

Sujith Ravi - Curriculum Vitae

CONTACT INFORMATION	1650 Nantucket Circle, Apt 325 Santa Clara, CA 95054, USA	<i>Tel:</i> (213) 453-8684 <i>E-mail:</i> sujith_ravi@yahoo.com http://www.sravi.org
RESEARCH INTERESTS	Large-scale inference for Natural Language Processing, Unsupervised/Semi-supervised Machine Learning, Information Retrieval, User Modeling, Computational Advertising, Computational Decipherment, Artificial Intelligence, Statistics.	
EDUCATION	Ph.D., Computer Science University of Southern California, Los Angeles, California Thesis: <i>Deciphering Natural Language</i> . Advisor: Kevin Knight	August 2006 – May 2011
	M.S., Computer Science University of Southern California, Los Angeles, California	August 2004 – May 2006
	B.Tech., Computer Science National Institute of Technology, Trichy, Tamil Nadu, India	July 2000 – May 2004
RESEARCH EXPERIENCE	Research Scientist Yahoo! Research Work on large-scale unsupervised and semi-supervised machine learning, structured prediction for natural language processing, information extraction, user modeling in social media, graph optimization algorithms for summarizing noisy data, computational advertising.	July 2011 – present Santa Clara, CA, USA
	Graduate Research Assistant , Advisor: Kevin Knight University of Southern California/Information Sciences Institute Work on unsupervised machine learning algorithms for structured prediction tasks (sequence labeling, word alignment, transliteration, machine translation), computational decipherment, exact and approximate inference for model minimization.	August 2007 – May 2011 Marina del Rey, CA, USA
	Research Intern , Mentor: Evgeniy Gabrilovich Yahoo! Research Work on statistical machine translation approach to computational advertising.	May 2009 – August 2009 Santa Clara, CA, USA
	Research Intern , Mentor: Marius Paşca Google Research Work on large-scale attribute and fact extraction from structured text on the Web.	May 2007 – August 2007 Mountain View, CA, USA
	Directed Research , Advisor: Jihie Kim University of Southern California/Information Sciences Institute Work on information extraction from online student discussion forums for pedagogical applications.	August 2006 – May 2007 Marina del Rey, CA, USA
	Directed Research , Advisor: Stacy Marsella University of Southern California/Information Sciences Institute Work on multi-agent framework to simulate and model social environments.	May 2005 – May 2006 Marina del Rey, CA, USA
	Directed Research , Advisor: Milind Tambe	January 2005 – May 2005

University of Southern California/Information Sciences Institute Marina del Rey, CA, USA
Work on policy randomization techniques for multi-agents in adversarial domains.

PATENT

- User Modeling for Conversation Settings in Social Media (co-inventor) – US Patent being filed.
- Automatic Generation of Bid Phrases for Online Advertising (co-inventor) – US Patent pending.

NEWS
ARTICLES/MEDIA

- Cracking the code of machine translation. *The New Scientist Magazine*. Volume 210, Issue 2817, 18 June 2011, Page 23.
- Cracking the code of machine translation. *ACM TechNews*. 24 June 2011.
- Article/interview on machine translation. *The Beijing News*. 26 June 2011.

WORKSHOP
ORGANIZATION

- **Co-organizer:** ACL 2012 workshop on Multilingual Modeling together with Jagadeesh Jagarlamudi (University of Maryland, USA), Hal Daume III (University of Maryland, USA) and Xiaojun Wan (Peking University, China)

PUBLICATIONS

JOURNAL ARTICLES

- **Sujith Ravi** and Kevin Knight. Does GIZA++ make search errors? *Computational Linguistics*, Volume 36, number 3, pages 295-302, 2010.
- **Sujith Ravi** and Kevin Knight. Attacking letter substitution ciphers with integer programming. *Cryptologia*, Volume 33, number 4, pages 321-334, 2009.

BOOK CHAPTER

- Jihie Kim, Erin Shaw, and **Sujith Ravi**. Mining student discussions to profile participation and scaffold learning. In Cristobal Romero, Sebastian Ventura, Mykola Pechenizkiy, and Ryan Baker (Eds.), *The Handbook of Educational Data Mining* (pages 299-310). CRC Press, 2010.

