

RISHIRAJ SAHA ROY

Computer Scientist at Adobe Research Labs India
[April 2014 - present]

Ph.D. [2009 – 2014]
Microsoft Research India Ph.D. Fellow
Computer Science and Engineering
Indian Institute of Technology Kharagpur



[**Advisors:** Prof. Niloy Ganguly, IIT Kharagpur;
Dr. Monojit Choudhury, Microsoft Research India]

E-mail : rishiraj.saharoy@gmail.com ; roy@adobe.com

Web : <http://cse.iitkgp.ac.in/~rishiraj> ;
<http://www.adobe.com/technology/people/bangalore/rishiraj-saha-roy.html>

RESEARCH INTERESTS

- **Ph.D. Thesis:** [Unsupervised Approaches to Syntactic Analysis of Web Search Queries \[Slides\]](#)
- **Ph.D. Research Areas:** Web query understanding – Query syntax, Query segmentation, Query intent, Query complexity
- **General Areas of Interest:** Information Retrieval, Text Mining, Complex Networks, Machine Learning, Linguistics

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)

PUBLICATION SUMMARY

- **Journals** : [Elsevier Web Semantics 2014](#), [JCMSE 2010](#), [IJACT 2009](#)
- **Conference long papers** : [SIGIR 2012](#), [ACL 2013](#), [Coling 2014](#), [Evolang 2012](#), [Evolang 2014](#), [ITC 2010](#), [IKE 2009](#), [SEDE 2009](#)
- **Conference short papers** : [SIGIR 2014](#), [WWW 2013](#), [WWW 2011](#), [ECCS 2013](#)
- **Workshops and doctoral consortia** : [WWW 2013](#), [QRU 2011 \(SIGIR\)](#), [CBRecSys 2014 \(RecSys\)](#)

PATENTS FILED

- **Rishiraj Saha Roy**, Guna Prasaad Jeganathan, Aishwarya Padmakumar, and Ponnurangam Kumaraguru, “Automated Linguistic Personalization of Ad Messages for Targeted Campaigns”, *Patent Application Number P4592-US, United States Patent Office*, filed on 10 December 2014 (*Patent pending*).
- Moumita Sinha, **Rishiraj Saha Roy**, and Ritwik Sinha, “Preference Mapping for Automated Attribute Selection in Campaign Design”, *Patent Application Number 4455US01, United States Patent Office*, filed on October 11, 2014 (*Patent pending*).

PUBLICATIONS FROM Ph.D. THESIS

- **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*, Elsevier, 2014 (*in press*).

- **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 (Coling '14)*, 23 – 29 August 2014, Dublin, Ireland, pages 1037 – 1046. [[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 (SIGIR '14)*, 6 - 11 July 2014, Gold Coast, Australia, pages 935 – 938. [[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 (Evolang X)*, 14 - 17 April 2014, Vienna, Austria, pages 286 – 293. [[Slides](#)] [[BibTeX](#)]
- **Rishiraj Saha Roy**, Monojit Choudhury, Prasenjit Majumder and Komal Agarwal, "[Overview and Datasets of 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. [[Slides](#)] [[BibTeX](#)] [[Data](#)]
- **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 (ECCS '13)*, 16 - 20 September 2013, Barcelona, Spain, page 75. [[Poster](#)] [[BibTeX](#)]
- 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 (ACL '13)*, 4 - 9 August 2013, Sofia, Bulgaria, pages 1713 – 1722. [[Slides](#)] [[Supplementary material](#)] [[Data](#)] [[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](#)]
- **Rishiraj Saha Roy**, "[Analyzing Linguistic Structure of Web Search Queries](#)", in *Doctoral Consortium of the 22nd International World Wide Web Conference (WWW '13)*, 13 - 17 May 2013, Rio de Janeiro, Brazil, pages 395 – 399 (companion). [[Slides](#)] [[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 (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 (Evolang IX)*, 13 – 16 March, 2012, Kyoto, Japan, pages 304 – 311. [**BEST RESEARCH POSTER AWARD**]. [[Poster](#)] [[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](#)]
- 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](#)]

MANUSCRIPTS IN PREPARATION

- **Rishiraj Saha Roy**, Smith Agarwal, Niloy Ganguly and Monojit Choudhury, "Syntactic Complexity of Web Queries through the Lenses of Language Models, Networks and Users", *communicated to PLOS ONE*, Public Library of Science (under review).
- **Rishiraj Saha Roy**, Anusha Suresh, Niloy Ganguly and Monojit Choudhury, "Nested Query Segmentation for Information Retrieval", *to be communicated to the 38th Annual ACM SIGIR Conference on Research and Development on Information Retrieval (SIGIR '15)*, 9 - 13 August 2015, Santiago, Chile (in preparation).

