

# Siddharth Patwardhan

144 Pine Wood Ln  
Los Gatos, CA 95032.

(347) 673-3838, (347) 414-7439  
sidd@patwardhans.net  
<http://www.patwardhans.net>

---

## HIGHLIGHTS

- ✓ More than eleven years of (academic and industrial) experience in Natural Language Processing research
  - ✓ Member of research team that created IBM's Watson system
  - ✓ Over 25 scientific publications, cited more than 1000 times
  - ✓ Member of program committees of top NLP conferences and reviewer for top scientific journals in AI
  - ✓ NCERT National Talent Search scholar
- 

## EDUCATION

*University of Utah, Salt Lake City, UT*  
**PhD, Computer Science**

**May 2010**

*University of Minnesota, Duluth, MN*  
**MS, Computer Science**

**Aug 2003**

*University of Pune, India*  
**BE, Computer Engineering**

**May 2001**

---

## EMPLOYMENT HISTORY

**NLP and Machine Learning Engineer**  
Apple, Cupertino, CA

**Sep 2016 – present**

**Research Staff Member**  
IBM Watson Group, Yorktown Heights, NY

**Oct 2014 – Aug 2016**

**Senior Software Engineer**  
IBM T. J. Watson Research Center, Yorktown Heights, NY

**Feb 2012 – Oct 2014**

**Post Doctoral Researcher**  
IBM T. J. Watson Research Center, Hawthorne, NY

**Oct 2009 – Feb 2012**

**Research Assistant**  
School of Computing, University of Utah, Salt Lake City, UT

**Aug 2003 – Oct 2009**

**Summer Research Intern**  
IBM T. J. Watson Research Center, Hawthorne, NY

**May 2007 – Aug 2007**

**Summer Research Intern**  
Division of Biomedical Informatics, Mayo Clinic, Rochester, MN

**May 2003 – Aug 2003**

**Teaching Assistant**  
Department of Computer Science, University of Minnesota, Duluth, MN

**Aug 2001 – May 2003**

---

## PROJECTS

**Watson for Healthcare:** Lead the research and development of question answering technology and NLP technology applied to electronic health records. **2011-2016**

**WatsonPaths:** Contributed to adaptation of IBM Watson technology to automatically answer "scenario-based" questions (in the medical domain), through inference over unstructured data. **2011-2014**

**IBM Jeopardy! Challenge:** Conducted research and development for the creation of IBM's Watson question answering system that competed (and won) against humans on *Jeopardy!* **2009-2011**

- Machine Reading Project:** Researched and developed information extraction components used within a DARPA-funded project on combining information extraction and reasoning to read, understand and answer queries about text documents. [Java, UIMA, RDF; publications at EACL '12 and COLING '10] **2009-2011**
- GLACIER Information Extraction System:** Created a unified probabilistic model for event extraction from text, using sentential event recognition and local contextual evidence to extract role fillers of events described in text. [C++ and Java; publication at EMNLP '09] **2009**
- PIPER Information Extraction System:** Lead the design and implementation of a two-phase system for extracting event information from text by combining global relevance cues with local contextual evidence. [C++; publication at EMNLP-CoNLL '07] **2007**
- Learning IE Patterns from the Web:** Created a system for improving the coverage of textual Information Extraction by learning new extraction patterns from the Web texts. [C++; publication at an ACL '06 workshop] **2006**
- Subsumption Hierarchy for Opinion Analysis:** Implemented a hierarchical structure and code for selecting the best combination of lexical and syntactic features in a system for detecting opinions in text. [C++; publication at EMNLP '06] **2006**
- Opinion Source Identification:** Developed a system for identifying the holders of opinions in free text using lexico-syntactic features in a sequence labeling machine learning system. [C++; OpinionFinder system demo at EMNLP '05] **2005**
- Generative Model for Role Labeling:** Contributed to a machine-learning system (trained using FrameNet) for automatically assigning semantic roles to words in text. [Java; participation in SENSEVAL-3 evaluation] **2004**
- WordNet::SenseRelate:** A system that determines the intended meaning of a word from its context using the notion of Semantic Relatedness of concepts. [Perl; system demo at ACL '05 and AAAI '05] **2004**
- Semantic Similarity of Biomedical Concepts:** An adaptation of existing Computational Linguistics research to the biomedical domain, for measuring the semantic similarity of biomedical terms and concepts. [Perl; article in Journal of Biomedical Informatics] **2003**
- WordNet::Similarity:** Lead the design and implementation of several algorithms for automatically measuring the semantic similarity and relatedness of English words and concepts. [Perl; system demo at NAACL '04 and AAAI '04] **2003**