REFEREED CONFERENCE PAPERS

- 2012: 4 papers under submission to various NLP/machine learning conferences (titles withheld for anonymity).
- **Sujith Ravi** and Kevin Knight. Deciphering foreign language. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT)*, 2011.
- **Sujith Ravi** and Kevin Knight. Deciphering foreign language. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT)*, 2011.
- **Sujith Ravi** and Kevin Knight. Bayesian inference for Zodiac and other homophonic ciphers. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics:*

Human Language Technologies (ACL-HLT), 2011.

- **Sujith Ravi**, Ashish Vaswani, Kevin Knight, and David Chiang. Fast, greedy model minimization for unsupervised tagging. In *Proceedings of the 23rd International Conference on Computational Linguistics (COLING)*, pages 940-948, 2010.
- **Sujith Ravi**, Jason Baldridge, and Kevin Knight. Minimized models and grammar-informed initialization for supertagging with highly ambiguous lexicons. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 495-503, 2010.
- David Chiang, Jonathan Graehl, Kevin Knight, Adam Pauls, and **Sujith Ravi**. Bayesian inference for finite-state transducers. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL/HLT)*, pages 447-455, 2010.
- **Sujith Ravi**, Andrei Z. Broder, Evgeniy Gabrilovich, Vanja Josifovski, Sandeep Pandey, and Bo Pang. Automatic generation of bid phrases for online advertising. In *Proceedings of the International Conference on Web Search and Data Mining (WSDM)*, pages 341-350, 2010.
- **Sujith Ravi** and Kevin Knight. Minimized models for unsupervised part-of-speech tagging. 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)*, pages 504-512, 2009. **Nominated for the Best Paper Award.**
- **Sujith Ravi** and Kevin Knight. Learning phoneme mappings for transliteration without parallel data. In *Proceedings of Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL/HLT)*, pages 37-45, 2009.
- **Sujith Ravi** and Kevin Knight. Probabilistic methods for a Japanese syllable cipher. In *Proceedings of the 22nd International Conference on the Computer Processing of Oriental Languages (ICCPOL)*, pages 270-281, 2009.
- **Sujith Ravi** and Kevin Knight. Attacking decipherment problems optimally with low-order n-gram models. In *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 812-819, 2008.
- **Sujith Ravi**, Kevin Knight, and Radu Soricut. Automatic prediction of parser accuracy. In *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 887-896, 2008.
- **Sujith Ravi** and Marius Pasca. Using structured text for large-scale attribute extraction. In *Proceedings of the 17th ACM Conference on Information and Knowledge Management (CIKM)*, pages 1183-1192, 2008.
- Jihie Kim, Erin Shaw, **Sujith Ravi**, Erin Tavano, Aniwat Arromratana, and Pankaj Sarda. Scaffolding on-line discussions with past discussions: An analysis and pilot study of Pedabot. In *Proceedings of the 9th International Conference on Intelligent Tutoring Systems (ITS)*, pages 343-352, 2008.
- **Sujith Ravi** and Jihie Kim. Profiling student interactions in threaded discussions with speech act classifiers. In *Proceedings of the 13th International Conference on Artificial Intelligence in Education (AIED)*, pages 357-364, 2007.

REFEREED WORKSHOP PAPERS

- Tugba Bodrumlu, Kevin Knight, and **Sujith Ravi**. A new objective function for word alignment. In *Proceedings of the NAACL/HLT Workshop on Integer Programming for Natural Language Processing*, pages 28-35, 2009.
- **Sujith Ravi**, Jihie Kim, and Erin Shaw. Mining on-line discussions: Assessing technical quality for student scaffolding and classifying messages for participation profiling. In *Proceedings of the Educational Data Mining Workshop in the 13th International Conference on Artificial Intelligence in Education (AIED)*, pages 70-79, 2007.