PUBLICATIONS OUTSIDE Ph.D. THESIS

- 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](#)]
- 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, “[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](#)]
- **Rishiraj Saha Roy** and Durga Toshniwal, “[Clustering Unstructured Text Documents Using Naïve 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](#)]
- 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 ISCA (International Society for Computers and their Applications) 18th International Conference on Software Engineering and Data Engineering 2009 (SEDE '09)*, 22 – 24 June 2009, Las Vegas, USA, pages 13 – 18. [[BibTeX](#)]

EDUCATIONAL BACKGROUND

- Ph.D. from Computer Science and Engineering, IIT Kharagpur [2009 – 2014]
- M.Tech. in Information Technology from IIT Roorkee with CGPA = 7.98 [2007 - 2009]
- B.E. in Information Technology from Jadavpur University, Kolkata, with CGPA = 8.76 [2003 - 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% [1989 - 2003]

AWARDS, FELLOWSHIPS AND GRANTS

- 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 Research 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)

PROJECTS

- **Clustering Unstructured Text Documents Using Naïve Bayesian Concept and Shape Pattern Matching** (July 2008 – June 2009, Indian Institute of Technology Roorkee, Roorkee) (Advisor: Prof. Durga Toshniwal) [*Masters' Project*]
- **Text Document Classification and a Search Tool for News Articles** (July 2008 – October 2008, Indian Institute of Technology Roorkee, Roorkee) (Advisor: Prof. Durga Toshniwal, Collaborator: Shweta Modi)

- **Analysis and Enhancement of TCP Performance in Mobile Ad Hoc Networks** (June 2006 – May 2007, Jadavpur University, Kolkata) (Advisor: Prof. Uttam Kumar Roy) [*Bachelors' Project*]

TEACHING ASSISTANTSHIPS

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

SKILLS

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

OTHER PROFESSIONAL ACTIVITIES

- Invited speaker 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 speaker on "Query and Document Understanding" at the Summer Workshop on Natural Language Processing and Data Mining, 18 July 2014, PESIT, Bengaluru, India.
- Task coordinator for the "Transliterated Search" track, with Dr. Monojit Choudhury (Microsoft Research India), Prof. Prasenjit Majumder (DAIICT) and Komal Agarwal (DAIICT), at the 5th Forum for IR Evaluation 2013 (FIRE '13), 4 – 6 December 2013, New Delhi, India.
- Invited speaker on "Query and Document Understanding" and 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.
- Invited poster on "Are Web Search Queries Evolving into a Language of their Own?" (with Prof. Niloy Ganguly, IIT Kharagpur and Dr. Monojit Choudhury, Microsoft Research India) at the 4th Forum for IR Evaluation 2012 (FIRE '12), 17 – 19 December 2012, Kolkata, India.

EXTRACURRICULAR ACHIEVEMENTS

- **Excellence in Character** (1999)
Obtained the Shri. C. Subramaniam Award for Best Character in School
- **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 Rotary Club of Salt Lake** (1999)
Obtained 2nd Position
- **General Quiz organized by Salt Lake Book fair Committee** (2000)
Obtained 2nd Position

REFERENCES

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 India
monojitc@microsoft.com
+91-80-66586214

Kalika Bali

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

Durga Toshniwal

Associate Professor
Computer Science and Engineering
Indian Institute of Technology Roorkee
durgafec@iitr.ernet.in
+91-1332-285687

PERSONAL DETAILS

Father's Name : Manoj Kumar Saha Roy
Mother's Name : (Late) Manjusree Saha Roy
Spouse's Name : Amrita Saha Roy
Date of Birth : 18 October 1984
Nationality : Indian
Gender : Male
Present Address : Flat #303, "Sri Paradise", #46, 6th Cross, HAL 3rd Stage,
Kodihalli, Bangalore – 560008, Karnataka.
Permanent Address : AH – 8, Sector – 2, Salt Lake, Kolkata – 700091, West Bengal.
Mobile Number : +91-9663088199/+91-9477588851