

Partha Pratim Talukdar

Suite 400A, 3401 Walnut Street
University of Pennsylvania, Philadelphia, PA-19104 (USA)
Email: partha[at]cis.upenn.edu
Web: <http://www.cis.upenn.edu/~partha>

Research Interests

I am primarily interested in Machine Learning and Computational Linguistics.

Academic Details

- **Doctoral Student** (*September 2004 - Present*)
CIS DEPARTMENT, UNIVERSITY OF PENNSYLVANIA
Advisors: Prof. Mark Liberman, Prof. Fernando Pereira.
- **B.E.(Hons.) Computer Science** (*Distinction*) (*August 1999 - June 2003*)
BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI (INDIA)

Awards & Honours

- Graduate Fellowship, CIS Department, University of Pennsylvania.
- Young Engineering Fellowship (YEF 2002-2003), awarded by Indian Institute of Science (IISc.), Bangalore to a total of 25 students in India.
- North Eastern Council (Govt. of India) Scholarship.
- Secured 2nd rank in the Engineering Entrance Test, Assam
- Secured state 3rd rank in 12th and 12th rank in the 10th standard Board Examinations.
- Received State and Central Government Scholarships for excellence in education.

Professional Experience

- **Research Intern** (*May 2008 - December 2008*)
Google, Mountain View (California).
Worked with Marius Pasca on large scale minimally supervised information extraction.
- **Research Intern** (*June 2006 - September 2006*)
Google, Mountain View (California).
Worked with Thorsten Brants and other members of the Machine Translation group at Google.
- **Research Intern** (*June 2005 - September 2005*)
Google, Mountain View (California).
Worked with Thorsten Brants on named entity extraction. Experimented with ways to combine labeled and unlabeled data using context pattern induction.
- **Research Engineer** (*July 2003 - June 2004*)
Hewlett-Packard Labs, Bangalore (India)
Worked on language processing and learning for building text-to-speech systems.

Publications

- Partha Pratim Talukdar, Joseph Reisinger, Marius Pasca, Deepak Ravichandran, Rahul Bhagat, Fernando Pereira. Weakly Supervised Acquisition of Labeled Class Instances using Graph Random Walks. *EMNLP 2008*, Honolulu, Hawaii.
- Partha Pratim Talukdar, Marie Jacob, Mohammad Salman Mehmood, Koby Crammer, Zack Ives, Fernando Pereira, Sudipto Guha. Learning to Create Data-Integrating Queries. 34th *International Conference on Very Large Databases (VLDB 2008)*, Auckland, New Zealand.
- Koby Crammer, Partha Pratim Talukdar, Fernando Pereira. A Rate-Distortion One-Class Model and its Applications to Clustering. *International Conference on Machine Learning (ICML) 2008*, Helsinki, Finland.
- Partha Pratim Talukdar, John Blitzer, Ted Sandler, Mark Dredze, Koby Crammer, Fernando Pereira. DRASO: Declaratively Regularized Alternating Structural Optimization. *ICML 2008 Workshop on Prior Knowledge for Text and Language Processing*, Helsinki, Finland.
- Kedar Bellare, Partha Pratim Talukdar, Giridhar Kumaran, Fernando Pereira, Mark Liberman, Andrew McCallum and Mark Dredze. Lightly-Supervised Attribute Extraction. *NIPS 2007 Workshop on Machine Learning for Web Search*, Whistler.
- Mark Dredze, John Blitzer, Partha Pratim Talukdar, Kuzman Ganchev, Joao Graca, and Fernando Pereira. Frustratingly Hard Domain Adaptation for Dependency Parsing. *CoNLL Shared Task Session of EMNLP-CoNLL 2007*, Prague.
- Koby Crammer, Mark Dredze, Kuzman Ganchev, Partha Pratim Talukdar and Steve Carroll. Automatic Code Assignment to Medical Text. *BioNLP 2007*, Prague.
- Partha Pratim Talukdar, Thorsten Brants, Mark Liberman and Fernando Pereira. A Context Pattern Induction Method for Named Entity Extraction. *Tenth Conference on Computational Natural Language Learning (CoNLL-X)*, New York City, June 8-9, 2006.
- Deepa S.R., A.G. Ramakrishnan, Kalika Bali, Partha Pratim Talukdar, "Automatic Generation of Compound Word Lexicon for Hindi Speech Synthesis", *Language Resources and Evaluation Conference (LREC 2004)*, Portugal, 26-28 May 2004.
- Kalika Bali, A.G.Ramakrishnan, Partha Pratim Talukdar, N. Sridhar Krishna, "Towards the Development of a Hindi Speech Synthesis System", *5th ISCA Speech Synthesis Workshop*, 14th-16th June 2004, Carnegie Mellon University, USA.
- N. Sridhar Krishna, Partha Pratim Talukdar, Kalika Bali, A.G. Ramakrishnan, "Duration Modeling for Hindi Text-to-Speech Synthesis, *8th International Conference on Spoken Language Language Processing (ICSLP)*, 4th-8th October 2004, Jeju Island, Korea.
- Satinder Singh, Partha Talukdar, Sridhar Krishna, Sandeep Manocha, Kalika Bali, Sitaram R.N.V., Optimal Creation of Speech Databases for Indian Language Speech Technology, *International Conference on Speech and Language Technology (O-COCOSDA)*, 17-19 November 2004, New Delhi, India.
- Sriram S., Partha Pratim Talukdar, Sameer Badaskar, Kalika Bali, A.G. Ramakrishnan, "Phonetic Distance Based Cross-lingual Search", *International Conference on Natural Language Processing*, 19-22 December 2004, Hyderabad India.

- K. Panchapagesan, Partha Pratim Talukdar, N. Sridhar Krishna, Kalika Bali, A.G. Ramakrishnan, "Hindi Text Normalization", *Fifth International Conference on Knowledge Based Computer Systems (KBCS)*, 19-22 December 2004, Hyderabad India.

Teaching Experience

- **CSE 112: Networked Life (Prof. Michael Kearns)** (Penn, Spring 2006): Duties included holding weekly recitation and office hour, setting and grading exams.
- **CSE 140: Introduction to Cognitive Science (Prof. Lyle Ungar & Prof. Virginia Richards)** (Penn, Fall 2005): Duties included holding weekly recitation and office hour, setting and grading homework and evaluating term paper.
- **BITS C481 Computer Networks** (BITS Pilani, Fall 2002): Duties included supervision of lab operations and preparation of a lab manual.

Invited Talks

- Hewlett Packard Labs (India), November 2007.
- Siemens Medical Research, January 2008.

Professional Activities

- Co-organizer of the Computational Linguistic Lunch (CLUNCH) at Penn, 2005-06.
- NESCAI 2008 PC member.
- Reviewer for ICON 2003, NESCAI 2006-2008, EMNLP 2008, NIPS 2008.

References

Available upon request.