

RISHIRAJ SAHA ROY

Post-doctoral Researcher
Max Planck Institute for Informatics (MPII), Saarbruecken, Germany
Advisor: [Prof. Dr. Gerhard Weikum](#), MPII
[November 2015 - Present]



Computer Scientist, Adobe Research, Bangalore, India
[April 2014 – September 2015]

Ph.D., *Microsoft Research Lab India Ph.D. Fellow*
Computer Science and Engineering, Indian Institute of Technology (IIT) Kharagpur, India
Advisors: [Prof. Niloy Ganguly](#), IIT Kharagpur, and [Dr. Monojit Choudhury](#), Microsoft Research
[December 2009 – March 2014]

E-mail : rishiraj@mpi-inf.mpg.de
Web : <http://people.mpi-inf.mpg.de/~rsaharo/>
Google Scholar : <https://goo.gl/tQo9TN>
DBLP : https://dblp.uni-trier.de/pers/hd/r/Roy:Rishiraj_Saha

RESEARCH INTERESTS

- **Current areas:** [Question-answering over knowledge bases and text](#), User-oriented online privacy and transparency
- **Ph.D. areas:** Query understanding for Web search (Query segmentation, intent, and complexity) (**Thesis:** [Unsupervised Approaches to Syntactic Analysis of Web Search Queries](#) [[Slides](#)])
- **General interests:** Information retrieval, Computational linguistics, Text mining

PUBLICATION SUMMARY

Conference long papers [WSDM '19](#), [WWW '18](#), [ICDAR '17](#), [SIGIR '17](#), [CICLing '15](#), [Coling '14](#), [Evolang '14](#), [ACL '13](#), [FIRE '13](#), [SIGIR '12](#), [Evolang '12](#), [ITC '10](#), [IKE '09](#), [SEDE '09](#)

Conference short papers [CIKM '18](#), [IJCNLP '17](#), [CIKM '17](#), [EMNLP '17](#), [CICLing '17](#), [ECIR '17](#), [ECIR '16](#), [WWW '15](#), [SIGIR '14](#), [WWW '13](#), [ECCS '13](#), [WWW '11](#)

Journals [IP&M '16](#), [Web Semantics '15](#), [JCMSE '10](#), [IJACT '09](#)

Others [HQA '18 \(WWW\)](#), [CBRecSys '14 \(RecSys\)](#), [WWW '13 DC](#), [QRU '11 \(SIGIR\)](#)

CONFERENCE LONG PAPERS

- Azin Ghazimatin, **Rishiraj Saha Roy**, and Gerhard Weikum, "FAIRY: A Framework for Understanding Relationships between Users' Actions and their Social Feeds", in *Proceedings of the 12th ACM International Conference on Web Search and Data Mining 2019 (WSDM '19)*, 11 – 15 February 2019, Melbourne, Australia (to appear).
- Abdalghani Abujabal, **Rishiraj Saha Roy**, Mohamed Yahya, and Gerhard Weikum, "[Never-Ending Learning for Open-Domain Question Answering over Knowledge Bases](#)", in

Proceedings of the 27th International World Wide Web Conference 2018 (WWW '18), 23 - 27 April 2018, Lyon, France, pages 1053 - 1062. [[Slides](#)] [[BibTeX](#)]

