
Thamar Solorio

Personal Information

Current Job Title: Assistant Professor
Affiliation: University of Alabama at Birmingham
Work Address: Department of Computer and Information Sciences
Birmingham, AL 35209
Homepage: <http://www.cis.uab.edu/solorio>
E-mail Address: solorio@uab.edu

Education

Ph.D. in Computer Science September 2005

Computer Science Department, Instituto Nacional de Astrofísica, Óptica y Electrónica, 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, 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

Professional Experience

Assistant Professor June 2009 –

Department of Computer and Information Sciences, University of Alabama at Birmingham

Research Associate September 2007–June 2009

Department of Computer Science, University of Texas at Dallas

Lecturer September 2005–August 2007

Department of Computer Science, University of Texas at El Paso

Research Interests

Natural Language Processing and Machine Learning: Information extraction, information retrieval, processing of bilingual discourse, automated language assessment using spontaneous language samples, computer forensics, authorship analysis

Research Grants

Human-Centered Computing, IIS, NSF September 2010–August 2013

\$21,055, Analysis of Language Impairment in Monolingual and Bilingual Children

Thamar Solorio (PI, UAB)

Computing Research Infrastructure, NSF February 2010–February 2012

\$21,992, Collaborative Research:CI-P: Creation of an annotated repository of multilingual and multigenre code switched data for several language pairs

Investigators: Thamar Solorio (PI, UAB)

Robust Intelligence and OISE, NSF

January 2010–July 2010

\$17,972, Young Investigators in the Americas Workshop
 Investigators: Thamar Solorio (PI, UAB)

Human-Centered Computing, IIS, NSF

September 2008–August 2010

\$110,493, Prediction of Language Status in Bilingual Children (PoLSi-BC)
 Thamar Solorio (PI, UAB), Yang Liu (co-PI, UTD)

National Center for Research Resources, NIH

September 2008–August 2009

\$36,000, Prediction of Language Status in Monolingual Children (PoLSi-MC)
 Thamar Solorio (PI, UTD), Yang Liu (co-PI, UTD)

Fellowships, Awards, and Honors

**Member of the National System of Researchers (SNI),
 CONACyT, Mexico**
 Level: Candidate

Since January 2011

Anthony Barnard Award, CIS, UAB
 Travel Grant

Fall 2009

CRA Academic Careers Workshop, NSF
 Travel Grant

February 2008

CRA-W Career Mentoring Workshop, NSF
 Travel Grant

June 2007

National Council for Science and Technology (CONACyT)
 Fellowship for PhD studies

September 2002- August 2005

National Council for Science and Technology (CONACyT)
 Fellowship for MS studies

August 2000- August 2002

Workshop Co-Chair

NAACL-HLT 2010, Young Investigators Workshop on Computational Approaches to Languages of the Americas

Program Committees

ACL-HLT 2011, The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies

CICLing-2011, 12th International Conference on Intelligent Text Processing and Computational Linguistics

CICLing-2010, 11th International Conference on Intelligent Text Processing and Computational Linguistics

RANLP-2009, Recent Advances in Natural Language Processing.

CICLing-2009, 10th International Conference on Intelligent Text Processing and Computational Linguistics

CICLing-2008, 9th International Conference on Intelligent Text Processing and Computational Linguistics

NAACL-HLT 2007, The Annual Conference of the North American Chapter of the Association for Computational Linguistics

CICLing-2007, 8th International Conference on Intelligent Text Processing and Computational Linguistics

CICLing-2006, 7th International Conference on Intelligent Text Processing and Computational Linguistics

Journal and Conference Reviewer

Transactions on Intelligent Systems and Technology, 2011

Concurrency and omputation: Practice and Experience, 2010

MICAI-2010, 10th Mexican International Conference on Artificial Intelligence

INFORMS Journal on Computing, 2010

MICAI-2009, 8th Mexican International Conference on Artificial Intelligence

Computación y Sistemas, Special issue: “Innovative applications of AI”, 2008

MICAI-2008, 7th Mexican International Conference on Artificial Intelligence

ACL Demos and Interactive Posters, 43rd Annual Meeting of the Association for Computational Linguistics, 2005

Services to Government Agencies

Panel reviewer for the National Science Foundation, 2010, 2009.

