Lecture Notes in Artificial Intelligence 6001

Edited by R. Goebel, J. Siekmann, and W. Wahlster

Subseries of Lecture Notes in Computer Science

Thiago Alexandre Salgueiro Pardo
António Branco Aldebaro Klautau
Renata Vieira Vera Lúcia Strube de Lima (Eds.)

Computational Processing of the Portuguese Language

9th International Conference, PROPOR 2010
Porto Alegre, RS, Brazil, April 27-30, 2010
Proceedings

Springer

Volume Editors

Thiago Alexandre Salgueiro Pardo
Universidade de São Paulo, São Carlos / SP, Brazil
E-mail: taspardo@icmc.usp.br

António Branco
Universidade de Lisboa, Portugal
E-mail: antonio.branco@di.fc.ul.pt

Aldebaro Klautau
Universidade Federal do Pará, Belém / PA, Brazil
E-mail: aldebaro@ufpa.br

Renata Vieira
PUCRS, Porto Alegre / RS, Brazil
E-mail: renata.vieira@pucrs.br

Vera Lúcia Strube de Lima
PUCRS, Porto Alegre / RS, Brazil
E-mail: vera.strube@pucrs.br

Library of Congress Control Number: 2010923533

CR Subject Classification (1998): I.2, H.3, H.4, I.4, I.5, H.2.8

LNCS Sublibrary: SL 7 – Artificial Intelligence

ISSN 0302-9743
ISBN-10 3-642-12319-8 Springer Berlin Heidelberg New York
ISBN-13 978-3-642-12319-1 Springer Berlin Heidelberg New York

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper 06/3180

Preface

The International Conference on Computational Processing of Portuguese—PROPOR—is the main event in the area of natural language processing that is focused on Portuguese and the theoretical and technological issues related to this language. It welcomes contributions for both written and spoken language processing.

The event is hosted in Brazil and in Portugal. The meetings have been held in Lisbon/Portugal (1993), Curitiba/Brazil (1996), Porto Alegre/Brazil (1998), Évora/Portugal (1999), Atibaia/Brazil (2000), Faro/Portugal (2003), Itatiaia/Brazil (2006) and Aveiro/Portugal (2008).

This meeting has been a highly productive forum for the progress of this area and to foster the cooperation among the researchers working on the automated processing of the Portuguese language. PROPOR brings together research groups, promoting the development of methodologies, resources and projects that can be shared among all researchers and practitioners in the field.

The ninth edition of this event was held in Porto Alegre, Brazil, at *Pontifícia Universidade Católica do Rio Grande do Sul* (PUCRS). It had two main tracks: one for language processing and another one for speech processing. This event hosted a special Demonstration Session and the first edition of the PhD and MSc Dissertation Contest, which aimed at recognizing the best academic work on processing of the Portuguese language in the last few years. This edition of the event featured tutorials on statistical machine translation and on speech recognition, as well as invited talks by renowned researchers of natural language processing.

A total of 48 submissions were received, 37 for the language track and 11 for the speech track, by authors from 10 countries: Brazil, China, Denmark, UK, Germany, Italy, Poland, Portugal, Spain and USA. Each submission was evaluated by at least three members from a multidisciplinary and international scientific committee.

This volume gathers a selection of the 21 best papers accepted to be presented at this meeting, of which 13 are full papers, corresponding to an acceptance rate of 27%. These papers cover the areas concerning applications for information handling and text processing, language processing, language resources, and speech recognition and synthesis.

We would like to express our thanks to everyone involved in the organization of the event, to the scientific committee members for their excellent work, to the researchers who kindly accepted to contribute to the event by delivering tutorials and invited talks, and to the institutions, organizations and funding agencies which allowed the realization of this event, namely, PUCRS, SBC (the Brazilian Computer Society), CEPLN (the SBC Special Interest Group on Natural Language Processing), CNPq (*Conselho Nacional de Desenvolvimento Científico e Tecnológico* – a Brazilian funding agency), NAACL (The North American Chapter of the Association for Computational Linguistics), ISCA (International Speech Communication Association), SIG-IL (the ISCA Special Interest Group on Iberian Languages) and CLARIN

(Common Language Resources and Technology Infrastructure – a large-scale pan-European collaborative effort to create, coordinate and make language resources and technology available and readily usable).

April 2010

Thiago Alexandre Salgueiro Pardo
António Branco
Aldebaro Klautau
Renata Vieira
Vera Lúcia Strube de Lima

Organization

General Chair

Vera Lúcia Strube de Lima	Pontifícia Universidade Católica do Rio Grande do Sul, Brazil

Program Chairs

António Branco	Universidade de Lisboa, Portugal - Language
Aldebaro Klautau	Universidade Federal do Pará, Brazil - Speech