- **Rishiraj Saha Roy**, Abhijeet Singh, Prashant Chawla, Shubham Saxena, and Atanu R. Sinha, "[Automatic Assignment of Topical Icons to Documents for Faster File Navigation](#)", in *Proceedings of the 14th IAPR International Conference on Document Analysis and Recognition 2017 (ICDAR '17)*, 9 - 15 November 2017, Kyoto, Japan. [[Data](#)] [[Poster](#)] [[BibTeX](#)]
- Asia Biega, **Rishiraj Saha Roy**, and Gerhard Weikum, "[Privacy through Solidarity: A User-Utility-Preserving Framework to Counter Profiling](#)", in *Proceedings of the 40th Annual ACM SIGIR Conference on Research and Development on Information Retrieval 2017 (SIGIR '17)*, 7 - 11 August 2017, Tokyo, Japan, pages 675 - 684. [[Slides by Asia Biega](#)] [[BibTeX](#)]
- **Rishiraj Saha Roy**, Aishwarya Padmakumar, Guna Prasaad Jeganathan and Ponnurangam Kumaraguru, "[Automated Linguistic Personalization of Targeted Marketing Messages Mining User-generated Text on Social Media](#)", in *Proceedings of the 16th International Conference on Intelligent Text Processing and Computational Linguistics 2015 (CICLing '15)*, 14 - 20 April 2015, Cairo, Egypt, pages 203 - 224, Volume 9042, Lecture Notes in Computer Science, Springer. [**BEST PAPER AWARD**] [[Slides](#)] [[BibTeX](#)] [[Data](#)]
- **Rishiraj Saha Roy**, Rahul Katare, Niloy Ganguly and Monojit Choudhury, "[Automatic Discovery of Adposition Typology](#)", in *Proceedings of the 25th International Conference on Computational Linguistics 2014 (Coling '14)*, 23 - 29 August 2014, Dublin, Ireland, pages 1037 - 1046. [[Poster](#)] [[BibTeX](#)]
- **Rishiraj Saha Roy**, M. Dastagiri Reddy, Niloy Ganguly and Monojit Choudhury, "[Understanding the Linguistic Structure and Evolution of Web Search Queries](#)", in *Proceedings of the 10th International Conference on the Evolution of Language 2014 (Evolang X)*, 14 - 17 April 2014, Vienna, Austria, pages 286 - 293. [[Slides](#)] [[BibTeX](#)]
- **Rishiraj Saha Roy**, Monojit Choudhury, Prasenjit Majumder and Komal Agarwal, "[Overview of the FIRE 2013 Track on Transliterated Search](#)", in *Proceedings of the Fifth Forum for Information Retrieval Evaluation 2013 (FIRE '13)*, 04 - 06 December 2013, New Delhi, India, pages 1 - 7. [[Slides](#)] [[BibTeX](#)] [[Data](#)]
- Rohan Ramanath, Monojit Choudhury, Kalika Bali and **Rishiraj Saha Roy**, "[Crowd Prefers the Middle Path: A New IAA Metric for Crowdsourcing Reveals Turker Biases in Query Segmentation](#)", in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics 2013 (ACL '13)*, 4 - 9 August 2013, Sofia, Bulgaria, pages 1713 - 1722. [[Slides](#)] [[Supplementary material](#)] [[Data](#)] [[BibTeX](#)]
- **Rishiraj Saha Roy**, Niloy Ganguly, Monojit Choudhury and Srivatsan Laxman, "[An IR-based Evaluation Framework for Web Search Query Segmentation](#)", in *Proceedings of the 35th Annual ACM SIGIR Conference on Research and Development on Information Retrieval 2012 (SIGIR '12)*, 12 - 16 August 2012, Portland, USA, pages 881 - 890. [[Slides](#)] [[Data](#)] [[BibTeX](#)]
- **Rishiraj Saha Roy**, Monojit Choudhury and Kalika Bali, "[Are Web Search Queries an Evolving Protolanguage?](#)", in *Proceedings of the 9th International Conference on the Evolution of Language 2012 (Evolang IX)*, 13 - 16 March, 2012, Kyoto, Japan, pages 304 - 311. [**BEST POSTER AWARD**] [[Poster](#)] [[BibTeX](#)]
- **Rishiraj Saha Roy** and Durga Toshniwal, "[Fuzzy Clustering of Text Documents Using Naïve Bayesian Concept](#)", in *Proceedings of the 1st International Conference on Recent Trends in Information, Telecommunication and Computing 2010 (ITC '10)*, 12 - 13 March 2010, Kochi, India, pages 55 - 59. [[BibTeX](#)]
- Durga Toshniwal and **Rishiraj Saha Roy**, "[A Hierarchical Clustering Scheme for Unstructured Text Data](#)", in *Proceedings of the International Conference on Information*

and Knowledge Engineering 2009 (IKE '09), 13 – 16 July 2009, Las Vegas, USA, pages 640 – 646. [[BibTeX](#)]

- Durga Toshniwal and **Rishiraj Saha Roy**, "[Using Shape Patterns for Clustering Unstructured Text Documents](#)", in *Proceedings of the 18th International Conference on Software Engineering and Data Engineering 2009 (SEDE '09)*, 22 – 24 June 2009, Las Vegas, USA, pages 13 – 18. [[BibTeX](#)]

