

Lecture Notes in Artificial Intelligence 3230

Edited by J. G. Carbonell and J. Siekmann

Subseries of Lecture Notes in Computer Science

José Luis Vicedo
Patricio Martínez-Barco Rafael Muñoz
Maximiliano Saiz Noeda (Eds.)

Advances in Natural Language Processing

4th International Conference, EsTAL 2004
Alicante, Spain, October 20-22, 2004
Proceedings

Series Editors

Jaime G. Carbonell, Carnegie Mellon University, Pittsburgh, PA, USA
Jörg Siekmann, University of Saarland, Saarbrücken, Germany

Volume Editors

José Luis Vicedo
Patricio Martínez-Barco
Rafael Muñoz
Maximiliano Saiz Noeda
Universidad de Alicante
Departamento de Lenguajes y Sistemas Informáticos
Carretera de San Vicente del Raspeig
03690 San Vicente del Raspeig, Alicante, Spain
E-mail: { vicedo;patricio;rafael;max }@dlsi.ua.es

Library of Congress Control Number: 2004113295

CR Subject Classification (1998): I.2.7, F.4.2-3, I.2, H.3, I.7

ISSN 0302-9743

ISBN 3-540-23498-5 Springer Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

Springer is a part of Springer Science+Business Media

springeronline.com

© Springer-Verlag Berlin Heidelberg 2004
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 11334347 06/3142 5 4 3 2 1 0

Preface

EsTAL – España for Natural Language Processing – continued on from the three previous conferences: FracTAL, held at the Université de Franch-Comté, Besançon (France) in December 1997, VexTAL, held at Venice International University, Cá Foscari (Italy), in November 1999, and PorTAL, held at the Universidade do Algarve, Faro (Portugal), in June 2002. The main goals of these conferences have been: (i) to bring together the international NLP community; (ii) to strengthen the position of local NLP research in the international NLP community; and (iii) to provide a forum for discussion of new research and applications.

EsTAL contributed to achieving these goals and increasing the already high international standing of these conferences, largely due to its Program Committee, composed of renowned researchers in the field of natural language processing and its applications. This clearly contributed to the significant number of papers submitted (72) by researchers from (18) different countries.

The scope of the conference was structured around the following main topics: (i) *computational linguistics research* (spoken and written language analysis and generation; pragmatics, discourse, semantics, syntax and morphology; lexical resources; word sense disambiguation; linguistic, mathematical, and psychological models of language; knowledge acquisition and representation; corpus-based and statistical language modelling; machine translation and translation aids; computational lexicography), and (ii) *monolingual and multilingual intelligent language processing and applications* (information retrieval, extraction and question answering; automatic summarization; document categorization; natural language interfaces; dialogue systems and evaluation of systems).

Each paper was revised by three reviewers from the Program Committee or by external referees designed by them. All those who contributed are mentioned on the following pages. The review process led to the selection of 42 papers for presentation. They have been published in this volume.

We would like to express here our thanks to all the reviewers for their quick and excellent work. We extend these thanks to our invited speakers, Walter Daelemans and Rada Mihalcea for their valuable contribution, which undoubtedly increased the interest in the conference. We are also indebted to a number of individuals for taking care of specific parts of the conference program. Specially, to Miguel Angel Varó who built and maintained all Web services for the conference.

Finally, we want to thank the University of Alicante, and, specially, the Office of the Vice-President for Extracurricular Activities (*Vicerrectorado de Extensión Universitaria*) and the Department of Software and Computing Systems (*Departamento de Lenguajes y Sistemas Informáticos*) because of their support of this conference.

Organization

Program and Conference Chairs

José L. Vicedo (University of Alicante, Spain)

Patricio Martinez-Barco (University of Alicante, Spain)

Organization Chairs

Rafael Muñoz (University of Alicante, Spain)

Maximiliano Saiz (University of Alicante, Spain)

Program Committee

Alfonso Ureña (University of Jaen, Spain)

Bernardo Magnini (ITC-irst, Italy)

Dan Moldovan (University of Texas at Dallas, USA)

Elisabete Ranchhod (University of Lisbon, Portugal)

German Rigau (University of the Basque Country, Spain)

Hans Uszkoreit (Saarland University at Saarbrücken, Germany)

Henri Zingle (University of Nice, France)

Horacio Rodríguez (Technical University of Catalonia, Spain)

Horacio Saggon (University of Sheffield, UK)

Igor Melcuk (University of Montreal, Canada)

Julio Gonzalo (Universidad Nacional de Educación a Distancia (UNED), Spain)

