

RISHIRAJ SAHA ROY

Post-doctoral Researcher
Max Planck Institute for Informatics, Saarbruecken, Germany
[November 2015 - Present]

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



Ph.D., Microsoft Research 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. thesis: Unsupervised Approaches to Syntactic Analysis of Web Search Queries

General areas: Information retrieval, Computational linguistics, Text mining

SELECTED PUBLICATION SUMMARY

Conference long papers WSDM '19, WWW '18, ICDAR '17, SIGIR '17, CILing '15, COLING '14, Evolang '14, ACL '13, FIRE '13, SIGIR '12, Evolang '12

Conference short papers CIKM '18, IJCNLP '17, CIKM '17, EMNLP '17, CILing '17, ECIR '17, ECIR '16, WWW '15, SIGIR '14, WWW '13, WWW '11

Journal papers IP&M '16, Web Semantics '15, IJACT '09

CONFERENCE LONG PAPERS

- Azin Ghazimatin, **Rishiraj Saha Roy**, and Gerhard Weikum, FAIRY: A Framework for Understanding Relationships between Users' Actions and their Social Feeds, WSDM 19.
- Abdalghani Abujabal, **Rishiraj Saha Roy**, Mohamed Yahya, and Gerhard Weikum, Never-Ending Learning for Open-Domain Question Answering over Knowledge Bases, WWW '18.

- **Rishiraj Saha Roy**, Abhijeet Singh, Prashant Chawla, Shubham Saxena, and Atanu R. Sinha, Automatic Assignment of Topical Icons to Documents for Faster File Navigation, ICDAR '17.
- Asia Biega, **Rishiraj Saha Roy**, and Gerhard Weikum, Privacy through Solidarity: A User-Utility-Preserving Framework to Counter Profiling, SIGIR 2017.
- **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, CICLing '15. [**BEST PAPER AWARD**]
- **Rishiraj Saha Roy**, Rahul Katare, Niloy Ganguly and Monojit Choudhury, Automatic Discovery of Adposition Typology, COLING '14.
- **Rishiraj Saha Roy**, Maliki Reddy Dastagiri Reddy, Niloy Ganguly, and Monojit Choudhury, Understanding the Linguistic Structure and Evolution of Web Search Queries, Evolang '14.
- **Rishiraj Saha Roy**, Monojit Choudhury, Prasenjit Majumder, and Komal Agarwal, Overview of the FIRE 2013 Track on Transliterated Search, FIRE '13.
- 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, ACL '13.
- **Rishiraj Saha Roy**, Niloy Ganguly, Monojit Choudhury, and Srivatsan Laxman, An IR-based Evaluation Framework for Web Search Query Segmentation, SIGIR '12.
- **Rishiraj Saha Roy**, Monojit Choudhury, and Kalika Bali, Are Web Search Queries an Evolving Protolanguage?, Evolang '12. [**BEST POSTER AWARD**]

CONFERENCE SHORT PAPERS

- Zhen Jia, Abdalghani Abujabal, **Rishiraj Saha Roy**, Jannik Strötgen, and Gerhard Weikum, TEQUILA: Temporal Question Answering over Knowledge Bases, CIKM '18.
- David Ziegler, Abdalghani Abujabal, **Rishiraj Saha Roy**, and Gerhard Weikum, Efficiency-aware Answering of Compositional Questions using Answer Type Prediction, IJCNLP '17.
- Sedigheh Eslami, Asia Biega, **Rishiraj Saha Roy**, and Gerhard Weikum, Privacy of Hidden Profiles: Utility-Preserving Profile Removal in Online Forums, CIKM '17.
- Abdalghani Abujabal, **Rishiraj Saha Roy**, Mohamed Yahya, and Gerhard Weikum, QUINT: Interpretable Question Answering over Knowledge Bases, EMNLP '17.
- Balaji Vasan Srinivasan, **Rishiraj Saha Roy**, Harsh Jhamtani, Natwar Modani, and Niyati Chhaya, Corpus-based Automatic Text Expansion, CICLing '17.
- Harsh Jhamtani, **Rishiraj Saha Roy**, Niyati Chhaya and Eric Nyberg, Leveraging Site Search Logs to Identify Missing Content on Enterprise Webpages, ECIR '17.
- **Rishiraj Saha Roy**, Anusha Suresh, Niloy Ganguly, and Monojit Choudhury, Improving Document Rankings for Long Queries with Nested Query Segmentation, ECIR '16.
- **Rishiraj Saha Roy**, Ritwik Sinha, Niyati Chhaya and Shiv Saini, Probabilistic Deduplication of Anonymous Web Traffic, WWW '15.

- **Rishiraj Saha Roy**, Yogarshi Vyas, Niloy Ganguly and Monojit Choudhury, Improving Unsupervised Query Segmentation using Parts-of-Speech Sequence Information, SIGIR '14.
- **Rishiraj Saha Roy**, Anusha Suresh, Niloy Ganguly and Monojit Choudhury, Place Value: Word Position Shifts Vital to Search Dynamics, WWW '13.
- Nikita Mishra, **Rishiraj Saha Roy**, Niloy Ganguly, Srivatsan Laxman and Monojit Choudhury, Unsupervised Query Segmentation Using only Query Logs, WWW '11.