CONFERENCE SHORT PAPERS

- Zhen Jia, Abdalghani Abujabal, **Rishiraj Saha Roy**, Jannik Strötgen, and Gerhard Weikum, "[TEQUILA: Temporal Question Answering over Knowledge Bases](#)", in *Proceedings of the 27th International Conference on Information and Knowledge Management 2018 (CIKM '18)*, 22 – 26 October 2018, Turin, Italy, pages 1807 – 1810. [[Poster](#)] [[Data](#)] [[Demo](#)] [[Code](#)] [[BibTeX](#)]
- David Ziegler, Abdalghani Abujabal, **Rishiraj Saha Roy**, and Gerhard Weikum, "[Efficiency-aware Answering of Compositional Questions using Answer Type Prediction](#)", in *Proceedings of the 8th International Joint Conference on Natural Language Processing 2017 (IJCNLP '17)*, 27 November – 1 December 2017, Taipei, Taiwan, pages 222 – 227. [[Poster](#)] [[BibTeX](#)]
- Sedigheh Eslami, Asia Biega, **Rishiraj Saha Roy**, and Gerhard Weikum, "[Privacy of Hidden Profiles: Utility-Preserving Profile Removal in Online Forums](#)", in *Proceedings of the 26th International Conference on Information and Knowledge Management 2017 (CIKM '17)*, 6 – 10 November 2017, Singapore, pages 2063 – 2066. [[Poster by Sedigheh Eslami](#)] [[BibTeX](#)]
- Abdalghani Abujabal, **Rishiraj Saha Roy**, Mohamed Yahya, and Gerhard Weikum, "[QUINT: Interpretable Question Answering over Knowledge Bases](#)", in *Proceedings of the 22nd Conference on Empirical Methods in Natural Language Processing 2017 (EMNLP '17)*, 7 - 11 September 2017, Copenhagen, Denmark (demo paper), pages 61 – 66. [[Demo](#)] [[Poster](#)] [[BibTeX](#)]
- Balaji Vasan Srinivasan, **Rishiraj Saha Roy**, Harsh Jhamtani, Natwar Modani, and Niyati Chhaya, "[Corpus-based Automatic Text Expansion](#)", in *Proceedings of the 18th International Conference on Intelligent Text Processing and Computational Linguistics 2017 (CICLing '17)*, 17 – 23 April 2017, Budapest, Hungary. [[Poster](#)] [[BibTeX](#)]
- Harsh Jhamtani, **Rishiraj Saha Roy**, Niyati Chhaya and Eric Nyberg, "[Leveraging Site Search Logs to Identify Missing Content on Enterprise Webpages](#)", in *Proceedings of the 39th European Conference on Information Retrieval 2017 (ECIR '17)*, 9 - 13 April 2017, Aberdeen, UK, pages 506 – 512. [[Poster](#)] [[BibTeX](#)]
- **Rishiraj Saha Roy**, Anusha Suresh, Niloy Ganguly and Monojit Choudhury, "[Improving Document Rankings for Long Queries with Nested Query Segmentation](#)", in *Proceedings of the 38th European Conference on Information Retrieval 2016 (ECIR '16)*, 20 - 23 March 2016, Padua, Italy, pages 775 – 781. [[Poster](#)] [[BibTeX](#)] [[Data](#)] [[Code](#)]
- **Rishiraj Saha Roy**, Ritwik Sinha, Niyati Chhaya and Shiv Saini, "[Probabilistic Deduplication of Anonymous Web Traffic](#)", in *Posters of the 24th International World Wide Web Conference 2015 (WWW '15)*, 18 – 22 May 2015, Florence, Italy, pages 103 – 104 (companion). [[Poster](#)] [[BibTeX](#)]
- **Rishiraj Saha Roy**, Yogarshi Vyas, Niloy Ganguly and Monojit Choudhury, "[Improving Unsupervised Query Segmentation using Parts-of-Speech Sequence Information](#)", in *Proceedings of the 37th Annual ACM SIGIR Conference on Research and Development on Information Retrieval 2014 (SIGIR '14)*, 6 - 11 July 2014, Gold Coast, Australia, pages 935 – 938. [[Poster](#)] [[BibTeX](#)]

