

Soumyadeep Roy

Complex Networks Research Lab, Department of Computer Science and Engineering, IIT Kharagpur, Kharagpur - 721302, West Bengal, India
soumyadeep.roy9@iitkgp.ac.in • <http://www.datanalytics101.com/> • github.com/roysoumya

EDUCATION

Indian Institute of Technology Kharagpur, West Bengal, India

- Ph.D. Candidate in Computer Science and Engineering Jan 2020 – Present
 - Thesis: Incorporating Domain Knowledge in Medical NLP Applications
 - Advisor: Prof. Niloy Ganguly and Prof. Shamik Sural
 - Focus areas: Medical NLP, Representation Learning for Genomics, Few-shot Learning
 - Supported by Institute Ph.D. Fellowship.
- M.S. (Research) in Computer Science and Engineering Jul 2017 – Nov 2019
 - Thesis: Computational Approaches for Online Reputation Monitoring [Modeling, Analysis and Recommendation]
 - Advisor: Prof. Niloy Ganguly and Prof. Shamik Sural
 - Cumulative GPA: 9.21 / 10.00 ; [Thesis](#)

Kalyani Government Engineering College, Kalyani, West Bengal, India

- B.Tech in Computer Science and Engineering Jul 2013 – Jun 2017
 - Thesis : Understanding Email Interactivity and Predicting User Response to Email
 - Advisor: Dr. Kousik Dasgupta and Mr. Binay Gupta
 - Cumulative GPA: 8.61 / 10.00

RESEARCH EXPERIENCE

L3S Research Center, Leibniz University Hannover, Germany

- Research Associate in *Leibniz AI Future Lab* Jan 2021 – Jun 2023
 - Advisor: Prof. Niloy Ganguly and Prof. Wolfgang Nejdl
 - Project: Data-driven patient subtyping of Parkinson's Disease

Indian Institute of Technology Kharagpur, West Bengal, India

- Senior Research Fellow in Dept. of Computer Science and Engineering May 2019 – Dec 2019
- Junior Research Fellow in Dept. of Computer Science and Engineering Jun 2017 – May 2019
 - Member of Complex Networks Research Group (CNeRG)
- Indian Academy of Sciences Summer Research Fellowship Programme Jun 2016 – Jul 2016
 - Project: Anomaly Detection and Summarization of various Extreme Weather Events over Indian sub-continent
 - Guide: Prof. Sudeshna Sarkar, Dept. of Computer Science and Engineering
 - [Technical Report](#), [Code](#)

Adobe Big Data Experience Lab, Bangalore, India

- Summer Research Internship May 2018 – Jun 2018
 - Project: Predicting Brand Personality from Web Content
 - Mentors: Mr. Anandhavelu Natarajan, Dr. Niyati Chhaya

PUBLICATIONS

JOURNALS

- [J.2] S. Roy, S. Mücke, M. Marscholke, H. Frieling, N. Ganguly and D. Wolff, “Decision Tree-based Approach to Robust Parkinson’s Disease Subtyping using Clinical Data of the Michael J. Fox Foundation LRRK2 Cross-sectional Study,” *Work-in-progress, collaboration with Medical School Hannover, Germany*
- [J.1] S. Roy, S. Sural, N. Chhaya, A. Natarajan and N. Ganguly, “An Integrated Approach for Improving Brand Consistency of Web Content: Modeling, Analysis and Recommendation,” in *ACM Transactions on the Web*, volume 15, article 2, Article 9 (May 2021), 25 pages. [doi](#) [arXiv code and data](#)

CONFERENCES

- [C.6] S. Roy, J. Wallat, S. Sundaram, W. Nejdl and N. Ganguly, “GeneMask: Fast Pretraining of Gene Sequences to Enable Few-Shot Learning,” in the *26th European Conference on Artificial Intelligence (ECAI 2023)*, 8 pages, Krakow, Poland, Oct 2023, doi: <https://doi.org/10.3233/FAIA230492> [arXiv](#) [code](#) [slides](#) **Received 3rd prize in poster presentation of IndoML 2023 at IIT Bombay**

