

Sujan Kumar Saha

Assistant Professor (Senior Grade)
Dept. of Computer Science and Engineering
Birla Institute of Technology, Mesra
Ranchi, Jharkhand, India
Phone: +91 9472711949
E-mail: sujan.kr.saha@gmail.com

Education

Doctor of Philosophy (Thesis submitted on 21st January, 2010)

Institute: Indian Institute of Technology Kharagpur.

Thesis Title: Machine learning methods for named entity recognition with limited resource.

Thesis Supervisors: Prof. Pabitra Mitra and Prof. Sudeshna Sarkar, Department of Computer Science and Engineering, Indian Institute of Technology Kharagpur.

Master of Technology in Computer Applications (2003 – 2005).

Institute: Indian Institute of Technology Delhi.

CGPA: 8.22/10.

Project Title: Ontology Based Text Information Retrieval.

Project Supervisor: Dr. Lipika Dey, Dept. of Mathematics, IIT Delhi.

Bachelor of Technology in Computer Science and Technology (1999 – 2003).

Institute: Kalyani Govt. Engineering College, West Bengal.

Percentage: 76.79.

Project Title: Clustering Based Face Detection.

Project Supervisor: Mr. Siddhartha Bhattacharya, Dept. of Comp. Sc. and Tech.

Higher Secondary - 10+2 (1999).

Board: West Bengal Council of Higher Secondary Education.

Institute: Kashiram Das Institution, Katwa, Burdwan, West Bengal.

Percentage: 83.60.

Secondary - 10th (1997).

Board: West Bengal Board of Secondary Education.

Institute: Kashiram Das Institution, Katwa, Burdwan, West Bengal.

Percentage: 81.13.

Research Interest

Machine Learning
Natural Language Processing
Text Mining
Information Retrieval

Publications

In International Journals:

1. **Saha Sujan Kumar**, Sarkar Sudeshna and Mitra Pabitra. 2009. Feature Selection Techniques for Maximum Entropy Based Biomedical Named Entity Recognition. Elsevier Journal of Biomedical Informatics, Vol. 42, No. 5, pages 905--911.
2. **Saha Sujan Kumar**, Sarkar Sudeshna and Mitra Pabitra. 2009. Hindi Named Entity Annotation Error Detection and Correction. Int. J. Language Forum, Vol. 35, No. 2, pages 73--92.
3. **Saha Sujan Kumar**, Ghosh Partha Sarathi, Sarkar Sudeshna and Mitra Pabitra. 2008. Hindi Named Entity Recognition Using Maximum Entropy and Transliteration. Polibits, Vol. 38, pages 33--41.
4. **Saha Sujan Kumar**, Narayan Shashi, Sarkar Sudeshna and Mitra Pabitra. A Kernel for Named Entity Recognition. Elsevier Pattern Recognition Letters (*accepted*).

In International Conference Proceedings:

1. **Saha Sujan Kumar**, Mitra Pabitra and Sarkar Sudeshna. 2008. Word Clustering and Word Selection Based Feature Reduction for MaxEnt Based NER in Hindi. Proc. of the Association for Computational Linguistics (ACL-08:HLT), pages 488--495, Columbus, Ohio.
2. **Saha Sujan Kumar**, Mitra Pabitra and Sarkar Sudeshna. 2009. A Semi-supervised Approach for Maximum Entropy Based Hindi Named Entity Recognition. Proc. of Pattern Recognition and Machine Intelligence (PReMI-09), LNCS Vol. 5909/2009, pages 225--230, New Delhi, India.
3. **Saha Sujan Kumar**, Sarkar Sudeshna and Mitra Pabitra. 2008. A Hybrid Feature Set Based Maximum Entropy Hindi Named Entity Recognition. Proc. of the 3rd International Joint Conference on Natural Language Processing (IJCNLP 2008), pages 343--349, Hyderabad, India.
4. **Saha Sujan Kumar**, Chatterji Sanjay, Dandapat Sandipan, Sarkar Sudeshna and Mitra Pabitra. 2008. A Hybrid Named Entity Recognition System for South and South East Asian Languages. Proc. of the IJCNLP 2008 Workshop on Named Entity Recognition for South and South East Asian Languages (NERSSAL 2008), pages 17 – 24, Hyderabad, India. (*Best result in the NERSSAL 2008 shared task*)
5. **Saha Sujan Kumar**, Sarkar Sudeshna and Mitra Pabitra. 2008. Gazetteer Preparation for Named Entity Recognition in Indian Languages. Proc. of the IJCNLP-08 Workshop on Asian Language Resources (ALR6 2008), pages 9--16, Hyderabad, India.
6. **Saha Sujan Kumar**, Sarkar Sudeshna and Mitra Pabitra. 2008. A Novel Semi-supervised Method for Named Entity Detection. Proc. of the 6th International Conference on Natural Language Processing (ICON 2008), Pune, India.
7. **Saha Sujan Kumar**, Sarkar Sudeshna and Mitra Pabitra. 2007. Transliteration Based Gazetteer Preparation for Named Entity Recognition in Hindi. Proc. of the 7th International Symposium on Natural Language Processing (SNLP 2007), Pattaya, Thailand.

8. **Saha Sujan Kumar**, De Lipika. 2006. Ontology Based Text Information Retrieval. Proc. of the International Conference on Emerging Applications of IT (EAIT 2006), Elsevier, pages 45--48, Kolkata, India. (*One of the selected best papers*)

Additional Experience

- Involved with the development of Hindi Named Entity Recognition system in the Microsoft Research India sponsored project “Named Entity Recognition, Parts of Speech tagging, G2P of Hindi” at Indian Institute of Technology Kharagpur (2006 – 2007).
- Involved with the development of Bengali Named Entity Recognition system in the Department of Information Technology, MCIT, Govt. of India sponsored project “Cross Language Information Access Portal for Indian Languages” at Indian Institute of Technology Kharagpur (2007 – 2008).
- Worked in the TNO Defence and Security sponsored project “Benchmark Text Mining Tools” at Indian Institute of Technology Kharagpur (December 2008 – May 2009).
- Organizing member of “Summer School on Natural Language Processing and Text Mining”, 2008, IIT Kharagpur, India.
- Served as teaching assistant in several courses like Programming and Data Structure Laboratory, Machine Learning, Natural Language Processing, at Computer Science and Engineering Department, IIT Kharagpur.

Personal Information

Present Address: Qtr No. A-307, Vikas Enclave, Near Booti More, Ranchi, Jharkhand, India.

Permanent Address: S/O. - Narayan Chandra Saha, Vill+PO. – Khajurdihi, Dist.– Burdwan, W.B. - 713518.

Date of Birth: January 23, 1982.

Sex: Male.

Marital Status: Married.