- **Rishiraj Saha Roy**, Niloy Ganguly and Monojit Choudhury, "[Structural Complexity of Web Search Queries through the Lenses of Positionality, Language Models and Networks](#)", in *Proceedings of the 10th European Conference on Complex Systems 2013 (ECCS '13)*, 16 - 20 September 2013, Barcelona, Spain, page 75. [[Poster](#)] [[BibTeX](#)]
- **Rishiraj Saha Roy**, Anusha Suresh, Niloy Ganguly and Monojit Choudhury, "[Place Value: Word Position Shifts Vital to Search Dynamics](#)", in *Posters of the 22nd International World Wide Web Conference 2013 (WWW '13)*, 13 - 17 May 2013, Rio de Janeiro, Brazil, pages 153 - 154 (*companion*). [[Poster](#)] [[BibTeX](#)]
- Nikita Mishra, **Rishiraj Saha Roy**, Niloy Ganguly, Srivatsan Laxman and Monojit Choudhury, "[Unsupervised Query Segmentation Using only Query Logs](#)", in *Posters of the 20th International World Wide Web Conference 2011 (WWW '11)*, 28 March - 1 April 2011, Hyderabad, India, pages 91 - 92 (*companion*). [[Poster](#)] [[BibTeX](#)]

JOURNALS

- **Rishiraj Saha Roy**, Smith Agarwal, Niloy Ganguly, and Monojit Choudhury, "[Syntactic Complexity of Web Search Queries through the Lenses of Language Models, Networks and Users](#)", in *Information Processing and Management (IP&M)*, Elsevier, Volume 52, Issue 5, September 2016, pages 923 - 948. [[Code](#)] [[BibTeX](#)]
- **Rishiraj Saha Roy**, Rahul Katare, Niloy Ganguly, Srivatsan Laxman and Monojit Choudhury, "[Discovering and understanding word level user intent in web search queries](#)", in *Web Semantics: Science, Services and Agents on the World Wide Web (Special Issue on Semantic Search)*, Elsevier, Volume 30, January 2015, pages 22 - 38. [[BibTeX](#)]
- Durga Toshniwal and **Rishiraj Saha Roy**, "[Shape pattern matching: A tool to cluster unstructured text documents](#)", in *Journal of Computational Methods in Science and Engineering (JCMSE)*, Volume 10, Supplement 1, IOS Press, September 2010, pages 73 - 84. [[BibTeX](#)]
- **Rishiraj Saha Roy** and Durga Toshniwal, "[Clustering Unstructured Text Documents Using Naive Bayesian Concept and Shape Pattern Matching](#)", in *International Journal of Advancements in Computing Technology (IJACT)*, AICIT Press, pages 52 - 63, Volume 1, Number 1, September 2009. [[BibTeX](#)]

OTHER PUBLICATIONS

- Abdalghani Abujabal, **Rishiraj Saha Roy**, Mohamed Yahya, and Gerhard Weikum, "[ComQA: A Community-sourced Dataset for Complex Factoid Question Answering with Paraphrase Clusters](#)", arXiv, 25 September 2018. [[Data](#)] [[BibTeX](#)]
- Zhen Jia, Abdalghani Abujabal, **Rishiraj Saha Roy**, Jannik Strötgen, and Gerhard Weikum, "[TempQuestions: A Benchmark for Temporal Question Answering](#)", in *Proceedings of the First International Workshop on Hybrid Question Answering with Structured and Unstructured Knowledge at WWW 2018 (HQA '18)*, 23 April 2018, Lyon, France, pages 1057 - 1062 (*WWW companion*). [[Data](#)] [[Slides](#)] [[BibTeX](#)]
- **Rishiraj Saha Roy**, Anusha Suresh, Niloy Ganguly, Monojit Choudhury, Deepak Shankar and Tanwita Nimiar, "[Improving Document Rankings for Long Queries with Nested Query Segmentation](#)", in *Microsoft Research Technical Report*, 31 December 2015, Report No. MSR-TR-2015-91. [[BibTeX](#)] [[Data](#)] [[Code](#)]
- Moumita Sinha and **Rishiraj Saha Roy**, "[Preference Mapping for Automated Recommendation of Product Attributes for Designing Marketing Content](#)", in *Proceedings of the 1st ACM RecSys Workshop on New Trends in Content-based*

