Arnab Bhattacharya

Professor
Dept. of Computer Science and Engineering,
Coordinator
ŚIKṢĀ: Study Centre for Indian Knowledge System for Holistic Advancement,

Indian Institute of Technology Kanpur
Kanpur, UP, India. PIN: 208016.

arnabb@iitk.ac.in, arnabb@cse.iitk.ac.in
https://www.cse.iitk.ac.in/users/arnabb/
+91-512-679-7650, +91-512-259-7650

Areas of Research Interest: Natural Language Processing, Data Mining, Databases, Information Retrieval, Artificial Intelligence.

Experience:

- Professor, Dept. of Computer Science and Engineering, Indian Institute of Technology (IIT), Kanpur, India. December 2020 – present.

- Associate Professor, Dept. of Computer Science and Engineering, Indian Institute of Technology (IIT), Kanpur, India. June 2014 – December 2020.

- Assistant Professor, Dept. of Computer Science and Engineering, Indian Institute of Technology (IIT), Kanpur, India. December 2007 – June 2014.

- Project Scientist, Dept. of Computer Science, University of California, Santa Barbara, USA. September 2007 – November 2007.

- Software Design Engineer, Texas Instruments (India) Ltd., Bangalore, India. July 2001 – July 2002.

Education:

- Ph.D. in Computer Science, Dept. of Computer Science, University of California, Santa Barbara, USA. 2007.

- M.S. in Computer Science, Dept. of Computer Science, University of California, Santa Barbara, USA. 2007.

- Bachelor of Computer Science and Engineering (B.C.S.E.), Jadavpur University, India. 2001.

Book:

- "Fundamentals of Database Indexing and Searching". Arnab Bhattacharya. *CRC Press*, 2014.

Selected Publications:

1. Legal Judgment Reimagined: PredEx and the Rise of Intelligent AI Interpretation in Indian Courts. Shubham Kumar Nigam, Anurag Sharma, Danush Khanna, Noel Shallum, Kripabandhu Ghosh, Arnab Bhattacharya. *Findings of the Association for Computational Linguistics (Findings of ACL)*, 2024, to appear, Bangkok, Thailand.

2. A Likelihood Ratio Test of Genetic Relationship among Languages. V. S. D. S. Mahesh Akavarapu, Arnab Bhattacharya. *Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, 2024, to appear, Mexico City, Mexico.

3. Automated Cognate Detection as a Supervised Link Prediction Task with Cognate Transformer. V. S. D. S. Mahesh Akavarapu, Arnab Bhattacharya. *The 18th Conf. of the European Chapter of the Association for Computational Linguistics (EACL)*, 2024, pages 965-975, Malta.

4. VeNoM: Approximate Subgraph Matching with Enhanced Neighbourhood Structural Information. Shubhangi Agarwal, Sourav Dutta, Arnab Bhattacharya. *ACM Joint Int. Conf. on Data Science & Management of Data (CoDS-COMAD)*, 2024, pages 18-26, Bengaluru, India.

5. Short-Term Fog Forecasting using Meteorological Observations at Airports in North India. Shruti Sharma, Kriti Bajaj, Prasad Deshpande, Arnab Bhattacharya, Shivam Tripathi. *ACM Joint Int. Conf. on Data Science & Management of Data (CoDS-COMAD)*, 2024, pages 307-315, Bengaluru, India. (Candidate for best paper)

6. Cognate Transformer for Automated Phonological Reconstruction and Cognate Reflex Prediction. V. S. D. S. Mahesh Akavarapu, Arnab Bhattacharya. *The 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2023, pages 6852-6862, Singapore.

7. Antarlekhaka: A Comprehensive Tool for Multi-task Natural Language Annotation. Hrishikesh Terdalkar, Arnab Bhattacharya. *3rd Workshop for Natural Language Processing Open Source Software (NLP-OSS) @ EMNLP 2023*, 2023, Singapore.

8. Vacaspati: A Diverse Corpus of Bangla Literature. Pramit Bhattacharyya, Joydeep Mondal, Subhadip Maji, Arnab Bhattacharya. *13th Int. Jt. Conf. on Natural Language Processing and 3rd Conf. of the Asia-Pacific Chapter of the Association for Computational Linguistics (IJCNLP-AACL)*, 2023, pages 1118-1130, Bali, Indonesia.

9. Development of a Fog Index to Study Relationships Between Fog and Climate Variables. Rakshit Paurwal, Shivam Tripathi, Arnab Bhattacharya. Book chapter in *Climate Change Impact on Water Resources* edited by P. V. Timbadiya and V. P. Singh and P. J. Sharma, pages 437-446. Springer, 2023.

