

KAPIL THADANI

400 W 119th St #13L, New York NY 10027 • (917) 573-8173 • kapil@cs.columbia.edu • www.cs.columbia.edu/~kapil

EDUCATION

- Columbia University, Graduate School of Arts & Sciences** New York, NY
Candidate for PhD (Computer Science) Dec 2012
Research focus: Computational Linguistics; Advisor: Dr. Kathleen McKeown
- Columbia University, Fu Foundation School of Engineering & Applied Science** New York, NY
Master of Science (Computer Science); Machine Learning/Thesis track May 2007
Cumulative GPA 4.17; Thesis: “Decreasing Textual Redundancy”
- University of Mumbai, Thadomal Shahani Engineering College** Mumbai, India
Bachelor of Engineering (Information Technology) May 2005
Cumulative percentage 65.4%; First Class standing

CURRENT PROJECTS

- Columbia University Natural Language Processing Group** New York, NY
Production of fluent sentences and parse structures from input text under formal operations like intersection, union, simplification etc. Applications to paraphrasing/entailment, compression/summarization, alignment and machine translation.
- Columbia University Machine Learning Group** New York, NY
Unsupervised discovery of latent tree structure in text, with applications to semantic correlation extraction, language modeling and unsupervised parsing.

EXPERIENCE

- Google Inc** New York, NY
Software Engineering Intern (Research) May 2011 – Aug 2011
Automatic closed-captioning of Youtube videos. Experimented with interpolating ensembles of topic-specific language models to improve audio transcription quality.
- Software Engineering Intern (Search Quality)* June 2009 – Aug 2009
Geopoint annotation within a named-entity annotation pipeline. Developed a bootstrapping approach to improve identification of location mentions using neighboring context words.
- Columbia University Natural Language Processing Group** New York, NY
Graduate Researcher Jan 2006 – June 2009
Part of SRI International’s distillation effort in the DARPA-sponsored Global Autonomous Language Exploitation (GALE) program for phases 1 through 3. Studied automated identification of redundancy within QA system responses and syntax-independent semantic role labeling over disfluent text.
- Columbia University Center for Computational Learning Systems** New York, NY
Graduate Researcher Sept 2006 – Dec 2006
Analysis of the relationship between patterns of service requests and major electrical outages in New York City in collaboration with Con Edison of New York.

Columbia University Fu Foundation School of Engineering & Applied Science

New York, NY

Teaching Assistant

- Artificial Intelligence; Dr. German Creamer Summer 2006
- Search Engine Technology; Dr. Dragomir Radev (U. Michigan) Spring 2007
- Artificial Intelligence; Dr. Kathleen McKeown Fall 2007
- Statistical Methods for Natural Language Processing; Dr. Sameer Maskey (IBM) Spring 2010

University of Mumbai/Geodesic Information Systems Ltd.

Mumbai, India

Part-time/Co-op

June 2004 – May 2005

Design and prototyping of an application for IM/conferencing over Bluetooth networks for Palm devices, developed in collaboration with Geodesic Information Systems for the commercially available Mundu™ Interoperable Messenger for Palm.

PUBLICATIONS

W. Wang, K. Thadani and K. McKeown. “Identifying Event Descriptions using Co-training with Online News Summaries”. To appear in IJCNLP 2011.

K. Thadani and K. McKeown. “Towards Strict Sentence Intersection: Decoding and Evaluation Strategies”. In the Workshop on Monolingual Text-to-Text Generation at ACL-HLT 2011.

K. Thadani and K. McKeown. “Optimal and Syntactically Informed Decoding for Monolingual Phrase-Based Alignment”. In ACL-HLT 2011.

Y. Petinot, K. McKeown and K. Thadani. “A Hierarchical Model for Web Summarization”. In ACL-HLT 2011.

K. McKeown, S. Rosenthal, K. Thadani and C. Moore. “Time-Efficient Creation of an Accurate Sentence Fusion Corpus”. In NAACL-HLT 2010.

M. Jha, J. Andreas, K. Thadani, S. Rosenthal and K. McKeown. “Corpus Creation for New Genres: A Crowdsourced Approach to PP Attachment”. In the Workshop for Creating Speech and Text Language Data using Amazon’s Mechanical Turk at NAACL-HLT 2010.

S. Rosenthal, W. J. Lipovsky, K. McKeown, K. Thadani and J. Andreas. “Toward Semi-Automated Annotation of Prepositional Phrase Attachment”. In LREC 2010.

K. Thadani and K. McKeown. “A Framework for Decreasing Textual Redundancy”. In COLING 2008.

T. Jebara, Y. Song and K. Thadani. “Density Estimation under Independent Similarly Distributed Sampling Assumptions”. In NIPS 2007.

T. Jebara, Y. Song and K. Thadani. “Spectral Clustering and Embedding with Hidden Markov Models”. In ECML 2007.

TECHNICAL SKILLS

Python, C++, Java, Matlab, Perl, Common Lisp, Lex & YACC, LaTeX, Max/MSP