

Thamar Solorio

May 2007

Personal Information

Name: Thamar Ivette Solorio
Date of Birth: August 3th, 1976
Affiliation: University of Texas at El Paso
Work Address: 500 W. University Drive
El Paso, Texas, 79968-0521
Telephone: 915.747.7941
Fax: 915.747.5030
Homepage: <http://www.cs.utep.edu/tsolorio/>
E-mail Address: tsolorio@utep.edu

Education

Ph.D. in Computer Science

September 2005

Computer Science Department, Instituto Nacional de Astrofísica, Óptica y Electrónica, Sta. María Tonantzintla, Puebla, México

Dissertation title: “*Improvement of Named Entity Tagging by Machine Learning*”

Advisor: Prof. Aurelio López López

M.S. in Computer Science

August 2002

Computer Science Department, Instituto Nacional de Astrofísica, Óptica y Electrónica, Sta. María Tonantzintla, Puebla, México

Thesis title: “*Using Unlabeled Data to Improve Classifier Accuracy*”

Advisor: Prof. Olac Fuentes

B.S. in Computer Systems Engineering

July 2000

Facultad de Ingeniería, Universidad Autónoma de Chihuahua, Chihuahua, México

Positions Held

January 2006-January 2007: Part-time Post-Doc, University of Texas at El Paso

September 2005-present: Lecturer, Department of Computer Science, University of Texas at El Paso

Research Interests

Natural Language Processing and Machine Learning

More specifically I'm interested in developing statistical methods for the automated processing of interlanguages, information extraction, and applications of NLP techniques to bioinformatics.

Participation in Research Projects

Beyond Words: Identification of Back-channel Communication Rules in Arabic and Development of Training Methods

Funded by DARPA

Participation: Research Assistant

Principal Investigator: Nigel Ward

Duration: 2005 – 2007

Formulation and Evaluation of Models for Question Answering from Spanish Texts using Linguistic Information at Different Levels

Funded by the National Council for Science and Technology (CONACYT) – Grant C02-43990

Participation: Research Assistant

Principal Investigator: Manuel Montes y Gómez
Duration: 2004 – 2007

Search, Extraction, Retrieval and Analysis of Information from Texts in Spanish

Funded by the National Council for Science and Technology (CONACYT) – Grant U39957-Y
Participation: Research Assistant
Principal Investigator: Aurelio López López
Duration: 2003 – 2006

Machine Learning for Astronomical Data Analysis

Funded by the National Council for Science and Technology (CONACYT) - Grant J31877A
Participation: Research Assistant
Principal Investigator: Olac Fuentes
Duration: 2000 – 2003

Program Committees

- NAACL-HLT 2007, The Annual Conference of the North American Chapter of the Association for Computational Linguistics.
- CICLing-2007, Eighth International Conference on Intelligent Text Processing and Computational Linguistics.
- CICLing-2006, Seventh International Conference on Intelligent Text Processing and Computational Linguistics.
- ACL Demos and Interactive Posters, 43rd Annual Meeting of the Association for Computational Linguistics, 2005.

Scholarships and Awards

- 2007: NSF Travel Grant, CRA-W Career Mentoring Workshop for Women in Computer Science and FCRC
- 2005: NSF Travel Grant, Association of Computational Linguistics (ACL) 2005 Student Research Workshop
- 2002-2005: Scholarship for PhD studies from the National Council for Science and Technology (CONACYT)
- 2000-2002: Scholarship for MS studies from the National Council for Science and Technology (CONACYT)

Other Research Related Activities

Principal Investigator on NSF Grant Submitted

Language Models for Discriminating Code-Switching Patterns of Bilingual Children with SLI, submitted December 2006.

Together with Olac Fuentes we co-direct the Machine Learning Research seminars. We meet weekly with graduate undergraduate and students to discuss work in the field of Machine Learning. One goal of the seminar is to motivate students to do research, and pursue advance degrees in computer science.