#### PROFESSIONAL SERVICE AND AWARDS

- Outstanding Technical Contribution Award**, IBM Watson Group (2014)
- AAAI Feigenbaum Prize** (to the Watson team), the Association for Advancement of Artificial Intelligence (2013)
- Area Chair (Information Retrieval and Question Answering)**, COLING '14
- Senior Program Committee Member**, IJCAI '16
- Publications Chair**, EMNLP '17, EMNLP '16
- Advisory Board**, DropThought Inc. (2011 – 2016)
- Program Committee**, ACL '17, EACL '17, AAAI '17, COLING '16, IJCAI '16, NAACL '16, ACL '16, AMIA CRI '16, \*SEM '16, EMNLP '15, ACL-IJCNLP '15, NAACL '15, \*SEM '15, EMNLP '14, ACL '14, EACL '14, LREC '14, SIGIR '14 Workshop on Semantic Matching in Information Retrieval, IJCNLP '13, ACL '13, NAACL-HLT '13, COLING '12, EMNLP-CoNLL '12, ACL '12, NAACL-HLT '12, RANLP '11 Workshop on Information Extraction and Knowledge Acquisition, EMNLP '11, ACL-HLT '11, EMNLP '10, COLING '10, NAACL-HLT '10, ACL '08
- Reviewer**, IEEE Transactions on Audio, Speech and Language Processing, Transactions of the Association for Computational Linguistics ('17, '16, '15, '14, '13), Artificial Intelligence ('15), Natural Language Engineering ('15, '13), Language Resources and Evaluation ('12), Data Mining and Knowledge Discovery ('11), Journal of Artificial Intelligence Research ('09), NESCAI ('08, '07)
- Auxiliary Reviewer**, AAAI '16, CIKM '14, AAAI '06.

**Session Chair**, CIKM '12, ACL '06 Workshop on *Information Extraction Beyond the Document*

**Graduate Student Advisory Committee**, School of Computing, University of Utah (2004)

**Most Outstanding Teaching Assistant**, University of Minnesota Duluth, MN (2003)

**National Talent Search Scholarship**, National Council for Educational Research & Training, India (1995 – 2001)

---

#### PATENTS

- D. Ferrucci, D. Gondek, A. Kalyanpur, A. Lally, and S. Patwardhan. **Fact Checking Using and Aiding Probabilistic Question Answering**. *U. S. Patent 8,959,043 B2*. February 2015.
- S. Berajawala, A. Levas, S. Patwardhan, and D. Taieb. **Collaborative Creation of Annotated Training Data**. *U. S. Patent Application 14/553,028*, filed November 2014. Patent Pending.
- B. Boguraev, D. Buchanan, J. Chu-Carroll, D. Ferrucci, A. Kalyanpur, J. W. Murdock, and S. Patwardhan. **Generating Secondary Questions in an Introspective Question Answering System**. *U. S. Patent Application 13/610,267*, filed September 2012. Patent Pending.
- 

#### JOURNAL ARTICLES

- B. Boguraev, S. Patwardhan, A. Kalyanpur, J. Chu-Carroll, and A. Lally. **Parallel and Nested Decomposition for Factoid Questions**. *Natural Language Engineering*, 20(04):441-468, October 2014.
- A. Kalyanpur, S. Patwardhan, B. Boguraev, A. Lally, and J. Chu-Carroll. **Fact-based Question Decomposition in DeepQA**. *IBM Journal of Research and Development*, 56(3/4):13:1-13:11, May/July 2012.
- A. Kalyanpur, B. Boguraev, S. Patwardhan, J. W. Murdock, A. Lally, C. Welty, J. Prager, B. Coppola, A. Fokoue-Nkoutche, L. Zhang, Y. Pan, and Z. Qiu. **Structured Data and Inference in DeepQA**. *IBM Journal of Research and Development*, 56(3/4):10:1-10:14, May/July 2012.
- A. Lally, J. Prager, M. McCord, B. Boguraev, S. Patwardhan, J. Fan, P. Fodor, and J. Chu-Carroll. **Question Analysis: How Watson Reads a Clue**. *IBM Journal of Research and Development*, 56(3/4):2:1-2:14, May/July 2012.
- T. Pedersen, S. Pakhomov, S. Patwardhan, and C. Chute. **Measures of Semantic Similarity and Relatedness in the Biomedical Domain**. *Journal of Biomedical Informatics*, 40(3):288-299, June 2007.
- T. Kwon, N. Dhruv, S. Patwardhan, and E. Kwon. **Common Data Format Archiving of Large-Scale Intelligent Transportation Systems Data for Efficient Storage, Retrieval, and Portability**. *Journal of the Transportation Research Board: Transportation Research Record 1836*, pages 111-117, 2003.

