

# Dwaipayan Roy

Assistant Professor

Department of Computational and Data Sciences  
Indian Institute of Science Education and Research, Kolkata  
✉ [dwaipayan.roy@iiserkol.ac.in](mailto:dwaipayan.roy@iiserkol.ac.in)  
📄 <https://dwaipayanroy.github.io/>

## Research Area

Information Retrieval, Text Processing

## Sub Area

- Word Embedding for Information Retrieval
- Scholarly Citation Recommendations
- Retrieval Models for Information Retrieval
- Knowledge Graph and Knowledge Base

## Experience

- 2020 - **Assistant Professor**, Indian Institute of Science Education and Research, Kolkata.
- 2019 - 2020 **Post-doctoral researcher**, *GESIS - Leibniz Institute for the Social Sciences*, Cologne, Germany.
- 2019 **Visiting Faculty**, *Ramakrishna Mission Vivekananda Educational and Research Institute*.
- Aug - Nov'18 **Visiting Scholar**, *IBM India Research Lab*, Delhi, India.
- 2014 - 2019 **Senior Research Fellow**, *Indian Statistical Institute*, Kolkata.
- 2014 - 2016 **Visiting Faculty**, *Ramakrishna Mission Vidyamandira*, Belurmath.
- Sep - Nov'16 **Visiting Researcher**, *Dublin City University*, Dublin, Ireland.
- May - Jul'16 **Visiting Scholar**, *Dublin City University*, Dublin, Ireland.

## Publications

**Roy, D., Bhatia, S., Jain, P. Information Asymmetry in Wikipedia Across Different Languages: A Statistical Analysis.** to appear at *Journal of the Association for Information Science and Technology (JASIST)*.

*Krämer, T., Carevic, Z., Roy, D., Klas, C., Mayr, P. ConSTR: A Contextual Search Term Recommender. ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL 2021) - accepted, to appear.*

*Carevic, Z., Roy, D., Mayr, P. Characteristics of Dataset Retrieval Sessions: Experiences from a Real-life Digital Library. 24th International Conference on Theory and Practice of Digital Libraries (TPDL 2020), pp. 185–193.*  
[https://doi.org/10.1007/978-3-030-54956-5\\_14](https://doi.org/10.1007/978-3-030-54956-5_14)

*Datta, S., Greene, D., Ganguly, D., Roy, D., Mitra, M. Where's the Why? In Search of Chains of Causes for Query Events. 28th Irish Conference on Artificial Intelligence and Cognitive Science (AICS 2020).*  
[http://ceur-ws.org/Vol-2771/AICS2020\\_paper\\_48.pdf](http://ceur-ws.org/Vol-2771/AICS2020_paper_48.pdf)

Datta, S., Ganguly, D., **Roy, D.**, Bonin, F., Jochim, C., Mitra, M. **Retrieving Potential Causes from a Query Event.** *43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2020)*, pp. 2949–2952.  
<https://doi.org/10.1145/3397271.3401207>

**Roy, D.**, Bhatia, S., Jain, P **A Topic-Aligned Multilingual Corpus of Wikipedia Articles for Studying Information Asymmetry in Low Resource Languages..** *12th Edition of its Language Resources and Evaluation Conference (LREC 2020)*, pp. 2373–2380.  
<https://doi.org/10.1145/3357384.3357859>

**Roy, D.**, Saha, S., Mitra, M., Sen, B., Ganguly, D. **I-REX: A Lucene Plugin for EXplainable IR.** *28th ACM International Conference on Information and Knowledge Management (CIKM 2019)*, pp. 2949–2952.  
<https://doi.org/10.1145/3357384.3357859>

Biswas, C., Ganguly, D., **Roy, D.** **Privacy Preserving Approximate K-means Clustering.** *28th ACM International Conference on Information and Knowledge Management (CIKM 2019)*, pp. 1321–1330.  
<https://doi.org/10.1145/3357384.3357969>

**Roy, D.**, Bhatia, S. **Analysis of Information Gap Across Different Language Editions of Wikipedia.** *Unpublished.*

**Roy, D.**, Bhatia, S., Mitra, M. **Selecting Discriminative Terms for Relevance Model.** *42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR-2019)*, pp. 1253–1256.  
<https://doi.org/10.1145/3331184.3331357>

Ganguly, D., Afti, H., **Roy, D.** **Word Embedding based Semantic Cross-Lingual Document Alignment in Comparable Corpora.** *10th Annual Meeting of the Forum for Information Retrieval Evaluation (FIRE 2018)*, pp. 28–34.  
<https://doi.org/10.1145/3293339.3293346>

**Roy, D.**, Ganguly, D., Bhatia, S., Bedathur, S., Mitra, M. **Using Word Embeddings for Information Retrieval: How Collection and Term Normalization Choices Affect Performance.** *27th ACM International Conference on Information and Knowledge Management (CIKM 2018)*, pp. 1835–1838.  
<https://doi.org/10.1145/3269206.3269277>