10. Task and Model Agnostic Adversarial Attack on Graph Neural Networks. Kartik Sharma, Samidha Verma, Sourav Medya, Arnab Bhattacharya, Sayan Ranu. *Association for the Advancement of Artifical Intelligence (AAAI)*, 2023, pages 15091-15099, Washington DC, USA.

11. Event-based Fog Climatology and Typology for Cities in Indo-Gangetic Plains. Prasad Deshpande, Deepika Meena, Shivam Tripathi, Arnab Bhattacharya, Mahendra K. Verma. *Urban Climate*, 2023, 51, 101642.

12. SpotSpam: Intention Analysis-driven SMS Spam Detection using BERT Embeddings. Christopher Oswald, Sona Elza Simon, Arnab Bhattacharya. *ACM Transactions on the Web*, 2022, 16(3), 3521-3526.

13. HLDC: Hindi Legal Documents Corpus. Arnav Kapoor, Mudit Dhawan, Anmol Goel, Arjun T H, Akshala Bhatnagar, Vibhu Agrawal, Amul Agrawal, Arnab Bhattacharya, Ponnurangam Kumaraguru, Ashutosh Modi. *Findings of the Association for Computational Linguistics (Findings of ACL)*, 2022, pages 3521-3526, Ireland.

14. Predictions of Cardio-Vascular Disease (CVD) using Ensemble Learning Methods. Christopher Oswald, Gadi Jaya Sathwika, Arnab Bhattacharya. *ACM Joint International Conference on Data Science & Management of Data (CoDS-COMAD)*, 2022, pages 292-293, Bengaluru, India.

15. Semantic Annotation and Querying Framework based on Semi-structured Ayurvedic Text. Hrishikesh Terdalkar, Arnab Bhattacharya, Madhulika Dubey, Ramamurthy S, Bhavna Naneria Singh. *World Sanskrit Conference (WSC)*, 2022, Australia.

16. VerSaChI: Finding Statistically Significant Subgraph Matches using Chebyshev's Inequality. Shubhangi Agarwal, Sourav Dutta, Arnab Bhattacharya. *International Conference on Information and Knowledge Management (CIKM)*, 2021, pages 545-554, Australia.

17. Computing and Maintaining Provenance of Query Result Probabilities in Uncertain Knowledge Graphs. Garima Gaur, Abhishek Dang, Arnab Bhattacharya, Srikanta Bedathur. *International Conference on Information and Knowledge Management (CIKM)*, 2021, pages 2812-2816, Australia.

18. Sangrahaka: A Tool for Annotating and Querying Knowledge Graphs. Hrishikesh Terdalkar, Arnab Bhattacharya. *29th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE)*, 2021, pages 1520-1524, Athens, Greece.

19. Indian Legal Documents Corpus (ILDC) for Court Judgment Prediction and Explanation. Vijit Malik, Rishabh Sanjay, Shubham Kumar Nigam, Kripabandhu Ghosh, Shouvik Kumar Guha, Arnab Bhattacharya, Ashutosh Modi. *Joint Conference of Association for Computational Linguistics and International Joint Conference Natural Language Processing (ACL-IJCNLP)*, 2021, pages 4046-4062, Bangkok, Thailand.

20. Exploring State-of-the-Art Nearest Neighbor (NN) Search Techniques. Parth Nagarkar, Arnab Bhattacharya, Omid Jafari. *ACM Joint International Conference on Data Science & Management of Data (CoDS-COMAD)*, 2021, pages 443-446, Bengaluru, India.

21. GraphReach: Position-Aware Graph Neural Network using Reachability Estimations. Sunil Nishad, Shubhangi Agarwal, Sayan Ranu, Arnab Bhattacharya. *International Joint Conference on Artificial Intelligence (IJCAI)*, 2021, pages 1527-1533, Montreal, Canada.

22. TIPS: Mining Top-K Locations to Minimize User-Inconvenience for Trajectory-Aware Services. Shubhadip Mitra, Priya Saraf, Arnab Bhattacharya. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 2021, 33(3), pages 1238-1250.

23. How and Why is an Answer (Still) Correct? Maintaining Provenance in Dynamic Knowledge Graphs. Garima Gaur, Arnab Bhattacharya, Srikanta Bedathur. *International Conference on Information and Knowledge Management (CIKM)*, 2020, pages 405-414, Virtual Event, Ireland.