#### REFEREED CONFERENCE PAPERS

- C. Shivade, P. Raghavan, and S. Patwardhan. **Addressing Limited Data for Textual Entailment across Domains**. In *Proceedings of the 54<sup>th</sup> Annual Meeting of the Association for Computational Linguistics*, Berlin, Germany, August 2016. [acceptance rate 28%]
- O. Melamud, D. McClosky, S. Patwardhan, and M. Bansal. **The Role of Context Types and Dimensionality in Learning Word Embeddings**. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technology*, San Diego, CA, June 2016. [acceptance rate 25%]
- A. Moschitti, S. Patwardhan, and C. Welty. **Long-Distance Time-Event Relation Extraction**. In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, pages 1330-1338, Nagoya, Japan, October 2013. [acceptance rate 23%]
- S. Patwardhan, B. Boguraev, A. Agarwal, A. Moschitti, and J. Chu-Carroll. **Labeling by Landscaping: Classifying Tokens in Context by Pruning and Decorating Trees**. In *CIKM '12: Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, Maui, HI, October 2012. [acceptance rate 13%]
- D. Hovy, J. Fan, A. Gliozzo, S. Patwardhan, and C. Welty. **When Did that Happen? - Linking Events and Relations to Timestamps**. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 185-193, Avignon, France, April 2012. [acceptance rate 27%]

- A. Kalyanpur, S. Patwardhan, B. Boguraev, A. Lally, and J. Chu-Carroll. **Fact-based Question Decomposition for Candidate Answer Re-ranking.** In *Proceedings of the 20<sup>th</sup> ACM Conference on Information and Knowledge Management*, Glasgow, UK, October 2011. [acceptance rate 35%]
- A. Moschitti, J. Chu-Carroll, S. Patwardhan, J. Fan, and G. Riccardi. **Using Syntactic and Semantic Structural Kernels for Classifying Definition Questions in Jeopardy!** In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, Edinburgh, UK, July 2011. [acceptance rate 23%]
- R. Kate, X. Luo, S. Patwardhan, M. Franz, R. Florian, R. Mooney, S. Roukos, and C. Welty. **Learning to Predict Readability using Diverse Linguistic Features.** In *Proceedings of the 23rd International Conference on Computational Linguistics*, Beijing, China, August 2010. [acceptance rate 22%]
- R. Bhagat, E. Hovy, and S. Patwardhan. **Acquiring Paraphrases from Text Corpora.** In *Proceedings of the Fifth International Conference on Knowledge Capture*, pages 161-168, Redondo Beach, CA, September 2009. [acceptance rate 26%]
- S. Patwardhan and E. Riloff. **A Unified Model of Phrasal and Sentential Evidence for Information Extraction.** In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 151-160, Singapore, August 2009. [acceptance rate 34%]
- Y. Park, S. Patwardhan, K. Visweswariah, and S. Gates. **An Empirical Analysis of Word Error Rate and Keyword Error Rate.** In *Proceedings of the International Conference on Spoken Language Processing*, pages 2070-2073, Brisbane, Australia, September 2008. [acceptance rate 60%]
- S. Patwardhan and E. Riloff. **Effective Information Extraction with Semantic Affinity Patterns and Relevant Regions.** In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 717-727, Prague, Czech Republic, June 2007. [acceptance rate 27%]
- E. Riloff, S. Patwardhan, and J. Wiebe. **Feature Subsumption for Opinion Analysis.** In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pages 440-448, Sydney, Australia, July 2006. [acceptance rate 31%]
- Y. Choi, C. Cardie, E. Riloff, and S. Patwardhan. **Identifying Sources of Opinions with Conditional Random Fields and Extraction Patterns.** In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 355-362, Vancouver, Canada, October 2005. [acceptance rate 32%]
- S. Patwardhan, S. Banerjee, and T. Pedersen. **Using Measures of Semantic Relatedness for Word Sense Disambiguation.** In *Proceedings of the Fourth International Conference on Intelligent Text Processing and Computational Linguistics*, pages 241-257, Mexico City, Mexico, February 2003. [acceptance rate 46%]
- N. Dhruv, T. Kwon, and S. Patwardhan. **CDF Archival of Large-Scaled ITS Data for Efficient Archival, Retrieval, and Portability.** In *Proceedings of the Transportation Research Board 82nd Annual Meeting*, Washington, DC, January 2003.

