

Siddharth Patwardhan

771 University Village, Salt Lake City, UT 84108.
(801) 403-9850

sidd@cs.utah.edu

<http://www.cs.utah.edu/~sidd>

EDUCATION

University of Utah, Salt Lake City, UT

PhD. in Computer Science

June 2008

GPA: 3.9/4.0

University of Minnesota Duluth, Duluth, MN

M.S. in Computer Science

Aug 2003

Thesis: "Incorporating Dictionary and Corpus Information into a Context Vector Measure of Semantic Relatedness"

GPA: 4.0/4.0

Maharashtra Institute of Technology, University of Pune, India

B.E. in Computer Science

May 2001

Final Year Project: "Generic Interface for Designing Embedded Systems"

Result: First Class

SELECTED PUBLICATIONS

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.

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.

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.

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.

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.

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.

S. Patwardhan. **Incorporating Dictionary and Corpus Information into a Context Vector Measure of Semantic Relatedness.** Master's thesis, University of Minnesota, Duluth, August 2003.

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.

RELEVANT EXPERIENCE

Research Assistant

Aug 2003 – present

School of Computing, University of Utah, Salt Lake City, UT

Assist Dr. Ellen Riloff on various projects in Opinion Analysis and in Information Extraction.

- Built a system (funded by LLNL) that improves Information Extraction coverage using rules learned from the Web.
- Created an opinion analysis and classification system as part of the AQUAINT project (funded by ARDA).
- Worked with Dr. Cindi Thompson to build a generative-model-based semantic role labeling system.

Summer Research Intern**May 2003 – Aug 2003***Division of Biomedical Informatics, Mayo Clinic, Rochester, MN*

Created a collection of Perl modules to measure the semantic similarity of medical concepts, based on the SNOMED-CT semantic hierarchy and statistics extracted from clinical texts. This work was done under the supervision of Dr. Serguei Pakhomov. Journal article to appear in JBI and published in U.Minn. Digital Technology Center Report DTC 2005/12.

Research Assistant**May 2002 – Aug 2002***Transportation Data Research Laboratory, Duluth, MN*

Worked with Minnesota Department of Transportation and Traffic Management Center to carry out research on analysis and processing of transportation data. Designed archive structure and wrote software to retrieve and archive data from data sensors. The principle focus of TDRL is to develop a large-scale centralized transportation data center that serves as a transportation information resource for MN/DOT, University researchers, and other government agencies.

Teaching Assistant**Sept 2001 – May 2003***Department of Computer Science, University of Minnesota Duluth, Duluth, MN*

Designed and prepared course material for undergraduate courses. Graded exams and assignments. Conducted labs and lectures for the courses (Computer Science 1, Natural Language Processing, Operating Systems Practicum).

PROJECTS

- Learning IE Patterns from the Web:** This system improves the coverage of an existing Information Extraction system by inferring new relevant extraction patterns from the Web (coded in **C++**, presented at a workshop in the ACL '06 conference). **2006**
- Subsumption Hierarchy for Opinion Analysis:** A hierarchy used for selecting the best combination of features for a text-classification system that detects opinions in text. (coded in **C++**, presented as a poster at the EMNLP '06 conference). **2006**
- Opinion Source Identification:** A system for identifying opinion-holders in free text, uses lexico-syntactic features in a machine learning system. (coded in **C++**, part of the OpinionFinder system demonstrated at the EMNLP '05 conference). **2005**
- Generative Model for Role Labeling:** A machine-learning system, trained using FrameNet, for automatically assigning semantic roles to words in text (coded in **Java**, participated in SENSEVAL-3). **2004**
- WordNet::SenseRelate:** A system that determines the intended meaning of a word from its context using the notion of Semantic Relatedness of concepts. Was demonstrated at ACL '05 and AAAI '05 conferences (coded in **Perl**, available for download at <http://senserelate.sourceforge.net>). **2004**
- Semantic Similarity of Biomedical Concepts:** A collection of Perl modules that adapt existing Computational Linguistics research for measuring semantic similarity of words and concepts to the biomedical domain (coded in **Perl**, an article to appear in the *Journal of Biomedical Informatics*). **2003**
- WordNet::Similarity:** A group of modules for measuring the semantic similarity or relatedness of English words and concepts. Was under the supervision of Dr. Ted Pedersen. Demonstrated at NAACL '04 and AAAI '04 conferences (coded in **Perl**, available for download at <http://wn-similarity.sourceforge.net>). **2003**
- Generic Interface for Embedded Systems:** A Visual-Basic-like programming language for building graphical interfaces for communicating with embedded systems (coded in **Java**, team of four students, undergraduate degree final-year project). **2001**

AWARDS

- Graduate Tuition Benefit.** Awarded by the Graduate School, University of Utah, Salt Lake City, UT. **2003 – 2007**
- Most Outstanding Teaching Assistant.** Awarded by the Graduate School, University of Minnesota Duluth, Duluth, MN. **2003**
- Graduate School Tuition Benefit.** Awarded by the Department of Computer Science, University of Minnesota Duluth, Duluth, MN. **2001 – 2003**
- National Talent Search Scholarship.** Awarded by the National Council for Educational Research and Training, New Delhi, India. **1995 – 2001**