Lidia Moreno (Polytechnical University of Valencia, Spain)

Manuel de Buenaga (European University of Madrid, Spain)

Manuel Palomar (University of Alicante, Spain)

Massimo Poesio (University of Essex, UK)

Nuno Mamede (INESC-ID Lisbon, Portugal)

Peter Greenfield (Centre Lucien Tesnière, Univ. of Franche-Comté, France)

Pierre-André Buvet (University of Paris 13, France)

Rodolfo Delmonte (University of Venice, Italy)

Ruslan Mitkov (University of Wolverhampton, UK)

Sanda Harabagiu (University of Texas at Dallas, USA)

Stephane Chaudiron (Ministry of Technology and Research, France)

Sylviane Cardey (Centre Lucien Tesnière, Univ. of Franche-Comté, France)

Victor Díaz (University of Seville, Spain)

Werner Winiwarter (University of Vienna, Austria)

Referees

Alfonso Ureña
Alicia Ageno
Ana García Serrano
Anabela Barreiro
Antonio Ferrandez
Armando Suarez
Bernardo Magnini
Carlo Strapparava
Damián López
Dan Moldovan
Diamantino Caseiro
Elisabete Ranchhod
Fernando Llopis
Finley Lacatusu
Georgiana Puscasu
German Rigau
Hans Uszkoreit
Henri Zingle
Horacio Rodríguez
Horacio Saggion
Igor Melcuk
José L. Vicedo
Juan A. Pérez-Ortiz
Julio Gonzalo
Laura Hasler
Le An Ha
Lidia Moreno
Luis Oliveira
Luisa Coheur

Manuel de Buenaga
Manuel García Vega
Manuel Palomar
María Teresa Martín
Massimo Poesio
Maximiliano Saiz Noeda
Miguel Angel García
Mijail Alexandrov
Mikel Forcada
Nuno Mamede
Olivia Sanchez
Patricio Martinez-Barco
Paul Morarescu
Paula Carvalho
Peter Greenfield
Pierre-André Buvet
Rafael Muñoz
Ramón López
Ricardo Ribeiro
Richard Evans
Rodolfo Delmonte
Ruslan Mitkov
Sanda Harabagiu
Stephane Chaudiron
Sylviane Cardey
Victor Díaz
Viktor Pekar
Werner Winiwarter
Zornitsa Kozareva

Table of Contents

Adaptive Selection of Base Classifiers in One-Against-All Learning for Large Multi-labeled Collections <i>Arturo Montejo Ráez, Luís Alfonso Ureña López, Ralf Steinberger</i>	1
Automatic Acquisition of Transfer Rules from Translation Examples <i>Werner Winiwarter</i>	13
Automatic Assessment of Open Ended Questions with a BLEU-Inspired Algorithm and Shallow NLP <i>Enrique Alfonseca, Diana Pérez</i>	25
Automatic Phonetic Alignment and Its Confidence Measures <i>Sérgio Paulo, Luís C. Oliveira</i>	36
Automatic Spelling Correction in Galician <i>M. Vilares, J. Otero, F.M. Barcala, E. Domínguez</i>	45
Baseline Methods for Automatic Disambiguation of Abbreviations in Jewish Law Documents <i>Yaakov HaCohen-Kerner, Ariel Kass, Ariel Peretz</i>	58
Bayes Decision Rules and Confidence Measures for Statistical Machine Translation <i>Nicola Ueffing, Hermann Ney</i>	70
Character Identification in Children Stories <i>Nuno Mamede, Pedro Chaleira</i>	82
Comparison and Evaluation of Two Approaches of a Multilayered QA System Applied to Temporality <i>E. Saquete, R. Muñoz, P. Martínez-Barco, J.L. Vicedo</i>	91
The Contents and Structure of the Context Base, and Its Application <i>Yusuke Takahashi, Ichiro Kobayashi, Michiaki Iwazume, Noriko Ito, Michio Sugeno</i>	103
Developing a Minimalist Parser for Free Word Order Languages with Discontinuous Constituency <i>Asad B. Sayeed, Stan Szpakowicz</i>	115
Developing Competitive HMM PoS Taggers Using Small Training Corpora <i>Muntsa Padró, Lluís Padró</i>	127

