaboration behaviors of scientists in Italy: A field level analysis. *Journal of Informetrics*, 7(2), 442–454.
 2. Costa, B.M.G., Pedro, E.S. & Macedo, G.R. (2013). Scientific collaboration in biotechnology: the case of the northeast region in Brazil. *Scientometrics*, 95(2), 571–592.
 3. De Souza, C.G. & Ferreira, M.L.A. (2012). Researchers profile, co-authorship pattern and knowledge organization in information science in Brazil. *Scientometrics*, 95(2), 673–687.
- ... (rest of references formatted consistently)