Ad-hoc reviewer for the National Science Foundation, 2008.

Publications

Book Chapters

- [1] Thamar Solorio. *HAREM, a primeira avaliação conjunta de sistemas de reconhecimento de entidades mencionadas para português: Documentação e actas do encontro*, chapter 15 MALINCHE: A NER system for Portuguese that reuses knowledge from Spanish. Number ISBN:978-989-20-0731-1. 2007.

Refereed Journal Articles

- [1] Thamar Solorio, Melissa Sherman, Yang Liu, Lisa Bedore, Elizabeth Peña, and Aquiles Iglesias. Analyzing language samples of Spanish-English bilingual children for the automated prediction of language dominance. *Natural Language Engineering*, Available on CJO.
- [2] M. Taufer, M-Y. Leung, T. Solorio, A. Licon, D. Mireles, D. Gomez-Leon, R. Araiza, and K.K. Johnson. RNAVLab: A unified environment for computational RNA structure analysis based on grid computing technology. *Parallel Computing*, 34:661–680, 2008.
- [3] Nigel G. Ward, Rafael Escalante, Yaffa Al Bayyari, and Thamar Solorio. Learning to show you’re listening. *Computer Assisted Language Learning*, (20):385–407, 2007.
- [4] 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):543–554, October 2005.

Refereed Conference and Workshop Papers

- [1] H. Jair Escalante, Manuel Montes y Gómez, and Thamar Solorio. Local histograms of character n-grams for authorship attribution. In *Proceedings of the 49th Annual Meeting for the Association for Computational Linguistics: Human Language Technologies*, Portland, Oregon, June 19–24 (to appear). ACL.
- [2] Aaron Blum, Brad Wardman, Thamar Solorio, and Gary Warner. Lexical feature based phishing url detection using online learning. In *Proceedings of the 3rd ACM workshop on Artificial intelligence and security*, AISec '10, pages 54–60, Chicago, Illinois, 2010. ACM.
- [3] Sangita Pillay and Thamar Solorio. Authorship attribution on web forum posts. In *eCrime Researchers Summit*, Dallas, TX, October 2010. APWG.
- [4] Gabriela Ramírez de-la Rosa, Manuel Montes y Gómez, Luis Villaseñor-Pineda, David Pinto-Avenidaño, and Thamar Solorio. Using information from the target language to improve crosslingual text classification. In *Proceedings of IceTAL 2010*, pages 305–313, Reykjavik, Iceland, August 2010.
- [5] Richa Tiwari, Chengcui Zhang, and Thamar Solorio. A supervised machine learning approach of extracting coexpression relationship among genes from literature. In *Proceedings of IEEE IRI 2010*, pages 98–103, Las Vegas, NV, 2010.
- [6] Keyur Gabani, Melissa Sherman, Thamar Solorio, Yang Liu, Lisa Bedore, and Elizabeth Peña. A corpus-based approach for the prediction of language impairment in monolingual English and Spanish-English bilingual children. In *North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL HLT) 2009*, pages 46–55, Boulder, Colorado, June 2009. Association for Computational Linguistics.
- [7] Thamar Solorio and Yang Liu. Part-of-speech tagging for English-Spanish code-switched text. In *Empirical Methods on Natural Language Processing, EMNLP-2008*, pages 1051–1060, Honolulu, Hawaii, October 2008. Association for Computational Linguistics.
- [8] Thamar Solorio and Yang Liu. Learning to predict code-switching points. In *Empirical Methods on Natural Language Processing, EMNLP-2008*, pages 973–981, Honolulu, Hawaii, October 2008. Association for Computational Linguistics.
- [9] Thamar Solorio and Yang Liu. Using language models to identify language impairment in Spanish-English bilingual children. In *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing*, pages 116–117, Columbus, Ohio, June 2008. Association for Computational Linguistics.
- [10] Michela Taufer, Thamar Solorio, Abel Licon, David Mireles, and Ming-Ying Leung. On the effectiveness of rebuilding RNA secondary structures from sequence chunks. In *The Seventh IEEE International Workshop on High-Performance Computational Biology, HiCOMB 2008*, Miami, Florida, April 2008.
- [11] 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)*, pages 45–48, Rochester, NY, April 2007. Association for Computational Linguistics.
- [12] 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*, volume 4394 of *Lecture Notes in Computer Science*, pages 75–84, Mexico City, February 2007.
- [13] 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.
- [14] 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*, pages 25–32, April 2006.
- [15] 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*, volume 3878 of *Lecture Notes in Computer Science*, pages 208–222, February 2006.