#### REFEREED WORKSHOP PAPERS

- P. Raghavan and S. Patwardhan. **Question Answering on Electronic Medical Records.** In *Podium Abstract Presentations at AMIA 2016 Summit on Clinical Research Informatics*, San Francisco, CA, March 2016.
- A. Gliozzo, O. Biran, S. Patwardhan, and K. McKeown. **Semantic Technologies in IBM Watson.** In *Proceedings of the Fourth Workshop on Teaching NLP and CL*, pages 85-92, Sofia, Bulgaria, August 2013.
- S. Patwardhan. **Combining Global Relevance Information with Local Contextual Clues for Event-Oriented Information Extraction.** In *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence (Doctoral Consortium)*, pages 1863-1864, Chicago, IL, July 2008.
- S. Patwardhan and E. Riloff. **Learning Domain-Specific Information Extraction Patterns from the Web.** In *Proceedings of the ACL 2006 Workshop on Information Extraction Beyond the Document*, pages 66-73, Sydney, Australia, July 2006.
- S. Patwardhan and T. Pedersen. **Using WordNet-based Context Vectors to Estimate the Semantic Relatedness of Concepts.** In *Proceedings of the EACL 2006 Workshop on Making Sense of Sense: Bringing Computational Linguistics and Psycholinguistics Together*, pages 1-8, Trento, Italy, April 2006.

## SYSTEM DEMOS AND NON-REFEREED PAPERS

- C. Shivade, P. Raghavan and S. Patwardhan. **Addressing Limited Data for Textual Entailment Across Domains.** In *Preprint arXiv: 1606.02638*, June 2016.
- O. Melamud, D. McClosky, S. Patwardhan and M. Bansal. **The Role of Context Types and Dimensionality in Learning Word Embeddings.** In *Preprint arXiv: 1601.00893*, January 2016.
- S. Patwardhan, S. Banerjee and T. Pedersen. **UMND1: Unsupervised Word Sense Disambiguation Using Contextual Semantic Relatedness.** In *SemEval-2007: Proceedings of the 4th International Workshop on Semantic Evaluations*, pages 390-393, Prague, Czech Republic, June 2007.
- T. Wilson, P. Hoffmann, S. Somasundaran, J. Kessler, J. Wiebe, Y. Choi, C. Cardie, E. Riloff, and S. Patwardhan. **OpinionFinder: A System for Subjectivity Analysis.** In *Proceedings of HLT/EMNLP 2005 Interactive Demonstrations*, pages 34-35, Vancouver, Canada, October 2005.
- S. Patwardhan, T. Pedersen, and S. Banerjee. **SenseRelate::TargetWord - A Generalized Framework for Word Sense Disambiguation.** In *Proceedings of the Twentieth National Conference on Artificial Intelligence (Intelligent Systems Demonstrations)*, pages 1692-1693, Pittsburgh, PA, July 2005.
- S. Patwardhan, T. Pedersen, and S. Banerjee. **SenseRelate::TargetWord - A Generalized Framework for Word Sense Disambiguation.** In *Proceedings of the ACL Interactive Poster and Demonstration Sessions*, pages 73-76, Ann Arbor, MI, June 2005.
- C. Thompson, S. Patwardhan, and C. Arnold. **Generative Models for Semantic Role Labeling.** In *Proceedings of SENSEVAL-3: Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text*, pages 235-238, Barcelona, Spain, July 2004.
- T. Pedersen, S. Patwardhan, and J. Michelizzi. **WordNet::Similarity - Measuring the Relatedness of Concepts.** In *Proceedings of the Nineteenth National Conference on Artificial Intelligence (Intelligent Systems Demonstrations)*, pages 1024-1025, San Jose, CA, July 2004.
- T. Pedersen, S. Patwardhan, and J. Michelizzi. **WordNet::Similarity - Measuring the Relatedness of Concepts.** In *Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics Demonstrations*, pages 38-41, Boston, MA, May 2004.