JOURNAL PAPERS

- **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, Information Processing and Management (IP&M), Elsevier, Volume 52, Issue 5, September '16.
- **Rishiraj Saha Roy**, Rahul Katare, Niloy Ganguly, Srivatsan Laxman and Monojit Choudhury, Discovering and understanding word level user intent in Web search queries, Journal of Web Semantics (Special Issue on Semantic Search), Elsevier, Volume 30, January '15.
- **Rishiraj Saha Roy** and Durga Toshniwal, Clustering Unstructured Text Documents Using Naive Bayesian Concept and Shape Pattern Matching, International Journal of Advancements in Computing Technology (IJACT), AICIT Press, Volume 1, Number 1, September '09.

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, September 2018.
- Zhen Jia, Abdalghani Abujabal, **Rishiraj Saha Roy**, Jannik Strötgen, and Gerhard Weikum, "TempQuestions: A Benchmark for Temporal Question Answering", HQA '18 (WWW Workshop).
- **Rishiraj Saha Roy**, Anusha Suresh, Niloy Ganguly, Monojit Choudhury, Deepak Shankar, and Tanwita Nimiar, "Improving Document Rankings for Long Queries with Nested Query Segmentation", Microsoft Research Technical Report, December 2015.
- Moumita Sinha and **Rishiraj Saha Roy**, "Preference Mapping for Automated Recommendation of Product Attributes for Designing Marketing Content", CBRecSys '14 (RecSys Workshop).
- **Rishiraj Saha Roy**, "Analyzing Linguistic Structure of Web Search Queries", Doctoral Consortium of WWW '13.
- **Rishiraj Saha Roy**, Niloy Ganguly, Monojit Choudhury and Naveen Kumar Singh, "Complex Network Analysis Reveals Kernel-Periphery Structure in Web Search Queries", QRU '11 (SIGIR Workshop).

PATENTS

- Balaji Vasani 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 Mar '18 (*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, published on 16 Mar '17 (*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 Jun '16 (*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 Apr '16 (*Patent published*).

PROFESSIONAL ACTIVITIES

- **Co-editor** of Dynamics On and Of Complex Networks III: Machine Learning and Statistical Physics Approaches, Springer, 16 June 2019 (with Fakhteh Ghanbarnejad, Fariba Karimi, Jean-Charles Delvenne, and Bivas Mitra).
- **Co-organizer** of the Twelfth Workshop on Dynamics on and of Complex Networks at NetSci 2019 (DOOCN-XII), 28 May 2019, Vermont, USA (with Fakhteh Ghanbarnejad, Parantapa Bhattacharya, and Sandipan Sikdar).
- **Co-organizer** of the Sixth Social Networking Workshop at COMSNETS 2019 (COMSNETS SN '19), 11 January 2019, Bangalore, India (with Saptarshi Ghosh).
- **Publications Chair** of the ACM India Joint International Conference on Data Science and Management of Data (CoDS-COMAD) 2019, 3 - 5 January 2019.
- **Co-organizer** of the Eleventh Workshop on Dynamics on and of Complex Networks at CCS 2018 (DOOCN-XI), 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-X), 19 June 2017, Indianapolis, USA (with Jean-Charles Delvenne, Fakhteh Ghanbarnejad, Fariba Karimi, and Bivas Mitra).
- **Conference Program Committees:** SIGIR (2019), NAACL-HLT (2019), EMNLP (2017, 2018), IJCNLP (2017)
- **Journal Review Boards:** ACM TALLIP (2018), CACM (2017), JASIST (2017), Elsevier IS (2017), Springer IS Frontiers (2017).
- **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)]

- **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 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 webpage** of the Complex Network Research Group (CNeRG) at IIT Kharagpur [February 2010 – March 2014]

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 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 talks on "Complexity in Question Answering" at Microsoft Research India (10 January 2019) and IIT Kharagpur (24 December 2018).
- 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 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 Microsoft Research India, 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

- **PhD Thesis:** Unsupervised Approaches to Syntactic Analysis of Web Search Queries (December 2009 – March 2014, IIT Kharagpur, Kharagpur, India) (Advisors: Prof. Niloy Ganguly and Dr. Monojit Choudhury)
- **Masters' Thesis:** Clustering Unstructured Text Documents Using Naive Bayesian Concept and Shape Pattern Matching (July 2008 – June 2009, IIT Roorkee, Roorkee, India) (Advisor: Prof. Durga Toshniwal)
- **Bachelors' 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)

INTERNSHIPS AND VISITS

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

P. T. O.

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 – 2018), 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)

EXTRA-CURRICULAR ACHIEVEMENTS

- Taekwondo (2009): Obtained Blue Belt; 2nd position in the Senior Bantamweight Category in the 2nd Open District Club Taekwondo Championship, Roorkee, India
- Face Painting Competition organized by Jadavpur University, India, 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

PERSONAL DETAILS

Date of Birth	18 October 1984
Father's Name	Manoj Kumar Saha Roy
Spouse's Name	Amrita Saha Roy
Nationality	Indian
Gender	Male
Languages	English, Bangla, and Hindi (Speak, read, and write)
Office Address	MPI for Informatics, Campus E1.4, 66123 Saarbruecken, Germany.
Home Address	Sankt-Michael-Strasse 5, 66424 Homburg, Germany.
Mobile Number	+49-176-22327841

REFERENCES

Prof. Dr. Gerhard Weikum

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

Dr. Monojit Choudhury

Researcher
Microsoft Research India
monojitc@microsoft.com
+91-80-66586214

Prof. Niloy Ganguly

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

Dr. Kalika Bali

Researcher
Microsoft Research India
kalikab@microsoft.com
+91-80-66566218