NAACL HLT 2010

Young Investigators Workshop on Computational Approaches to Languages of the Americas

Proceedings of the Workshop

June 6, 2010
Los Angeles, California

USB memory sticks produced by
Omnipress Inc.
2600 Anderson Street
Madison, WI 53707
USA

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

Introduction

Welcome to the First Young Investigators Workshop on Computational Approaches to Languages of the Americas. This workshop will be held on June 6, 2010 in Los Angeles, immediately following the 11th Annual Conference of the North American Chapter of the Association for Computational Linguistics, NAACL-HLT 2010. The goal of this workshop is to bring together researchers from all of the Americas developing human language technologies that are interested in establishing international collaborations. We believe a more interactive community within the Americas can contribute to the advancement of the field, not only with respect to the improvement of performance on specific areas of NLP but more important, with respect to motivating the growth of its community by providing a conducive collaboration infrastructure that facilitates the active involvement of researchers in the field.

We are very excited about the response to the call for papers. We received a total of 21 submissions from 8 countries. The final program brings together researchers from Argentina, Brazil, Colombia, Costa Rica, Mexico, Uruguay and the USA. The contributions in the proceedings are of three types: research papers, project overviews and opinion papers. The research papers include recent advances in topics from opinion mining, to textual entailment, to adaptation of NLP approaches to software engineering. The survey papers present an overview of larger research projects by a single university or research group. These overviews present interesting efforts in dialogue systems, text simplification, language generation, and corpus based approaches to verb subcategorization and relation extraction. The proceedings also include two opinion papers that describe the research situation of the NLP communities in Costa Rica and Brazil. All contributions describe how international collaborations can push research forward by either listing the resources and/or experience sought or what specific resources and experience can be contributed. In sum, these proceedings provide a broad coverage of research on computational linguistics south of the Rio Bravo addressing three different languages: Spanish, Brazilian Portuguese and English.

We would like to thank the program committee members for their support in spreading the call for papers and providing a conscientious and timely review. Without their support this workshop would have not been as successful. We would also like to thank the NAACL-HLT 2010 Workshop Chairs, David Traum and Richard Sproat for all their help and great overseeing of the logistics of this workshop. Lastly, we were able to offer travel support and full conference registration waivers due to the very generous support of the NAACL Executive Board, and the Information and Intelligent Systems Directorate and the Office of International Science and Engineering of the National Science Foundation (USA) award number 1008711.

As part of the one day workshop, the program will also include a panel discussion to brainstorm on ways to promote a more interactive community on this side of the globe, and the possibility of having more workshops of this kind. A summary from this panel will be available on the workshop website soon after the event: (http://groups.google.com/group/naacl-2010-yi-workshop).

We are are looking forward to a great event and hope that initiatives like these will eventually lead to a stronger and tighter computational linguistics research community on the Americas.

Thamar Solorio and Ted Pedersen

Organizers:

Thamar Solorio, University of Alabama at Birmingham, USA
Ted Pedersen, University of Minnesota–Duluth, USA

Program Committee:

Laura Alonso Alemani, Universidad Nacional de Córdoba, Argentina
John Atkinson, Universidad de Concepción, Chile
Diego Burgos, Instituto Tecnológico Metropolitano, Colombia
Vitor Carvalho, Microsoft Bing, USA
Maria das Graças Volpe Nunes, Universidade de São Paulo, Brazil
Ana Feldman, Montclair State University, USA
Caroline Gasperin, Universidade de São Paulo, Brazil
Alexander Gelbukh, CIC, IPN, Mexico
Carlos Gómez Gallo, Harvard, USA
Agustin Gravano, Universidad de Buenos Aires, Argentina
Diana Inpken, University of Ottawa, Canada
Greg Kondrak, University of Alberta, Canada
Jorge Antonio Leoni de León, Universidad de Costa Rica, Costa Rica
Aurelio López López, INAOE, Mexico
Lucia Helena Machado Rino, Universidade Federal de São Carlos, Brazil
Rada Mihalcea, University of North Texas, USA
Raymond Mooney, University of Texas at Austin, USA
Manuel Montes y Gómez, INAOE, Mexico
Thiago A. S. Pardo, Universidade de São Paulo, Brazil
Renata Vieira, Pontifícia Universidade Católica do Rio Grande do Sul, Brazil
Luis Villaseñor-Pineda, INAOE, Mexico
Dina Wonsever, Universidad de la Republica, Uruguay

Table of Contents

Workshop Program

Sunday, June 6, 2010

Session 1

8:45–9:00 Opening Remarks

9:00–9:30 *Computational Linguistics in Brazil: An Overview*
Thiago Pardo, Caroline Gasperin, Helena de Medeiros Caseli and Maria das Graças Nunes

9:30–10:00 *Data-driven computational linguistics at FaMAF-UNC, Argentina*
Laura Alonso Alemany and Gabriel Infante-Lopez

10:00–10:30 *Variable-Length Markov Models and Ambiguous Words in Portuguese*
Fabio Natanael Kepler and Marcelo Finger

10:30–11:00 **Break**

Session 2

11:00–11:30 *Using Common Sense to generate culturally contextualized Machine Translation*
Helena de Medeiros Caseli, Bruno Akio Sugiyama and Junia Coutinho Anacleto

11:30–12:30 **Poster Session**

Human Language Technology for Text-based Analysis of Psychotherapy Sessions in the Spanish Language
Horacio Saggion, Elena Stein-Sparvieri, David Maldavsky and Sandra Szasz

Computational Linguistics in Costa Rica: an overview
Jorge Antonio Leoni de León

Fostering Digital Inclusion and Accessibility: The PorSimples project for Simplification of Portuguese Texts
Sandra Aluisio and Caroline Gasperin

Opinion Identification in Spanish Texts
Aiala Rosá, Dina Wonsever and Jean-Luc Minel

Sunday, June 6, 2010 (continued)

A Machine Learning Approach for Recognizing Textual Entailment in Spanish
Julio Castillo

The emergence of the modern concept of introspection: a quantitative linguistic analysis
Iván Raskovsky, Diego Fernández Slezak, Carlos Diuk and Guillermo A. Cecchi

Combining CBIR and NLP for Multilingual Terminology Alignment and Cross-Language Image Indexing
Diego Burgos

IRASubcat, a highly parametrizable, language independent tool for the acquisition of verbal subcategorization information from corpus
Ivana Romina Altamirano and Laura Alonso Alemany

The TermiNet Project: an Overview
Ariani Di Felippo

Automated Detection of Language Issues Affecting Accuracy, Ambiguity and Verifiability in Software Requirements Written in Natural Language
Allan Berrocal Rojas and Elena Gabriela Barrantes Sliesarieva

Recognition and extraction of definitional contexts in Spanish for sketching a lexical network
Cesar Aguilar, Olga Acosta and Gerardo Sierra

12:30–2:00 **Lunch**

Session 3

2:00–2:30 *Computational Linguistics for helping Requirements Elicitation: a dream about Automated Software Development*
Carlos Mario Zapata Jaramillo

2:30–3:00 *Text Generation for Brazilian Portuguese: the Surface Realization Task*
Eder Novais, Thiago Tadeu and Ivandre Paraboni

3:00–3:30 **Break**

Sunday, June 6, 2010 (continued)

Session 4

3:30–4:00 *Dialogue Systems for Virtual Environments*
Luciana Benotti, Paula Estrella and Carlos Areces

Panel Session

4:00–5:00 Challenges and Opportunities for Conducting Research and Forming Collaborations in the Americas

5:00–5:30 Concluding Discussion