## TECHNICAL REPORTS AND THESES

- A. Lally, S. Bagchi, M. Barborak, D. Buchanan, J. Chu-Carroll, D. Ferrucci, M. Glass, A. Kalyanpur, E. Mueller, J. W. Murdock, S. Patwardhan, J. Prager and C. Welty. **WatsonPaths: Scenario-based Question Answering and Inference over Unstructured Information.** Research Report RC25489 (WAT1409-048), IBM T. J. Watson Research Center, September 2014.
- S. Patwardhan. **Widening the Field of View of Information Extraction through Sentential Event Recognition.** PhD thesis, University of Utah, September 2009.
- T. Pedersen, S. Pakhomov, and S. Patwardhan. **Measures of Semantic Similarity and Relatedness in the Medical Domain.** Research Report DTC 2005/12, University of Minnesota Digital Technology Center, May 2005.
- T. Pedersen, S. Banerjee, and S. Patwardhan. **Maximizing Semantic Relatedness to Perform Word Sense Disambiguation.** Research Report UMSI 2005/25, University of Minnesota Supercomputing Institute, March 2005.
- S. Patwardhan. **Incorporating Dictionary and Corpus Information into a Context Vector Measure of Semantic Relatedness.** Master's thesis, University of Minnesota, Duluth, August 2003.

---

## INVITED PRESENTATIONS

### **Semantic Technologies for Automatic Question Answering**

Visiting Professor, ECI 2016 Winter School  
*University of Buenos Aires, Argentina*

**Jul 2016**

### **Cognitive Computing: Question Answering Technologies Behind Watson**

Guest Instructor, IBM Online Course: <http://ibm.com/watsonacademy>  
*IBM, Armonk NY*

**Mar 2015**

<b>Case Study on IBM Watson</b> Invited Guest Commentator, MBA Course: "Technology and Operations Management" <i>Harvard Business School, Boston MA</i>	<b>Nov 2015</b>
<b>UIMA in IBM Watson</b> Guest Lecturer, Computer Science Course: "Semantic Technologies in IBM Watson" <i>Columbia University, New York NY</i>	<b>Oct 2015</b>
<b>Building Applications with Watson Experience Manager</b> Guest Lecturer, Computer Science Course: "Natural Language Processing with Watson" <i>Rensselaer Polytechnic Institute, Troy NY</i>	<b>Oct 2014</b>
<b>Building IBM's Watson: Technical Insights into the Jeopardy-winning DeepQA Technology</b> Invited Speaker, First Workshop on Cognitive Systems and Watson <i>National University of Singapore, Singapore</i>	<b>Sept 2014</b>
<b>Jeopardy! "Special" Questions: Exploring the Boundaries of Watson</b> Guest Lecturer, Computer Science Course: "Natural Language Processing with Watson" <i>Rensselaer Polytechnic Institute, Troy NY</i>	<b>March 2014</b>
<b>Watson Beyond Jeopardy: Adaptation to the Medical Domain</b> Invited Speaker, Computer Science Colloquium <i>Rensselaer Polytechnic Institute, Troy NY</i>	<b>March 2014</b>
<b>UIMA in IBM Watson</b> Guest Lecturer, Computer Science Course: "Semantic Technologies in IBM Watson" <i>Columbia University, New York NY</i>	<b>March 2014</b>
<b>UIMA in IBM Watson</b> Guest Lecturer, Computer Science Course: "Semantic Technologies in IBM Watson" <i>Columbia University, New York NY</i>	<b>March 2013</b>
<b>Building Watson: An Overview of the DeepQA Project for the Jeopardy! Challenge</b> Invited Speaker, Computer Science Colloquium <i>University of Utah, Salt Lake City UT</i>	<b>Feb 2012</b>
<b>Building Watson: An Overview of the DeepQA Project for the Jeopardy! Challenge</b> Invited Speaker, External Speaker Series <i>Knight Capital Group, Jersey City NJ</i>	<b>Oct 2011</b>

---

## TEACHING

Visiting Professor, <b>Semantic Technologies for Automatic Question Answering</b> <i>ECI 2016 Winter School, University of Buenos Aires</i>	<b>Jul 2016</b>
Guest Instructor, <b>Cognitive Computing: Question Answering Technologies Behind Watson</b> <i>IBM Online Course: <a href="http://ibm.com/watsonacademy">http://ibm.com/watsonacademy</a></i>	<b>Mar 2015</b>
Teaching Assistant, <b>Machine Learning</b> <i>School of Computing, University of Utah</i>	<b>Fall 2003</b>
Teaching Assistant, <b>Operating Systems Practicum</b> <i>Department of Computer Science, University of Minnesota Duluth</i>	<b>Spring 2003</b>
Teaching Assistant, <b>Introduction to Natural Language Processing</b> <i>Department of Computer Science, University of Minnesota Duluth</i>	<b>Fall 2002</b>
Teaching Assistant, <b>Computer Science II</b> <i>Department of Computer Science, University of Minnesota Duluth</i>	<b>Spring 2002</b>
Teaching Assistant, <b>Computer Science I</b> <i>Department of Computer Science, University of Minnesota Duluth</i>	<b>Fall 2001</b>

---