

Sameer S. Pradhan

CONTACT INFORMATION 700 Huron Avenue Voice: (720) 771-2945
Apartment 14F Internet: <http://cemantix.org>
Cambridge, MA 02138 E-mail: sameer@cemantix.org
USA

RESEARCH INTERESTS Natural Language Processing, Machine Learning, Annotation Science

EDUCATION **Ph.D., Computer Science**, 2005
University of Colorado, Boulder, CO
Thesis: Robust Semantic Role Labeling
Advisors: Professors Wayne Ward, James H. Martin and Daniel Jurafsky

M.S., Industrial Engineering, 1997
Alfred University, Alfred, NY
Thesis: Virtual Manufacturing Information System Using Java and JDBC

B.E., Production Engineering, 1994
Victoria Jubilee Technical Institute, Bombay
Thesis: Rejection Analysis of Manufactured Components

RESEARCH EXPERIENCE **Scientist** October 2005 – present
Speech and Language Processing,
BBN Technologies

- Data representation, management and planning for the DARPA GALE multi-site collaborative effort – OntoNotes.
- Developing robust statistical models for semantic role labeling, coreference resolution and word sense disambiguation

Research Assistant May 2000 - September 2005
Center for Spoken Language Research,
Department of Computer Science,
University of Colorado at Boulder

- Improving the state-of-the-art in robust, open-domain natural language semantic role labeling.
- Implemented a semantic role tagger (ASSERT) that assigns a semantic argument structure to plain text.
- Designed and implemented an open-domain natural language question answering system – CUAQ.
- Investigated metrics for assigning semantic confidence to spoken dialog interactions.

Research Assistant January 2000 - May 2000
Department of Math & Computer Science,
Colorado School of Mines

- Investigated fuzzy clustering algorithms for retrieving documents on the Internet.

Research Assistant

November 1996 - September 1997

Department of Industrial Engineering,
Alfred University

- o Designed and implemented a prototype three-tier virtual manufacturing information system using Java Remote Method Invocation (RMI) and Java Database Connectivity (JDBC) APIs.

TEACHING EXPERIENCE

Teaching Assistant

January 1999 - January 2000

Department of Math & Computer Science,
Colorado School of Mines

- o Provided teaching assistance for Programming Languages (MACS 400), C++ Programming (MACS 261A/C), Design and Analysis of Algorithms (MACS 406) and Software Engineering (MACS 567).

PUBLICATIONS

2009

- o **OntoNotes: A Large Training Corpus for Enhanced Processing**
Ralph Weischedel, Eduard Hovy, Mitchell Marcus, Martha Palmer, Robert Belvin, **Sameer Pradhan**, Lance Ramshaw and Nianwen Xue.
Book chapter for DARPA Global Automatic Language Exploitation (GALE), submitted.

2008

- o **Towards Robust Semantic Role Labeling**
Sameer Pradhan, Wayne Ward and James H. Martin
Computational Linguistics: Special Issue on Semantic Role Labeling, Vol. 34, No. 2, June 2008.

2007

- o **OntoNotes: A Unified Relational Semantic Representation**
Sameer Pradhan, Eduard Hovy, Mitchell Marcus, Martha Palmer, Lance Ramshaw, and Ralph Weischedel
International Journal of Semantic Computing, Vol. 1, No. 4, 2007.
- o **Unrestricted Coreference: Identifying Entities and Events in OntoNotes**
Sameer Pradhan, Lance Ramshaw, Ralph Weischedel, Jessica MacBride and Linnea Micciulla
in Proceedings of the International Conference on Semantic Computing, Irvine, CA, Sep. 17-19, 2007.
- o **OntoNotes: A Unified Relational Semantic Representation [Best Paper Award]**
Sameer Pradhan, Eduard Hovy, Mitchell Marcus, Martha Palmer, Lance Ramshaw and Ralph Weischedel
in Proceedings of the International Conference on Semantic Computing, Irvine, CA, Sep. 17-19, 2007.
- o **SemEval-2007 Task-17: English Lexical Sample, SRL and All Words**
Sameer Pradhan, Eduard Loper, Dmitriy Dligach and Martha Palmer
in Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007), Prague, Czech Republic, June, 2007.
- o **Towards Robust Semantic Role Labeling**
Sameer Pradhan, Wayne Ward and James H. Martin
in Proceedings of the Human Language Technology Conference (HLT/NAACL 2007),

Rochester, NY, April 22-27, 2007.