Teaching Experience

- Elementary Data Structures and Algorithms. UTEP Fall 2005, Spring 2006, Fall 2006, Spring 2007.
- Computer Programming for Science & Engineering. UTEP Fall 2006, Spring 2007.
- Artificial Intelligence. UTEP Fall 2006.
- Software Engineering I. UTEP Fall 2005, Spring 2006.
- Computer Architecture I. UTEP Fall 2005.
- Selected Topics on Artificial Intelligence. Instituto Tecnológico de Apizaco, graduate level, 2003.
- Natural Language Processing. Instituto Tecnológico de Apizaco, graduate level, 2003.
- Object Oriented Programming in Java. Universidad Popular Autónoma del Estado de Puebla 2002.
- Teaching Assistant, Automata Theory and Formal Languages. Instituto Nacional de Astrofísica, Óptica y Electrónica, graduate level, assistant for professor Aurelio López, 2004.
- Teaching Assistant, Machine Learning. Instituto Nacional de Astrofísica, Óptica y Electrónica, graduate level, assistant for professor Olac Fuentes, 2004.

- Teaching Assistant, Research Seminar II. Instituto Nacional de Astrofísica, Óptica y Electrónica, graduate level, assistant for professor Olac Fuentes, 2003.
- Teaching Assistant, Automata Theory and Formal Languages. Instituto Nacional de Astrofísica, Óptica y Electrónica, graduate level, assistant for professor Ariel Carrasco, 2001.

Student Advising

M.S. Students

1. María Petra Paredes, Sumarización Automática de Documentos, M.S in Computer Science, Instituto Tecnológico de Apizaco, 2004.

Undergraduate Students

1. David Mireles, Genetic Algorithms for the prediction of secondary structure of RNA.
2. Juan Carlos Franco, Part-of-Speech tagging of code-switched discourse in English-Spanish.
3. David Vera, co-supervised with Olac Fuentes, Machine learning for the prediction of turn-ends.

Publications

- [1] Nigel G. Ward, Rafael Escalante, Yaffa Al Bayyari, and Thamar Solorio. Learning to show you're listening. *Computer Assisted Language Learning*, submitted.
- [2] Olac Fuentes, David Vera, and Thamar Solorio. A filter-based approach to detect end-of-utterances from prosody in dialog systems. In *Proceedings of the The Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT 2007)*, Rochester, NY, April 2007 (to appear).
- [3] Juan C. Franco and Thamar Solorio. Baby steps towards a language model for Spanglish. In Alexander Gelbukh, editor, *The Eighth International Conference on Intelligent Text Processing and Computational Linguistics CICLing-2007*, Lecture Notes in Computer Science LNCS 4394, pages 75–84, Mexico City, Mexico, February 2007. Springer.
- [4] Thamar Solorio, Olac Fuentes, Nigel Ward, and Yaffa Al Bayyari. Prosodic feature generation for back-channel prediction. In *The Ninth International Conference on Spoken Language Processing, INTERSPEECH 2006*, Pittsburgh, Pennsylvania, September 2006.
- [5] Ted Pedersen, Anagha Kulkarni, Roxana Angheluta, Zornista Kozareva, and Thamar Solorio. Improving name discrimination : A language salad approach. In *Proceedings of the EACL 2006 Workshop on Cross-Language Knowledge Induction*, April 2006.
- [6] Ted Pedersen, Anagha Kulkarni, Roxana Angheluta, Zornista Kozareva, and Thamar Solorio. An unsupervised language independent method of name discrimination using second order co-occurrence features. In Alexander Gelbukh, editor, *Computational Linguistics and Intelligent Text Processing: Seventh International Conference, CICLing 2006*, Lecture Notes in Computer Science. Springer, February 2006.
- [7] Thamar Solorio. Exploiting named entity taggers in a second language. In Chris Callison-Burch and Stephen Wan, editors, *Student Research Workshop at the 43rd Annual Meeting of the Association for Computational Linguistics*, pages 25–30, Ann Arbor, Michigan, June 2005. The Association for Computational Linguistics.
- [8] Thamar Solorio, Olac Fuentes, Roberto Terlevich, and Elena Terlevich. An active instance-based machine learning method for stellar population studies. *Monthly Notices of the Royal Astronomical Society*, 363(2), October 2005.
- [9] Thamar Solorio and Aurelio López López. Learning named entity recognition in Portuguese from Spanish. In Alexander Gelbukh, editor, *Computational Linguistics and Intelligent Text Processing: Sixth International Conference, CICLing 2005*, volume 3406 of *Lecture Notes in Computer Science*. Springer, February 2005.
- [10] Thamar Solorio, Manuel Pérez Coutiño, Manuel Montes y Gómez, Luis Villaseñor Pineda, and Aurelio López López. Question classification in Spanish and Portuguese. In Alexander Gelbukh, editor, *Computational Linguistics and Intelligent Text Processing: Sixth International Conference, CICLing 2005*, volume 3406 of *Lecture Notes in Computer Science*. Springer, February 2005.
- [11] Thamar Solorio and Aurelio López López. Adapting a named entity recognition system for Spanish to Portuguese. In Guillermo de Ita, Olac Fuentes, and Mauricio Osorio, editors, *IX Iberoamerican Workshops on Artificial Intelligence: Workshop on Herramientas y Recursos Lingüísticos para el Español y el Portugués*, pages 292–297, Puebla, Mexico, November 2004.