Tutorial Chair

Maria das Graças Volpe Nunes	Universidade de São Paulo, Brazil

Editorial Chair

Thiago Alexandre Salgueiro Pardo	Universidade de São Paulo, Brazil

PhD and MSc Dissertation Contest Chair

David de Matos	Instituto de Engenharia de Sistemas e Computadores, Portugal

Demo Session Chair

António Teixeira	Universidade de Aveiro, Portugal

Local Organizing Chair

Renata Vieira	Pontifícia Universidade Católica do Rio Grande do Sul, Brazil

Program Committee

Alexandre Agustini	Pontifícia Universidade Católica do Rio Grande do Sul, Brazil
Aline Villavicencio	Universidade Federal do Rio Grande do Sul, Brazil
Amália Mendes	Universidade de Lisboa, Portugal
Andre Adami	Universidade de Caxias do Sul, Brazil
António Ribeiro	European Commission, Joint Research Centre, Italy
António Serralheiro	Instituto de Engenharia de Sistemas e Computadores / Instituto Superior Técnico, Portugal
António Teixeira	Universidade de Aveiro, Portugal
Antonio Bonafonte	Universitat Politècnia de Catalunya, Spain
Augusto Silva	Universidade Católica Portuguesa, Portugal
Bento Dias da Silva	Universidade Estadual Paulista, Brazil
Berthold Crysmann	Universität Bonn / Universität des Saarlandes, Germany
Carlos Prolo	Pontifícia Universidade Católica do Rio Grande do Sul, Brazil
Caroline Hagège	Xerox Research Centre, France
Celso Kaestner	Universidade Tecnológica Federal do Paraná, Brazil
Climent Nadeu	Universitat Politècnia de Catalunya, Spain
Cristina Martins	Universidade de Coimbra, Portugal
Daniela Braga	Microsoft Language Development Center, Portugal
Dante Barone	Universidade Federal do Rio Grande do Sul, Brazil
David de Matos	Instituto de Engenharia de Sistemas e Computadores, Portugal
Diamantino Freitas	Universidade do Porto, Portugal
Edmilson Morais	Vocalize, Brazil
Eric Laporte	Université Paris-Est Marne-la-Vallée, France
Fábio Violaro	Universidade Estadual de Campinas, Brazil
Fernando Resende	Universidade Federal do Rio de Janeiro, Brazil
Florence Amardeilh	Mondeca, France
Gabriel Pereira Lopes	Universidade Nova de Lisboa, Portugal
Gael Harry Dias	Universidade da Beira Interior, Portugal
Horacio Saggion	University of Sheffield, UK
Irene Rodrigues	Universidade de Évora, Portugal
Isabel Trancoso	Instituto de Engenharia de Sistemas e Computadores / Instituto Superior Técnico, Portugal
Ivandré Paraboni	Universidade de São Paulo, Brazil
Ivandro Sanches	Centro Universitário da FEI, Brazil
Jason Baldridge	University of Texas, USA
Jean-Luc Minel	Université Paris Ouest - Nanterre La Défense, France
João Balsa	Universidade de Lisboa, Portugal
João Paulo Neto	Instituto de Engenharia de Sistemas e Computadores, Portugal
Jorge Baptista	Universidade do Algarve, Portugal
Julia Hirschberg	Columbia University, USA

Lúcia Rino	Universidade Federal de São Carlos, Brazil
Luísa Coheur	Instituto de Engenharia de Sistemas e Computadores, Portugal
Luiz Pizzato	University of Sydney, Australia
Marcelo Finger	Universidade de São Paulo, Brazil
Marco Gonzalez	Pontifícia Universidade Católica do Rio Grande do Sul, Brazil
Maria Viana	Universidade de Lisboa, Portugal
Maria das Graças Volpe Nunes	Universidade de São Paulo, Brazil
Maria Francisca Xavier	Universidade Nova de Lisboa, Portugal
Maximiliano Saiz Noeda	Universidad de Alicante, Spain
Miguel Filgueiras	Universidade do Porto, Portugal
Nestor Yoma	Universidad de Chile, Chile
Nuno Mamede	Instituto de Engenharia de Sistemas e Computadores / Instituto Superior Técnico, Portugal
Nuno Cavalheiro Marques	Universidade Nova de Lisboa, Portugal
Pablo Gamallo	Universidad de Santiago de Compostela, Spain
Paulo Quaresma	Universidade de Évora, Portugal
Plínio Barbosa	Universidade Estadual de Campinas, Brazil
Ranniery Maia	Toshiba Research Europe Limited, UK
Renata Vieira	Pontifícia Universidade Católica do Rio Grande do Sul, Brazil
Sandra Aluisio	Universidade de São Paulo, Brazil
Stanley Loh	Universidade Católica de Pelotas, Brazil
Steven Bird	University of Melbourne, Australia
Thiago Alexandre Salgueiro Pardo	Universidade de São Paulo, Brazil
Tracy King	Microsoft, USA
Violeta Quental	Pontifícia Universidade Católica do Rio de Janeiro, Brazil
Viviane Orengo	Universidade Federal do Rio Grande do Sul, Brazil

Additional Reviewers

Aldebaro Klautau	Universidade Federal do Pará, Brazil
Belinda Maia	Universidade do Porto, Portugal
Fábio Kepler	Universidade de São Paulo, Brazil
Guillaume Jacquet	Xerox Research Centre, France
Jussara Vieira	Vocalize, Brazil
Marta Costa-jussà	Barcelona Media, Spain
Milagros Fernández Gavilanes	Universidade da Coruña, Spain
Simon Zwarts	Macquarie University, Australia

Steering Committee

Maria das Graças Volpe Nunes	Universidade de São Paulo, Brazil
Vera Lucia Strube de Lima	Pontifícia Universidade Católica do Rio Grande do Sul, Brazil
Isabel Trancoso	Instituto de Engenharia de Sistemas e Computadores / Instituto Superior Técnico, Portugal
Violeta Quental	Pontifícia Universidade Católica do Rio de Janeiro, Brazil
António Teixeira	Universidade de Aveiro, Portugal

Table of Contents

Applications: Information Handling

Applications: Text Processing

Language Processing

Language Resources

Speech Recognition

Speech Synthesis