

# **STUDENT RESEARCH WORKSHOP**

*held in conjunction with  
The International Conference RANLP - 2009*

# **PROCEEDINGS**

Edited by

Irina Temnikova, Ivelina Nikolova  
and Natalia Konstantinova

Borovets, Bulgaria

14-15 September 2009

**Student Research  
Workshop**

**PROCEEDINGS**

Borovets, Bulgaria  
14-15 September 2009

ISBN 978-954-452-014-4

Designed and Printed by INCOMA Ltd.  
Shoumen, Bulgaria

## ORGANISING COMMITTEE

**The Student Research Workshop associated with RANLP-09 is organised by:**

**Irina Temnikova**, Research Group in Computational Linguistics  
Research Institute in Information and Language Processing (RIILP)  
University of Wolverhampton, United Kingdom

**Ivelina Nikolova**, Linguistic Modelling Department, Institute for Parallel Processing,  
Bulgarian Academy of Sciences, Bulgaria

**Natalia Konstantinova**, Research Group in Computational Linguistics  
Research Institute in Information and Language Processing (RIILP)  
University of Wolverhampton, United Kingdom

## PROGRAMME COMMITTEE

**Itziar Aldabe** (University of the Basque Country, Spain)  
**Galia Angelova** (Bulgarian Academy of Sciences, Bulgaria )  
**Chris Biemann** (Powerset / Microsoft, San Francisco, USA)  
**Iustin Dornescu** (University of Wolverhampton, UK)  
**Le An Ha** (University of Wolverhampton, UK)  
**Laura Hasler** (University of Wolverhampton, UK)  
**Diana Inkpen** (University of Ottawa, Canada)  
**Natalia Konstantinova** (University of Wolverhampton, UK)  
**Sandra Kübler** (Indiana University, USA)  
**Elina Lagoudaki** (Imperial College, UK)  
**Wolfgang Maier** (University of Tübingen, Germany)  
**Dalila Mekhaldi** (University of Wolverhampton, UK)  
**Preslav Nakov** (National University of Singapore, Singapore)  
**Ivelina Nikolova** (Bulgarian Academy of Sciences, Bulgaria)  
**Constantin Orăsan** (University of Wolverhampton, UK)  
**Petya Osenova** (Bulgarian Academy of Sciences, Bulgaria)  
**Elena Paskaleva** (Bulgarian Academy of Sciences, Bulgaria)  
**Viktor Pekar** (Oxford University Press, UK)  
**Jelena Prokić** (University of Groningen, The Netherlands)  
**Georg Rehm** (vionto GmbH, Berlin, Germany)  
**Lucia Specia** (University of Wolverhampton, UK)  
**Irina Temnikova** (University of Wolverhampton, UK)  
**Cristina Toledo** (University of Málaga, Spain)  
**Pinar Wennerberg** (Siemens AG, Germany)



## PREFACE

The Recent Advances in Natural Language Processing (RANLP) conference, already in its seventh year and ranked among the most influential NLP conferences, has always been a meeting venue for scientists coming from all over the world. In the last few years there has been increasing interest from students and young researchers from the NLP field. For this reason, for the first time in RANLP history, a competitive Student Workshop featuring peer reviewing and printed proceedings is held in conjunction with the International conference RANLP 2009.

The aim of the workshop is to provide an excellent opportunity for students at all levels (Bachelor, Master, and PhD) to present their work in progress or completed projects to an international research audience and receive feedback from senior researchers. We have received 29 high quality submissions, among which only 3 papers have been accepted as regular oral papers, and 13 as posters. Each submission has been reviewed by at least 2 reviewers, who are experts in their field, in order to supply detailed and helpful comments. The papers' topics cover a broad selection of research areas, such as:

- Anaphora Resolution;
- Information Extraction;
- Information Retrieval;
- Machine Translation;
- Multiple-Choice Questions Generation;
- Opinion Mining;
- Parsing;
- Part-of-Speech Tagging;
- Temporal Processing;
- Text Segmentation;
- Text Summarization;
- Textual Entailment;
- Word Sense Disambiguation.

We are also glad to admit that our authors comprise a very international group with students coming from: Argentina, Bulgaria, China, France, India, Iran, Italy, Portugal, Romania, Spain, Tunisia, Turkey, United Kingdom and United States.

We would like to thank the authors for submitting their papers and the members of the Programme Committee for their efforts to provide exhaustive reviews.

We are especially grateful to the RANLP Chairs Prof. Galia Angelova and Prof. Ruslan Mitkov, and to Dr. Constantin Orăsan for their indispensable support and encouragement during the Workshop organisation.

We hope the participants will receive invaluable feedback about their work. Enjoy the Workshop!