2006

- **Robust Semantic Role Labeling**
Sameer Pradhan
PhD Thesis, University of Colorado, Boulder, CO, 2006

2005

- **Support Vector Learning for Semantic Argument Classification**
Sameer Pradhan, Kadri Hacioglu, Valerie Krugler, Wayne Ward, James H. Martin, Daniel Jurafsky
Machine Learning Journal, Volume 60, Issue 1-3, 11–39, September 2005.
- **Semantic Role Labeling Using Different Syntactic Views**
Sameer Pradhan, Wayne Ward, Kadri Hacioglu, James H. Martin and Daniel Jurafsky
in Proceedings of the Association for Computational Linguistics (ACL-2005), Ann Arbor, MI, June 25-30, 2005.
- **Semantic Role Chunking Combining Complementary Syntactic Views**
Sameer Pradhan, Kadri Hacioglu, Wayne Ward, James H. Martin and Daniel Jurafsky
Proceedings of the Ninth Conference on Computational Natural Language Learning (CoNLL-2005), Ann Arbor, MI, June 29-30, 2005.

2004

- **Shallow Semantic Parsing using Support Vector Machines**
Sameer Pradhan, Wayne Ward, Kadri Hacioglu, James H. Martin and Daniel Jurafsky
Proceedings of the Human Language Technology Conference/North American chapter of the Association for Computational Linguistics annual meeting (HLT/NAACL-2004), Boston, MA, May 2-7, 2004.
- **Parsing Arguments of Nominalizations in English and Chinese**
Sameer Pradhan, Honglin Sun, Wayne Ward, James H. Martin and Daniel Jurafsky
Proceedings of the Human Language Technology Conference/North American chapter of the Association for Computational Linguistics annual meeting (HLT/NAACL-2004), Boston, MA, May 2-7, 2004.
- **Semantic Role Labeling by Tagging Syntactic Chunks**
Kadri Hacioglu, **Sameer Pradhan**, Wayne Ward, James H. Martin and Daniel Jurafsky
Proceedings of the Eighth Conference on Natural Language Learning (CONLL-2004), Boston, MA, May 6-7, 2004
- **Mixing Weak Learners in Semantic Parsing**
Rodney Nielsen and **Sameer Pradhan**
Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-2004), Barcelona, Spain, July 25-26, 2004.

2003

- **Semantic Role Parsing: Adding Semantic Structure to Unstructured Text**
Sameer Pradhan, Kadri Hacioglu, Wayne Ward, James H. Martin, Daniel Jurafsky
Proceedings of the International Conference on Data Mining (ICDM-2003), Melbourne, FL, Nov. 19-22, 2003.

Before 2003

- **Building a Foundation System for Producing Short Answers to Factual Questions**
Sameer S. Pradhan, Gabriel Illouz, Sasha J. Blair-Goldensohn, Andrew Hazen Schlaikjer, Valerie Krugler, Elena Filatova, Pablo A. Duboue, Hong Yu, Rebecca J. Passonneau, Steven Bethard, Vasileios Hatzivassiloglou, Wayne Ward, Dan Jurafsky, Kathleen R. McKeown, and James H. Martin
Proceedings of Text REtrieval Conference (TREC), Gaithersburg, MD, Nov. 19-22, 2002.
- **Using Semantic Representations in Question Answering**
Sameer Pradhan, Valerie Krugler Wayne Ward, Daniel Jurafsky and James H. Martin
Proceedings of the International Conference on Natural Language Processing (ICON-2002), Bombay, India, Dec. 2002.
- **Estimating Semantic Confidence for Spoken Dialogue Systems**
Sameer Pradhan and Wayne Ward
Proceedings of the International Conference on Acoustic Speech and Signal Processing (ICASSP-2002), Orlando, Florida, May 2002.
- **University of Colorado Dialog Systems for Travel and Navigation**
Bryan Pellom, Wayne Ward, John Hansen, Kadri Hacioglu, Jianping Zhang, Xiuyang Yu and **Sameer Pradhan**
Proceedings of the Human Language Technology Conference (HLT-2001), San Diego, March 2001.
- **The CU Communicator: An Architecture for Dialogue Systems**
Bryan Pellom, Wayne Ward and **Sameer Pradhan**
Proceedings of the sixth International Conference on Spoken Language Processing, Beijing, China, Oct. 2000.
- **Virtual Manufacturing Information System Using Java and JDBC**
Sameer Pradhan and Wilfred Huang
International Journal of Computers and Industrial Engineering, Vol. 35 Nos. 1-2, 1998.