- [C.5] S. Roy, N. Ganguly, S. Sural and K. Rudra, “Interpretable Clinical Trial Search using Pubmed Citation Network,” in the *2023 International Conference on Digital Health (ICDH)*, pp. 328 - 338, Chicago, United States, Jul 2023, doi: 10.1109/ICDH60066.2023.00056 [pdf](#) [code](#) [slides](#)
Candidate for Best Student Paper
- [C.4] S. Roy, S. Chakraborty, A. Mandal, P. Sharma, G. Balde, A. Natarajan, M. Khosla, S. Sural and N. Ganguly, “Developing Knowledge-Aware Neural Models for Medical Forum Question Classification,” in the 30th ACM International Conference on Information and Knowledge Management, 1-5 November 2021, Online, 4 pages [doi](#) [arXiv](#) [code](#) [slides](#) [video](#)
- [C.3] S. Roy, K. Rudra, N. Agrawal, S. Sural and N. Ganguly, “Towards an Aspect-Based Ranking Model for Clinical Trial Search,” in the *8th International Conference on Computational Data and Social Networks (CSoNet 2019)*, pp. 209 - 222, Ho Chi Minh City, Vietnam, Nov 2019. [doi](#) [pdf](#) [code](#) [slides](#)
- [C.2] S. Roy, N. Ganguly, S. Sural, N. Chhaya and A. Natarajan, “Understanding Brand Consistency from Web Content,” in *Proceedings of the 10th ACM Conference on Web Science (WebSci 2019)*, pp. 245 - 253, Boston, MA, USA, Jul 2019. [doi](#) [pdf](#), [data](#), [slides](#)
- [C.1] S. Roy, “Automated EBM-oriented Summarization of Active or Recruiting Clinical Trials,” in the *10th International Conference on COMMunication Systems & NETWORKS (COMSNETS 2018)*, 2 pages, Bangalore, India, Jan 2018. **Best Graduate Forum Presentation Award**

BOOK CHAPTER

- [B.1] S. Roy, N. Pal, K. Dasgupta, and B. Gupta, “Understanding Email Interactivity and Predicting User Response to email,” in *Methodologies and Application Issues of Contemporary Computing Framework (CICBA 2018)*, pp. 69 - 79, West Bengal, India, Jul 2018. [doi](#) [pdf](#), [code](#), [slides](#)

PROFESSIONAL ACTIVITIES (TALKS/ TRAVEL GRANTS/ REVIEWING)

- Reviewed for Core A* conferences: EMNLP 2022, AAAI 2023, ACL 2023, EMNLP 2023, AAAI 2024. Also reviewed for EACL 2023 (Core A), IEEE Transactions on Cognitive and Developmental Systems
- Received the Microsoft Research Travel Grant for attending ECAI 2023 at Krakow, Poland from Sept. 30 to Oct. 5, 2023
- Performed the role of *Subject Matter Expert* for the “Data Science and Machine Learning Track” of Pycon India 2020, and part of “Mentorship Team” of Pycon India 2021.
- Performed the role of PC member of ECML-PKDD 2020 Applied Data Science Track.
- Presented the ACM WebSci 2019 paper titled *Understanding Brand Consistency from Web Content* at the “Out-of-India Exhibition” track of India HCI 2019 held at the Indian School of Business, Hyderabad, India on November 1 to 3, 2019
- Received IIT Kharagpur’s *Institute Travel Grant* and CNeRG Travel Grant, for attending ACM WebSci 2019 at Northeastern University, Boston, MA, United States from July 1 - 6, 2019.
- Performed the role of *CodeChef Campus Chapter Representative* of Kalyani Government Engineering College from 2015 to 2017. Co-founded the college programming club, KeyGENcoders, which is still running to date.

TEACHING (TA-SHIP)

- **IIT Kharagpur:** Programming and Data Structure Lab (Spring 2018), Department Website Team (Autumn 2018 to Spring 2020), Information Retrieval (Autumn 2020)
- **Leibniz University Hannover, Germany:** Natural Language Processing (Winter 2021), Foundations of Information Retrieval (Summer 2022, 2023)

MENTORING

- **3 M.Tech Thesis:** (i) *Multidimensional Retrieval and Ranking Model for Clinical Trials Search* by Nikhil Agrawal (2019), (ii) *Query-Focused Summarization in Medical Domain: A Comparative Study* by Gunjan Balde (2020), and (iii) *Understanding Medical Knowledge Graph for NLP Applications* by Sourav Saha (2021)
- **1 B.Tech Thesis** titled *Improving Question Generation of Questions Answering Based Summary Evaluation Metric* by Sudip Chakraborty (2022)

REFERENCES

Can be provided on request.