- [12] Olac Fuentes, Thamar Solorio, Roberto Terlevich, and Elena Terlevich. Analysis of galactic spectra using active instance-based learning and domain knowledge. In Christian Lemaître, Carlos Reyes, and Jesús A. González, editors, *Advances in Artificial Intelligence – IBERAMIA 2004*, Lecture Notes in Artificial Intelligence 3315, pages 215–224, Puebla, Mexico, November 2004. Springer.
- [13] Manuel Pérez Coutiño, Thamar Solorio, Manuel Montes y Gómez, Aurelio López López, and Luis Villaseñor Pineda. Question answering for Spanish based on lexical and context annotation. In Christian Lemaître, Carlos Reyes, and Jesús A. González, editors, *Advances in Artificial Intelligence – IBERAMIA 2004*, Lecture Notes in Artificial Intelligence 3315, pages 325–333, Puebla, Mexico, November 2004. Springer.
- [14] Manuel Pérez-Coutiño, Thamar Solorio, Manuel Montes y Gómez, Aurelio López López, and Luis Villaseñor Pineda. The use of lexical context in question answering for Spanish. In Carol Peters and Francesca Borri, editors, *Working Notes for the Cross Language Evaluation Forum Workshop, (CLEF-2004)*, Bath, England, September 2004. ISTI-CNR.
- [15] Elena Terlevich, Roberto Terlevich, Juan P. Torres Papaqui, Trilce Estrada Piedra, Olac Fuentes, Thamar Solorio, and Sandro Bressan. Computer science approach to the stellar fabric of violent starforming regions in AGN. In Th. Storchi Bergmann, L. C. Ho, and H. R. Schmitt, editors, *IAU Symposium, the interplay among black holes, stars and ISM in Galactic Nuclei*, number 222. International Astronomical Union, 2004.
- [16] Thamar Solorio, Manuel Pérez Coutiño, Manuel Montes y Gómez, Luis Villaseñor Pineda, and Aurelio López López. A language independent method for question classification. In *The 20th International Conference on Computational Linguistics, COLING-04*, volume II, pages 1374–1380, Geneva, Switzerland, August 2004.
- [17] Manuel Pérez Coutiño, Thamar Solorio, Manuel Montes y Gómez, Aurelio López López, and Luis Villaseñor Pineda. Toward a document model for question answering systems. In Jesus Favela, Ernestina Menasalvas, and Edgar Chávez, editors, *Advances in Web Intelligence: Second International Atlantic Web Intelligent Conference AWIC 2004*, Lecture Notes in Artificial Intelligence 3034, pages 145–154, Cancun, Mexico, May 2004. Springer.
- [18] Olac Fuentes and Thamar Solorio. An optimization algorithm based on active and instance-based learning. In R. Monroy, G. Arroyo-Figueroa, L. E. Sucar, and H. Sossa, editors, *MICAI 2004: Advances in Artificial Intelligence, Third Mexican International Conference on Artificial Intelligence*, Lecture Notes in Artificial Intelligence 2972, pages 242–251, Mexico City, Mexico, April 2004. Springer.
- [19] Thamar Solorio and Aurelio López López. Learning named entity classifiers using support vector machines. In Alexander Gelbukh, editor, *Computational Linguistics and Intelligent Text Processing: Fifth International Conference, CICLing 2004*, volume 2945/2004 of *Lecture Notes in Computer Science 2945*, pages 158–167. Springer, February 2004.
- [20] Thamar Solorio, Olac Fuentes, Roberto Terlevich, Elena Terlevich, and Sandro Bressan. Automated determination of stellar population parameters in galaxies using active instance-based learning. In François Oshsenbein, Mark G. Allen, and Daniel Egret, editors, *Astronomical Data Analysis Software and Systems XIII (ADASS XIII)*, volume 314 of *Astronomical Society of the Pacific Conference Series*, pages 609–612, Strasbourg, France, October 2003. A.S.P.
- [21] Olac Fuentes and Thamar Solorio. Principal component analysis and active instance-based machine learning for interferogram analysis. In *Proc. of the IASTED International Conference on Artificial Intelligence and Applications (AIA2003)*, Benalmádena, Spain, September 2003.
- [22] Thamar Solorio. Using unlabeled data to improve classifier accuracy. M.S. thesis, National Institute of Astrophysics, Optics, and Electronics (INAOE), Puebla, Mexico, August 2002.
- [23] Thamar Solorio and Olac Fuentes. Improving classification accuracy of large test sets using the ordered classification algorithm. In F.J. Garijo, J.C. Riquelme, and M. Toro (Eds.), editors, *Advances in Artificial Intelligence - IBERAMIA 2002: 8th Ibero-American Conference on AI*, Lecture Notes in Computer Science 2527, pages 70–79, Seville, Spain, November 2002. Springer.
- [24] Thamar Solorio and Olac Fuentes. Taking advantage of unlabeled data with the ordered classification algorithm. In *Proceedings of IASTED International Conference on Artificial Intelligence and Soft Computing*, Banff, Canada, July 2002.
- [25] Thamar Solorio and Olac Fuentes. Using unlabeled data to improve the automated prediction of stellar atmospheric parameters. In David A. Bohlender, Daniel Durand, and T. H. Handley, editors, *Astronomical Data Analysis Software and Systems XI*, pages 405–408, Victoria, B.C., Canada, October 2001. Astronomical Society of the Pacific Conference Series Volume 281.