AWARDS AND FELLOWSHIPS

Best Paper Award

- For *OntoNotes: A Unified Relational Semantic Representation* at the International Conference on Semantic Computing, 2007.

NCST Scholarship

- Award and Scholarship for ranking in top 1 percentile at the nation-wide entrance exam of National Center for Software Technology, 1996.

National Merit Scholarship

- For Higher Secondary Education, India. 1988.

INVITED TALKS

OntoNotes: A Unified Relational Semantic Representation – Unified Linguistic Annotation (ULA) Workshop at Brandeis University, 2007.

THESIS SUPERVISION

Committee Member

- Alex Baron, Ph.D., Computer Science, Brandeis University, 2007.
Adviser James Pustejovsky

- Linnea Micciulla, Ph.D., Linguistics, Boston University, TBD.
Adviser Cathy O'Connor

PROFESSIONAL
ACTIVITIES

Guest Editorial Board

- Computational Linguistics Journal – Special Issue on Semantic Role Labeling

Organizing Committee

- Linguistic Annotation Workshop (LAW II), 2008
- SemEval tasks on Word Sense Disambiguation and Semantic Role Labeling, 2007

Program Committee Member

- International Conference on Global Interoperability for Language Resources

Executive Committee Member/Collaborator

- Special Interest Group for Annotation (SIGANN)

Tutorials

- OntoNotes: The 90% Solution – HLT/NAACL Conference, Boulder, CO, 2009

Guest Lectures

- Spring 2009: Topics in Natural Language Processing, Department of Computer Science, Brandies University

Invited Team Member

- Johns Hopkins Workshop on *Exploiting Lexical & Encyclopedic Resources for Entity Disambiguation*, 2007

Journal Reviewer

- Computational Linguistics Journal
- Journal of Artificial Intelligence Research
- ACM Transactions on Asian Language Information Processing
- Knowledge and Information Systems: An International Journal

Conference Reviewer

- HLT/NAACL: 2006, 2007
- ACL: 2006, 2008, 2009
- EMNLP: 2007
- ICGL: 2007
- SemEval: 2007

RESEARCH
SOFTWARE

Automatic Statistical SEMantic Role Tagger (ASSERT)

- A toolkit for assigning arguments to all verb predicates in a document. This is available under LGPL at <http://cemantix.org/assert>. It has been downloaded by hundreds of researchers worldwide, and about a dozen published papers have used it in their experiments.

OntoNotes database tool (ontonotes-db-tool)

- o This is the tool that reads in the different layers of semantics from the individual OntoNotes data files, performs various consistency checks on the objects, and adds them to a MySQL database. This tool provides an API that would significantly reduce data processing/cleaning time for researchers using OntoNotes. This is distributed along with the OntoNotes corpus through LDC, but is also available for download separately to provide bug-fixes and incremental updates.

OTHER
EMPLOYMENT

Software Consultant
IMI Systems

September 1997 - January 1999

- o Involved in the design and implementation of a three-tier marketing information system, at MCI Inc., using JDK 1.1, JDBC on Windows NT 4.0 and INFORMIX database.

Information Technology Consultant
Nagarwalla Consultants

September 1995 - March 1996

- o Involved in the design and implementation of a distribution resource planning system for a large scale manufacturing organization.

Engineer
Chemtex Engineering

July 1994 - September 1995

- o Developing material procurement policy and system for a turnkey project.

CITIZENSHIP

Citizen of India. Permanent resident of U.S.

REFERENCES

Aravind Joshi, Ph.D. joshi@seas.upenn.edu
Henry Salvatori Professor of Computer and Cognitive Science, Phone: (215) 898-0359
University of Pennsylvania, Fax: (215) 898-0587
Philadelphia, PA 19104

James Martin, Ph.D. martin@colorado.edu
Professor of Computer Science, Phone: (303) 492 3552
University of Colorado, Fax: (303) 492-2844
Boulder, CO 80309

Mitchell Marcus, Ph.D. mitch@cis.upenn.edu
RCA Professor of Artificial Intelligence, Phone: (215) 898-2538
University of Pennsylvania, Fax: (215) 898-0587
Philadelphia, PA 19104

Martha Palmer, Ph.D. mpalmer@colorado.edu
Associate Professor of Linguists and Computer Science, Phone: (303) 492 1300
University of Colorado, Fax: (303) 492-4416
Boulder, CO 80309

Wayne Ward, Ph.D. whw@colorado.edu
Research Professor, Department of Computer Science, Phone: (303) 735 5070
University of Colorado, Fax: (303) 735-5072
Boulder, CO 80309