24. ChiSeL: Graph Similarity Search using Chi-Squared Statistics in Large Probabilistic Graphs. Shubhangi Agarwal, Sourav Dutta, Arnab Bhattacharya. *Proceedings of the VLDB Endowment (PVLDB)*, 2020, 13(10), pages 1654-1668.

25. NetClus: A Scalable Framework to Mine Top-K Locations for Placement of Trajectory-Aware Services. Shubhadip Mitra, Priya Saraf, Richa Sharma, Arnab Bhattacharya, Sayan Ranu. *ACM Joint International Conference on Data Science & Management of Data (CoDS-COMAD)*, 2019, pages 27-35, Kolkata, India.

26. Framework for Question-Answering in Sanskrit through Automated Construction of Knowledge Graphs. Hrishikesh Terdalkar, Arnab Bhattacharya. *6th International Sanskrit Computational Linguistics Symposium (ISCLS)*, 2019, pages 98-117, Kharagpur, India.

27. RAQ: Relationship-Aware Graph Querying in Large Networks. Jithin Vachery, Akhil Arora, Sayan Ranu, Arnab Bhattacharya. *International World Wide Web Conference (WWW)*, 2019, pages 1886-1896, San Francisco, USA.

28. HD-Index: Pushing the Scalability-Accuracy Boundary for Approximate kNN Search in High-Dimensional Spaces. Akhil Arora, Sakshi Sinha, Piyush Kumar, Arnab Bhattacharya. *Proceedings of the VLDB Endowment (PVLDB)*, 2018, 11(8), pages 906-919.

29. Finding Largest Rectangle inside a Digital Object and Rectangularization. Apurba Sarkar, Arindam Biswas, Mousumi Dutt, Arnab Bhattacharya. *Journal of Computer and System Sciences*, 2018, 95, pages 204-217.

30. Image Management for Biological Data. Arnab Bhattacharya, Vebjorn Ljosa. Book chapter in *Encyclopedia of Database Systems (2nd Edition)* edited by L. Liu and M. T. Özsu. Springer, 2018.

31. MineAr: Using Crowd Knowledge for Mining Association Rules in the Health Domain. Milan Someswar, Arnab Bhattacharya. *ACM Joint International Conference on Data Science & Management of Data (CoDS-COMAD)*, 2018, pages 108-117, Goa, India.

32. Finding Shell Company Accounts using Anomaly Detection. Devendra K. Luna, Girish K. Palshikar, Manoj Apte, Arnab Bhattacharya. *ACM Joint International Conference on Data Science & Management of Data (CoDS-COMAD)*, 2018, pages 167-174, Goa, India.

33. Tracking the Impact of Fact Deletions on Knowledge Graph Queries using Provenance Polynomials. Garima Gaur, Srikanta J. Bedathur, Arnab Bhattacharya. *International Conference on Information and Knowledge Management (CIKM)*, 2017, pages 2079-2082, Singapore.

34. SkyGraph: Retrieving Regions of Interest using Skyline Subgraph Queries. Shiladitya Pande, Sayan Ranu, Arnab Bhattacharya. *Proceedings of the VLDB Endowment (PVLDB)*, 2017, 10(11), pages 1382-1393.

35. NetClus: A Scalable Framework for Locating Top-K Sites for Placement of Trajectory-Aware Services. Shubhadip Mitra, Priya Saraf, Richa Sharma, Arnab Bhattacharya, Sayan Ranu, Harsh Bhandari. *International Conference on Data Engineering (ICDE)*, 2017, pages 87-90, San Diego, USA.

36. K-Dominant Skyline Join Queries: Extending the Join Paradigm to K-Dominant Skylines. Anuradha Awasthi, Arnab Bhattacharya, Sanchit Gupta, Ujjwal K. Singh. *International Conference on Data Engineering (ICDE)*, 2017, pages 99-102, San Diego, USA.

37. Neighbor-Aware Search for Approximate Labeled Graph Matching using the Chi-Square Statistics. Sourav Dutta, Pratik Nayek, Arnab Bhattacharya. *International World Wide Web Conference (WWW)*, 2017, pages 1281-1290, Perth, Australia.

38. Automatic Grading and Feedback using Program Repair for Introductory Programming Courses. Sagar Parihar, Ziyaan Dadachanji, Praveen Kumar Singh, Rajdeep Das, Amey Karkare, Arnab Bhattacharya. *ACM Conference on Innovation and Technology in Computer Science Education (ITiCSE)*, 2017, pages 92-97, Bologna, Italy.

39. GARUDA: A System for Large-Scale Mining of Statistically Significant Connected Subgraphs. Satyajit Bhadange, Akhil Arora, Arnab Bhattacharya. *Proceedings of the VLDB Endowment (PVLDB)*, 2016, 9(13), pages 1449-1452.