**Roy, D.**, Ganguly, D., Mitra, M., Jones, G. **Estimating Gaussian Mixture Models in the Local Neighbourhood of Embedded Word Vectors for Query Performance Prediction.** *Elsevier Journal on Information Processing & Management (IP&M) – Volume 56, Issue 3*, pp. 1026–1045.  
<https://doi.org/10.1016/j.ipm.2018.10.009>

**Roy, D.**, Mitra, M., Ganguly, D. **To Clean or not to Clean: Document Pre-processing and Reproducibility.** *ACM Journal of Data and Information Quality – Volume 10 Issue 4, Article No. 18*, pp. 18:1–18:25.  
<https://doi.org/10.1145/3242180>

**Roy, D.** An Improved Test Collection and Baselines for Bibliographic Citation Recommendation. *26th ACM International Conference on Information and Knowledge Management (CIKM 2017)*, pp. 2271–2274.  
<https://doi.org/10.1145/3132847.3133085>

**Roy, D.** Word Embedding based Approaches for Information Retrieval. *Seventh BCS-IRSG Symposium on Future Directions in Information Access (FDIA 2017)*, pp. 1–4.  
<https://dx.doi.org/10.14236/ewic/FDIA2017.9>

**Roy, D., Ganguly, D., Mitra, M., Jones, G.** Word Embedding based Relevance Feedback using Kernel Density Estimation. *25th ACM International Conference on Information and Knowledge Management (CIKM 2016)*, pp. 1281–1290.  
<https://doi.org/10.1145/2983323.2983750>

**Roy, D., Ganguly, D., Mitra, M., Jones, G.** Representing Documents and Queries as Sets of Word Embedded Vectors for Information Retrieval. *SIGIR 2016 Workshop on Neural Information Retrieval (NeuralIR – 2016)*.  
<https://www.microsoft.com/en-us/research/event/neuir2016/>

**Roy, D., Paul, D., Garain, U., Mitra, M.** Using Word Embeddings for Automatic Query Expansion. *SIGIR 2016 Workshop on Neural Information Retrieval (NeuralIR – 2016)*.  
<https://www.microsoft.com/en-us/research/event/neuir2016/>

**Roy, D., Ray, K., Mitra, M.** From a Scholarly Big Dataset to a Test Collection for Bibliographic Citation Recommendation. *Scholarly Big Data: AI Perspectives, Challenges, and Ideas - AAAI'16 Workshop*, pp. 705–710.  
<https://www.aaai.org/ocs/index.php/WS/AAAIW16/paper/view/12635>

Ganguly, D., **Roy, D.**, Mitra, M., Jones, G. Word Embedding based Generalized Language Model for Information Retrieval. *38th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR-2015)*, pp. 795–798.  
<https://doi.org/10.1145/2766462.2767780>

Bandyopadhyay, A., **Roy, D.**, Mitra, M., Saha, S. Named Entity Recognition from Tweets. *16th LWA workshops: KDML, IR and FGWM (LWA 2014)*, pp. 218–225.  
<http://ceur-ws.org/Vol-1226/paper33.pdf>

**Roy, D., Bandyopadhyay, A., Mitra, M.** A Simple Context Dependent Suggestion System. *21st Text Retrieval Conference (TREC 2013)*.  
<https://trec.nist.gov/trec/pubs/trec22/papers/isi-context.pdf>

Sing, J.K., **Roy, D.**, Basu, D.K., Nasipuri, M. Generalized Diagonal 2D FLDA for Efficient Face Recognition. *International Conference on Communications, Devices and Intelligent Systems (CODIS-2012)*, pp. 636–639.  
<https://doi.org/10.1109/CODIS.2012.6422283> (Master's thesis)

- **INDCON 2020-2021 - Industry Academia Conclave** - invited panelist.
- **AICTE sponsored STTP on Machine Learning - Word Embeddings and Information Retrieval** - invited speaker.
- **Improving Information Retrieval Performance using Word Embedding: Invited talk at TBML 2018** - invited speaker.
  - Theoretical Basis of Machine Learning - 2018 (TBML - 2018), International Centre for Theoretical Sciences, Bengaluru
- **Word Embedding and Information Retrieval:** Presented at:
  - Indian Statistical Institute, Kolkata
- **A Survey of Information Retrieval Approaches with Embedded Word Vectors: Tutorial at ICON 2017:** Presented at:
  - 14<sup>th</sup> International Conference on Natural Language Processing (ICON - 2017), Jadavpur University, Kolkata
- **Introduction to HDFS and MapReduce:** Presented at:
  - Ramakrishna Mission Vivekananda University
  - Indian Statistical Institute, CVPR Departmental Colloquium
  - Department of Computer Science, Jadavpur University

---

## Academic Involvements

### Organizing Committee

- ACM International Conference on Information and Knowledge Management (CIKM-2020).
- Forum for Information Retrieval Evaluation (FIRE) 2012 - 2021.

