Mahesh Joshi

CONTACT INFORMATION

5000 Forbes Ave
Gates Hillman Complex, 5404
Pittsburgh, PA 15213

Voice: +1 (412) 623-9677
E-mail: maheshj@cs.cmu.edu
Web: http://www.cs.cmu.edu/~maheshj

EDUCATION

Ph.D. in Language and Information Technologies (May 2013, expected)
Carnegie Mellon University, Pittsburgh, PA

- **Thesis Title**: Multi-Domain Learning: Analysis, and Methods for Multi-Attribute Domains
- **Thesis Advisors**: Carolyn P. Rosé and William W. Cohen

Master of Science, Language Technologies (2008)
Carnegie Mellon University, Pittsburgh, PA

Master of Science, Computer Science (2006)
University of Minnesota Duluth, Duluth, MN

- **Thesis Title**: Kernel Methods for Word Sense Disambiguation and Abbreviation Expansion in the Medical Domain
- **Thesis Advisors**: Richard Maclin and Ted Pedersen

Bachelor of Engineering, Computer Engineering (2001)
Government College of Engineering, Pune, India

RESEARCH INTERESTS

Natural language processing (opinion mining and sentiment analysis, dialogue analysis, NLP for education, biomedical informatics), machine learning (multi-domain learning, domain adaptation, semi-supervised learning)

RESEARCH EXPERIENCE

Graduate Research Assistant [Sep 2006–present]
Language Technologies Institute, Carnegie Mellon University

- Multi-domain learning, learning in the presence of multiple domain assignments
- Opinion mining and sentiment analysis
- Language technologies for analysis of collaborative conversations
- Contributor to TagHelper Tools (http://www.cs.cmu.edu/~cprose/TagHelper.html)

Graduate Research Assistant [May 2006–Jul 2006]
Department of Computer Science, University of Minnesota Duluth

- Unsupervised word sense disambiguation
- Contributor to SenseClusters (http://senseclusters.sourceforge.net)

WORK EXPERIENCE

Software Engineering Intern [Jun 2010–Sep 2010]
Google Inc., Mountain View, CA

- Worked with the Comparison Ads team
- Developed techniques for sentiment analysis of reviews related to financial products such as credit cards

Research Intern [May 2005–Jul 2005]
Division of Biomedical Informatics, Mayo Clinic, Rochester MN.

- Automatic expansion of ambiguous acronyms in the biomedical domain
- Developed WSDGate (http://wsdgate.sourceforge.net)

Member of Technical Staff [Jul 2001–Jun 2004]
Persistent Systems, Pune, India

- Worked with the Data Management team for the Cerity Network Data System for Pharmaceutical QA/QC

Refereed Publications

1. **Mahesh Joshi**, Mark Dredze, William W. Cohen, and Carolyn P. Rosé. Multi-Domain Learning: When Do Domains Matter? In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP–CoNLL)*, 2012.
2. Ted Pedersen, Satanjeev Banerjee, Bridget McInnes, Saiyam Kohli, **Mahesh Joshi**, and Ying Liu. The Ngram Statistics Package (Text::NSP) – A Flexible Tool for Identifying Ngrams, Collocations, and Word Associations. (Demonstration System). In *Proceedings of Multiword Expressions : from Parsing and generation to the Real World (MWE 2011), an ACL HLT 2011 Workshop*, 2011.
3. **Mahesh Joshi**, Dipanjan Das, Kevin Gimpel, and Noah A. Smith. Movie Reviews and Revenues: An Experiment in Text Regression. In *Proceedings of Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics (HLT–NAACL)*, 2010.
4. **Mahesh Joshi** and Carolyn P. Rosé. Generalizing Dependency Features for Opinion Mining. In *Proceedings of the joint conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing (ACL–IJCNLP)*, 2009.
5. Shilpa Arora, **Mahesh Joshi**, and Carolyn P. Rosé. Identifying Types of Claims in Online Customer Reviews. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics (HLT–NAACL), Companion Volume: Short Papers*, 2009.
6. **Mahesh Joshi**, Yi-Chia Wang, John Wilkerson, and Carolyn P. Rosé. A Needs Analysis for Instructional Support in LegSim. In *Proceedings of the 2008 International Conference of the Learning Sciences (ICLS), Poster Session*, 2008.
7. Yi-Chia Wang, **Mahesh Joshi**, Carolyn P. Rosé, and William Cohen. Recovering Implicit Thread Structure in Newsgroup Style Conversations. In *Proceedings of the International Conference on Weblogs and Social Media (ICWSM)*, 2008.
8. Sourish Chaudhuri, Rohit Kumar, **Mahesh Joshi**, Elon Terrell, Fred Higgs, Vincent Aleven, and Carolyn P. Rosé. It's Not Easy Being Green: Supporting Collaborative "Green Design" Learning. In *Proceedings of Intelligent Tutoring Systems (ITS)*, 2008.
9. **Mahesh Joshi** and Carolyn P. Rosé. Using Transactivity in Conversation for Summarization of Educational Dialog. In *Proceedings of the Workshop on Speech and Language Technology in Education (SLaTE)*, 2007.
10. Rohit Kumar, Gahgene Gweon, **Mahesh Joshi**, Yue Cui, and Carolyn P. Rosé Supporting Students Working Together on Math with Social Dialogue. In *Proceedings of the Workshop on Speech and Language Technology in Education (SLaTE)*, 2007.
11. Rohit Kumar, Carolyn P. Rosé, Yi-Chia Wang, **Mahesh Joshi**, and Allen Robinson. Tutorial Dialogue as Adaptive Collaborative Learning Support. In *Proceedings of Artificial Intelligence in Education (AIED)*, 2007.
12. Yi-Chia Wang, **Mahesh Joshi**, and Carolyn P. Rosé. A Feature Based Approach to Leveraging Context for Classifying Newsgroup Style Discussion Segments. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL), Companion Volume Proceedings of the Demo and Poster Sessions*, 2007.