- HONORS & AWARDS
- **Phi Kappa Phi Member** Spring 2010 – *present*
University of Southern California
 - **Outstanding Research Assistant Award** Fall 2008
Computer Science Department, University of Southern California
 - **Best Paper Award** Fall 2008
Annual Information Sciences Institute (ISI) Graduate Research Symposium
 - **Best Presentation Award (Runner-up)** Fall 2008
Annual Information Sciences Institute (ISI) Graduate Research Symposium
 - **Outstanding Teaching Assistant Award** Spring 2007
Computer Science Department, University of Southern California
 - **Outstanding Academic Achievement Award (Graduate Student)** Spring 2006
Awarded to International students by the University of Southern California Spring 2011

- SELECTED INVITED TALKS
- **Microsoft Research**. Redmond, WA. May 2011
Host: Kristina Toutanova.
Talk Title: *Deciphering Natural Language*.
 - **SDL Language Weaver**. Los Angeles, CA. May 2011
Host: Daniel Marcu.
Talk Title: *Deciphering Natural Language*.
 - **IBM T. J. Watson Research**. Yorktown Heights, NY. April 2011
Host: Yaser Al-Onaizan.
Talk Title: *Deciphering Natural Language*.
 - **Google Research**. Mountain View, CA. April 2011
Host: Marius Paşca.
Talk Title: *Deciphering Natural Language*.
 - **Yahoo! Research**. Santa Clara, CA. April 2011
Host: Evgeniy Gabrilovich.
Talk Title: *Deciphering Natural Language*.
 - **AT&T Labs Research**. Florham Park, NJ. April 2011
Host: Dan Melamed.

Talk Title: *Deciphering Natural Language*.

- **University of Maryland.** College Park, MD. March 2011
Host: Philip Resnik.
Talk Title: *Deciphering Natural Language*.
- **Massachusetts Institute of Technology.** Cambridge, MA. May 2010
Host: Prof. Regina Barzilay, MIT Computer Science & Artificial Intelligence Lab.
Talk Title: *Deciphering Natural Language*.
- **University of Southern California.** Los Angeles, CA. March 2010
Host: Computer Science Department
Talk Title: *Minimized Models for Natural Language Processing*.
Selected as one of the speakers at the Annual Research Review session (a one-day event showcasing current research at USC; attended by faculty/students from USC and other universities, alumni and people from the industry).
- **University of Washington.** Seattle, WA. July 2008
Host: Prof. Mari Ostendorf, Department of Electrical Engineering
Talk Title: *Attacking Decipherment Problems Optimally with Low-Order N-gram Models*.

TEACHING
EXPERIENCE

Graduate Teaching Assistant, University of Southern California Spring 2007
CS 571: Web Technologies, a popular and highly attended graduate-level course offered by the Computer Science Department. Helped design and evaluate assignments and exams, provided support for course instructors.

Graduate Teaching Assistant, University of Southern California Fall 2006
CS 573: Advanced Artificial Intelligence, an advanced graduate-level course and a core-requirement for Ph.D. students working in the area of artificial intelligence. Delivered lectures, helped design projects, assignments and exams.

Class Instructor, National Institute of Technology Fall 2003
Conducted classes on “Programming in C language” for first year undergraduate students.

PROFESSIONAL
ACTIVITIES

- **Senior Program Committee Member:** IJCAI (2011)
- **Program Committee Member:** ACL 2012, EMNLP 2012, NAACL 2012, EACL 2012, IJCNLP (2011), IEKA - Workshop on Information Extraction and Knowledge Acquisition (RANLP 2011), BUCC - Workshop on Building and Using Comparable Corpora (LREC 2012, ACL 2011, LREC 2010), NAACL/HLT Student Research Workshop (2012, 2010)
- **Journal/Conference Reviewer:** ACL 2012, EMNLP 2012, NAACL 2012, EACL 2012, Journal on Pattern Recognition (2011), IJCNLP (2011), AAAI (2011), ACL (2011), IJCAI (2011), ACM TIST - Transactions on Intelligent Systems and Technology Journal (2011), AMTA (2010), NAACL (2010), LREC (2010), ISWC (2008), IUI (2008)
- **Organizer:** USC/ISI Natural Language Seminar (2008 – 2009)
- **Co-Organizer:** 3rd annual ISI Graduate Research Symposium (2008)
- **Volunteer:** Southern California Linux Expo (SCALE 2005)
- **Organizing Committee Member:** VORTEX, a national-level technical symposium conducted

by the Department of Computer Science and Engineering, National Institute of Technology,
Trichy, India (2002 – 2004)

PROGRAMMING SKILLS Python, C++, Java, Perl, MapReduce programming with Hadoop.

LANGUAGES Fluent in English, Hindi, Malayalam. Minimal reading/writing skills in Arabic.

REFERENCES Available upon request.