Exploring the Use of Target-Language Information to Train the Part-of-Speech Tagger of Machine Translation Systems <i>Felipe Sánchez-Martínez, Juan Antonio Pérez-Ortiz, Mikel L. Forcada</i>	137
Expressive Power and Consistency Properties of State-of-the-Art Natural Language Parsers <i>Gabriel Infante-Lopez, Maarten de Rijke</i>	149
An Independent Domain Dialogue System Through a Service Manager <i>Márcio Mourão, Renato Cassaca, Nuno Mamede</i>	161
Information Retrieval in Digital Theses Based on Natural Language Processing Tools <i>Rocío Abascal, Béatrice Rumpler, Jean-Marie Pinon</i>	172
Integrating Conceptual Density with WordNet Domains and CALD Glosses for Noun Sense Disambiguation <i>Davide Buscaldi, Paolo Rosso, Francesco Masulli</i>	183
Intertwining Deep Syntactic Processing and Named Entity Detection <i>Caroline Brun, Caroline Hagège</i>	195
Language Understanding Using n-multigram Models <i>Lluís Hurtado, Encarna Segarra, Fernando García, Emilio Sanchis</i>	207
Multi-label Text Classification Using Multinomial Models <i>David Vilar, María José Castro, Emilio Sanchis</i>	220
A Multi-use Incremental Syntax-Semantic Interface <i>Luísa Coheur, Nuno Mamede, Gabriel G. Bès</i>	231
Multiword Expression Translation Using Generative Dependency Grammar <i>Stefan Diaconescu</i>	243
Named Entity Recognition Through Corpus Transformation and System Combination <i>José A. Troyano, Vicente Carrillo, Fernando Enríquez, Francisco J. Galán</i>	255
One Size Fits All? A Simple Technique to Perform Several NLP Tasks <i>Daniel Gayo-Avello, Darío Álvarez-Gutiérrez, José Gayo-Avello</i>	267
Ontology-Based Feature Transformations: A Data-Driven Approach <i>Filip Ginter, Sampo Pyysalo, Jorma Boberg, Jouni Järvinen, Tapio Salakoski</i>	280
On the Quality of Lexical Resources for Word Sense Disambiguation <i>Lluís Màrquez, Mariona Taulé, Lluís Padró, Luis Villarejo, Maria Antònia Martí</i>	292

Reuse of Free Online MT Engines to Develop a Meta-system of Multilingual Machine Translation <i>Vo Trung Hung</i>	303
Semantic-Aided Anaphora Resolution in Large Corpora Development <i>Maximiliano Saiz-Noeda, Borja Navarro, Rubén Izquierdo</i>	314
SemRol: Recognition of Semantic Roles <i>P. Moreda, M. Palomar, A. Suárez</i>	328
Significance of Syntactic Features for Word Sense Disambiguation <i>Sasi Kanth Ala, Narayana Murthy Kavi</i>	340
SisHiTra : A Hybrid Machine Translation System from Spanish to Catalan <i>José R. Navarro, Jorge González, David Picó, Francisco Casacuberta, Joan M. de Val, Ferran Fabregat, Ferran Pla, Jesús Tomás</i>	349
Smoothing and Word Sense Disambiguation <i>Eneko Agirre, David Martínez</i>	360
Spelling Correction for Search Engine Queries <i>Bruno Martins, Mário J. Silva</i>	372
A Statistical Study of the WPT-03 Corpus <i>Bruno Martins, Mário J. Silva</i>	384
A Study of Chunk-Based and Keyword-Based Approaches for Generating Headlines <i>Enrique Alfonseca, José María Guirao, Antonio Moreno-Sandoval</i>	395
Suffixal and Prefixal Morpholexical Relationships of the Spanish <i>Octavio Santana, José Pérez, Francisco Carreras, Gustavo Rodríguez</i> .	407
SuPor: An Environment for AS of Texts in Brazilian Portuguese <i>Lucia Helena Machado Rino, Marcelo Módolo</i>	419
Systemic Analysis Applied to Problem Solving: The Case of the Past Participle in French <i>Séverine Vienne, Sylviane Cardey, Peter Greenfield</i>	431
The Merging Problem in Distributed Information Retrieval and the 2-Step RSV Merging Algorithm <i>Fernando Martínez-Santiago, Miguel Angel García Cumberras, L. Alfonso Ureña López</i>	442
Unsupervised Training of a Finite-State Sliding-Window Part-of-Speech Tagger <i>Enrique Sánchez-Villamil, Mikel L. Forcada, Rafael C. Carrasco</i>	454

Using Seed Words to Learn to Categorize Chinese Text	
<i>Zhu Jingbo, Chen Wenliang, Yao Tianshun</i>	464
On Word Frequency Information and Negative Evidence in Naive Bayes	
Text Classification	
<i>Karl-Michael Schneider</i>	474
Author Index	487