13. Yi-Chia Wang, Carolyn P. Rosé, **Mahesh Joshi**, Frank Fischer, Armin Weinberger, and Karsten Stegmann. Context Based Classification for Automatic Collaborative Learning Process Analysis. In *Proceedings of Artificial Intelligence in Education (AIED), Poster Session*, 2007.
14. **Mahesh Joshi**, Serguei Pakhomov, Ted Pedersen, and Christopher Chute. A Comparative Study of Supervised Learning as Applied to Acronym Expansion in Clinical Reports. In *Proceedings of the Annual Symposium of the American Medical Informatics Association (AMIA)*, 2006.
15. **Mahesh Joshi**, Serguei Pakhomov, Ted Pedersen, Richard Maclin, and Christopher Chute. An End-to-end Supervised Target-Word Sense Disambiguation System. In *Proceedings of the 21st national conference on Artificial intelligence (AAAI), Intelligent Systems Demonstrations*, 2006.
16. **Mahesh Joshi**, Ted Pedersen, Richard Maclin, and Serguei Pakhomov. Kernel Methods for Word Sense Disambiguation, and Acronym Expansion. In *Proceedings of the 21st national conference on Artificial intelligence (AAAI), Student Abstract Papers*, 2006.
17. **Mahesh Joshi**, Ted Pedersen, and Richard Maclin. A Comparative Study of Support Vector Machines Applied to the Supervised Word Sense Disambiguation Problem in the Medical Domain. In *Proceedings of the Indian International Conference on Artificial Intelligence (IICAI)*, 2005.

Computer Skills

- Languages: Java, C++, Python, Perl
- Software/Packages/Frameworks: MapReduce (Google, Hadoop), Matlab, R, Mercurial/SVN/CVS
- Data Mining Toolkits: MinorThird, Scikit-Learn, Weka
- Operating Systems: Linux, Mac OS, Windows

Scholarships and Awards

- Honorable Mention at the LTI Student Research Symposium at Carnegie Mellon University (2007)
- Graduate Research Assistantship, Carnegie Mellon University (Fall 2006–present)
- Graduate Teaching Assistantship, Carnegie Mellon University (Fall 2008, Spring 2010)
- Graduate Research Assistantship, University of Minnesota Duluth (2005–2006)
- Graduate Teaching Assistantship, University of Minnesota Duluth (2004–2005)
- National Talent Search Scholarship, India (1995–2001)
- Dhirubhai Ambani Foundation Undergraduate Scholarship (1997–2001)
- National Merit Scholarship, India (1997-98)
- Maharashtra State Merit Scholarship (1997-98)

Teaching Experience

Teaching Assistant, Spring 2010
• Machine Learning in Practice
Carnegie Mellon University

Teaching Assistant, Fall 2008
• Research Design and Writing
Carnegie Mellon University

Teaching Assistant, Spring 2005
• Operating Systems Practicum
• Visual Basic .NET
University of Minnesota Duluth

Teaching Assistant, Fall 2004
• Database Management Systems
• Visual Basic .NET
University of Minnesota Duluth

Professional Service

- Reviewer, Journal of Machine Learning Research
- NLP co-chair, Student Research Workshop, HLT–NAACL 2010

Community Service

- Presenter and facilitator for "Life Skills and Problem Solving" workshops at the Shuman Juvenile Detention Center (2007–2012)

References

Carolyn P. Rosé
Associate Professor
Language Technologies Institute
Human-Computer Interaction Institute
Carnegie Mellon University
Mailing Address:
5000 Forbes Ave.
5404 Gates Hillman Complex
Pittsburgh, PA 15213, USA
Voice: +1 (412) 268-7130
Fax: +1 (412) 268-6298
E-mail: cprose@cs.cmu.edu
Web: www.cs.cmu.edu/~cprose

William W. Cohen
Research Professor
Machine Learning Department
Carnegie Mellon University

Mailing Address:
5000 Forbes Ave.
6105 Gates Hillman Complex
Pittsburgh, PA 15213, USA
Voice: +1 (412) 268-7664
Fax: +1 (412) 268-2205
E-mail: wcohen@cs.cmu.edu
Web: www.cs.cmu.edu/~wcohen

Noah A. Smith
Associate Professor
Language Technologies Institute
Carnegie Mellon University
Mailing Address:
5000 Forbes Ave.
5404 Gates Hillman Complex
Pittsburgh, PA 15213, USA
Voice: +1 (412) 268-4963
Fax: +1 (412) 268-6298
E-mail: nasmith@cs.cmu.edu
Web: www.cs.cmu.edu/~nasmith

Mark Dredze
Assistant Research Professor
Department of Computer Science
Johns Hopkins University
Mailing Address:
HLTCOE
810 Wyman Park Drive
Baltimore, MD 21211, USA
Voice: +1 (410) 516-6786
Fax: +1 (410) 516-6700
E-mail: mdredze@cs.jhu.edu
Web: www.cs.jhu.edu/~mdredze