Recommender Systems 2014 (CBRecSys '14), 6 October 2014, Silicon Valley, USA, pages 64 – 67. [[Slides](#)] [[BibTeX](#)]

- **Rishiraj Saha Roy**, "[Analyzing Linguistic Structure of Web Search Queries](#)", in *Doctoral Consortium of the 22nd International World Wide Web Conference 2013 (WWW '13)*, 13 – 17 May 2013, Rio de Janeiro, Brazil, pages 395 – 399 (*companion*). [[Slides](#)] [[BibTeX](#)]
- **Rishiraj Saha Roy**, Niloy Ganguly, Monojit Choudhury and Naveen Kumar Singh, "[Complex Network Analysis Reveals Kernel-Periphery Structure in Web Search Queries](#)", in *Proceedings of the 2nd ACM SIGIR Workshop on Query Representation and Understanding 2011 (QRU '11)*, 28 July 2011, Beijing, China, pages 5 – 8. [[Slides](#)] [[Poster](#)] [[BibTeX](#)]

PATENTS

- Balaji Vasan Srinivasan, **Rishiraj Saha Roy**, Niyati Chhaya, Natwar Modani, and Harsh Jhamtani, "[Expanding Input Content Utilizing Previously-generated Content](#)", *USPTO Publication No. US20180060287A1*, published on 01 March 2018 (*Patent published*).
- **Rishiraj Saha Roy**, Abhijeet Singh, Prashant Chawla, Shubham Saxena and Atanu R. Sinha, "[Assigning of Topical Icons to Documents to Improve File Navigation](#)", *USPTO Publication No. US20170076151A1*, filed on 16 March 2017 (*Patent published*).
- **Rishiraj Saha Roy**, Guna Prasaad Jeganathan, Aishwarya Padmakumar, and Ponnurangam Kumaraguru, "[Linguistic Personalization of Messages for Targeted Campaigns](#)", *USPTO Publication No. US20160171560A1*, published on 16 June 2016 (*Patent published*).
- Moumita Sinha, **Rishiraj Saha Roy**, and Ritwik Sinha, "[Preference Mapping for Automated Attribute-Selection in Campaign Design](#)", *USPTO Publication No. US20160117737A1*, published on 28 April 2016 (*Patent published*).

PROFESSIONAL ACTIVITIES

- Co-organizer of the [Sixth Social Networking Workshop at COMSNETS 2019 \(COMSNETS SN '19\)](#), 11 January 2019, Bangalore, India (with Saptarshi Ghosh).
- Co-organizer of the [Eleventh Workshop on Dynamics on and of Complex Networks at CCS 2018 \(DOOCN '18\)](#), 26 September 2018, Thessaloniki, Greece (with Jean-Charles Delvenne, Fakhteh Ghanbarnejad, Fariba Karimi, Bivas Mitra, and Parantapa Bhattacharya).
- Co-organizer of the [Fifth Social Networking Workshop at COMSNETS 2018 \(COMSNETS SN '18\)](#), 7 January 2018, Bangalore, India (with Moumita Sinha).
- Co-organizer of the [Tenth Workshop on Dynamics on and of Complex Networks at NetSci 2017 \(DOOCN '17\)](#), 19 June 2017, Indianapolis, USA (with Jean-Charles Delvenne, Fakhteh Ghanbarnejad, Fariba Karimi, and Bivas Mitra).
- Co-organizer of the shared task on "Transliterated Search" at the [Fifth Forum for IR Evaluation 2013 \(FIRE '13\)](#), 4 – 6 December 2013, New Delhi, India (with Monojit Choudhury, Prasenjit Majumder, and Komal Agarwal).
- Student volunteer at the [MSR India and IRSI Pre-FIRE Workshop on Multilingual Information Retrieval](#), 15 – 17 June 2013, Bengaluru, India.
- Student volunteer for the [22nd International World Wide Web Conference 2013 \(WWW '13\)](#), 13 – 17 May 2013, Rio de Janeiro, Brazil WWW 2013, Rio de Janeiro, Brazil.
- Maintained Complex Network Research Group's (CNeRG) webpage at IIT Kharagpur almost throughout duration of Ph.D. [February 2010 – March 2014]
- **Review Committees:** [EMNLP](#) (2017, 2018), [ESWC](#) (2018), [ACM TALLIP](#) (2018), [CACM](#) (2017), [JASIST](#) (2017), [JCNLP](#) (2017), [Elsevier IS](#) (2017), [FIRE](#) (2016, 2017, 2018), [Springer IS Frontiers](#) (2017), [CODS-COMAD](#) (2018, 2019), [ICDCN](#) (2019), [Springer](#)