40. SMS: Stable Matching Algorithm using Skylines. Rohit Anurag, Arnab Bhattacharya. *International Conference on Scientific and Statistical Database Management (SSDBM)*, 2016, pages 24:1-24:4, Budapest, Hungary.

41. SkyCover: Finding Range-Constrained Approximate Skylines with Bounded Quality Guarantees. Shubhendu Aggarwal, Shubhadip Mitra, Arnab Bhattacharya. *International Conference on Management of Data (COMAD)*, 2016, pages 1-12, Pune, India.

42. Finding Largest Rectangle inside a Digital Object. Apurba Sarkar, Arindam Biswas, Mousumi Dutt, Arnab Bhattacharya. *Computational Topology in Image Context (CTIC)*, 2016, pages 157-169, Marseille, France.

43. Probabilistic Aggregate Skyline Join Queries: Skylines with Aggregate Operations over Existential Uncertain Relations. Arnab Bhattacharya, Shrikant Awate. *International Conference on Scientific and Statistical Database Management (SSDBM)*, 2015, pages 5:1-5:12, San Diego, USA.

44. Trajectory Aware Macro-cell Planning for Mobile Users. Shubhadip Mitra, Sayan Ranu, Vinay Kolar, Arnab Bhattacharya, Ravi Kokku, Aditya Telang, Sriram Raghavan. *IEEE International Conference on Computer Communications (INFOCOM)*, 2015, 792-800, Hong Kong, China.

45. Generation of Random Triangular Digital Curves using Combinatorial Techniques. Apurba Sarkar, Arindam Biswas, Mousumi Dutt, Arnab Bhattacharya. *International Conference on Pattern Recognition and Machine Intelligence (PReMI)*, 2015, pages 136-145, Warsaw, Poland.

46. Using Social Connections to Improve Collaborative Filtering. Kanish Manuja, Arnab Bhattacharya. *IKDD Conference on Data Sciences (CoDS)*, 2015, pages 140-141, Bengaluru, India.

47. Generation of Random Digital Curves using Combinatorial Techniques. Apurba Sarkar, Arindam Biswas, Mousumi Dutt, Arnab Bhattacharya. *Conference on Algorithms and Discrete Applied Mathematics (CALDAM)*, 2015, pages 286-297, Kanpur, India.

48. Mining Statistically Significant Connected Subgraphs in Vertex Labeled Graphs. Akhil Arora, Mayank Sachan, Arnab Bhattacharya. *SIGMOD International Conference on Management of Data (SIGMOD)*, 2014, pages 1003-1014, Snowbird, USA.

49. Efficient and Effective Route Planning in Road Networks with Probabilistic Data using Skyline Paths. Arzoo Katiyar, Arnab Bhattacharya, Shubhadip Mitra. *IKDD Conference on Data Sciences (CoDS)*, 2014, New Delhi, India.

50. Emotion Recognition from Audio and Visual Data using F-score based Fusion. Abhishek Gera, Arnab Bhattacharya. *IKDD Conference on Data Sciences (CoDS)*, 2014, New Delhi, India.

51. RCached-tree: An Index Structure for Efficiently Answering Popular Queries. Manash Pal, Arnab Bhattacharya, Debjyoti Paul. *International Conference on Information and Knowledge Management (CIKM)*, 2013, pages 1173-1176, San Francisco, USA.

52. Efficient Edit Distance based String Similarity Search using Deletion Neighborhoods. Shashwat Mishra, Tejas Gandhi, Akhil Arora, Arnab Bhattacharya. *EDBT/ICDT Workshops*, 2013, pages 375-383, Genoa, Italy.

53. Hybrid HBase: Leveraging Flash SSDs to Improve Cost per Throughput of HBase. Anurag Awasthi, Avani Nandini, Arnab Bhattacharya, Priya Sehgal. *International Conference on Management of Data (COMAD)*, 2012, pages 68-79, Pune, India.

54. A Plant Identification System using Shape and Morphological Features on Segmented Leaflets: Team IITK, CLEF 2012. Akhil Arora, Ankit Gupta, Nitesh Bagmar, Shashwat Mishra, Arnab Bhattacharya. *CLEF (Online Notes/Labs/Workshop)*, 2012, Rome, Italy.

55. Mining Statistically Significant Substrings using the Chi-Square Statistic. Mayank Sachan, Arnab Bhattacharya. *Proceedings of the VLDB Endowment (PVLDB)*, 2012, 5(10), pages 1052-1063.