- [16] 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, ACL-2005*, pages 25–30, Ann Arbor, Michigan, June 2005.
- [17] 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*, pages 762–768, February 2005.
- [18] 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*, pages 612–619, February 2005.
- [19] 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: The IX Ibero-American Conference on Artificial Intelligence*, volume 3315 of *Lecture Notes in Artificial Intelligence 3315*, pages 215–224, Puebla, Mexico, November 2004.
- [20] 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*, volume 3315 of *Lecture Notes in Artificial Intelligence 3315*, pages 325–333, Puebla, Mexico, November 2004.
- [21] 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.
- [22] 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.
- [23] 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*, volume 3034 of *Lecture Notes in Artificial Intelligence 3034*, pages 145–154, Cancun, Mexico, May 2004.
- [24] 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*, volume 2972 of *Lecture Notes in Artificial Intelligence 2972*, pages 242–251, Mexico City, Mexico, April 2004.
- [25] 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, February 2004.
- [26] 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 Artificial Intelligence*, volume 2527 of *Lecture Notes in Computer Science 2527*, pages 70–79, Seville, Spain, November 2002.

Selected Invited Talks

Exploring New Frontiers in Natural Language Processing (NLP): Applications of NLP to Communication Disorders and Bilingual Discourse. Research Group in Computational Linguistics, University of Wolverhampton. May 2009.

Exploring corpus-based approaches for distinguishing children with and without language impairment. Callier Center, The University of Texas at Dallas. November 2008.

The Automated Processing of Bilingual Discourse. Language and Information Technologies, University of North Texas. July 2008.

Processing Code-Switched Text. Department of Computer Sciences, The University of Texas at Austin. May 2008.

Processing Code-Switched Text. Department of of Computer and Information Sciences, University of Delaware. February 2008.

Teaching Experience

Natural Language Processing. UAB Spring 2010, 2011

Artificial Intelligence. UAB Fall 2010, 2009

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 in 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

Professional Affiliations

- Association for Computing Machinery (ACM)
- Association for Computational Linguistics (ACL)

Student Advising and Mentoring

Ph.D. Students

1. Binod Gyawali
2. Sangita Pillay
3. Gabriela Ramirez de la Rosa

M.S. Students

1. Keyur Gabani, UTD
2. Debangana Dey, UAB
3. Melissa Sherman, UTD
4. María Petra Paredes, Sumarización Automática de Documentos, M.S in Computer Science, Instituto Tecnológico de Apizaco, 2004.

Undergraduate Students

1. Kenny G. Blackmon, UAB
2. Anthony Bowman, UAB
3. Juan Carlos Franco, UTEP
4. Wesley Johnson, UAB
5. Brenda Medina, UTEP, (CRA-W DMP Student Awardee 2008)

6. David Mireles, UTEP, (Microsoft Scholar, Google Scholar, DSSI-2007 Fellow)
7. Maria Patterson, UAB
8. Cameron Stanley, UAB
9. David Vera, UTEP,
10. Nicolle Whitman, UTEP, (LSMAP Scholar)