[Sadhana](#) (2018), [SMERP](#) (2017, 2018), [IJCAT](#) (2016), [ACM Compute](#) (2015), [IDRBT DC](#) (2015)

- **Mentorships:** 2 Masters' students [MPI for Informatics (2)], 2 Bachelors' students [MPI for Informatics (2)], and 7 interns [Adobe Research India (5) and MPI for Informatics (2)]

PROFESSIONAL ACHIEVEMENTS

- Selected to be among the top-80 researchers from European and Asian universities for attending the Google NLP Summit at Zürich, Switzerland (September 2017)
- **Best Paper Award** at the 16th International Conference on Intelligent Text Processing and Computational Linguistics 2015 (CICLing '15) at Cairo, Egypt (April 2015)
- ACM SIGIR Student Travel Support for attending the 35th Annual SIGIR Conference on Research and Development in Information Retrieval (SIGIR '12) at Portland, USA (August 2012)
- **Best Poster Award** at Evolang IX, Kyoto, Japan, sponsored by the John Benjamins Publishing Co. (March 2012)
- Evolang Student Grant for attending the 9th International Conference on the Evolution of Language (Evolang IX) at Kyoto, Japan (March 2012)
- **Microsoft Research Lab India Ph.D. Fellowship** for three years (August 2011)
- NIXI Fellowship for attending the 20th International World Wide Web Conference (WWW '11) at Hyderabad, India (March 2011)

INVITED TALKS AND POSTERS

- Invited talk on "[Answer Type Prediction for Question Answering](#)" at Adobe Research, 16 December 2016, Bangalore, India.
- Invited talk on "[Syntactic Complexity of Web Search Queries through the Lenses of Language Models, Networks and Users](#)" at the [9th Satellite on Dynamics on and of Complex Networks 2016 \(DOOCN '16\)](#), 20 September 2016, Amsterdam, The Netherlands.
- Invited poster on "[Unsupervised Approaches to Syntactic Analysis of Web Search Queries](#)" at the [Xerox Research Centre India \(XRCI\) Open PhD Symposium 2015](#), 23 January 2015, Bangalore, India.
- Invited talk on "[Experiment Design and Evaluation for Information Retrieval](#)" at the [Pre-FIRE Workshop on Information Retrieval for Transliterated Queries](#), 2 August 2014, Microsoft Research Lab India, Bangalore.
- Invited talk on "[Query and Document Understanding](#)" at the [Summer Workshop on Natural Language Processing and Data Mining](#), 18 July 2014, PESIT, Bengaluru, India.
- Invited talk on "[Query and Document Understanding](#)" at the [MSR India and IRSI Pre-FIRE Workshop on Multilingual Information Retrieval](#), 15 – 17 June 2013, Bengaluru, India.
- Invited poster on "[Are Web Search Queries Evolving into a Language of their Own?](#)" at the [4th Forum for IR Evaluation 2012 \(FIRE '12\)](#), 17 – 19 December 2012, Kolkata, India.

EDUCATIONAL BACKGROUND

- Ph.D. from Department of Computer Science and Engineering (CSE), IIT Kharagpur [December 2009 – March 2014]
- Master of Technology (M.Tech.) in Information Technology from Department of Electronics and Computer Engineering (ECE), IIT Roorkee, with CGPA = 7.98 [July 2007 – July 2009]

- Bachelor of Engineering (B.E.) in Information Technology from Department of Information Technology (IT), Jadavpur University, Kolkata, with CGPA = 8.76 [July 2003 – June 2007]
- Schooling from Bhavan's Gangabux Kanoria Vidyamandir (BGKV, also known as "Bharatiya Vidya Bhavan"), Kolkata, with CBSE Class X marks = 89.80% and CBSE Class XII marks = 89.40% [April 1989 – June 2003]