56. Mining Statistically Significant Substrings Based on the Chi-Square Measure. Sourav Dutta, Arnab Bhattacharya. Book chapter in *Pattern Discovery Using Sequence Data Mining: Applications and Studies* edited by P. Kumar, P. R. Krishna and S. B. Raju. IGI Global, 2012.

57. Minimally Infrequent Itemset Mining using Pattern-Growth Paradigm and Residual Trees. Ashish Gupta, Akshay Mittal, Arnab Bhattacharya. *International Conference on Management of Data (COMAD)*, 2011, pages 57-68, Bengaluru, India. (Best paper)

58. Caching Stars in the Sky: A Semantic Caching Approach to Accelerate Skyline Queries. Arnab Bhattacharya, B. Palvali Teja, Sourav Dutta. *International Conference on Database and Expert Systems Applications (DEXA)*, 2011, pages 493-501, Toulouse, France.

59. A Continuous Query System for Dynamic Route Planning. Nirmesh Malviya, Samuel Madden, Arnab Bhattacharya. *International Conference on Data Engineering (ICDE)*, 2011, pages 792-803, Hannover, Germany.

60. Finding the Bias and Prestige of Nodes in Networks based on Trust Scores. Abhinav Mishra, Arnab Bhattacharya. *International World Wide Web Conference (WWW)*, 2011, pages 567-576, Hyderabad, India.

61. Aggregate Skyline Join Queries: Skylines with Aggregate Operations over Multiple Relations. Arnab Bhattacharya, B. Palvali Teja. *International Conference on Management of Data (COMAD)*, 2010, pages 15-26, Nagpur, India. (Best student paper)

62. INSTRUCT: Space-Efficient Structure for Indexing and Complete Query Management of String Databases. Sourav Dutta, Arnab Bhattacharya. *International Conference on Management of Data (COMAD)*, 2010, pages 27-38, Nagpur, India.

63. Simulated Evolution and Learning. *Proceedings of the 8th International Conference on Simulated Evolution and Learning (SEAL)*. Co-edited by K. Deb, A. Bhattacharya, N. Chakraborti, P. Chakroborty, S. Das, J. Dutta, S. K. Gupta, A. Jain, V. Aggarwal, J. Branke, S. J. Louis, K. C. Tan, Springer, 2010.

64. Minimum Spanning Tree on Spatio-Temporal Networks. Viswanath Gunturi, Shashi Shekhar, Arnab Bhattacharya. *International Conference on Database and Expert Systems Applications (DEXA)*, 2010, pages 149-158, Bilbao, Spain.

65. Finding Top-k Similar Pairs of Objects Annotated with Terms from an Ontology. Arnab Bhattacharya, Abhishek Bhowmick, Ambuj K. Singh. *International Conference on Scientific and Statistical Database Management (SSDBM)*, 2010, pages 214-232, Heidelberg, Germany.

66. Most Significant Substring Mining based on Chi-square Measure. Sourav Dutta, Arnab Bhattacharya. *Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*, 2010, pages 319-327, Hyderabad, India.

67. Querying Spatial Patterns. Vishwakarma Singh, Arnab Bhattacharya, Ambuj K. Singh. *International Conference on Extending Database Technology (EDBT)*, 2010, pages 418-429, Lausanne, Switzerland.

68. Image Management for Biological Data. Arnab Bhattacharya, Vebjorn Ljosa. Book chapter in *Encyclopedia of Database Systems* edited by M. T. Özsu and L. Liu. Springer, 2009.

69. On Low Distortion Embeddings of Statistical Distance Measures into Low Dimensional Spaces. Arnab Bhattacharya, Purushottam Kar, Manjish Pal. *International Conference on Database and Expert Systems Applications (DEXA)*, 2009, pages 164-172, Linz, Austria.

70. FTDP-17 Mutations in Tau Alter the Regulation of Microtubule Dynamics: An "Alternative Core" Model for Normal and Pathological Tau Action. Adria LeBoeuf, Sasha F. Levy, Michelle Gaylord, Arnab Bhattacharya, Ambuj K. Singh, Mary Ann Jordan, Leslie Wilson, Stuart C. Feinstein. *Journal of Biological Chemistry*, 2008, 283(52), pages 36406-36415.

71. A General Modeling and Visualization Tool for Comparing Different Members of a Group: Application to Studying Tau-Mediated Regulation of Microtubule Dynamics. Arnab Bhattacharya, Sasha Levy, Adria LeBoeuf, Michelle Gaylord, Leslie Wilson, Ambuj K. Singh, Stuart C. Feinstein. *BMC Bioinformatics*, 2008, 9, page 339.