### Program Chair

- ACL Empirical Methods in Natural Language Processing (EMNLP)
- ACM International Conference on Web Search and Data Mining (WSDM)
- ACM International Conference on Special Interest Group on Information Retrieval (SIGIR)
- ACM International Conference on Information and Knowledge Management (CIKM)
- ACM India Joint International Conference on Data Science & Management of Data (CODS-COMOD)
- Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT)
- Information Processing & Management (Elsevier Journal)
- ACM Conference on Research and Development in Information Retrieval (SIGIR)
- International Conference on Natural Language Processing (ICON)
- Forum for Information Retrieval Evaluation (FIRE)
- International Conference on Emerging Applications of Information Technology (EAIT)
- International Conference on Computational Linguistics (COLING)

---

## Academic Visits

- 2018 August - October **IBM Research, Delhi, India:** Visiting Researcher
- 2016 October - November **Dublin City University, Dublin Ireland:** Visiting Researcher

- 2015 May - June **Dublin City University, Dublin Ireland:** Visiting Scholar

---

## Academic Achievements

- Received **ACM SIGIR Travel Grant** to attend CIKM 2018 in Turin, Italy.
- Received **Microsoft Research Travel Grant** to attend CIKM 2017 in Singapore.
- Received **Google Research Travel Grant** to attend CIKM 2016 in Indianapolis, USA.
- Received **Microsoft Research Travel Grant** to attend CIKM 2016 in Indianapolis, USA.
- Received **ACM SIGIR Travel Grant** to attend CIKM 2016 in Indianapolis, USA.
- Received **ISI Travel Grant** to attend AAAI 2016 in Arizona, US.
- Received **ACM SIGIR Travel Grant** to attend SIGIR 2015 in Santiago, Chile.
- Receiving **ISI Senior Research Fellowship** from 2014 till date
- Received **ISI Junior Research Fellowship** from 2012 to 2014
- Qualified All India Research Fellowship examination organized by Indian Statistical Institute.
- Qualified in West Bengal Joint Entrance for Computer Applications, ranked among top 10.
- Stood first in B.Sc. Computer Applications in Calcutta University, 2006-2009 batch.

---

## Teaching Experiences

- *Course Instructor* -
  - **Programming and Data Structures - II** - Spring 2021 - IISER, Kolkata.
  - **Introduction to Computer Programming** - Autumn 2020 - IISER, Kolkata.
  - **Programming in Python** - Autumn 2020, 2021 - IISER, Kolkata.
  - **Data Structures II** - Fall 2014 - 2016 - Ramakrishna Mission Vidyamandira, Belurmath.
  - **Information Retrieval** - Autumn 2019 - Ramakrishna Mission Vivekananda Research Institute, Belurmath.
- *Teaching Assistant* -
  - **Programming and Data Structures Lab** - Fall 2014 - Indian Statistical Institute, Kolkata

---

## Technical Skills Expertise

**Languages** Expert: C, Java, Script **Familiar:** C++, Python

**Platforms** Expert: Unix **Familiar:** Windows

**Tools** Expert: Lucene **Familiar:** Elastic Search

---

## Education

2019 June **PhD in Computer Science**, *Indian Statistical Institute, Kolkata.*

2014–2019 **Senior Research Fellow**, *Indian Statistical Institute, Kolkata.*

2012–2014 **Junior Research Fellow**, *Indian Statistical Institute, Kolkata.*

2009–2012 **Master of Computer Applications**, *Jadavpur University*.  
2006–2009 **B.Sc.(Major) in Computer Applications**, *University of Calcutta*.  
2006 **Higher Secondary(10+2)**, *W.B.C.H.S.E.*  
2004 **Secondary(10)**, *W.B.B.S.E.*

---

## References

- **Dr. Mandar Mitra** Associate Professor, Indian Statistical Institute, Kolkata, India.  
email: *mandar@isical.ac.in*
- **Prof. Gareth Jones** Professor, Dublin City University, Dublin, Ireland.  
email: *gareth.jones@dcu.ie*
- **Dr. Debasis Ganguly** Lecturer, University of Glasgow, Scotland.  
email: *Debasis.Ganguly@glasgow.ac.uk*
- **Dr. Sumit Bhatia** Machine Learning Scientist, Adobe Inc. India.  
email: *Sumit.Bhatia@adobe.com*
- **Dr. Philipp Mayr-Schlegel** Researcher, Team leader (Information & Data Retrieval),  
GESIS - Leibniz-Institute for the Social Sciences, Cologne, Germany.  
email: *Philipp.Mayr@gesis.org*