THESES

- [Unsupervised Approaches to Syntactic Analysis of Web Search Queries](#) [Slides] (December 2009 – March 2014, IIT Kharagpur, Kharagpur, India) (Advisors: Prof. Niloy Ganguly and Dr. Monojit Choudhury) [PhD Thesis]
- **Clustering Unstructured Text Documents Using Naïve Bayesian Concept and Shape Pattern Matching** (July 2008 – June 2009, IIT Roorkee, Roorkee, India) (Advisor: Prof. Durga Toshniwal) [Masters' Thesis]
- **Analysis and Enhancement of TCP Performance in Mobile Ad Hoc Networks** (June 2006 – May 2007, Jadavpur University, Kolkata, India) (Advisor: Prof. Uttam Kumar Roy, Collaborators: Mainak Ghosh and Prajit Kumar Das) [Bachelors' Thesis]

INTERNSHIPS AND VISITS

- Internships at Microsoft Research Lab India, Bangalore (May 2010 - July 2010, October 2011 - January 2012)
- Visits to Microsoft Research Lab India, Bangalore (November - December 2010, March - April 2012, June - July 2013)

TEACHING ASSISTANTSHIPS

- **Information Retrieval**, CSE, IIT Kharagpur: Autumn 2013, Autumn 2012, Autumn 2011
- **Complex Networks**, CSE, IIT Kharagpur: Spring 2013, Spring 2012
- **Programming and Data Structures**, CSE, IIT Kharagpur: Spring 2011, Autumn 2010
- **Data Mining**, ECE, IIT Roorkee: Spring 2009
- **Object Oriented Programming**, ECE, IIT Roorkee: Autumn 2008

SKILLS

- **Programming Languages:** Python (2013, 2015 – 2017), R (2014 – 2015), C (2003 – 2013), C# (2011 – 2012), C++ (2004 – 2008), Java (Core) (2005 – 2011, 2014 – 2015)
- **Specialized Software:** Apache Spark (2016 – 2017), Lucene (2011 – 2013), Octave (2013 – 2014), MATLAB (2010), WEKA (2008 – 2009)
- **Speak, Read, and Write:** English, Bangla, and Hindi

EXTRACURRICULAR ACHIEVEMENTS

- **Taekwondo** (2009): Obtained Blue Belt; 2nd position in the Senior Bantamweight Category in the 2nd Open District Club Taekwondo Championship, Roorkee, Uttarakhand
- **Face Painting Competition organized by Jadavpur University, Kolkata during its cultural fest Sanskriti** (2007): Obtained 1st Position
- **General Quiz organized by Salt Lake Book Fair Committee** (2000): Obtained 2nd Position

- **Excellence in Character** (1999): Obtained the Shri. C. Subramaniam Award for the Best Character in School
- **General Quiz organized by Rotary Club of Salt Lake** (1999): Obtained 2nd Position

REFERENCES

Gerhard Weikum

Research Director and Professor
Databases and Information Systems Group
Max Planck Institute for Informatics
weikum@mpi-inf.mpg.de
+49-681-93255000

Niloy Ganguly

Professor
Computer Science and Engineering
Indian Institute of Technology Kharagpur
niloy@cse.iitkgp.ernet.in
+91-3222-283460

Monojit Choudhury

Researcher
Multilingual Systems Research Group
Microsoft Research Lab India
monojitc@microsoft.com
+91-80-66586214

Kalika Bali

Researcher
Technology for Emerging Markets Group
Microsoft Research Lab India
kalikab@microsoft.com
+91-80-66566218

PERSONAL DETAILS

Father's Name : Manoj Kumar Saha Roy
Mother's Name : Manjusree Saha Roy
Spouse's Name : Amrita Saha Roy
Date of Birth : 18 October 1984
Nationality : Indian
Gender : Male
Office Address : Room #407, Max Planck Institute for Informatics, Campus E1.4,
66123 Saarbruecken, Germany.
Current Home Address : Sankt-Michael-Strasse 5, 66424 Homburg, Germany.
Permanent Address : AH - 8, Sector - 2, Salt Lake, Kolkata - 700091, West Bengal, India.
Mobile Number : +49-176-22327841