72. Efficient Computation of Statistical Significance of Query Results in Databases. Vishwakarma Singh, Arnab Bhattacharya, Ambuj K. Singh. *International Conference on Scientific and Statistical Database Management (SSDBM)*, 2008, pages 509-516, Hong Kong, China.

73. MIST: Distributed Indexing and Querying in Sensor Networks using Statistical Models. Arnab Bhattacharya, Anand Meka, Ambuj K. Singh. *International Conference on Very Large Data Bases (VLDB)*, 2007, pages 854-865, Vienna, Austria.

74. Indexing Spatially Sensitive Distance Measures Using Multi-Resolution Lower Bounds. Vebjorn Ljosa, Arnab Bhattacharya, Ambuj K. Singh. *International Conference on Extending Database Technology (EDBT)*, 2006, pages 865-883, Munich, Germany.

75. LB-Index: A Multi-Resolution Index Structure for Images. Vebjorn Ljosa, Arnab Bhattacharya, Ambuj K. Singh. *International Conference on Data Engineering (ICDE)*, 2006, pages 144-145, Atlanta, USA.

76. ViVo: Visual Vocabulary Construction for Mining Biomedical Images. Arnab Bhattacharya, Vebjorn Ljosa, Jia-Yu Pan, Mark R. Verardo, Hyung-Jeong Yang, Christos Faloutsos, Ambuj K. Singh. *International Conference on Data Mining (ICDM)*, 2005, pages 50-57, Houston, USA. (One of the top five student papers)

77. ProGreSS: Simultaneous Searching of Protein Databases by Sequence and Structure. Arnab Bhattacharya, Tolga Can, Tamer Kahveci, Ambuj K. Singh, Yuan-Fang Wang. *Pacific Symposium on Biocomputing (PSB)*, 2004, pages 264-275, Hawaii, USA.

Selected Invited Talks:

1. "Can Indian Languages Benefit from the Generative AI Wave? An Overview of Vision and Possibilities for GPTs for Bharata" at Flipkart Data Science and Analytics, 2024.

2. "How Indian Languages can Benefit from the AI Wave" at Assam AI Initiative webinar, 2024.

3. "Artificial Intelligence: The Path Forward" at Interra Information Technologies, 2023.

4. "Artificial Intelligence" at the Kendriya Vidyalaya, IIT Kanpur, Kanpur, 2023.

5. "Generative Adversarial Networks (GAN)" at National Workshop on Machine Learning and Data Science Trends and Techniques (MLDS'23), Tiruchirapalli, 2023.

6. "Bhāratīya Names" at World Association of Hindu Academicians (WAHA) Conference, Faridabad, 2023.

7. "Improving Graph Neural Networks through Position-Aware Embeddings" at Vortex, NIT Tiruchirapalli, 2023.

8. "Position-Aware Graph Neural Networks using Reachability Estimations" at Data Science and Machine Learning in Real World series, IIIT Kottayam, 2022.

9. "Towards Artificial Vikramaditya: Can Machines Deliver and Explain Fair Judgments?" at the IIITD-TIH Monthly Seminar Series on Legal Informatics, 2021.

10. "Database Systems and Data Structures" at the Delhi Public School Teacher Seminar, 2021.

11. "Ensemble Methods and their Applications for Fog Prediction" as part of Expert Talk Series at CDAC Patna, 2021.

12. "Computer Science and Artificial Intelligence" at the Delhi Public School, Kalyanpur, Kanpur, 2021.

13. "Big Data Analytics and Machine Learning" at the Machine Learning Workshop at SPGC conference by ONGC, 2020.

14. "Querying and Mining of Graphs" at the Machine Learning Summit at Data Labs, CapitalOne, 2019.

15. "Querying Statistically Significant Subgraphs" at the NetApp Corporation, Bengaluru, India, 2019.

16. "Graph Querying using Statistical Significance" at the Indian Institute of Science, Engineering and Technology, Shibpur, 2019.

17. "Data Mining" at the Andhra Pradesh Human Resource Development Institute (APHRDI), 2018.

18. "Trajectory Aware Service Location Problems" at the NetApp Corporation, Bengaluru, India, 2016.

19. "Mining Statistically Significant Connected Subgraphs" at the NetApp Corporation, Bengaluru, India, 2015.

20. "Mining Statistically Significant Substructures based on the Chi-square Statistic" at IBM, New Delhi, India, 2015.

21. "Mining Statistically Significant Substrings based on the Chi-square Measure" at the NetApp Corporation, Bengaluru, India, 2014.