- [26] Tamar Solorio and Olac Fuentes. Improving classifier accuracy using unlabeled data. In M.H. Hamza, editor, *Proc. of the IASTED International Conference on Artificial Intelligence and Applications*, pages 409–412, Marbella, Spain, September 2001.

References

Ann Gates, Professor and Chairman
Department of Computer Science
University of Texas at El Paso
El Paso, TX, 79968-0521
Tel. 915-747-6952
agates@utep.edu

Aurelio López, Associate Professor and Chairman
Computer Science Department
Instituto Nacional de Astrofísica, Óptica y Electrónica
Tonantzintla, Puebla, México 72840
Tel. (52-222)266-3100 X 8314
allopez@inaoep.mx

Rada Mihalcea, Assistant Professor
Department of Computer Science & Engineering
University of North Texas
Denton, Texas 76203
Tel. 940-369-7630
rada@cs.unt.edu

Nigel Ward, Associate Professor
Department of Computer Science
University of Texas at El Paso
El Paso, TX, 79968-0521
Tel. 915-747-6827
nigel@utep.edu

Ted Pedersen, Associate Professor
Department of Computer Science
University of Minnesota, Duluth
Duluth MN 55812-2496
Tel. 218-726-8770
tpederse@d.umn.edu

Roberto J. Terlevich, Professor
Institute of Astronomy, University of Cambridge
Cambridge CB3 0FD United Kingdom

rjt@ast.cam.ac.uk

Instituto Nacional de Astrofísica, Óptica y Electrónica
Tonantzintla, Puebla, México 72840
Tel. +52 22 22 663100 X 2307
rjt@inaoep.mx

Manuel Montes-y-Gómez, Associate Professor
Computer Science Department
Instituto Nacional de Astrofísica, Óptica y Electrónica
Tonantzintla, Puebla, México 72840
Tel. +52(222)-266-3100 X 8218
mmontesg@inaoep.mx

Bess Fjordbak, Assistant Professor
Speech-Language Pathology
University of Texas at El Paso
El Paso, TX, 79902
Tel. 915-747-7278
bsfjordbak@utep.edu