Irina Temnikova, Ivelina Nikolova and Natalia Konstantinova  
Organisers of the Student Workshop, held in conjunction with  
The International Conference RANLP-09



## Table of Contents

<i>Effect of Minimal Semantics on Dependency Parsing</i> Bharat Ram AMBATI, Pujitha GADE, Chaitanya GSK and Samar HUSAIN.....	1
<i>Mink: An Incremental Data-Driven Dependency Parser with Integrated Conversion to Semantics</i> Rachael CANTRELL.....	6
<i>A Study of Machine Learning Algorithms for Recognizing Textual Entailment</i> Julio Javier CASTILLO.....	12
<i>LOGICON: A System for Extracting Semantic Structure using Partial Parsing</i> Kais DUKES.....	18
<i>Framework for using a Natural Language Approach to Object Identification</i> Mosa Emhamed ELBENDAK.....	23
<i>Improving the Output from Software that Generates Multiple Choice Question (MCQ) Test Items Automatically using Controlled Rhetorical Structure Theory</i> Robert Michael FOSTER.....	29
<i>Normalized Accessor Variety Combined with Conditional Random Fields in Chinese Word Segmentation</i> Saike HE, Taozheng ZHANG, Xue BAI, Xiaojie WANG and Yuan DONG.....	35
<i>Event Ordering. Temporal Annotation on Top of the BulTreeBank</i> Laska LASKOVA.....	40
<i>Ambiguous Arabic Words Disambiguation: The Results</i> Laroussi MERHBEN, Anis ZOUAGHI and Mounir ZRIGUI.....	45
<i>ZAC.PB: An Annotated Corpus for Zero Anaphora Resolution in Portuguese</i> Simone PEREIRA.....	53
<i>A Rule-Based Approach to the Identification of Spanish Zero Pronouns</i> Luz RELLO and Iustina ILISEI.....	60
<i>Exploring Context Variation and Lexicon Coverage in Projection-based Approach for Term Translation</i> Raphaël RUBINO.....	66
<i>An Evaluation of Output Quality of Machine Translation Program</i> Mitra SHAHABI.....	71
<i>A Two-stage Bootstrapping Algorithm for Relation Extraction</i> Ang SUN.....	76
<i>Context Driven XML Retrieval</i> Aneliya TINCHEVA.....	83
<i>Hierarchical Discourse Parsing Based on Similarity Metrics</i> Ravikiran VADLAPUDI, Poornima MALEPATI and Suman YELATI.....	89





# Workshop Programme

14 September 2009

Poster presentations - 17:15–18:30

*Normalized Accessor Variety Combined with Conditional Random Fields in Chinese Word Segmentation*

Saike HE, Taozheng ZHANG, Xue BAI, Xiaojie WANG and Yuan DONG

*LOGICON: A System for Extracting Semantic Structure using Partial Parsing*

Kais DUKES

*An Evaluation of Output Quality of Machine Translation Program*

Mitra SHAHABI

*Ambiguous Arabic Words Disambiguation: The Results*

Laroussi MERHBEN, Anis ZOUAGHI and Mounir ZRIGUI

*Hierarchical Discourse Parsing Based on Similarity Metrics*

Ravikiran VADLAPUDI, Poornima MALEPATI and Suman YELATI

*Effect of Minimal Semantics on Dependency Parsing*

Bharat Ram AMBATI, Pujitha GADE, Chaitanya GSK and Samar HUSAIN

*A Study of Machine Learning Algorithms for Recognizing Textual Entailment*

Julio Javier CASTILLO

*Exploring Context Variation and Lexicon Coverage in Projection-Based Approach for Term Translation*

Raphaël RUBINO

*ZAC.PB: An Annotated Corpus for Zero Anaphora Resolution in Portuguese*

Simone PEREIRA

*A Rule-Based Approach to the Identification of Spanish Zero Pronouns*

Luz RELLO and Iustina ILISEI

*Context Driven XML Retrieval*

Aneliya TINCHEVA

*Framework for using a Natural Language Approach to Object Identification*

Mosa Emhamed ELBENDAK

*Improving the Output from Software that Generates Multiple Choice Question (MCQ) Test Items Automatically using Controlled Rhetorical Structure Theory*

Robert Michael FOSTER

**15 September 2009**

**Oral Presentations**

18:15–18:45 *Event Ordering. Temporal Annotation on Top of the BulTreeBank*  
Laska LASKOVA

17:15–17:45 *A Two-stage Bootstrapping Algorithm for Relation Extraction*  
Ang SUN

17:45–18:15 *Mink: An Incremental Data-Driven Dependency Parser with Integrated Conversion to Semantics*  
Rachael CANTRELL