22. "Mining Statistically Significant Substructures based on the Chi-square Statistic" at the Indian Statistical Institute, Kolkata, India, 2014.

23. "Skylines: Databases' Answer to Multiple Preferences" at the NetApp Corporation, Bengaluru, India, 2013.

24. "Skylines: Databases' Answer to Multiple Preferences" at the Dept. of Computer Science and Engineering, Indian Institute of Technology, Kanpur, India, 2012.

25. "Finding the Bias and Prestige of Nodes in Networks based on Trust Scores" at Yahoo! Labs, Bengaluru, India, 2011.

26. "Earth Mover's Distance: An Adaptable and Universally Applicable Distance Measure" at the Dept. of Computer Science, Andhra University, Vishakhapatnam, India, 2010.

27. "Earth Mover's Distance: An Adaptable and Universally Applicable Distance Measure" at Tata Consultancy Services (TCS), Gurgaon, India, 2010.

28. "On Earth Mover's Distance: A Spatially Sensitive Distance Measure" at the Dept. of Computer Science, Free University of Bozen-Bolzano, Italy, 2009.

29. "Distributed Indexing and Querying in Sensor Networks using Statistical Models" at the Dept. of Computer Science, Université Libre de Bruxelles, Belgium, 2008.

Patents:

1. Multiple Criteria Decision Analysis

   - US patent number US8504581B2: 2013
   - India patent number INDEL20123027A: 2014

2. Multiple Criteria Decision Analysis in Distributed Databases

   - Global patent number WO2015104591A1: 2015

Important Courses Taught:

1. Computational Linguistics for Indian Languages

2. Data Mining

3. Indexing and Searching Techniques in Databases

4. Information Retrieval

5. Skyline Queries in Databases

6. Data-Driven Program Analysis

7. Topics in Biocomputing

8. Artificial Intelligence, Machine Learning and Deep Learning

9. Principles of Database Systems

10. Fundamentals of Computing

11. Computing Laboratory

Awards, Scholarships and Certificates:

1. IBM Faculty Research Award, 2014.

2. Award from Yahoo! Faculty Research and Engagement Program, 2011.

3. Best paper award at the International Conference on Management of Data (COMAD), 2011 for the paper "Minimally Infrequent Itemset Mining using Pattern-Growth Paradigm and Residual Trees".

4. Best student paper award at the International Conference on Management of Data (COMAD), 2010 for the paper "Aggregate Skyline Join Queries: Skylines with Aggregate Operations over Multiple Relations".

5. One of the top-five student paper awards at the International Conference on Data Mining (ICDM), 2005 for the paper "ViVo: Visual Vocabulary Construction for Mining Biomedical Images".

6. ICDM Student Travel Award sponsored by IBM at the International Conference on Data Mining (ICDM), 2005 awarded to the top five student papers.

Major Sponsored Projects:

1. "Sanskrit Knowledge Accessor" from MeitY, Govt. of India, 2024-2026.

2. "Forecasting Weather and Climate using Dynamics and Artificial Intelligence" from DORA, IIT Kanpur, 2024-2026.

3. "Enhancing Disaster Resilience through Innovative Fog Monitoring and Forecasting" from Coalition for Disaster Resilient Infrastructure (CDRIN), 2024-2025.

4. "Advancement of NLP Techniques for Indian Languages with Focus on Bangla and Hindi" from SERB, 2024-2026.

5. "Designing IKS Courses and Training Teachers on Computational Linguistics, Tarka and Science of Kalā" from IKS division, AICTE, 2023-2025.

6. "Gaṇita, Darśana, Nīti, Kalā and Āyurveda: Towards Integration of IKS into Holistic School Education" from IKS division, AICTE, 2023-2024.

7. "Text-to-Text Translation among Indian Languages using Sanskrit-based Interlingua Representation" from MeitY, Govt. of India, 2022-2025.

8. "Automated Question-Answering System for Ramayana" from IKS division, AICTE, 2022-2024.

9. "Creation of Science and Technology Content for Indic Wikipedia" from National Council for Science and Technology Communication (NCSTC), DST, 2020-2023.

10. "Short and Long-term Fog Predictions using Data Science" under STARS scheme from MHRD, 2020-2023.

11. "Scalable Spatio-Temporal Measurement and Analysis of Air Pollution Data for Delhi-NCR using Vehicle Mounted Sensors" under IMPRINT-II scheme from SERB, 2019-2022.

12. "NYAYA: A Legal Assistance System for Legal Experts and the Common Man in India" from SERB, India, 2019-2022.

13. "Continuous Monitoring of Sampreeti Setu (New Jubilee Bridge): Instrumentation, Design and Health Assessment" from Eastern Railway Zone, Indian Railways, 2018-2023.

14. "Development of Optimal Eco Driving System in HEV/PHEV based on Vehicle Environment" from KEIIT (Korea Evaluation Institute of Industrial Technology), 2018-2021.

15. "Development of Novel Materials and Methods for Removal of Relcalcitrant Organics from Water" from Indo-Taiwan Programme in Science and Technology, 2017-2020.

16. "Provenance in Graph Databases" from IBM, India, 2016-2018.

17. "Identifying Fake Product Listings and Sellers" from Flipkart, India, 2016-2018.

18. "Mining Statistically Significant Substructures using the Chi-Square Measure and Setup of Big Data Lab" from IBM, India, 2014-2017.

19. "Extending Skyline Queries to Distributed and Uncertain Databases" from SERB, India, 2014-2017.

20. "Development of Air Quality Index (AQI) for Indian Cities" from Central Pollution Control Board (CPCB), India, 2014-2015.

21. "Deciphering the BMP Signaling Network in Developing Bone: An Inter-disciplinary Approach Combining Bioinformatic Data Mining Tools along with Molecular, Genetic and Developmental Biology" from DBT, Govt. of India, 2013-2016.

22. "Flash-Aware Optimizations for Columnar Databases" from NetApp Corporation, 2011-2018.

23. "Reputation Framework for Ad Networks" from Yahoo! Research, 2011-2011.

Students Advised:

- Ph.D. students: Graduated: 5, In progress: 6
- M.Tech. and M.S. students: Graduated: 101, In progress: 9

Chairs and Organizers:

- Program Chair for the 7th International Sanskrit Computational Linguistics Symposium (IS-CLS), 2024.

- Publications Chair for the 2nd International Conference on AI-ML Systems (AIML), 2022.

- Program Chair for the 27th International Conference on Database Systems for Advanced Applications (DASFAA), 2022.

- Publications Chair for the 1st International Conference on AI-ML Systems (AIML), 2021.

- Program Chair for the ACM India Joint International Conference on Data Science & Management of Data (CoDS-COMAD), 7th ACM IKDD CODS and 25th COMAD, 2020.

- Organizer for FIRE 2019 AILA Track: Artificial Intelligence for Legal Assistance at Forum for Information Retrieval (FIRE), 2019.

- Workshop Organizer of 2nd Joint International Workshop on Graph Data Management Experiences & Systems (GRADES) and Network Data Analytics (NDA) co-located with SIGMOD 2019.

- Workshop Organizer of 1st Workshop on Legal Data Analytics and Mining (LeDAM) co-located with CIKM 2018.

- Workshop Organizer of 1st Joint International Workshop on Graph Data Management Experiences & Systems (GRADES) and Network Data Analytics (NDA) co-located with SIGMOD 2018.

- Workshop Organizer of 2nd International Workshop on Network Data Analytics (NDA@SIGMOD) co-located with SIGMOD 2017.

- Organizer for FIRE 2017 IRLeD Track: Information Retrieval from Legal Documents at Forum for Information Retrieval (FIRE), 2017.

- Program Chair for the 19th International Conference on Management of Data (COMAD), 2013.

- Publications Chair for the 18th International Conference on Management of Data (COMAD), 2012.

- Program Chair for the 8th International Conference on Simulated Evolution and Learning (SEAL), 2010.

- Publicity Chair for the 14th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), 2010.

Professional Activities:

- Jury for Artificial Intelligence category for Aegis Graham Bell Awards, 2021-22.

- Panelist for Webinar on AI and 5G organized by NIELIT, 2021.

- Editor in Graph Data Mining and Management part of the journal Frontiers in Big Data since 2021.

- Panelist for Webinar on AI organized by NIELIT and MEITY, 2020.

- Associate Editor for the journal Distributed and Parallel Databases (DAPD) since 2020.

- Review Editor in Data Mining and Management part of the journal Frontiers in Big Data since 2018.

- Executive Member of the Computer Society of India's (CSI) Special Interest Group in Data (SIGDATA) since 2012.

- Member of the CODATA National Committee since 2016.

- Member of the Association for Computing Machinery (ACM) since 2010.

- Member of the Institute of Electrical and Electronics Engineers (IEEE) since 2005.

- Reviewer and Program Committee member for many international journals and conferences, including VLDB, ICDE, TKDD, etc.

- Panelist for discussions on AI and Its Impact on Jobs at IIT Kanpur.